

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE CIÊNCIAS BÁSICAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS:
BIOQUÍMICA

Guilherme Bauer Negrini

**BIOINFORMÁTICA DE CIRCUNAS: ANÁLISE EM MODELOS ANIMAIS DO
TRANSTORNO DO ESPECTRO AUTISTA E DESENVOLVIMENTO DE UM
MÉTODO DE AVALIAÇÃO DA USABILIDADE DOS SOFTWARES UTILIZADOS
PARA IDENTIFICAÇÃO**

Porto Alegre

2022

Guilherme Bauer Negrini

**BIOINFORMÁTICA DE CIRCURNAS: ANÁLISE EM MODELOS ANIMAIS DO
TRANSTORNO DO ESPECTRO AUTISTA E DESENVOLVIMENTO DE UM
MÉTODO DE AVALIAÇÃO DA USABILIDADE DOS SOFTWARES UTILIZADOS
PARA IDENTIFICAÇÃO**

Tese apresentada ao Programa de Pós-Graduação em Ciências Biológicas: Bioquímica do Instituto de Ciências Básicas da Saúde da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de doutor em Bioquímica.

Orientadora: Prof.^a. Dr.^a. Carmem Juracy
Silveira Gottfried

Coorientador: Prof. Dr. Guilherme Cordenonsi
da Fonseca

Porto Alegre

2022

CIP - Catalogação na Publicação

Negrini, Guilherme Bauer

Bioinformática de circRNAs: análise em modelos animais do Transtorno do Espectro Autista e desenvolvimento de um método de avaliação da usabilidade dos softwares de identificação utilizados / Guilherme Bauer Negrini. -- 2022.

159 f.

Orientadora: Carmem Juracy Silveira Gottfried.

Coorientador: Guilherme Cordenonsi da Fonseca.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Instituto de Ciências Básicas da Saúde, Programa de Pós-Graduação em Ciências Biológicas: Bioquímica, Porto Alegre, BR-RS, 2022.

1. RNAs circulares. 2. Transtorno do Espectro Autista. 3. Bioinformática. 4. Usabilidade de software. 5. Heurísticas de usabilidade. I. Gottfried, Carmem Juracy Silveira, orient. II. da Fonseca, Guilherme Cordenonsi, coorient. III. Título.

Guilherme Bauer Negrini

**BIOINFORMÁTICA DE CIRCRNAS: ANÁLISE EM MODELOS ANIMAIS DO
TRANSTORNO DO ESPECTRO AUTISTA E DESENVOLVIMENTO DE UM
MÉTODO DE AVALIAÇÃO DA USABILIDADE DOS SOFTWARES UTILIZADOS
PARA IDENTIFICAÇÃO**

Tese de Doutorado apresentada ao Instituto de Ciências Básicas da Saúde da
Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção o
título de doutor em Bioquímica.

Aprovado em: 14 de dezembro de 2022.

BANCA EXAMINADORA

Prof. Dr. Dinler Amaral Antunes - Universidade de Houston
(Departamento de Biologia e Bioquímica)

Prof. Dr. Geancarlo Zanatta - Universidade Federal do Rio Grande do Sul
(Departamento de Biofísica)

Prof.^a. Dra.^a. Clarissa Severino Gama - Universidade Federal do Rio Grande do Sul
(Departamento de Psiquiatria e Medicina Legal)

Prof.^a. Dr.^a. Carmem Juracy Silveira Gottfried – Universidade Federal do Rio Grande
do Sul (orientadora)

Prof. Dr. Guilherme Cordenonsi da Fonseca – Laboratório Nacional de Computação
Científica (coorientador)

Agradecimentos

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Instituto Nacional de Ciência e Tecnologia em NeuroImunoModulação pelo financiamento disponibilizado.

À Universidade Federal do Rio Grande do Sul, pela excelência na formação e pela estrutura que desfrutei ao longo de 10 anos. Com carinho, destaco o PPG Bioquímica e seus profissionais pelo auxílio aos alunos, especialmente em tempos desafiadores.

À Carmem, minha orientadora, pela oportunidade dada a mim ainda na graduação. Obrigado pelos incentivos e por vibrar com cada vitória dos alunos do nosso grupo de pesquisa. Sou muito grato pelos teus ensinamentos ao longo de tantos anos e pelo teu apoio na vida científica e pessoal.

Ao Guilherme, meu coorientador, a gratidão por ter me apresentado o mundo da bioinformática. Muito obrigado por pacientemente ter me ensinado tantas vezes.

À professora Juliana, por tornar possível um trabalho tão interdisciplinar.

Aos meus colegas de laboratório, em especial à Iohanna, Júlio e Brum, obrigado não apenas pelas contribuições científicas, mas pela amizade e companhia. Em um sistema tão duro, como a academia, é a humanidade diária dos nossos iguais que nos dá força.

À Alice, Júlia e Sofia, pela amizade e pelo melhor grupo de trabalho.

Ao Zeca e à Ju, os melhores amigos que eu poderia ter, e à pequena Aurora por entender a ausência do dindo.

À minha família, pelo amor inesgotável. Aos meus pais, Ismael e Marilene, por compreenderem e apoiarem a minha jornada. Aos meus sogros, Marco e Eliane, pelo incentivo e companhia diária.

À minha esposa, Luísa, a maior fonte de inspiração e força durante estes anos. Muito obrigado por me levar tão longe e por compartilhar tantos sonhos comigo.

Nem todos que vagueiam estão perdidos.

J.R.R. Tolkien

SUMÁRIO

PARTE I	1
RESUMO.....	2
ABSTRACT.....	3
LISTA DE ILUSTRAÇÕES.....	4
LISTA DE TABELAS.....	5
LISTA DE ABREVIATURAS E SIGLAS.....	6
NOTA SOBRE A NOMENCLATURA GENÉTICA.....	7
1. INTRODUÇÃO	8
1.1. Transtorno do Espectro Autista	8
1.1.1. Prevalência	9
1.1.2. Fatores de risco	10
1.1.3. Diagnóstico	12
1.1.4. Alterações na expressão gênica	13
1.2. Modelos animais do Transtorno do Espectro Autista.....	15
1.2.1. Tipos de modelos.....	16
1.3. RNAs circulares.....	19
1.3.1. Biogênese e características	19
1.3.2. Expressão gênica e funções no sistema nervoso central	21
1.3.3. RNAs circulares no Transtorno do Espectro Autista	22
1.4. Identificação de RNAs circulares por softwares de bioinformática	23
1.4.1. Principais softwares	23
1.4.2. Usabilidade de softwares de identificação de circRNAs	24
1.5. Qualidade de software.....	25
1.5.1. Métodos de avaliação de usabilidade de software.....	25
1.5.2. Qualidade e usabilidade de softwares de bioinformática	27
2. JUSTIFICATIVA	29

3. OBJETIVOS	30
3.1. Objetivo geral	30
3.2. Objetivos específicos	30
PARTE II	31
CAPÍTULO I.....	32
<i>A comprehensive bioinformatics analysis of circRNA expression in the brain of distinct mouse models of autism.....</i>	<i>32</i>
CAPÍTULO II.....	65
<i>Resultados adicionais: ciRS-7/CDR1as is up-regulated in the prefrontal cortex of rats of the animal model of autism induced by valproic acid.....</i>	<i>65</i>
CAPÍTULO III.....	80
<i>Usability evaluation of circRNA identification tools: Development of a heuristic-based framework and analysis</i>	<i>80</i>
PARTE III	116
4. DISCUSSÃO	117
5. CONSIDERAÇÕES FINAIS E CONCLUSÕES	128
6. PERSPECTIVAS	130
REFERÊNCIAS BIBLIOGRÁFICAS.....	131
ANEXOS	147
ANEXO I - Parecer de projeto de doutorado encaminhado para avaliação pelo Programa de Pós-Graduação em Ciências Biológicas: Bioquímica	147
ANEXO II – Ficha de avaliação heurística da usabilidade de software de identificação de circRNA.....	150

PARTE I

RESUMO

O Transtorno do Espectro Autista (TEA) é um transtorno do neurodesenvolvimento caracterizado por prejuízos na comunicação e interação social, comportamentos repetitivos e interesses restritos. Estima-se que, mundialmente, 1% da população é afetada pela condição. A etiologia do TEA é desconhecida, porém sabe-se que existe uma contribuição complexa de fatores ambientais e genéticos para o desenvolvimento do transtorno. Alterações transcritômicas são consistentemente observadas no sistema nervoso central de indivíduos com TEA e não estão restritas a expressão de genes codificantes. Uma classe de RNAs não-codificantes de interesse especialmente recente é a dos RNAs circulares (circRNAs). Os circRNAs são abundantes no encéfalo e podem desempenhar funções variadas como competir pela ligação de miRNAs, estimular a transcrição do gene parental e ser traduzido em proteínas funcionais. O envolvimento dos circRNAs no TEA ainda é pouco explorado. Recentemente foram feitas as primeiras observações de circRNAs diferencialmente expressos (CDE) no córtex cerebral de indivíduos com TEA. Adicionalmente, CDE também foram observados em dois modelos animais de TEA. Modelos animais são essenciais para o estudo de transtornos do neurodesenvolvimento e podem auxiliar os pesquisadores a investigar a etiologia e os mecanismos biológicos subjacentes. A detecção em massa e a descoberta de novos circRNAs depende quase exclusivamente de softwares de identificação de circRNA. Entretanto, possíveis problemas na usabilidade dessas ferramentas são relatados na literatura e podem prejudicar os usuários ao consumir tempo e dificultar a reprodutibilidade das análises. Considerando o potencial envolvimento de circRNAs no TEA e a dependência metodológica de softwares para identificação, a presente tese analisa por bioinformática a expressão gênica de circRNAs no encéfalo de diferentes modelos animais de TEA, assim como desenvolver um método de avaliação e melhoramento da usabilidade das ferramentas de bioinformática utilizadas para a identificação dos circRNAs. No Capítulo I, identificamos milhares de circRNAs em diferentes modelos animais de TEA em camundongos e demonstramos que as alterações na expressão estão relacionadas principalmente às funcionalidades sinápticas. No Capítulo II, identificamos parcialmente a sequência do circRNA ciRS-7 no modelo animal de ratos induzidos por ácido valproico. Demonstramos que a expressão de ciRS-7 está elevada, assim como encontrada nos modelos Ash1L, BTBR e Kmt5b no Capítulo I, e que funções sinápticas estão entre os possíveis grupos funcionais afetados. Por fim, no Capítulo III, identificamos os problemas de usabilidade das ferramentas de identificação de circRNAs utilizadas nos Capítulos I e II. Também criamos um conjunto de regras para auxiliar os desenvolvedores desse tipo de software para criarem ferramentas mais utilizáveis para os usuários finais. Em conjunto, os resultados apresentam um panorama do envolvimento de circRNAs no TEA a partir de modelos animais. Além disso, apresentamos um novo método de avaliação de usabilidade de software que deve possibilitar que futuras análises estejam alinhadas com as necessidades e expectativas de uso de ferramentas de identificação de circRNAs.

ABSTRACT

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by impairments in communication and social interaction, repetitive behaviors, and restricted interests. It is estimated that the condition affects 1% of the population worldwide. The etiology of ASD is unknown, but it is known that there is a complex contribution of environmental and genetic factors to the development of the disorder. Transcriptomic alterations are consistently observed in the central nervous system of individuals with ASD and are not restricted to the expression of coding genes. Circular RNAs (circRNAs) are a class of noncoding RNAs of recent interest. CircRNAs are abundant in the brain and can perform varied functions, such as competing for binding miRNAs, stimulating parental gene transcription, and being translated into functional proteins. The involvement of circRNAs in ASD is still poorly explored. Recently, the first observations of differentially expressed circRNAs (DECs) were made in the cerebral cortex of individuals with ASD. Additionally, DECs were also observed in two animal models of ASD. Animal models are essential for studying neurodevelopmental disorders and can help researchers investigate the etiology and underlying biological mechanisms. The mass detection and discovery of novel circRNAs rely almost exclusively on circRNA identification software. However, potential problems in the usability of these tools are reported in the literature and may harm users by consuming time and hindering the reproducibility of analyses. Considering the potential involvement of circRNAs in ASD and the methodological dependence on software for identification, this thesis analyzes by bioinformatics the gene expression of circRNAs in the brain of different animal models of ASD, as well as develops a method to evaluate and improve the usability of bioinformatics tools used for circRNA identification. In Chapter I, we identify thousands of circRNAs in different animal models of ASD in mice and demonstrate that the changes in expression are mainly related to synaptic functionalities. In Chapter II, we partially identify the sequence of the circRNA ciRS-7 in the valproic acid-induced rat model. We demonstrate that ciRS-7 expression is elevated, as found in the *Ash1L*, *BTBR*, and *Kmt5b* models in Chapter I, and that synaptic functions are among the possible functional groups affected. Finally, in Chapter III, we identify the usability problems of the circRNAs identification tools used in Chapters I and II. We also create a set of rules to assist developers of such software in creating more usable tools for end users. Taken together, the results present an overview of the involvement of circRNAs in ASD from the perspective of animal models. In addition, we present a new software usability evaluation method that should enable future analyses to be aligned with the needs and expectations of using circRNAs identification tools.

LISTA DE ILUSTRAÇÕES

Figura 1 - Biogênese de circRNAs.	20
Figura 2 - Principais funções dos circRNAs.	21
Figura 3 – Exemplo de uma violação heurística referente ao auxílio no diagnóstico e recuperação de erros.	26
Figura 4 - Visão geral dos principais resultados obtidos ao longo do trabalho.....	128

LISTA DE TABELAS

Tabela 1 – Análises fenotípicas nas principais espécies de modelos animais de TEA.	16
Tabela 2 – Principais modelos animais genéticos e ambientais de TEA em camundongos.....	18

LISTA DE ABREVIATURAS E SIGLAS

ADDM - Monitoramento do Autismo e das Deficiências de Desenvolvimento, do inglês *Autism and Developmental Disabilities Monitoring*

CDC – Centros de Controle e Prevenção de Doenças, do inglês *Centers for Disease Control and Prevention*

CDE – circRNAs diferencialmente expressos

circRNA – ácido ribonucleico circular, do inglês *circular ribonucleic acid*

DSM-5 - Manual Diagnóstico e Estatístico de Transtornos Mentais, do inglês *Diagnostic and Statistical Manual of Mental Disorders*

EUA – Estados Unidos da América

lncRNA – ácido ribonucleico longo não-codificante, do inglês *long noncoding RNA*

miRNA – micro ácido ribonucleico, do inglês *micro ribonucleic acid*

ncRNA – ácido ribonucleico não-codificante, do inglês *noncoding RNA*

RNA – ácido ribonucleico, do inglês *ribonucleic acid*

RNA_m – ácido ribonucleico mensageiro, do inglês *messenger ribonucleic acid*

RNA_r – ácido ribonucleico ribossomal, do inglês *ribosomal ribonucleic acid*

RNA-Seq – sequenciamento de RNA

SNC - sistema nervoso central

TDAH - Transtorno do déficit de atenção com hiperatividade

TEA – Transtorno do Espectro Autista

VPA – ácido valproico, do inglês *valproic acid*.

NOTA SOBRE A NOMENCLATURA GENÉTICA

Na presente tese são referenciados genes e proteínas das espécies *Homo sapiens*, *Mus musculus* e *Rattus norvegicus*. A convenção (SPLENDORE, 2005) apresentada a seguir foi utilizada para comunicar de forma precisa ao leitor a distinção entre genes e proteínas em diferentes espécies.

1. Símbolos de genes humanos são grafados em itálico e letras maiúsculas. Por exemplo, o gene *Fragile X Mental Retardation 1* é grafado como *FMR1*.
2. Em genes de espécies não-humanas, os símbolos são grafados em itálico e com apenas a primeira letra maiúscula. Por exemplo, *Fmr1* em camundongos.
3. As proteínas usualmente recebem o mesmo símbolo do gene, porém não são grafadas em itálico. Por exemplo, a proteína codificada pelo gene *FMR1* em humanos é a FMR1. Em camundongos, semelhante à regra 2, a proteína resultante é a Fmr1.

1. INTRODUÇÃO

1.1. Transtorno do Espectro Autista

O Transtorno do Espectro Autista (TEA) é definido no Manual Diagnóstico e Estatístico de Transtornos Mentais (DSM-5) como um transtorno do neurodesenvolvimento caracterizado por prejuízos na comunicação e interação social, comportamentos repetitivos e interesses restritos (AMERICAN PSYCHIATRIC ASSOCIATION, 2013). Os transtornos do neurodesenvolvimento são um grupo de condições com início no período de desenvolvimento que podem acarretar condições crônicas persistentes pelo resto da vida (FRANCÉS *et al.*, 2022). Além do TEA, os transtornos do neurodesenvolvimento também compreendem a deficiência intelectual; transtornos de comunicação; transtorno do déficit de atenção com hiperatividade (TDAH); transtornos motores do neurodesenvolvimento, e transtornos de aprendizagem específicos. Uma característica desse grupo é o alto grau de comorbidade ou sobreposição de alguns sintomas (MORRIS-ROSENDAHL; CROCCO, 2020).

O termo autismo é derivado do grego *autos*, que significa “de si mesmo, em relação a si mesmo”. Foi utilizado pela primeira vez em 1911, pelo psiquiatra suíço Paul Bleuler, como uma das características comumente observadas em crianças com esquizofrenia (KUHN, 2004). Com uma definição ainda muito influenciada por Bleuler, a psiquiatra soviética Grunya Sukhareva (ou Ssucharewa, a depender da transliteração) utilizou o termo autismo, em 1926, para descrever com detalhes as características autistas de seis crianças com “transtorno de personalidade esquizoide” (POSAR; VISCONTI, 2017). Em retrospecto, o estudo de Sukhareva surpreende pela descrição de características relacionadas aos critérios de diagnóstico adotados hoje pelo DSM-5:

Todas as crianças afetadas se mantêm separadas de seus pares, têm dificuldade de se adaptar e nunca são completamente elas mesmas entre outras crianças. [...] manifestam uma tendência à solidão e evitam outras pessoas a partir da primeira infância; elas se mantêm separadas, evitam jogos comunitários. [...] existe uma certa platidão e superficialidade das emoções. [...] uma tendência ao automatismo que se manifesta como adesão às tarefas que foram iniciadas e como inflexibilidade psíquica com dificuldade de adaptação à novidade. [...] falta de expressividade facial e de movimentos expressivos. (SSUCHAREWA, 1926b, p. 255 – 256, tradução do autor).

Além disso, Suhkareva menciona a presença de alterações sensoriais, como sensibilidade ao ruído (G. E. SSUCHAREWA, 1926a), uma característica incluída apenas recentemente nos critérios diagnósticos.

Deve-se mencionar que, apesar do pioneirismo, as descrições de Suhkareva só começaram a ganhar reconhecimento recentemente (MANOUILENKO; BEJEROT, 2015). Até então, a descrição de “autistas psicopatas”, por Hans Asperger em 1944 (LYONS; FITZGERALD, 2007), e de crianças com diminuída capacidade de estabelecer contato afetivo, por Leo Kanner em 1943 (KANNER, 1968), eram reconhecidas como as primeiras descrições detalhadas do TEA.

1.1.1. Prevalência

As primeiras investigações epidemiológicas do TEA foram realizadas no final dos anos da década de 1960 e início da década de 1970. Inicialmente, a prevalência estimada em estudos europeus era em torno de 5 a cada 10.000 indivíduos na população (GILLBERG; WING, 1999). No ano de 2000, o órgão Centros de Controle e Prevenção de Doenças dos Estados Unidos da América (*Centers for Disease Control and Prevention – CDC*), estabeleceu o programa Monitoramento do Autismo e das Deficiências de Desenvolvimento (*Autism and Developmental Disabilities Monitoring - ADDM*) para coletar dados, estimar a prevalência, e descrever a população infantil com TEA nos EUA. Os primeiros resultados do monitoramento, referentes aos anos de 2000 e 2002, revelaram que aproximadamente 1 a cada 150 crianças de até 8 anos de idade era diagnosticada com autismo. Atualmente, a prevalência estimada para os EUA é de 1 a cada 44 crianças com até 8 anos de idade (MAENNER *et al.*, 2021). No mundo, estima-se que o TEA possa afetar 1% da população (ZEIDAN *et al.*, 2022).

O aumento observado da prevalência ao longo dos anos é multifatorial e ainda não está completamente esclarecido. Entre conhecidos fatores de influência estão a diferença metodológica dos monitoramentos, a mudança dos critérios de diagnóstico, aumento da conscientização pública e aumento do esforço de detecção (ZEIDAN *et al.*, 2022). Em uma proposta de quantificar o efeito não-etiológico associado ao aumento da prevalência, Hansel, Schendel e Parner (2015) estimaram que 60% do aumento observado entre 1980 e 1991 na Dinamarca podia ser atribuído à mudança dos critérios de diagnóstico e inclusão de pacientes não-ambulatoriais.

Outro ponto de atenção sobre a prevalência está relacionado à aparente maior incidência em indivíduos de sexo biológico masculino do que feminino (MAENNER *et al.*, 2021). Tal proporção pode variar de 0,8 a 6, a depender dos métodos de estudo (ZEIDAN *et al.*, 2022). Entre as razões especuladas estão possíveis fatores genéticos e hormonais (BARON-COHEN *et al.*, 2011), e o subdiagnóstico de pessoas do sexo biológico feminino devido à falta de sensibilidade dos protocolos de avaliação para características dessa população (LOOMES; HULL; MANDY, 2017).

A influência de fatores sociodemográficos, de raça e de etnia também não está completamente elucidada. O fato da maior parte dos estudos epidemiológicos de larga escala serem realizados em países do norte global dificulta a comparabilidade com estudos mais localizados e com menor tamanho amostral (ZEIDAN *et al.*, 2022). No mais recente relatório da ADDM, a prevalência entre crianças de origem hispânica foi menor do que a de crianças brancas ou negras, enquanto crianças negras foram mais diagnosticadas com deficiência intelectual (MAENNER *et al.*, 2021). A provável causa dessa variação não parece ser de origem etiológica e sim da interação entre os determinantes sociais de saúde, os quais modificam os padrões de busca e acesso ao diagnóstico e aos cuidados (ZEIDAN *et al.*, 2022).

1.1.2. Fatores de risco

Apesar da etiologia ser desconhecida, o TEA possui uma complexa e evidente combinação de fatores de risco genéticos e ambientais (CHASTE; LEBOYER, 2012; GOTTFRIED *et al.*, 2015). As evidências genéticas do transtorno começaram a ser desvendadas com estudos em gêmeos. Folstein e Rutter (1977) observaram que a incidência de TEA entre irmãos era 50 vezes superior à média, sendo que gêmeos monozigóticos possuíam maior probabilidade de compartilhar o diagnóstico do que gêmeos dizigóticos. Além disso, o TEA parece compartilhar uma carga hereditária com outros transtornos do neurodesenvolvimento, como evidenciado pela observação de que um irmão gêmeo monozigótico de uma criança diagnosticada com TEA tem 44% de probabilidade de ser diagnosticado com TDAH. Entretanto, a probabilidade é de 15% para gêmeos dizigóticos (LICHTENSTEIN *et al.*, 2010). Adicionalmente, algumas doenças sindrômicas de causa monogênica apresentam elevado risco de comorbidade com TEA. As mais estudadas são a síndrome do X frágil, causada por mutação do gene *FMR1* (KAISER-MCCAW *et al.*, 1980); a síndrome de Rett, por

mutação do gene *MECP2* (MELONI *et al.*, 2000); e a esclerose tuberosa, causada por mutações nos genes *TSC1* ou *TSC2* (GREEN; JOHNSON; YATES, 1994).

Diversos genes de risco já foram implicados no transtorno. Recentemente, um estudo de sequenciamento de exoma demonstrou a maior frequência de 53 variantes *de novo* disruptivas em indivíduos com TEA (SATTERSTROM *et al.*, 2020). Apesar da gama de alterações já associadas, os genes identificados podem ser relacionados principalmente à formação sináptica e regulação da transcrição gênica (DE RUBEIS *et al.*, 2014; SATTERSTROM *et al.*, 2020). Entre os genes que codificam proteínas sinápticas, destacam-se a família das neuroliginas, proteínas de adesão celular pós-sinápticas (TROBIANI *et al.*, 2020); e das SHANK, principais proteínas de estruturação da densidade pós-sináptica de sinapses glutamatérgicas (MONTEIRO; FENG, 2017). Em relação à transcrição gênica, genes de risco comumente associados são *UBE3A*, uma ubiquitina-proteína ligase e co-ativadora transcricional (VATSA; JANA, 2018); e *CHD8*, uma helicase que atua na remodelação da cromatina (XU, Q. *et al.*, 2018). Em conjunto, as observações fortalecem o caráter poligênico e heterogêneo do TEA.

Entre os fatores de risco ambientais, destaca-se a exposição materna a substâncias medicamentosas ou teratogênicas e outros eventos adversos durante a gestação. Nessa categoria, uma das associações mais robustas documentadas é o uso do medicamento ácido valproico (VPA), um anticonvulsivante e estabilizador de humor, durante a gestação. O uso da substância está associado ao aumento do risco de malformações congênitas, atrasos no desenvolvimento, prejuízo cognitivo e TEA (GENTILE, 2014; ROULLET; LAI; FOSTER, 2013).

A infecção materna durante a gravidez é também um conhecido fator de risco ambiental para o desenvolvimento de transtornos do neurodesenvolvimento ou neuropsiquiátricos, tais como TEA, esquizofrenia e transtorno bipolar (BROWN; MEYER, 2018). A estimulação do sistema imunológico materno, independente do agente infeccioso, pode desencadear uma cascata de respostas celulares e humorais capazes de interferir no neurodesenvolvimento (ESTES; MCALLISTER, 2016; PARKER-ATHILL; TAN, 2010). A combinação de fatores genéticos de predisposição e da janela de exposição durante a gestação é uma das principais hipóteses para o desencadeamento de transtornos distintos com causa semelhante (ESTES; MCALLISTER, 2016; KNUESEL *et al.*, 2014).

