

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

***IDENTIFICAÇÃO DE RNAS QUIMÉRICOS E RNAS CIRCULARES EM DADOS DE
TRANSCRIPTOMA DE ORGANOIDES CEREBRAIS E SEUS IMPACTOS NO
NEURODESENVOLVIMENTO***

Aluno: Bruno Siegel Guerra, M.Sc.

Orientador: Roberto H. Herai, Ph.D

Curitiba, julho/2024

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

***IDENTIFICAÇÃO DE RNAS QUIMÉRICOS E RNAS CIRCULARES EM DADOS DE
TRANSCRIPTOMA DE ORGANOIDES CEREBRAIS E SEUS IMPACTOS NO
NEURODESENVOLVIMENTO***

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciências da Saúde como requisito parcial para obtenção do título de Doutor em Ciências da Saúde.

Orientador: Prof. Dr. Roberto H. Herai

Curitiba, julho/2024



Pontifícia Universidade Católica do Paraná
Escola de Medicina e Ciências da Vida
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 **05 dias do mês de julho de 2024 às 13:30**, realizou-se a sessão aberta de Defesa de Tese “**IDENTIFICAÇÃO DE RNAs QUIMÉRICOS E RNAs CIRCULARES EM DADOS DE TRANSCRIPTOMA DE ORGANOIDES CEREBRAIS E SEUS IMPACTOS NO NEURODESENVOLVIMENTO**” apresentado por **Bruno Siegel Guerra** 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. Edson Emilio Scalabrin – PUCPR	
Prof. Dr. Eduardo Del Bosco Brunetti Cunha - Laboratório VitaGene	
Profa. Dra. Janaina Sena Boyce - University of California San Diego	
Profa. Dra. Rafaela Rosa-Ribeiro – UNICENTRO	

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. Edson Emilio Scalabrin	Conceito: Aprovado
Prof. Dr. Eduardo Del Bosco Brunetti Cunha	Conceito: Aprovado
Profa. Dra. Janaina Sena Boyce	Conceito: Aprovado
Profa. Dra. Rafaela Rosa-Ribeiro	Conceito: Aprovado

Parecer Final: Aprovado

Observações da Banca Examinadora: Realizar as correções indicadas pela banca. Todas as alterações devem ser verificadas e analisadas pelo orientador do projeto, sem a necessidade de reavaliação por parte da banca de avaliação.

Prof. Dr. Roberto Hirochi Herai
 Presidente da Banca Examinadora

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

AGRADECIMENTOS

Primeiramente, agradeço a todos os docentes e funcionários do Programa de Pós Graduação em Ciências da Saúde (PPGCS) da Escola de Medicina da PUCPR. Local onde passei os últimos anos de minha vida e que foram essências na minha formação como profissional em ciências da saúde.

Agradeço a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) por ser a principal agencia financiadora de minha formação acadêmica e profissional.

Ao pesquisador Alysson Muotri, coordenador do Laboratório de Células Tronco da Universidade da California San Diego (UCSD, USA), pela colaboração e recepção durante meu doutorado sanduiche em seu laboratório. Em especial a todo o corpo profissional que conheci durante esse período: Janaína Sena de Souza, Wirla Pontes, Sandra e Ronaldo. Vocês foram essenciais.

Agraço aos meus colegas do Heraí Lab's e do Laboratório Experimental Multiusuário (LEM) da PUCPR pelos momentos de descontração, café, discussões e aprendizados.

Ao meu orientador e referencia acadêmica, Prof. Dr. Roberto Heraí, sou muito grato por todas as oportunidades de crescimento que o senhor me proporcionou. Seguirei com a melhor lição que aprendi com o senhor: a coragem para pensar diferente.

A minha família, que mesmo não entendendo muito deste tema tão complexo, me ouviu com carinho, cuidado e amor, isso foi essencial. Obrigado pelo apoio incondicional e serem meu chão para crescimento pessoal e profissional. Amo vocês.

Por fim, agradeço a minha querida noiva, Carolina, que com sua parceria e cuidado, me deram as forças necessárias para concluir essa etapa da minha vida e abrir novas oportunidades. Estamos apenas começando.

DEDICATÓRIA

Às crianças e famílias que enfrentam corajosamente os desafios das condições neurológicas do neurodesenvolvimento, espero que este trabalho contribua para um futuro melhor todos vocês.

FOLHA INFORMATIVA

A tese foi elaborada seguindo a normatização estabelecida pela ABNT. Este trabalho foi realizado nas instalações do Laboratório Experimental Multiusuário (LEM) do Programa de Pós-Graduação em Ciências da Saúde da PUCPR em colaboração com o Prof. Dr. Alysson Muotri e sua equipe na Universidade da Califórnia San Diego (UCSD). Declaramos que não há conflito de interesse na realização da presente tese.

RESUMO

Introdução: O desenvolvimento cerebral ocorre desde os primeiros dias de vida já na vida intrauterina e pode ser afetado por desequilíbrios genotípicos ou fenotípicos, resultando em prejuízos para as capacidades sensoriais, cognitivas, de motricidade ou de comunicação do indivíduo. O uso de tecnologias capazes de mimetizar o neurodesenvolvimento são essenciais para o estudo e compreensão do complexo contexto genético em que as condições se desenvolvem. Dois genes desempenham um papel chave no tecido cerebral e mutações podem impactar o funcionamento desse órgão: o gene *MECP2* na síndrome de Rett e o gene *NOVA1* regulador de splicing alternativo. Nesse contexto, a expressão de moléculas de RNA não canônicas ainda precisa ser compreendida. Como é o caso das fusões de genes a nível de RNA com o *trans-splicing*, e o de RNAs circulares. O uso de tecnologias capazes de reproduzir o neurodesenvolvimento, aliado às formas de detectar a presença de moléculas de RNA torna-se um estudo com potencial alto de impacto na vida de pessoas com essas condições. **Metodologia:** Células tronco pluripotente induzidas (iPSC) de pacientes com mutações em *MECP2* e *NOVA1* foram cultivadas *in-vitro* reproduzindo o desenvolvimento cerebral através de organoides corticais. O DNA e RNA desses organoides foram processados para detecção de casos de *trans-splicing* e RNA circulares. Novos métodos foram desenvolvidos para auxiliar esse processo. Validação dos candidatos detectados foi realizado por PCR e Sequenciamento de Sanger. Por fim, o seu impacto no neurodesenvolvimento foi correlacionado com informações presentes na literatura. **Resultados:** O cultivo dos organoides corticais foi capaz de reproduzir o neurodesenvolvimento esperado para o tecido cortical. Identificamos 308.551 casos de *trans-splicing* e 403 casos de RNA circular em todas as amostras testadas. Desenvolvemos uma técnica capaz de filtrar a maioria dos falsos-positivos presentes nos dados e validamos 20

candidatos por técnicas de PCR e 8 deles por sequenciamento Sanger. Identificamos que 5 casos de fusões de genes que tem relação teórica com o neurodesenvolvimento. E identificamos o RNA circular *MAP2* que tem papel essencial na formação de microtúbulos da neurogênese.

Conclusão: A presença dessas moléculas de RNA parece ter papel importante no desenvolvimento cerebral. Apesar de não encontrar uma correlação clara entre os dois genes alvo, identificamos diversos candidatos que podem ajudar a elucidar o complexo processo do neurodesenvolvimento.

Palavras-chave: Organoides corticais; Neurodesenvolvimento; *trans-splicing*; CircRNA; transcriptoma

ABSTRACT AND KEYWORDS

Introduction: Brain development occurs from the early days of life, even during intrauterine life, and can be affected by genotypic or phenotypic imbalances, resulting in impairments to an individual's sensory, cognitive, motor, or communication skills. The use of technologies capable of mimicking neurodevelopment is essential for the study and understanding of the complex genetic context in which conditions develop. Two genes play a key role in brain tissue, and mutations can impact the functioning of this organ: the *MECP2* gene in Rett syndrome and the *NOVA1* gene, which regulates alternative splicing. In this context, the expression of novel RNA molecules still needs to be understood, such as gene fusions at the RNA level with trans-splicing and circular RNAs. The use of technologies capable of reproducing neurodevelopment, along with ways to detect the presence of RNA molecules, becomes a study with high potential impact on the lives of people with these conditions. **Methodology:** iPSC cells from patients with mutations in *MECP2* and *NOVA1* were cultured *in-vitro*, reproducing brain development through cortical organoids. The DNA and RNA of these organoids were then processed to detect cases of trans-splicing and circular RNA. New methods were developed to assist in this process. Validation of detected candidates was performed by PCR and Sanger sequencing. Finally, their impact on neurodevelopment was correlated with information present in literature. **Results:** The development of cortical organoids was able to reproduce the expected neurodevelopment for cortical tissue. We identified 308,551 cases of trans-splicing and 403 cases of circular RNA in all tested samples. We developed a technique capable of filtering out most of the false positives present in the data and validated 20 candidates by PCR techniques and 8 of them by Sanger sequencing. We identified 5 cases of gene fusions that have a theoretical relationship with neurodevelopment. And we identified the circular RNA MAP2, which plays an essential role in

the formation of neurogenesis microtubules. **Conclusion:** The presence of these RNA molecules appears to play an important role in brain development. Although we did not find a clear correlation between the two target genes, we identified several candidates that could help elucidate the complex process of neurodevelopment.

Keywords: Cortical organoids; Neurodevelopment; trans-splicing; CircRNA; transcriptome

RESUMO POPULAR

O desenvolvimento do cérebro começa nos primeiros dias de vida, ainda no útero, e pode ser afetado por alterações genéticas, causando problemas sensoriais, cognitivos, motores ou de comunicação. Para entender melhor como essas alterações impactam o cérebro, usamos tecnologias que se aproximam o desenvolvimento cerebral. Dois genes importantes para o funcionamento do cérebro, *MECP2* e *NOVA1*, foram estudados por sua relação com condições como a síndrome de Rett. Neste estudo, foram utilizadas células de pacientes com mutações nesses genes para criar mini cérebros em laboratório, chamados de organoides corticais. Essas estruturas permitem simular o desenvolvimento do cérebro humano e estudar moléculas específicas de RNA. Foram detectados milhares de casos de fusões de genes e RNAs circulares, que são tipos especiais de RNA, nos organoides corticais. Desenvolvemos uma nova técnica para filtrar resultados falsos e confirmaram alguns desses RNAs usando métodos avançados de laboratório. Entre os achados, destacam-se cinco fusões de genes possivelmente relacionadas ao desenvolvimento cerebral e um RNA circular, chamado *MAP2*, que é crucial para a formação de estruturas importantes nos neurônios. A descoberta desses RNAs pode ajudar a entender melhor o desenvolvimento do cérebro e as condições que afetam esse processo. Esses resultados são um passo importante para futuros estudos que podem levar a novas terapias e melhorias na qualidade de vida para pessoas com distúrbios do desenvolvimento cerebral.

LISTA DE FIGURAS

Figura 1 – Posição do gene MECP2 no cromossomo X, estrutura dos domínios presentes e da posição da mutação QX83.....	26
Figura 2 – Posição do gene NOVA1 no cromossomo 14, estrutura dos domínios presentes e da posição da mutação 200 de substituição.....	29
Figura 3 – Resumo esquemático do processo de biogênese de Trans-splicing.....	34
Figura 4 – Resumo esquemático do processo de cultivo, diferenciação e maturação dos organoides desenvolvidos a partir de células de hiPSC.	43
Figura 5 – Resumo esquemático do processo de criação de um dataset positivo para avaliação das ferramentas de detecção de trans-splicing.....	49
Figura 6 - Resumo esquemático do processo de elaboração dos primers para validação in-vitro dos candidatos.....	51
Figura 7 - Fluxograma de seleção para RSL para ferramentas computacionais de detecção de Trans-splicing.	54
Figura 8 - Estratégias de detecção de TS em relação ao método de alimento usado.	56
Figura 9 – Estratégia de detecção de TS em relação a qual alinhador foi construído o método de detecção.	57
Figura 10 - Distribuição de filtros presentes nas ferramentas para detecção de Trans-splicing.	61
Figura 11 – Gráficos de barras indicando o número de casos positivos de cada ferramenta para cada tipo de teste desenhado.....	64

Figura 12 – Gráfico de barra de erro comparando as ferramentas em relação as medidas do índice de Jaccard.	65
Figura 13 - Processo de cultivo e diferenciação dos organoides corticais, separado em quatro momentos para os organoides de NOVA1.	67
Figura 14 – Fotos do marcador celular para a expressão gene NES nos organoides corticais NOVA1.....	69
Figura 15 - Fotos do marcador celular para a expressão gene MAP2 nos organoides corticais NOVA1.....	70
Figura 16 - Fotos do marcador celular para a expressão de DAPI nos organoides corticais NOVA1.....	71
Figura 17 - Correlação entre candidatos em cada cromossomo e número de gene nos cromossomos segundo o NCBI.	75
Figura 18 - Características dos candidatos a Trans-splicing em relação a capacidade de tradução e distancia genômica entre os pares.....	77
Figura 19 – Resumo em forma gráfica do processo de filtragem in-silico para validação.	79
Figura 20 - Gráfico de barras dos níveis de expressão dos candidatos por RT-PCR.....	83

LISTA DE TABELAS

Tabela 1 - Resumo das amostras por gene, acrónimo e descrição	41
Tabela 2 - Resumo das amostras na detecção de Trans-splicing e CircRNA.....	73
Tabela 3 - Resumo dos candidatos validados	81
Tabela 4 - Características gerais das fusões intracromossomais	85
Tabela 5 - Características gerais das fusões intercromossomais.	90
Tabela 6 - Características gerais dos RNA circulares.	95

LISTA DE SIGLAS

RTT – Síndrome de Rett

TEA – Transtorno do Espectro Autista

UCSD – Universidade da California San Diego

MF – Mapping First

AF – Assembling First

RSL - Revisão sistemática de Literatura

TS – *Trans-Splicing*

CircRNA – *Circular RNA*

lncRNA – Long non-coding RNA

KdV - Koolen-de Vries

SUMÁRIO

1	INTRODUÇÃO	17
2	JUSTIFICATIVA.....	21
3	REVISÃO BIBLIOGRAFICA.....	22
3.1	Modelos e ferramentas para o estudo do neurodesenvolvimento.....	22
3.2	O gene <i>MECP2</i>	25
3.3	O gene <i>NOVA1</i>	28
3.4	Transcriptoma.....	30
3.5	Fusão de genes a nível de RNA.....	32
3.6	RNAs Circulares.....	37
4	OBJETIVOS	39
4.1	Objetivos Específicos.....	39
5	MATERIAL E MÉTODOS	40
5.1	Cultivo, diferenciação e caracterização dos organoides.....	40
5.2	Detecção e filtragem <i>in-silico</i>	44
5.3	Validação <i>in-vitro</i>	50
5.4	Análise funcional dos candidatos em dados públicos	52
6	RESULTADOS E DISCUSSÃO	54
6.1	Seleção de Ferramentas via RSL.....	54
6.2	Cultivo, diferenciação e caracterização dos organoides.....	66
6.3	Detecção e filtragem de candidatos de TS e CircRNA	72
6.4	Validação <i>in-vitro</i> dos candidatos	80

6.5	Caracterização dos candidatos validados	84
6.5.1	Fusões do mesmo cromossomo.....	84
6.5.2	Fusões de cromossomos diferentes.....	90
6.5.3	RNAs Circulares	94
7	CONCLUSÃO	99
8	CONSIDERAÇÕES FINAIS	101
	REFERÊNCIAS	102
	APÊNDICES	109

1 INTRODUÇÃO

A embriogênese do cérebro é um processo fascinante que ocorre durante o desenvolvimento embrionário, começando nos estágios iniciais da gestação. Esse intrincado processo envolve uma sequência coordenada de eventos que resultam na formação do sistema nervoso central. Inicialmente, o tubo neural é estabelecido, a partir do qual se desenvolverão as estruturas cerebrais fundamentais. Ao longo das semanas seguintes, ocorrem proliferação celular, migração neuronal, diferenciação e formação de conexões sinápticas. A complexidade da embriogênese cerebral é evidenciada pela delicada regulação genética e ambiental que determina a formação de regiões específicas do cérebro, cada uma com funções distintas. Interrupções nesse processo, seja por fatores genéticos, ambientais ou outros, podem levar a malformações cerebrais, destacando a importância crucial desse período para o desenvolvimento saudável do sistema nervoso e, por conseguinte, para as funções cognitivas e sensoriais ao longo da vida (MACHADO; HAERTEL MACHADO, 2013).

Nesse contexto, a reprogramação de células de pacientes com desordens neurológicas, como a síndrome de Rett (RTT), destaca-se como um modelo de análise impactante no estudo do neurodesenvolvimento. A RTT, resultante de mutações no cromossomo X no *locus* do gene *MECP2*, evidencia a importância crítica dessa proteína como fator de transcrição essencial para o desenvolvimento saudável e a manutenção do sistema nervoso central (MARCHETTO et al., 2010b; NA et al., 2013). Além disso, o modelo associado ao gene *NOVA1* oferece outra perspectiva valiosa, revelando alterações no cenário do desenvolvimento cerebral. O *NOVA1* desempenha um papel crucial como fator de transcrição para elementos do genoma, especialmente RNAs, e sua associação a condições como esquizofrenia e espectro autista (TEA)

destaca a complexidade das interações genéticas que podem influenciar o desenvolvimento neurobiológico (TRAN et al., 2018).

Os modelos mencionados desempenham um papel chave na manipulação da produção de moléculas de RNA e, por conseguinte, de proteínas, oferecendo insights valiosos sobre os mecanismos subjacentes a desordens neurológicas. No entanto, é importante ressaltar que a complexidade do transcriptoma cerebral vai além da síntese proteica convencional. Diversas classes de RNA ainda não foram completamente caracterizadas, e algumas desempenham funções essenciais que extrapolam a simples síntese de proteínas. Isso é evidente em transcritos que exibem características como fusões de gene a nível de RNA (*trans-splicing*) e os RNA de conformação circular (circRNA). Essas formas de RNA emergem como componentes fundamentais na regulação fina da expressão gênica e podem desempenhar papéis cruciais no desenvolvimento cerebral e na patogênese de condições neurológicas (TANG et al., 2019a; CHEN et al., 2020).

As moléculas de RNA resultantes do *trans-splicing* apresentam uma intrigante capacidade de gerar proteínas quiméricas que desempenham diversas funções no genoma, incluindo o controle preciso da expressão gênica. Entretanto, sua presença fora do contexto de rearranjos a nível de DNA e seu papel específico no desenvolvimento cerebral continuam a exigir investigações mais detalhadas (BABICEANU et al., 2016). A compreensão desses mecanismos intrincados pode fornecer *insights* cruciais sobre a complexidade regulatória do transcriptoma cerebral, contribuindo para uma visão mais abrangente das bases moleculares do neurodesenvolvimento.

No que diz respeito aos circRNA, sua abundância em pacientes com transtorno do espectro autista (TEA) suscita questões intrigantes sobre seu papel na patogênese dessa condição. Essas moléculas circulares foram identificadas como reguladores significativos de funções genômicas, atuando como "esponjas" para microRNAs ou proteínas, efetivamente sequestrando essas moléculas do contexto celular e modificando a expressão gênica (GUERRA et al., 2020).

Esses exemplos ilustram a intrincada complexidade do sistema no qual essas moléculas de RNA estão imersas, destacando que cada nova descoberta representa um avanço significativo no entendimento do desenvolvimento cerebral. A utilização desses modelos, aliada às técnicas biotecnológicas mencionadas anteriormente, permite não apenas a compreensão aprofundada das bases moleculares das desordens neurológicas, mas também oferece insights valiosos para potenciais estratégias terapêuticas que visam corrigir alterações no neurodesenvolvimento intrauterino. Essa abordagem integrativa, que combina modelos computacionais da bioinformática com ferramentas biotecnológicas avançadas, representa um passo significativo em direção ao entendimento e à intervenção eficaz em distúrbios neurológicos complexos. Embora a dificuldade computacional da bioinformática para abordar estas questões e a própria complexidade na análise do neurodesenvolvimento humano sejam desafios inegáveis, é possível conceber experimentos e métodos que ofereçam suporte à investigação do neurodesenvolvimento.

Nesse contexto, surge o presente doutorado, que busca a idealização de experimentos utilizando o melhor modelo *in-vitro* disponível para simular o desenvolvimento cerebral. A abordagem inclui a exploração do contexto genético por meio de candidatos de dois tipos de RNA não canônicos. Este enfoque visa agregar novos conhecimentos à ciência, com o objetivo principal de aprimorar a qualidade de vida das pessoas afetadas por desordens neurológicas. Ao adotar essa abordagem inovadora, o projeto visa preencher lacunas na compreensão do papel

dessas moléculas não canônicas no desenvolvimento cerebral, proporcionando *insights* valiosos que podem direcionar estratégias terapêuticas mais eficazes no futuro. Essa iniciativa reflete o compromisso contínuo com a pesquisa e a busca por soluções que possam impactar positivamente na qualidade de vida de indivíduos que enfrentam desafios associados a distúrbios neurológicos.

2 JUSTIFICATIVA

A caracterização das moléculas de *trans-splicing* já foi estabelecida em diversos contextos, sobretudo em doenças vinculadas ao câncer e outras condições resultantes de rearranjos cromossômicos. No entanto, em tecidos e células normais, há uma lacuna de dados que evidenciem a ocorrência expandida do *trans-splicing*, especialmente no contexto de transtornos mentais. No que concerne aos circRNAs, embora tenham sido identificados e caracterizados em tecidos do neurodesenvolvimento, suas implicações específicas para condições patológicas ainda carecem de uma análise mais aprofundada.

Além disso, as informações relacionadas às mutações nos genes *MECP2* e *NOVA1*, suas origens e o potencial papel dessas mutações em condições específicas, como transtornos mentais, demandam uma abordagem mais detalhada. Considerando essas limitações, o presente trabalho foi concebido pelos autores deste projeto com o objetivo de preencher essas lacunas e contribuir significativamente para a elucidação do cenário do neurodesenvolvimento.

Ao abordar especificamente as mutações nos genes *MECP2* e *NOVA1*, o projeto busca oferecer uma compreensão mais profunda das implicações dessas mutações em condições neurológicas específicas. A pesquisa proposta visa preencher as lacunas existentes no conhecimento sobre o *trans-splicing* e circRNAs, especialmente no contexto dos transtornos mentais.

Assim, esse trabalho almeja contribuir de maneira substancial para o avanço do entendimento sobre o papel dessas moléculas no desenvolvimento cerebral, lançando luz sobre possíveis alvos terapêuticos e estratégias para melhorar a qualidade de vida daqueles afetados por desordens neurológicas.

3 REVISÃO BIBLIOGRAFICA

3.1 Modelos e ferramentas para o estudo do neurodesenvolvimento

Células-tronco de pluripotência induzida (iPSC) representam entidades somáticas que foram submetidas a um processo de reprogramação para alcançar um estado pluripotente, conferindo-lhes a habilidade intrínseca de diferenciar-se em diversos fenótipos celulares característicos do corpo humano. Este fenômeno de reprogramação é meteticulosamente desencadeado pela indução de expressão de genes específicos em células adultas, frequentemente por meio da aplicação de fatores de transcrição notáveis, exemplificados pelos fatores de Yamanaka (*MYC*, *OCT3/4*, *SOX2* e *KIF4*). A ação destes fatores de transcrição opera uma reconfiguração genômica nas células-alvo, notadamente em células adultas, como fibroblastos, empregando um sistema retroviral capaz de integrar seu material genético no genoma da célula hospedeira. O resultado alcançado culmina na instauração de um estado celular assemelhado ao das células-tronco embrionárias, ostentando a notável capacidade de diferenciação em um espectro diversificado de linhagens celulares distintas. Importante ressaltar que essa diferenciação ocorre enquanto preserva a singularidade da estrutura genética inerente ao organismo original, conforme evidenciado nos estudos de Takahashi et al. (2007).

A aplicação de hiPSCs, ou seja, iPSCs derivadas de fontes humanas, ostenta notável potencial nos domínios da medicina regenerativa e investigativa. Esse tipo de célula destaca-se pela sua capacidade de proliferação indefinida e pela habilidade intrínseca de diferenciar-se em diversas linhagens celulares dentro do organismo humano, conforme delineado por Mahla (2016). As hiPSCs podem ser direcionadas para a diferenciação em uma ampla variedade de tipos celulares cerebrais, incluindo neurônios, astrócitos e oligodendrócitos. Tais células

revelam-se instrumentais no estudo da morfogênese cerebral, investigação das conexões neuronais e compreensão das vias de sinalização subjacentes ao desenvolvimento do sistema nervoso. Ademais, sua aplicação é valiosa na análise das disparidades individuais na resposta cerebral a estímulos externos e na exploração da relação entre o desenvolvimento cerebral e a função cognitiva. Destaca-se, ainda, que essa tecnologia se configura como um modelo *in-vitro* para a investigação das bases moleculares de doenças durante o processo de diferenciação celular em pacientes portadores de diversas condições de interesse, conforme evidenciado por Marchetto et al. (2010a).

A investigação de patologias mediante o emprego de células iPS pode resultar em avanços significativos quando associada à técnica de edição genômica proporcionada pela tecnologia CRISPR/Cas9. Este sistema integrado, constituído por uma sequência de DNA originária de genomas bacterianos (CRISPR) e uma enzima guia (Cas9), demonstra competência na modificação de regiões específicas do DNA, culminando na inativação ou ativação de genes-chave vinculados a condições de interesse (HSU; LANDER; ZHANG, 2014). A título de exemplo, destaca-se a modulação de ganho ou perda de função em genes associados ao neurodesenvolvimento, exemplificado pela intervenção no gene *MECP2* no estudo da Síndrome de Rett (RTT) (HUONG LE et al., 2019). No entanto, para conceber um modelo *in-vitro* que recapitule o desenvolvimento de um órgão humano, é imperativo que tal sistema possua a capacidade de reproduzir fielmente os elementos constituintes da origem, incluindo regiões com estruturas especializadas e distintos tipos celulares. Para atender a essa exigência, emergem os organoides multicelulares, estruturas tridimensionais cultivadas *in-vitro* que emulam de maneira próxima a reprodução *in vivo* de órgãos de interesse.

Sob as condições e influências propícias ao desenvolvimento, os organoides se configuram como modelos promissores para replicar estruturas mais intrincadas do cérebro humano,

conforme indicado por Lancaster et al. (2013). Organoides corticais, em particular, emulam as principais zonas de origem de neurônios e progenitores durante o neurodesenvolvimento, incluindo a zona ventricular (VZ) e subventricular (SVZ). Estes apresentam uma organização inicial que se assemelha ao córtex cerebral humano, notadamente evidenciada pela presença proeminente de células gliais radiais externas (oRGs). Além disso, as células neuronais corticais nos organoides demonstram diferenciação em diversos subtipos, mantendo uma separação espacial moderada, enquanto zonas intermediárias, como a zona intermediária (IZ), também são recapituladas. A configuração tridimensional (3D) dessas estruturas representa uma vantagem crucial, permitindo a reprodução de tipos celulares mais complexos que requerem interações intrincadas em um contexto multicelular. Este cenário é evidenciado pela maturação neural e crescimento axonal, eventos que ocorrem nos modelos de organoides entre 2 a 3 meses de desenvolvimento (GIANDOMENICO; SUTCLIFFE; LANCASTER, 2021). Nesse contexto de interações complexas, torna-se evidente que mutações em genes-chave podem exercer impacto significativo no desenvolvimento do órgão em estudo, reverberando nos fenótipos e genótipos do sistema neurológico (PHATNANI; MANIATIS, 2015).

A integração sinérgica das três tecnologias fundamentais - hIPSc, CRISPR e organoides - proporciona um arcabouço robusto para a minuciosa investigação das diversas condições genéticas que impactam o desenvolvimento do organismo humano, notadamente no que tange ao desenvolvimento cerebral. Ilustrativamente, destacam-se os organoides desenvolvidos para abordar a RTT, os quais são concebidos a partir de hIPSc obtidas de pacientes acometidos por essa condição, nas quais mutações específicas são induzidas por meio da técnica CRISPR. Esse enfoque possibilita uma avaliação meticulosa do desenvolvimento da RTT, como corroborado pelos estudos de Marchetto et al. (2010a).

3.2 O gene *MECP2*

O gene *MECP2*, que significa *Methyl-CpG-binding protein 2* em inglês, é um gene localizado no braço longo do cromossomo X (Xq28). Composto por quatro éxons, esse gene codifica a proteína *MECP2*, que apresenta dois domínios proteicos distintos: o MBD (*methyl-cytosine-binding domain*) e o TRD (*transcriptional repression domain*) (Figura-1A). Esta proteína atua como um regulador epigenético, desempenhando funções essenciais na modulação da expressão gênica, na formação e manutenção de sinapses no cérebro e no desenvolvimento do sistema nervoso. Mutações no gene *MECP2* podem levar a distúrbios neurológicos graves, sendo mais conhecido por sua relação com o TEA e a RTT (LEWIS et al., 1992).

Caracterizada por sua estrutura intrinsecamente desordenada, ausente de conformações secundárias e terciárias, a proteína *MECP2* demonstra capacidade de interação com diversos parceiros fisiológicos, promovendo interações proteína-proteína e ajustando sua conformação a diferentes estruturas moleculares por meio de seus domínios. O domínio MBD assume uma configuração de cunha, associando-se a regiões metiladas de CpG nas fitas de DNA. Uma vez posicionado, o domínio TRD tem a capacidade de recrutar histona deacetilase (HDAC), resultando na remoção de grupos acetil e na modulação da estrutura da cromatina. Dessa forma, a proteína regula a transcrição de genes próximos a essas regiões metiladas, processo conhecido como repressão transcricional (LEWIS et al., 1992).

