

Universidade de São Paulo  
Faculdade de Medicina de Ribeirão Preto

**A heterogeneidade fenotípica no Transtorno do Espectro do Autismo e sua associação com o impacto de variantes exônicas deletérias**

Catarina dos Santos Gomes

Trabalho de conclusão de curso apresentado à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para a obtenção do título de Bacharela em Ciências Biomédicas - Ênfase em Ciências Básicas da Saúde

**Orientadora:**  
Profa. Dra. Helena Brentani

**Co-orientador:**  
Prof. Dr. Ricardo Zorzetto Nicoliello Vêncio

Ribeirão Preto  
2020

Autorizo a reprodução e divulgação total deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.  
A inclusão deste trabalho foi aprovada pela Comissão Coordenadora do Curso em sua 57<sup>a</sup> Sessão Ordinária, realizada em 04/10/2021.



# Agradecimentos

Em primeiro lugar gostaria de agradecer aos meus pais, Deise e Cristiano (*in memoriam*) por todo o apoio e esforço depositados em mim e nos meus estudos ao longo da minha vida. Sem os incentivos de vocês eu não seria a primeira pessoa da família a ingressar e concluir um curso em uma das melhores Universidades da América Latina. Agradeço também aos meus avós, Mariângela e Francisco por todo o apoio emocional ao longo do último ano.

Agradeço à Prof<sup>a</sup> Dra. Helena Brentani por ter me dado a oportunidade de trabalhar no seu grupo de pesquisa, de colaborar com outras pessoas, por acreditar no meu trabalho e por me ensinar novas formas de entender a ciência e a psiquiatria.

Ao Prof. Dr. Ricardo Vêncio, por toda paciência, tempo, disponibilidade, confiança e pelos valiosos ensinamentos sobre rigor científico e reflexões de vida que levarei comigo sempre. Sem a sua co-orientação, este trabalho não seria possível.

À Prof<sup>a</sup> Dra. Tie Koide por ter me acolhido no LaBiSisMi durante meus primeiros anos de graduação. À todos os colegas do LaBiSisMi, especialmente ao Alan Lorenzetti e João Almeida por todo o acompanhamento, (muita) paciência e dedicação ao ensinar. Vocês foram um divisor de águas na minha vida e me fizeram decidir construir uma carreira na bioinformática. Sem a passagem pelo LaBiSisMi eu não seria capaz de fazer tudo o que faço hoje e foi sem dúvidas, a melhor iniciação científica que alguém pode fazer.

Aos demais excelentes professores com quem tive a oportunidade de aprender e que se esforçam muito para oferecer uma estrutura diferenciada e única na formação de um cientista por meio do curso de Ciências Biomédicas: Eduardo Brandt de Oliveira, Angela Kaysel Cruz, Davi José de Almeida Moraes, Adriano Silva Sebollela, Vania Luiza Deperon Bonato, Alline Cristina Campos, Lucila Leico Kagohara Elias, Eurico de Arruda Neto, Leticia Fröhlich Archangelo, Vitor Marcel Faça, Luis Lamberti Silva, Mariana Kiomy Osako, e todos os demais.

Aos meus colegas do PsysBio: Victor, Cleiton, Verônica, Leonardo, Arthur e Vinícius pela convivência harmoniosa e também pelas noites de discussão de projetos no laboratório. Agradeço especialmente à Joana Portolese por toda a parceria, amizade, apoio, empatia e momentos de diversão durante esse último ano e meio. Você me abriu os olhos para a importância da pesquisa translacional na área de autismo e espero que possamos trabalhar em colaboração por muito mais tempo.

Aos meus amigos da graduação, Giovanna, Gabriel Martins, Ana Clara e Giseli, que fizeram os momentos mais difíceis se tornarem muito mais leves durante esses quatro anos.

*"The intensity of the conviction that a hypothesis is true has no bearing on whether it is true or not."*