1.1.3. Diagnóstico

Os critérios de diagnósticos atualmente apresentados no DSM-5 são focados nos sintomas presentes em apenas dois domínios: 1) déficits na comunicação social e na interação social. Indivíduos podem apresentar, por exemplo, prejuízos na reciprocidade emocional e comunicação não-verbal, dificuldades no início, desenvolvimento e manutenção de relacionamentos; 2) restrição de comportamentos e interesses e a presença de comportamentos repetitivos. Tais alterações podem incluir movimentos ou falas estereotipadas, dificuldade na flexibilização de rotinas, interesses restritos, hipo- ou hiper-responsividade a estímulos sensoriais, entre outras manifestações (AMERICAN PSYCHIATRIC ASSOCIATION, 2013).

Apesar dos diversos avanços na compreensão da neurobiologia e da genética no TEA, o diagnóstico permanece baseado na identificação e avaliação de sintomas e comportamentos clínicos (HYMAN *et al.*, 2020). Em geral, questionários dirigidos aos responsáveis estão entre os métodos de rastreamento mais utilizados no mundo (HYMAN *et al.*, 2020; MARLOW; SERVILI; TOMLINSON, 2019). No Brasil, a Lei nº 13.438 de 26 de abril de 2017 refere a obrigatoriedade da aplicação, nos primeiros dezoito meses de vida, de instrumentos ou protocolos para detectar, em consulta pediátrica, o risco para o desenvolvimento psíquico infantil (BRASIL, 2017).

Após a identificação do risco, geralmente a partir do rastreamento, observação pelos responsáveis ou pela comunidade escolar, deve-se proceder com a busca de um diagnóstico clínico (HYMAN *et al.*, 2020). No mundo, a média de idade para o diagnóstico é de 60 meses, ou 43 quando considerados apenas estudos com crianças menores de 10 anos (VAN 'T HOF *et al.*, 2021). A detecção precoce do TEA é fundamental, pois pode proporcionar um acompanhamento e cuidado terapêutico antecipado, aumentar as chances de melhoramento das habilidades linguísticas, cognitivas e sintomas associados (CLARK *et al.*, 2018; DAWSON; BURNER, 2011; ROGERS *et al.*, 2014). Além dos aspectos sociais e psíquicos, o TEA também chama atenção pelos efeitos econômicos. Estima-se que, nos EUA, os custos individuais de uma pessoa com TEA em saúde, educação e serviços especiais pode variar de 1,4 a 2,4 milhões de dólares estadunidenses (BUESCHER *et al.*, 2014). Dessa forma, o diagnóstico precoce e o contínuo auxílio à família e aos indivíduos com TEA são necessidades críticas da sociedade.

A complexidade etiológica e fenotípica do TEA impulsionou diversos esforços para a identificação de possíveis biomarcadores, os quais poderiam auxiliar no diagnóstico precoce, estratificação de subgrupos, e acompanhamento de respostas à tratamentos (KLIN, 2018). As propostas podem ser categorizadas em domínios tão diversos quanto os afetados no transtorno. Possíveis biomarcadores metabólicos incluem moléculas envolvidas com estresse oxidativo, como redução da glutatona oxidada no plasma e encéfalo (CHEN *et al.*, 2021; ROSE *et al.*, 2012). No componente imunológico, o aumento de citocinas pró-inflamatórias, como IL-6 e IL-1 β , é frequentemente observado no plasma ou soro de crianças com TEA (MASI *et al.*, 2015). Em termos anatômicos e funcionais, estudos de neuroimagem avaliaram alterações estruturais e de conectividade em áreas conhecidamente envolvidas no TEA, como amígdala, córtex e cerebelo, na busca de marcadores oriundos de técnicas não-invasivas (LI; KARNATH; XU, 2017). Como mencionado anteriormente, alterações genéticas são comuns no TEA e, portanto, também são grande foco de propostas de biomarcadores. Classificações diagnósticas com base em variação no número de cópias ou polimorfismos de nucleotídeo único, por exemplo, estão entre as mais propostas devido ao grau de hereditariedade observado no transtorno (AR *et al.*, 2022).

Apesar dos esforços, atualmente não existem biomarcadores de rastreamento, diagnóstico ou tratamento, validados em uso clínico para o TEA (AR *et al.*, 2022; GOLDANI *et al.*, 2014). O caráter heterogêneo do transtorno influencia na baixa concordância entre os achados e na aparente ausência de uma assinatura biológica única (KLIN, 2018). Ainda assim, a busca de biomarcadores contribui para evidenciar a complexa interação entre fatores genéticos e ambientais.

1.1.4. Alterações na expressão gênica

Os estudos iniciais sobre a contribuição genética para o TEA focaram principalmente nas alterações estruturais do DNA. Atualmente, uma nova gama de alterações é descrita a partir de estudos transcritômicos, ou seja, da expressão gênica. Apesar da heterogeneidade e das centenas de genes de risco envolvidos, as desregulações transcritômicas no TEA possuem padrões compartilhados e são consistentemente observadas no córtex frontal e temporal de indivíduos (GANDAL *et al.*, 2018; SUN *et al.*, 2016; VOINEAGU *et al.*, 2011). Quando analisados funcionalmente, os módulos de expressão gênica desregulados são em sua maior

parte relacionados à funções sinápticas, imunológicas, mitocondriais e gliais (HE *et al.*, 2019; LOMBARDO *et al.*, 2017; VOINEAGU *et al.*, 2011)

Recentemente, Gandal *et al.* (2022) expandiu a análise da transcrição cortical para um número abrangente de áreas, incluindo regiões sensoriais. A expressão gênica em onze áreas corticais de amostras *post-mortem* de indivíduos neurotípicos e com TEA foi analisada a partir do sequenciamento de RNA. Os resultados encontrados demonstram uma alteração generalizada no córtex, incluindo uma atenuação de esperadas diferenças transcritômicas entre diferentes regiões corticais. Além disso, as alterações da expressão gênica convergiram funcionalmente para a infrarregulação (*down-regulation*) de módulos associados à comunicação sináptica e barreira hematoencefálica, e suprarregulação (*up-regulation*) de astrócitos reativos e resposta imunológica.

As alterações transcritômicas no TEA não estão restritas a genes com função codificante. RNAs não-codificantes (ncRNAs) são moléculas de RNA não traduzidas em proteínas e vem sendo associadas a diversos transtornos, inclusive neuropsiquiátricos (NIE *et al.*, 2019). Entre as classes de ncRNAs destacam-se os miRNAs, pequenas moléculas de aproximadamente 22 nucleotídeos cuja principal função é atuar no silenciamento e repressão pós-transcricional da expressão de genes alvos (WAHID *et al.*, 2010).

No TEA, alterações em miRNAs e redes de correção foram demonstradas no encéfalo de pacientes, sendo que os alvos dos miRNAs estavam relacionados com conhecidos genes de risco para o transtorno (WU *et al.*, 2016). Recentemente, nosso grupo de pesquisa relatou níveis circulantes elevados dos miR-134-5p e miR-138-5p em indivíduos com TEA, sendo que o miR-134-5p também se encontrou elevado no sangue de ratos do modelo animal de TEA, caracterizando um importante resultado translacional (HIRSCH *et al.*, 2018). Em uma proposta de biomarcadores, Hicks *et al.* (2016) demonstrou que 14 miRNAs estão diferencialmente expressos na saliva de indivíduos com TEA e seus alvos preditos estão classificados em categorias funcionais de ativação da transcrição gênica e desenvolvimento neuronal.

RNAs longos não-codificantes (lncRNA), uma categoria de ncRNA de geralmente mais de 200 nucleotídeos, também já foram observados alterados no TEA. Ziats e Rennert (2013) identificaram mais de 200 lncRNAs diferencialmente expressos

em amostras post-mortem do córtex pré-frontal e cerebelo de indivíduos com TEA. Interessantemente, lncRNAs também estão alterados em células circulantes, como leucócitos periféricos, nas quais mais de 3.000 estão diferencialmente expressos (WANG *et al.*, 2015). Não surpreende que as possíveis vias reguladas por tais lncRNA incluem funções conhecidamente associadas ao TEA, como transmissão sináptica e migração celular cortical.

O corpo de evidências da desregulação transcritômica no transtorno impressiona pela diversidade dos achados moleculares. Não apenas a expressão gênica tradicional está alterada, como também de importantes reguladores dessa expressão, como os ncRNAs.

1.2. Modelos animais do Transtorno do Espectro Autista

Os modelos animais são essenciais para estudar doenças neurodegenerativas, neuropsiquiátricas e do neurodesenvolvimento. Idealmente, modelos podem auxiliar os pesquisadores a investigar a etiologia, os mecanismos biológicos e as respostas à intervenções de forma ética, econômica e controlada (MUKHERJEE *et al.*, 2022).

Modelos animais são geralmente avaliados sobre o prisma de três características principais: 1) validade de construto, ou seja, a relevância dos métodos pelos quais um modelo é construído. Deve-se buscar reproduzir a circunstância que, em humanos, leva a determinada condição; 2) validade de face, relacionada à habilidade do modelo de reproduzir características da condição estudada, sejam elas anatômicas, bioquímicas, moleculares ou comportamentais; 3) validade preditiva, ou seja, a similaridade de reações e respostas entre o modelo e o ser humano quando expostos a um tratamento farmacológico (CHADMAN; YANG; CRAWLEY, 2009; NESTLER; HYMAN, 2010; VAN DER STAAY, 2006).

1.2.1. Tipos de modelos

Diversos modelos animais de TEA foram criados para serem utilizados como ferramentas de estudo. A heterogeneidade do transtorno também é refletida na diversidade dos modelos, os quais buscam reproduzir as características comportamentais, anatômicas, bioquímicas e moleculares sob diferentes aspectos das validades apresentadas. A Tabela 1 ilustra as principais espécies de modelos invertebrados e vertebrados utilizados por pesquisadores que, coletivamente, fizeram contribuições significativas para compreender os mecanismos subjacentes do TEA (BANERJEE; BHAT; RIORDAN, 2014).

Tabela 1 – Análises fenotípicas nas principais espécies de modelos animais de TEA

Fenótipo	Modelos animais				
	Primata não-humano	Camundongo	Rato	Zebrafish	<i>Drosophila</i>
Análises genéticas		+	+	+	+
Análises moleculares		+	+	+	+
Comportamento repetitivo	+	+	+		
Comunicação social	+	+	+		
Cognição		+	+		
Prejuízos na vocalização		+			

Fonte: adaptada de (BANERJEE; BHAT; RIORDAN, 2014).

As categorias de classificação dos modelos de TEA estão em constante mudança devido à criação de novos modelos e popularização de alguns em detrimento de outros. Por exemplo, Belzung *et al.* (2005) classificou, no início dos anos 2000, os modelos em quatro categorias: 1) modelos de anormalidades neuropeptídicas; 2) modelos que reproduzem marcas epigenéticas que aumentam o risco de TEA em humanos; 3) modelos de lesões neonatais em áreas cerebrais afetadas em indivíduos com TEA; 4) modelos de doenças genéticas humanas associadas ao TEA. Atualmente, a maioria dos modelos animais de TEA mais utilizados pode ser categorizada amplamente em 1) modelos manipulados

geneticamente ou 2) induzidos ambientalmente (ERGAZ; WEINSTEIN-FUDIM; ORNOY, 2016).

Os modelos genéticos são geralmente reproduzidos pela manipulação de genes sindrômicos, como *Fmr1* e *Mecp2* (SCHROEDER *et al.*, 2017). Com o avanço das tecnologias de manipulação gênica, foram criados diversos modelos baseados em genes associados fenotipicamente, isso é, que não possuem uma doença humana reconhecida, mas que reproduzem características e validades de interesse ao TEA (ERGAZ; WEINSTEIN-FUDIM; ORNOY, 2016). Linhagens isogênicas de animais que exibem naturalmente comportamentos relacionados ao TEA também podem se enquadrar nesta categoria.

Os modelos ambientais são baseados na exposição, usualmente materna, a fatores de risco ambientais associados (PATTERSON, 2011). Nessa categoria, um dos modelos mais utilizados é o induzido por exposição pré-natal ao ácido valproico (VPA). Sua validade de construto é baseada nos efeitos teratogênicos e no aumento de risco para o TEA em seres humanos (CHRISTIANSON; CHESLER; KROMBERG, 1994; CHRISTENSEN *et al.*, 2013). Modelos baseados na ativação imunitária materna (AIM) também podem ser classificados na categoria ambiental. A maioria dos modelos de AIM avalia, sobre a prole, os efeitos do processo inflamatório e não dos vírus específicos (ERGAZ; WEINSTEIN-FUDIM; ORNOY, 2016). Dessa forma, a indução de uma resposta inflamatória sistêmica com imunoestimulante PolyI:C é uma das formas mais comuns de modelo (LAMMERT; LUKENS, 2019). A Tabela 2 lista os principais modelos animais utilizados na literatura, com destaque para aqueles que são analisados na presente tese.

Tabela 2 – Principais modelos animais genéticos e ambientais de TEA em camundongos

Modelo animal	Tipo	Descrição	Principais características	Referências
Ash1L	Genético	Gene associado. Silenciamento do gene <i>Ash1L</i> .	Prejuízos sociais; comportamentos repetitivos; prejuízos na poda sináptica; desequilíbrio excitatório/inibitório cortical.	(GAO <i>et al.</i> , 2021; QIN <i>et al.</i> , 2021; YAN <i>et al.</i> , 2022)
BTBR T+ tf/J	Genético.	Isogênico.	Prejuízos sociais; comportamentos repetitivos; alterações na expressão gênica; desequilíbrio excitatório/inibitório cortical.	(MCFARLANE <i>et al.</i> , 2008; MEYZA; BLANCHARD, 2017)
Fmr1	Genético.	Sindrômico (síndrome do X frágil). Nocaute do gene <i>Fmr1</i> .	Prejuízos sociais; comportamentos repetitivos; plasticidade sináptica hipocampal alterada.	(PIETROPAOLO <i>et al.</i> , 2011; ZANG <i>et al.</i> , 2009)
Kmt5b	Genético.	Gene associado. Silenciamento do gene <i>Kmt5b</i> .	Prejuízos sociais; disfunção sináptica; disfunção do reparo do DNA.	(WANG, Z.-J. <i>et al.</i> , 2021)
Mecp2	Genético.	Sindrômico (síndrome de Rett). Nocaute do gene <i>Mecp2</i> .	Comportamento social anormal; comportamentos repetitivos; redução da ansiedade.	(EZEONWUKA; RASTEGAR, 2014; GUY <i>et al.</i> , 2007)
Ptchd1	Genético.	Gene associado. Nocaute do gene <i>Ptchd1</i> .	Disfunção na estrutura sináptica excitatória; disfunção cognitiva; hiperatividade; déficits na filtragem sensorial auditiva.	(PASTORE <i>et al.</i> , 2022; UNG <i>et al.</i> , 2018)
Poly I:C	Ambiental.	AIM por exposição a poly I:C.	Prejuízos sociais; comportamentos repetitivos; disfunção sináptica.	(HADDAD; PATEL; SCHMID, 2020)
Shank3	Genético.	Gene associado eindrômico (síndrome de Phelan-McDermid). Nocaute do gene <i>Shank3</i> .	Prejuízos sociais leves; interesses restritos; disfunção sináptica.	(AMAL <i>et al.</i> , 2020; PEÇA <i>et al.</i> , 2011)
VPA	Ambiental.	Exposição pré-natal ao VPA.	Prejuízos sociais; comportamentos repetitivos.	(BARONIO <i>et al.</i> , 2015; KOLOZSI <i>et al.</i> , 2009)

Fonte: autoria própria (2022).

Notas: Destaque em **negrito** indica que o modelo foi analisado em um dos capítulos da presente tese.

1.3. RNAs circulares

1.3.1. Biogênese e características

Os RNAs circulares (circRNAs) são moléculas endógenas de ncRNA de fita simples que formam um *loop* covalentemente fechado nas extremidades, ao contrário das moléculas lineares de RNA (SALZMAN *et al.*, 2012). A descrição de moléculas circulares de RNA remonta aos anos 1970, quando foram identificados viroides de plantas (SANGER *et al.*, 1976). No mesmo período, RNAs com estrutura circular foram observados por microscopia eletrônica em células humanas (HSU; COCA-PRADOS, 1979). Até recentemente, acreditava-se que os circRNAs eram subprodutos do processo de *splicing* sem função biológica.

Descobertas substanciais sobre a biogênese e mecanismo de ação dos circRNAs foram feitas com os recentes avanços da tecnologia de RNA-Seq, como métodos de depleção de RNA ribossomal e algoritmos de análise bioinformática. No início dos anos 2010, diversos estudos demonstraram que os circRNAs são expressos em abundância em mamíferos, são o transcrito predominante de alguns genes e estão envolvidos na regulação da expressão gênica através de mecanismos complexos (JECK *et al.*, 2013; MEMCZAK *et al.*, 2013; SALZMAN *et al.*, 2012). Entre as características interessantes descobertas até então estão a estabilidade e resistência à degradação (ZHANG, Y. *et al.*, 2016), expressão específica em tecidos e células (MEMCZAK *et al.*, 2013a; SALZMAN *et al.*, 2013), papel como esponjas miRNA (HANSEN *et al.*, 2013a; PIWECKA *et al.*, 2017a), e capacidade de regular a expressão do gene parental (LI, Z. *et al.*, 2015).

Os circRNAs são transcritos pela RNA Polimerase II como transcritos primários (pré-mRNA ou lncRNA) e posteriormente sofrem um processo de circularização através de um *splicing* diferencial conhecido como *backsplicing*. O *backsplicing* parece ser o principal mecanismo de biogênese de circRNAs e é favorecido por elementos ALU invertidos presentes em íntrons (JECK *et al.*, 2013). Tais sequências favorecem a justaposição de dois íntrons de um transcrito, aproximando um sítio doador a um sítio aceptor de *splicing*. Diferente do *splicing* canônico, que une um sítio *upstream 5'* (doador) com um sítio *downstream 3'* (aceptor) em uma ordem sequencial para produzir um RNA, o *backsplicing* ocorre em uma orientação inversa que liga um sítio *downstream 5'* (doador) a um sítio *upstream 3'* (aceptor) para produzir um circRNA

(JECK *et al.*, 2013; ZHANG, X. O. *et al.*, 2016). A Figura 1 ilustra de forma simplificada esse processo.

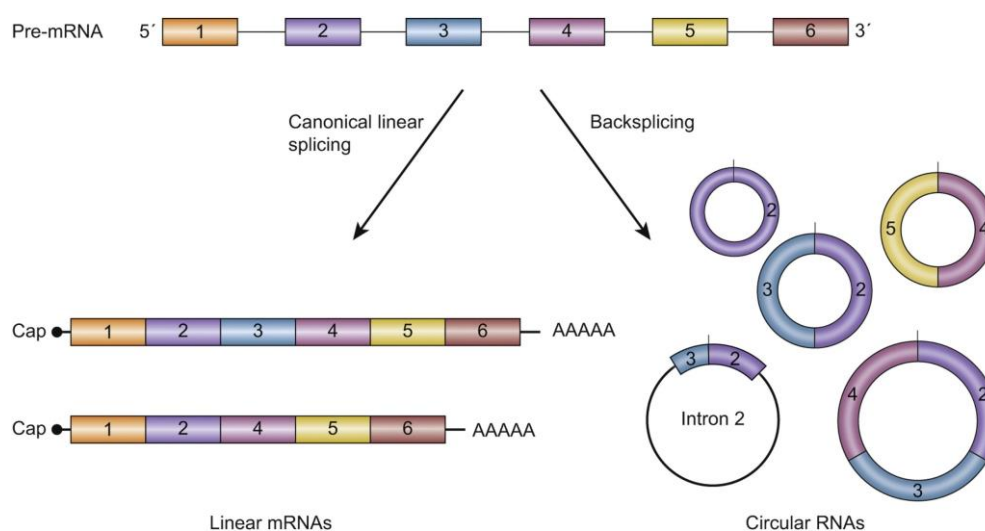


Figura 1 - Biogênese de circRNAs. Na esquerda, processo canônico de *splicing* que consiste na junção linear de éxons após a remoção de introns. Na direita, processo de *backsplicing* em consequência da junção de um sítio de *splicing downstream 5'* e *upstream 3'*. CircRNAs formados podem envolver apenas um éxon, múltiplos éxons e preservação de introns. Junção de *backsplicing* é indicada por barras cinzas. Fonte: (HUANG *et al.*, 2017).

A Figura 2 ilustra quatro funções conhecidas de circRNAs. A função mais conhecida, embora não seja a mais generalizável (YOU *et al.*, 2015), é a capacidade dos circRNAs de interagirem com miRNAs e causarem o aumento da expressão gênica dos alvos desses miRNAs. Essa função é conhecida como esponja de miRNAs, pois é observada a existência de múltiplos sítios de ligação para um miRNA específico (MEMCZAK *et al.*, 2013; PANDA, 2018) (Figura 2a). Alguns circRNAs também já tiveram demonstrada a sua capacidade de interagir com proteínas (Figura 2b-c). CircRNAs exônicos-intrônicos, como circEIF3J e circPAIP2, podem interagir com o complexo RNA-proteína U1 snRNP e promover a transcrição de seus genes parentais (LI, Z. *et al.*, 2015). Por fim, a presença de fases abertas de leituras e sítios internos de entrada do ribossomo indicam o potencial de tradução da sequência de alguns circRNAs em proteínas (SINHA *et al.*, 2022) (Figura 2d).

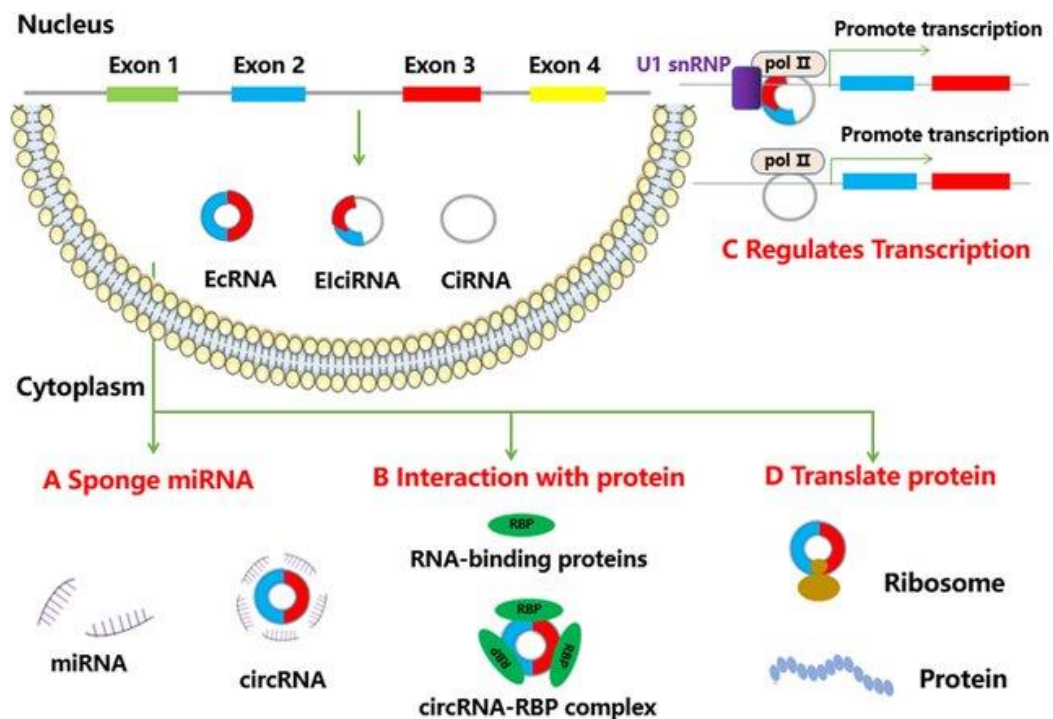


Figura 2 - Principais funções dos circRNAs. CircRNAs podem (a) atuar como esponjas de miRNA ao competir pela ligação dessas moléculas; (b) interagir com proteínas ligantes de RNA para exercer funções de regulação da transcrição ou transporte extracelular; (c) estimular a transcrição gênica de seus genes parentais através de complexos de regulação transcricional; (d) ser traduzidos em proteínas funcionais. EcRNA: circRNA exônico, ElciRNA: circRNA exônico-intrônico, CiRNA: circRNA intrônico. Fonte: (ZHANG, X. *et al.*, 2020)

1.3.2. Expressão gênica e funções no sistema nervoso central

Interessantemente, circRNAs são abundantes no sistema nervoso central (SNC). Além disso, muitos circRNAs são expressos exclusivamente no tecido nervoso e são provenientes de genes que codificam proteínas com funções sinápticas, como o receptor neurotrófico tirosina-cinase 2 (NTRK2) e a proteína de arcabouço pós-sináptica HOMER1 (MAASS *et al.*, 2017; YOU *et al.*, 2015). Além da regulação espacial, a expressão de circRNAs também é temporalmente regulada no desenvolvimento e envelhecimento de mamíferos. Durante o neurodesenvolvimento do neocórtex fetal humano, 39% dos genes expressam circRNAs, além dos respectivos transcritos lineares (CHEN; HUANG; JANITZ, 2019). Curiosamente, o número de diferentes circRNAs encontrados no encéfalo adulto é aproximadamente 25% menor do que o encontrado no encéfalo fetal, sugerindo um importante papel dessa classe de ncRNAs no neurodesenvolvimento (XIA *et al.*, 2017).

Até o momento, o envolvimento de circRNAs ou do eixo de interação circRNAs-miRNA-RNA já foi demonstrado em diversas doenças humanas, especialmente

carcinomas, e sua implicação em condições e patologias do SNC vem sendo desvendada através de estudos em modelos animais e tecidos humanos (RONG *et al.*, 2017). O primeiro circRNA com função de esponja descrito, ciRS-7 (*circular RNA sponge for miR-7*, ou CDR1as – *Cerebellar Degeneration-Related Protein 1 antisense*), possivelmente participa de processos-chave em condições neuropsiquiátricas e neurodegenerativas. O nocaute experimental do *locus Cdr1os* em camundongos provoca alterações na transmissão sináptica e prejuízos sensório-motores, um fenótipo característico de transtornos neuropsiquiátricos e do neurodesenvolvimento, como o TEA (PIWECKA *et al.*, 2017). O ciRS-7 atua como regulador negativo do miRNA miR-7 e sua redução, e consequente aumento da expressão de miR-7, parece estar relacionada à prejuízos na taxa de depuração de peptídeos β -amiloide na doença de Alzheimer (ZHAO *et al.*, 2016). Adicionalmente, a interação ciRS-7-miR-7 também pode estar relacionada à regulações da expressão da proteína alfa-sinucleína na doença de Parkinson (JUNN *et al.*, 2009).