Além da função mencionada, o *MECP2* pode atuar como um ativador transcricional, como evidenciado pelo recrutamento de HDAC3 para a redução da transcrição de outros genes, possivelmente promovendo uma regulação positiva de outros elementos no genoma (CHAHROUR et al., 2008; TILLOTSON; BIRD, 2020). Adicionalmente, para além dessas

funções, o MECP2 também desempenha papéis na compactação da cromatina e no processamento de microRNAs (miRNAs) (TILLOTSON; BIRD, 2020).

O papel preponderante de MECP2 como um dos principais reguladores da transcrição no cérebro confere a este gene uma função essencial no funcionamento e desenvolvimento típico do sistema nervoso central (SNC) em mamíferos. Mutações neste gene constituem a principal causa etiológica para condições de neurodesenvolvimento, como a RTT (AMIR et al., 1999). Um exemplo é a mutação destacada na Figura-1B. A mutação, induzida por CRISPR-Cas9, ocasiona a substituição de um aminoácido (glutamina, Q) por um stop-códon (X) na posição 83 da sequência proteica. Esta mutação resulta em uma versão truncada da proteína MECP2, privada de seu principal domínio de interação com outros elementos do genoma (MBD). Indivíduos portadores da mutação QX83 tendem a manifestar uma forma mais grave dos sintomas de RTT em comparação com outras mutações existentes (TILLOTSON; BIRD, 2020).

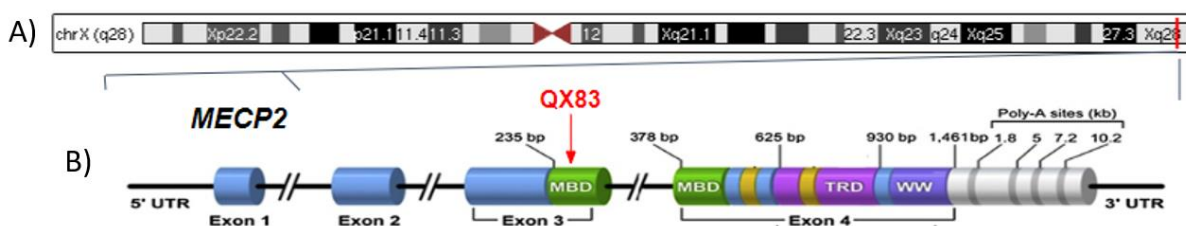


Figura 1 – Posição do gene *MECP2* no cromossomo X, estrutura dos domínios presentes e da posição da mutação QX83.

A) A posição do gene *MECP2* no genoma *Homo sapiens* está no braço longo do cromossomo X na banda de número 28 e apresenta um total de 12 éxons segundo a referência GRCh38.p14. **B)** Posição da mutação de deleção QX83 no domínio MBD do gene *MECP2*, que causa uma inativação desse gene na célula de origem. Fonte: o autor, 2024.

A RTT é uma condição que afeta predominantemente crianças do sexo feminino, com uma incidência estimada de 1 em cada 10.000 nascimentos, e tem como característica principal a deficiência intelectual (RICCERI; DE FILIPPIS; LAVIOLA, 2013). Crianças afetadas por esta síndrome apresentam um desenvolvimento inicial típico entre seis e dezoito meses de idade; entretanto, a partir desse ponto, observa-se uma regressão em diversas funções cognitivas e motoras. Características frequentemente identificadas incluem a perda de habilidades linguísticas, movimentos estereotipados das mãos, o desenvolvimento de microcefalia, convulsões, TEA e ataxia. Após a fase inicial de regressão, os sintomas tendem a estabilizar, permitindo que os pacientes sobrevivam até a idade adulta (AMIR et al., 1999).

O gene MECP2 desempenha um papel crucial no contexto da RTT, sendo que um gene específico afetado é o BDNF (*Brain-derived neurotrophic factor*). Este gene codifica a proteína BDNF, uma neurotrofina responsável por fatores de crescimento no sistema nervoso. Esses fatores são essenciais para a sobrevivência, maturação e formação de neurônios, e qualquer alteração na expressão desse gene desencadeia uma cascata de modificações no desenvolvimento neuronal (BINDER; SCHARFMAN, 2004; LI; POZZO-MILLER, 2014).

3.3 O gene *NOVA1*

O gene *NOVA1* (*neuro-oncological ventral antigen 1*) está localizado no cromossomo 14q12, com expressão preponderante no tecido cerebral, notadamente no córtex. Composto por 5 éxons, o gene codifica a proteína *NOVA1*, que apresenta três domínios proteicos de função similar: KH1, KH2 e KH3 (*K homology*) (Figura-2A). O domínio KH, presente em diversas proteínas, desempenha a função de se ligar a regiões específicas do RNA mensageiro (mRNA) conhecidas como RNA *binding protein* (RBP). *NOVA1* atua como um regulador de *splicing* alternativo em neurônios, permitindo que um único gene produza diferentes isoformas proteicas com potenciais funções distintas. Esta capacidade deriva da habilidade da proteína em se unir a sequências específicas presentes em pré-mRNA, conhecidas como motif YCAY. Este motif, uma sequência de quatro nucleotídeos, é reconhecido por proteínas de ligação a RNAs alvos específicos, resultando em alteração na transcrição (BUCKANOVICH; DARNELL, 1997; JENSEN et al., 2000). Estudos funcionais indicam que mais de 200 RNAs são regulados por *NOVA1* no cérebro de camundongos (ULE et al., 2003).

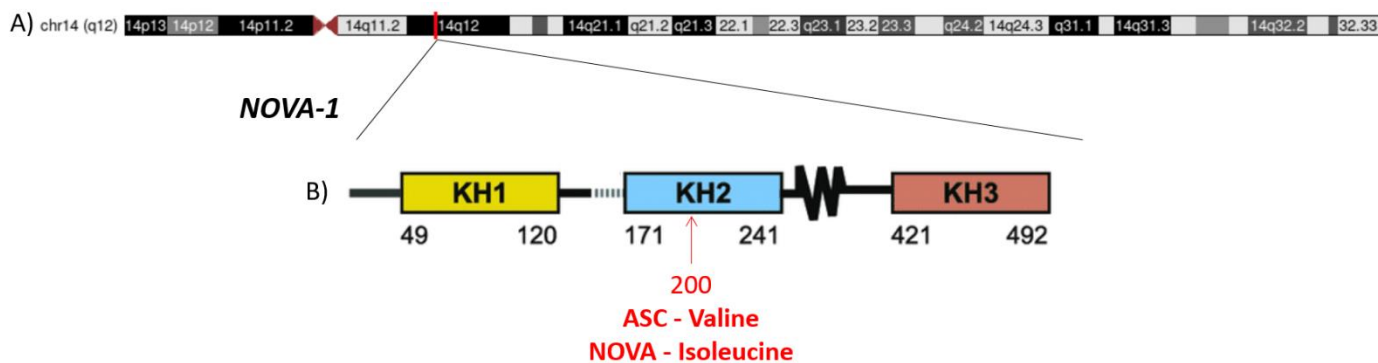


Figura 2 – Posição do gene *NOVA1* no cromossomo 14, estrutura dos domínios presentes e da posição da mutação 200 de substituição.

A) A posição do gene *NOVA1* no genoma *Homo sapiens* está no braço longo do cromossomo 14 na banda de número 12 e apresenta um total de 10 éxons segundo a referência GRCh38.p14. **B)** Posição da mutação de substituição no aminoácido 200, onde uma das linhagens apresenta uma isoleucina (NOVA) no lugar de uma valina (ASC). Fonte: o autor, 2024.

O gene *NOVA1* exibe uma característica que o torna um alvo de interesse no estudo da evolução da espécie humana. Ao longo do processo evolutivo, os neandertais apresentaram aproximadamente 200 genes codificadores de proteínas distintos dos encontrados nos humanos atuais. Este gene, previamente associado a condições como TEA e esquizofrenia, é capaz de codificar mais de 100 novas proteínas em neandertais, mesmo com apenas uma mutação pontual de valina para isoleucina (TRUJILLO et al., 2021a).

No âmbito do neurodesenvolvimento, estudos conduzidos por Trujillo e colaboradores (2021) revelaram que a introdução do gene *NOVA1* com a variante neandertal arcaica (isoleucina) em organoides do córtex cerebral resultou em um desenvolvimento mais lento e em uma superfície celular mais complexa em comparação com a variante moderna de valina na mesma posição (Figura-2B). Este achado destaca o impacto potencial de variações genéticas específicas, mesmo aquelas envolvendo apenas uma única mutação, na complexidade e desenvolvimento do córtex cerebral.

3.4 Transcriptoma

Conforme postulado pelo dogma central da biologia molecular, proposto por Francis Crick, o fluxo da informação genética ocorre por meio dos processos de replicação, transcrição e tradução. No âmbito da transcrição, esse procedimento refere-se à síntese de RNA a partir de uma fita de DNA. Posteriormente, esse RNA pode ser traduzido em uma proteína, desempenhando uma função específica no organismo, ou permanecer como uma molécula de RNA, cumprindo diversas outras funções. No contexto funcional dos RNAs, emerge uma área de pesquisa dedicada à sua compreensão, denominada transcriptômica. Essa disciplina visa examinar o conjunto completo de transcritos em uma célula, buscando quantificar ou qualificar essas moléculas em estágios específicos do desenvolvimento celular, condições de interesse ou amostras biológicas. As três principais funções do transcriptoma incluem catalogar todas as espécies de transcritos, abrangendo RNAs mensageiros, RNAs não codificantes e RNAs pequenos; determinar a estrutura transcricional de genes-alvo em termos de regiões codificantes, padrões de *splicing* e modificações pós-transcricionais; e, por fim, quantificar os níveis de expressão de cada transcrito durante condições ou períodos de interesse.

No estudo contemporâneo do transcriptoma, uma das técnicas aplicadas é a RNA-seq, uma tecnologia de *deep-sequencing*, na qual são gerados múltiplos fragmentos de uma mesma região. Esse procedimento é viabilizado por sequenciadores de nova geração (NGS), ou segunda geração, que executam vários processos de sequenciamento simultaneamente. Ao término desse processo, é gerado um volume expressivo de fragmentos, permitindo que os pesquisadores identifiquem até mesmo os eventos mais raros dentro do contexto genético.

A preparação de uma biblioteca de RNA-seq envolve passos essenciais para garantir a qualidade e reprodutibilidade do experimento. Inicialmente, o RNA é isolado da amostra de

interesse e tratado com DNase para reduzir a presença de DNA. Aproveitando as características moleculares dos RNAs mensageiros, que apresentam uma cauda 3' poly(A), é possível capturar esses RNAs utilizando uma estrutura contendo uma base complementar com Poly(T). Uma etapa de purificação e remoção de RNA ribossomal também é realizada, uma vez que o rRNA compreende 90% do conteúdo de RNA em uma célula. O RNA é extraído e, em seguida, transcrito em cDNA devido à maior estabilidade que a fita dupla proporciona, possibilitando a amplificação pela DNA polimerase. Por fim, os fragmentos são selecionados com base no tamanho de interesse para a análise do transcriptoma e, em seguida, submetidos a uma análise computacional.

O tratamento computacional dos fragmentos de RNA-seq pode seguir dois principais caminhos: mapeamento ou montagem. Nos processamentos baseados em mapeamento, o volume de fragmentos é comparado com um genoma ou transcriptoma de referência por meio de alinhamento de sequências, permitindo a identificação da origem genômica dos genes de interesse. Já nos processamentos baseados em montagem, ocorre a reconstrução do fragmento completo, utilizando os trechos de fragmentos menores que se sobrepõem.

Cada técnica apresenta suas vantagens e desvantagens no contexto do processamento de RNA-seq. No caso do mapeamento, é um processo mais rápido e capaz de quantificar a expressão gênica usando uma referência estabelecida. No entanto, se o estudo estiver relacionado a um organismo com uma referência ainda não determinada, o mapeamento não conseguirá identificar a origem dos elementos com precisão. Nesse caso, a aplicação da técnica de montagem é necessária, embora seja mais custosa computacionalmente, permitindo iniciar o processo de reconstrução de um organismo ainda não caracterizado.

3.5 Fusão de genes a nível de RNA

Transcritos quiméricos a nível de RNA constituem uma classe rara de moléculas geradas por fusão de genes no DNA, seja por meio de rearranjos cromossômicos ou, ainda, a nível de RNA, através do processo de fusão de dois ou mais transcritos de RNA, conhecido como *trans-splicing* (HERAI; YAMAGISHI, 2009). RNAs quiméricos originados pelo mecanismo de *trans-splicing* envolvem a união de éxons ou íntrons de duas moléculas precursoras de RNA mensageiro durante o processo de *splicing*, resultando em uma única molécula de mRNA formada a partir de múltiplos RNAs (Figura-3) (LI et al., 2009). O *trans-splicing* é classificado em duas categorias, dependendo da origem dos éxons e íntrons nos pré-mRNAs que formam a fusão. A primeira é a intracromossomal, ocorrendo quando esses elementos se originam do mesmo cromossomo, e a segunda é a intercromossomal, quando os fragmentos provêm de cromossomos de regiões distintas. As moléculas de RNA resultantes do *trans-splicing* podem dar origem a uma proteína quimérica ou desempenhar outras funções no genoma, como o controle da expressão gênica.

O fenômeno do *trans-splicing* foi inicialmente caracterizado em contextos nos quais rearranjos cromossômicos são frequentes, como observado em tecidos cancerosos. Um exemplo é a fusão intercromossomal JAZF1-SUZ12, que ocorre entre os cromossomos 7 e 17, identificada em casos de sarcoma do endométrio (HRZENJAK, 2016). Outro caso relevante é a fusão intercromossomal BCR-ABL, encontrada em leucemia, originada de uma translocação entre os cromossomos 9 e 22 (SALESSE; VERFAILLIE, 2002).

Em 2016, Babiceanu e colaboradores identificaram 238 casos de *trans-splicing* em mais de 300 bibliotecas de RNA-seq de células não cancerígenas. Esses resultados destacam que a presença de RNAs quiméricos não se restringe a situações em que rearranjos cromossômicos estão presentes (BABICEANU et al., 2016). Um exemplo adicional é o transcrito intracromossomal DU4L-BCAP29, originado do cromossomo 7, identificado no contexto do neurodesenvolvimento, onde promove a diferenciação neuronal em crianças. Isso demonstra que esses elementos também estão presentes em um contexto de neurodesenvolvimento (TANG et al., 2019b).

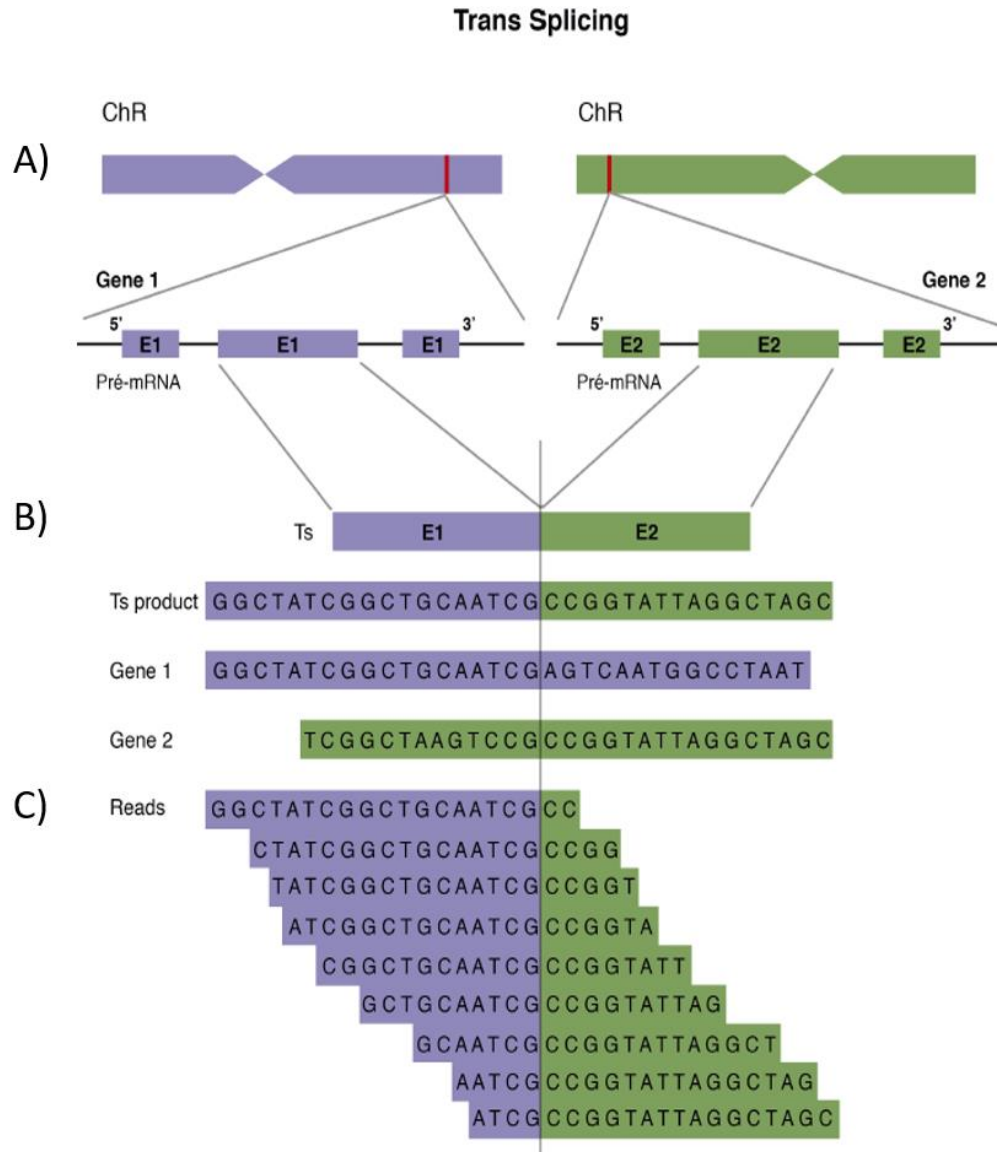


Figura 3 – Resumo esquemático do processo de biogênese de *Trans-splicing*.

A) Visão geral do processo de *trans-splicing* e como pode ser visualizado após o processo de detecção. Dois genes de regiões genômicas distintas, ao passar pelo processo de transcrição, formam seus pré-mRNAs contendo seus éxons e íntrons, representado pela cor verde e roxa. Por algum mecanismo desconhecido, éxons desses dois genes são unidos para formar uma nova molécula química contendo informação de ambas as estruturas originárias. **B)** O produto resultante desse processo pode ser identificado por sequências de RNA-seq que se alinham às duas regiões simultaneamente. **C)** As reads desse processo também apresentam uma característica típica dessa fusão, que é a representação dessas reads através de uma escada, mostrando diferentes reads, de diferentes tamanhos, oriundas do mesmo processo de *trans-splicing*. Fonte: o autor, 2024.

A qualificação e quantificação da presença de *trans-splicing* em amostras biológicas são facilitadas pela tecnologia de RNA-seq. Um protocolo de identificação pode ser sintetizado em três etapas principais: i) alinhamento das sequências; ii) detecção e validação *in silico*; e iii) filtragem dos potenciais candidatos e falsos positivos.

O processo de estudo de um transcriptoma inicia-se com o alinhamento de fragmentos de sequências de RNAm extraídos do alvo, comparando-os a um genoma ou transcriptoma de referência. Isso possibilita a identificação das regiões genômicas de origem de cada fragmento. Esses fragmentos são inicialmente tratados para manter um padrão de qualidade, evitando alinhamentos espúrios. Em seguida, os fragmentos podem seguir dois caminhos: i) mapeá-los diretamente contra uma referência (abordagem *mapping-first*); ou ii) remontar o transcrito original usando uma técnica de montagem *de-novo* para fragmentos sobrepostos (abordagem *assembly-first*). Independentemente da técnica escolhida, a próxima etapa é identificar as regiões cromossômicas de origem dos fragmentos encontrados (HAAS et al., 2019).

No contexto do *trans-splicing*, buscamos fragmentos que não se alinham a uma única região, mas sim a duas ou mais regiões simultaneamente. Esses fragmentos, que começam em um gene e continuam em outro, às vezes de regiões distantes no genoma, são denominados reads discordantes (KUMAR et al., 2016), podendo incluir os chamados *split-reads*. Nos *split-reads*, há um par de fragmentos de uma biblioteca de RNA-seq *paired-end*, em que cada fragmento do par está presente em uma região cromossômica. Outro exemplo são os chamados *spanning-reads*, que iniciam em uma região e continuam em outro *loci*. A combinação desses dois tipos de fragmentos, detectados após um processo de alinhamento ou montagem, fornece suporte para a identificação de casos de *trans-splicing*.

Finalmente, a última etapa do processo de detecção é a filtragem e, quando necessário, a classificação dos candidatos identificados, uma vez que todo o procedimento mencionado é

conhecido por gerar um volume significativo de falsos-positivos. Uma abordagem para realizar essa filtragem é considerar, inicialmente, um número mínimo de fragmentos detectados por candidato a *trans-splicing*. A presença simultânea de *split-reads* e *spanning-reads* para um mesmo candidato é um indicativo de confiabilidade. Além disso, podem ser realizadas validações utilizando softwares de apoio, como BLAT. Todo o processo de detecção implica em um custo computacional elevado, e o processamento, detecção e filtragem dos candidatos demandam ferramentas capazes de lidar com o volume e complexidade dos dados.

3.6 RNAs Circulares

RNAs circulares (circRNA) constituem uma classe estrutural de RNA que exibe uma conformação não linear em sua estrutura. Durante o processo de *splicing*, as extremidades 3' e 5' de um pré-mRNA são covalentemente unidas em um *loop* contínuo, conferindo ao circRNA uma conformação circular (GUERRA et al., 2020). Esse processo de união dessas extremidades é denominado *back-splicing*. Em resumo, o pré-mRNA é processado pela maquinaria do spliceossomo, resultando em uma ligação direta da região doadora 5' *downstream* de um éxon com a receptora 3' *upstream*. O processo de *back-splicing* é considerado o principal mecanismo de formação de circRNA (ZHENG et al., 2016).

Os RNAs circulares mais comuns têm sua origem a partir de éxons; no entanto, não é incomum encontrar aqueles compostos por fragmentos de íntrons e éxons, ou mesmo formados exclusivamente por íntrons (GUERRA et al., 2020). Os circRNA estão amplamente distribuídos em diversas espécies e exibem conservação evolutiva, sendo identificados em vírus, mamíferos e até mesmo em plantas, destacando sua relevância no contexto evolutivo do genoma (TAN; LIM, 2021).

Inicialmente, os RNAs circulares foram categorizados como subprodutos do processo transcricional do RNA, sem considerar a possibilidade de que sua estrutura pudesse codificar proteínas. Isso ocorre devido à presença de regiões de ligações enzimáticas nas extremidades 3' ou 5' em estruturas lineares, as quais não estão presentes nos RNAs circulares devido à sua conformação peculiar. No entanto, é importante destacar que vários genes que dão origem a *circRNA* são genes codificadores de proteínas, e estudos recentes conseguiram demonstrar *in vivo* a capacidade de tradução desses RNAs circulares (PAMUDURTI et al., 2017).

Os RNAs circulares desempenham funções biológicas bem-caracterizadas no genoma, incluindo a regulação transcricional e o sequestro de microRNAs. No que diz respeito à

regulação transcricional, os *circRNAs* têm a capacidade de se ligar a *loci* específicos do DNA/RNA, formando estruturas em *loop* que resultam na pausa da transcrição desses elementos, afetando o recrutamento de fatores de *splicing* (CONN et al., 2017). Alguns *circRNAs*, que apresentam éxons e íntrons em sua conformação, podem se associar à região da subunidade menor U1 da maquinaria de *splicing* por meio de interação RNA-RNA e, assim, interagir com genes-alvo, aumentando a expressão destes (LI et al., 2015).

No contexto do sequestro de microRNAs, muitos *circRNAs* apresentam sítios conservados de ligação para microRNAs em sua estrutura circular. No caso específico do *circRNA* CDR1AS, que contém aproximadamente 63 sítios para o RNA miR-7, sua presença eleva os níveis desse microRNA no contexto celular, correlacionando-se com a progressão tumoral em humanos (YU et al., 2016; ZHONG et al., 2019; ZOU et al., 2019).

4 OBJETIVOS

Analisar eventos de *trans-splicing* e circRNA no transcriptoma de células derivadas de indivíduos com mutações em genes-chave do neurodesenvolvimento.

4.1 Objetivos Específicos

- Desenvolver ou aprimorar técnicas de bioinformática para a detecção *in silico* de *trans-splicing* e circRNA no transcriptoma dos organoides, e analisar os dados de sequenciamento das amostras de RTT e NOVA1 para identificar moléculas de interesse.
- Realizar buscas sistemáticas em literatura especializada sobre ferramentas disponíveis, desenvolvendo testes para avaliar a performance das ferramentas existentes.
- Filtrar os candidatos de *trans-splicing* e circRNA com base em sua confiabilidade, e cultivar linhagens celulares *in-vitro* originadas de fibroblastos de pacientes com mutações e controle nos genes MECP2 e NOVA-1.
- Reproduzir os protocolos de desenvolvimento de organoides cerebrais, caracterizar as estruturas de interesse através de marcadores para validar seu desenvolvimento cerebral, e utilizar técnicas de biologia molecular para validar qualitativa e quantitativamente os transcritos detectados.
- Caracterizar a função biológica dos candidatos de *trans-splicing* e circRNA utilizando bancos de dados públicos, ferramentas de bioinformática e análises estatísticas.

5 MATERIAL E MÉTODOS

5.1 Cultivo, diferenciação e caracterização dos organoides

Quatro linhagens humanas de células-tronco pluripotentes induzidas (hiPSC) foram utilizadas neste projeto, geradas por colaboradores na UCSD (Laboratório do Dr. Alysson Muotri). Dessas, duas são para o gene MECP2 (controle e afetado), e duas são para o gene NOVA1 (controle e afetado), cada uma contendo duas e três réplicas técnicas, respectivamente (ver Tabela 1). A linhagem com mutação de deleção no gene MECP2 está localizada no aminoácido da posição 83 (QX83), e seu controle complementar apresenta a versão sem a mutação (WT83). Nesse grupo, as células hiPSC foram diferenciadas em células progenitoras neuronais e, em seguida, em neurônios adultos e astrócitos, selecionados por meio de citometria de fluxo (FAC-Sorting). Para o segundo grupo, a mutação no gene NOVA-1 carrega características ancestrais de um aminoácido isoleucina na posição 200 (NOVA), e sua versão controle tem características modernas, com um aminoácido valina na mesma posição (ASC). Ambas as edições foram feitas utilizando a técnica CRISPR/Cas9. As quatro linhagens apresentam o gene em homozigose e foram amplamente caracterizadas e validadas em trabalhos anteriores (MARCHETTO et al., 2010a; TRUJILLO et al., 2021b).

Tabela 1 - Resumo das amostras por gene, acrónimo e descrição

Gene	Acrónimo	Descrição
<i>NOVA-1</i>	ASC	Controle <i>NOVA1</i> – 3 réplicas técnicas – Clone A, B e C
	NOVA	Mutação de substituição no gene <i>NOVA1</i> – 3 réplicas técnicas – Clone A, B e C
<i>MECP2</i>	WT-Astro	Controle do gene <i>MECP2</i> – Astrócitos – 2 clones biológicos (C6 e C9)
	WT-Neuro	Controle do gene <i>MECP2</i> – Neurônios – 2 clones biológicos (C6 e C9)
	QX-Astro	Mutação de deleção no <i>MECP2</i> – Astrócitos – 2 clones biológicos (C9 e C2)
	QX-Neuro	Mutação de deleção no <i>MECP2</i> – Neurônios – 2 clones biológicos (C9 e C2)

Fonte: o autor, 2024.

Para o descongelamento e início do cultivo celular das células hiPSC, mantendo o controle de qualidade, as células foram rapidamente descongeladas em banho-maria a 37°C e transferidas para um tubo cônico de 15 ml contendo meio de cultura pré-aquecido. Em seguida, as células foram centrifugadas e ressuspensas em meio fresco. O controle de qualidade das células foi realizado pela avaliação da viabilidade celular e pela detecção de contaminação microbiológica utilizando kits específicos. Para iniciar o cultivo celular, as células foram transferidas para placas de cultura e mantidas a 37°C e 5% de CO₂. O monitoramento do crescimento celular e a análise periódica da expressão de marcadores pluripotentes foram realizados para avaliar a qualidade do cultivo celular. Uma vez estabilizado e validado o processo de descongelamento e viabilidade celular, é possível iniciar o protocolo de diferenciação em organoides.