- Peter B. Medawar



# Resumo

O transtorno do espectro do autismo (TEA) é um transtorno do neurodesenvolvimento de origem poligênica e multifatorial, caracterizado por alterações na sociabilização e presença de padrões restritivos e repetitivos de comportamento, de interesses ou de atividades. Estudos do Sequenciamento Completo do Exoma permitiram grande avanço no entendimento das bases moleculares do TEA e hoje, diversos genes e tipos de variantes já foram apontados como relevantes. Entretanto, ainda há um grande desafio em transformar essa lista de genes em informações sólidas que permitam o entendimento sistemático dos mecanismos e das vias biológicas que de fato contribuem para a apresentação clínica no TEA. Esse desafio se deve em parte à grande heterogeneidade de apresentação fenotípica observada. Portanto, para investigar transtornos mentais altamente heterogêneos faz-se necessário a análise conjunta de genes, moléculas, vias biológicas, circuitos neurais e dimensões de comportamento. Estudos que investigam relações genética-fenótipo existentes no TEA analisam a associação entre a carga mutacional e instrumentos de avaliação fenotípica utilizados na clínica de uma forma individual, desconsiderando os aspectos globais de heterogeneidade da apresentação fenotípica que contribuem fortemente para o espectro de gravidade nos indivíduos. Este trabalho foi dividido em duas etapas. Na primeira etapa, buscamos analisar diferentes instrumentos de avaliação fenotípica utilizados na prática clínica em pacientes afetados pelo TEA e por meio do uso da Análise de Componentes Principais buscamos componentes principais que melhor caracterizassem a heterogeneidade fenotípica observada como um *continuum* de gravidade. Utilizando a abordagem de Modelagem Normativa pudemos criar um escore de gravidade com base nos componentes contribuintes para a heterogeneidade fenotípica. Na segunda etapa, verificamos se o *continuum* de gravidade sistematizado poderia ser influenciado por variantes deletérias em grupos de genes específicos. Neste estudo, analisamos os dados fenotípicos de 2.744 pacientes com TEA provenientes da coorte *Simons Simplex Collection* (SSC), da base de dados da *Simons Foundation for Autism Research Initiative* assim como dados de variantes exônicas *de novo* e raras herdadas deletérias em Genes Essenciais e Genes Não Essenciais para o desenvolvimento. Empregando a combinação de técnicas de análise de dados multivariada e de modelagem normativa em instrumentos fenotípicos que refletem os sintomas core do TEA e seus fatores agravantes, foi possível construir um Escore de Gravidade Multidimensional que permite a comparação da gravidade de apresentação fenotípica observada entre indivíduos no que diz respeito à problemas de funcionalidade social, regulação do comportamento e problemas de comunicação. Na segunda parte do estudo, a análise de variantes exônicas que afetam os indivíduos da mesma coorte, indicou que a interação entre variantes do tipo *de novo* de perda de função e *de novo* não sinônimas deletérias em Genes Não Essenciais estão associadas ao Escore de Gravidade Multidimensional em meninas ( $p$ -valor = 0.0259; regressão beta). Ademais, foi possível verificar que a ocorrência dessas variantes em genes que participam de módulos de co-expressão em 16 diferentes regiões cerebrais também estão associadas ao escore. É possível concluir que a utilização de modelos normativos em conjunto com técnicas de análise multivariada permitem

a análise do TEA como um transtorno multidimensional do ponto de vista fenotípico e que variantes do tipo *de novo* de perda de função e *de novo* não sinônimas deletérias estão associadas à gravidade global de apresentação clínica quando consideramos grupos de genes específicos. No entanto, para aumentar o poder explicativo do escore de gravidade, outros fatores genéticos, epigenéticos e ambientais precisam ser investigados.