1.3.3. RNAs circulares no Transtorno do Espectro Autista

Mais recentemente, uma nova camada de complexidade foi adicionada ao TEA por dois estudos demonstrando que a expressão de circRNAs está alterada em amostras de encefálicas de indivíduos com TEA. Gokool, Anwar e Voineagu (2020) demonstraram que redes de co-expressão de circRNAs envolvidos com sinapses excitatórias e inibitórias estão suprarreguladas, sendo essa a primeira identificação de alterações na expressão de circRNAs no transtorno. Adicionalmente, Chen *et al.* (2020) identificaram a expressão diferencial de 60 circRNAs no córtex cerebral, com possíveis alterações em eixos regulatórios circRNA-miRNA-mRNA envolvidos na expressão de proteínas sinápticas inibitórias.

Esse recente e ainda escasso corpo de evidências para o envolvimento de circRNAs no TEA é reforçado pela identificação em dois modelos animais do transtorno. Gasparini *et al.* (2020) detectou mais de 4.000 candidatos no hipocampo do modelo BTBR. Considerando os circRNAs expressos diferencialmente, os camundongos BTBR tiveram uma diminuição geral na expressão. Wang *et al.* (2021) mostrou que mais de mil circRNAs estão diferencialmente expressos no modelo animal de TEA induzido por VPA e que as redes de interação previstas estão

relacionadas com vias biológicas alteradas no transtorno. Em conjunto, os quatro estudos mencionados constituem, até o momento, os indícios do envolvimento de circRNAs no TEA, seja em seres humanos ou em modelos animais do transtorno.

1.4. Identificação de RNAs circulares por softwares de bioinformática

Conforme mencionado anteriormente, os avanços tecnológicos das técnicas de sequenciamento de RNA e o estabelecimento de protocolos de depleção de RNAr contribuíram para a obtenção de dados propícios à identificação de circRNAs através de análises por bioinformática (KULCHESKI; CHRISTOFF; MARGIS, 2016). Diversos métodos e algoritmos para identificação de circRNAs em dados de RNA-Seq vêm sendo propostos, a maioria com o objetivo de detectar a junção de *backsplicing*, onde há uma ordem inesperada de éxons (JECK; SHARPLESS, 2014).

1.4.1. Principais softwares

Entre as principais ferramentas disponíveis na literatura, destacam-se o find_circ, CIRCexplorer, CIRI e CirComPara. O find_circ é uma abordagem computacional para identificar circRNAs *de novo* em qualquer região do genoma através da detecção de fragmentos de leitura (*reads*) alinhados na orientação reversa, diferente do alinhamento linear (MEMCZAK *et al.*, 2013). O CIRCexplorer também possui como estratégia a detecção de reordenação de éxons indicativa de junção de *backsplicing*, necessitando do genoma de referência para realizar um alinhamento convencional e, com os *reads* não mapeados resultantes, um alinhamento em ordem reversa (ZHANG, X. O. *et al.*, 2016). O *pipeline* CIRI propõe um robusto método de identificação *de novo* de candidatos, particularmente útil para genomas com problemas de anotação (GAO; WANG; ZHAO, 2015). Por fim, o CirComPara é uma ferramenta bioinformática automatizada para comparação dos múltiplos métodos de detecção, anotação e quantificação de circRNAs citados acima (GAFFO *et al.*, 2017, 2022). Destaca-se que a combinação de múltiplos métodos é proposta como forma de diminuir o número de falsos positivos e direcionar a validação experimental dos circRNAs identificados (HANSEN *et al.*, 2015).

1.4.2. Usabilidade de softwares de identificação de circRNAs

A usabilidade pode ser definida como o grau em que usuários podem usar um produto ou sistema para atingir metas estabelecidas com eficácia, eficiência e satisfação em um contexto de uso específico (ISO/IEC, 2010). Questões de usabilidade são um obstáculo para uma experiência de usuário bem sucedida e podem levar a interações frustrantes e demoradas com um produto de software (BOLCHINI *et al.*, 2009). Este fator é dividido em seis subcaracterísticas: reconhecimento da adequação, capacidade de aprendizagem, operabilidade, proteção de erros do usuário, estética da interface do usuário e acessibilidade (ISO/IEC, 2010).

O número de ferramentas de identificação circRNA tem crescido rapidamente desde 2012. Até hoje, há mais de 40 programas de computador relacionados à identificação circRNA disponíveis em repositórios on-line (CHEN, L. *et al.*, 2020). Entretanto, as avaliações de qualidade dessas ferramentas são restritas a comparar métricas de desempenho, tais como precisão, sensibilidade e uso de memória e núcleos de processamento entre múltiplos algoritmos (HANSEN, 2018; HANSEN *et al.*, 2016; ZENG *et al.*, 2017; ZHANG, J. *et al.*, 2020).

Curiosamente, alguns problemas potenciais de usabilidade podem ser observados a partir de estudos anteriores. Apesar de não descrever em detalhes, Hansen *et al.* (2016) relataram falha na implementação dos softwares de identificação de circRNA KNIFE (SZABO *et al.*, 2015) e segemehl (HOFFMANN *et al.*, 2014), apesar de executarem com sucesso outros algoritmos no mesmo conjunto de dados. Zeng *et al.* (2017) enfatizaram os problemas de instalação existentes para as ferramentas de identificação circRNA com mais dependências e a necessidade de desenvolver guias de usuário abrangentes e interfaces agradáveis. Para demonstrar a eficácia de seu sistema dedicado a melhorar a reprodutibilidade de análises bioinformáticas, Menegidio *et al.* (2018) optaram por implementar o *pipeline* de identificação CirComPara (GAFFO *et al.*, 2017) devido aos vários problemas de dependências de instalação que podem ser encontrados.

Chen *et al.* (2020) fornece um panorama geral das características das ferramentas dedicadas à identificação e análise de circRNAs. Quase todas possuem o código-fonte público na plataforma GitHub e as principais linguagens de

programação utilizadas são Python, R e Perl. A maior parte dos *pipelines* necessita de um sistema operacional baseado em Linux e possui algum tipo de gerenciador de instalação e dependências, como Conda, Python package index (pip) ou BiocManager (Bioconductor). Entretanto, é destacado que, embora a instalação possa ser simples, algumas habilidades em informática podem ser necessárias (por exemplo, programação, conhecimento do sistema Linux). Nesse sentido, ferramentas com maior usabilidade e interfaces bem projetadas são geralmente necessárias para usuários sem treinamento computacional avançado (CHEN, L. *et al.*, 2020).

1.5. Qualidade de software

A ISO/IEC 25010 define a qualidade de um sistema como o grau no qual o sistema satisfaz as necessidades declaradas e implícitas de seus vários interessados, fornecendo assim valor (ISO/IEC, 2010). De forma semelhante, o *Institute of Electrical and Electronics Engineers (IEEE) Standard Glossary of Software Engineering Terminology* define a qualidade dos produtos de software como 1) o grau em que um sistema, componente ou processo atende aos requisitos especificados e 2) o grau em que um sistema, componente ou processo atende às necessidades ou expectativas de um usuário (IEEE, 1990).

Diferentes modelos foram projetados para identificar quais características afetam a qualidade do software. Os primeiros modelos de qualidade foram propostos no final da década de 1970 e definiram fatores de qualidade similares como: correção, confiabilidade, integridade, usabilidade, manutenção, flexibilidade, reutilização e portabilidade (BOEHM; BROWN; LIPOW, 1976; MCCALL; RICHARDS; WALTERS, 1977). Atualmente, a ISO/IEC 25010, descreve um modelo de qualidade de produto composto por oito características. Entre essas características está a usabilidade, uma das principais propriedades que contribuem para a qualidade do software (FAULKNER; CULWIN, 2000; RODER, 2012; XENOS, 2001).

1.5.1. Métodos de avaliação de usabilidade de software

Os métodos para identificar problemas de usabilidade podem ser amplamente categorizados em métodos de inspeção (sem a participação do usuário final) e

métodos de teste (ou empíricos, com a participação do usuário final). O método de inspeção mais comum é a avaliação heurística, na qual diferentes avaliadores inspecionam e identificam potenciais problemas de usabilidade guiados por uma lista de princípios, ou seja, heurísticas (LAUESEN; MUSGROVE, 2005).

As heurísticas mais utilizadas na literatura para avaliar a usabilidade são as de Jakob Nielsen. Os trabalhos de Nielsen (NIELSEN, 1992, 1994; NIELSEN; MOLICH, 1990a) buscaram sintetizar um pequeno conjunto de regras que abrangem os principais pontos a serem considerados quando se avalia a usabilidade de um software. A Figura 3 ilustra um exemplo simples e reconhecível de um problema de usabilidade. Segundo uma das heurísticas de Nielsen, boas mensagens de erro devem sempre auxiliar o usuário a compreender o que causou o problema e como é possível solucioná-lo. No cenário proposto, uma das mensagens não atinge esse objetivo, sendo, portanto, considerada como uma violação desse princípio.

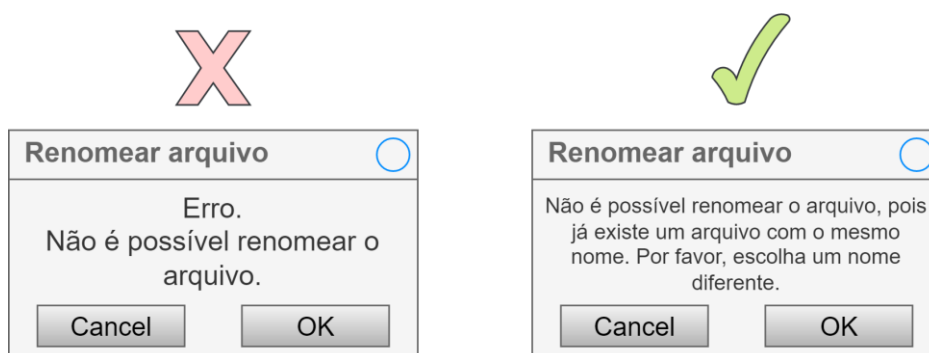


Figura 3 – Exemplo de uma violação heurística referente ao auxílio no diagnóstico e recuperação de erros. Considere um cenário no qual o usuário realiza a tentativa de renomear um arquivo do sistema. Na esquerda, a mensagem de erro exibida informa que não foi possível completar a ação, sem fornecer detalhes. Na direita, a mensagem exibida informa ao usuário a causa do erro e fornece uma solução. Fonte: autoria própria (2022).

Em relação aos métodos empíricos, os mais utilizados são os testes de usuário. Eles são caracterizados pela interação dos usuários finais com o software, seguindo uma lista de tarefas pré-definidas, onde a observação dessas interações ajuda a identificar problemas de usabilidade (PAZ; POW-SANG, 2016). A comparação entre estes métodos foi amplamente feita no passado, destacando os pontos fortes e fracos de cada um deles. Os testes com usuários frequentemente encontram problemas mais graves e reúnem o parecer de usuários-chave, enquanto a avaliação heurística encontra mais problemas de usabilidade em geral e é certamente menos dispendiosa

e mais fácil de executar várias vezes ao longo de um processo iterativo de desenvolvimento de software (JEFFRIES *et al.*, 1991; MAGUIRE; ISHERWOOD, 2018; WANG; CALDWELL, 2002).

1.5.2. Qualidade e usabilidade de softwares de bioinformática

Garantir a qualidade de software em bioinformática tem sido uma preocupação de longa data (GIANNOULATOU *et al.*, 2014; KAMALI *et al.*, 2015). No entanto, a maioria dos estudos tem se concentrado apenas na reprodutibilidade e desempenho do computacional (BADER *et al.*, 2005; IOANNIDIS *et al.*, 2009; O'RAWE *et al.*, 2013). Análises comparativas são bastante comuns neste campo, mas, como observado por Aniba, Poch e Thompson (2010), estas avaliações se concentram principalmente na velocidade ou precisão computacional, negligenciando características igualmente importantes como confiabilidade, acessibilidade, portabilidade, compatibilidade e estabilidade.

Estudos de caso com ferramentas web de bioinformática aumentaram a conscientização sobre algumas questões de usabilidade. Ao empregar tanto métodos de inspeção como métodos empíricos, Bolchini *et al.* (2009) descobriram que formular corretamente uma consulta de pesquisa e interpretar os resultados estão entre os obstáculos mais comuns à usabilidade. Mannapperuma, Street e Waterworth (2019) utilizaram as famosas heurísticas de Nielsen (NIELSEN, 1994) para avaliar suas ferramentas web depois de aplicar princípios de design a seu processo de desenvolvimento de software. Um método interessante de avaliação de usabilidade foi desenvolvido por Mirel e Wright (2009) ao adaptar e integrar as heurísticas de Nielsen com processos organizacionais, mas novamente, o resultado focou apenas na interface do usuário. Preocupados com questões de usabilidade de software, List, Ebert e Albrecht (2017) propuseram dez regras gerais para melhorar a usabilidade em biologia computacional, enfatizando tópicos essenciais como formatação de dados de entrada e saída, seleção cuidadosa dos parâmetros expostos e material de documentação sob medida.

A maioria das avaliações de usabilidade e diretrizes propostas concentram-se na interface gráfica do usuário e nas ferramentas de visualização de dados a partir de recursos web. Ainda assim, vários softwares de bioinformática são, na verdade, *scripts*

de linha de comando, como os utilizados nas várias etapas dos *pipelines* de análise de RNA-Seq. Com base em sua própria experiência com problemas de usabilidade, Seemann (2013) propôs dez recomendações para o desenvolvimento de ferramentas de linha de comando de bioinformática, tais como validação de entradas e parâmetros, comunicação de erros e verificação de dependência. Para ajudar programadores iniciantes a construir um melhor software de linha de comando em bioinformática, Georgeson *et al.* (2019) desenvolveram a Bioinitio, uma ferramenta que automatiza a criação de projetos seguindo as melhores práticas de programação como análise de argumentos, registros, numeração de versões e teste de código.

Em síntese, problemas gerais de qualidade, e principalmente usabilidade, são há muito tempo conhecidos e documentados na bioinformática. O estado atual da área prejudica os usuários, consumindo tempo e esforço, e dificulta a reprodutibilidade pela falta de métodos para obter e utilizar software desenvolvido por terceiros (BELMANN *et al.*, 2015). Dado que a maior parte do software acadêmico é desenvolvido por grupos de pesquisa subfinanciados e sem treinamento formal em engenharia de software (LIST; EBERT; ALBRECHT, 2017; MANGUL *et al.*, 2019), é possível imaginar que esforços para qualificar esse processo possam oferecer benefícios a curto e longo prazo.

2. JUSTIFICATIVA

Modelos animais são amplamente utilizados no estudo do TEA, pois permitem a análise de diferentes fatores de indução causadores de um fenótipo em comum, o qual é semelhante ao TEA manifestado em seres humanos. O compartilhamento de fenótipos comportamentais e moleculares pode indicar um compartilhamento de alterações subjacentes de grande significado biológico, as quais podem ser chave para a compreensão da etiologia e da manifestação de características no transtorno. Considerando o contexto emergente do envolvimento dos circRNAs em diversas patologias, sua conhecida abundância durante o neurodesenvolvimento, e a escassez de estudos sobre o envolvimento dessas moléculas no TEA, é de fundamental importância uma investigação do referido tema.

A detecção de circRNAs em larga escala só é possível através do uso de ferramentas de bioinformática desenvolvidas para identificar candidatos a circRNAs em bibliotecas de sequenciamento de RNA. Além disso, a identificação confiável dessas moléculas necessita do uso de diversos softwares de bioinformática para a detecção por diferentes métodos. Entretanto, em bioinformática, sabe-se que existe um problema generalizado na usabilidade de ferramentas de software, o qual pode impactar a adoção e utilização correta das ferramentas por pesquisadores.

Dessa forma, considerando:

- O uso de modelos animais para o estudo do TEA;
- O potencial de envolvimento de circRNAs no transtorno;
- A dependência metodológica de softwares de bioinformática para identificação de circRNAs;
- A importância da usabilidade de software para as análises e a existência de relatos na literatura de problemas nesse quesito em ferramentas de identificação de circRNA.

Justifica-se a presente tese como uma proposta interdisciplinar e integrativa de análise do perfil de expressão de circRNAs em modelos animais de TEA e identificação dos principais problemas de usabilidade dos softwares utilizados.

3. OBJETIVOS

3.1. Objetivo geral

Analisar por bioinformática a expressão gênica de circRNAs em bibliotecas de RNA-Seq de tecido encefálico de diferentes modelos animais de TEA, assim como desenvolver um método de avaliação e melhoramento da usabilidade das ferramentas de bioinformática utilizadas para a identificação dos circRNAs.

3.2. Objetivos específicos

Capítulo I: análise do perfil de expressão de circRNAs em tecido encefálico de camundongos de diferentes modelos animais de TEA.

- Identificar, por bioinformática, os circRNAs presentes em diferentes regiões encefálicas de diferentes modelos animais de TEA;
- Analisar a expressão gênica diferencial dos circRNAs presentes em cada modelo;
- Analisar funcionalmente os circRNAs diferencialmente expressos;
- Identificar similaridades entre as alterações em diferentes modelos e entre seres humanos.

Capítulo II: análise da expressão gênica do circRNA ciRS-7/CDR1as no córtex pré-frontal de ratos do modelo animal de TEA induzido por exposição pré-natal ao VPA.

- Reconstruir *de novo* a sequência do circRNA ciRS-7 na espécie *Rattus norvegicus*;
- Avaliar a expressão gênica de ciRS-7 no córtex pré-frontal de ratos do modelo animal de TEA induzido por VPA;
- Construir possíveis redes de interação circRNA-miRNA-RNA_m para as diferentes hipóteses de regulação gênica promovidas pelo ciRS-7.

Capítulo III: avaliação de usabilidade de ferramentas de bioinformática de identificação de circRNAs.

- Identificar problemas de usabilidade em softwares de identificação de circRNAs a partir da literatura e de experiências internas do grupo de pesquisa;
- Desenvolver um método de avaliação de usabilidade baseado em heurísticas;
- Identificar e recomendar soluções para os principais problemas de usabilidade encontrados.

PARTE II

CAPÍTULO I

A comprehensive bioinformatics analysis of circRNA expression in the brain of distinct mouse models of autism

Situação: manuscrito preparado para submissão.

CAPÍTULO II

Resultados adicionais: ciRS-7/CDR1as is up-regulated in the prefrontal cortex of rats of the animal model of autism induced by valproic acid

Situação: dados preliminares.

CAPÍTULO III

Usability evaluation of circRNA identification tools: Development of a heuristic-based framework and analysis

Situação: artigo publicado em agosto de 2022.

Periódico: Computers in Biology and Medicine (Elsevier).

Versão publicada disponível em:

<https://doi.org/10.1016/j.combiomed.2022.105785>

TITLE

Usability evaluation of circRNA identification tools: development of a heuristic-based framework and analysis

AUTHORS AND AFFILIATIONS

Guilherme Bauer-Negrini^{a,b,c}, Guilherme Cordenonsi da Fonseca^{c,d}, Carmem Gottfried^{a,b,c}, Juliana Herbert^e.

^aTranslational Research Group in Autism Spectrum Disorder (GETTEA), Universidade Federal do Rio Grande do Sul (UFRGS). Rua Ramiro Barcelos, 2600, CEP: 90035-003, Porto Alegre - Rio Grande do Sul, Brazil

^bDepartment of Biochemistry, Universidade Federal do Rio Grande do Sul (UFRGS), Rua Ramiro Barcelos, 2600, CEP: 90035-003, Porto Alegre - Rio Grande do Sul, Brazil

^cNational Institute of Science and Technology on Neuroimmunomodulation (INCT-NIM), Instituto Oswaldo Cruz, Fundação Oswaldo Cruz. Av. Brasil, 4365, CEP: 21040-900, Rio de Janeiro - Rio de Janeiro, Brazil

^dBioinformatics Laboratory, National Laboratory for Scientific Computing (LNCC). Av. Getúlio Vargas, 333, CEP: 25651-075, Petrópolis - Rio de Janeiro, Brazil.

^eInformation Technology and Healthcare Management Graduate Program, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Rua Sarmiento Leite, 245, CEP: 90050-170, Porto Alegre, Brazil.

Corresponding author

© 2022. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <https://creativecommons.org/licenses/by-nc-nd/4.0/>
This is the accepted manuscript version. Final and published version can be found at: <https://doi.org/10.1016/j.combiomed.2022.105785>

Guilherme Bauer-Negrini

Department of Biochemistry, Universidade Federal do Rio Grande do Sul (UFRGS), Rua Ramiro Barcelos, 2600, CEP: 90035-003, Porto Alegre – Rio Grande do Sul, Brazil.

<negrini.guilherme@gmail.com >

ABSTRACT

Background and Objective: Circular RNAs (circRNAs) are endogenous molecules of non-coding RNA that form a covalently closed loop at the 3' and 5' ends. Recently, the role of these molecules in the regulation of gene expression and their involvement in several human pathologies has gained notoriety. The identification of circRNAs is highly dependent on computational methods for analyzing RNA sequencing data. However, bioinformatics software is known to be problematic in terms of usability. Evidence points out that tools for identifying circRNAs can have such problems, negatively impacting researchers in this field. Here we present a heuristic-based framework for evaluating the usability of command-line circRNA identification software. **Methods:** We used heuristics evaluation to comprehensively identify the usability issues in a sample of circRNA identification tools. **Results:** We identified 46 usability issues presented individually in four tools. Most of the issues had cosmetic or minor severity. These are unlikely to challenge experienced users but may cause inconvenience for novice users. We also identified severe issues with the potential to harm users regardless of their experience. The areas most affected were the documentation and the installability of the tools. **Conclusions:** With the proposed framework, we

© 2022. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <https://creativecommons.org/licenses/by-nc-nd/4.0/>
This is the accepted manuscript version. Final and published version can be found at: <https://doi.org/10.1016/j.combiomed.2022.105785>

formally describe, for the first time, the usability problems that can affect users in this area of circRNA research. We hope that our framework can help researchers evaluate their software's usability during development.

KEYWORDS: usability, bioinformatics, circRNA, heuristic evaluation, command-line interface.

1. INTRODUCTION

Usability is the degree to which specified users can use a product or system to achieve established goals with effectiveness, efficiency, and satisfaction in a specified context of use [1]. Usability issues are an obstacle to a successful user experience and can lead to frustrating and time-consuming interactions with a software product [2]. In bioinformatics, usability and general software quality issues are frequently observed in many applications. Some authors highlight as major underlying causes for these issues the lack of formal training in software engineering [3,4] and the lack of academic incentives and economic resources for quality assurance [5].

We believe that usability issues can be especially problematic in a rapidly growing field of study, such as circular RNA (circRNA) research, because novice users from different backgrounds encounter an extra barrier to using new software tools. Based on our experience using circRNA identification tools and related software and inspired by observations of usability problems reported in previous studies in the literature, we believe that these issues are negatively impacting researchers. To the best of our knowledge, no thorough usability evaluations have been performed on circRNA identification tools. Moreover, there

are no specific methods to help authors check the usability of their circRNA-related software during development. The existing general usability principles and best practices guidelines for bioinformatics software provide insightful help to software developers but can fail to identify domain-specific issues effectively. Therefore, we propose a heuristic-based framework for evaluating the usability of circRNA command-line identification software and comprehensively identifying the usability issues in a sample of circRNA identification tools. First, we review the software quality and usability in the bioinformatics field. We also present the current state of software quality evaluations on general circRNA identification tools and highlight potential usability issues already observed by other authors. Next, we describe the development of our framework and its use to identify usability issues in circRNA identification tools. Finally, we discuss the main usability issues found and how to reduce their impact on future circRNA research.

2. BACKGROUND

2.1. Software quality and usability

The ISO/IEC 25010 standard defines the quality of a system as the degree to which the system satisfies the stated and implied needs of its various stakeholders, therefore providing value [1]. The standard describes a product quality model comprised of eight quality characteristics: 1) Functional Stability; 2) Performance Efficiency; 3) Compatibility; 4) Usability; 5) Reliability; 6) Security; 7) Maintainability; 8) Portability.

The usability factor has been present since the first quality models, and it is considered one of the major characteristics contributing to software quality [6–

8]. Moreover, usability is detailed into six sub-characteristics: appropriateness, recognizability, learnability, operability, user error protection, user interface aesthetics, and accessibility (ISO/IEC 25010). Methods for finding usability problems can be broadly categorized into inspection methods (without end-user participation) and test (or empirical) methods (with the end-user involvement). The most common inspection method is heuristic evaluation, in which different evaluators inspect and identify potential usability problems guided by a list of principles, *i.e.*, heuristics [9]. The main advantage of heuristic evaluation is that it is a discount method, *i.e.*, it requires fewer resources and time than user testing [10].

2.2. Software quality and usability in bioinformatics software

Ensuring software quality in bioinformatics has been a long-standing concern [11,12]. However, most studies have focused only on reproducibility, computer performance, or efficiency issues [13–15]. Benchmarking analyses are relatively common in this field, but, as stated by Aniba et al. (2010), these assessments mainly focus on computation speed or accuracy while neglecting equally essential features such as reliability, accessibility, portability, compatibility, and stability.

Case studies on bioinformatics web tools have increased awareness about some usability issues. Bolchini et al. (2009) found that properly formulating a search query and interpreting search results are the most common usability obstacles in bioinformatics web resources. Mannapperuma et al. (2019) used the famous Nielsen's heuristics [18] to evaluate their web tools after applying design

principles to its development process. Concerned with software usability issues, List et al. (2017) proposed ten general rules for improving usability in computational biology stressing essential topics such as input and output data formatting, careful selection of exposed parameters, and tailored documentation material.

Most usability evaluations and proposed guidelines concentrate on web resources' graphical user interfaces and data visualization tools. However, most bioinformatics software is command-line scripts executed in a system terminal window. Based on his own experience with usability problems, Seemann (2013) proposed ten recommendations for developing bioinformatics command-line tools, such as input and parameter validations, error communication, and dependency checking. Similarly, to aid beginner programmers in building better bioinformatics command-line software, Georgeson et al. (2019) developed Bioinitio. This tool automates project creation following programming best practices like argument parsing, logging, version numbering, and code testing.