Os organoides corticais foram desenvolvidos conforme protocolo estabelecido e testado por Negraes et al. (2021) e Trujillo et al. (2019). Este protocolo mantém os organoides em meio de cultura preparado com fatores de crescimento e diferenciação que promovem a proliferação, maturação, glicogênese e atividade eletrofisiológica. Primeiramente, as colônias celulares de

hiPSC descongeladas foram posicionadas em uma placa de 6 cm revestida com Matrigel e, uma vez atingido cerca de 70-80% de confluência, o protocolo é iniciado (ver Figura 4). As colônias foram então dissociadas utilizando Accutase (Life Technologies) e ressuspensas em meio mTeSR1 suplementado com 10 μ M de SB (inibidor de TGF β 1), 1 μ M de dorsomorfina (Tocris) e 5 μ M de inibidor de ROCK (StemCell). Uma vez concluído o processo, as células foram filtradas e centrifugadas por 3 minutos a 0,6 rpm, o sobrenadante foi descartado e o pellet suspenso novamente em meio semelhante ao anterior. Na última etapa, as células foram quantificadas e selecionamos um mínimo de 3×10^6 células em um poço da AggreWell 800 (StemCell), previamente tratado com solução antiaderente. No dia seguinte, as esferas formadas foram transferidas para poços menores e mantidas durante todo o processo em um shaker de rotação a 95 rpm, temperatura de 37°C, 5% de CO₂ e ambiente úmido. Assim que o processo de agregação estiver estabilizado, é possível iniciar o protocolo de diferenciação dos organoides, com foco no desenvolvimento cortical.

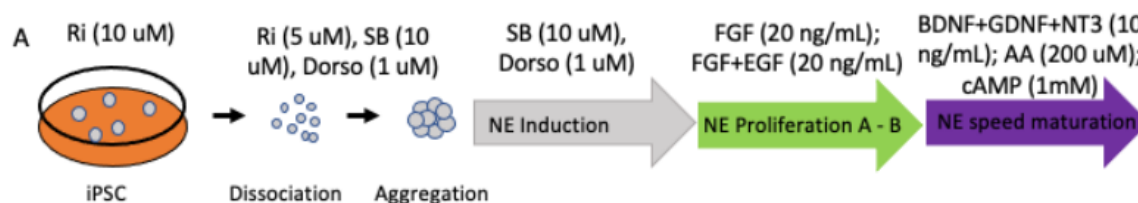


Figura 4 – Resumo esquemático do processo de cultivo, diferenciação e maturação dos organoides desenvolvidos a partir de células de hiPSC.

As células foram primeiramente tratadas com inibidor ROCK (Ri) por, pelo menos, uma hora. Em seguida, durante o processo de dissociação, as células foram tratadas com **PBS**, Ri e Dorso por aproximadamente 10-20 minutos. Com isso, foi possível coletar o material suspenso e filtrar por um tamanho mínimo de 40 μm . As células foram, então, contadas para uma quantidade mínima de 3×10^6 unidades, e o processo de agregação se iniciou para formação das esferas neurais. Durante os próximos 28 dias, foi aplicado um protocolo rigoroso de indução, proliferação e maturação das esferas neurais em organoides corticais, seguindo trocas sucessivas de meio com diluição exata de fatores de crescimento e diferenciação. Fonte: de Souza JS (2024).

Após três dias, um novo meio de cultura neurobasal, denominado M1, foi utilizado até o nono dia, contendo 1X GlutaMAX (Gibco), 1X Gem21 (Gibco), 1X de N2 NeuroPlex (Gemini), 1X NEAA (Gemini), 1X Pen/Strep (Gibco), 10 μM SB e 1 μM de dorsomorfina. Este meio foi trocado em dias alternados durante 6 dias, como parte do processo de indução neuronal (NE) dos organoides corticais.

Um segundo meio de cultura neurobasal (M2) foi empregado por treze dias, contendo os suplementos do meio anterior, adicionando fatores de diferenciação neuronal, como 20 ng/mL FGF2 (R&D Systems) e 50 ng/mL EGF2 (R&D Systems). Nesta etapa, o objetivo foi alcançar a proliferação neuronal e o crescimento dos organoides.

Por fim, foi aplicado um protocolo de aceleração da maturação neuronal, utilizando 10 ng/ml de BDNF (Peprotech), 10 $\mu\text{g/ml}$ de GDNF (Peprotech), 10 ng/ml de NT3 (Peprotech), 200 μM de ácido ascórbico (Sigma) e 1 mM de dibutyryl-cAMP (Sigma). O processo completo levou,

em média, 30 dias, com sucessivas trocas de meio, acompanhadas de medições do tamanho e formato dos organoides, sempre mantendo os padrões de qualidade de desenvolvimento.

Para validar se os organoides cultivados desenvolveram as estruturas essenciais para serem considerados cerebrais, aplicamos a técnica de imunistoquímica, utilizando marcadores específicos do neurodesenvolvimento. Os organoides foram fixados em paraformaldeído 4% (PFA), e, em seguida, com sucrose 30%, permeabilizados e bloqueados com 0,1% de Triton X-100 e 3% de FBS em PBS. Posteriormente, foram incubados com anticorpos primários durante a noite a 4°C. Os anticorpos selecionados foram: *MAP2* (Galinha - 1:1000), marcado com AlexaA-555 para neurônios; *FOXG1* (coelho – 1:500), marcado com Alex647 para o telencéfalo; *NESTIN* (rato – 1:500), marcado com AlexaA-488 para o sistema nervoso central; e, por fim, DAPI para o núcleo. Após serem lavadas com PBS, as amostras foram incubadas com os anticorpos secundários (Life Technologies) por duas horas em temperatura ambiente. As placas foram montadas e analisadas em um microscópio de fluorescência (Axio Observer Apotome, Zeiss).

5.2 Detecção e filtragem *in-silico*

Para a análise do transcriptoma dos organoides, o RNA total com cauda poli(A) foi extraído das quatro linhagens celulares utilizando o kit RNeasy (Qiagen). As amostras com qualidade mínima foram submetidas ao sequenciamento Illumina HiSeq 2500 para gerar *reads* de 100 pares de base *paired-end*, com um mínimo de 40 milhões de fragmentos sequenciados por amostra. Posteriormente, os dados foram processados através de um rigoroso controle de qualidade para remover contaminantes, reads de baixa qualidade e fragmentos indesejados ou artefatos de sequenciamento, utilizando os softwares NGS QC Toolkit, Cutadapt e TrapFilter (MARTIN, 2011; PATEL; JAIN, 2012).

Com as *reads* em uma qualidade mínima para análises de transcriptoma, iniciamos um processo de seleção da ferramenta a ser aplicada na detecção dos candidatos a *trans-splicing* e circRNA. Optamos por realizar uma revisão sistemática da literatura (RSL), buscando identificar as melhores ferramentas em termos de sensibilidade, desempenho e reprodutibilidade. O método para a RSL baseou-se no protocolo desenvolvido por Kitchengam em 2007 para engenharia de software (KEELE, 2007). Nosso objetivo principal foi identificar o estado da arte de todas as ferramentas bioinformáticas publicamente disponíveis em bancos de dados capazes de detectar eventos *de trans-splicing* em dados de RNA-seq. Essa revisão incluiu ferramentas de todas as plataformas e linguagens de programação.

Os critérios de inclusão foram: ferramentas publicadas até setembro de 2020; apenas publicações em inglês; e apenas ferramentas que propõem a descoberta de moléculas quiméricas, transcritos de fusão ou *trans-splicing*. Foram excluídas desta revisão as ferramentas anteriores a 2010, bem como as que ainda não foram publicadas ou finalizadas, ou que foram descontinuadas.

As bases de dados públicas definidas para a busca foram o *PubMed* e o *Scopus*. O *PubMed* é um mecanismo de busca gratuito desenvolvido e mantido pelo Centro Nacional de Biologia (NCBI), na Biblioteca Nacional de Medicina (NLM) dos Estados Unidos, que possui mais de 27 milhões de citações e resumos de estudos científicos em biomedicina e biotecnologia (<https://www.ncbi.nlm.nih.gov/pubmed/>). O *Scopus* é outra base de dados pública que oferece uma ampla quantidade de publicações de quase todas as áreas do conhecimento, com aproximadamente 69 milhões de registros disponíveis desde 2004 (<https://www.scopus.com>). Para garantir que nossa revisão incluísse todas as possíveis ou a maioria das ferramentas disponíveis para a comunidade científica, aplicamos termos de busca e sinônimos para as palavras-chave relevantes:

(trans-splicing OR chimeric OR fusion OR gene fusion) AND (transcript OR gene OR molecule OR splice junction) AND (software OR program OR tool OR script OR algorithm OR detection OR splice detection) AND (ngs OR next generation sequencing OR high throughput sequencing OR RNA-seq OR transcriptome OR sequencing)

Para avaliar o desempenho das ferramentas selecionadas na detecção de moléculas de trans-splicing (TS), criamos um conjunto de dados positivos contendo um número fixo de leituras, a fim de verificar o impacto das estratégias de alinhamento, a influência dos filtros na sensibilidade e os métodos de pontuação do relatório dos candidatos. Este conjunto de dados foi construído de acordo com o tipo mais comum de dados RNA-seq disponíveis na literatura até o momento: dados pareados, com um tamanho de leitura de 100nt e com pontuação de qualidade ótima nas leituras (Figura-5).

Selecionamos dez genes de fusão já caracterizados e validados na literatura no banco de dados NCBI. A sequência de DNA da isoforma mais frequente das fusões escolhidas foi extraída como arquivos fasta e separada em dois arquivos diferentes: um contendo a parte esquerda da fusão e outro com a parte direita. Desenvolvemos um script em Python para concatenar esses arquivos em um único arquivo de fusão fasta, mantendo quatro características principais: 1) os pontos de quebra para a fusão foram fixados; 2) 20 leituras divididas e 20 leituras abrangentes para cada gene; 3) número aleatório de tamanho de fragmento para cada gene entre 10-90nt; e 4) um tamanho mínimo de âncora de 10nt.

Seguindo a linha de teste de diferentes estruturas em bibliotecas RNA, construímos quatro conjuntos de dados: 1) estruturas Éxon-Éxon respeitando os pontos de quebra canônicos; 2)

estruturas Éxon-Broken Éxon, onde o ponto de quebra pode estar presente em qualquer parte do éxon; 3) estruturas Éxon-Íntron; e 4) leituras Íntron-Íntron.

Como etapa final, convertimos este arquivo fasta para FastQ e criamos uma pontuação artificial com BBMAP, definida como o valor máximo para a pontuação, para não descartar nenhuma leitura devido à baixa qualidade. Esse processo resultou em um conjunto de dados contendo um total de 1.600 leituras pareadas com 100nt, com *reads* de fusão para teste da qualidade das ferramentas. O arquivo contendo as sequencias gerado para esses testes, assim como o script em Python, pode ser encontrado em anexo ou online no GitHub do autor ([link](#)).

Os critérios de escolha das ferramentas levaram em consideração aquela que apresentou a melhor performance nos testes de performance, e também em termos de usabilidade, como por exemplo: consistência nos resultados, multiprocessamento e velocidade. Calculamos o índice de Jaccard, que é uma métrica usada para quantificar a similaridade entre dois conjuntos de dados. No caso os dados com candidatos positivos e o que a ferramenta devolvia como casos identificados. A média dos quatro tipos de fusões foi então calculada, e os valores comparados por uma análise de ANOVA. Aqueles que apresentaram índice de pelo menos 0.8 de similaridade, foram então selecionados para a análise de usabilidade.

Optamos por abordar essa questão usando três metodologias distintas e analisar, primeiro, a interseção dos dados e, em seguida, os melhores candidatos de cada método. Utilizamos uma metodologia de mapeamento com a ferramenta ARRIBA (UHRIG et al., 2021), uma tecnologia de construção de fragmentos *de novo* com a ferramenta JAFFA (DAVIDSON; MAJEWSKI; OSHLACK, 2015), e uma metodologia própria de filtragem pela ferramenta BLAT. Para os candidatos de cada método, foi avaliada manualmente a sua qualidade: número de *reads* que davam suporte à fusão ou à circularização, genes de origem e qualidade dos alinhamentos. Ferramentas de suporte nessa filtragem, como o BLAT, IGV e SAMTools, foram

utilizadas (ROBINSON et al., 2011). Uma análise descritiva de cada amostra foi desenvolvida ao final do processo. Todas as análises foram executadas em ambiente com sistema operacional Linux versão 16.04.

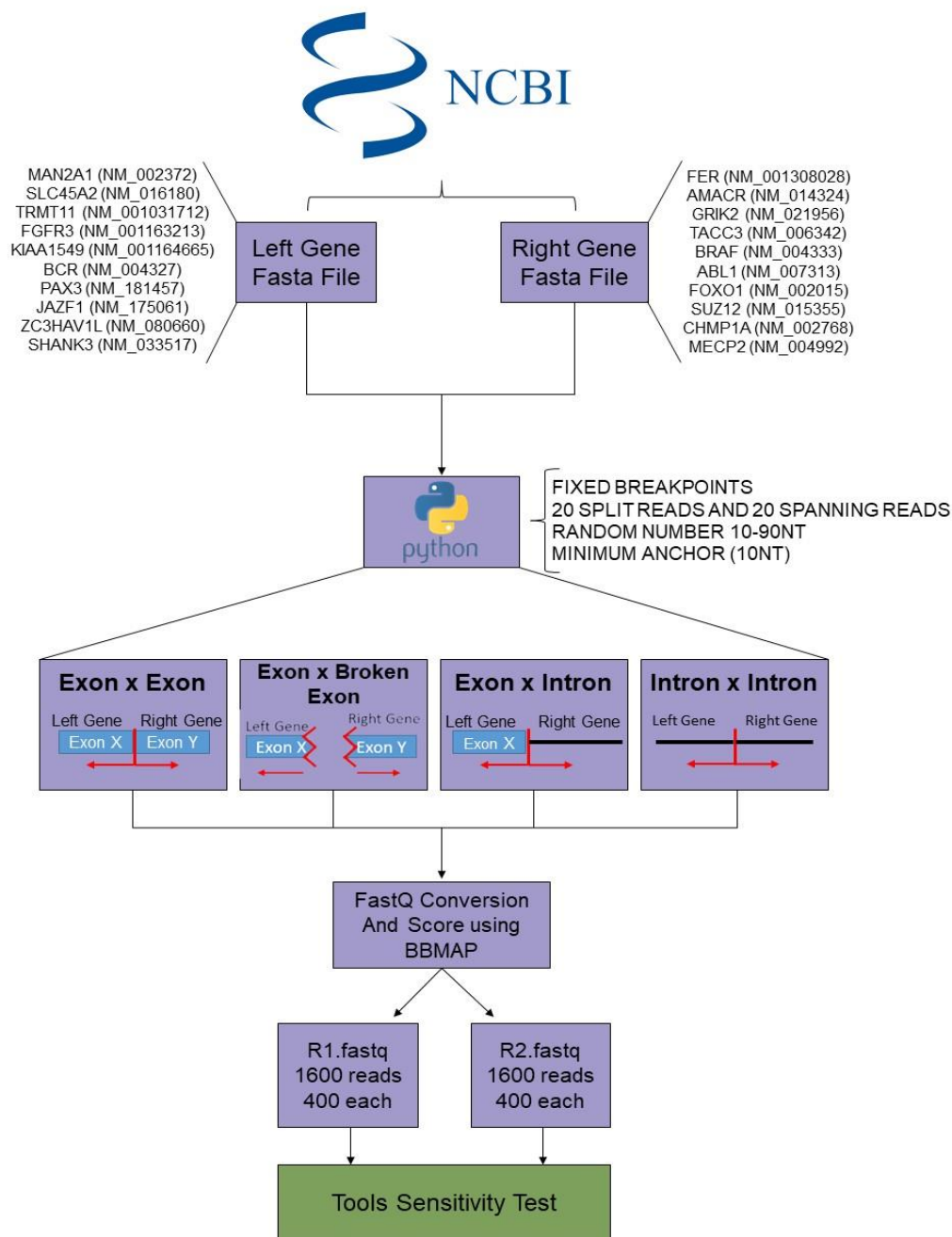


Figura 5 – Resumo esquemático do processo de criação de um *dataset* positivo para avaliação das ferramentas de detecção de *trans-splicing*.

Genes que já foram caracterizados como participantes de fusão a nível de RNA foram selecionados e suas sequências de RNA foram coletadas do NCBI. Um script em Python foi desenvolvido capaz de juntar as *reads* em fusões respeitando algumas características que um verdadeiro caso de TS precisa desenvolver. Após, quatro categorias de fusões foram reproduzidas para testar a capacidade das ferramentas em detectar diferentes cenários. Finalmente, 40 reads de cada fusão foram geradas e utilizadas no processo de testes de sensibilidade para as ferramentas selecionadas na RSL. Fonte: O autor (2024).

A metodologia de RNA-seq permite não apenas a busca por elementos ainda não descobertos no genoma, como é o caso do *trans-splicing* e circRNA, mas também a medição dos níveis de expressão gênica entre as amostras ou tratamentos. Para atingir esse objetivo, conduzimos a análise de expressão diferencial de genes. As amostras foram quantificadas utilizando a ferramenta featureCounts com anotação do GENCODE (versão 38) e normalizadas para o número de *reads* e o tamanho dos genes, utilizando a biblioteca DESeq2 para obtenção de métricas como expressão média para cada amostra, disponível para o software estatístico R versão 4.1.2 (LIAO; SMYTH; SHI, 2014; LOVE; HUBER; ANDERS, 2014).

A análise de enriquecimento funcional de vias metabólicas que eram de interesse para a pesquisa foi realizada utilizando o pacote topGO do banco de dados do GENE ONTOLOGY. Consideramos o nível de significância de 5% para as análises estatísticas realizadas neste trabalho.

5.3 Validação *in-vitro*

Para confirmar os resultados da etapa de detecção e filtragem *in silico*, realizamos a validação *in-vitro* dos principais candidatos identificados até o momento. O DNA total foi extraído dos organoides utilizando o protocolo DNeasy da Qiagen, assim como o RNA, utilizando o kit RNeasy (Qiagen). Para o RNA, sintetizamos o cDNA coletando 1 µg de RNA total e utilizamos o kit QuantiTec Reverse Transcription (Qiagen), seguindo as instruções do fabricante, e quantificamos o resultado utilizando o NanoDrop OneC Microvolume (ThermoScientific).

Com auxílio da ferramenta Primer3, projetamos *primers* específicos para os genes dos nossos candidatos que apresentavam uma das duas características: o par precisava conter a fusão dentro do seu amplicon (Tipo-1); ou a estrutura de um dos pares dos *primers* precisava conter a fusão ou circularização (Tipo-2) (ver Figura 6). Dessa forma, garantimos mais de uma prova da

existência do evento. Também foram avaliadas características como tamanho do amplicon, para sequenciamento por Sanger, e temperaturas de anelamento e desnaturação.

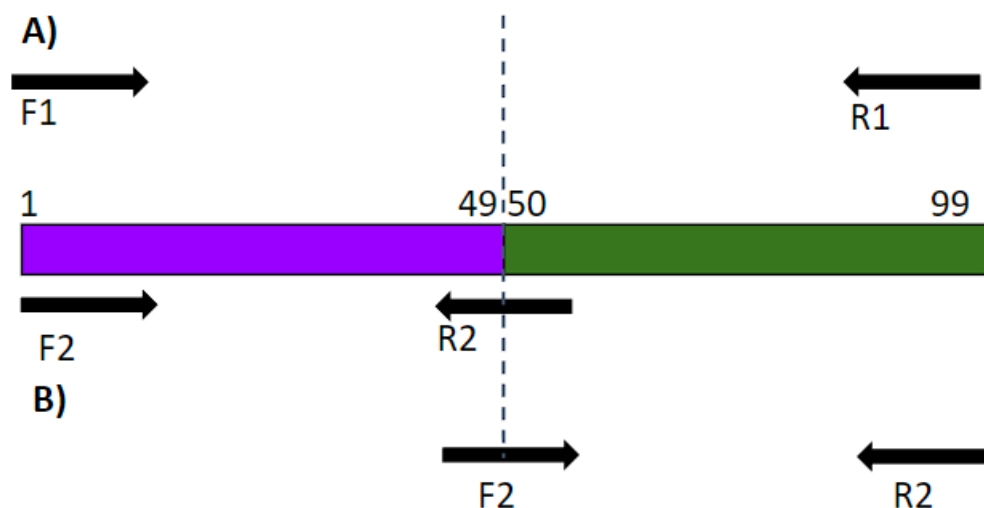


Figura 6 - Resumo esquemático do processo de elaboração dos *primers* para validação *in-vitro* dos candidatos.

A) Duas estratégias foram pensadas para validar a existência do processo de fusão, a primeira contém dois *primers* complementares no início dos genes parentais, ambos apontando em direção à fusão desejada. A segunda estratégia está representada em **B**, onde um dos pares do *primers* foi construído contendo a sequência que representava a região de fusão do candidato de *trans-splicing*. Fonte: O autor (2024).

Dois tipos de PCR foram executados neste trabalho: o PCR qualitativo e o PCR quantitativo, utilizando a técnica de qRT-PCR. Para a versão qualitativa, utilizamos o PCR Master Mix 2X (ThermoScientific), e para a versão quantitativa utilizamos o SYBRTM Green PCR Master Mix (ThermoScientific). Em todas as reações, incluímos um controle positivo contendo o gene de controle endógeno GAPDH e um controle negativo que não continha amostra – todas as análises foram realizadas em duplicata.

Para verificar a diferença de expressão utilizando a técnica de qRT-PCR, calculamos o delta-delta Ct, utilizando como referência o gene de controle positivo. Géis de agarose a 2% foram utilizados com uma *ladder* de 50bp (ThermoScientific) a 100V por 50 minutos para adicionar um nível extra de verificação. Por fim, os candidatos que foram validados pelas duas técnicas de PCR e que apresentaram resultados de qualidade mínima foram submetidos a sequenciamento por Sanger utilizando produtos de PCR não purificados, executado pela empresa GeneWiz (San Diego, California).

5.4 Análise funcional dos candidatos em dados públicos

Por fim, buscamos verificar a associação funcional dos candidatos e seu possível impacto no neurodesenvolvimento, bem como em doenças relacionadas aos genes estudados. Inicialmente, estudamos as relações estruturais das sequências de DNA e RNA utilizando as ferramentas BLAST e ENSEMBL, que permitem identificar homologias, prever estruturas secundárias e quantificar a expressão em diferentes tecidos ou condições.

Posteriormente, verificamos a estrutura e as características das regiões genômicas e funcionais dos candidatos, utilizando bases públicas de dados como *PubMed*, *Genecards* e chaves de funções biológicas com a técnica de ontologia gênica (GO). Essas chaves levam em

conta a função molecular dos genes, os processos biológicos e os componentes celulares de cada um dos candidatos, permitindo inferências biológicas a partir de grupos de genes.

Uma terceira etapa aplicada foi a análise de redes de interação com o auxílio do banco de dados STRING, para verificar se algum dos genes de fusão ou circularização apresenta relação de interação gênica ou proteína-proteína com os genes do neurodesenvolvimento.

Por fim, aplicamos a ferramenta JASPAR para análise da regulação gênica, que permite identificar elementos regulatórios em sequências de DNA e prever como esses elementos podem afetar a expressão dos genes.

6 RESULTADOS E DISCUSSÃO

6.1 Seleção de Ferramentas via RSL

A revisão sistemática possibilitou a identificação de 1.242 registros no banco de dados PubMed e 1.487 registros no banco de dados Scopus (Figura-7). Em seguida, os resultados iniciais foram filtrados manualmente por título e resumo, resultando na seleção de um total de 97 e 93 registros dos bancos de dados PubMed e Scopus, respectivamente. Posteriormente, uma análise do texto completo foi realizada, e apenas as ferramentas que apareceram em ambos os conjuntos de dados foram incluídas nesta revisão. A lista final contou com 34 ferramentas (Tabela Suplementar-1).

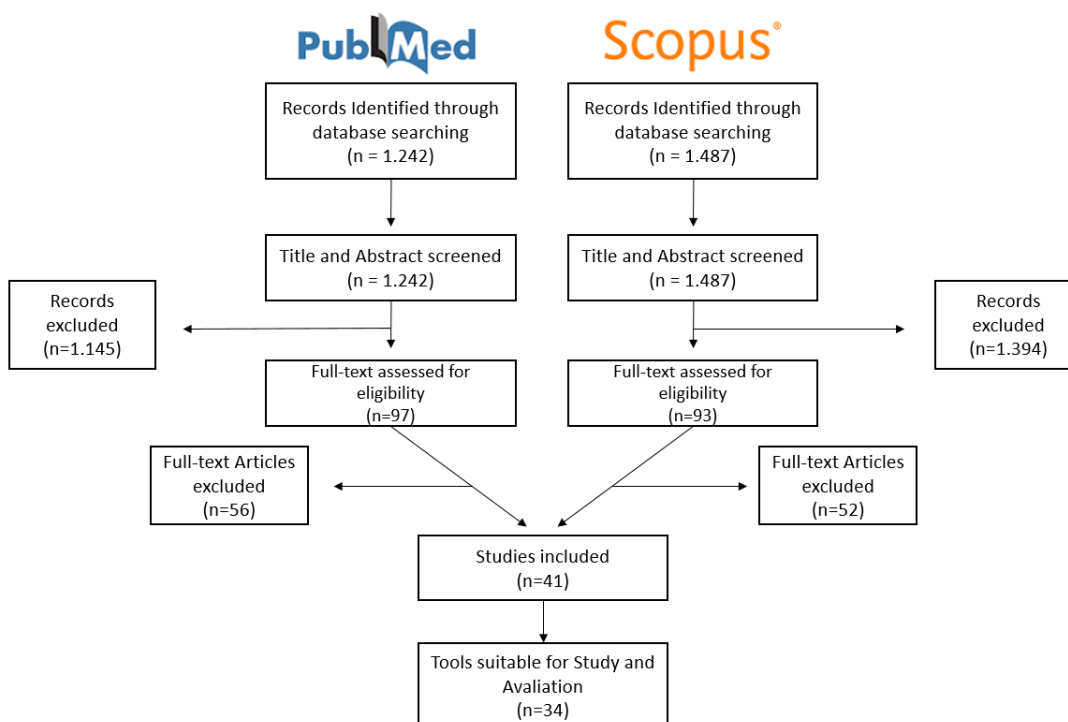


Figura 7 - Fluxograma de seleção para RSL para ferramentas computacionais de detecção de *Trans-splicing*.

As bases foram verificadas através de um script próprio para varredura de artigos usando a chave de busca escolhida. As ferramentas finais foram amplamente analisadas pelas informações que existiam no artigo ou manual de execução. Todas as informações levantadas estão registradas na tabela suplementar 1. Fonte: o autor, 2024.

Analizamos as 34 ferramentas atualmente disponíveis para mapear suas características e métodos de detecção de transcritos quiméricos. As ferramentas identificadas foram então submetidas a um conjunto conhecido de dados simulados de RNA-seq para avaliar sua capacidade de detectar transcritos quiméricos, o que foi crucial para avaliar sua integridade e eficácia (Figura-5). Nesta seção, abordaremos o processo de detecção conforme as três etapas mencionadas previamente: (i) Alinhamento de *reads*, (ii) detecção *in silico* de candidatos a *trans-splicing* e (iii) filtragem de candidatos falsos positivos.

O primeiro passo na detecção de fusões é o alinhamento de reads com uma sequência de referência de DNA ou RNA. Aqui, as ferramentas geralmente empregam duas estratégias distintas: "Mapping first" (MF) ou "Assembly first" (AF). A estratégia MF baseia-se no alinhamento direto de leituras de RNA-seq a uma sequência de referência, seguido pela identificação de pontos de quebra de fusão com base nos padrões de alinhamento resultantes. Já a estratégia AF envolve a montagem de novo das leituras de entrada em contigs, seguida pelo alinhamento desses contigs a uma sequência de referência. Das duas abordagens mencionadas, o MF emerge como a predominante no campo da detecção de transcritos de fusão. Das 34 ferramentas identificadas em nossa revisão, 24 (73,5%) são baseadas na estratégia MF, enquanto 7 (20,6%) utilizam a estratégia AF (Figura-8). As razões prováveis para a predominância de ferramentas baseadas em MF incluem: (i) maior velocidade e eficiência das estratégias de mapeamento, que consomem menos recursos computacionais em comparação com a montagem de contigs; (ii) essas ferramentas foram as primeiras a serem desenvolvidas e estabelecidas como referências para outras semelhantes baseadas em MF; e (iii) a montagem de reads curtas é mais demorada e suscetível a erros (BENOIT-PILVEN et al., 2018). Além disso, algumas ferramentas, como JAFFA (DAVIDSON; MAJEWSKI; OSHLACK, 2015) e INTEGRATE (ZHANG et al., 2016), adotam uma abordagem híbrida de MF e AF em seu funcionamento. Em

um contexto em que os recursos computacionais não são uma limitação, as abordagens híbridas podem representar uma melhoria na detecção de fusões; no entanto, ainda carecem de comparações detalhadas nesse aspecto.

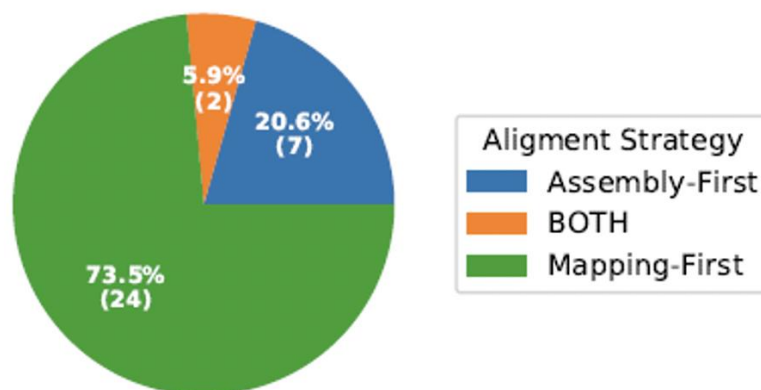


Figura 8 - Estratégias de detecção de TS em relação ao método de alinhamento usado.