**Palavras-chaves:** Transtorno do Espectro do Autismo; Heterogeneidade Fenotípica; Sequenciamento Completo do Exoma

# Abstract

Autism spectrum disorder (ASD) is a neurodevelopmental disorder of polygenic and multifactorial origin, characterized by changes in sociability and the presence of restrictive and repetitive patterns of behavior. Analysis of Whole Exome Sequencing (WES) has been contributing to a better understanding of the molecular bases of ASD and today, several genes and types of variants have been pointed out as relevant. However, there is still a great challenge in unifying those lists of genes in information which allows a systematic understanding of the mechanisms and biological pathways that contribute to the clinical presentation in ASD. This challenge is due in part to the great heterogeneity of the phenotypic presentation observed. To investigate highly heterogeneous mental disorders, it is necessary to jointly analyze genes, molecules, biological pathways, neural circuits and dimensions of behavior. Studies investigating existing genetic-phenotype relationships in ASD analyze the association between mutational burden and phenotypic assessment instruments used in the clinic on an individual basis, disregarding the global aspects of heterogeneity of the phenotypic presentation that strongly contribute to the spectrum of severity in individuals. This work was performed in two steps. In the first one, we seek to analyze different instruments for phenotypic evaluation used in clinical practice in patients affected by ASD and through the use of Principal Component Analysis we seek main components that best characterize the phenotypic heterogeneity observed as a continuum of severity. Through the Normative Modeling approach it was possible to create a severity score based on the components that contribute to severity. In the second step, we checked whether the systematized severity continuum could be associated with deleterious variants in groups of genes and specific biological pathways. For this purpose, we analyzed the phenotypic data of 2.744 ASD patients from the Simons Simplex Collection (SSC) cohort, from the Simons Foundation for Autism Research Initiative. We also analyzed data from *de novo* and rare inherited damaging exonic variants in Essential and Non-Essential Genes for development. When combining multivariate data analysis techniques and normative modeling, it was possible to create a Multidimensional Severity Score that allows the comparison of the severity of phenotypic presentation observed in individuals when considering social functionality, behavior problems and communication problems. In the second part of the study, the analysis of exonic variants that affect individuals in the same cohort, indicated that the interaction between *de novo* loss of function variants and *de novo* nonsense damaging variants in Non-Essential Genes, are slightly associated with the Multidimensional Severity Score in females ( $p$ -value = 0.0259; beta regression). Furthermore, it was possible to verify that the occurrence of these variants in genes that participate in co-expression modules in 16 different brain regions is also associated with the score. It is possible to conclude that the use of normative models with multivariate analysis techniques allows the analysis of ASD as a phenotypic multidimensional disorder and that *de novo* loss of function and *de novo* non-synonymous damaging variants are partially associated with the global severity of clinical presentation when considering specific groups of genes. However, to increase the explanatory power of the severity score, other genetic, epigenetic

and environmental factors need to be investigated.

**Key-words:** Autism Spectrum Disorder; Phenotypic Heterogeneity; Whole Exome Sequencing;

# Lista de ilustrações

Figura 1 – Resumo do método . . . . .	15
Figura 2 – Representação gráfica de uma distribuição normal multivariada, em duas dimensões . . . . .	16
Figura 3 – Representação gráfica de uma distribuição normal multivariada, em três dimensões. . . . .	18
Figura 4 – Variância explicada por cada componente principal. . . . .	25
Figura 5 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e a VABS Comunicação. . . . .	30
Figura 6 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e a VABS Socialização. . . . .	30
Figura 7 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e a VABS Habilidades de Vida Diárias. . . . .	31
Figura 8 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e o QI. . . . .	31
Figura 9 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e a ADI-R Socialização. . . . .	32
Figura 10 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e a CBCL Problemas Internalizantes. . . . .	32
Figura 11 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e a CBCL Problemas Externalizantes. . . . .	33
Figura 12 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e a ADI-R Comportamentos Restritos e Repetitivos. . . . .	33
Figura 13 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e a ADI-R Comunicação. . . . .	34
Figura 14 – Implicações clínicas e conceituais da representação dimensional de indivíduos autistas. . . . .	35
Figura 15 – Eixo de variação fenotípica que representa a gravidade da apresentação clínica de forma geral. . . . .	36
Figura 16 – Distribuição dos valores de Escore de Gravidade Multidimensional nos 2.301 pacientes da SSC. . . . .	38
Figura 17 – Efeitos da interação entre variantes <i>de novo</i> de perda de função e <i>de novo</i> não sinônimas deletérias em Genes Não Essenciais na média marginal estimada do escore de gravidade multidimensional. . . . .	41