2.3. Quality assessment of circRNA identification tools

Circular RNAs (circRNAs) are endogenous molecules of non-coding RNA (ncRNA) that form a covalently closed loop at the 3' and 5' ends, unlike linear RNA molecules [21]. Until recently, it was believed that circRNAs were by-products of the splicing process with no biological function. With recent advances in RNA sequencing technology (RNA-Seq) and the development of circRNA detection algorithms, substantial discoveries about circRNA biogenesis and its involvement in human health and disease have been made [22–24].

The number of circRNA identification tools has been rapidly growing since 2012. To date, there are more than 40 computer programs related to circRNA identification available at online repositories [25]. Nevertheless, quality evaluations of these tools are restricted to comparing performance metrics, such as precision, sensitivity, and RAM and CPU usage among multiple algorithms [26–29].

Interestingly, some potential usability issues can be observed from previous performance assessment studies. Although not describing it in detail, Hansen et al. (2016) reported failing to implement KNIFE [30] and segemehl [31] identification pipelines despite successfully running other algorithms on their sequencing datasets. Zeng et al. (2017) emphasized the existing installation problems for circRNA identification tools with more dependencies and the need to develop comprehensive user guides and user-friendly outputs. To demonstrate the effectiveness of their Docker contained system dedicated to improving reproducibility for bioinformatics analyses, Menegidio et al. (2018) chose to implement the circRNA identification pipeline CirComPara [33] because several dependency issues can be encountered during its implementation.

3. MATERIALS AND METHODS

3.1. Heuristic framework sources and development

Three sets of existing heuristics and guidelines were selected as references for developing the circRNA software domain-specific heuristics. They were: 1) Nielsen's (1994) ten usability heuristics for user interface design; List's et al. (2017) ten simple rules for developing usable software in Computational

Biology; 3) Seemann's (2013) ten recommendations for creating usable bioinformatics command-line software.

To broadly group similar heuristics into core principles, the above-mentioned sets of heuristics were mapped to ISO 25010 usability sub-characteristics (Supplementary Table 1). Based on that mapping, the ISO 25010 standard usability sub-characteristics was extended and shaped into a new model according to the following alterations and rationale:

- i. Appropriateness recognizability and learnability were combined into a new dimension named Hierarchical Documentation. Karimzadeh and Hoffman (2018) suggested that bioinformatics software documentation should consist of hierarchically organized components. Documentation with different levels of abstraction can help users recognize if the software is appropriate for their needs and learn how to use it effectively, therefore comprising the original ISO sub-characteristics definitions.
- ii. Operability and user error protection were maintained with definitions similar to their originals. They are related to fundamental aspects of usability and user satisfaction, *i.e.*, how easy it is to use the system and how users are protected from their errors.
- iii. User interface aesthetics was renamed Visibility, inspired by Nielsen's first heuristic. Command-line tools have limited interaction and aesthetic options. Users are often responsible for recovering from errors based on their information. Therefore, correctly displaying information and feedback in the terminal window is crucial.

- iv. Installability, originally a sub-characteristic of the portability aspect of the ISO 25010 product quality model, was added. Given that the end-user installs the software, the installability can affect the resulting functional appropriateness and operability (ISO/IEC 25010). It was considered a major factor to be accounted for since other authors have already reported installability issues in circRNA identification tools [28,32].
- v. Accessibility was removed. Usability and accessibility are usually considered separated concepts with complementary aspects [35]. Due to the limited interaction options between users and command-line interfaces, we assumed that accessibility is a factor mainly covered by the assistive tools and options provided by the operating system.

The final evaluation framework composed of five usability dimensions and 15 domain-specific heuristics designed by the authors is presented and described in the Results section.

3.2. Heuristic evaluation environment

A Docker image based on a Linux-based operating system (Ubuntu version 18.04 LTS) was used to guarantee a standardized environment for evaluating all circRNA identification tools. Each tool was evaluated in a clean environment (container), to prevent compatibility issues caused by the installation requirements of other tools. Essential utilities and packages (*wget*, *git*, *curl*, *vim*, *build-essential*, *pkg-config*, *software-properties-common*) were already installed for the evaluators. However, care was taken to preserve the image close to a

newly installed system. Scripts to install some libraries and packages (*htslib*, *conda*) were provided for evaluators if needed.

A small test dataset of *Homo sapiens* data was provided, including reference genome, annotation, one paired-ended library, and genome indexes (*bowtie*, *bowtie2*, *bwa*, *STAR*). Genome, annotation, and RNA-Seq data were originally obtained from CirComPara test data [33].

3.3. Heuristic evaluation process

Heuristic evaluation was performed independently by one of the authors and two external researchers, all trained in heuristic evaluation and bioinformatics. According to Nielsen (1992), two to three double specialists are sufficient to find between 80% and 90% of the usability problems. Four widely used circRNA identification tools were selected for evaluation (Table 1). These tools were already extensively evaluated in terms of computational performance and accuracy [26,28,29], but never usability.

Table 1 – circRNA identification tools selected for evaluation.

Name	Version	URL	Reference
find_circ	v1.2	https://github.com/marvin-jens/find_circ	Memczak et al., 2013
CIRI2	v2.0.6	https://sourceforge.net/projects/ciri/files/CIRI2 /	Gao et al., 2015; Yuan et al., 2018
MapSplice2	v2.2.1	http://www.netlab.uky.edu/p/bioinfo/MapSplice2	Wang et al., 2010

After a meeting for general guidance on the use of heuristics, the evaluators proceeded independently and registered the heuristic violations and severities into the evaluation sheet (Supplementary Table 2). Next, heuristic violations were compiled into a single list of issues for each tool. After, the evaluators discussed and reassessed the violations that 1) had not been found by all and 2) whose severity ratings differed by more than one level. Finally, the final severity of each violation was calculated as the mean severity score, *i.e.*, the arithmetic mean of the individual ratings. Mean severity score was rounded up to the nearest integer. Final severity scores of less than 0.5 were discarded.

4. RESULTS

4.1. Framework to evaluate the usability of circRNA-related tools

We have organized the proposed framework into a five-dimension usability evaluation model and its respective heuristics. We present the attributed heuristics with their name and a detailed description focused on the circRNA research domain for each dimension (Table 2).

Table 2 - Heuristic-based model for usability evaluation of circRNAs identification tools.

Id	Dimension/heuristics	Description
1	Hierarchical documentation	The degree to which the software documentation is presented hierarchically with different levels of abstraction and specific objectives.

H1	Comprehensive and hierarchically organized documentation	Ideally, technical details and benchmarking analysis should be described in a scientific publication. A quick start guide must include basic installation and usage instructions. Use-cases and tutorials can help users understand how a tool can be used. Built-in help (-h option) is appreciated as a readily available reference for command-line parameters. Code comments are suited for experienced users and collaborating software developers. Finally, to help users easily recognize if the software is appropriate for their needs, documentation is expected to address if the identification tool supports single-end and/or paired-end reads; uses <i>de novo</i> and/or annotation-based identification and detects exonic/intronic/exon-intron circRNAs.
H2	Software license and citation	A license document specifies guidelines for using and distributing software. It must be compatible with any third-party code that might have been used and should address if, or how, other developers can incorporate the existing code into their software. A citation source such as a published article or a repository's DOI must be provided so users can reference the tool used.
H3	Versioning protocol and information about updates.	Code can, and should, change since its first release as some functionalities are added or bugs are fixed. Changes must be clearly described when a new version is released, especially if there is any impact on the reliability of circRNAs identified by previous versions.
II Installability		The degree to which the software can be successfully installed and/or uninstalled by novice and advanced users.
H4	Software dependencies and package requirements	All installation requirements and versions must be clearly described as if users did not have any prerequisites installed, including programming language interpreter/compiler and even the appropriate operating system.
H5	Easy installation	The installation process frequently poses a significant usability barrier, but there are several ways to make it less troublesome to users. For example, a <i>makefile</i> can automate compilation; a requirements file can be used to pull and install dependencies from the programming language package manager, and a Docker image can package software and its dependencies as a container.

H6	Test data	A small and validated data set for testing must be available. A complete test run should uncover any missing requirements, environment incompatibilities, or conflicting results. This can also improve the operability of the software.
III Operability		The degree to which the software has attributes that make it easy to operate and control for novice and advanced users.
H7	Standard input and output file formats	Standardization is essential since it is highly recommended that users analyze the same data with multiple circRNA identification tools. Standard file formats used for RNA-Seq data analysis are FASTQ for sequencing reads, FASTA for reference genomes, GTF/GFF for genome annotation and SAM/BAM for alignment information. To output custom data, e.g., back-spliced junction read counts, tabular data formats such as comma-separated values or tab-delimited text are preferred.
H8	Supports input from both command-line interface and configuration file.	Command-line parameters are excellent for quickly running a tool with few inputs and allowing easy automation through bash scripts. Ideally, command-line parameters must be named. The commands can get long and complex depending on how many optional parameters and file paths are passed as input. Therefore, a template configuration file where users can set options and define directory paths to genome references, annotations, and indexes is advisable.
H9	Default parameters for novice users and optional parameters for more experienced users	A careful selection of default values should cover most usage scenarios to avoid confusing novice users with excessive customization. Advanced options are suited for skilled users who wish to have more control over performance metrics of circRNA detection and computational resource usage.
IV User error protection		The degree to which the software protects users from making errors.
H10	Validation of input parameters	Before starting the analysis, the tool must check if all mandatory parameters are set, if all paths to directories or files exist and if data types are as expected. Early checking of all parameters could avoid breaking downstream analyses of a sequential circRNA identification pipeline after an apparent successful start.

H11	Prevents accidental file overwrite	It is common for users to incorrectly set the same output file when running many samples or unintentionally re-run a previous analysis when browsing through command-line interface history. Raising a warning message and halting execution if an identifier for a given sample already exists avoids wasting users' time and computational resources.
V Visibility		The degree to which the software's activities and status are appropriately informed to the users.
H12	Updated information about the analysis progress	Users must be informed about the workflow with messages regarding which step is under execution and how many steps are left until completion. All information printed on the terminal should be stamped with the date and time data, especially when an analysis is completed. Reporting this information can help users plan their next experiment and decide if the average analysis duration suits their needs.
H13	Verbose mode	Printing too many lines can overwhelm users by making it difficult to follow the analysis steps. Therefore, information must be kept concise by default. However, a verbose mode should be available so that advanced users can access detailed information like the output from an underlying genome aligner or index builder algorithm.
H14	Detailed and intelligible error messages	When an error occurs, printing the stack trace can provide helpful insights for experienced users familiar with the programming language used. Nevertheless, whenever a predictable exception is caught, users should be informed of what went wrong and how to recover possibly.
H15	Logging option	A log file can help users document their experiments. The option to record details about each execution should include information such as software version, sample identifier, selected parameters, and terminal output.

A severity scale was adapted from Nielsen (1994) and updated with examples from the studied domain (Table 3).

Table 3 - Severity rating for usability problems.

Severity	Definition
<p>© 2022. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/ This is the accepted manuscript version. Final and published version can be found at: https://doi.org/10.1016/j.combiomed.2022.105785</p>	

0	None: not a usability problem.
1	Cosmetic: subtle problems; users can easily proceed; lower priority issue.
2	Minor: the issue has minor effects on usability; it may hinder novice users; most users can proceed based on their experience or by consulting the documentation.
3	Major: the issue has significant effects on usability; it can cause significant delay or frustration; users may not find information or help on documentation; users may proceed based on external help; some degree of programming skills may be required to proceed; novice users can experience several challenges and fail to proceed.
4	Critical: the issue may prevent the complete execution of the tool/pipeline; the issue can be solved by experienced users but requires advanced programming skills and/or several hours; users are unlikely to insist on using the tool.

4.2. Usability evaluation of circRNA identification tools

The evaluators identified 113 heuristic violations distributed among the four circRNA identification tools individually. After consolidating a list of unique violations for each tool, 51 were registered. For each violation, the evaluators recommended a possible solution. Five issues were discarded, and 46 issues remained after calculating the mean severity score of each violation (Table 4). A more detailed description, including discarded issues and code output errors where relevant, is available in Supplementary Table 3.

Table 4 - Compiled list of 46 heuristic violations present in four circRNA identification tools.

Tool	Issue	Recommendation	Mean severity score	Severity class
Hierarchical documentation				
<i>H1 - Comprehensive and hierarchically organized documentation</i>				
CIRI2	The documentation is not organized. CIRI repository contains other related tools. The difference between all tools is	Describe in the main README the differences among all tools. Detailed information about each	1.67	Minor

	not clearly described and could confuse novice users.	tool could be placed inside each folder.		
find_circ	The example commands do not run as expected. Samtools “-o” flag and “.bam” extension are missing from both documentation and makefile code.	Update the documentation with tested commands.	4.00	Critical
find_circ	The example commands do not run as expected. The command for getting the unmapped reads and piping them through unmapped2anchors.py fails to complete. The command must be split and bam index built to proceed. Samtools index build is not described in the documentation.	Update the documentation with tested commands.	4.00	Critical
KNIFE	The information is complete but scattered among different folders. This hinders the overall understanding.	A compiled manual/README should be available regarding all aspects of the tool.	2.00	Minor
MapSplice2	The most recent version is not available on the Downloads page, only on the home page. The user guide refers to the “latest version” but links to the outdated version.	Update the URL for the most recent version on all pages.	3.33	Major
MapSplice2	Use cases are not described. The tool can be used to detect circRNAs and other types of spliced products. Use cases could help elucidate which flags to set for each case.	Design a few use cases demonstrating the command-line arguments needed.	1.00	Cosmetic

H2 - Software license and citation

KNIFE	License is not available.	Choose and make available an appropriate license.	1.00	Cosmetic
-------	---------------------------	---	------	----------

H3 - Versioning protocol and information about updates.

CIRI2	Release notes are not provided for each new version.	Provide release notes with details about the new version.	1.67	Minor
-------	--	---	------	-------

Installability

H4 - Software dependencies and package requirements

CIRI2	BWA-MEM version not specified.	Specify in the documentation the tested versions of each dependency.	1.33	Cosmetic
find_circ	Dependencies versions are not specified (samtools, pysam, numpy, and bowtie). Different samtools versions could cause reported errors in H1 and H6.	Specify in the documentation the tested versions of each dependency.	1.33	Cosmetic
KNIFE	Dependencies and versions are listed. However, it is not described how to obtain or install them.	Describe the commands to install the dependencies or point to their repositories.	0.67	Cosmetic
MapSplice2	System requirements list g++ 4.3.3 or higher. However, the latest version (g++ 7) fails to compile the code.	Update the tool to work with an upgraded version of g++ or fix the installation instructions.	4.00	Critical
MapSplice2	"Python 2.4.3 or higher" could mislead users into thinking that Python 3.x would also work.	Describe that the tool only works with Python 2 versions.	1.33	Cosmetic
<i>H5 - Easy installation</i>				
CIRI2	The user must install BWA.	Dependencies could be packaged with the tool, or an installation script could be available.	1.00	Cosmetic
find_circ	The user must install dependencies. The process can be different for each dependency, harming novice users.	Dependencies could be packaged with the tool, or an installation script could be available.	1.67	Minor
KNIFE	Dependencies must be installed by the user. The process can be different for each dependency, harming novice users.	Dependencies could be packaged with the tool, or an installation script could be available.	2.00	Minor
KNIFE	Common transcriptomes and indexes are made available. However, users need to download files from a MEGA repository. This platform has a download limit. It is also not trivial or possible to download using the command line interface.	The index files should be available in a more convenient repository for users.	2.67	Major
MapSplice2	Makefile is present; however, it fails to build. The installation described in the documentation is not possible. Refer to the H4 issue. The installation had to be performed by unofficial	The problem seems related to fixing the g++/gcc version. Updating the code to work with newer versions of g++ is advisable.	4.00	Critical

instructions (package manager) for the evaluation to proceed.

H6 – Test data

find_circ	Test data is provided, but it fails to validate the pipeline. The commands in the makefile are missing flags (refer to H1 issues).	The makefile code should be revised.	3.33	Major
KNIFE	Test data is available, but it is not easy to find. Users must inspect the repository and manually download it from the v1.1 release.	Test data should be packaged with each release.	1.67	Minor
MapSplice2	Test data is not available.	Provide test data to help users validate the installation and use.	2.00	Minor

Operability

H7 - Standard input and output file formats

MapSplice2	The tool uses standard FASTA format, but the user is required to perform preprocessing steps to split one chromosome per file and rename reference genome files.	Support a more standard reference file containing all chromosomes in a single file.	2.00	Minor
------------	--	---	------	-------

H8 - Supports input from both command-line interface and configuration file

CIRI2	Supports only command-line.	Support commands from a config file. Configuration files are a common solution to organize parameters and values in data pipelines.	1.00	Cosmetic
find_circ	Supports only command-line.	Support commands from a config file. Configuration files are a common solution to organize parameters and values in data pipelines.	1.00	Cosmetic
KNIFE	Supports only command-line.	Support commands from a config file. Configuration files are a common solution to organize parameters and values in data pipelines.	1.00	Cosmetic
KNIFE	Parameters are not named; this impairs the command readability.	Change the input to named parameters. Positional parameters can confuse novice users.	1.67	Minor

MapSplice2	Supports only command-line.	Support commands from a config file. Configuration files are a common solution to organize parameters and values in data pipelines.	1.00	Cosmetic
------------	-----------------------------	---	------	----------

H9 - Default parameters for novice users and optional parameters for more experienced users

MapSplice2	Default parameters are defined, but the user must set the correct flags to detect circRNAs.	A use case for circRNA detection could be available at the docs.	0.67	Cosmetic
------------	---	--	------	----------

User error protection

H10 - Validation of input parameters

CIRI2	Some inputs are not checked. Example: -U and -E options accept non-numerical input and do not raise warnings.	The type of each input parameter should be checked before running.	1.00	Cosmetic
KNIFE	The tool raises an exception for files not found, but the execution is not halted.	Check all inputs before the analysis and stop the run if one is missing.	1.67	Minor
KNIFE	The mode parameter is not checked and accepts any string. Execution is not halted.	Check all inputs, including variable type.	1.33	Cosmetic
MapSplice2	"--gene-gtf" input is not checked and entering an invalid filename does not produce any warning or error. Users may think that no circRNAs were detected, but it was actually a filename error.	Validate --gene-gtf input if the parameter is set.	2.00	Minor

H11 - Prevents accidental file overwrite

CIRI2	Detects output file existence and informs the user but overwrites it.	Wait for user confirmation before overwriting the files.	2.00	Minor
find_circ	Overwrites the output directory at each run, not informing the user.	The tool should first check if the output folder already exists and raise a warning if so.	2.00	Minor
KNIFE	Overwrites the output directory at each run, not informing the user.	The tool should first check if the output folder already exists and raise a warning if so.	2.00	Minor
MapSplice2	Overwrites the output directory at each run, not informing the user.	The tool should first check if the output folder already exists and raise a warning if so.	2.00	Minor

Visibility

H12 - Updated information about the analysis progress

find_circ	Timestamp and total running time are not printed on the terminal.	Add timestamp and running time information.	1.33	Cosmetic
find_circ	Information about execution steps is not printed on the terminal.	Add information about the workflow.	1.33	Cosmetic

H13 - Verbose mode

find_circ	Not available, but the information printed on the terminal is hardly overwhelming.	A verbose flag could be set to print in real time the output from bowtie and samtools. Default information could be higher level and inform about the execution steps.	1.00	Cosmetic
-----------	--	--	------	----------

H14 - Detailed and intelligible error messages

find_circ	Messages from makefile errors are not clear enough.	Common errors are captured, and clear messages are displayed. However, errors from makefile could confuse most users (refer to H1 issue).	3.00	Major
KNIFE	Some error messages are unclear and do not inform users how to proceed. This can happen with uncaught errors of underlying tools (e.g., bowtie)	Test the bowtie error conditions to rewrite the messages for the captured errors.	2.33	Minor
MapSplice2	The error message is unclear, whether it is the directory or the file that is missing.	Separate file and folder checks to provide more precise information.	0.67	Cosmetic

H15 - Logging option

CIR12	An execution and a separate error log are provided. However, logging information is limited, and it does not contain information about the run. Additionally, some errors registered in the error log only contain information about the line where the error happened without further details.	Add more details, such as the parameters used and the sequence of calls that lead to the error.	0.67	Cosmetic
find_circ	Logging information is limited to bowtie output.	Include more information in the log file, such as commands used and errors.	1.67	Minor
KNIFE	Logs are not generated. Users must redirect the output to a file.	Logs should be generated by default.	1.33	Cosmetic

MapSplice2	A comprehensive log folder is present, and each step has its own log file. However, a general log registering the tool version and command/inputs could help users document their experiment.	Add a general log file with top-level information (command, inputs, version).	0.67	Cosmetic
------------	---	---	------	----------

Usability problems were found in all dimensions, with Installability being the most affected with 13 violations and Operability being the least affected with seven violations (Figure 1A). Individually, the most common sources of usability issues are comprehensive and hierarchically sorted documentation (H1), software dependencies and package requirements (H4), easy installation (H5), and supports input from both command-line interface and configuration file (H8) (Figure 1B). Moreover, problems in preventing accidental file overwriting (H11) and logging option (H15) also stand out as they are present in all the analyzed tools. Few problems were found related to software license and citation (H2), versioning protocol and information about updates (H3), standard input and output file formats (H7), default parameters for novice users and optional parameters for more experienced users (H9), and verbose mode (H13), which were violated by only one tool each (Figure 1B).

Most of the violations correspond to problems of cosmetic and minor severity, with all dimensions affected in these categories (Figure 1C). However, major and critical usability issues have also been observed. These severe usability problems are mainly concentrated in Installability and Hierarchical documentation dimensions. In particular, critical issues are only observed in comprehensive and hierarchically organized documentation (H1), software

dependencies and package requirements (H4), and easy installation (H5), which are also the most common usability problems (Table 4).

A detailed examination of the distribution of violations and severities among the tools reveals that all tools are affected by cosmetic and minor issues (Figure 1D). The critical problems affect the Hierarchical documentation of `find_circ` and the Installability of MapSplice2 (Figure 1D). These issues are related to incorrect commands described in the documentation (H1), incorrect versions of dependencies (H4), and failure of the installation process (H5) (Table 4). Another concern is the major severity issues present in all tools except CIRI2 (Figure 1D). In this case, usability problems are related to the availability of

outdated versions, difficulties in obtaining files from repositories, test data used, and unclear error messages (Table 4).

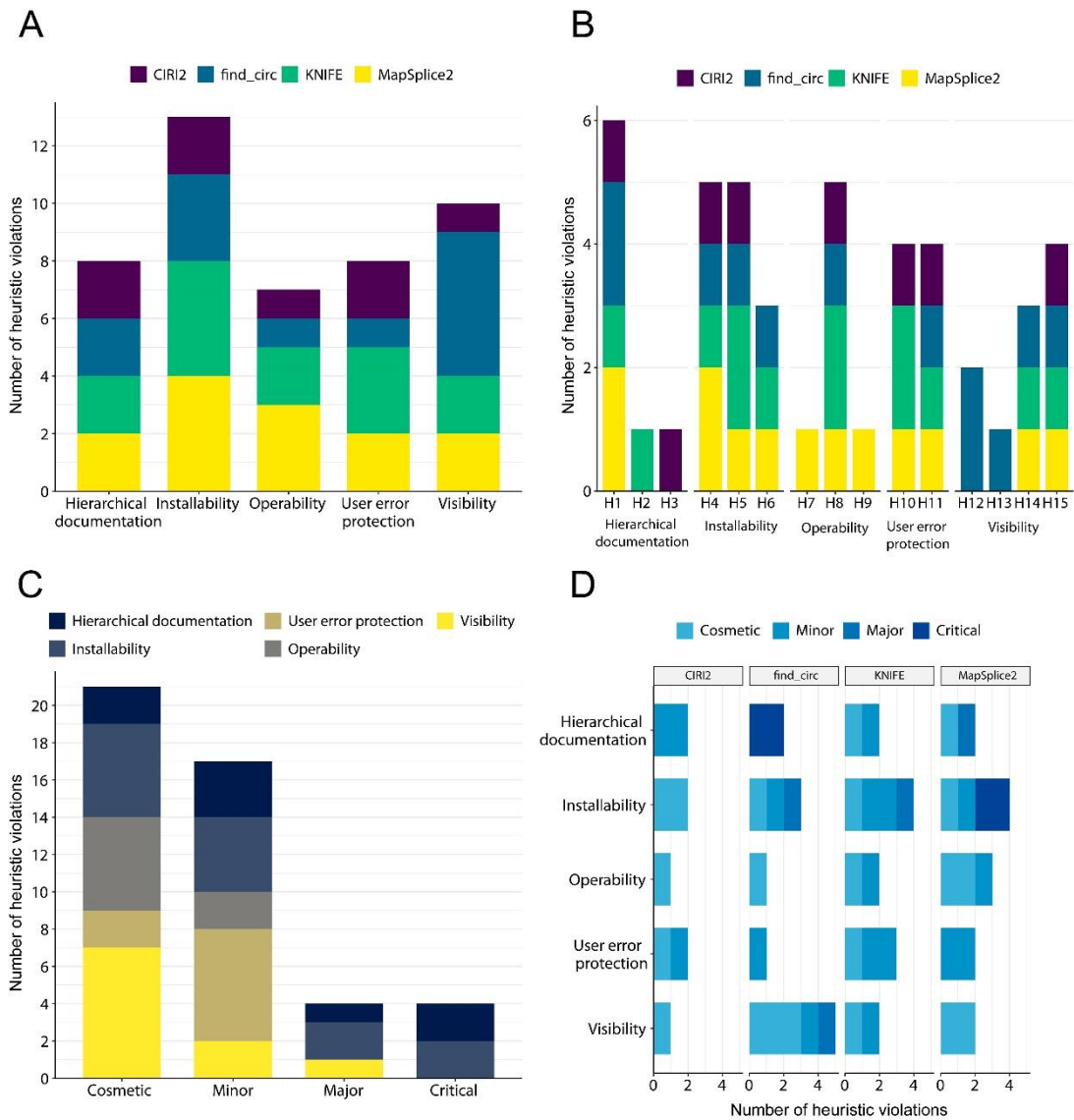


Figure 1 – Distribution of 46 heuristic violations of circRNA identification tools. The number of heuristic violations is portrayed from different aspects. (A) Heuristic violations across dimensions. Violations are present in all dimensions, and the contribution of each tool to the total is highlighted. **(B)** Heuristic violations by individual heuristics. The dimensions in (A) are broken down to the heuristic level to show the contribution of each tool and heuristic to the total. **(C)** Heuristic violations by severity. Cosmetic and minor issues are widespread among the dimensions, while major and critical problems are fewer in number and more concentrated in some dimensions. **(D)** Distribution of heuristic violations by severity across dimensions for each tool. Detailed

analysis shows that all tools are affected by cosmetic and minor issues, while major and critical violations are more concentrated on tools and dimensions.