Um gráfico de porcentagem indicando a principal estratégia de cada ferramenta para tratar o processo de alinhamento das sequências de RNA-seq. Fonte: o autor, 2024.

Para realizar o alinhamento preciso dos reads RNA-seq com uma referência, os softwares de detecção de fusão dependem de algoritmos de alinhamento específicos. Algumas ferramentas utilizam alinhadores RNA-seq de terceiros em suas pipelines, comumente gerenciando arquivos de saída em formatos BAM ou SAM, antes de prosseguir com a categorização e filtragem dos alinhamentos (CARRARA et al., 2013; KUMAR et al., 2016; SINGH; LI, 2021). Notavelmente, escolhas populares incluem Bowtie/Bowtie2 (41%) e Burrows-Wheeler Aligner (BWA - 17%). Por outro lado, algumas ferramentas desenvolvem seus próprios alinhadores de transcriptoma ou genoma, integrados como parte de um sistema mais amplo (Figura-8).

Essa etapa é crucial na detecção de fusões genéticas a partir de dados RNA-seq, pois um alinhamento inadequado comprometeria a precisão da detecção, tornando essa fase, juntamente

com a filtragem, a parte mais sensível das ferramentas, que variam em sua eficácia na detecção de transcritos quiméricos.

Alinhadores como Bowtie e BWA, originalmente projetados para alinhar DNA, enfrentam desafios ao lidar com lacunas de tamanho intrônico, frequentemente encontradas em genes distantes em transcritos de fusão. Por outro lado, ferramentas capazes de lidar adequadamente com reads de junção, como STAR e TopHat2, demonstraram melhor desempenho (BARUZZO et al., 2017).

Entretanto, algumas ferramentas mais antigas, como TopHat, Bowtie e BWA, enfrentam problemas de compatibilidade com outras ferramentas, instalação relacionada a bibliotecas e arquiteturas de computador, além de questões de manutenção de software. Os desenvolvedores do TopHat e Bowtie recomendam o uso de suas versões atualizadas, TopHat2 e Bowtie2, que abordam diretamente as limitações das versões anteriores. Contudo, nem todas as ferramentas de detecção de fusão implementaram essas atualizações em suas pipelines, comprometendo todo o processo de análise (BARUZZO et al., 2017).

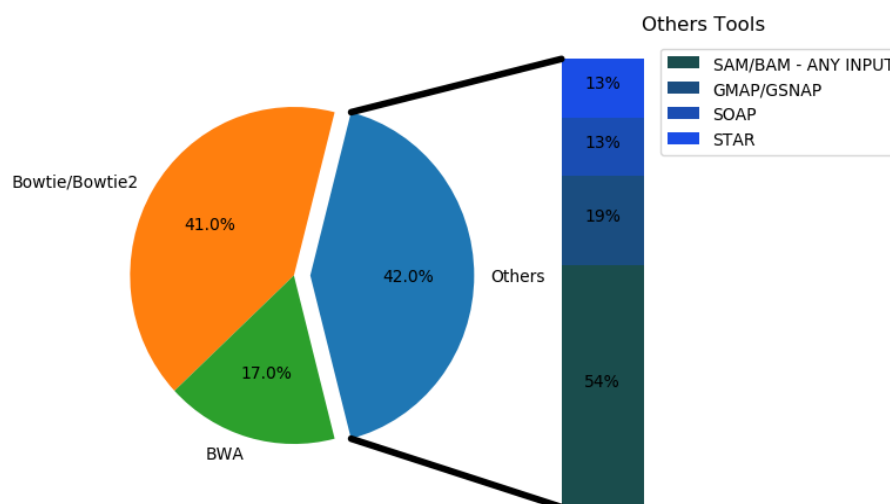


Figura 9 – Estratégia de detecção de TS em relação a qual alinhador foi construído o método de detecção.

Um gráfico de porcentagem seguido de um de barra indicando também a porcentagem do tipo de alinhador utilizado pelas ferramentas. Fonte: o autor, 2024.

Continuando com nossa metodologia de detecção de fusão em três etapas, a segunda etapa assume a responsabilidade pela identificação e classificação das *reads* mapeadas geradas na fase de alinhamento inicial. Os métodos atuais de WGS e RNA-seq oferecem dois tipos de *reads*: *single-end reads* (SE) e *paired-end reads* (PE). Os *single-end reads* representam fragmentos que são sequenciados de apenas uma extremidade, enquanto os *paired-end reads* consistem em fragmentos com orientações de ambas as extremidades contidas nas *reads*.

Além disso, os resultados do processo de alinhamento podem ser categorizados em dois tipos de saídas: *reads* concordantes, que se alinham corretamente à referência, e *reads* discordantes, que se alinham incorretamente devido a inconsistências com a referência. As *reads* discordantes exibem tamanho ou orientação de extensão inesperada, características frequentemente presentes em transcritos de fusão, destacando sua importância no processo de detecção. Quando um conjunto de *reads* SE ou PE discordantes abriga dois genes diferentes, com uma junção de fusão entre eles, temos um candidato a gene de fusão, denominado *reads split*. Além disso, as *reads* PE que se alinham a duas entidades de referência distintas - genes diferentes - também fornecem evidências para o evento de fusão e são chamadas de *reads spanning*. Se considerarmos apenas as *reads split*, o método de SE pode ser suficiente para detectar transcritos de fusão. No entanto, diferentes estudos demonstraram uma sensibilidade e desempenho aprimorados usando PE e uma combinação de *reads split* e *spanning*.

Além da importância da classificação precisa do tipo de *reads* para a detecção de transcritos de fusão, o mecanismo de formação da fusão também desempenha um papel crucial no processo. Os dados de NGS podem ser analisados em nível de DNA com WGS (sequenciamento completo do genoma) ou RNA-seq. O WGS pode oferecer insights detalhados sobre as origens da molécula de fusão - como rearranjos cromossômicos - que ocorrem em nível de DNA. No entanto, essa abordagem demanda uma quantidade significativa de sequenciamento e poder

computacional extensivo. Por outro lado, o RNA-seq fornece informações sobre os eventos de fusão transcritos, como *cis* e *trans-splicing*. Quando as *reads* estão em dois cromossomos diferentes, temos uma fusão intercromossômica; para *reads* no mesmo cromossomo, elas são denominadas fusões intracromossômicas; e se a fusão ocorre entre dois genes adjacentes, isso é conhecido como *read through*. Devido ao custo comparativamente mais baixo e à análise mais rápida, o RNA-seq se tornou um método popular para a detecção de transcritos de fusão. Algumas ferramentas utilizam uma combinação de WGS e RNA-seq para prever as moléculas de fusão, com capacidade de diferenciar eventos que ocorrem em nível de DNA ou apenas em nível de RNA.

A etapa final no *pipeline* de detecção de fusão de transcritos é o processo de filtragem e pontuação, implementado pela maioria das ferramentas para reduzir o número de falsos positivos e aumentar a confiança nas fusões verdadeiras detectadas. No nosso *pipeline* de três etapas, esse último estágio é crucial para garantir a qualidade geral da detecção.

Se uma ferramenta não implementar filtros adequados ou não aplicar nenhum filtro, o número de falsos positivos de fusão pode aumentar significativamente, comprometendo a confiabilidade da tecnologia *in silico*; fusões reais podem ser obscurecidas nos resultados e a qualidade do processo é prejudicada. Por outro lado, se uma ferramenta empregar filtros rigorosos na tentativa de remover apenas os falsos positivos, existe o risco significativo de perder eventos verdadeiros, já que esse é um aspecto inerente à exploração de dados (CARRARA et al., 2013).

Até o momento, a presença dessas moléculas no contexto genético tem sido considerada rara, e métodos restritivos podem impedir novas descobertas. Recentemente, dois autores avaliaram os efeitos de diversos elementos no processo de detecção, especialmente aqueles que afetam a precisão e os recursos computacionais (LIU et al., 2016; SINGH; LI, 2021). Esses

estudos oferecem insights valiosos como um guia para os usuários escolherem as ferramentas mais adequadas com base em suas necessidades, embora haja limitações na discussão sobre o processo de filtragem e seus efeitos. Em resumo, as ferramentas empregam diferentes estratégias de filtragem e pontuação para equilibrar entre especificidade e sensibilidade na abordagem desse desafio.

O primeiro e mais comum conjunto de filtros são os métodos de contagem, nos quais as ferramentas consideram apenas os candidatos a transcritos de fusão que atingem um número mínimo de *reads* de suporte. Em nossa pesquisa, constatamos que 94% das ferramentas utilizam esse parâmetro para determinar seus candidatos (Figura 10). As *reads* de suporte podem ser identificadas como *split-reads* ou *spanning-reads*, e um número mínimo - geralmente entre 2 e 3 - é necessário para que a ferramenta identifique o candidato de fusão. As ferramentas combinam esses dois tipos de leituras e buscam um valor mínimo (KIM; SALZBERG, 2011). Tipicamente, o usuário pode definir esse critério durante a configuração da ferramenta. Além disso, algumas ferramentas empregam técnicas de normalização para a contagem de leituras, como medidas de FFPM (fragmentos de fusão por milhão de *reads*). Por exemplo, um filtro de 0,1 FFPM mostrará apenas candidatos com pelo menos uma *read* de suporte de fusão por 10 milhões de *reads* totais (DOBIN et al., 2012). Esse método é uma maneira de eliminar artefatos de fusão devido à variação na profundidade de sequenciamento entre diferentes bibliotecas de RNA-seq.

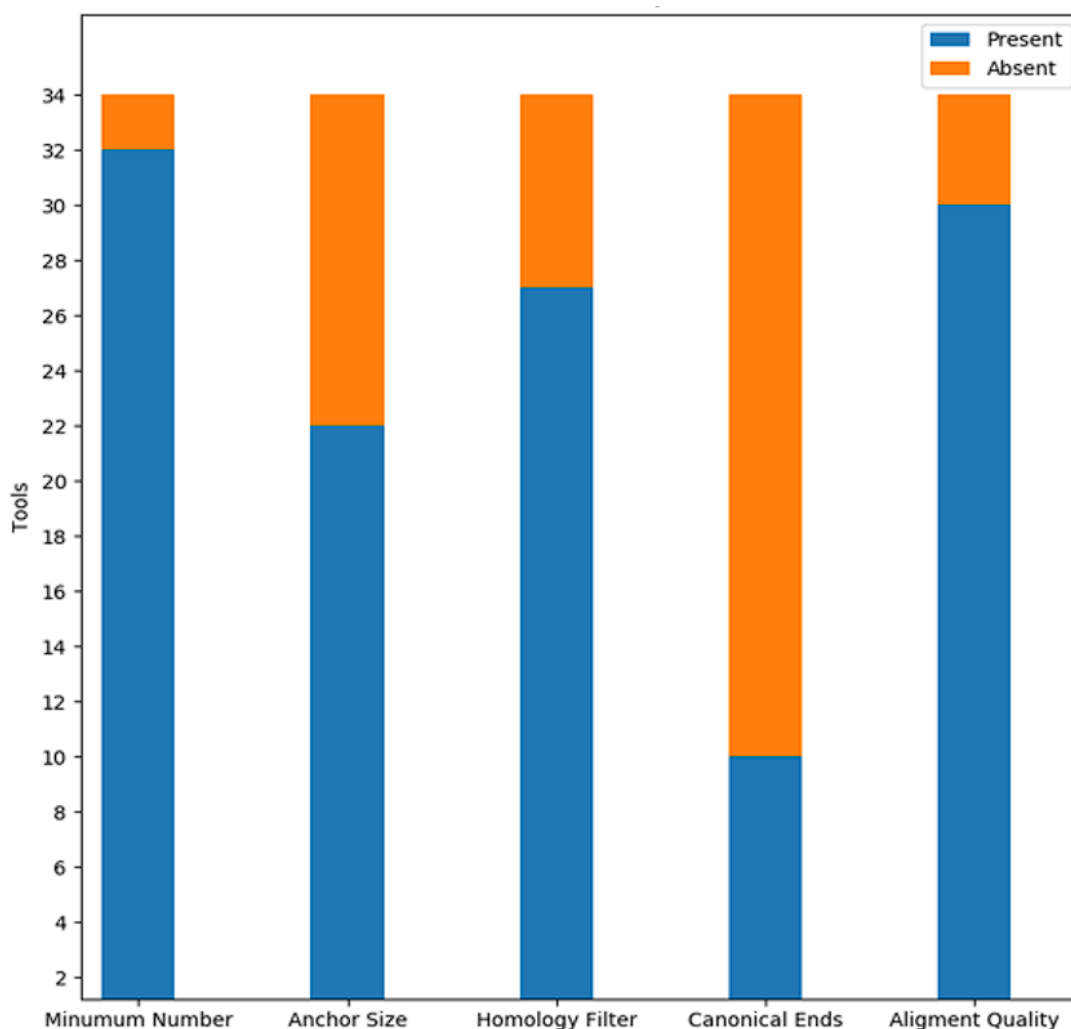


Figura 10 - Distribuição de filtros presentes nas ferramentas para detecção de *Trans-splicing*.

Gráfico de barras indicando a quantidade de ferramentas que apresentavam o determinado filtro. As mais presentes foram número mínimo de *reads* e qualidade do alinhamento. A que teve menor presença foi de regiões de *splicing* canônico. Fonte: o autor, 2024.

A segunda categoria de filtros diz respeito às características de alinhamento dos candidatos. As ferramentas geralmente incorporam filtros durante a fase de alinhamento para remover *reads* com base na qualidade do procedimento de alinhamento. Carrara et al. observaram uma correlação entre a pontuação de qualidade e a detecção falsa de fusão; resultados semelhantes foram observados nos estudos de Liu et al. e Kumar et al. em 2016. Essas descobertas destacam que uma detecção sensível de candidatos requer uma boa qualidade das leituras mapeadas.

Para *reads* divididas que se alinham em ambos os lados de um ponto de quebra de fusão de junção, o número de bases em cada lado da leitura é denominado tamanho da âncora. Esse é um filtro no qual, se o número de bases for menor que um limite predefinido, o candidato tende a ser um falso positivo. Identificamos 22 ferramentas (Figura 10) que suportam *split-reads*, as quais apresentam um tamanho de âncora maior, reduzindo os alinhamentos espúrios que possuem baixo suporte na junção de fusão. Além disso, uma abordagem para identificar candidatos verdadeiros positivos é a presença de sítios de *splicing* nos fragmentos de *split-reads* ou *spanning-reads*.

Algumas ferramentas consideram que, se o spliceossomo for responsável pela formação de transcritos de fusão, os sítios de *splicing* estarão presentes nos candidatos resultantes. Por esse motivo, se uma *read* não apresenta um sítio de *splicing* popular, como GT-AG, GC-AG e AT-AC, essa fusão provavelmente é um falso positivo. No entanto, isso deve ser considerado com cautela. Não há consenso científico sobre o papel do spliceossomo na biogênese de transcritos de fusão, e o filtro de candidatos com base nessa suposição pode limitar a descoberta de novos mecanismos de fusões de RNA.

A terceira e última categoria de filtros refere-se ao contexto genético em que as fusões propostas estão envolvidas. As duas opções populares são o filtro de homologia e o filtro de *read-through* de fusão. A maioria das ferramentas tende a remover sequências altamente semelhantes, como *reads* discordantes de genes parálogos. Essa estratégia pode reduzir a presença de falsos positivos nas análises, mas também pode filtrar fusões verdadeiras, uma vez que o *trans-splicing* pode produzir essas fusões.

Quanto às fusões de *read-through*, elas não são necessariamente removidas da análise, mas são sinalizadas nos resultados. Esse tipo de leitura é difícil de detectar, uma vez que a proximidade dos genes envolvidos pode indicar um falso candidato. Genes sobrepostos, *splicing*

alternativo e fusões estão todos muito próximos na identificação. Portanto, como solução final, a evidência experimental é necessária.

Para concluir a etapa de revisão sistemática de literatura, realizamos testes de sensibilidade nas ferramentas em que conseguimos instalar e reproduzir o processo de detecção. Descobrimos que algumas ferramentas não apresentavam mais suporte ou seus autores não mantinham seu funcionamento, e erros durante o processo de instalação eram comuns. Por esse motivo, também incluímos informações sobre ferramentas que não recebem mais manutenção e que devem ser evitadas.

Utilizando um arquivo de dados simulado (Figura 5), que continha fusões positivas de *trans-splicing*, testamos 12 ferramentas para validar qual apresentava a maior sensibilidade (Figura 13). Das 400 *reads* criadas para as quatro categorias propostas, inicialmente analisamos as fusões a nível de Éxon-Éxon, onde identificamos que o FusionCatcher (85%) e o STAR-Fusion (81,5%) foram os que apresentaram melhor desempenho. Para as fusões a nível de Íntron-Íntron, as principais ferramentas foram o STAR-Fusion (92%) e o JAFFA (88,5%). No caso de Éxon-BrokenÉxon, mais uma vez as ferramentas FusionCatcher (90%) e STAR-Fusion (87%) destacaram-se ao detectar o maior número de casos; neste cenário, também demonstramos que o JAFFA (83%) obteve um desempenho elevado em comparação com as outras ferramentas. Por fim, para o grupo de éxon-íntron, as ferramentas STAR-Fusion (92%) e JAFFA (88%) foram as que apresentaram os melhores resultados de detecção.

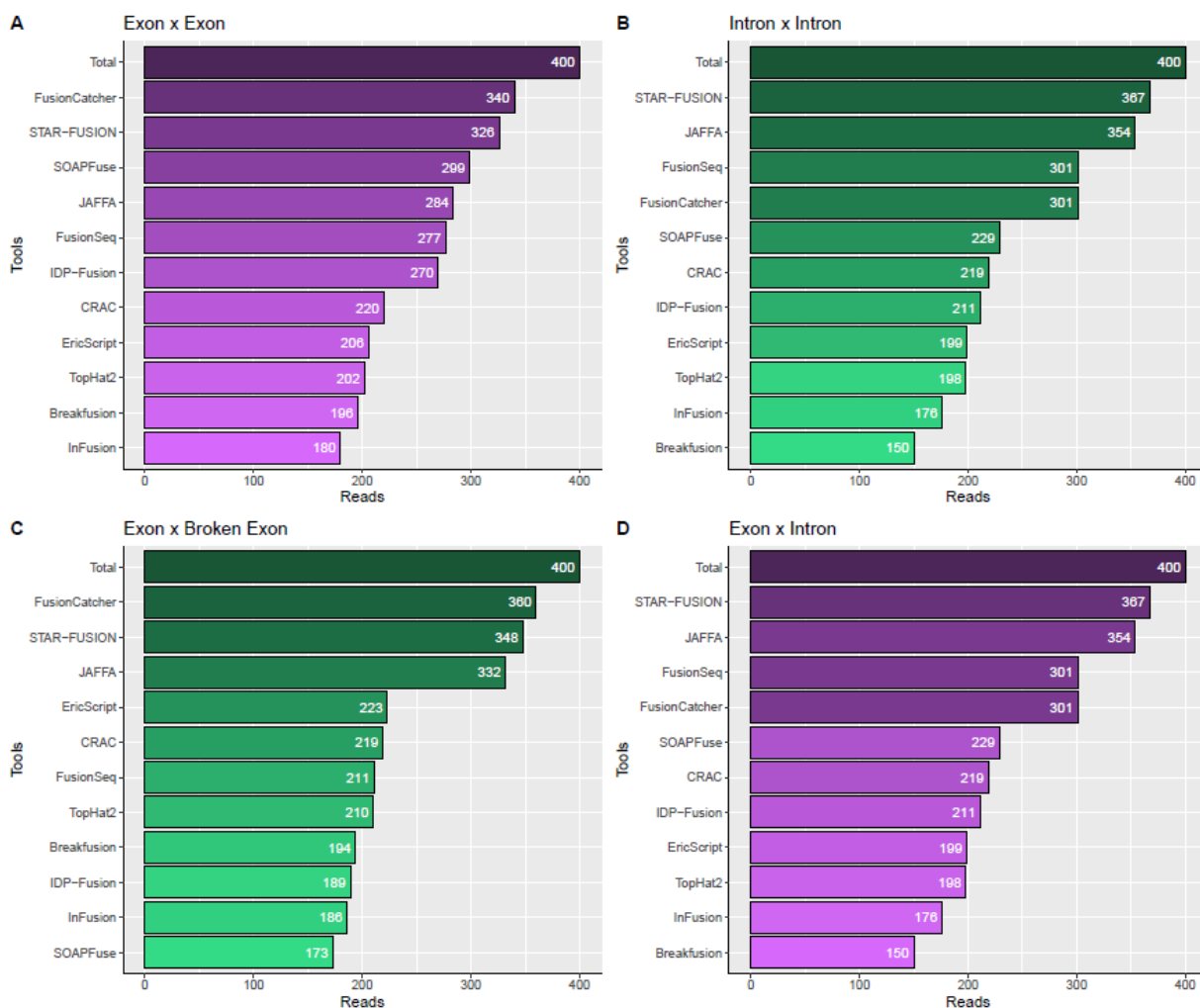


Figura 11 – Gráficos de barras indicando o número de casos positivos de cada ferramenta para cada tipo de teste desenhado.

Em cada quadrante do gráfico, um tipo de construção é apresentado. Todas as ferramentas foram testadas uma vez, e os resultados foram analisados somando o número total de casos positivos identificados dentro de cada tipo de *trans-splicing*. JAFFA e STAR-Fusion apresentaram o maior número de casos positivos entre os quatro casos. Fonte: o autor, 2024.

Para confirmar a performance dessas ferramentas, calculamos o índice de Jaccard e comparamos os quatro casos de uso para obter uma métrica geral de similaridade. As três ferramentas com maior média de performance foram: Star-Fusion com 0.89 (± 0.06); seguida de JAFFA com 0.84 (± 0.1) e FusionCatcher com 0.82 (± 0.06). Um teste de ANOVA foi executado para validar a significância estatística dessas comparações, seguido de um teste de Post-hoc. Na

tabela suplementar 2 estão os resultados detalhados dessa comparação. De maneira geral as três melhores metodologias foram estatisticamente diferentes das piores. Contudo, entre elas não houve diferença estatística. Desse modo, optamos por utilizar as ferramentas com melhor performance em similaridade, seguido das vantagens de usabilidade que as selecionadas apresentavam.

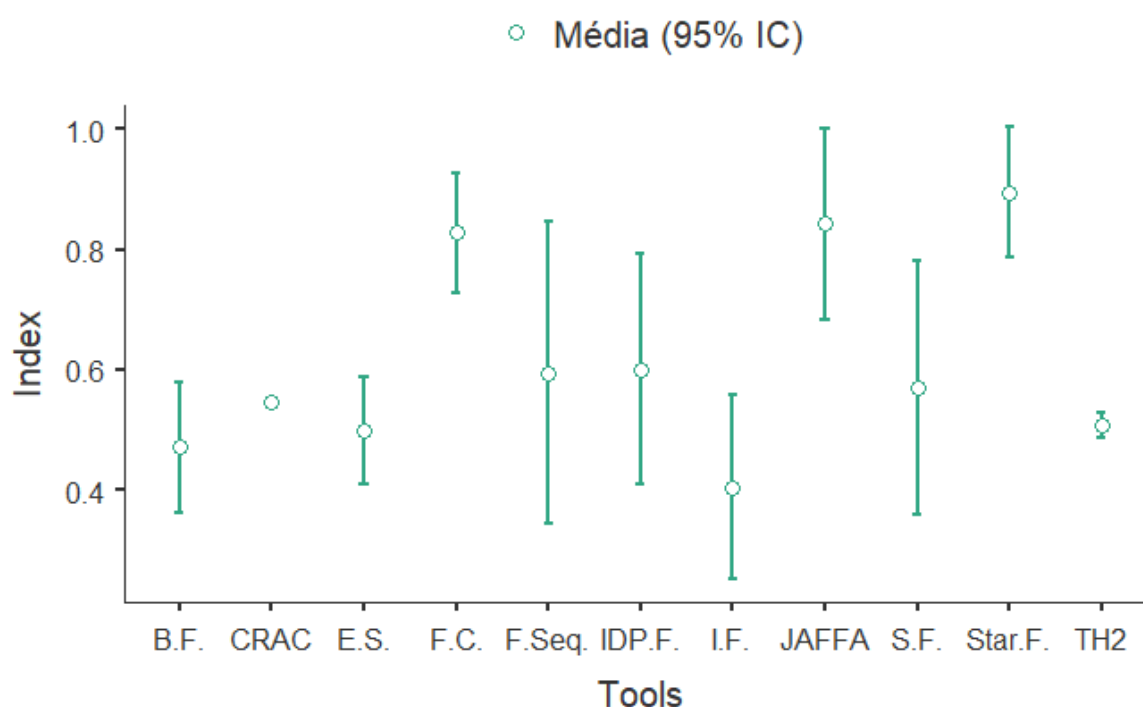


Figura 12 – Gráfico de barra de erro comparando as ferramentas em relação as medidas do índice de Jaccard.

Para cada ferramenta, a média do índice foi calculado levando em consideração os quatro casos de interesse, as bolas em verde representam a média e as barras o intervalo de confiança. Os pressupostos da ANOVA foram validados através do teste de normalidade e homogeneidade das variâncias, e correções foram aplicadas conforme necessário. Cálculo realizado no software estatístico JAMOVI. Fonte: o autor, 2024.

6.2 Cultivo, diferenciação e caracterização dos organoides

As amostras utilizadas neste estudo estão detalhadas na Tabela 1. As amostras referentes ao gene *MECP2* já estavam disponíveis no laboratório para análise, portanto, não foi necessário realizar um novo cultivo dessas células. Para as células relacionadas ao gene *NOVA-1*, foi executado o protocolo completo de cultivo dos organoides, com réplicas biológicas em triplicata. Além disso, para o gene *MECP2*, foram utilizados clones diferentes, proporcionando um nível adicional de controle para avaliar a presença dos candidatos.

O processo de diferenciação dos organoides foi documentado por meio de fotografias para avaliar o tamanho e a morfologia em cada etapa. A Figura 13 ilustra que, antes do início da diferenciação, as colônias de hiPSC não apresentavam sinais de diferenciação celular. Posteriormente, a conformação circular das esferas e o aumento de tamanho indicaram a transição de neuro esferas para organoides corticais entre as trocas de meio M2 e M3. Como parte do controle de qualidade, esferas que se condensavam no fundo da placa foram removidas durante o processo. Além disso, testes regulares de contaminação por micoplasma foram realizados no início (dia 1), no meio (dia 15) e no final (dia 30) do cultivo para garantir a integridade dos resultados de expressão gênica. Todos os testes realizados foram negativos para a presença de micoplasma, conforme indicado pelo teste da MycoAlert (Lonza).

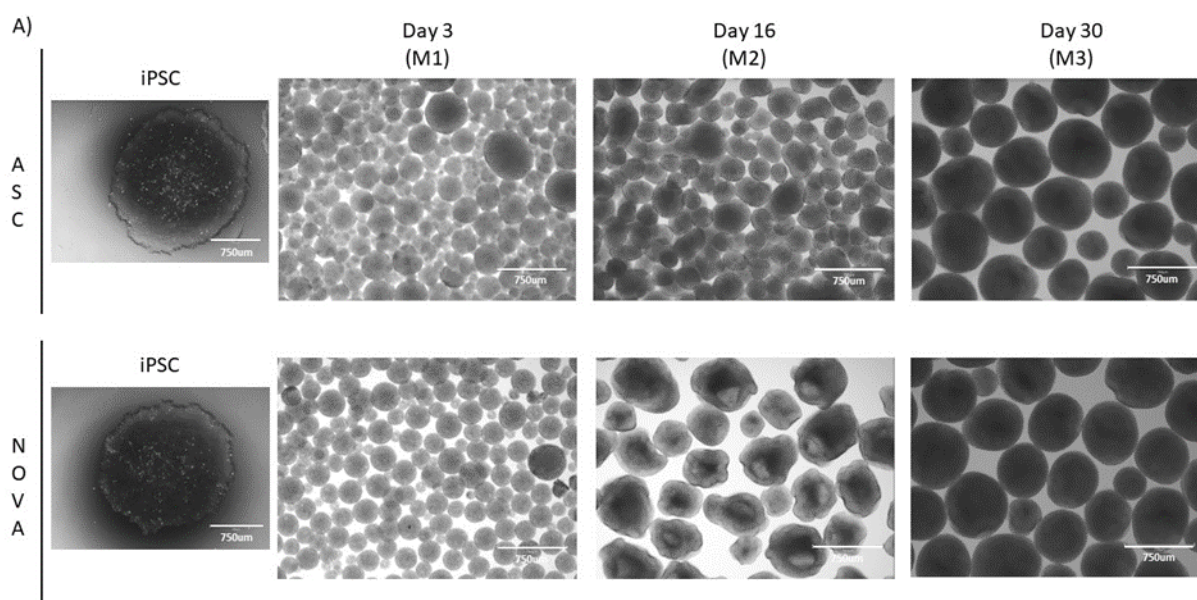


Figura 13 - Processo de cultivo e diferenciação dos organoides corticais, separado em quatro momentos para os organoides de NOVA1.