5. DISCUSSION

General problems with bioinformatics software quality, especially usability, have long been known and documented [11,12]. Here, we focus on developing a methodology to evaluate the usability problems present in the specific domain of circRNAs identification tools. We have selected three sets of heuristics and recommendations from the literature as a starting point. The criteria for this selection were that Nielsen's heuristics usually serve as a basis for creating new heuristics as they present general concepts and are probably the most used heuristics for interface design. At the same time, the List's and Seemann's guidelines are general principles for developing computational biology software and command-line software, thus being closely related to circRNA identification tools. Our heuristics can be easily generalized to other domains of bioinformatics or scientific computing software based on a command-line interface. However, their definitions and characterizations are originally focused on concrete examples of the domain of circRNA identification tools.

The heuristic evaluation showed that Installability was the dimension with the highest number of aggregate violations. Moreover, this was the dimension most affected by issues classified as major or critical. Not surprisingly, the Installability of bioinformatics tools is known to be problematic. Installing dependencies is a source of great frustration for users of bioinformatics software [5,39]. Here we demonstrate that circRNAs identification tools often fail to specify

the required dependencies correctly. The consequences of such a problem can vary. In the best-case scenario, the dependencies used are stable enough not to cause critical errors. However, in severe cases, using incorrect dependencies prevents the installation or use of the tool, as observed with `find_circ` and `MapSplice2`. A general solution to these problems can be installation by package managers. This alternative usually provides a simplified process that removes the need for the user to obtain each dependency. As shown by Mangul et al. (2019), omics tools distributed without package managers are subject to installation problems in 32% of the cases. In fact, the `MapSplice2` distribution available in the Bioconda [40] package manager was the solution to proceed with our evaluation of the tool after the source installation failed. Another alternative is the containerization of software, which has already been demonstrated efficiently by Menegidio et al. (2018) to reduce the complexity of setting up the `CirComPara` circRNA identification pipeline.

Documentation problems were also frequent. Issues of this kind have the potential to be especially frustrating for novice users as the official documentation may represent their only source of information. A particularly critical case was observed with `find_circ` because some command examples failed to reproduce as expected. Aghajani et al. (2019) empirically demonstrated that code examples are the most frequent documentation correctness issues. The redundancy between the code to be tested and the documentation that must be updated creates the risk of inconsistency [41]; therefore, developers should have a well-planned release protocol. As Karimzadeh & Hoffman (2018) emphasized in their

comprehensive guide to documentation in bioinformatics software, complete
© 2022. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
<https://creativecommons.org/licenses/by-nc-nd/4.0/>
This is the accepted manuscript version. Final and published version can be found at:
<https://doi.org/10.1016/j.combiomed.2022.105785>

documentation does not solely benefit the user. Additionally, it benefits reviewers who must assess the reproducibility of the software during the peer-review process. Lastly, it also helps the developers by saving their time providing support to users.

The low observation of some problems highlights the maturity of the tools in many respects. Most of the repositories had a license, which is fundamental to establishing how users can use and modify the available software. Although this is not a criterion of our analysis, we see it as a positive fact that all licenses so far allow, with minor restrictions, the use and modification of software for research purposes. Another positive feature was the compliance with standardized input and output formats. The combination of different tools has been proven beneficial to increase the accuracy of circRNAs identification [27]. Therefore, standardization helps users repeatedly use the same input files.

This study has potential limitations. First, there is no agreement on the minimum number of evaluators needed for a heuristic evaluation. Some authors consider that five is enough to detect most issues [42,43]. However, Nielsen (1992) considers that this number can be reduced to three in the case of double experienced evaluators, as is the case in our study. Second, heuristic evaluation is not the only method for usability evaluation. User testing often finds more severe problems, while heuristic evaluation finds quantitatively more problems [44–46]. Ideally, both methods should be adopted as complementary practices. However, considering the particular characteristics of the software development environment in academia, heuristic evaluation seems to be the most cost-

effective alternative. In academia, scientific software is developed by small groups of students, often without formal training in software engineering or human-computer interaction, and with few resources to develop and maintain software products [4,5]. Therefore, a minimal set of heuristics has the potential to guide developers themselves in a low-cost usability evaluation that can, and should, be performed multiple times during the stages of the iterative software development process. Third, our goal was to identify representative issues and discuss solutions for the study area, not criticize any individual tool. We believe usability is a determining factor for the use of circRNA identification tools, but it should not be the only factor considered by users. Aspects such as precision and performance, which have already been evaluated in other studies, should also be considered by users according to their needs.

Future works on the usability of circRNA-related software should not be limited to identification pipelines. Vromman et al. (2020) described the lack of content overlap among multiple databases, ambiguous nomenclature, outdated content, and connectivity problems as significant issues observed in circRNA online databases. As circRNA investigations are relying heavily on computational methods, it is becoming more necessary to improve the usability of these bioinformatics software products. Experienced researchers could work more efficiently with more usable tools, and novice users may become more interested in exploring this growing field of study. Moreover, usable tools are more prone to adoption by the community and can have more citations [5].

6. CONCLUSIONS

We developed a set of 15 usability heuristics dedicated to circRNAs identification tools. With the proposed framework, we formally describe, for the first time, the usability problems that affect users in this area of circRNA research. Most of the problems were considered to have cosmetic or minor severity. Such issues are unlikely to prove disastrous for experienced users but may cause inconvenience for beginners. The heuristic evaluation also found some problems to be severe with a greater potential to harm most users. Software development in academia, including bioinformatics, does not always follow best practices and is not always funded with enough resources, which can affect the quality of the final product. We hope that our heuristic framework dedicated to circRNA identification tools can assist both users and developers to achieve their goals through more usable software.

ACKNOWLEDGEMENTS

The authors would like to thank the external heuristic evaluators for their careful and voluntary contribution. We would also like to acknowledge Luana Müller (Pravaler - Brazil) for performing an early qualitative evaluation of our heuristics set. Finally, we would like to thank the developers of the analyzed tools for contributing to the community by making open-source code available.

FUNDING SOURCES

This study was financed in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Brazil (Financial Code 001), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Instituto Nacional de Ciência e Tecnologia em NeuroImunoModulação (INCT-NIM).

© 2022. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <https://creativecommons.org/licenses/by-nc-nd/4.0/>. This is the accepted manuscript version. Final and published version can be found at: <https://doi.org/10.1016/j.combiomed.2022.105785>

AUTHOR CONTRIBUTIONS

Guilherme Bauer-Negrini: Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft. Guilherme Cordenonsi da Fonseca: Conceptualization, Writing - Review & Editing, Supervision. Carmem Gottfried: Conceptualization, Writing - Review & Editing, Supervision. Juliana Herbert: Conceptualization, Methodology, Writing - Review & Editing, Supervision.

All authors have approved the final version of the manuscript.

DATA AVAILABILITY

Additional data is available in the Supplementary Materials. Files to reproduce an evaluation environment are available at: www.github.com/gbnegrini/circrna-usability.

COMPETING INTERESTS

The authors declare that they have no competing interests.

REFERENCES

- [1] International Organization For Standardization, Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) – System and software quality models, ISO/IEC 25010:2011. (2011).
- [2] D. Bolchini, A. Finkelstein, V. Perrone, S. Nagl, Better bioinformatics through usability analysis, *Bioinformatics*. 25 (2009) 406–412. <https://doi.org/10.1093/bioinformatics/btn633>.
- [3] B. Lawlor, P. Walsh, Engineering bioinformatics: Building reliability, performance and productivity into bioinformatics software, *Bioengineered*. (2015). <https://doi.org/10.1080/21655979.2015.1050162>.

- [4] M. List, P. Ebert, F. Albrecht, Ten Simple Rules for Developing Usable Software in Computational Biology, *PLOS Comput. Biol.* 13 (2017) e1005265. <https://doi.org/10.1371/journal.pcbi.1005265>.
- [5] S. Mangul, L.S. Martin, E. Eskin, R. Blekhman, Improving the usability and archival stability of bioinformatics software, *Genome Biol.* 20 (2019) 47. <https://doi.org/10.1186/s13059-019-1649-8>.
- [6] M. Xenos, Usability perspective in software quality, in: *Usability Eng. Work. 8th Panhellenic Conf. Informatics with Int. Particip.* South. Cyprus, 2001.
- [7] X. Faulkner, F. Culwin, Enter the Usability Engineer: Integrating HCI and Software Engineering, in: *Proc. 5th Annu. SIGCSE/SIGCUE ITiCSEconference Innov. Technol. Comput. Sci. Educ.*, Association for Computing Machinery, New York, NY, USA, 2000: pp. 61–64. <https://doi.org/10.1145/343048.343076>.
- [8] H. Roder, Specifying usability features with patterns and templates, in: *2012 1st Int. Work. Usability Access. Focus. Requir. Eng. UsARE 2012 - Proc.*, 2012. <https://doi.org/10.1109/UsARE.2012.6226790>.
- [9] S. Lauesen, M.P. Musgrove, Heuristic Evaluation of User Interfaces versus Usability Testing, in: *User Interface Des. a Softw. Eng. Perspect.*, 2005.
- [10] J. Nielsen, Finding usability problems through heuristic evaluation, in: *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. - CHI '92*, ACM Press, New York, New York, USA, 1992: pp. 373–380. <https://doi.org/10.1145/142750.142834>.
- [11] A.H. Kamali, E. Giannoulatou, T.Y. Chen, M.A. Charleston, A.L. McEwan, J.W.K. Ho, How to test bioinformatics software?, *Biophys. Rev.* (2015). <https://doi.org/10.1007/s12551-015-0177-3>.
- [12] E. Giannoulatou, S.H. Park, D.T. Humphreys, J.W.K. Ho, Verification and validation of bioinformatics software without a gold standard: A case study of BWA and Bowtie, *BMC Bioinformatics.* (2014). <https://doi.org/10.1186/1471-2105-15-S16-S15>.

- [13] J. O’Rawe, T. Jiang, G. Sun, Y. Wu, W. Wang, J. Hu, P. Bodily, L. Tian, H. Hakonarson, W.E. Johnson, Z. Wei, K. Wang, G.J. Lyon, Low concordance of multiple variant-calling pipelines: Practical implications for exome and genome sequencing, *Genome Med.* 5 (2013) 28. <https://doi.org/10.1186/gm432>.
- [14] J.P.A. Ioannidis, D.B. Allison, C.A. Ball, I. Coulibaly, X. Cui, A.C. Culhane, M. Falchi, C. Furlanello, L. Game, G. Jurman, J. Mangion, T. Mehta, M. Nitzberg, G.P. Page, E. Petretto, V. Van Noort, Repeatability of published microarray gene expression analyses, *Nat. Genet.* (2009). <https://doi.org/10.1038/ng.295>.
- [15] D.A. Bader, Y. Li, T. Li, V. Sachdeva, BioPerf: A benchmark suite to evaluate high-performance computer architecture on bioinformatics applications, in: *Proc. 2005 IEEE Int. Symp. Workload Charact. IISWC-2005*, 2005. <https://doi.org/10.1109/IISWC.2005.1526013>.
- [16] M.R. Aniba, O. Poch, J.D. Thompson, Issues in bioinformatics benchmarking: The case study of multiple sequence alignment, *Nucleic Acids Res.* 38 (2010) 7353–7363. <https://doi.org/10.1093/nar/gkq625>.
- [17] C. Mannapperuma, N. Street, J. Waterworth, Designing Usable Bioinformatics Tools for Specialized Users, in: *Adv. Intell. Syst. Comput.*, Springer Verlag, 2019: pp. 649–670. https://doi.org/10.1007/978-3-030-11890-7_62.
- [18] J. Nielsen, Enhancing the explanatory power of usability heuristics, in: *Conf. Hum. Factors Comput. Syst. - Proc.*, 1994. <https://doi.org/10.1145/259963.260333>.
- [19] T. Seemann, Ten recommendations for creating usable bioinformatics command line software, *BioMed Central Ltd.*, 2013. <https://doi.org/10.1186/2047-217X-2-15>.
- [20] P. Georgeson, A. Syme, C. Sloggett, J. Chung, H. Dashnow, M. Milton, A. Lonsdale, D. Powell, T. Seemann, B. Pope, Bionitio: demonstrating and

facilitating best practices for bioinformatics command-line software, *Gigascience*. (2019). <https://doi.org/10.1093/gigascience/giz109>.

- [21] R. Awasthi, A.K. Singh, G. Mishra, A. Maurya, D.K. Chellappan, G. Gupta, P.M. Hansbro, K. Dua, An overview of circular RNAs, in: *Adv. Exp. Med. Biol.*, Springer New York LLC, 2018: pp. 3–14. https://doi.org/10.1007/978-981-13-1426-1_1.
- [22] C.J. Zhuo, W.H. Hou, D.G. Jiang, H.J. Tian, L.N. Wang, F. Jia, C.H. Zhou, J.J. Zhu, Circular RNAs in early brain development and their influence and clinical significance in neuropsychiatric disorders, *Neural Regen. Res.* 15 (2020) 817–823. <https://doi.org/10.4103/1673-5374.268969>.
- [23] D. Rong, H. Sun, Z. Li, S. Liu, C. Dong, K. Fu, W. Tang, H. Cao, An emerging function of circRNA-miRNAs-mRNA axis in human diseases, *Oncotarget*. 8 (2017). <https://doi.org/10.18632/oncotarget.19154>.
- [24] S. Memczak, M. Jens, A. Elefsinioti, F. Torti, J. Krueger, A. Rybak, L. Maier, S.D. Mackowiak, L.H. Gregersen, M. Munschauer, A. Loewer, U. Ziebold, M. Landthaler, C. Kocks, F. Le Noble, N. Rajewsky, Circular RNAs are a large class of animal RNAs with regulatory potency, *Nature*. 495 (2013) 333–338. <https://doi.org/10.1038/nature11928>.
- [25] L. Chen, C. Wang, H. Sun, J. Wang, Y. Liang, Y. Wang, G. Wong, The bioinformatics toolbox for circRNA discovery and analysis, *Brief. Bioinform.* (2020). <https://doi.org/10.1093/bib/bbaa001>.
- [26] T.B. Hansen, M.T. Venø, C.K. Damgaard, J. Kjems, Comparison of circular RNA prediction tools, *Nucleic Acids Res.* 44 (2016) e58–e58. <https://doi.org/10.1093/nar/gkv1458>.
- [27] T.B. Hansen, Improved circRNA identification by combining prediction algorithms, *Front. Cell Dev. Biol.* (2018). <https://doi.org/10.3389/fcell.2018.00020>.
- [28] X. Zeng, W. Lin, M. Guo, Q. Zou, A comprehensive overview and evaluation of circular RNA detection tools, *PLoS Comput. Biol.* (2017).

<https://doi.org/10.1371/journal.pcbi.1005420>.

- [29] J. Zhang, S. Chen, J. Yang, F. Zhao, Accurate quantification of circular RNAs identifies extensive circular isoform switching events, *Nat. Commun.* 11 (2020) 90. <https://doi.org/10.1038/s41467-019-13840-9>.
- [30] L. Szabo, R. Morey, N.J. Palpant, P.L. Wang, N. Afari, C. Jiang, M.M. Parast, C.E. Murry, L.C. Laurent, J. Salzman, Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development, *Genome Biol.* 16 (2015). <https://doi.org/10.1186/s13059-015-0690-5>.
- [31] S. Hoffmann, C. Otto, G. Doose, A. Tanzer, D. Langenberger, S. Christ, M. Kunz, L.M. Holdt, D. Teupser, J. Hackermüller, P.F. Stadler, A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection, *Genome Biol.* (2014). <https://doi.org/10.1186/gb-2014-15-2-r34>.
- [32] F.B. Menegidio, D.L. Jabes, R. Costa de Oliveira, L.R. Nunes, Dugong: a Docker image, based on Ubuntu Linux, focused on reproducibility and replicability for bioinformatics analyses, *Bioinformatics.* 34 (2018) 514–515. <https://doi.org/10.1093/bioinformatics/btx554>.
- [33] E. Gaffo, A. Bonizzato, G. te Kronnie, S. Bortoluzzi, CirComPara: A multi-method comparative bioinformatics pipeline to detect and study circRNAs from RNA-seq data, *Non-Coding RNA.* 3 (2017). <https://doi.org/10.3390/ncrna3010008>.
- [34] M. Karimzadeh, M.M. Hoffman, Top considerations for creating bioinformatics software documentation, *Brief. Bioinform.* 19 (2018) 693–699. <https://doi.org/10.1093/bib/bbw134>.
- [35] A.R. Casare, C.G. Da Silva, P.S. Martins, R.L.O. Moraes, Usability heuristics and accessibility guidelines: A comparison of heuristic evaluation and WCAG, in: *Proc. ACM Symp. Appl. Comput.*, 2016. <https://doi.org/10.1145/2851613.2851913>.
- [36] G. Yuan, Z. Jinyang, Z. Fangqing, Circular RNA identification based on

- multiple seed matching, *Brief. Bioinform.* 19 (2018) 803–810. <https://doi.org/10.1093/BIB/BBX014>.
- [37] Y. Gao, J. Wang, F. Zhao, CIRI: an efficient and unbiased algorithm for de novo circular RNA identification, *Genome Biol.* 16 (2015) 4. <https://doi.org/10.1186/s13059-014-0571-3>.
- [38] K. Wang, D. Singh, Z. Zeng, S.J. Coleman, Y. Huang, G.L. Savich, X. He, P. Mieczkowski, S.A. Grimm, C.M. Perou, J.N. MacLeod, D.Y. Chiang, J.F. Prins, J. Liu, MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery, *Nucleic Acids Res.* 38 (2010) e178. <https://doi.org/10.1093/nar/gkq622>.
- [39] P. Belmann, J. Dröge, A. Bremges, A.C. McHardy, A. Sczyrba, M.D. Barton, Bioboxes: Standardised containers for interchangeable bioinformatics software, *Gigascience.* (2015). <https://doi.org/10.1186/s13742-015-0087-0>.
- [40] B. Grüning, R. Dale, A. Sjödin, B.A. Chapman, J. Rowe, C.H. Tomkins-Tinch, R. Valieris, J. Köster, T.B. Team, Bioconda: sustainable and comprehensive software distribution for the life sciences, *Nat. Methods.* 15 (2018) 475–476. <https://doi.org/10.1038/s41592-018-0046-7>.
- [41] M. Nassif, A. Hernandez, A. Sridharan, M.P. Robillard, Generating Unit Tests for Documentation, *IEEE Trans. Softw. Eng.* (2021) 1. <https://doi.org/10.1109/TSE.2021.3087087>.
- [42] D. Vingen, E.J. Andrews, M. Ferati, Usability in Patient-Oriented Drug Interaction Checkers—A Scandinavian Sampling and Heuristic Evaluation, *Informatics.* 7 (2020). <https://doi.org/10.3390/informatics7040042>.
- [43] J. Nielsen, R. Molich, Heuristic Evaluation of User Interfaces, in: *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Association for Computing Machinery, New York, NY, USA, 1990: pp. 249–256. <https://doi.org/10.1145/97243.97281>.
- [44] R. Jeffries, J.R. Miller, C. Wharton, K.M. Uyeda, User interface evaluation

- in the real world: A comparison of four techniques, in: Conf. Hum. Factors Comput. Syst. - Proc., 1991. <https://doi.org/10.1145/108844.108862>.
- [45] M. Maguire, P. Isherwood, A comparison of user testing and heuristic evaluation methods for identifying website usability problems, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2018. https://doi.org/10.1007/978-3-319-91797-9_31.
- [46] E. Wang, B. Caldwell, An Empirical Study of Usability Testing: Heuristic Evaluation Vs. User Testing, Proc. Hum. Factors Ergon. Soc. Annu. Meet. 46 (2002) 774–778. <https://doi.org/10.1177/154193120204600802>.
- [47] M. Vromman, J. Vandesompele, P.-J. Volders, Closing the circle: current state and perspectives of circular RNA databases, Brief. Bioinform. (2020). <https://doi.org/10.1093/bib/bbz175>.

PARTE III

4. DISCUSSÃO

O Transtorno do Espectro Autista é uma condição multifatorial na qual um dos principais eixos de alteração é a expressão gênica no sistema nervoso central. Neste trabalho, utilizamos ferramentas de bioinformática e dados de sequenciamento de RNA para investigar o possível envolvimento de circRNAs no transtorno e fazer considerações metodológicas sobre os softwares de identificação utilizados.

A avaliação proposta da expressão de circRNAs só foi possível devido à disponibilidade de dados públicos de RNA-Seq, em especial de modelos animais de TEA. Iniciamos no Capítulo I com uma visão abrangente, na qual utilizamos dados de seis diferentes modelos animais de camundongos para investigar o perfil de expressão de circRNAs no encéfalo. Tal análise do panorama geral foi limitada a modelos animais cujas bibliotecas de RNA-Seq passaram pelo processo de remoção de RNA ribossomal, o qual favorece a detecção de circRNAs na amostra (SZABO; SALZMAN, 2016).

Nós identificamos mais de 130.000 circRNAs únicos em camundongos. Características genômicas como distribuição cromossômica, maior origem exônica e abundância de isoformas concordam com observações gerais para estas moléculas (GOKOOL; ANWAR; VOINEAGU, 2020; JECK *et al.*, 2013a; XU *et al.*, 2017). Quase a metade estava registrada em bancos de dados dedicados, sendo a maioria encontrada no circAtlas (WU; JI; ZHAO, 2020), o banco de dados mais atualizado no momento. O restante dos circRNAs não anotados podem ser consequências da diversidade de isoformas, uma vez que adotamos um critério relativamente rigoroso para consideração de sobreposição entre coordenadas (GRUHL *et al.*, 2021).

Os conjuntos de dados utilizados possuíam variações técnicas significativas, incluindo profundidade de sequenciamento, tamanho de *reads* e layout de biblioteca. *Reads* alinhados à junção de *backsplicing* são tipicamente apenas 0,1% do total de fragmentos de um experimento de sequenciamento (GRUNER *et al.*, 2016), fazendo com que a capacidade de identificação seja proporcional à profundidade do sequenciamento. O tamanho dos *reads* também impacta na detecção, sendo demonstrada uma relação direta entre maior sensibilidade e maior tamanho (NGUYEN; NGUYEN; VU, 2022). Além disso, o layout *paired-end* também parece ser capaz de aumentar a sensibilidade e a precisão da detecção (NGUYEN; NGUYEN; VU, 2022; SZABO; SALZMAN, 2016). A influência desses fatores fica evidente quando

comparamos os resultados no número de circRNAs detectados entre os conjuntos VPA (129.503 circRNAs, *reads* de 150 pb, layout *paired-ended*) e PTCHD1 (275 circRNAs, *reads* de 40 pb, layout *single-ended*). Infelizmente, o cenário ideal de análise não é sempre possível ao se utilizar dados públicos. Portanto, tomamos alguns cuidados metodológicos para uma análise robusta, como o uso de mais de uma ferramenta de detecção, correção de variáveis latentes e a análise estatística dos conjuntos de dados separadamente.

As amostras de conjuntos de dados de alta qualidade (e.g. layout *paired-end* e *reads* maiores ou iguais a 100 pb) exibiram uma concordância considerável em circRNAs compartilhados, o que é esperado para amostras da mesma espécie. Quando os dados são analisados por regiões específicas do encéfalo, o número de circRNAs em comum é de 11%. Similarmente, em humanos, a sobreposição de circRNAs identificados em quatro regiões cerebrais (córtex pré-frontal anterior, giro parahipocampal, giro frontal inferior, lóbulo temporal superior) foi de cerca de 19% (LO *et al.*, 2020). A comparação transcritômica entre áreas é de especial interesse no estudo do TEA, dado que o perfil de expressão é significativamente diferente entre regiões. Por exemplo, as alterações transcritômicas são mais pronunciadas no córtex do que no cerebelo de indivíduos com TEA (PARIKSHAK *et al.*, 2013, 2016). Porém, na presente tese, a análise adicional dessas diferenças é limitada pelo pequeno número de conjuntos de dados por região e pode ser influenciada principalmente pelos fatores técnicos do sequenciamento.

A análise da expressão gênica diferencial não revelou uma mudança onipresente e consistente, relativa à circRNAs individuais, entre os modelos. Entretanto, o quadro muda levemente quando os circRNAs diferencialmente expressos (CDE) são analisados no nível do gene parental. Nesse caso, alguns poucos genes parentais de CDE são compartilhados entre todos ou pelo menos a maioria dos modelos. Mais uma vez, as isoformas podem desempenhar um papel no perfil de expressão. Análises futuras poderão se concentrar em possíveis *locus* cuja regulação da transcrição dá origem a diferentes isoformas de circRNAs relacionadas (GRUHL *et al.*, 2021).

A maioria dos CDE de cada modelo parece ser exclusiva. Entretanto, alguns CDE são compartilhados, especialmente entre modelos induzidos geneticamente. Em especial, o modelo BTBR compartilha mais de 30 CDE com os modelos Ash1L e

Shank3. Já com o modelo VPA, compartilha em torno de 20. Tal resultado é curioso em termos absolutos, considerando que VPA e BTBR são dois dos modelos com mais circRNAs identificados e com amostras em melhor condição técnica. É possível que, além das diferenças técnicas e de regiões encefálicas, o maior compartilhamento de CDE entre modelos genéticos esteja associado a mecanismos subjacentes da etiologia, os quais são mais semelhantes entre si do que com um modelo ambiental.

Observamos que os perfis funcionais dos genes parentais dos CDE estão relacionados a temas biológicos conhecidos por serem alterados no TEA, apesar desta sobreposição limitada entre circRNAs individuais nos modelos. Em particular, os CDE infrarregulados estão associados à estrutura sináptica em todos os modelos, exceto no *Ptchd1*. A disfunção sináptica é generalizada no TEA (GUANG *et al.*, 2018), e este fenômeno é reproduzido por vários modelos animais, tais como os aqui analisados. Ratos do modelo animal VPA apresentam redução de marcadores neuronais GABAérgicos, aumentos de marcadores neuronais glutamatérgicos e de marcadores pós-sinápticos excitatórios no córtex pré-frontal (KIM *et al.*, 2013). Camundongos do modelo Shank3 apresentam disfunções sinápticas eletrofisiológicas, como diminuição da potenciação de longa duração e da amplitude de correntes excitatórias pós-sinápticas no hipocampo (BOZDAGI *et al.*, 2010; WANG *et al.*, 2011). Camundongos *Ash1L* apresentam menor inibição GABAérgica e maior excitação glutamatérgica no córtex pré-frontal (QIN *et al.*, 2021). A análise proteômica do córtex e hipocampo de camundongos BTBR mostra um nível elevado de proteínas sinápticas, como spinofilina e Sinapsina 1 (MEYZA; BLANCHARD, 2017). Camundongos *Kmt5b* apresentam redução de proteínas pós-sinápticas, como receptores de glutamato, no córtex pré-frontal (WANG, Z.-J. *et al.*, 2021).

Curiosamente, circRNAs são especialmente abundantes em sinapses (RYBAK-WOLF *et al.*, 2014). Além disso, a maioria dos circRNAs neurais são dinamicamente modulados durante a plasticidade neural e se originam de genes que codificam proteínas sinápticas (YOU *et al.*, 2015). Os CDE podem ser apenas mais uma consequência dessa disfunção sináptica generalizada, ou podem ser reguladores pós-transcricionais responsáveis por algumas dessas alterações. As possibilidades de regulação sináptica são complexas. Mecanismos de regulação semelhantes ao exercido pelo circRNA *ciRS-7* podem tanto aumentar a expressão de genes alvos, ao diminuir a disponibilidade de miRNAs livres, quanto diminuir a expressão, caso o

circRNA atue como um transportador de miRNA (PIWECKA *et al.*, 2017). Além disso, devemos considerar a possibilidade dos circRNAs aumentarem a expressão de seus genes parentais através da interação com proteínas ligantes de RNA. Alguns circRNAs exônicos-intrônicos podem se associar com o complexo de pequenas ribonucleoproteínas nucleares U1 (U1 snRNP) e subsequentemente interagir com a RNA Polimerase II nos promotores dos genes parentais (LI, Z. *et al.*, 2015). Outro possível mecanismo de regulação exercido por circRNAs é através da perturbação da transcrição canônica linear. Ao competir por sítios de *splicing* e usar a mesma maquinaria transcricional, circRNAs podem reduzir a expressão de suas contrapartes lineares (AUFIERO *et al.*, 2018; LIANG *et al.*, 2017). No momento, podemos apenas levantar a hipótese de que as mudanças observadas nos CDE de modelos animais reforçam a gama de alterações sinápticas presentes no TEA.

Outro termo de destaque na análise funcional é relacionado às histonas. As histonas são proteínas encontradas no núcleo de células eucarióticas e que interagem com o DNA para auxiliar na organização e compactação da cromatina (MARIÑO-RAMÍREZ *et al.*, 2005). A acessibilidade transcricional do genoma pode ser transientemente modificada através da metilação ou acetilação das histonas (CAMPOS; REINBERG, 2009). Curiosamente, apesar de serem modelos de tipos diferentes, *Kmt5b* e VPA apresentam genes parentais de CDE enriquecidos para genes relacionados à ligação de histonas. O gene *KMT5B* é um gene de risco para o TEA e dá origem a uma proteína metiltransferase de histonas (WANG, Z.-J. *et al.*, 2021). Já o ácido valproico é um conhecido inibidor de desacetilases de histonas (DE CAMPOS VIDAL; MELLO, 2020; GURVICH *et al.*, 2004). As semelhantes alterações transcricionais em nível de circRNAs podem ser um reflexo dos mecanismos subjacentes de alteração da cromatina que estão atuando em modelos com validades de construto tão distintas.

Uma análise mais detalhada da possível fonte celular das alterações observadas em camundongos revelou que os genes parentais afetados estão sobrerrepresentados em marcadores celulares de microglia e astrócitos. As disfunções neuronais estão entre as mais relatadas no TEA, mas também já foram observadas alterações gliais, tais como proporção e localização celular alteradas nas camadas corticais (GZIELO; NIKIFORUK, 2021). Dados *in vitro* mostram que astrócitos expostos ao VPA são capazes de reduzir a transmissão sináptica especificamente de

neurônio inibitórios (TAKEDA *et al.*, 2021). Microglia e astrócitos são especialmente interessantes no contexto sináptico, pois podem modular a formação e função sináptica (PETRELLI; PUCCI; BEZZI, 2016). Podemos formular a hipótese de que a disfunção dos circRNAs na TEA afeta não apenas os componentes sinápticos neuronais, mas também a maquinaria glial associada.

O valor translacional é um aspecto crítico da pesquisa de modelos animais. A este respeito, encontramos sobreposições de ambos os genes parentais com genes associados a TEA e diretamente com circRNAs alterados no cérebro de indivíduos autistas. Até onde sabemos, esta é a primeira comparação direta de potenciais CDE compartilhados entre modelos animais e humanos no contexto do TEA. Este compartilhamento, embora bastante limitado, mostra que modelos genéticos como de silenciamento dos genes de risco *Shank3* e *Ash1L* podem modelar outro aspecto molecular do TEA. É possível, na verdade, que mais CDE se sobreponham entre humanos e modelos animais. O critério que utilizamos para buscar correspondentes ortólogos foi o mais conservador possível. Gruhl *et al.* (2021) descreve três níveis de comparação de *loci* de circRNAs compartilhados entre espécies: gene parental, *locus* de circRNA no gene e compartilhamento de éxons de início e término. A estratégia utilizada por nós (compartilhamento de início e término), pode necessitar de um elevado grau de conservação evolucionária entre os sítios de *splicing*. Portanto, uma análise menos restrigente, como em nível de *locus* de circRNA, poderia encontrar o compartilhamento de outras isoformas.

A ausência de um conjunto de circRNAs consistentemente alterado em todos os modelos dificulta uma proposta de biomarcador universal. Entretanto, análises futuras podem ser focadas na expressão de um conjunto minimamente representativo comum a alguns modelos e seres humanos. A identificação de circRNAs candidatos no SNC de modelos animais pode fornecer pistas para futuros biomarcadores em humanos, dado que circRNAs são relativamente conservados e têm o potencial de cruzar a barreira hematoencefálica (LI, Y. *et al.*, 2015; YANG *et al.*, 2021).

Construímos uma rede de interação predita entre circRNA-miRNA-mRNA para explorar outro aspecto das vias compartilhadas possivelmente alteradas. A análise de ontologia genética demonstrou o enriquecimento para dois termos: receptores acoplados à proteína G, os quais controlam diversas funções celulares, e estão significativamente alterados em transtornos psiquiátricos, tais como TEA,

esquizofrenia, transtorno bipolar e transtorno depressivo maior (MONFARED *et al.*, 2021); e problemas de alimentação, os quais também parecem afetar os jovens com TEA (BARASKEWICH *et al.*, 2021). Portanto, a rede ilustra um potencial mecanismo regulatório cujos alvos estão associados ao TEA.

Deve-se ressaltar que a relação entre a expressão dos circRNAs e suas contrapartes lineares ainda não está elucidada (SEKAR; LIANG, 2019), porém sugere-se que a expressão da forma circular possa ser um mecanismo de diminuição da expressão gênica de RNAm lineares que incluiriam os éxons envolvidos (ASHWAL-FLUSS *et al.*, 2014; ZHANG *et al.*, 2014). Múltiplos estudos demonstram uma relação linear inversa entre a abundância de circRNAs e seus respectivos RNAm (ASHWAL-FLUSS *et al.*, 2014; JECK *et al.*, 2013; RYBAK-WOLF *et al.*, 2014), porém relações diretas (RYBAK-WOLF *et al.*, 2014) ou mesmo nenhuma relação já foram relatadas (XU, K. *et al.*, 2018). Comparações dessa natureza não foram realizadas no Capítulo I, no qual focamos em analisar apenas o panorama geral dos circRNAs e suas possíveis consequências *per se*. Porém, no Capítulo II, focamos em um eixo de regulação circRNA-miRNA-mRNA específico e mais bem elucidado, possibilitando algumas considerações sobre esse aspecto.

Após a análise do panorama geral de circRNAs em camundongos de modelos animais de TEA, iniciamos no Capítulo II uma análise mais aprofundada e focada em um modelo animal específico. O modelo animal de TEA induzido por exposição pré-natal ao VPA em ratos é utilizado há anos pelo nosso grupo de pesquisa para a investigação de mecanismos associados ao TEA. Demonstramos previamente diversas alterações comportamentais, moleculares e celulares induzidas pelo VPA em ratos, assim como a prevenção de algumas alterações pela administração intraútero do polifenol resveratrol (BAMBINI-JUNIOR *et al.*, 2011, 2014; DECKMANN *et al.*, 2021; FONTES-DUTRA *et al.*, 2018; HIRSCH *et al.*, 2018; SANTOS-TERRA *et al.*, 2022). Portanto, nosso interesse na análise da expressão de circRNAs nessa espécie e modelo específico estava alinhado com o objetivo geral do grupo de pesquisa e nos permitiria investigar possíveis vias biológicas com mais detalhes utilizando o tratamento com resveratrol. Porém, os resultados apresentados no Capítulo II são apenas adicionais e ainda carecem de uma maior investigação. As análises experimentais planejadas para esse capítulo foram, infelizmente, adiadas em decorrência na pandemia de SARS-CoV-2 em 2020.

A busca por conjuntos de dados de RNA-Seq publicamente disponíveis para o modelo VPA em ratos resultou em apenas um resultado. Além disso, as amostras sequenciadas não foram processadas com um protocolo de remoção de RNA ribossomal, o qual costuma resultar em uma detecção mais abundante de circRNAs (SZABO; SALZMAN, 2016). Dessa forma, decidimos investigar especificamente a expressão de um dos circRNAs mais abundantes no sistema nervoso central.

Apesar de ser também o circRNA melhor descrito em humanos (RYBAK-WOLF *et al.*, 2014), o ciRS-7/CDR1as não está anotado no genoma da espécie *Rattus norvegicus*. No genoma do rato, esta sequência tem extensas regiões de nucleotídeos desconhecidos (lacunas de sequenciamento), prejudicando a anotação e identificação deste circRNA, especialmente por ferramentas que dependem da anotação de coordenadas genômicas. Portanto, realizamos uma reconstrução *de novo* do transcrito ciRS-7 para a avaliação da sua expressão no modelo animal. A determinação parcial e inédita da sequência permitiu a quantificação do aumento da expressão gênica relativa desse importante circRNA no modelo animal de TEA induzido por exposição pré-natal ao VPA.

Apesar de estudos sobre a atuação de circRNAs como esponjas de miRNAs demonstrarem a suprarregulação de genes-alvos (por potencialmente diminuir a disponibilidade de miRNAs e sua consequente ação por RNAm-alvos) (HANSEN *et al.*, 2013b; MEMCZAK *et al.*, 2013b; XIE *et al.*, 2019), um mecanismo diferente já foi proposto para a ligação entre ciRS-7 e miR-7. Como demonstrado por Piwecka *et al.* (2017), o ciRS-7/CDR1as pode atuar na estabilização e transporte de um conjunto de miR-7, por consequência levando à infrarregulação de alvos. Considerando ambas as hipóteses, demonstramos que as expressões gênicas de diferentes alvos potenciais do miR-7 estão aumentadas ou diminuídas no córtex pré-frontal de animais do modelo animal de TEA induzido por VPA.

A análise funcional dos genes suprarregulados (hipótese: ↑ ciRS-7 ↓ miR-7 ↑ alvos) demonstrou enriquecimento de genes envolvidos no metabolismo de ácidos graxos. Alterações no metabolismo de ácidos graxos, e metabolismo de lipídeos no geral, já foram associadas ao TEA (USUI *et al.*, 2020) e também são encontradas em modelos animais do transtorno (CHEN *et al.*, 2016; KERR; GILMARTIN; ROCHE, 2016). Curiosamente, desse grupo de genes, o *Id3* também se encontra com a

expressão gênica elevada em indivíduos e animais do modelo de síndrome de Rett, doença síndrômica associada ao TEA (PEDDADA; YASUI; LASALLE, 2006).

Já a análise funcional dos genes infrarregulados (hipótese \uparrow ciRS-7 \uparrow miR-7 \downarrow alvos) revelou enriquecimento de termos relacionados principalmente à componentes ou funções sinápticas. Nossas análises demonstraram que *Shank2* é um potencial alvo do miR-7 e está infrarregulada no modelo VPA. Ainda no grupo de genes infrarregulados, destaca-se o gene *Dhcr24*, único alvo de miR-7 identificado como validado tanto em ratos quanto em humanos ao final do processo de filtragem de alvos. A 24-desidrocolesterol redutase (DHCR24) é uma oxirredutase que converte desmosterol em colesterol. Além dessa função na biossíntese de colesterol, a DHCR24 está envolvida na formação de jangadas lipídicas, regulação de estresse oxidativo, e possui atividades neuroprotetivas e anti-apoptóticas (LUU *et al.*, 2014). A caracterização de variantes genéticas de genes relacionados ao metabolismo de lipídeos e esteroides em indivíduos com TEA identificou correlações entre variantes do gene *DHCR24* (e outros relacionados) com a presença e severidade do transtorno (HALL *et al.*, 2015). Curiosamente, mutações do gene *DHCR24* causam a Síndrome de Smith-Lemli-Opitz, uma condição na qual os indivíduos são frequentemente diagnosticados com TEA (BUKELIS *et al.*, 2007; SIKORA *et al.*, 2006).

Em conjunto, os resultados dos Capítulos I e II fornecem uma descrição geral da expressão de circRNAs nos modelos animais e do possível envolvimento específico do circRNA ciRS-7 no modelo animal induzido por VPA em ratos. Curiosamente, o circRNA *Cdr1os*, correspondente em camundongos, não foi detectado diferencialmente expresso no modelo VPA em camundongos, analisado no Capítulo I. Apesar disso, o circRNA *Cdr1os* está suprarregulado nos modelos Ash1L, BTBR e Kmt5b. Análises adicionais serão necessárias para determinar a contribuição dessa suprarregulação. Porém, a presença dessa alteração em modelos distintos nos faz considerar que esse circRNA pode ser um interessante ponto de convergência no transtorno.

Como pode-se observar, os resultados apresentados na presente tese dependem exclusivamente de métodos de análise de bioinformática. A necessidade de uso de múltiplas ferramentas de identificação de circRNAs para as análises propiciou uma compreensão ampla sobre as particularidades de cada software. Em conjunto, o treinamento do autor em disciplinas de engenharia de software e interação

humano-computador tornou possível a investigação de conhecidos problemas de usabilidade anteriormente relatados na literatura (CHEN, L. *et al.*, 2020; HANSEN *et al.*, 2016; MENEGIDIO *et al.*, 2018; ZENG *et al.*, 2017).

No Capítulo III, nós focamos no desenvolvimento de um método para avaliar os problemas de usabilidade presentes no domínio específico das ferramentas de identificação das circRNAs. Seleccionamos três conjuntos de heurísticas e recomendações da literatura como ponto de partida. Os critérios para esta seleção foram que as heurísticas de Nielsen (1992) apresentam conceitos generalizáveis e as diretrizes de List, Ebert e Albrecht (2017) e de Seemann (2013) são princípios para o desenvolvimento de software de biologia computacional e software de linha de comando, estando assim intimamente relacionadas às ferramentas de identificação circRNA. As experiências adquiridas durante as análises descritas nos Capítulos I e II permitiram a criação de um conjunto de 15 heurísticas dedicadas à identificação e solução dos problemas de usabilidade observados nos softwares de identificação de circRNAs.

A avaliação heurística mostrou que a Instabilidade era a dimensão com o maior número de violações agregadas. Além disso, esta foi a dimensão mais afetada por problemas classificados como majoritários ou críticos. Não surpreendentemente, sabe-se que a instabilidade de ferramentas bioinformáticas é problemática. A instalação de dependências é uma fonte de grande frustração para os usuários de software de bioinformática (BELMANN *et al.*, 2015; MANGUL *et al.*, 2019). Aqui, demonstramos que as ferramentas de identificação circRNAs frequentemente falham em especificar corretamente as dependências necessárias. As consequências de tal problema podem variar. Na melhor das hipóteses, as dependências utilizadas são estáveis o suficiente para não causar erros críticos. Entretanto, em casos graves, o uso de dependências incorretas impede a instalação ou o uso da ferramenta, conforme observado com `find_circ` e `MapSplice2`.

Uma solução geral para estes problemas de instalação é o uso de gerenciadores de pacotes. Esta alternativa geralmente proporciona um processo simplificado que elimina a necessidade de o usuário obter cada dependência. Como mostrado previamente por Mangul *et al.* (2019), as ferramentas ômicas distribuídas sem gerenciadores de pacotes estão sujeitas a problemas de instalação em 32% dos casos. Particularmente, a distribuição do `MapSplice2` disponível no gerenciador de

pacotes Bioconda (GRÜNING *et al.*, 2018) foi a solução para prosseguir tanto com a avaliação heurística quanto com as análises realizadas no Capítulo II, após o procedimento de instalação padrão falhar. Uma alternativa são os contêineres de software, já utilizados por Menegidio *et al.* (2018) para reduzir a complexidade da instalação do *pipeline* CirComPara. De fato, o uso do CirComPara2 disponibilizado como contêiner foi a alternativa utilizada por nós no Capítulo I para executar diferentes ferramentas em análises de larga escala.

Também foram frequentes os problemas de documentação. Dificuldades deste tipo têm o potencial de ser especialmente frustrantes para os usuários iniciantes, pois a documentação oficial pode representar sua única fonte de informação. Um caso particularmente crítico foi observado com o `find_circ`, no qual alguns exemplos de comandos computacionais não se reproduziram como esperado. Segundo a avaliação empírica de Aghajani *et al.* (2019), exemplos de comandos são os mais frequentes problemas de exatidão de documentação. A redundância entre o código a ser testado e a documentação que deve ser atualizada cria o risco de inconsistência (NASSIF *et al.*, 2021). Portanto, os desenvolvedores devem ter um protocolo de atualização bem planejado. Como enfatizado por Karimzadeh e Hoffman (2018), a documentação completa não beneficia apenas o usuário. Adicionalmente, ela beneficia os revisores que devem avaliar a reprodutibilidade do software durante o processo de revisão por pares. Finalmente, ela também ajuda os desenvolvedores, economizando seu tempo fornecendo suporte aos usuários.

A baixa observação de alguns problemas destaca a maturidade das ferramentas em muitos aspectos. A maioria dos repositórios tinha uma licença, a qual é fundamental para estabelecer como os usuários podem usar e modificar o software disponível. Embora este não seja um critério de nossa análise, vemos como um fato positivo que todas as licenças até agora permitem, com pequenas restrições, o uso e modificação do software para fins de pesquisa. Outra característica positiva foi a conformidade com formatos de entrada e saída padronizados. A combinação de diferentes ferramentas provou ser benéfica para aumentar a precisão da identificação dos circRNAs (HANSEN, 2018). Portanto, a padronização ajuda os usuários a usar repetidamente os mesmos arquivos de entrada em diferentes ferramentas.

Nossas análises de usabilidade possuem algumas limitações. Primeiro, não há acordo sobre o número mínimo de avaliadores necessários para uma avaliação

heurística. Alguns autores consideram que cinco são suficientes para detectar a maioria dos problemas (NIELSEN; MOLICH, 1990b; VINGEN; ANDREWS; FERATI, 2020). Entretanto, Nielsen (1992) considera que este número pode ser reduzido para três no caso de avaliadores com dupla experiência, como é o caso da avaliação apresentada no Capítulo III. Em segundo lugar, a avaliação heurística não é o único método para a avaliação da usabilidade. Os testes com usuários frequentemente encontram problemas mais graves, enquanto a avaliação heurística encontra quantitativamente mais problemas (JEFFRIES *et al.*, 1991; MAGUIRE; ISHERWOOD, 2018; WANG; CALDWELL, 2002). Idealmente, ambos os métodos deveriam ser adotados como práticas complementares. Entretanto, considerando as características particulares do ambiente de desenvolvimento de software na academia, a avaliação heurística parece ser a alternativa mais econômica. No meio acadêmico, o software científico é desenvolvido por pequenos grupos de estudantes, muitas vezes sem treinamento formal em engenharia de software ou interação humano-computador, e com poucos recursos para desenvolver e manter produtos de software (LIST; EBERT; ALBRECHT, 2017; MANGUL *et al.*, 2019). Portanto, um conjunto mínimo de heurísticas tem o potencial de orientar os próprios desenvolvedores em uma avaliação de usabilidade de baixo custo que pode, e deve, ser realizada várias vezes durante as etapas do processo iterativo de desenvolvimento de software. Em terceiro lugar, nosso objetivo era identificar questões representativas e discutir soluções para a área de estudo, e não criticar nenhuma ferramenta individual. Acreditamos que a usabilidade é um fator determinante para o uso de ferramentas de identificação circRNA, mas não deve ser o único fator considerado pelos usuários. Aspectos como precisão e desempenho, que já foram avaliados em outros estudos, também devem ser considerados pelos usuários de acordo com suas necessidades.

5. CONSIDERAÇÕES FINAIS E CONCLUSÕES

A presente tese avaliou o perfil de expressão de circRNAs no sistema nervoso central de diversos modelos animais de TEA. Os resultados obtidos se fundamentam em uma extensa gama de análises de bioinformática com foco em processos de descoberta de conhecimento. Além disso, o uso exaustivo de ferramentas de identificação de circRNAs permitiu com que o presente autor contribuísse de forma interdisciplinar para o melhoramento dos softwares utilizados. A Figura 4 resume brevemente os principais resultados e as associações entre os capítulos da tese.

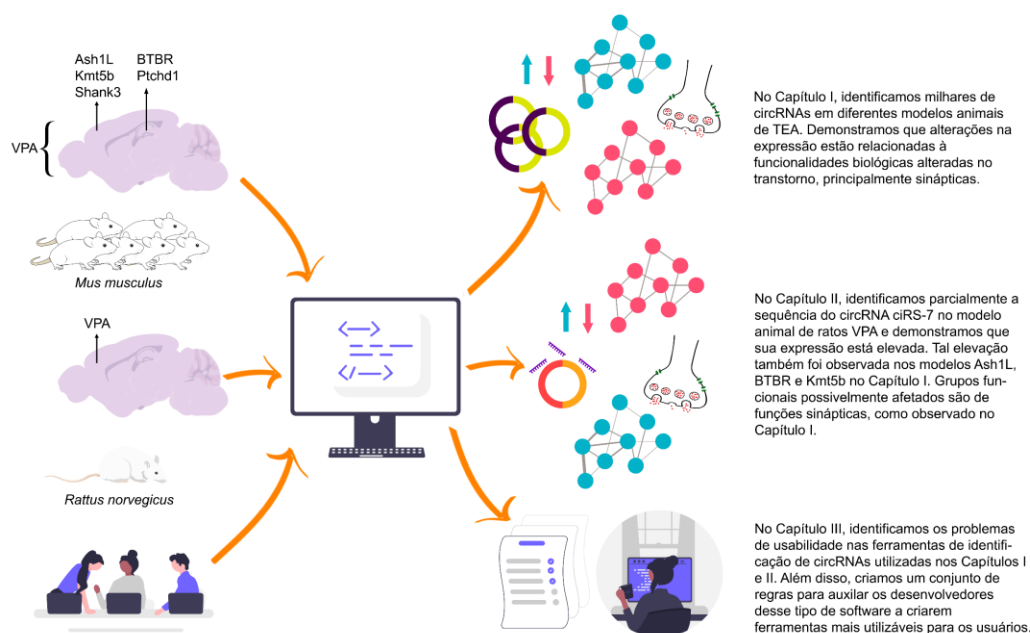


Figura 4 - Visão geral dos principais resultados obtidos ao longo do trabalho. Na esquerda, principais componentes metodológicos e objetos de estudo. Em posição central está a representação das análises de bioinformáticas que sustentaram os resultados obtidos, os quais estão resumidos na direita. Fonte: autoria própria (2022).

No Capítulo I, é possível observar o panorama geral das alterações transcritômicas de circRNAs em modelos animais do TEA. Demonstrou-se que circRNAs alterados estão relacionados a vias biológicas em comum entre diversos modelos animais. Tal observação reforça que o TEA se apresenta de forma heterogênea e multifatorial, mas compartilha alterações fundamentais e difundidas. Além disso, a disponibilização das coordenadas dos circRNAs identificados pode ser

um importante instrumento para outros pesquisadores da área realizarem investigações mais aprofundadas em determinado modelo animal.

No Capítulo II, a análise é focada especificamente no modelo animal induzido por VPA em ratos. Apesar de ser apenas um compilado de resultados adicionais, o capítulo fornece uma possível rota de interesse experimental que envolve uma rede circRNA-miRNA-mRNA. As análises de bioinformática realizadas serão fundamentais para a validação experimental das hipóteses levantadas para a regulação de alvos via ação de esponja de miRNAs. Adicionalmente, a determinação parcial da sequência de ciRS-7 na espécie *Rattus norvegicus* pode auxiliar outros pesquisadores na caracterização desse circRNA em outros contextos na espécie.

Por último, o Capítulo III finaliza as análises com a descrição formal e detalhada dos principais problemas de usabilidade encontrados durante os experimentos realizados nos Capítulos I e II. A área da bioinformática é conhecida por problemas dessa natureza e demonstramos exatamente quais são os pontos mais prejudicados no caso das ferramentas de identificação de circRNAs. Apenas o registro dos problemas não teria efeitos práticos. Sendo assim, desenvolvemos um conjunto de regras capazes de auxiliar os desenvolvedores a identificarem e solucionarem os problemas durante o processo de desenvolvimento de software.

Em conjunto, os resultados apresentados são apenas o início de uma complexa investigação de mecanismos regulatórios não usuais. Podemos observar que os circRNAs são de fato outra classe de ncRNAs associada ao TEA. Porém, assim como na maior parte das observações no transtorno, resta elucidar o que são causas e o que são consequências da desregulação transcritômica. A utilização de modelos animais para o estudo desses mecanismos será fundamental e demonstramos uma potencial validade de face no quesito disfunção da expressão de circRNAs. Por fim, apresentamos um método de avaliação de usabilidade de software que pode permitir que, idealmente, futuras análises estejam alinhadas com as necessidades e expectativas de uso de nós e outros pesquisadores.