Duas linhagens celulares precisaram ser cultivadas novamente por não apresentar material disponível para validação dos candidatos. Ambas eram de mutações no gene *NOVA-1* da linha controle (ASC) e da linhagem com mutação (NOVA). Primeiramente, vemos as células de hiPSC fixadas após descongelamento, em espera do início do processo de diferenciação. Após três dias de iniciado o protocolo, as células foram expostas a três diferentes meios de cultivos, cada um com fatores e objetivos diferentes, por um tempo determinado. A cada passagem de meio, a progressão do protocolo era registrada através de fotos, durante 30 dias. Fotos coletadas no microscópio eletrônico EVOS 20x – escala 750µm. Fonte: o autor (2024).

O processo de cultivo e diferenciação de organoides a partir de células de hiPSC envolveu etapas de descongelamento, crescimento, diferenciação e coleta. Após a análise das características fisiológicas dos organoides, é plausível concluir que, inicialmente, o protocolo de desenvolvimento alcançou seu objetivo de gerar organoides corticais. Isso se evidencia pela ausência de sinais de diferenciação nas hiPSC, além da conformação das esferas e do tamanho estarem conforme o esperado pelo protocolo. Como próxima etapa, é necessário confirmar a presença de marcadores corticais para validar a existência das regiões esperadas para esse desenvolvimento celular.

Um experimento de imunoistoquímica foi conduzido para confirmar a presença de marcadores do neurodesenvolvimento, assim como a formação do córtex cerebral nos organoides (Figura 14). O primeiro marcador, em verde, foi para o gene *NES*, que está presente no desenvolvimento de proteínas do filamento intermediário de células do sistema nervoso, indicando o desenvolvimento radial do axônio. Esse gene é responsável por codificar uma proteína conhecida como nestina. Esta proteína intermediária do citoesqueleto é encontrada em células progenitoras neurais e um marcador de células tronco. Sua principal aplicação é validar células progenitoras neurais em estudos do desenvolvimento do sistema nervoso e sua presença é um indicativo de células em proliferação e de uma estabilidade do citoesqueleto celular.

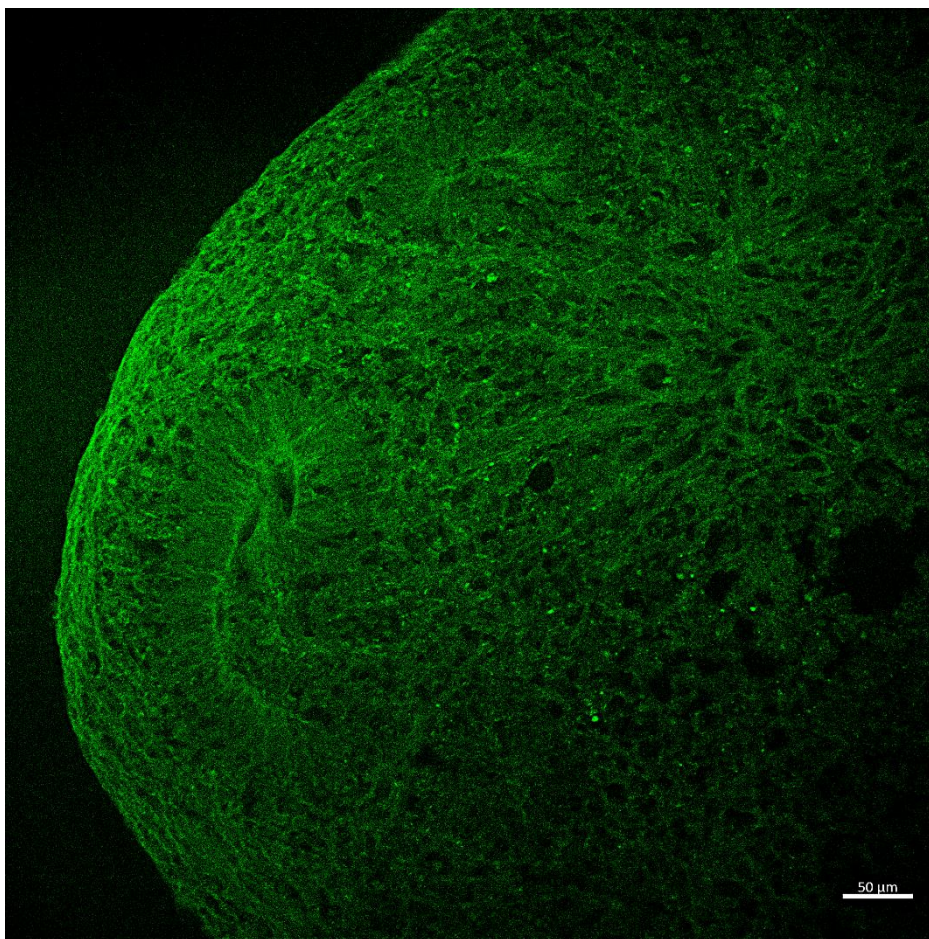


Figura 14 – Fotos do marcador celular para a expressão gene *NES* nos organoides corticais *NOVA1*.

No marcador com a fluorescência na cor verde, é possível observar o filamento intermediário de células do axônio. Imagem editada onde o brilho e a nitidez foram ajustados. Imagem coletada no microscópio eletrônico Axio Observer Apotome (Zeiss) – escala de 50 μm. Fonte: o autor (2024).

O segundo ensaio, em vermelho, foi utilizado o marcador para o gene *MAP2*, responsável pela proteína MAP2, que é associada a microtúbulos de neurônios maduros na placa cortical na região subventricular cerebral (Figura 15). Essa proteína também ajuda na organização do citoesqueleto dendrítico, facilitando o transporte de moléculas ao longo dessa estrutura, uma função essencial para a sinapse e plasticidade neural. Durante a neurogênese, MAP2 é expresso nas fases iniciais de diferenciação neuronal, marcando uma transição de células progenitoras para neurônios maduros.

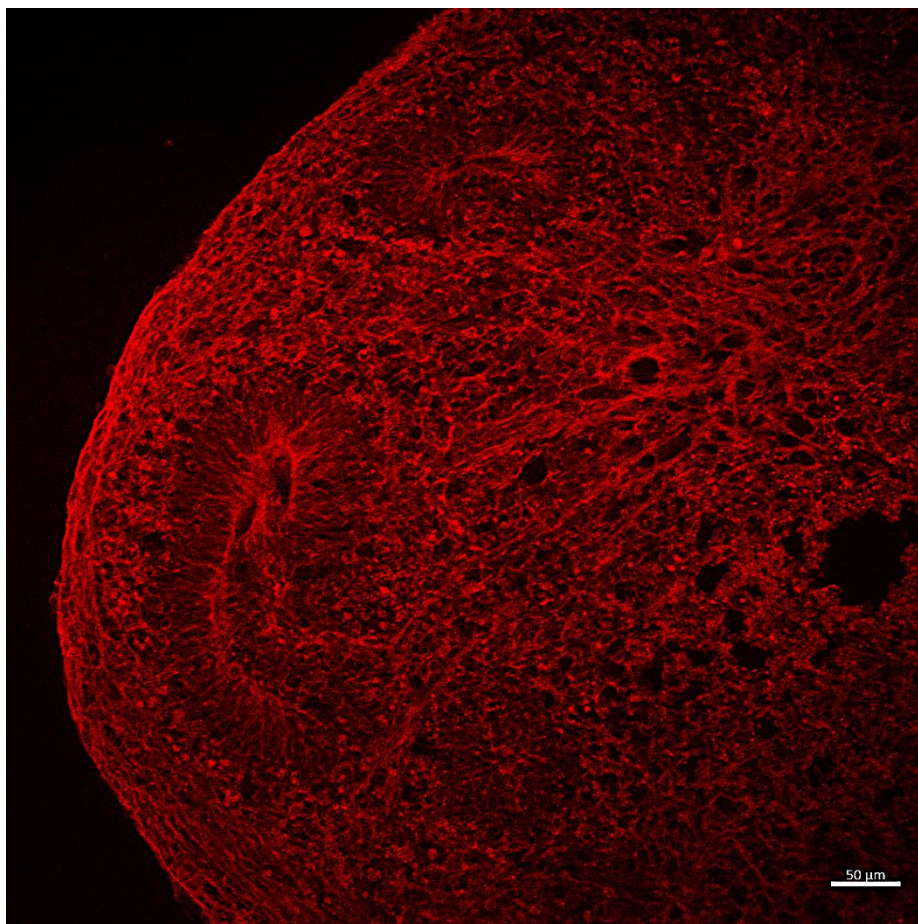


Figura 15 - Fotos do marcador celular para a expressão gene *MAP2* nos organoides corticais *NOVA1*.

No marcador com a fluorescência na cor vermelha, é possível observar a presença de neurônios através de marcação do microtúbulo da placa cortical. Imagem editada onde o brilho e a nitidez foram ajustados. Imagem coletada no microscópio eletrônico Axio Observer Apotome (Zeiss) – escala de 50 μm. Fonte: o autor (2024).

Por fim, em azul, foi o corante DAPI, que se liga fortemente a regiões do DNA ricas em timina e adenina e possui capacidade de atravessar a membrana intacta da célula, permitindo a visualização da morfologia nuclear intacta (Figura 16). Considerando que essa estrutura esteja intacta, é possível também realizar uma contagem e análise da viabilidade celular no experimento alvo.

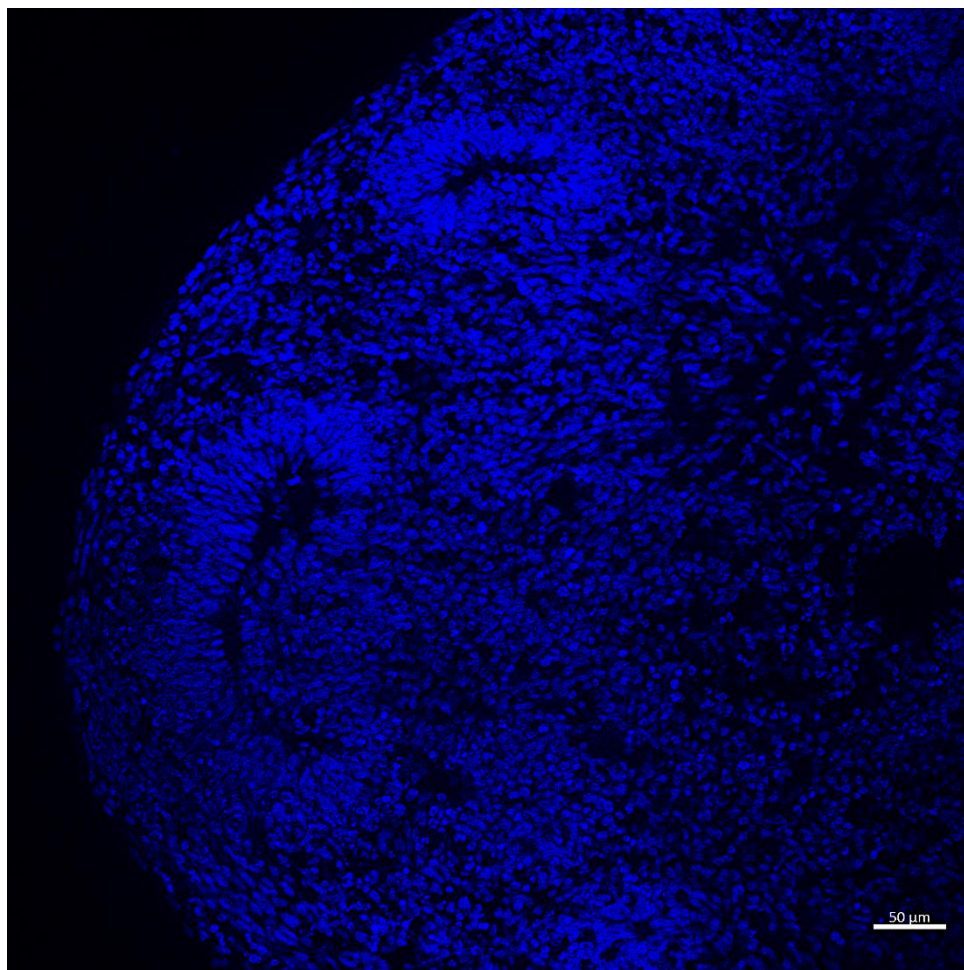


Figura 16 - Fotos do marcador celular para a expressão de DAPI nos organoides corticais NOVA1.

No marcador com a fluorescência na cor azul, é possível observar a estrutura celular intacta, que é o caso do núcleo celular. Imagem editada onde o brilho e a nitidez foram ajustados. Imagem coletada no microscópio eletrônico Axio Observer Apotome (Zeiss) – escala de 50 μm . Fonte: o autor (2024).

Com os resultados do experimento de imunoistoquímica, foi possível observar duas marcações cruciais para o desenvolvimento cortical: o desenvolvimento radial do axônio (*NES*) e a presença de neurônios maduros na placa cortical da região subventricular (*MAP2*). Além disso, a marcação da estrutura celular preservada através do núcleo intacto também foi identificada (DAPI). Com esses resultados, reforça-se o argumento de que o protocolo de desenvolvimento dos organoides ocorreu conforme o esperado, demonstrando a capacidade de simular o desenvolvimento cortical até certo ponto em um ambiente *in-vitro* de cultivo celular.

6.3 Detecção e filtragem de candidatos de TS e CircRNA

Com base nos resultados obtidos durante a RSL, optamos por adotar uma estratégia de detecção e filtragem para nosso projeto. Decidimos utilizar duas ferramentas desenvolvidas por outros autores, ARRIBA e JAFFA, além de uma ferramenta de filtragem desenvolvida por nossa equipe. Escolhemos a ferramenta ARRIBA devido à sua metodologia de alinhamento, enquanto a ferramenta JAFFA foi selecionada por ser uma técnica de montagem. Embora ARRIBA não tenha sido mencionada na RSL, ela se baseia no mesmo princípio de detecção que a STAR-Fusion, apresentando melhorias significativas na exibição dos resultados e nas opções de filtragem disponíveis.

Por fim, nossa ferramenta de filtragem representa uma solução para processar grandes volumes de dados provenientes do processo de detecção. Desenvolvemos um script em Python para realizar uma avaliação estatística dos candidatos, identificando os potenciais verdadeiros positivos e gerando arquivos BAM para validação posterior em ferramentas como BLAT ou IGV.

Durante a etapa de detecção e filtragem *in silico* dos candidatos, após o controle de qualidade e a limpeza dos dados, realizamos o mapeamento das amostras no genoma humano versão GRCh38 (hg38). Para as metodologias de alinhamento ao genoma, utilizamos a ferramenta STAR, tanto para o ARRIBA quanto para o BLAT.

No caso da metodologia de montagem de *reads de novo*, partimos das *reads* após o processo de controle de qualidade. Na Tabela 2, apresentamos a totalidade dos candidatos de *trans-splicing* e circRNA identificados no projeto por meio de qualquer uma das técnicas. Entre as amostras, aquela que registrou a maior porcentagem de *reads* de fusão ($16,43\% \pm 0,85$) e o maior número de circRNAs ($66,7 \pm 14,19$) foi a ASC. Ao mesmo tempo, essa amostra apresentou a

menor porcentagem de mapeamento único ($69,1\% \pm 1,0$). No total, foram identificados 308.551 candidatos de *trans-splicing* e 407 candidatos de circRNA.

Tabela 2 - Resumo das amostras na detecção de *Trans-splicing* e CircRNA

Amostra	# Reads	% Mapeados única	% quimérico	# Trans-splicing	# CircRNA
QX83 (C9) Astrócito	95.186.835	90,84%	0,65%	282	57
QX83 (C9) Neurônio	61.813.052	83,54%	1,17%	6104	24
QX83 (C2) Astrócito	9.588.208	91,55%	0,85%	44	5
QX83 (C2) Neurônio	155.940.827	92,29%	0,87%	605	67
WT83 (C6) Astrócito	68.066.245	89,67%	0,79%	8068	22
WT83 (C6) Neurônio	47.544.743	91,51%	0,63%	2361	24
WT83 (C9) Astrócito	65.318.082	89,31%	1,14%	469	18
WT83 (C9) Neurônio	14.785.555	93,66%	0,66%	157	9
ASC_A	38.206.786	68,99%	16,43%	104870	82
ASC_B	24.981.995	67,93%	17,29%	57092	54
ASC_C	48.909.740	69,93%	15,58%	127876	64
NOVA_A	31.797.151	92,46%	1,34%	332	17
NOVA_B	38.463.672	92,95%	1,02%	605	17
NOVA_C	29.672.003	92,71%	1,22%	335	19
Total				308551	407

Fonte: o autor, 2024.

A presença intrínseca de artefatos em experimentos de RNA-seq tornou-se evidente, especialmente considerando que a linhagem ASC apresentou a maior proporção de quiméricos. Além disso, a menor quantidade de *reads* mapeadas exclusivamente em todos os experimentos sugere que algo atípico ocorreu com essa linhagem em particular. Uma das razões principais

pode ser atribuída à complexidade do processo de preparação das amostras, que envolve múltiplas etapas de amplificação e manipulação, suscetíveis a erros de sequenciamento, viés de expressão e variações técnicas. O fato de que a linhagem ASC exibiu a menor proporção de *reads* mapeadas exclusivamente ($69,1\% \pm 1,0$) sugere uma possível redundância nos dados gerados durante esse processo. Concluimos, portanto, que essa variação é de natureza técnica, não refletindo necessariamente a variabilidade do gene NOVA-1, dado que esta linhagem serve como controle.

Um método para validar o número de candidatos identificados em nossas amostras foi correlacionar o número de genes que originaram as fusões em cada cromossomo com a quantidade total de genes anotados em cada cromossomo, conforme a anotação mais recente disponível no NCBI. Esperávamos que quanto maior o número de genes em um cromossomo, maior seria a quantidade de candidatos provenientes desse mesmo cromossomo. A figura 17 apresenta a correlação entre essas duas variáveis quantitativas. Os resultados demonstraram uma correlação positiva forte entre o número de genes em cada cromossomo e o número de candidatos identificados (R de Pearson = 0,93; $p < 0,001$), o que está em consonância com o comportamento biológico esperado para esses dados. Esse método serve como uma etapa de validação da qualidade na detecção do fenômeno.

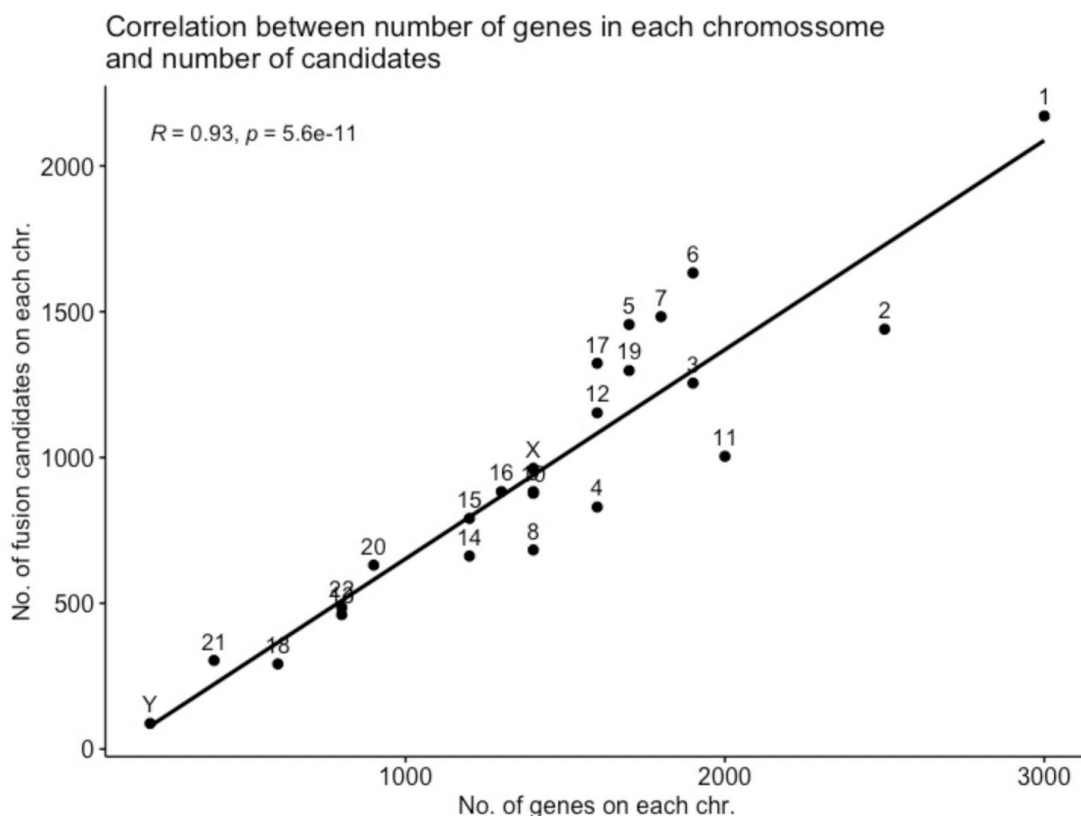


Figura 17 - Correlação entre candidatos em cada cromossomo e número de gene nos cromossomos segundo o NCBI.

Correlação de Pearson apresentada através de um gráfico de dispersão. O número de genes por cromossomo foi coletado do NCBI junto com o número de candidatos de todas as amostras, a normalidade dos dados foi verificada por teste de Shapiro-Wilk, e a correlação de Pearson foi escolhida. Fonte: o autor, 2024.

Entre as características dos candidatos detectados, duas são relevantes para compreender o funcionamento do TS no genoma humano: o frame de leitura das *reads* detectadas e se os candidatos são originados do mesmo cromossomo ou de cromossomos diferentes. Em relação ao frame de leitura, se a maioria dos candidatos apresentar uma estrutura de transcrição intacta, ou seja, estiverem *in-frame*, isso sugere que essas fusões podem ser traduzidas em proteínas. Por outro lado, se o sentido de leitura não estiver intacto, caracterizamos isso como uma fusão *out-of-frame*, indicando que provavelmente esse RNA possui outra função dentro do genoma que não seja a de ser traduzido em uma proteína.

Identificamos que 59,8% dos candidatos estavam *out-of-frame*, enquanto 40,2% apresentavam a característica *in-frame* (Figura 18A). Quanto à natureza da posição dos genes parentais da fusão em relação ao cromossomo de origem, os candidatos poderiam ser do tipo intracromossomal e intercromossomal. Descobrimos que a maioria das fusões ocorre entre genes dentro do mesmo cromossomo, representando 55,3% dos casos. Dentre essas fusões, constatamos que 32,9% ocorrem entre genes considerados vizinhos no contexto genético. Esse dado evidencia que muitos dos casos de TS detectados na literatura ocorrem entre genes que são naturalmente semelhantes devido a eventos evolutivos, como a duplicação genética.

A natureza desses candidatos levanta a questão de saber se genes semelhantes realmente passam pelo processo de fusão ou se são apenas casos de falsos positivos detectados pela metodologia de busca. Por fim, em relação às fusões entre genes de cromossomos diferentes, identificamos que representavam 44,7% dos casos em nossa detecção de candidatos (Figura 18B).

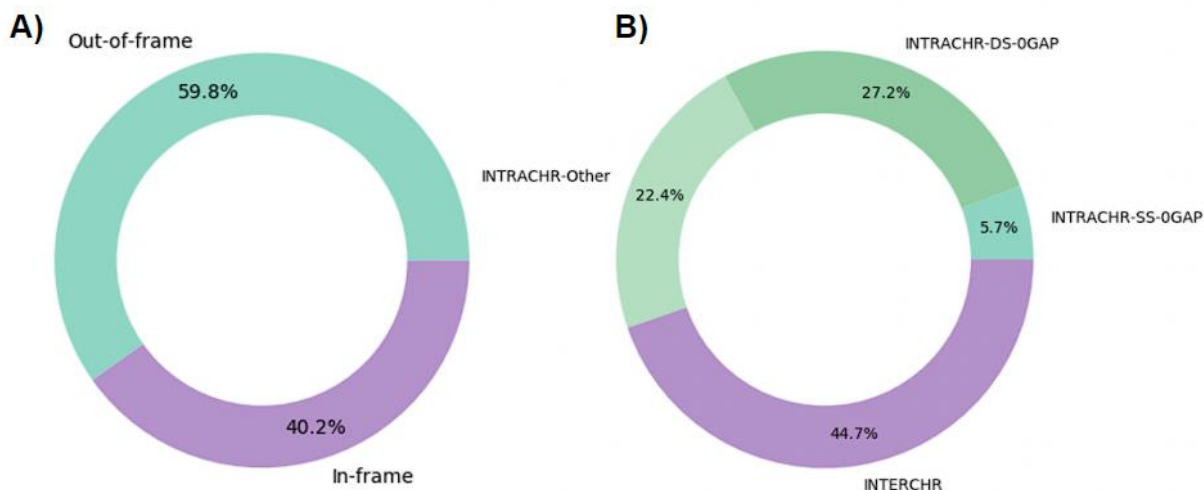


Figura 18 - Características dos candidatos a *Trans-splicing* em relação a capacidade de tradução e distancia genômica entre os pares.

Dois gráficos de pizza para exemplificar a natureza dos candidatos a TS detectados nas amostras. A) *Out-of-frame* e *in-frame* são características de função em relação a capacidade da tradução em proteínas. B) Os dois tipos de INTRACHR-0GAP representam candidatos que estavam a uma distância genômica de 0 gap entre os genes. DS representa funções no mesmo sentido de leitura e SS em sentidos opostos. Fonte: o autor, 2024.

Em seguida, iniciamos o processo de filtragem dos candidatos com base em métricas que ajudam a excluir falsos positivos. Primeiramente, nos concentramos em filtros quantitativos para auxiliar na validação, como exigir um número mínimo de tipos de *reads* nos candidatos que suportam a fusão ou a circularização - *split reads* e *junction reads*. Além disso, os candidatos precisavam aparecer em pelo menos duas réplicas técnicas que estávamos trabalhando. Propriedades de alinhamento que indicavam um possível falso positivo, como alinhamentos em regiões repetitivas do genoma ou com múltiplos *hits*, também foram descartadas. Por fim, aplicamos uma seleção qualitativa dos candidatos resultantes da última etapa. Analisamos os candidatos com base em suas características de alinhamento e visualização, utilizando ferramentas como IGV, bem como os genes de origem dos candidatos.

Para a etapa de validação, aproximadamente 80 candidatos com suporte de várias *reads* e contexto de interesse foram selecionados para validação *in-vitro*. Na Figura 19, há um resumo

das amostras que geraram os candidatos. Foram escolhidos 55 candidatos de TS e 25 candidatos de CircRNA. A amostra ASC apresentou o maior número, com 22 casos, a maioria deles sendo intercromossomal (17). Quanto às amostras Rett, WTNeuro e QXAstro, apresentaram 13 e 12 candidatos cada, com um maior número de RNAs circulares detectadas entre todas as amostras.

A metodologia que identificou o maior número de potenciais candidatos foi ARRIBA, que utiliza como técnica de alinhamento a ferramenta STAR. Contudo, mesmo sendo considerada um padrão ouro na detecção de fusões de genes, identificamos 6 candidatos com alta confiabilidade que a ferramenta de alinhamento não detectou, reforçando a necessidade de uma abordagem metodológica com mais de uma ferramenta ou técnica (HAAS et al., 2019).

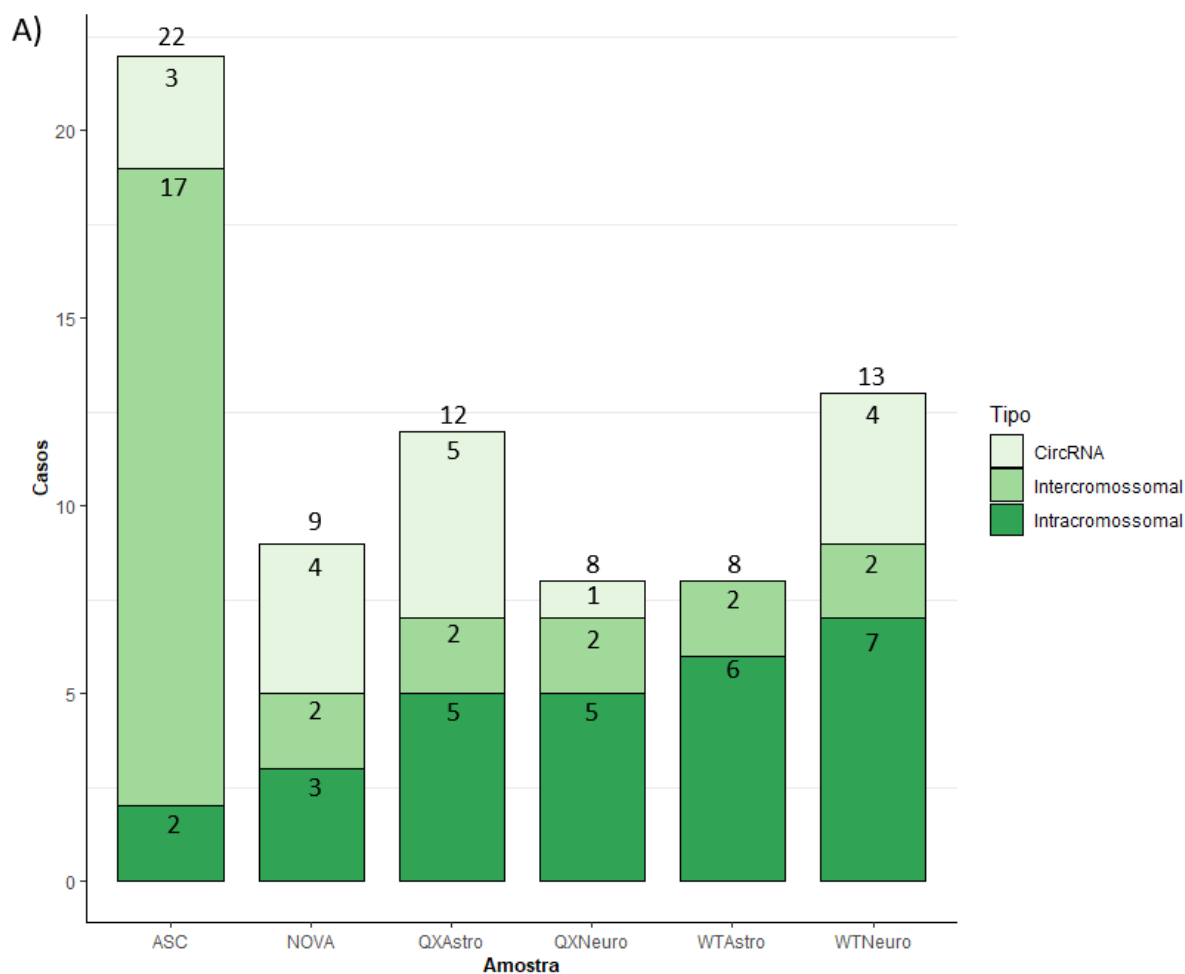


Figura 19 – Resumo em forma gráfica do processo de filtragem *in-silico* para validação.