6. PERSPECTIVAS

Assim como em diversos aspectos da vida cotidiana, a pandemia de SARS-CoV-2 nos últimos anos teve profundas consequências para a presente tese. Em resumo, validações experimentais previstas para os Capítulos I e II precisaram ser adiadas, ou descartadas, para uma adequação dos objetivos da tese. Além disso, testes de usabilidade contemplando usuários, previstos para o Capítulo III, não puderam ser realizadas devido às restrições de encontros presenciais. Dessa forma, as principais perspectivas conjecturadas estão relacionadas à continuidade de experimentos, os quais podem elucidar pontos ainda não esclarecidos dos resultados. A saber, pretende-se:

- Validar experimentalmente a expressão do ciRS-7 em *Rattus norvegicus* pela técnica de RT-qPCR com uso de *primers* divergentes e tratamento com RNase R.
- Determinar a sequência completa do ciRS-7 em *Rattus norvegicus* pela técnica de sequenciamento de Sanger.
- Avaliar por RT-qPCR a expressão gênica da rede de ncRNAs envolvendo ciRS-7, miR-7, miR-671, Cyrano e genes-alvos em *Rattus norvegicus*.
- Avaliar, no modelo animal induzido por VPA em ratos, a expressão de circRNAs de interesse identificados na análise de camundongos, bem como os possíveis efeitos do resveratrol na expressão.
- Submeter correções e melhorias de código nos repositórios públicos das ferramentas de identificação de circRNA de acordo com os resultados encontrados pela avaliação de usabilidade.

REFERÊNCIAS BIBLIOGRÁFICAS

AMAL, Haitham *et al.* Shank3 mutation in a mouse model of autism leads to changes in the S-nitroso-proteome and affects key proteins involved in vesicle release and synaptic function. **Molecular Psychiatry**, [s. l.], v. 25, n. 8, p. 1835–1848, 2020.

AMERICAN PSYCHIATRIC ASSOCIATION. **Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5)**. [S. l.: s. n.], 2013.

ANIBA, Mohamed Radhouene; POCH, Olivier; THOMPSON, Julie D. Issues in bioinformatics benchmarking: The case study of multiple sequence alignment. **Nucleic Acids Research**, [s. l.], v. 38, n. 21, p. 7353–7363, 2010.

AR, Jensen *et al.* Modern Biomarkers for Autism Spectrum Disorder: Future Directions. **Molecular diagnosis & therapy**, [s. l.], v. 26, n. 5, 2022. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/35759118/>. Acesso em: 15 nov. 2022.

ASHWAL-FLUSS, Reut *et al.* CircRNA Biogenesis competes with Pre-mRNA splicing. **Molecular Cell**, [s. l.], v. 56, n. 1, p. 55–66, 2014.

AUFIERO, Simona *et al.* Cardiac circRNAs arise mainly from constitutive exons rather than alternatively spliced exons. **RNA (New York, N.Y.)**, [s. l.], v. 24, n. 6, p. 815–827, 2018.

BADER, David A. *et al.* BioPerf: A benchmark suite to evaluate high-performance computer architecture on bioinformatics applications. *Em:* , 2005. **Proceedings of the 2005 IEEE International Symposium on Workload Characterization, IISWC-2005**. [S. l.: s. n.], 2005.

BAMBINI-JUNIOR, Victorio *et al.* Animal model of autism induced by prenatal exposure to valproate: Behavioral changes and liver parameters. **Brain Research**, [s. l.], v. 1408, p. 8–16, 2011.

BAMBINI-JUNIOR, Victorio *et al.* Resveratrol prevents social deficits in animal model of autism induced by valproic acid. **Neuroscience letters**, [s. l.], v. 583, p. 176–181, 2014.

BANERJEE, Swati; BHAT, Manzoor; RIORDAN, Maeveen. Genetic aspects of autism spectrum disorders: insights from animal models. **Frontiers in Cellular Neuroscience**, [s. l.], v. 8, 2014. Disponível em: <https://www.frontiersin.org/articles/10.3389/fncel.2014.00058>. Acesso em: 15 nov. 2022.

BARASKEWICH, Jessica *et al.* Feeding and eating problems in children and adolescents with autism: A scoping review. **Autism**, [s. l.], v. 25, n. 6, p. 1505, 2021.

BARON-COHEN, Simon *et al.* Why are autism spectrum conditions more prevalent in males?. **PLoS biology**, [s. l.], v. 9, n. 6, p. e1001081, 2011.

BARONIO, Diego *et al.* Effects of an H3R antagonist on the animal model of autism induced by prenatal exposure to valproic acid. **PloS one**, [s. l.], v. 10, n. 1, p. e0116363, 2015.

BELMANN, Peter *et al.* **Bioboxes: Standardised containers for interchangeable bioinformatics software.** [S. l.: s. n.], 2015.

BELZUNG, Catherine *et al.* Rodent models for autism: A critical review. **Drug Discovery Today: Disease Models**, [s. l.], v. 2, n. 2, p. 93–101, 2005.

BOEHM, B. W.; BROWN, J. R.; LIPOW, M. Quantitative evaluation of software quality. *Em:* , 1976. **Proceedings - International Conference on Software Engineering.** [S. l.: s. n.], 1976.

BOLCHINI, Davide *et al.* Better bioinformatics through usability analysis. **Bioinformatics**, [s. l.], v. 25, n. 3, p. 406–412, 2009.

BOZDAGI, Ozlem *et al.* Haploinsufficiency of the autism-associated Shank3 gene leads to deficits in synaptic function, social interaction, and social communication. **Molecular Autism**, [s. l.], v. 1, n. 1, p. 15, 2010.

BRASIL. **LEI Nº 13.438, DE 26 DE ABRIL DE 2017.** 26 abr. 2017. Disponível em: <https://www2.camara.leg.br/legin/fed/lei/2017/lei-13438-26-abril-2017-784640-publicacaooriginal-152405-pl.html>.

BROWN, Alan S.; MEYER, Urs. Maternal Immune Activation and Neuropsychiatric Illness: A Translational Research Perspective. **The American Journal of Psychiatry**, [s. l.], v. 175, n. 11, p. 1073–1083, 2018.

BUESCHER, Ariane V. S. *et al.* Costs of autism spectrum disorders in the United Kingdom and the United States. **JAMA pediatrics**, [s. l.], v. 168, n. 8, p. 721–728, 2014.

BUKELIS, Irena *et al.* Smith-Lemli-Opitz syndrome and autism spectrum disorder. *Em:* , 2007. **American Journal of Psychiatry.** [S. l.]: Am J Psychiatry, 2007. p. 1655–1661.

CAMPOS, Eric I.; REINBERG, Danny. Histones: Annotating Chromatin. **Annual Review of Genetics**, [s. l.], v. 43, n. 1, p. 559–599, 2009.

CHADMAN, Kathryn K.; YANG, Mu; CRAWLEY, Jacqueline N. Criteria for Validating Mouse Models of Psychiatric Diseases. **American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics**, [s. l.], v. 150B, n. 1, p. 1–11, 2009.

CHASTE, Pauline; LEBOYER, Marion. Autism risk factors: genes, environment, and gene-environment interactions. **Dialogues in Clinical Neuroscience**, [s. l.], v. 14, n. 3, p. 281–292, 2012.

CHEN, Yen-Ju *et al.* Genome-wide, integrative analysis of circular RNA dysregulation and the corresponding circular RNA-microRNA-mRNA regulatory axes in autism. **Genome Research**, [s. l.], v. 30, n. 3, p. 375–391, 2020.

CHEN, Jianling *et al.* Increased expression of fatty acid synthase and acetyl-CoA carboxylase in the prefrontal cortex and cerebellum in the valproic acid model of autism. **Experimental and Therapeutic Medicine**, [s. l.], v. 12, n. 3, p. 1293–1298, 2016.

CHEN, Lei *et al.* Oxidative stress marker aberrations in children with autism spectrum disorder: a systematic review and meta-analysis of 87 studies (N=9109). **Translational Psychiatry**, [s. l.], v. 11, n. 1, p. 15, 2021.

CHEN, Liang *et al.* The bioinformatics toolbox for circRNA discovery and analysis. **Briefings in Bioinformatics**, [s. l.], 2020.

CHEN, Bei Jun; HUANG, Suleigh; JANITZ, Michael. Changes in circular RNA expression patterns during human foetal brain development. **Genomics**, [s. l.], v. 111, n. 4, p. 753–758, 2019.

CHRISTENSEN, Jakob *et al.* Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. **JAMA**, [s. l.], v. 309, n. 16, p. 1696–1703, 2013.

CHRISTIANSON, A. L.; CHESLER, N.; KROMBERG, J. G. Fetal valproate syndrome: clinical and neuro-developmental features in two sibling pairs. **Developmental Medicine and Child Neurology**, [s. l.], v. 36, n. 4, p. 361–369, 1994.

CLARK, Megan Louise Erin *et al.* School Age Outcomes of Children Diagnosed Early and Later with Autism Spectrum Disorder. **Journal of Autism and Developmental Disorders**, [s. l.], v. 48, n. 1, p. 92–102, 2018.

DAWSON, Geraldine; BURNER, Karen. Behavioral interventions in children and adolescents with autism spectrum disorder: a review of recent findings. **Current Opinion in Pediatrics**, [s. l.], v. 23, n. 6, p. 616–620, 2011.

DE CAMPOS VIDAL, Benedicto; MELLO, Maria Luiza S. Sodium valproate (VPA) interactions with DNA and histones. **International Journal of Biological Macromolecules**, [s. l.], v. 163, p. 219–231, 2020.

DE RUBEIS, Silvia *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. **Nature**, [s. l.], v. 515, n. 7526, p. 209–215, 2014.

DECKMANN, Iohanna *et al.* Resveratrol prevents brain edema, blood–brain barrier permeability, and altered aquaporin profile in autism animal model. **International Journal of Developmental Neuroscience**, [s. l.], v. 81, n. 7, p. 579–604, 2021.

ERGAZ, Zivanit; WEINSTEIN-FUDIM, Liza; ORNOY, Asher. Genetic and non-genetic animal models for autism spectrum disorders (ASD). **Reproductive Toxicology**, [s. l.], v. 64, 44th Annual Conference of the European Teratology Society, p. 116–140, 2016.

ESTES, Myka L.; MCALLISTER, A. Kimberley. Maternal immune activation: Implications for neuropsychiatric disorders. **Science (New York, N.Y.)**, [s. l.], v. 353, n. 6301, p. 772–777, 2016.

EZEONWUKA, Chinelo D.; RASTEGAR, Mojgan. MeCP2-Related Diseases and Animal Models. **Diseases (Basel, Switzerland)**, [s. l.], v. 2, n. 1, p. 45–70, 2014.

FAULKNER, Kristine; CULWIN, Fintan. Enter the Usability Engineer: Integrating HCI and Software Engineering. *Em:* , 2000, New York, NY, USA. **Proceedings of the 5th Annual SIGCSE/SIGCUE ITiCSEconference on Innovation and Technology in**

Computer Science Education. New York, NY, USA: Association for Computing Machinery, 2000. p. 61–64.

FOLSTEIN, S.; RUTTER, M. Infantile autism: a genetic study of 21 twin pairs. **Journal of Child Psychology and Psychiatry, and Allied Disciplines**, [s. l.], v. 18, n. 4, p. 297–321, 1977.

FONTES-DUTRA, Mellanie *et al.* Resveratrol prevents cellular and behavioral sensory alterations in the animal model of autism induced by valproic acid. **Frontiers in Synaptic Neuroscience**, [s. l.], v. 10, p. 9, 2018.

FRANCÉS, Lorena *et al.* Current state of knowledge on the prevalence of neurodevelopmental disorders in childhood according to the DSM-5: a systematic review in accordance with the PRISMA criteria. **Child and Adolescent Psychiatry and Mental Health**, [s. l.], v. 16, p. 27, 2022.

G. E. SSUCHAREWA, Dr. Die schizoiden Psychopathien im Kindesalter. (Part 1 of 2). **European Neurology**, [s. l.], v. 60, n. 3–4, p. 235–247, 1926a.

G. E. SSUCHAREWA, Dr. Die schizoiden Psychopathien im Kindesalter. (Part 2 of 2). **European Neurology**, [s. l.], v. 60, n. 3–4, p. 248–261, 1926b.

GAFFO, Enrico *et al.* CirComPara: A multi-method comparative bioinformatics pipeline to detect and study circRNAs from RNA-seq data. **Non-coding RNA**, [s. l.], v. 3, n. 1, 2017.

GAFFO, Enrico *et al.* Sensitive, reliable and robust circRNA detection from RNA-seq with CirComPara2. **Briefings in Bioinformatics**, [s. l.], v. 23, n. 1, p. bbab418, 2022.

GANDAL, Michael J. *et al.* Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. **Nature**, [s. l.], p. 1–8, 2022.

GANDAL, Michael J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. **Science (New York, N.Y.)**, [s. l.], v. 362, n. 6420, p. eaat8127, 2018.

GAO, Yuen *et al.* Loss of histone methyltransferase ASH1L in the developing mouse brain causes autistic-like behaviors. **Communications Biology**, [s. l.], v. 4, n. 1, p. 756, 2021.

GAO, Yuan; WANG, Jinfeng; ZHAO, Fangqing. CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. **Genome Biology**, [s. l.], v. 16, n. 1, p. 4, 2015.

GASPARINI, Silvia *et al.* Differential Expression of Hippocampal Circular RNAs in the BTBR Mouse Model for Autism Spectrum Disorder. **Molecular Neurobiology**, [s. l.], 2020.

GENTILE, Salvatore. Risks of neurobehavioral teratogenicity associated with prenatal exposure to valproate monotherapy: a systematic review with regulatory repercussions. **CNS spectrums**, [s. l.], v. 19, n. 4, p. 305–315, 2014.

GEORGESON, Peter *et al.* Bionitio: demonstrating and facilitating best practices for bioinformatics command-line software. **GigaScience**, [s. l.], 2019.

GIANNOULATOU, Eleni *et al.* Verification and validation of bioinformatics software without a gold standard: A case study of BWA and Bowtie. **BMC Bioinformatics**, [s. l.], 2014.

GILLBERG, C; WING, L. Autism: not an extremely rare disorder. **Acta psychiatrica Scandinavica**, [s. l.], v. 99, n. 6, p. 399–406, 1999.

GOKOOL, Akira; ANWAR, Firoz; VOINEAGU, Irina. The Landscape of Circular RNA Expression in the Human Brain. **Biological Psychiatry**, [s. l.], v. 87, n. 3, p. 294–304, 2020.

GOLDANI, Andre A S *et al.* Biomarkers in autism. **Frontiers in psychiatry**, [s. l.], v. 5, p. 100, 2014.

GOTTFRIED, Carmem *et al.* The Impact of Neuroimmune Alterations in Autism Spectrum Disorder. **Frontiers in psychiatry**, [s. l.], v. 6, p. 121, 2015.

GREEN, Andrew J.; JOHNSON, Philip H.; YATES, John R.W. The tuberous sclerosis gene on chromosome 9q34 acts as a growth suppressor. **Human Molecular Genetics**, [s. l.], v. 3, n. 10, p. 1833–1834, 1994.

GRUHL, Franziska *et al.* Circular RNA repertoires are associated with evolutionarily young transposable elements. **eLife**, [s. l.], v. 10, p. e67991, 2021.

GRUNER, Hannah *et al.* CircRNA accumulation in the aging mouse brain. **Scientific Reports**, [s. l.], v. 6, n. 1, p. 38907, 2016.

GRÜNING, Björn *et al.* Bioconda: sustainable and comprehensive software distribution for the life sciences. **Nature Methods**, [s. l.], v. 15, n. 7, p. 475–476, 2018.

GUANG, Shiqi *et al.* Synaptopathology involved in autism spectrum disorder. **Frontiers in Cellular Neuroscience**, [s. l.], v. 12, 2018. Disponível em: Acesso em: 20 mar. 2020.

GURVICH, Nadia *et al.* Histone deacetylase is a target of valproic acid-mediated cellular differentiation. **Cancer Research**, [s. l.], v. 64, n. 3, p. 1079–1086, 2004.

GUY, Jacky *et al.* Reversal of neurological defects in a mouse model of Rett syndrome. **Science (New York, N.Y.)**, [s. l.], v. 315, n. 5815, p. 1143–1147, 2007.

GZIELO, Kinga; NIKIFORUK, Agnieszka. Astroglia in Autism Spectrum Disorder. **International Journal of Molecular Sciences**, [s. l.], v. 22, n. 21, p. 11544, 2021.

HADDAD, Faraj L.; PATEL, Salonee V.; SCHMID, Susanne. Maternal Immune Activation by Poly I:C as a preclinical Model for Neurodevelopmental Disorders: A focus on Autism and Schizophrenia. **Neuroscience and Biobehavioral Reviews**, [s. l.], v. 113, p. 546–567, 2020.

HALL, Trevor A. *et al.* Lipid and sterol gene sequence variation in autism and correlates with neurodevelopmental status: A pilot study. **New Horizons in Translational Medicine**, [s. l.], v. 2, n. 6–7, p. 137–146, 2015.

HANSEN, Thomas B. *et al.* Comparison of circular RNA prediction tools. **Nucleic Acids Research**, [s. l.], 2015.

HANSEN, Thomas B. *et al.* Comparison of circular RNA prediction tools. **Nucleic Acids Research**, [s. l.], v. 44, n. 6, p. e58–e58, 2016.

HANSEN, Thomas B. Improved circRNA identification by combining prediction algorithms. **Frontiers in Cell and Developmental Biology**, [s. l.], 2018.

HANSEN, Thomas B. *et al.* Natural RNA circles function as efficient microRNA sponges. **Nature**, [s. l.], v. 495, n. 7441, p. 384–388, 2013a.

HANSEN, Thomas B. *et al.* Natural RNA circles function as efficient microRNA sponges. **Nature**, [s. l.], v. 495, n. 7441, p. 384–388, 2013b.

HANSEN, Stefan N.; SCHENDEL, Diana E.; PARNER, Erik T. Explaining the Increase in the Prevalence of Autism Spectrum Disorders: The Proportion Attributable to Changes in Reporting Practices. **JAMA Pediatrics**, [s. l.], v. 169, n. 1, p. 56–62, 2015.

HE, Yi *et al.* An integrated transcriptomic analysis of autism spectrum disorder. **Scientific Reports**, [s. l.], v. 9, n. 1, p. 11818, 2019.

HICKS, Steven D. *et al.* Salivary miRNA profiles identify children with autism spectrum disorder, correlate with adaptive behavior, and implicate ASD candidate genes involved in neurodevelopment. **BMC Pediatrics**, [s. l.], 2016.

HIRSCH, Mauro Mozael *et al.* Behavioral alterations in autism model induced by valproic acid and translational analysis of circulating microRNA. **Food and Chemical Toxicology**, [s. l.], v. 115, p. 336–343, 2018.

HOFFMANN, Steve *et al.* A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. **Genome Biology**, [s. l.], 2014.

HSU, Ming TA; COCA-PRADOS, Miguel. Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells [24]. **Nature**, [s. l.], v. 280, n. 5720, p. 339–340, 1979.

HUANG, S. *et al.* The emerging role of circular RNAs in transcriptome regulation. **Genomics**, [s. l.], v. 109, n. 5, p. 401–407, 2017.

HYMAN, Susan L. *et al.* Identification, Evaluation, and Management of Children With Autism Spectrum Disorder. **Pediatrics**, [s. l.], v. 145, n. 1, p. e20193447, 2020.

IEEE. IEEE Standard Glossary of Software Engineering Terminology (IEEE Std 610.12-1990). Los Alamitos. **CA: IEEE Computer Society**, [s. l.], 1990.

IOANNIDIS, John P.A. *et al.* Repeatability of published microarray gene expression analyses. **Nature Genetics**, [s. l.], 2009.

ISO/IEC. **ISO/IEC 25010 System and software quality models**. [S. l.]: ISO/IEC, 2010.

JECK, William R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. **RNA**, [s. l.], v. 19, n. 2, p. 141–157, 2013a.

JECK, William R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. **RNA**, [s. l.], v. 19, n. 2, p. 141–157, 2013b.

JECK, William R.; SHARPLESS, Norman E. **Detecting and characterizing circular RNAs**. [S. l.]: Nature Publishing Group, 2014.

JEFFRIES, Robin *et al.* User interface evaluation in the real world: A comparison of four techniques. *Em:* , 1991. **Conference on Human Factors in Computing Systems - Proceedings**. [S. l.: s. n.], 1991.

JUNN, Eunsung *et al.* Repression of α -synuclein expression and toxicity by microRNA-7. **Proceedings of the National Academy of Sciences of the United States of America**, [s. l.], v. 106, n. 31, p. 13052–13057, 2009.

KAISER-MCCAW, Barbara *et al.* Fragile X-linked mental retardation. **American Journal of Medical Genetics**, [s. l.], v. 7, n. 4, p. 503–505, 1980.

KAMALI, Amir Hossein *et al.* **How to test bioinformatics software?**. [S. l.: s. n.], 2015.

KANNER, L. Autistic disturbances of affective contact. **Acta paedopsychiatrica**, [s. l.], v. 35, p. 100–136, 1968.

KARIMZADEH, Mehran; HOFFMAN, Michael M. Top considerations for creating bioinformatics software documentation. **Briefings in bioinformatics**, [s. l.], v. 19, n. 4, p. 693–699, 2018.

KERR, Daniel M.; GILMARTIN, Aoife; ROCHE, Michelle. Pharmacological inhibition of fatty acid amide hydrolase attenuates social behavioural deficits in male rats prenatally exposed to valproic acid. **Pharmacological Research**, [s. l.], v. 113, p. 228–235, 2016.

KIM, Ki Chan *et al.* Male-specific alteration in excitatory post-synaptic development and social interaction in pre-natal valproic acid exposure model of autism spectrum disorder. **Journal of Neurochemistry**, [s. l.], v. 124, n. 6, p. 832–843, 2013.

KLIN, Ami. Biomarkers in Autism Spectrum Disorder: Challenges, Advances, and the Need for Biomarkers of Relevance to Public Health. **Focus: Journal of Life Long Learning in Psychiatry**, [s. l.], v. 16, n. 2, p. 135–142, 2018.

KNUESEL, Irene *et al.* Maternal immune activation and abnormal brain development across CNS disorders. **Nature Reviews. Neurology**, [s. l.], v. 10, n. 11, p. 643–660, 2014.

KOLOZSI, E. *et al.* Prenatal exposure to valproic acid leads to reduced expression of synaptic adhesion molecule neuroligin 3 in mice. **Neuroscience**, [s. l.], v. 163, n. 4, p. 1201–1210, 2009.

KUHN, Roland. Eugen Bleuler's concepts of psychopathology. **History of psychiatry**, [s. l.], v. 15, n. 59 Pt 3, p. 361–366, 2004.

KULCHESKI, Franceli Rodrigues; CHRISTOFF, Ana Paula; MARGIS, Rogerio. **Circular RNAs are miRNA sponges and can be used as a new class of biomarker**. [S. l.]: Elsevier B.V., 2016.

LAMMERT, Catherine R.; LUKENS, John R. Modeling autism-related disorders in mice with maternal immune activation (MIA). **Methods in molecular biology (Clifton, N.J.)**, [s. l.], v. 1960, p. 227–236, 2019.

LAUESEN, Soren; MUSGROVE, Mimi Pave. Heuristic Evaluation of User Interfaces versus Usability Testing. *Em: USER INTERFACE DESIGN: A SOFTWARE ENGINEERING PERSPECTIVE*. [S. l.: s. n.], 2005.

LI, Yan *et al.* Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. **Cell Research**, [s. l.], v. 25, n. 8, p. 981–984, 2015.

LI, Zhaoyong *et al.* Exon-intron circular RNAs regulate transcription in the nucleus. **Nature Structural and Molecular Biology**, [s. l.], v. 22, n. 3, p. 256–264, 2015.

LI, Dongyun; KARNATH, Hans-Otto; XU, Xiu. Candidate Biomarkers in Children with Autism Spectrum Disorder: A Review of MRI Studies. **Neuroscience Bulletin**, [s. l.], v. 33, n. 2, p. 219–237, 2017.

LIANG, Dongming *et al.* The Output of Protein-Coding Genes Shifts to Circular RNAs When the Pre-mRNA Processing Machinery Is Limiting. **Molecular Cell**, [s. l.], v. 68, n. 5, p. 940-954.e3, 2017.

LICHTENSTEIN, Paul *et al.* The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. **The American Journal of Psychiatry**, [s. l.], v. 167, n. 11, p. 1357–1363, 2010.

LIST, Markus; EBERT, Peter; ALBRECHT, Felipe. Ten Simple Rules for Developing Usable Software in Computational Biology. **PLOS Computational Biology**, [s. l.], v. 13, n. 1, p. e1005265, 2017.

LO, IJu *et al.* Linking the association between circRNAs and Alzheimer's disease progression by multi-tissue circular RNA characterization. **RNA Biology**, [s. l.], v. 17, n. 12, p. 1789–1797, 2020.

LOMBARDO, Michael V. *et al.* Hierarchical cortical transcriptome disorganization in autism. **Molecular Autism**, [s. l.], v. 8, n. 1, p. 29, 2017.

LOOMES, Rachel; HULL, Laura; MANDY, William Polmeare Locke. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. **Journal of the American Academy of Child and Adolescent Psychiatry**, [s. l.], v. 56, n. 6, p. 466–474, 2017.

LUU, Winnie *et al.* Signaling regulates activity of DHCR24, the final enzyme in cholesterol synthesis[s]. **Journal of Lipid Research**, [s. l.], v. 55, n. 3, p. 410–420, 2014.

LYONS, Viktoria; FITZGERALD, Michael. Asperger (1906-1980) and Kanner (1894-1981), the two pioneers of autism [2]. **Journal of Autism and Developmental Disorders**, [s. l.], v. 37, n. 10, p. 2022–2023, 2007.

MAASS, Philipp G. *et al.* A map of human circular RNAs in clinically relevant tissues. **Journal of Molecular Medicine**, [s. l.], v. 95, n. 11, p. 1179–1189, 2017.

MAENNER, Matthew J. *et al.* Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2018. **Morbidity and Mortality Weekly Report. Surveillance Summaries (Washington, D.C.: 2002)**, [s. l.], v. 70, n. 11, p. 1–16, 2021.