De 308551 candidatos detectados na primeira etapa desse projeto, ficamos com um número total de 80 candidatos para validação na próxima etapa. A maioria dos candidatos foi eliminada por não apresentar um número mínimo de suporte a fusão/circularização. O segundo filtro que eliminou um número significativo de candidatos foi o de *reads* que alinhavam múltiplas regiões ao mesmo tempo. Fonte: o autor, 2024.

6.4 Validação *in-vitro* dos candidatos

A última etapa de detecção foi a validação *in-vitro* dos candidatos, usando material das mesmas células que geraram os dados para análise. O método escolhido foi a PCR qualitativa, seguida de sequenciamento por Sanger. Primeiramente, realizamos a extração do DNA e RNA de todas as células disponíveis, utilizando os kits descritos no método e seguindo as orientações dos fabricantes. Apenas amostras de DNA e RNA que apresentavam qualidade mínima foram utilizadas, com uma relação 260/280 maior que 1.8 para DNA e 2.0 para RNA.

Para o RNA, realizamos o processo de transformação em cDNA, considerado mais estável para os próximos passos. No PCR qualitativo, testamos temperaturas de anelamento entre 60°C e 64°C, assim como temperaturas de desnaturação (94°C – 98°C) e, principalmente, o tempo de amplificação. Todos os candidatos de *trans-splicing* foram avaliados quanto à presença em RNA, DNA, controle negativo e controle positivo. Quanto aos candidatos de circRNA, verificamos a presença em RNA, controle negativo e controle positivo. Em ambos os protocolos de PCR qualitativo e quantitativo, verificamos o tamanho da banda esperada por meio de gel de agarose a 2%.

Iniciamos o processo com um total de 80 candidatos e, ao final, obtivemos 20 candidatos validados, pelo menos, por meio das técnicas de PCR qualitativo e quantitativo. Dentre esses, foi possível realizar o sequenciamento por Sanger em 8 candidatos. Apesar do número expressivo de candidatos provenientes da amostra ASC, foi constatado que esta apresentou o maior número de falsos positivos, embora o processo de filtragem tenha indicado que os candidatos eram verdadeiros.

Identificamos 12 candidatos de *trans-splicing*, dos quais 7 são de fusão intercromossomais e 5 de fusão intracromossomais, além de 8 candidatos de circRNA. O número de candidatos intercromossomais é surpreendente, pois muitos autores da área têm dificuldade em validar a

maioria dos casos desse tipo, devido à sua considerada maior raridade no genoma (TANG et al., 2019).

Os casos de fusões intracromossômicas demandam uma análise cuidadosa, pois genes colocalizados e conservados podem compartilhar função e, conseqüentemente, sequências de nucleotídeos semelhantes, levantando dúvidas sobre a distância mínima para considerar uma fusão de gene legítima (FOFLONKER; BLABY-HAAS, 2021). Os resultados dessa etapa estão apresentados na Tabela 3, e fotos dos géis, assim como os resultados do sequenciamento por Sanger, estão disponíveis no apêndice deste trabalho, referenciados na mesma Tabela 3.

Tabela 3 - Resumo dos candidatos validados

GENE	Tipo	Amostra	GEL	DNA	SANGER
<i>KANSL1:ARL17A</i>	Intracromossomal	WTA - WTN	✓		✓
<i>FUS:EWSR1</i>	Intercromossomal	ASC	✓	✓	
<i>UQCR10-C1orf194</i>	Intercromossomal	WTN - NOVA	✓		✓
<i>ACTB:NUCB1</i>	Intercromossomal	ASC	✓		✓
<i>CNN3:CALM3</i>	Intercromossomal	ASC	✓		
<i>FBX032:SLBP</i>	Intercromossomal	WTA	✓		✓
<i>COL11A1:TP53BP1</i>	Intercromossomal	QXNeuro	✓		
<i>PODXL2:FYN</i>	Intercromossomal	QXNeuro	✓	✓	✓
<i>RP11-384F7.2:LSMAP</i>	Intracromossomal	NOVA	✓	✓	
<i>RP11-680G10:GSE1</i>	Intracromossomal	NOVA	✓		✓
<i>TPST1:GS1-124K5</i>	Intracromossomal	QXA	✓		✓
<i>COL8A1:CMSS1</i>	Intracromossomal	QXA	✓		
<i>DCBLD2</i>	CircRNA	QXA	✓		
<i>CRIM1</i>	CircRNA	QXA	✓		
<i>ENST00000640700</i>	CircRNA	QXA	✓		
<i>LPAR1</i>	CircRNA	QXA	✓		

<i>AMFR</i>	CircRNA	QXA	✓		
<i>CCDC173</i>	CircRNA	WTN	✓		
<i>MAP2</i>	CircRNA	WTN	✓		✓
<i>ADK</i>	CircRNA	WTN	✓		

Além da métrica qualitativa da presença dos candidatos, também avaliamos a métrica do seu nível de expressão dentro das amostras, em comparação com um gene de homeostase celular. Na Figura 20, apresentamos a expressão dos candidatos, medida por RT-qPCR e normalizada pelo método Delta-Ct. No eixo y, quanto maior o valor do log(delta-ct), menor é a expressão em relação ao gene de controle.

Observamos que os candidatos de circRNA, como *CCDC173* e *ADK*, apresentam um nível de expressão próximo ao do *GAPDH*. Em contrapartida, entre os candidatos de *trans-splicing*, a fusão intercromossomal *CNN3:CALM* foi a mais abundantemente expressa. No entanto, de maneira geral, a expressão desses elementos no genoma é baixa, por dois motivos principais: I) o número de *reads* que suportavam a fusão já era baixo, inclusive de alguns dos genes fora do estado de fusão; e II) esses candidatos possuem um papel de silenciamento de outros elementos no genoma, já que os domínios proteicos dessas fusões estão, em sua maioria, fora do quadro de leitura, possivelmente resultando em uma proteína não funcional.

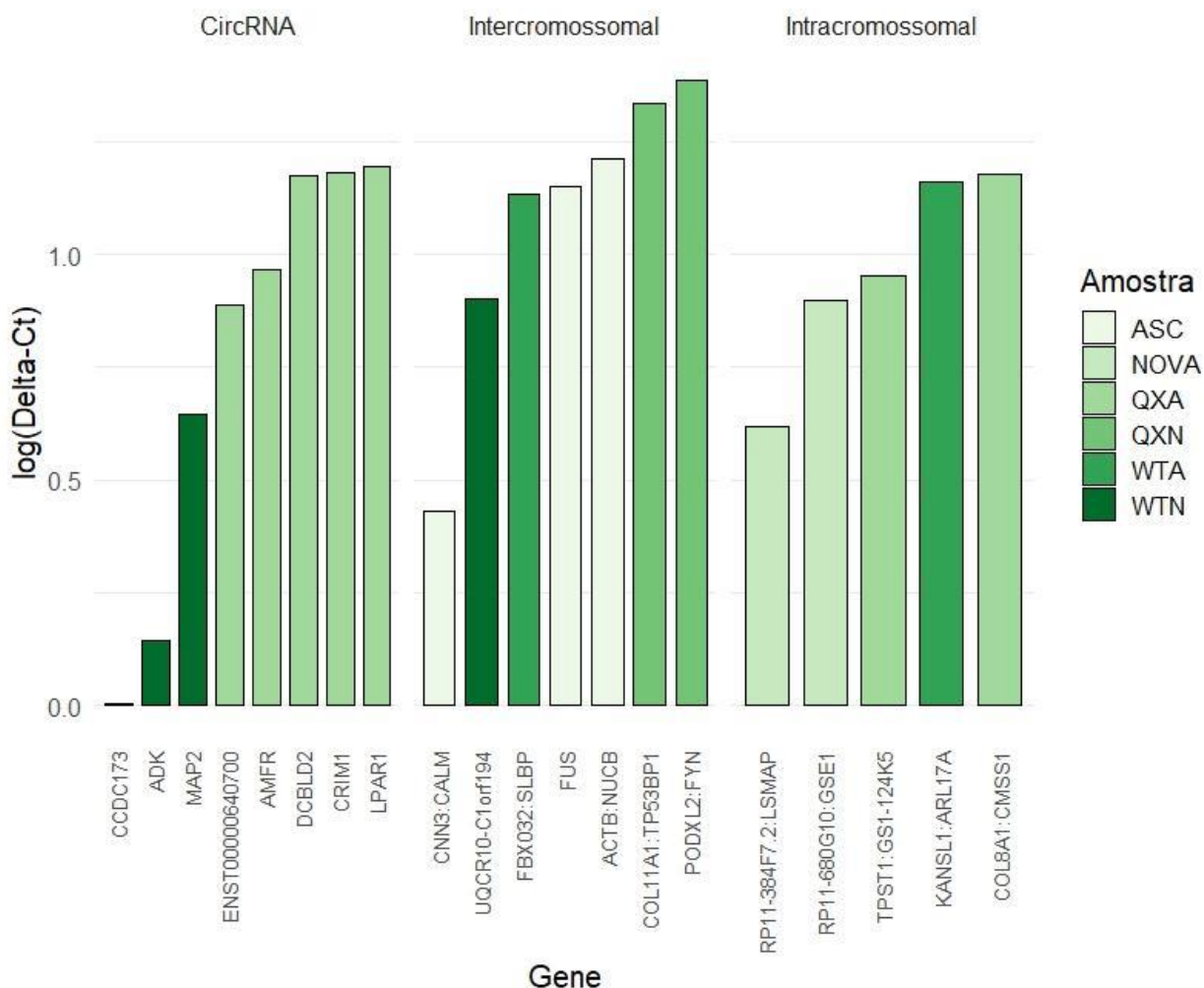


Figura 20 - Gráfico de barras dos níveis de expressão dos candidatos por RT-PCR.

Os níveis de expressão foram medidos pela métrica de ΔCt onde a média entre as réplicas técnicas do ciclo qPCR de cada candidato foi subtraída pela média entre as réplicas técnicas do gene de controle *GAPDH* (YAKOUB; SADEK, 2018). Quanto maior o valor do $\log(\Delta Ct)$, menor é a expressão relativa do candidato ao gene de controle. Aplicamos uma transformação logarítmica nos dados para aproximar da distribuição normal e facilitar a visualização dos dados. Desenvolvido no programa estatístico R usando o pacote ggplot. Fonte: o autor, 2024.

6.5 Caracterização dos candidatos validados

Na última etapa deste projeto, unimos conhecimentos de biologia molecular, genética e técnicas de bioinformática para identificar uma possível função no neurodesenvolvimento desses candidatos. Uma segunda contribuição deste projeto é a confirmação e caracterização de fusões ocorrendo em tecidos não neoplásicos, uma vez que, atualmente, na literatura, acredita-se que o *trans-splicing* seja exclusivo em condições de rearranjos cromossômicos.

Inicialmente, abordamos as fusões de RNA intracromossomais, onde foram identificados 5 casos. Isso foi seguido pelas fusões de RNA intercromossomais, com 7 casos, e, por fim, 8 candidatos de CircRNA. No material em anexo é possível encontrar a tabela suplementar 3 contendo informações dos genes que geraram as fusões, como regiões em que estão no genoma e função de cada gene identificado.

6.5.1 Fusões do mesmo cromossomo

As fusões de um mesmo cromossomo ocorreram em menor número quando comparadas as fusões de cromossomos diferente. Na tabela-4 as informações de que os genes formadores da fusão, e outros dados relevantes estão exemplificados.

Tabela 4 - Características gerais das fusões intracromossômicas

<i>Trans-Splicing</i>	Tipo1- Tipo2	Amostra	Expressão (Delta-ct) em relação a GAPDH	Função dos genes formadores	Descrito Antes?
<i>KANSL1:ARL17A</i>	Protein Coding- Protein Coding	WTA	14,52	Regulação de proliferação celular e transporte intracelular de proteínas	Câncer e Biomarcador
<i>RP11- 384F7.2:LSAMP</i>	lncRNA – Protein coding	NOVA	4,16	Adesão hemofílica durante o neurodesenvolvimento	-
<i>RP11- 680G10.1:GSE1</i>	lncRNA – Protein coding	NOVA	7,90	Possível oncogene no câncer de mama e endometrial	Câncer
<i>TPST1:GS1- 124K5</i>	Protein Coding - lncRNA	QXA	8,93	Síntese de proteínas sulfatadas e associação a carcinoma pulmonar	-
<i>COL8A1:CMSS1</i>	Protein Coding- Protein Coding	QXA	15,08	Síntese de colágeno e ligações de ácidos nucleicos e RNA	-

A fusão *KANSL1:ARL17A* foi detectada em duas amostras (WT-Astro e WT-Neuron) de astrócitos e neurônios do caso controle para a mutação *MECP2*. O gene *KANSL1* está posicionado adjacente ao gene *ARL17A*, e essa fusão pode representar uma continuação de leitura, já que o intervalo entre esses dois genes é de apenas 476.306 kb. No entanto, é importante notar que esses genes desempenham funções diferentes.

O *KANSL1* está envolvido na acetilação de histonas e atua como regulador da proliferação celular e mitose. Mutações nesse gene foram associadas à síndrome neurológica Koolen-de

Vries (KdV), que afeta o desenvolvimento intelectual (KEEN et al., 2017). Por outro lado, o gene *ARL17A* é responsável pelo transporte intracelular de proteínas e vesículas. Essa fusão foi previamente identificada em casos de linfoma linfoblástico (LÓPEZ-NIEVA et al., 2019) e como biomarcador para um gene de predisposição ao câncer na população de origem europeia (ZHOU et al., 2017).

A fusão *RP11-384F7.2:LSAMP* foi detectada em todas as amostras de organoides analisadas neste projeto. O gene *RP11-384F7.2* é classificado como um RNA longo não codificante (lncRNA), e sua função ainda não foi validada experimentalmente. A única informação encontrada até o momento é que ele atua como um promotor de outro gene, *LSAMP*, que é precisamente o parceiro de fusão identificado. Por sua vez, o gene *LSAMP* é codificador de proteínas e desempenha um papel crucial na superfície neuronal, facilitando a seleção de adesão hemofílica durante o neurodesenvolvimento. Embora essa fusão ainda não tenha sido relatada na literatura, a interação entre esses dois genes é reconhecida pelo fator de enriquecimento da expressão de *LSAMP* por *RP11-384F7.2*. Além disso, *LSAMP* foi caracterizado como um supressor tumoral do neuroblastoma e associado a polimorfismos genéticos em casos de esquizofrenia (KOIDO et al., 2014; MARTINEZ-MONLEON et al., 2023).

A fusão *RP11-680G10.1:GSE1* foi identificada em três amostras diferentes: NOVA, WT-Neuron e QX-Neuron. O gene *RP11-680G10.1* está localizado adjacente ao gene *GSE1*, com uma distância de apenas 290,913 kb, o que sugere a possibilidade de uma fusão de leitura direta entre eles. O *RP11-680G10.1* é classificado como um RNA longo não codificante (lncRNA) e está situado em uma região promotora, sugerindo que pode atuar como um promotor de outros genes. Por outro lado, o gene *GSE1* codifica uma proteína e desempenha um papel no sistema endossomal e em processos relacionados ao transporte intracelular.

O gene *GSE1*, por sua vez, é um gene codificador de proteínas que foi implicado em diversos casos de câncer no sistema nervoso, incluindo meningioma e meduloblastoma (WONG et al., 2020; YOUNGBLOOD; GÜNEL, 2020). Estudos demonstraram que a proteína GSE1 está envolvida na formação de complexos de acetilase de histonas, o que sugere seu papel na regulação da expressão gênica por meio da modificação da cromatina. Além disso, a fusão *RP11-680G10.1:GSE1* também foi identificada em um estudo anterior conduzido por Yao et al. (2019), que investigou diversas fusões em câncer endometrial, destacando a relevância potencial dessa fusão em diferentes contextos biológicos e patológicos.

A identificação da fusão *TPST1:GS1-124K5* na amostra QX-astro, astrócitos que apresentavam a mutação no gene *MECP2*, é um achado interessante. O gene *TPST1* é conhecido por codificar uma enzima chamada tirosil proteína sulfotransferase 1, que desempenha um papel crucial na síntese de proteínas sulfatadas. Essas proteínas sulfatadas têm várias funções biológicas importantes, incluindo a regulação da interação entre proteínas e moléculas de sinalização extracelular. Estudos também relacionaram o gene *TPST1* a certos tipos de câncer, como carcinoma pulmonar, sugerindo um possível papel desse gene na carcinogênese.

Entre as fusões de genes de um mesmo cromossomo, a *TPST1:GS1-124K5* foi a que apresentou a menor distância entre os genes: 208,548kb. A proximidade dos genes envolvidos na fusão no mesmo cromossomo, juntamente com a associação com a mutação no gene *MECP2*, pode sugerir um potencial papel desses eventos na patogênese ou na regulação de processos biológicos específicos nos astrócitos. A compreensão dos efeitos dessa fusão e sua relação com a mutação do gene *MECP2* pode fornecer insights importantes sobre os mecanismos subjacentes às condições neurológicas associadas a essa mutação, como a síndrome de Rett.

O fato de o gene *GS1-124K5* ser um *lncRNA* sem funções moleculares documentadas até o momento ressalta ainda mais a importância de investigar o papel dessa fusão e seu possível

envolvimento em condições neurológicas. FAN et al. (2021) caracterizaram esse gene como fator de risco para o desenvolvimento da doença de Parkinson em pacientes com menos de 50 anos. A caracterização dessa fusão pode abrir novas perspectivas na compreensão dos mecanismos moleculares subjacentes à doença de Parkinson e outras condições neurológicas associadas. Essa fusão é inédita e ainda não descrita em dados documentados na literatura científica.

Finalmente, a fusão *COL8A1:CMSS1* foi identificada em uma única amostra, denominada QX-astro, com uma distância aproximada de 401,954kb entre os dois genes. O gene *COL8A1*, responsável pela codificação de uma das duas cadeias de colágeno tipo VIII, desempenha um papel crucial na formação da membrana basal do endotélio. Em contraste, o gene *CMSS1*, embora seja codificador de proteínas, carece de informações detalhadas sobre sua função fisiológica, limitando-se a dados sobre ligações de ácidos nucleicos e RNA. Embora esta fusão esteja registrada em alguns bancos de dados de *trans-splicing*, sua caracterização na literatura científica permanece pendente. E, embora haja informações limitadas sobre a função do gene *CMSS1*, sua associação com ácidos nucleicos e RNA sugere um possível papel na regulação da expressão gênica ou em processos moleculares relacionados ao RNA. A detecção dessa fusão em uma amostra de astrócitos pode indicar sua relevância potencial no contexto do sistema nervoso central.

As fusões intracromossômicas identificadas neste estudo exibiram algumas características distintivas, incluindo: I) fusões entre genes próximos; II) fusões entre genes de regiões promotoras; e III) fusões detectadas em todas as amostras examinadas. No que diz respeito ao primeiro ponto, observa-se claramente o impacto da duplicação gênica na detecção de *trans-splicing*. Esses casos, caracterizados por sequências com alta similaridade, são interpretados

como uma fusão de dois genes, embora possa ser o resultado da incapacidade dos alinhadores em distinguir adequadamente a região esperada para cada gene.

A detecção de fusões intracromossômicas entre genes próximos é interessante e pode ser atribuída à proximidade física desses genes no genoma. A duplicação de genes pode resultar em regiões genômicas com alta similaridade, levando a alinhamentos ambíguos ou mistos durante o processo de detecção de *trans-splicing*. Isso pode resultar em uma aparência de fusão entre dois genes adjacentes, quando na verdade são apenas regiões semelhantes sendo interpretadas erroneamente como uma fusão. Essa observação destaca a importância de considerar a estrutura genômica e a organização dos genes ao interpretar os resultados de detecção de fusões intracromossômicas. Além disso, a detecção dessas fusões em todas as amostras testadas sugere uma possível relevância biológica desses eventos no contexto específico das células ou tecidos investigados.

No segundo ponto, fusões entre genes que compartilham regiões promotoras sugerem uma relação funcional entre os dois genes, onde o transcrito resultante pode desempenhar um papel importante na regulação da expressão do gene principal. Nesses casos, a fusão pode não gerar uma molécula completamente nova, mas sim um transcrito necessário para o funcionamento adequado do gene alvo. No entanto, o conhecimento da existência desses promotores pode representar um potencial terapêutico para certas doenças do neurodesenvolvimento, como exemplificado pelo caso de *RP11-384F7.2:LSAMP*. Além disso, a detecção de fusões presentes em todas as amostras testadas sugere que esses eventos de *trans-splicing* podem ser comuns ou até mesmo essenciais para o funcionamento normal das células ou tecidos investigados. Isso ressalta a importância de estudar o *trans-splicing* como um mecanismo de regulação gênica e como forma de diversificação do transcrito em diferentes contextos biológicos.

6.5.2 Fusões de cromossomos diferentes

As fusões de cromossomos diferentes ocorreram em maior número quando comparadas as fusões de um mesmo cromossomo. Na tabela-5 as informações de que os genes formadores da fusão, e outros dados relevantes estão exemplificados.

Tabela 5 - Características gerais das fusões intercromossomais.

<i>Trans-Splicing</i>	Tipo1-Tipo2	Amostra	Expressão (Delta-ct)	Função dos genes formadores	Descrito Antes?
<i>FUS:EWSR1</i>	Protein Coding-Protein Coding	ASC	14,13	Processamento de pré-mRNA e regulação da expressão genica	Neoplasia epitelióide
<i>UQCR10:C1orf194</i>	Protein Coding-Protein Coding	WTN	7,99	Subunidade do complexo mitocondrial, transporte celular	Ependimoma
<i>ACTB:NUCB1</i>	Protein Coding-Protein Coding	ASC	16,29	Sinalização intracelular e proteína de ligação de cálcio	-
<i>CNN3:CALM3</i>	Protein Coding-Protein Coding	ASC	2,70	Regulação durante remodelagem neuronal, plasticidade dos dendrocitos e ligações com cálcio	-
<i>FBX032:SLBP</i>	Protein Coding-Protein Coding	QXA	13,57	Processamento eficiente e transporte de histonas	-
<i>COL11A1:TP53BP1</i>	Protein Coding-Protein Coding	QXN	21,67	Reparação de DNA e limitação de recombinação homóloga	-
<i>PODXL2:FYN</i>	Protein Coding-Protein Coding	QXN	24,35	Controle do crescimento celular e proteína de superfície celular	-

A primeira fusão intercromossômica detectada foi a *FUS:EWSR1*, principalmente nas amostras ASC. No entanto, algumas variações contendo o gene *FUS* ou *EWSR1* também foram identificadas em outras amostras, como WT-Astro e WT-Neuron. O gene *FUS*, localizado no cromossomo 16, é um gene codificador de proteínas que desempenha um papel crucial no processamento de pré-mRNA e na exportação do mRNA maduro para o citoplasma. Ele integra um complexo ribonucleoproteico conhecido como hnRNP e, portanto, está envolvido em diversos processos celulares.

O mesmo padrão é observado para o gene *EWSR1*, localizado no cromossomo 22, que codifica uma proteína multifuncional responsável por regular a expressão gênica, a sinalização celular e o processamento e transporte de RNA. Essa fusão já foi documentada como uma causa de translocações cromossômicas entre esses genes e outros fatores de transcrição, resultando na formação de proteínas quiméricas, particularmente em casos de neoplasias epitelioides (DERMAWAN et al., 2022).

A fusão *UQCR10:C1orf194* foi detectada em apenas uma amostra, proveniente de neurônios do grupo controle para a mutação em *MECP2*. O gene *UQCR10*, localizado no cromossomo 22, é responsável pela codificação de uma proteína que integra a subunidade do complexo mitocondrial III, formando o segmento médio da cadeia respiratória na membrana mitocondrial. Por outro lado, *C1orf194*, presente no cromossomo 1, também é um gene codificador de proteínas, embora sua função ainda não tenha sido completamente elucidada. Entretanto, evidências preliminares sugerem que este gene desempenha um papel no transporte e exportação de heme no citoplasma e no citoesqueleto celular. Essa fusão já foi previamente caracterizada em casos de câncer de ependimoma da medula espinhal, onde foi identificada como uma translocação entre o cromossomo 22 e o cromossomo 1 (OLSEN et al., 2016).

A fusão *ACTB:NUCB1* foi identificada nas amostras de células estaminais adultas (ASC), que serviram como grupo controle para o gene *NOVA-1*. O gene *ACTB*, localizado no cromossomo 7, codifica uma proteína de actina essencial para a estrutura celular, integridade e sinalização intercelular. Mutações neste gene foram previamente associadas à síndrome de Baraitser-Winter 1, caracterizada por déficits intelectuais e fenotipo facial específico em pacientes (YATES et al., 2017).

O gene *NUCB1*, localizado no cromossomo 17, desempenha uma função semelhante ao seu par, sendo responsável pela regulação da homeostase de cálcio no complexo de Golgi e participando ativamente na sinalização celular. Até o momento, essa fusão específica entre *ACTB* e *NUCB1* não foi descrita na literatura. No entanto, o gene *ACTB* já foi mencionado em outros estudos, possivelmente devido à sua função estrutural relacionada à actina.

A fusão *CNN3:CALM3* foi identificada em amostras de células estaminais adultas (ASC). O gene *CNN3*, localizado no cromossomo 1, codifica a proteína Calponina 3, que desempenha um papel potencial na regulação dos filamentos de actina durante a remodelação neuronal no cérebro. Foi encontrado também em espinhas dendríticas de neurônios do hipocampo adulto, onde regula a plasticidade dendrítica. A eliminação sistemática do gene *Cnn3* em camundongos resulta em letalidade embrionária e neonatal devido a defeitos no desenvolvimento do sistema nervoso central (FERHAT et al., 2003; RAMI et al., 2006; FLEMMING et al., 2015). Por outro lado, o gene *CALM3* codifica uma proteína que atua em ligações de cálcio e como cofator enzimático, sendo essencial para a regulação do ciclo celular e citocinese. Até o momento, essa fusão específica entre *CNN3* e *CALM3* não foi caracterizada na literatura.

A fusão *FBXO32:SLBP* foi identificada em apenas uma amostra - WT Astrócito. O gene *FBXO32*, localizado no cromossomo 8, codifica uma proteína da família F-BOX, a qual é encontrada expressa em níveis elevados em casos de atrofia muscular. Por outro lado, o gene

SLBP, situado no cromossomo 4, codifica uma proteína que se liga a estruturas de loop de histonas de RNA. Essa função é essencial para o processamento, transporte e tradução eficientes dessa classe de RNA. Até o momento, essa fusão específica entre *FBXO32* e *SLBP* ainda não foi descrita na literatura.

A fusão *COL11A1:TP53BP1* foi detectada em apenas uma amostra - QX Neurônios. O gene *COL11A1*, localizado no cromossomo 1, codifica uma das cadeias alfa do colágeno tipo XI, que é essencial para a estruturação do colágeno fibrilar menor. Mutações neste gene estão associadas à síndrome de Stickler-Marshall, ambas causando problemas de visão, audição e anomalias articulares e faciais no tecido conjuntivo (O'BRIEN; PHILLIPS, 2021). Por outro lado, o gene *TP53BP1*, situado no cromossomo 15, codifica uma proteína envolvida na reparação de quebras de dupla hélice de DNA, promovendo a ligação de extremidades não homólogas. Até o momento, essa fusão específica entre *COL11A1* e *TP53BP1* ainda não foi descrita na literatura.

A fusão entre os genes *PODXL2* e *FYN* representa uma interessante combinação de proteínas envolvidas na regulação do crescimento celular e na interação com a superfície celular. O gene *PODXL2* codifica uma proteína de superfície celular pertencente à família CD34, que desempenha papéis importantes na adesão celular, migração e diferenciação. Por outro lado, o gene *FYN* codifica uma tirosina quinase associada à membrana, implicada no controle do crescimento celular, na diferenciação celular e na transdução de sinais.