MAGUIRE, Martin; ISHERWOOD, Paul. A comparison of user testing and heuristic evaluation methods for identifying website usability problems. *Em: , 2018. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [S. l.: s. n.], 2018.

MANGUL, Serghei *et al.* Improving the usability and archival stability of bioinformatics software. **Genome Biology**, [s. l.], v. 20, n. 1, p. 47, 2019.

MANNAPPERUMA, Chanaka; STREET, Nathaniel; WATERWORTH, John. Designing Usable Bioinformatics Tools for Specialized Users. *Em: ADVANCES IN INTELLIGENT SYSTEMS AND COMPUTING*. [S. l.]: Springer Verlag, 2019. v. 918, p. 649–670.

MANOUILENKO, Irina; BEJEROT, Susanne. Sukhareva--Prior to Asperger and Kanner. **Nordic Journal of Psychiatry**, [s. l.], v. 69, n. 6, p. 479–482, 2015.

MARIÑO-RAMÍREZ, Leonardo *et al.* Histone structure and nucleosome stability. **Expert review of proteomics**, [s. l.], v. 2, n. 5, p. 719–729, 2005.

MARLOW, Marguerite; SERVILI, Chiara; TOMLINSON, Mark. A review of screening tools for the identification of autism spectrum disorders and developmental delay in infants and young children: recommendations for use in low- and middle-income countries. **Autism Research: Official Journal of the International Society for Autism Research**, [s. l.], v. 12, n. 2, p. 176–199, 2019.

MASI, A. *et al.* Cytokine aberrations in autism spectrum disorder: a systematic review and meta-analysis. **Molecular Psychiatry**, [s. l.], v. 20, n. 4, p. 440–446, 2015.

MCCALL, Jim a.; RICHARDS, Paul K.; WALTERS, Gene F. Factors in Software Quality. **at'I Tech. Information Servicel**, [s. l.], 1977.

MCFARLANE, H. G. *et al.* Autism-like behavioral phenotypes in BTBR T+tf/J mice. **Genes, Brain, and Behavior**, [s. l.], v. 7, n. 2, p. 152–163, 2008.

MELONI, Ilaria *et al.* A Mutation in the Rett Syndrome Gene, MECP2, Causes X-Linked Mental Retardation and Progressive Spasticity in Males. **The American Journal of Human Genetics**, [s. l.], v. 67, n. 4, p. 982–985, 2000.

MEMCZAK, Sebastian *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. **Nature**, [s. l.], v. 495, n. 7441, p. 333–338, 2013a.

MEMCZAK, Sebastian *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. **Nature**, [s. l.], v. 495, n. 7441, p. 333–338, 2013b.

MENEGIDIO, Fabiano B *et al.* Dugong: a Docker image, based on Ubuntu Linux, focused on reproducibility and replicability for bioinformatics analyses. **Bioinformatics**, [s. l.], v. 34, n. 3, p. 514–515, 2018.

MEYZA, K. Z.; BLANCHARD, D. C. The BTBR mouse model of idiopathic autism – current view on mechanisms. **Neuroscience and biobehavioral reviews**, [s. l.], v. 76, n. Pt A, p. 99–110, 2017.

MIREL, Barbara; WRIGHT, Zach. Heuristic Evaluations of Bioinformatics Tools: A Development Case. *Em: LECTURE NOTES IN COMPUTER SCIENCE (INCLUDING SUBSERIES LECTURE NOTES IN ARTIFICIAL INTELLIGENCE AND LECTURE NOTES IN BIOINFORMATICS)*. [S. l.]: Springer, Berlin, Heidelberg, 2009. v. 5610 LNCS, p. 329–338.

MONFARED, Roudabeh Vakil *et al.* Transcriptome Profiling of Dysregulated GPCRs Reveals Overlapping Patterns across Psychiatric Disorders and Age-Disease Interactions. **Cells**, [s. l.], v. 10, n. 11, p. 2967, 2021.

MONTEIRO, Patricia; FENG, Guoping. SHANK proteins: Roles at the synapse and in autism spectrum disorder. **Nature Reviews Neuroscience**, [s. l.], 2017.

MORRIS-ROSENDAHL, Deborah J.; CROCQ, Marc-Antoine. Neurodevelopmental disorders—the history and future of a diagnostic concept. **Dialogues in Clinical Neuroscience**, [s. l.], v. 22, n. 1, p. 65–72, 2020.

MUKHERJEE, P. *et al.* Role of animal models in biomedical research: a review. **Laboratory Animal Research**, [s. l.], v. 38, n. 1, p. 18, 2022.

NASSIF, Mathieu *et al.* Generating Unit Tests for Documentation. **IEEE Transactions on Software Engineering**, [s. l.], p. 1, 2021.

NESTLER, Eric J.; HYMAN, Steven E. Animal Models of Neuropsychiatric Disorders. **Nature neuroscience**, [s. l.], v. 13, n. 10, p. 1161–1169, 2010.

NGUYEN, Manh Hung; NGUYEN, Ha-Nam; VU, Trung Nghia. Evaluation of methods to detect circular RNAs from single-end RNA-sequencing data. **BMC Genomics**, [s. l.], v. 23, n. 1, p. 106, 2022.

NIE, Jun Hua *et al.* **Roles of non-coding RNAs in normal human brain development, brain tumor, and neuropsychiatric disorders**. [S. l.]: MDPI AG, 2019.

NIELSEN, Jakob. Enhancing the explanatory power of usability heuristics. *Em:* , 1994. **Conference on Human Factors in Computing Systems - Proceedings**. [S. l.: s. n.], 1994.

NIELSEN, Jakob. Finding usability problems through heuristic evaluation. *Em:* , 1992, New York, New York, USA. **Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '92**. New York, New York, USA: ACM Press, 1992. p. 373–380. Disponível em: <http://portal.acm.org/citation.cfm?doid=142750.142834>.

NIELSEN, Jakob; MOLICH, Rolf. Heuristic Evaluation of User Interfaces. *Em:* , 1990a, New York, NY, USA. **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 1990. p. 249–256. Disponível em: <https://doi.org/10.1145/97243.97281>.

NIELSEN, Jakob; MOLICH, Rolf. Heuristic Evaluation of User Interfaces. *Em:* , 1990b, New York, NY, USA. **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 1990. p. 249–256.

O'RAWE, Jason *et al.* Low concordance of multiple variant-calling pipelines: Practical implications for exome and genome sequencing. **Genome Medicine**, [s. l.], v. 5, n. 3, p. 28, 2013.

PANDA, Amaresh Chandra. Circular RNAs Act as miRNA Sponges. **Advances in Experimental Medicine and Biology**, [s. l.], v. 1087, p. 67–79, 2018.

PARIKSHAK, Neelroop N. *et al.* Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. **Nature**, [s. l.], v. 540, n. 7633, p. 423–427, 2016.

PARIKSHAK, Neelroop N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. **Cell**, [s. l.], v. 155, n. 5, p. 1008–1021, 2013.

PARKER-ATHILL, E. Carla; TAN, Jun. Maternal immune activation and autism spectrum disorder: interleukin-6 signaling as a key mechanistic pathway. **Neuro-Signals**, [s. l.], v. 18, n. 2, p. 113–128, 2010.

PASTORE, Stephen F. *et al.* PTCHD1: Identification and Neurodevelopmental Contributions of an Autism Spectrum Disorder and Intellectual Disability Susceptibility Gene. **Genes**, [s. l.], v. 13, n. 3, p. 527, 2022.

PATTERSON, Paul H. Modeling Autistic Features in Animals. **Pediatric Research**, [s. l.], v. 69, n. 8, p. 34–40, 2011.

PAZ, Freddy; POW-SANG, José Antonio. Usability Evaluation Methods for Software Development: A Systematic Mapping Review. *Em:* , 2016. **Proceedings - 8th International Conference on Advanced Software Engineering and Its Applications, ASEA 2015**. [S. l.: s. n.], 2016.

PEÇA, João *et al.* Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. **Nature**, [s. l.], v. 472, n. 7344, p. 437–442, 2011.

PEDDADA, Sailaja; YASUI, Dag H.; LASALLE, Janine M. Inhibitors of differentiation (ID1, ID2, ID3 and ID4) genes are neuronal targets of MeCP2 that are elevated in Rett syndrome. **Human Molecular Genetics**, [s. l.], v. 15, n. 12, p. 2003–2014, 2006.

PETRELLI, Francesco; PUCCI, Luca; BEZZI, Paola. Astrocytes and Microglia and Their Potential Link with Autism Spectrum Disorders. **Frontiers in Cellular Neuroscience**, [s. l.], v. 10, 2016. Disponível em: <https://www.frontiersin.org/articles/10.3389/fncel.2016.00021>. Acesso em: 11 nov. 2022.

PIETROPAOLO, Susanna *et al.* Genetic-background modulation of core and variable autistic-like symptoms in Fmr1 knock-out mice. **PloS One**, [s. l.], v. 6, n. 2, p. e17073, 2011.

PIWECKA, Monika *et al.* Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. **Science**, [s. l.], v. 357, n. 6357, 2017a. Disponível em: Acesso em: 12 dez. 2019.

PIWECKA, Monika *et al.* Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. **Science**, [s. l.], v. 357, n. 6357, 2017b.

POSAR, Annio; VISCONTI, Paola. Tribute to Grunya Efimovna Sukhareva, the Woman who First Described Infantile Autism. **Journal of Pediatric Neurosciences**, [s. l.], v. 12, n. 3, p. 300–301, 2017.

QIN, Luye *et al.* Deficiency of autism risk factor ASH1L in prefrontal cortex induces epigenetic aberrations and seizures. **Nature Communications**, [s. l.], v. 12, n. 1, p. 6589, 2021.

RODER, Holger. Specifying usability features with patterns and templates. *Em: , 2012. 2012 1st International Workshop on Usability and Accessibility Focused Requirements Engineering, UsARE 2012 - Proceedings*. [S. l.: s. n.], 2012.

ROGERS, S. J. *et al.* Autism Treatment in the First Year of Life: A Pilot Study of Infant Start, a Parent-Implemented Intervention for Symptomatic Infants. **Journal of autism and developmental disorders**, [s. l.], v. 44, n. 12, p. 2981–2995, 2014.

RONG, Dawei *et al.* An emerging function of circRNA-miRNAs-mRNA axis in human diseases. **Oncotarget**, [s. l.], v. 8, n. 42, 2017.

ROSE, S. *et al.* Evidence of oxidative damage and inflammation associated with low glutathione redox status in the autism brain. **Translational Psychiatry**, [s. l.], v. 2, p. e134, 2012.

ROULLET, Florence I; LAI, Jonathan K Y; FOSTER, Jane a. In utero exposure to valproic acid and autism--a current review of clinical and animal studies. **Neurotoxicology and teratology**, [s. l.], v. 36, p. 47–56, 2013.

RYBAK-WOLF, Agnieszka *et al.* Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. **Molecular Cell**, [s. l.], v. 58, n. 5, p. 870–885, 2014a.

RYBAK-WOLF, Agnieszka *et al.* Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. **Molecular Cell**, [s. l.], v. 58, n. 5, p. 870–885, 2014b.

SALZMAN, Julia *et al.* Cell-Type Specific Features of Circular RNA Expression. **PLoS Genetics**, [s. l.], v. 9, n. 9, 2013.

SALZMAN, Julia *et al.* Circular RNAs Are the Predominant Transcript Isoform from Hundreds of Human Genes in Diverse Cell Types. **PLoS ONE**, [s. l.], v. 7, n. 2, p. e30733, 2012.

SANGER, H. L. *et al.* Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. **Proceedings of the National Academy of Sciences of the United States of America**, [s. l.], v. 73, n. 11, p. 3852–3856, 1976.

SANTOS-TERRA, Júlio *et al.* Resveratrol Prevents Cytoarchitectural and Interneuronal Alterations in the Valproic Acid Rat Model of Autism. **International Journal of Molecular Sciences**, [s. l.], v. 23, n. 8, p. 4075, 2022.

SATTERSTROM, F. Kyle *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. **Cell**, [s. l.], v. 180, n. 3, p. 568-584.e23, 2020.

SCHROEDER, Jan C. *et al.* Genetic Animal Models for Autism Spectrum Disorder. **Current Topics in Behavioral Neurosciences**, [s. l.], v. 30, p. 311–324, 2017.

SEEMANN, Torsten. **Ten recommendations for creating usable bioinformatics command line software**. [S. l.]: BioMed Central Ltd., 2013.

SEKAR, Shobana; LIANG, Winnie S. **Circular RNA expression and function in the brain**. [S. l.]: KeAi Communications Co., 2019.

SIKORA, Darryn M. *et al.* The near universal presence of autism spectrum disorders in children with Smith-Lemli-Opitz syndrome. **American Journal of Medical Genetics, Part A**, [s. l.], v. 140, n. 14, p. 1511–1518, 2006.

SINHA, Tanvi *et al.* Circular RNA translation, a path to hidden proteome. **Wiley interdisciplinary reviews. RNA**, [s. l.], v. 13, n. 1, p. e1685, 2022.

SPLENDORE, Alessandra. Para que existem as regras de nomenclatura genética?. **Revista Brasileira de Hematologia e Hemoterapia**, [s. l.], v. 27, p. 148–152, 2005.

SUN, Wenjie *et al.* Histone Acetylome-wide Association Study of Autism Spectrum Disorder. **Cell**, [s. l.], v. 167, n. 5, p. 1385-1397.e11, 2016.

SZABO, Linda *et al.* Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development. **Genome Biology**, [s. l.], v. 16, n. 1, 2015.

SZABO, Linda; SALZMAN, Julia. Detecting circular RNAs: bioinformatic and experimental challenges. **Nature Reviews Genetics**, [s. l.], v. 17, n. 11, p. 679–692, 2016.

TAKEDA, Kotomi *et al.* Valproic acid-exposed astrocytes impair inhibitory synapse formation and function. **Scientific Reports**, [s. l.], v. 11, n. 1, p. 23, 2021.

TROBIANI, Laura *et al.* The neuroligins and the synaptic pathway in Autism Spectrum Disorder. **Neuroscience and Biobehavioral Reviews**, [s. l.], v. 119, p. 37–51, 2020.

UNG, D. C. *et al.* Ptchd1 deficiency induces excitatory synaptic and cognitive dysfunctions in mouse. **Molecular Psychiatry**, [s. l.], v. 23, n. 5, p. 1356–1367, 2018.

USUI, Noriyoshi *et al.* VLDL-specific increases of fatty acids in autism spectrum disorder correlate with social interaction. **EBioMedicine**, [s. l.], v. 58, 2020.

VAN DER STAAY, F. Josef. Animal models of behavioral dysfunctions: basic concepts and classifications, and an evaluation strategy. **Brain Research Reviews**, [s. l.], v. 52, n. 1, p. 131–159, 2006.

VAN 'T HOF, Maarten *et al.* Age at autism spectrum disorder diagnosis: A systematic review and meta-analysis from 2012 to 2019. **Autism: The International Journal of Research and Practice**, [s. l.], v. 25, n. 4, p. 862–873, 2021.

VATSA, Naman; JANA, Nihar Ranjan. UBE3A and Its Link With Autism. **Frontiers in Molecular Neuroscience**, [s. l.], v. 11, 2018. Disponível em: <https://www.frontiersin.org/articles/10.3389/fnmol.2018.00448>. Acesso em: 14 nov. 2022.

VINGEN, David; ANDREWS, Elias J; FERATI, Mexhid. Usability in Patient-Oriented Drug Interaction Checkers—A Scandinavian Sampling and Heuristic Evaluation. **Informatics**, [s. l.], v. 7, n. 4, 2020.

VOINEAGU, Irina *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. **Nature**, [s. l.], v. 474, n. 7351, p. 380–384, 2011.

WAHID, Fazli *et al.* **MicroRNAs: Synthesis, mechanism, function, and recent clinical trials**. [S. l.: s. n.], 2010.

WANG, Zi-Jun *et al.* Autism risk gene KMT5B deficiency in prefrontal cortex induces synaptic dysfunction and social deficits via alterations of DNA repair and gene transcription. **Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology**, [s. l.], v. 46, n. 9, p. 1617–1626, 2021.

WANG, Ji *et al.* Comprehensive circRNA Expression Profile and Construction of circRNAs-Related ceRNA Network in a Mouse Model of Autism. **Frontiers in Genetics**, [s. l.], v. 11, p. 623584, 2021.

WANG, Y. *et al.* Genome-wide differential expression of synaptic long noncoding RNAs in autism spectrum disorder. **Translational Psychiatry**, [s. l.], v. 5, p. e660, 2015.

WANG, Xiaoming *et al.* Synaptic dysfunction and abnormal behaviors in mice lacking major isoforms of Shank3. **Human Molecular Genetics**, [s. l.], v. 20, n. 15, p. 3093–3108, 2011.

WANG, Enlie; CALDWELL, Barrett. An Empirical Study of Usability Testing: Heuristic Evaluation Vs. User Testing. **Proceedings of the Human Factors and Ergonomics Society Annual Meeting**, [s. l.], v. 46, n. 8, p. 774–778, 2002.

WU, Ye E. *et al.* Genome-wide, integrative analysis implicates microRNA dysregulation in autism spectrum disorder. **Nature Neuroscience**, [s. l.], 2016.

WU, Wanying; JI, Peifeng; ZHAO, Fangqing. CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. **Genome Biology**, [s. l.], v. 21, n. 1, p. 101, 2020.

XENOS, Michalis. Usability perspective in software quality. *Em:* , 2001. **Usability Engineering Workshop, The 8th Panhellenic Conference on Informatics with International Participation, Southern Cyprus**. [S. l.: s. n.], 2001.

XIA, Siyu *et al.* Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. **Briefings in Bioinformatics**, [s. l.], v. 18, n. 6, p. 984–992, 2017.

XIE, Binhui *et al.* CircRNA has_circ_0078710 acts as the sponge of microRNA-31 involved in hepatocellular carcinoma progression. **Gene**, [s. l.], v. 683, p. 253–261, 2019.

XU, Kaiyu *et al.* Annotation and functional clustering of circRNA expression in rhesus macaque brain during aging. **Cell Discovery**, [s. l.], v. 4, n. 1, 2018.

XU, Qiong *et al.* Autism-associated CHD8 deficiency impairs axon development and migration of cortical neurons. **Molecular Autism**, [s. l.], v. 9, n. 1, p. 65, 2018.

XU, Tianyi *et al.* Circular RNA expression profiles and features in human tissues: a study using RNA-seq data. **BMC Genomics**, [s. l.], v. 18, n. Suppl 6, p. 680, 2017.

YAN, Yuze *et al.* ASH1L haploinsufficiency results in autistic-like phenotypes in mice and links Eph receptor gene to autism spectrum disorder. **Neuron**, [s. l.], v. 110, n. 7, p. 1156-1172.e9, 2022.

YANG, Ruicheng *et al.* Non-coding RNAs: the extensive and interactive regulators of the blood-brain barrier permeability. **RNA biology**, [s. l.], v. 18, n. sup1, p. 108–116, 2021.

YOU, Xintian *et al.* Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. **Nature Neuroscience**, [s. l.], v. 18, n. 4, p. 603–610, 2015.

ZANG, Julie B. *et al.* A mouse model of the human Fragile X syndrome I304N mutation. **PLoS genetics**, [s. l.], v. 5, n. 12, p. e1000758, 2009.

ZEIDAN, Jinan *et al.* Global prevalence of autism: A systematic review update. **Autism Research**, [s. l.], v. 15, n. 5, p. 778–790, 2022.

ZENG, Xiangxiang *et al.* **A comprehensive overview and evaluation of circular RNA detection tools**. [S. l.: s. n.], 2017.

ZHANG, Jinyang *et al.* Accurate quantification of circular RNAs identifies extensive circular isoform switching events. **Nature Communications**, [s. l.], v. 11, n. 1, p. 90, 2020.

ZHANG, Xiaoqing *et al.* Circular RNAs and esophageal cancer. **Cancer Cell International**, [s. l.], v. 20, n. 1, p. 362, 2020.

ZHANG, Xiao Ou *et al.* Complementary sequence-mediated exon circularization. **Cell**, [s. l.], v. 159, n. 1, p. 134–147, 2014.

ZHANG, Xiao Ou *et al.* Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. **Genome Research**, [s. l.], v. 26, n. 9, p. 1277–1287, 2016a.

ZHANG, Xiao Ou *et al.* Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. **Genome Research**, [s. l.], v. 26, n. 9, p. 1277–1287, 2016b.

ZHANG, Yang *et al.* The Biogenesis of Nascent Circular RNAs. **Cell Reports**, [s. l.], v. 15, n. 3, p. 611–624, 2016.

ZHAO, Yuhai *et al.* Deficiency in the ubiquitin conjugating enzyme UBE2A in Alzheimer's Disease (AD) is linked to deficits in a natural circular miRNA-7 sponge (circRNA; ciRS-7). **Genes**, [s. l.], v. 7, n. 12, 2016.

ZIATS, Mark N.; RENNERT, Owen M. Aberrant expression of long noncoding RNAs in autistic brain. **Journal of molecular neuroscience: MN**, [s. l.], v. 49, n. 3, p. 589–593, 2013.

ANEXOS

ANEXO I - Parecer de projeto de doutorado encaminhado para avaliação pelo Programa de Pós-Graduação em Ciências Biológicas: Bioquímica

PPG
Ciências
Biológicas
Bioquímica

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE CIÊNCIAS BÁSICAS DA SAÚDE

PARECER COSUBSTANCIADO

O projeto de pesquisa intitulado “ANÁLISE DE BIOINFORMÁTICA E EXPERIMENTAL DO PERFIL DE CIRCNRNAS NO SISTEMA NERVOSO CENTRAL DE RATOS DO MODELO ANIMAL DE AUTISMO INDUZIDO POR EXPOSIÇÃO PRÉ-NATAL AO ÁCIDO VALPROICO”, do aluno **Guilherme Bauer Negrini**, sob a orientação da Profa. Dra. **Carmem Juracy Silveira Gottfried**, vinculado ao PPG Ciências Biológicas – Bioquímica desta Universidade, apresentado para apreciação da Comissão de Pós-Graduação, apresenta:

- Um tema relevante e com mérito científico uma vez que propõe caracterizar o perfil de circRNAs no córtex pré-frontal e no corpo caloso de animais do modelo animal de autismo induzido por exposição pré-natal ao ácido valproico.
- Uma introdução fundamentada e com objetivos adequados ao estudo proposto. A metodologia é coerente para que os objetivos sejam alcançados, havendo até este momento uma infraestrutura adequada no Departamento para sua realização.
- Um cronograma de execução que prevê o período de 48 meses para a realização do projeto.

O projeto atende as exigências do “n” amostral, do detalhamento dos procedimentos experimentais e do tratamento dos rejeitos.

Sendo assim, somos de parecer favorável à aprovação do presente projeto.

Parecer aprovado em 02/03/2020



Prof. Daniel Pens Gelain

Prof. DANIEL PENS GELAIN
Depto. de Bioquímica / ICBS
UFRGS / POA



Prof. Carlos Alberto Gonçalves
Coordenador do PPG em C.B.:
Bioquímica-ICBS/UFRGS

Parecer de Projeto de Doutorado

O projeto do aluno **Guilherme Bauer Negrini**, intitulado: “ANÁLISE DE BIOINFORMÁTICA E EXPERIMENTAL DO PERFIL DE CIRCURNAS NO SISTEMA NERVOSO CENTRAL DE RATOS DO MODELO ANIMAL DE AUTISMO INDUZIDO POR EXPOSIÇÃO PRÉ-NATAL AO ÁCIDO VALPROICO”, orientado pela Profa. Dra. Carmem Juracy Silveira Gottfried, encaminhado para avaliação pelo PPG - Bioquímica teve a seguinte análise e parecer:

ANÁLISE DO PROJETO

1. Mérito científico

Relevante

Sugestões de alterações:

2. Fundamentação

Adequada

Sugestões de alterações:

3. Objetivos

Bem definidos

Sugestões de alterações:

4. Metodologia

Adequada

Sugestões de alterações:

5. Cálculo do número amostral

Realizado

Não aplicável

Não realizado:

6. Considerações Éticas

- Descritas
 Não aplicável
 Sugestões de alterações:

7. Descarte dos Resíduos Químicos e Biológicos

- Descritos
 Não aplicável
 Sugestões de alterações:

8. Cronograma de Execução

- Apresentado
 Não apresentado

9. Referências Bibliográficas

- Pertinentes
 Sugestões de alterações:

PARECER FINAL:


- Aprovado
 Diligência

Porto Alegre, __02__ de __março__ de 2020



Daniel Pens Gelain

Prof. DANIEL PENS GELAIN
Depto. de Bioquímica / ICBS
UFRGS / FOA



Prof. Carlos Alberto Gonçalves
Coordenador do PPG em C.B.:
Bioquímica-ICBS/UFRGS

ANEXO II – Ficha de avaliação heurística da usabilidade de software de identificação de circRNA

Session information	
Evaluator	
Software	
Date	

Hierarchical documentation				
Id	Heuristic	Issues	Recommendations	Severity
H1	Comprehensive and hierarchically organized documentation			
H2	Software license and citation			
H3	Versioning protocol and information about updates.			
Installability				
Id	Heuristic	Issues	Recommendations	Severity
H4	Software dependencies and package requirements			
H5	Easy installation			
H6	Test data			
Operability				
Id	Heuristic	Issues	Recommendations	Severity
H7	Standard input and output file formats			
H8	Supports input from both command-line interface and configuration file.			

H9	Default parameters for novice users and optional parameters for more experienced users			
User error protection				
Id	Heuristic	Issues	Recommendations	Severity
H10	Validation of input parameters			
H11	Prevents accidental file overwrite			
Visibility				
Id	Heuristic	Issues	Recommendations	Severity
H12	Updated information about the analysis progress			
H13	Verbose mode			
H14	Detailed and intelligible error messages			
H15	Logging option			

Severity scale	
0	None: not a usability problem.
1	Cosmetic: subtle problems; users can easily proceed; lower priority issue.
2	Minor: the issue has minor effects on usability; it may hinder novice users; most users can proceed based on their experience or by consulting the documentation.
3	Major: the issue has significant effects on usability; it can cause significant delay or frustration; users may not find information or help on documentation; users may proceed based on external help; some degree of programming skills may be required to proceed; novice users can experience several challenges and fail to proceed.
4	Critical: the issue may prevent the complete execution of the tool/pipeline; the issue can be solved by experienced users but requires advanced programming skills and/or several hours; users are unlikely to insist on using the tool.