As fusões intercromossômicas validadas apresentaram algumas características distintas das fusões intracromossômicas discutidas anteriormente. Notavelmente, nenhum dos genes envolvidos foi identificado como um promotor ou uma região de lncRNA. Todos os genes associados a fusões intercromossômicas codificam proteínas funcionais com funções bem definidas no organismo, o que descarta a possibilidade de uma fusão ser meramente uma continuação da leitura ou resultar de regiões com alta similaridade de sequência. Apenas duas

das oito fusões apresentaram dados previamente observados na literatura, destacando a dificuldade em validar esse tipo de fusão e a preferência de outros autores por focar apenas em fusões intracromossômicas (TANG et al., 2019a).

6.5.3 RNAs Circulares

Os RNAs circulares identificados neste estudo foram investigados para auxiliar na compreensão do contexto do neurodesenvolvimento em que estão envolvidos. Ao contrário dos casos de fusão, os RNAs circulares já foram amplamente caracterizados em vários tipos de tecidos, e sua detecção não é uma novidade fora do contexto neoplásico. No entanto, a maioria dos estudos que demonstram a presença de circRNAs são baseados em linhagens celulares hESCs e Hs68, conhecidas por apresentarem diversos problemas de genotipagem. Neste estudo, identificamos quais circRNAs estavam presentes em nossas amostras, determinando se já foram caracterizados em tecido do córtex cerebral ou não, e avaliando seu impacto, de acordo com a literatura, nas condições estudadas neste projeto. Os resultados e informações levantadas estão resumidos na Tabela-6.

Tabela 6 - Características gerais dos RNA circulares.

Gene	Região (chr/pos)	CircID	Expressão (Delta-ct)	Função
<i>DCBLD2</i>	chr3:98536612	hsa_circ_15296	14,93	proteína envolvida em regulação negativa do crescimento celular e restauro de feridas
<i>CRIM1</i>	chr2:36623756	hsa_circ_16396	15,18	proteína transmembrana ligante para fator de crescimento
ENST00000640700	chr12:52692605	-	7,73	LncRNA
<i>LPAR1</i>	chr9:113703700	hsa_circ_06523	15,67	codifica receptor de ácido lisofosfatídico (LPA) - sinalização celular
<i>AMFR</i>	chr16:56401355	hsa_circ_07148	9,21	codifica receptor transmembrana do fator de motilidade autocrio
<i>CCDC173</i>	chr2:170505691	hsa_circ_01594	1,01	proteína associada a cílios e flagelos nos testículos
<i>MAP2</i>	chr2:210557348	hsa_circ_28398	4,43	proteínas associadas a microtúbulos essencial na neurogênese. proteínas cito esqueléticas neuronais específicas que são enriquecidas em dendritos, implicando um papel na determinação e estabilização da forma dendrítica durante o desenvolvimento neuronal
<i>ADK</i>	chr10:76153898	hsa_circ_25380	1,39	enzima responsável pela transformação do ATP em adenosina

O RNA circular *DCBLD2*, localizado no cromossomo 3, foi identificado em amostras de QX-Astrócito, e já foi documentado que seu transcrito linear de RNA tem a capacidade de se circularizar. O gene *DCBLD2*, em seu estado linear, codifica uma proteína envolvida na regulação negativa do crescimento celular e no reparo de danos. No entanto, quando em sua forma circular, essa molécula provavelmente não codifica uma proteína, uma vez que sua sequência não possui uma região de leitura aberta (ORF) e seu tamanho seria menor do que 100 aa. Este RNA circular já foi descrito anteriormente no tecido cerebral sem condições síndrômicas e em alguns casos de oligodendroma (ZHANG et al., 2014; SALZMAN, 2016).

O RNA circular *CRIM1*, localizado no cromossomo 2, também foi caracterizado em amostras de QX-Astrócito, e seu transcrito circular já foi descrito anteriormente. O gene *CRIM1*, em sua forma linear, codifica uma proteína transmembrana envolvida no fator de crescimento celular. Em sua forma circular, foi observada a capacidade de codificação de proteína, devido à presença de uma região de leitura aberta (ORF) em sua sequência de RNA, juntamente com diversos domínios conservados. Este RNA circular já foi previamente documentado em casos de glioblastoma e oligodendroma (SONG et al., 2016).

O circRNA ENST00000640700, localizado no cromossomo 12, também foi identificado em amostras de QX-Astrócito. Até o momento, não foi descrito na literatura como um RNA capaz de circularização. Este gene representa um *lncRNA*, considerado um transcrito novo, e ainda não possui função descrita nos bancos de dados. Não foram encontradas informações sobre a presença de ORF ou domínios conservados, uma vez que se trata de um *lncRNA*.

O circRNA *LAPR1*, localizado no cromossomo 9, foi caracterizado em amostras de QX-Astrócito, e seu transcrito circular já foi descrito. O gene *LAPR1* codifica uma proteína que atua como receptor de ácido lisofosfatídico, desempenhando um papel importante na sinalização

celular. No entanto, a probabilidade deste RNA circular gerar uma proteína é baixa, uma vez que apresenta vários códons de parada e um tamanho pequeno. O circRNA *LAPR1* já foi identificado como um promotor de proliferação e diferenciação osteogênica em células-tronco, através da repressão do gene *PTEN* (XIN et al., 2021).

O circRNA *AMFR*, localizado no cromossomo 16, foi o último a ser caracterizado em QX-Astrócito, e seu transcrito circular já foi descrito. O gene *AMFR*, em sua forma linear, codifica uma proteína transmembrana glicosilada que atua como receptor para o fator de mobilidade autócrina (AMF), estimulando assim a motilidade tumoral, particularmente quando secretada por células tumorais. No entanto, este circRNA tem uma baixa probabilidade de ser traduzido em uma proteína, uma vez que não apresenta uma região de leitura aberta (ORF) estruturada. Este transcrito já foi identificado em casos de carcinoma de células escamosas de cabeça e pescoço (SHEN; WANG; YE, 2021).

O circRNA *CCDC173*, localizado no cromossomo 2, foi caracterizado em WT-Neurônios, e seu transcrito circular já foi descrito. O gene *CCDC173*, em sua forma linear, codifica uma proteína associada a cílios e flagelos. Assim como os casos anteriores, este circRNA também apresenta uma baixa probabilidade de codificar uma proteína, pois não possui uma região de leitura aberta (ORF) significativa. Este transcrito foi identificado como marcador e alvo terapêutico para hepatoblastoma (SONG et al., 2016).

O circRNA *MAP2*, localizado no cromossomo 2, foi identificado como um candidato em WT-Neurônios e foi descrito como capaz de circularização. O gene *MAP2*, em sua forma linear, codifica uma proteína que pertence à família de proteínas associadas a microtúbulos. Essas proteínas desempenham um papel crucial na montagem de microtúbulos, um processo essencial na neurogênese. Em roedores, produtos de genes semelhantes são proteínas específicas do citoesqueleto neuronal, enriquecidas em dendritos, sugerindo um papel na determinação e

estabilização da forma dendrítica durante o desenvolvimento neuronal. No entanto, este circRNA não apresenta uma região de leitura aberta (ORF), o que sugere uma baixa probabilidade de codificar uma proteína. Embora este RNA circular tenha sido previamente detectado em outros estudos, esta é a primeira vez que ele é observado em um contexto de neurônios e neurodesenvolvimento (SONG et al., 2016).

O último circRNA identificado foi o *ADK*, localizado no cromossomo 10, caracterizado na amostra WT-Neurônios e foi descrito como capaz de circularização. O gene *ADK*, em sua forma linear, codifica uma enzima responsável pela conversão de ATP em adenosina. Este circRNA apresenta uma região de leitura aberta (ORF) estruturada, porém curta, o que sugere uma baixa probabilidade de ser transcrito em uma proteína funcional. Este RNA circular já foi descrito anteriormente em casos de glioblastoma e oligodendroma (SONG et al., 2016).

O RNA circular *MAP2* emerge como um dos principais focos de estudo nesse projeto, pois ele já foi caracterizado como um regulador no neurodesenvolvimento, atuando em diversos níveis de regulação gênica e proteica. Das três principais funções desenvolvida por ele estão: 1) Morfologia neuronal; 2) formação de sinapses e 3) plasticidade sináptica. O nível de expressão ou a sua presença nesse contexto precisa ser investigado com maior profundidade, uma vez que não consideramos que o desenho de experimento abordado aqui tenha sido o ideal para medir a expressão de um RNA circular.

De maneira geral, os elementos não canônicos de RNA, como o TS e o CircRNA, são de complexa detecção e validação. O tipo de ferramenta aplicado, a metodologia de filtro e pôr fim a validação escolhida devem ser revistos em todo novo experimento desse tipo. Pequenas variações em qualquer uma dessas técnicas pode impactar o resultado final obtido pelo pesquisador. E isso não exclui essa presente tese. Na parte de detecção, novas tecnologias emergentes devem auxiliar a tratar as reads discordantes e a quantidade de dados analisados. Os

métodos de filtros aplicados aqui são limitados ao conhecimento prévio do que deveriam ser a estrutura desses RNAs, uma nova descoberta da sua biogênese pode influenciar na descoberta de novos tipos de moléculas. E por fim, a validação precisa de auxílio de novas técnicas de biologia molecular, uma vez a variabilidade desse tipo de experimento é conhecida e pode induzir a identificação de falsos-positivos uma vez que são casos raros no neurodesenvolvimento.

7 CONCLUSÃO

A presente tese tinha como objetivo principal enriquecer o conhecimento genético sobre dois tipos de molécula no neurodesenvolvimento. Para tal, foi necessário validar quais técnicas poderiam simular o desenvolvimento do córtex cerebral, quais ferramentas seriam capazes de identificar e validar os casos de interesse, bem como o possível papel dessas moléculas no genótipo e fenótipo estudado. Aqui estão as principais conclusões:

- Existem técnicas capazes de detectar moléculas de *Trans-splicing* e circRNA em dados de RNA-seq. Com as ferramentas ARRIBA, JAFFA e a nossa técnica de detecção, identificamos mais de 300.000 mil candidatos de *TS* e 400 candidatos de circRNA.
- Desenvolvemos uma técnica de validação de ferramentas para *TS* através de um script capaz de simular *reads* contendo características dessa detecção. A ferramenta em Python está disponível *online* no GitHub ([link](#)).
- Durante a etapa de filtragem de candidatos *in silico*, foi possível reduzir o número de candidatos falsos-positivos, dadas as características das fusões ou das circularizações, em mais de 95%. A nossa ferramenta auxiliar desenvolvida nessa tese pôde ajudar na detecção mais livre de hipóteses e filtros que podem impactar a detecção de novos candidatos.

- Os organoides corticais cultivados nesta tese do gene *NOVA1* mimetizaram o desenvolvimento do córtex cerebral. Observamos camadas estruturais típicas do neurodesenvolvimento e validamos a sua função molecular com a presença marcadores de MAP2, NES e DAPI.
- Existem casos de *TS* em tecido do desenvolvimento típico. Identificamos 13 fusões a nível de RNA que estão presentes no tecido do córtex cerebral. Evidenciamos que esse tipo de molécula não é exclusivo de casos de câncer e pode desenvolver um papel no neurodesenvolvimento, como é o caso de 5 candidatos que identificamos como responsáveis por alguma função para a homeostase do cérebro.
- Existem casos de CircRNA em tecido do córtex cerebral. Identificamos 8 candidatos que são inéditos em organoides e que podem ter um impacto no neurodesenvolvimento - como o caso de *MAP2* na neurogenese.
- Não encontramos correlação direta entre as mutações em *MECP2* e a biogênese de moléculas de *trans-splicing* ou circRNA. Mais estudos nessa área são necessários.
- Não encontramos correlação direta entre as mutações em *NOVA1* e a biogênese de moléculas de *trans-splicing* ou circRNA. Mais estudos nessa área são necessários.

8 CONSIDERAÇÕES FINAIS

O estudo de moléculas de TS e CircRNA é um tópico que se preocupa com o estudo de casos raros e de impacto incerto na genética humana. Ter a capacidade tecnológica de conseguir investigar nesse nível é um privilégio científico. Nesse trabalho conseguimos identificar e validar a presença de moléculas que até alguns anos atrás eram desconhecidas ao homem. Contudo, essa exploração também ainda leva a novas perguntas, como a principal delas como a sua função biológica. Acreditamos que essa tese sirva como ponto de partida para novas explorações na área de neurodesenvolvimento, em especial dos casos afetados por um desbalanço desse processo.

As técnicas aqui apresentadas têm potencial de auxiliar a investigação nesse campo. Contudo ainda está em aberto qual seria o padrão ouro na detecção desse tipo de moléculas de RNA. Não conseguimos propor uma forma sensível o bastante para facilitar o trabalho de qualquer investigador molecular. A taxa de validação *in-vitro* desse projeto ainda pode ser considerada baixa, principalmente pela dificuldade em levantar casos verdadeiros positivos nas amostras coletadas. Os próximos esforços precisam ser direcionados em anotar maiores características dos casos verdadeiros, e incorporar essas características nas ferramentas de busca, aumentando assim a sensibilidade e especificidade do método.

REFERÊNCIAS

- AMIR, R. E. et al. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics* 1999 23:2, v. 23, n. 2, p. 185–188, out. 1999. Disponível em: <https://www.nature.com/articles/ng1099_185>. Acesso em: 10 jun. 2022.
- BABICEANU, M. et al. Recurrent chimeric fusion RNAs in non-cancer tissues and cells. *Nucleic acids research*, v. 44, n. 6, p. 2859–2872, 2 fev. 2016. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/26837576/>>. Acesso em: 7 jun. 2022.
- BARUZZO, G. et al. Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nature Methods*, v. 14, n. 2, p. 135–139, 1 fev. 2017. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/272058/>>. Acesso em: 20 jun. 2020.
- BENOIT-PILVEN, C. et al. Complementarity of assembly-first and mapping-first approaches for alternative splicing annotation and differential analysis from RNAseq data. *Scientific Reports*, v. 8, n. 1, p. 1–13, 1 dez. 2018. Disponível em: <<https://www.nature.com/articles/s41598-018-21770-7>>. Acesso em: 20 jun. 2020.
- BINDER, D. K.; SCHARFMAN, H. E. Brain-derived Neurotrophic Factor. *Growth factors (Chur, Switzerland)*, v. 22, n. 3, p. 123, set. 2004. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/154526/>>. Acesso em: 7 abr. 2023.
- BUKANOVICH, R. J.; DARNELL, R. B. The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Molecular and cellular biology*, v. 17, n. 6, p. 3194–3201, 1 jun. 1997. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/9154818/>>. Acesso em: 7 abr. 2023.
- CARRARA, M. et al. State-of-the-Art Fusion-Finder Algorithms Sensitivity and Specificity. *BioMed research international*, v. 2013, p. 340620, 17 fev. 2013.
- CHAHROUR, M. et al. MeCP2, a Key Contributor to Neurological Disease, Activates and Represses Transcription. *Science (New York, N.Y.)*, v. 320, n. 5880, p. 1224, 5 maio 2008. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/1843785/>>. Acesso em: 7 abr. 2023.
- CHEN, Y. J. et al. Genome-wide, integrative analysis of circular RNA dysregulation and the corresponding circular RNAmicroRNA-mRNA regulatory axes in autism. *Genome Research*, v. 30, n. 3, p. 375–391, 3 mar. 2020. Disponível em: <<https://genome.cshlp.org/content/early/2020/03/03/gr.255463.119>>. Acesso em: 23 fev. 2021.
- CONN, V. M. et al. A circRNA from SEPALLATA3 regulates splicing of its cognate mRNA through R-loop formation. *Nature plants*, v. 3, 18 abr. 2017. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/28418376/>>. Acesso em: 7 abr. 2023.
- DAVIDSON, N. M.; MAJEWSKI, I. J.; OSHLACK, A. JAFFA: High sensitivity transcriptome-focused fusion gene detection. *Genome Medicine*, v. 7, n. 1, p. 43, 11 dez. 2015. Disponível em: <<http://genomemedicine.com/content/7/1/43>>. Acesso em: 20 jun. 2020.
- DERMAWAN, J. K. et al. Comprehensive genomic profiling of EWSR1/FUS-CREB translocation-associated tumors uncovers prognostically significant recurrent genetic alterations

and methylation-transcriptional correlates. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc.*, v. 35, n. 8, p. 1055, 1 ago. 2022. Disponível em: </pmc/articles/PMC9329182/>. Acesso em: 8 abr. 2023.

DOBIN, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, v. 29, n. 1, p. 15–21, 25 out. 2012. Disponível em: <https://doi.org/10.1093/bioinformatics/bts635>.

FERHAT, L. et al. Increased levels of acidic calponin during dendritic spine plasticity after pilocarpine-induced seizures. *Hippocampus*, v. 13, n. 7, p. 845–858, 1 jan. 2003. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1002/hipo.10136>. Acesso em: 8 abr. 2023.

FLEMMING, A. et al. A Conditional Knockout Mouse Model Reveals That Calponin-3 Is Dispensable for Early B Cell Development. *PLOS ONE*, v. 10, n. 6, p. e0128385, 5 jun. 2015. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0128385>. Acesso em: 8 abr. 2023.

GIANDOMENICO, S. L.; SUTCLIFFE, M.; LANCASTER, M. A. Generation and long-term culture of advanced cerebral organoids for studying later stages of neural development. *Nature protocols*, v. 16, n. 2, p. 579–602, 1 fev. 2021. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/33328611/>. Acesso em: 7 abr. 2023.

GUERRA, B. S. et al. *Biogenesis of circular RNAs and their role in cellular and molecular phenotypes of neurological disorders* *Seminars in Cell and Developmental Biology* Elsevier Ltd, , 3 set. 2020. .

HAAS, B. J. et al. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biology*, v. 20, n. 1, p. 1–16, 21 out. 2019. Disponível em: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1842-9>. Acesso em: 5 maio. 2022.

HERAI, R. H.; YAMAGISHI, M. E. B. Detection of human interchromosomal trans-splicing in sequence databanks. *Briefings in Bioinformatics*, v. 11, n. 2, p. 198–209, 2 dez. 2009. Disponível em: <https://doi.org/10.1093/bib/bbp041>.

HRZENJAK, A. *JAZF1/SUZ12 gene fusion in endometrial stromal sarcomas* *Orphanet Journal of Rare Diseases* BioMed Central Ltd., , 16 fev. 2016. .

HSU, P. D.; LANDER, E. S.; ZHANG, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, v. 157, n. 6, p. 1262, 6 jun. 2014. Disponível em: </pmc/articles/PMC4343198/>. Acesso em: 7 abr. 2023.

HUONG LE, T. T. et al. Efficient and Precise CRISPR/Cas9-Mediated MECP2 Modifications in Human-Induced Pluripotent Stem Cells. *Frontiers in genetics*, v. 10, n. JUL, 2019. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/31333716/>. Acesso em: 7 abr. 2023.

JENSEN, K. B. et al. Nova-1 Regulates Neuron-Specific Alternative Splicing and Is Essential for Neuronal Viability. *Neuron*, v. 25, n. 2, p. 359–371, 1 fev. 2000.

KEELE, S. *Guidelines for performing systematic literature reviews in software engineering*. Disponível em: <https://www.researchgate.net/publication/302924724_Guidelines_for_performing_Systematic_Literature_Reviews_in_Software_Engineering>. Acesso em: 10 mar. 2020.

- KEEN, C. et al. 10-year-old female with intragenic KANSL1 mutation, no KANSL1-related intellectual disability, and preserved verbal intelligence. *American journal of medical genetics. Part A*, v. 173, n. 3, p. 762–765, 1 mar. 2017. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/28211987/>>. Acesso em: 8 abr. 2023.
- KIM, D.; SALZBERG, S. L. TopHat-Fusion: An algorithm for discovery of novel fusion transcripts. *Genome Biology*, v. 12, n. 8, p. R72, jan. 2011.
- KOIDO, K. et al. Associations between polymorphisms of LSAMP gene and schizophrenia. *Psychiatry research*, v. 215, n. 3, p. 797–798, 30 mar. 2014. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/24491686/>>. Acesso em: 8 abr. 2023.
- KUMAR, S. et al. *Identifying fusion transcripts using next generation sequencing* Wiley Interdisciplinary Reviews: RNABlackwell Publishing Ltd, , 1 nov. 2016. . Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/24491686/>>. Acesso em: 29 jun. 2020.
- LANCASTER, M. A. et al. Cerebral organoids model human brain development and microcephaly. *Nature*, v. 501, n. 7467, p. 373–379, 2013. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/23995685/>>. Acesso em: 10 jun. 2022.
- LEWIS, J. D. et al. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to Methylated DNA. *Cell*, v. 69, n. 6, p. 905–914, 12 jun. 1992. Disponível em: <<http://www.cell.com/article/0092867492906100/fulltext>>. Acesso em: 7 abr. 2023.
- LI, H. et al. Gene fusions and RNA trans-splicing in normal and neoplastic human cells. *Cell cycle (Georgetown, Tex.)*, v. 8, n. 2, p. 218–222, 15 jan. 2009. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/19158498/>>. Acesso em: 7 abr. 2023.
- LI, W.; POZZO-MILLER, L. BDNF deregulation in Rett syndrome. *Neuropharmacology*, v. 76, n. 00, p. 737–746, jan. 2014. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/24491686/>>. Acesso em: 7 abr. 2023.
- LI, Z. et al. Exon-intron circular RNAs regulate transcription in the nucleus. *Nature structural & molecular biology*, v. 22, n. 3, p. 256–264, 6 mar. 2015. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/25664725/>>. Acesso em: 7 abr. 2023.
- LIAO, Y.; SMYTH, G. K.; SHI, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, v. 30, n. 7, p. 923–930, 1 abr. 2014. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/24227677/>>. Acesso em: 3 maio. 2022.
- LIU, S. et al. Comprehensive Evaluation of Fusion Transcript Detection Algorithms and a Meta-Caller to Combine Top Performing Methods in Paired-End RNA-Seq Data. *Nucleic acids research*, v. 44, n. 5, p. e47, mar. 2016.
- LÓPEZ-NIEVA, P. et al. Detection of novel fusion-transcripts by RNA-Seq in T-cell lymphoblastic lymphoma. *Scientific Reports* 2019 9:1, v. 9, n. 1, p. 1–11, 26 mar. 2019. Disponível em: <<https://www.nature.com/articles/s41598-019-41675-3>>. Acesso em: 8 ago. 2022.
- 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. 1–21, 5 dez. 2014. Disponível

em: <<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8>>. Acesso em: 3 maio. 2022.

MACHADO, Â.; HAERTEL MACHADO, L. *Neuroanatomia Funcional*. [s.l: s.n.]

MAHLA, R. S. Stem Cells Applications in Regenerative Medicine and Disease Therapeutics. *International Journal of Cell Biology*, v. 2016, 2016. Disponível em: <[pmc/articles/PMC4969512/](https://pubmed.ncbi.nlm.nih.gov/21074045/)>. Acesso em: 7 abr. 2023.

MARCHETTO, M. C. N. et al. A model for neural development and treatment of rett syndrome using human induced pluripotent stem cells. *Cell*, v. 143, n. 4, p. 527–539, 2010a.

MARCHETTO, M. C. N. et al. A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell*, v. 143, n. 4, p. 527–539, 12 nov. 2010b. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/21074045/>>. Acesso em: 7 abr. 2023.

MARTIN, M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet.journal*, v. 17, n. 1, p. 10, maio 2011.

MARTINEZ-MONLEON, A. et al. Identification of recurrent 3q13.31 chromosomal rearrangement indicates LSAMP as a tumor suppressor gene in neuroblastoma. *International journal of oncology*, v. 62, n. 2, 1 fev. 2023. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/36601748/>>. Acesso em: 8 abr. 2023.

NA, E. S. et al. The impact of MeCP2 loss- or gain-of-function on synaptic plasticity. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology*, v. 38, n. 1, p. 212–219, jan. 2013. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/22781840/>>. Acesso em: 7 abr. 2023.

OLSEN, T. K. et al. Novel fusion genes and chimeric transcripts in ependymal tumors. *Genes, chromosomes & cancer*, v. 55, n. 12, p. 944–953, 1 dez. 2016. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/27401149/>>. Acesso em: 8 ago. 2022.

PAMUDURTI, N. R. et al. Translation of CircRNAs. *Molecular Cell*, v. 66, n. 1, p. 9, 4 abr. 2017. Disponível em: <[pmc/articles/PMC5387669/](https://pubmed.ncbi.nlm.nih.gov/27401149/)>. Acesso em: 7 abr. 2023.

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, jan. 2012.

PHATNANI, H.; MANIATIS, T. Astrocytes in Neurodegenerative Disease. *Cold Spring Harbor Perspectives in Biology*, v. 7, n. 6, p. 1–18, 2015. Disponível em: <[pmc/articles/PMC4448607/](https://pubmed.ncbi.nlm.nih.gov/27401149/)>. Acesso em: 7 abr. 2023.

RAMI, G. et al. Change in the shape and density of dendritic spines caused by overexpression of acidic calponin in cultured hippocampal neurons. *Hippocampus*, v. 16, n. 2, p. 183–197, 1 fev. 2006. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1002/hipo.20145>>. Acesso em: 8 abr. 2023.

RICCERI, L.; DE FILIPPIS, B.; LAVIOLA, G. Rett syndrome treatment in mouse models: Searching for effective targets and strategies. *Neuropharmacology*, v. 68, p. 106–115, maio 2013.

ROBINSON, J. T. et al. Integrative Genomics Viewer. *Nature biotechnology*, v. 29, n. 1, p. 24,

jan. 2011. Disponível em: </pmc/articles/PMC3346182/>. Acesso em: 3 maio. 2022.

SALESSE, S.; VERFAILLIE, C. M. BCR/ABL: From molecular mechanisms of leukemia induction to treatment of chronic myelogenous leukemia. *Oncogene*, v. 21, n. 56 REV. ISS. 7, p. 8547–8559, 9 dez. 2002.

SALZMAN, J. Circular RNA Expression: Its Potential Regulation and Function. *Trends in genetics : TIG*, v. 32, n. 5, p. 309–316, 1 maio 2016. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/27050930/>. Acesso em: 8 abr. 2023.

SHEN, Z.; WANG, L.; YE, D. The expression profile and clinical application value of hsa_circ_0016148 in head and neck squamous cell carcinoma. *Journal of clinical laboratory analysis*, v. 35, n. 11, 1 nov. 2021. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/34592791/>. Acesso em: 8 abr. 2023.

SINGH, S.; LI, H. Comparative study of bioinformatic tools for the identification of chimeric RNAs from RNA Sequencing. *RNA biology*, v. 18, n. sup1, p. 254–267, 2021. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/34142643/>. Acesso em: 8 abr. 2023.

SONG, X. et al. Circular RNA profile in gliomas revealed by identification tool UROBORUS. *Nucleic acids research*, v. 44, n. 9, 19 maio 2016. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/26873924/>. Acesso em: 8 abr. 2023.

TAKAHASHI, K. et al. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, v. 131, n. 5, p. 861–872, 30 nov. 2007. Disponível em: <http://www.cell.com/article/S0092867407014717/fulltext>. Acesso em: 7 abr. 2023.

TAN, K. E.; LIM, Y. Y. Viruses join the circular RNA world. *The FEBS journal*, v. 288, n. 15, p. 4488–4502, 1 ago. 2021. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/33236482/>. Acesso em: 7 abr. 2023.

TANG, Y. et al. Identification of chimeric RNAs in human infant brains and their implications in neural differentiation. *The international journal of biochemistry & cell biology*, v. 111, p. 19–26, 1 jun. 2019a. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/30959201/>. Acesso em: 6 maio. 2022.

TANG, Y. et al. Identification of chimeric RNAs in human infant brains and their implications in neural differentiation. *International Journal of Biochemistry and Cell Biology*, v. 111, p. 19–26, 1 jun. 2019b.

TILLOTSON, R.; BIRD, A. The Molecular Basis of MeCP2 Function in the Brain. *Journal of Molecular Biology*, v. 432, n. 6, p. 1602–1623, 13 mar. 2020.

TRAN, S. S. et al. Widespread RNA editing dysregulation in brains from autistic individuals. *Nature Neuroscience* 2018 22:1, v. 22, n. 1, p. 25–36, 17 dez. 2018. Disponível em: <https://www.nature.com/articles/s41593-018-0287-x>. Acesso em: 6 maio. 2022.

TRUJILLO, C. A. et al. Pharmacological reversal of synaptic and network pathology in human MECP2-KO neurons and cortical organoids. *EMBO Molecular Medicine*, v. 13, n. 1, p. e12523, 11 jan. 2021a. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.15252/emmm.202012523>. Acesso em: 1 jul. 2022.

TRUJILLO, C. A. et al. Reintroduction of the archaic variant of NOVA1 in cortical organoids alters neurodevelopment. *Science (New York, N.Y.)*, v. 371, n. 6530, 12 fev. 2021b. Disponível em: <[pmc/articles/PMC8006534/](https://pubmed.ncbi.nlm.nih.gov/33441414/)>. Acesso em: 3 maio. 2022.

UHRIG, S. et al. Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Research*, v. 31, n. 3, p. 448–460, 1 jan. 2021. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/33441414/>>. Acesso em: 29 abr. 2021.

ULE, J. et al. CLIP identifies Nova-regulated RNA networks in the brain. *Science (New York, N.Y.)*, v. 302, n. 5648, p. 1212–1215, 14 nov. 2003. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/14615540/>>. Acesso em: 7 abr. 2023.

WONG, G. C. H. et al. Clinical and mutational profiles of adult medulloblastoma groups. *Acta Neuropathologica Communications*, v. 8, n. 1, p. 1–14, 1 dez. 2020. Disponível em: <<https://actaneurocomms.biomedcentral.com/articles/10.1186/s40478-020-01066-6>>. Acesso em: 8 abr. 2023.

XIN, W. et al. Hsa_circ_0066523 promotes the proliferation and osteogenic differentiation of bone mesenchymal stem cells by repressing PTEN. *Bone & joint research*, v. 10, n. 8, p. 526–535, 1 ago. 2021. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/34402627/>>. Acesso em: 8 abr. 2023.

YAKOUB, A. M.; SADEK, M. Development and Characterization of Human Cerebral Organoids: An Optimized Protocol. *Cell transplantation*, v. 27, n. 3, p. 393–406, mar. 2018.

YATES, T. M. et al. Baraitser–Winter cerebrofrontofacial syndrome. *Clinical Genetics*, v. 92, n. 1, p. 3–9, 1 jul. 2017.

YOUNGBLOOD, M. W.; GÜNEL, M. Molecular genetics of meningiomas. *Handbook of clinical neurology*, v. 169, p. 101–119, 1 jan. 2020. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/32553282/>>. Acesso em: 8 abr. 2023.

YU, L. et al. The circular RNA Cdr1as act as an oncogene in hepatocellular carcinoma through targeting miR-7 expression. *PLoS ONE*, v. 11, n. 7, 1 jul. 2016. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/27391479/>>. Acesso em: 25 jun. 2020.

ZHANG, J. et al. INTEGRATE: Gene fusion discovery using whole genome and transcriptome data. *Genome Research*, v. 26, n. 1, p. 108–118, 1 jan. 2016. Disponível em: <[pmc/articles/PMC4691743/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/25242744/)>. Acesso em: 20 jun. 2020.

ZHANG, X. O. et al. Complementary sequence-mediated exon circularization. *Cell*, v. 159, n. 1, p. 134–147, 25 set. 2014. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/25242744/>>. Acesso em: 8 abr. 2023.

ZHENG, Q. et al. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nature Communications*, v. 7, n. 1, p. 1–13, 6 abr. 2016.

ZHONG, Q. et al. Circular RNA CDR1as sponges miR-7-5p to enhance E2F3 stability and promote the growth of nasopharyngeal carcinoma. *Cancer cell international*, v. 19, n. 1, 1 out. 2019. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/31582908/>>. Acesso em: 7 abr. 2023.

ZHOU, J. X. et al. Identification of KANSARL as the first cancer predisposition fusion gene

specific to the population of European ancestry origin. *Oncotarget*, v. 8, n. 31, p. 50594, 8 ago. 2017. Disponível em: </pmc/articles/PMC5584173/>. Acesso em: 8 ago. 2022.

ZOU, Y. et al. The Role of Circular RNA CDR1as/ciRS-7 in Regulating Tumor Microenvironment: A Pan-Cancer Analysis. *Biomolecules*, v. 9, n. 9, 1 set. 2019. Disponível em: </pmc/articles/PMC6770779/>. Acesso em: 7 abr. 2023.

APÊNDICES

Tabela suplementar 1

Ferramenta	Ano	RNA-seq	WGS	Alinhador	Mapping ou Assembly First?	Formato das Reads (Single-end/Paired-end)	FILTROS
MapSplice	2010	x		Bowtie	Assembly First	Paired-End	Alignment Quality / Anchor Size / Canonical Filters
FusionQ	2013	x	x	Bowtie	Assembly First	Paired-End	Minimum Number Filter / Homology Filter / Distribution Filter / Expression Levels
NFuse	2012	x	x	Bowtie 2	Assembly First	Paired-End	Canonical Ends / Minimum Number Filter
NCLscan	2016	x		BWA	Assembly First	Paired-End	Minimum Number Filter / Distance Filter
SQUID	2018	x	x	BWA	Assembly First	Paired-End	Minimum Number Filter
IDP-Fusion	2015	x	x	GMAP	Assembly First	Paired-End AND LONG READS	Minimum Number Filter / Long Reads Support
TAP	2018	x	x	GMAP and BWA	Assembly First	Paired-End	Minimum Number Filter
JAFFA	2015	x		Bowtie 2	BOTH	Paired-End	Minimum Number Filter / Homology Filter
INTEGRATE	2016	x	x	GSNAP / Bowtie / STAR	BOTH	Paired-End	Minimum Number Filter / Homology Filter / WGS Support
GFusion	2017	x	x	Bowtie	Mapping First	Paired-End	Map Quality Score / Minimum Number Filter / Realignment
Barnacle	2013	x	x	Bowtie	Mapping First	Paired-End	Map Quality Score / Minimum Number Filter / Realignment
ShortFuse	2011	x	x	Bowtie	Mapping First	Paired-End	Minimum Number Filter / Distance Filter
DeFuse	2011	x	x	Bowtie	Mapping First	Paired-End	Minimum Number Filter /

							Distance Filter
TopHat-Fusion	2011	x	x	Bowtie	Mapping First	Paired-End	Minimum Number Filter
ChimeraScan	2011	x	x	Bowtie	Mapping First	Paired-End	Minimum Number Filter
Bellerophon	2012	x	x	Bowtie	Mapping First	Paired-End	Minimum Number Filter / Homology Filter / PCR Filter
FusionCatcher	2014	x		Bowtie	Mapping First	Paired-End	Homology Filter / Multiple Alignment Filter
TopHat2	2013	x	x	Bowtie / Bowtie 2	Mapping First	Paired-End	Canonical Ends / Minimum Number Filter / Intron Filter
InFusion	2016	x	x	Bowtie 2	Mapping First	Paired-End	Distance Filter / Homology Filter / Minimum Number Filter
PRADA	2014	x	x	BWA	Mapping First	Paired-End	Minimum Number Filter / Homology Filter
EricScript	2012	x	x	BWA	Mapping First	Paired-End	BLAT Filter / Minimum Number Filter / Homology Filter
BreakID	2019	x		BWA	Mapping First	Paired-End	Minimum Number Filter
FusionMap	2011	x	x	GSNAP	Mapping First	Paired-End	Homology Filter
FusionSeq	2010	x	x	SAM/BAM - ANY	Mapping First	Paired-End	Homology Filter / PCR Filter / Candidates Score
ChimPipe	2017	x	x	SAM/BAM - ANY	Mapping First	Paired-End	Minimum Number Filter / Canonical Ends / High Exonic Similarity (Homology Filter)
CRAC	2013	x	x	SAM/BAM - ANY	Mapping First	Paired-End	k-mer
FuSEQ	2018	x	x	SAM/BAM - ANY	Mapping First	Paired-End	k-mer / Homology Filter / Minimum Number Filter
Pegasus	2017	x	x	SAM/BAM - ANY	Mapping First	Paired-End	BLAT / Minimum Number Filter
Bellerophon	2013	x		SAM/BAM - ANY	Mapping First	Paired-End	Distance Filter / Orientation Filter

Comrad	2011	x	x	SAM/BAM - ANY	Mapping First	Paired-End	Homology Filter / Minimum Number Filter
Breakfusion	2012	x	x	SAM/BAM - ANY	Mapping First	Paired-End	Minimum Number Filter
SOAPfusion	2013	x	x	SOAP-Aligner	Mapping First	Paired-End	BWA / Homology Filter / Minimum Number Filter
SOAPFuse	2013	x		SOAP2	Mapping First	Paired-End	BWA / Homology Filter / Minimum Number Filter
STAR-Fusion	2017	x		STAR	Mapping First	Paired-End	

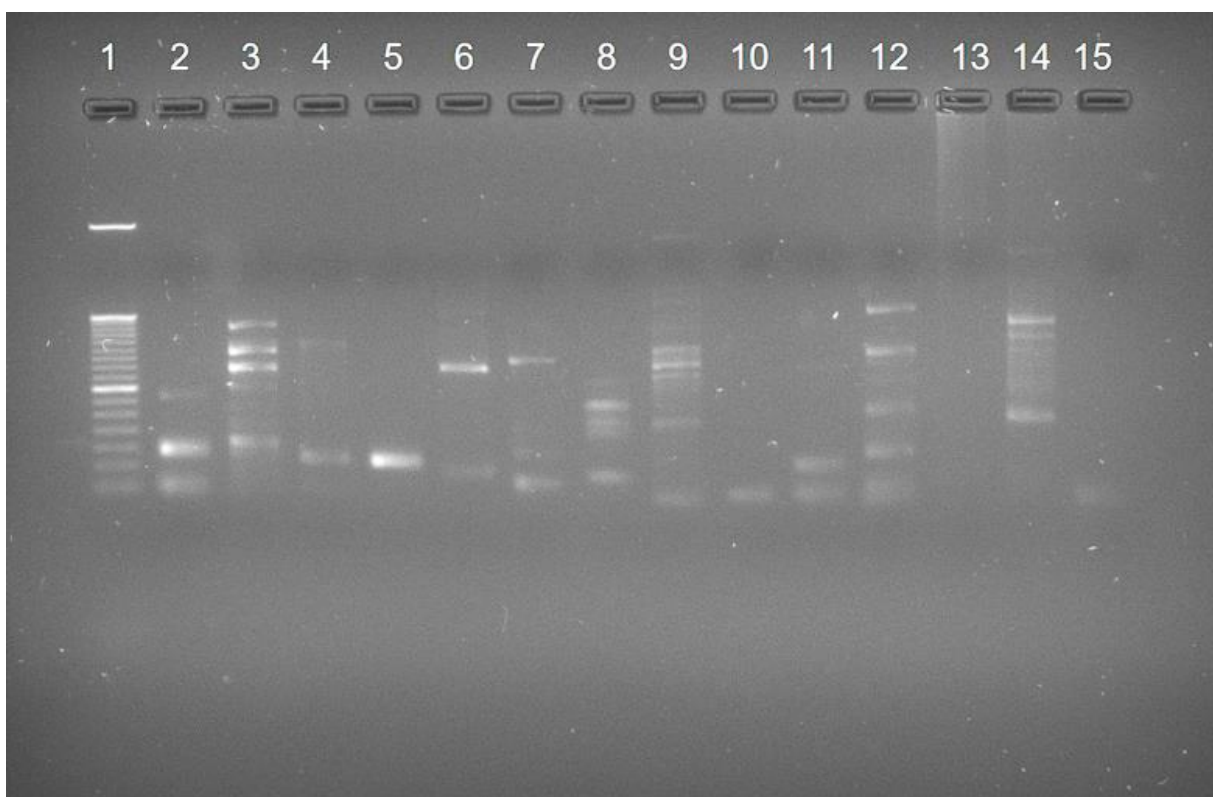


Figura 1A – Géis ASC e NOVA – 1) LADDER; 2) FUS:EWSR; 3) ACTB:NUB (T1); 4) ACTB:NUB (T2); 5) CNN:CALM3 (T1); 6) RP11:LSMAP (T1); 7) RP11:LSMAP (T2); 8) RP11:GSE3; 9) CNN:CALM3 (T2); 10) FUS:EWSR (DNA); 11) ACTB:NUB (DNA); 12) CNN:CALM3 (DNA); 13) RP11:LSMAP (DNA); RP11:GSE3 (DNA)

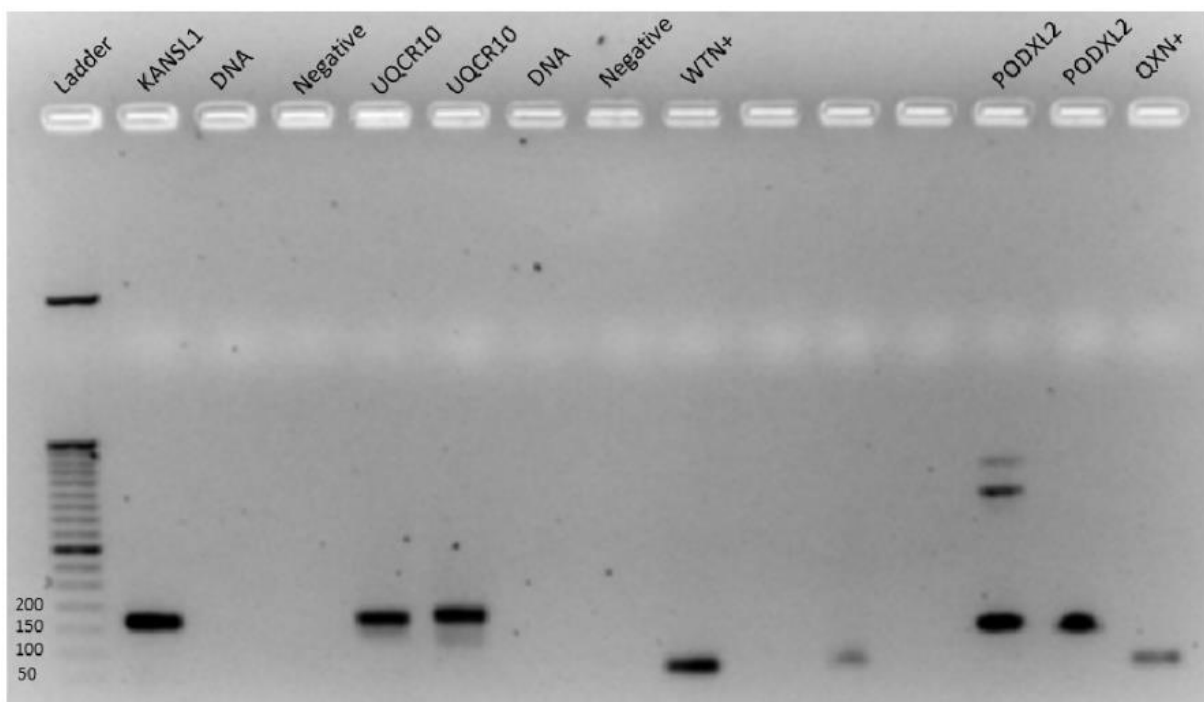


Figura 2A – Géis WT E QXN

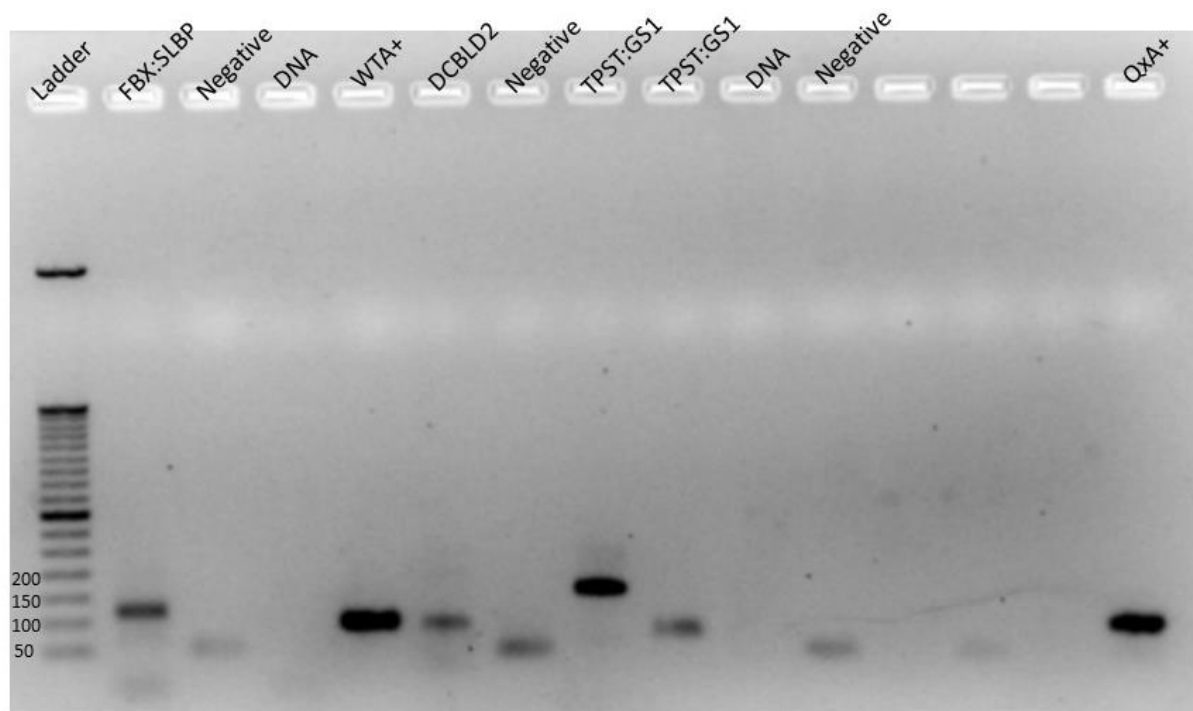


Figura 3A – Géis WT E QXA

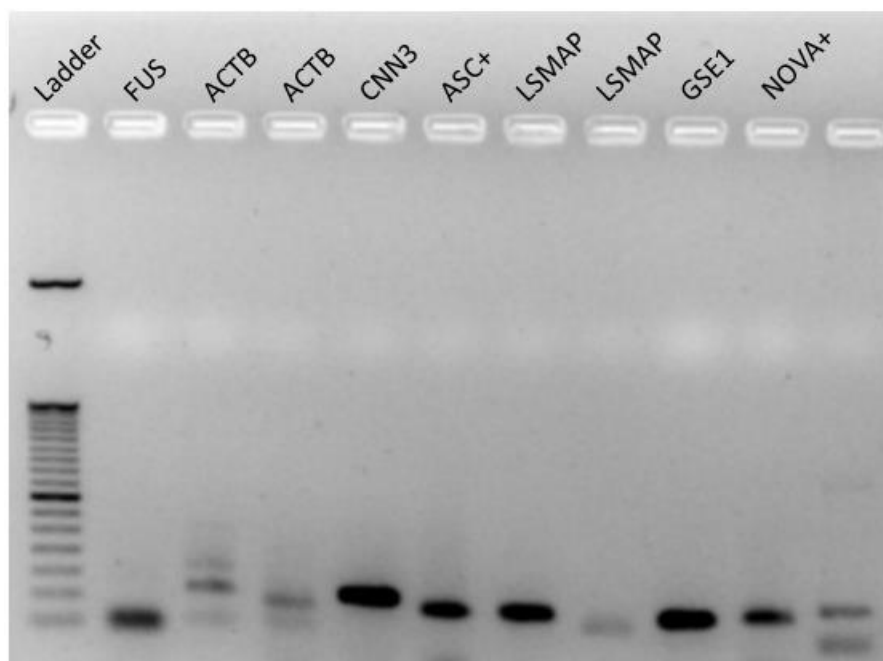


Figura 3A – Géis WT E QXA

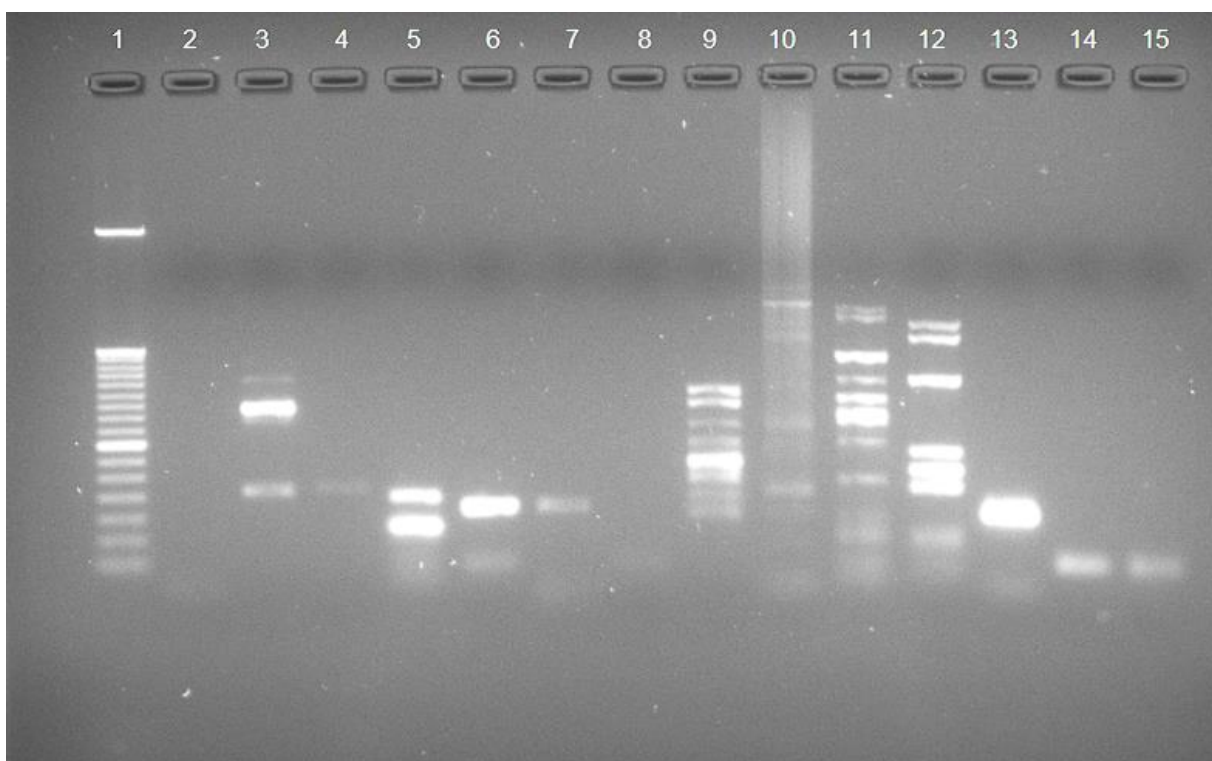


Figura 4A – CIRCRNA – 1) LADDER; 3) COL8A:CMSS1; 4) CIRC CRIM1 (QXA); 5) CIRC LPAR (QXA); 6) CIRC AMFR; 7) CIRC MAP2; 8) RP11:GSE3; 9) CNN:CALM3 (T2); 13) PODXL2:FYN (DNA); 14) FBXO23:SLBP; 15) FBX032:SLBP

[illegible]

Figura 11A – Sequenciamento Sanger para a fusão RP11-680G10:GSE1

Tabela suplementar 3

Gene-1	Região (chr/p os)	Tipo	Função	Gene-2	Região	Tipo	Função	Descrit o Antes?
KANSL1 (ENSG00000120071)	17/46094560	Prot ein-Codi ng	Acetilação de Histonas, regulador de proliferaç ão celular e mitose	ARL17A (ENSG00000185829)	17/46570869	Prot ein-Codi ng	Atividade de ligação GTP, transporte intracelular de proteínas e vesículas	Linfom a linfobl ástico [56] e Biomar cador [57]
RP11-384F7.2 (ENST00000484092)	3/117997182	Lnc RNA	Sem informaç ões	LSAMP (ENSG00000185565)	3/116444955	Prot ein-Codi ng	Glicoproteín a de superfície neuronal que age na adesão homofílica durante neurodesen volvimento	-
RP11-680G10.1 (ENSG00000261567)	16/85357643	Lnc RNA	Sem informaç ões	GSE1 (ENSG00000131149)	16/85633914	Prot ein-Codi ng	Formação de complexo deacetilase de histonas, associado como oncogene no câncer de mama	Câncer endom etrial [63]
TPST1 (ENSG00000169902)	7/66286709	Prot ein-Codi ng	Síntese de proteínas sulfatadas . Associado a carcinoma pulmonar	GS1-124K5 (ENSG00000244657)	7/66495255	Lnc RNA	Sem informações	-
COL8A1 (ENSG00000144810)	3/99745021	Prot ein-Codi ng	Uma das duas cadeias de colágeno tipo VIII. Compone nte essencial	CMSS1 (ENSG00000184220)	3/100146973	Prot ein-Codi ng	Ligações de ácidos nucleicos e RNA	-

			da membran a basal do endotélio					
FUS (ENSG00000089280)	16/311 86836	Prot ein- Codi ng	Proteína Multifunci onal do complexo ribonucle oproteico heterogên eo nuclear (hnRNP). Processa mento do pré-mRNA e exportaçã o do mRNA para o citoplasm a	EWSR1 (ENSG00000182944)	22/292 92137	Prot ein- Codi ng	Regulação da expressão genica, sinalização celular e processame nto e transporte de RNA	Neopla sia epiteli oide [65]
UQCR10 (ENSG00000184076)	22/297 67384	Prot ein- Codi ng	Subunidad e do complexo mitocondr ial III. Forma o segmento médio da cadeia respiratóri a na membran a Mitocondr ial	C1orf194 (ENSG00000179902)	1/1091 08012	Prot ein- Codi ng	Previsão de atividades relacionadas à ligação e transporte de heme, além de estar envolvido na exportação de heme, e está localizado no citoplasma e citoesquelet o.	Ependi moma [66]
ACTB (ENSG00000075624)	7/5528 281	Prot ein- Codi ng	Proteína de Actina, que auxiliam na motilidad e celular, estrutura,	NUCB1 (ENSG00000104805)	17/815 11108	Prot ein- Codi ng	Proteínas de ligação de cálcio. Papel na homeostase do cálcio no complexo de	-

			integridade e sinalização intercelular.				Golgi e sinalização	
CNN3 (ENSG00000117519.16)	1/94903524	Protein-Coding	Regula filamento de actina durante remodelagem neuronal. Também presente na plasticidade de dendritos	CALM3 (ENSG00000160014)	19/46601074	Protein-Coding	Ligações ao cálcio e atua como cofator enzimático. Essencial na regulação do ciclo celular e citocinese	-
FBX032	8/123540899	Protein-Coding	Proteína da família F-BOX	SLBP (ENSG00000163950)	4/1703700	Protein-Coding	Ligação a estrutura de loop de histonas de RNA. Essencial para o processamento eficiente, transporte, tradução e transporte dos RNAs histona	-
COL11A1 (ENST00000370096)	1/103074618	Protein-Coding	Codifica uma cadeia alfa do colágeno do tipo XI - colágeno fibrilar menor	TP53BP1 (ENSG00000067369)	15/43470066	Protein-Coding	Reparação de quebras de dupla hélice do DNA, promove a junção de extremidades não homólogas e limitando a recombinação homóloga	-

PODXL2 (ENSG00000114631)	3/1276 29289	Prot ein- Codi ng	Proteína de superfície celular que faz parte da família CD34 e é um ligante para selectinas vasculares	FYN (ENSG00000010810)	6/1117 20062	Prot ein- Codi ng	membro da família de oncogenes de tirosina quinase. Ele codifica uma tirosina quinase associada à membrana que tem sido implicada no controle do crescimento celular	-
---	-----------------	----------------------------	---	--	-----------------	----------------------------	---	---

ANEXOS

SEMIC XXVII (2019) – Susanna Rachel - ANÁLISE DE EXPRESSÃO GÊNICA DE MOLÉCULAS REGULADORAS DA ATIVIDADE DE RETROTRANSPOSONS DO TIPO LINE-1 EM TRANSPOSONS DO ESPECTRO DO AUTISMO. Orientador: Adriano Ferrasa e Bruno Siegel Guerra e Roberto Herai

SEMIC XXX (2022) – Davi Cecconi Borges Massa Checan - CARACTERIZAÇÃO DE TRANSCRITOS QUIMÉRICOS E RNA CIRCULARES EM DADOS DE PACIENTES COM SÍNDROME DE RETT. Orientador: Bruno Siegel Guerra e Roberto Herai

29º SIICUSP (2021) – Davi Cecconi Borges Massa Checan - CARACTERIZAÇÃO DE TRANSCRITOS QUIMÉRICOS E RNA CIRCULARES EM DADOS DE PACIENTES COM SÍNDROME DE RETT. Orientador: Bruno Siegel Guerra e Roberto Herai

X-MEETING XP (2021) – Thomas Creczynski Pasa – CHIMERIC RNA DETECTION TOOL, A NO-FILTERS APPROACH FOR THE DETECTION OF NEW FUSIONS. Orientador: Bruno Siegel Guerra e Roberto Herai

GUERRA, B. S. et al. **Biogenesis of circular RNAs and their role in cellular and molecular phenotypes of neurological disorders. Seminars in Cell and Developmental Biology** Elsevier Ltd, 3 set. 2020.

ADAMS, J.W., NEGRAES, P.D., TRUONG, J. *et al.* **Impact of alcohol exposure on neural development and network formation in human cortical organoids.** *Mol Psychiatry* (2022).

<https://doi.org/10.1038/s41380-022-01862-7>

SOUZA, Andressa & MARCHIORO, Stella & BOHN, Loyse & OLIVEIRA, Miguel & GUERRA, Bruno & ROMANI, Rafael & CARNEIRO, João & RIELLA, Miguel. (2022). **Perfil epidemiológico de pacientes com injúria renal aguda e COVID-19 na primeira onda da pandemia no Brasil.** *BioSCIENCE*. <https://doi.org/10.55684/80.2.3>

JW Adams, A Vinokur, JS de Souza, C Austria, BS Guerra, RH Herai *et al.* **Loss of GTF2I promotes neuronal apoptosis and synaptic reduction in human cellular models of neurodevelopment.** *Cell Reports* (2024). <https://doi.org/10.1016/j.celrep.2024.113867>