# Lista de tabelas

Tabela 1 – Médias e desvios padrão dos escores obtidos pelos pacientes com TEA provenientes da coorte SSC nos testes, escalas e questionários utilizados como entrada para a PCA. . . . .	23
Tabela 2 – Números de variantes do tipo <i>de novo</i> nos 2.301 indivíduos utilizados nas análises de associação genética . . . . .	24
Tabela 3 – Números de variantes do tipo rara herdada os 2.301 indivíduos utilizados nas análises de associação genética. . . . .	24
Tabela 4 – <i>Eigenvalues</i> e variância explicados por cada componente principal ao usar os subitens da VABS, ADI-R, QI total e CBCL internalizante e externalizante avaliados em 2.744 pacientes como medidas de entrada. . . . .	26
Tabela 5 – <i>Eigenvalues</i> e variância explicados por cada componente principal ao usar os subitens da VABS, ADI-R, QI total e CBCL internalizante e externalizante avaliados em 823 pacientes como medidas de entrada. . . . .	26
Tabela 6 – <i>Eigenvalues</i> e variância explicados por cada componente principal ao usar os subitens da VABS, ADI-R, QI total e CBCL internalizante e externalizante avaliados em 1.098 pacientes como medidas de entrada. . . . .	27
Tabela 7 – <i>Eigenvalues</i> e variância explicados por cada componente principal ao usar os subitens da VABS, ADI-R, QI total e CBCL internalizante e externalizante avaliados em 1.372 pacientes como medidas de entrada. . . . .	27
Tabela 8 – <i>Eigenvalues</i> e variância explicados por cada componente principal ao usar os subitens da VABS, ADI-R, QI total e CBCL internalizante e externalizante avaliados em 1.646 pacientes como medidas de entrada. . . . .	28
Tabela 9 – <i>Eigenvalues</i> e variância explicados por cada componente principal ao usar os subitens da VABS, ADI-R, QI total e CBCL internalizante e externalizante avaliados em 1.912 pacientes como medidas de entrada. . . . .	28
Tabela 10 – <i>Eigenvalues</i> e variância explicados por cada componente principal ao usar os subitens da VABS, ADI-R, QI total e CBCL internalizante e externalizante avaliados em 2.195 pacientes como medidas de entrada. . . . .	29
Tabela 11 – Associação do Escore de Gravidade Multidimensional e o total de variantes. . . . .	38
Tabela 12 – Associação do Escore de Gravidade Multidimensional e o total de variantes em Genes Não Essenciais. . . . .	38
Tabela 13 – Associação do Escore de Gravidade Multidimensional e o total de variantes em Genes Essenciais. . . . .	39
Tabela 14 – Associação do Escore de Gravidade Multidimensional e o total de variantes raras herdadas. . . . .	39
Tabela 15 – Associação do Escore de Gravidade Multidimensional e o total de variantes <i>de novo</i> de perda de função ( <i>Loss of Function</i> , LoF). . . . .	39

Tabela 16 – Associação do Escore de Gravidade Multidimensional e o total de variantes <i>de novo</i> não sinônimas deletérias. . . . .	39
Tabela 17 – Associação do Escore de Gravidade Multidimensional e o total de variantes raras herdadas em Genes Essenciais. . . . .	40
Tabela 18 – Associação do Escore de Gravidade Multidimensional e o total de variantes <i>de novo</i> de perda de função e <i>de novo</i> não sinônimas deletérias em Genes Essenciais. . . . .	40
Tabela 19 – Associação do Escore de Gravidade Multidimensional e o total de variantes raras herdadas em Genes Não Essenciais. . . . .	40
Tabela 20 – Associação do Escore de Gravidade Multidimensional e o total de variantes <i>de novo</i> de perda de função e <i>de novo</i> não sinônimas deletérias em Genes Não Essenciais. . . . .	40
Tabela 21 – Associação do Escore de Gravidade Multidimensional e o total de variantes em genes do módulo M01. . . . .	42
Tabela 22 – Associação do Escore de Gravidade Multidimensional e o total de variantes em genes do módulo M11. . . . .	42
Tabela 23 – Associação do Escore de Gravidade Multidimensional e o total de variantes em genes do módulo M12. . . . .	43
Tabela 24 – Associação do Escore de Gravidade Multidimensional e o total de variantes em genes do módulo M13. . . . .	43
Tabela 25 – Associação do Escore de Gravidade Multidimensional e o total de variantes em genes do módulo M31. . . . .	43

# **Lista de abreviaturas e siglas**

ADI-R	<i>Autism Diagnostic Interview-Revised</i>
ASDBC	<i>Autism Spectrum Disorder Behaviour Checklist</i>
CBCL	<i>Child Behavior Checklist</i>
CARS	<i>Childhood Autism Rating Scales</i>
CNVs	<i>Copy Number Variations</i>
dnNSD	<i>de novo Nonsense Damaging</i>
dnLoF	<i>de novo Loss of Function</i>
EGs	<i>Essential genes</i>
GWAS	<i>Genome-Wide Association Studies</i>
LGD	<i>Likely Gene-disruptive</i>
DSM-5	Manual de Diagnóstico e Estatístico de Transtornos Mentais
MAF	<i>Minor Allele Frequency</i>
NEGs	<i>Non Essential genes</i>
PCA	<i>Principal Component Analysis</i>
PC	<i>Principal Component</i>
QI	Quociente de Inteligência
SCE	Sequenciamento Completo do Exoma
SFARI	<i>Simons Foundation Autism Research Initiative</i>
SSC	<i>Simons Simplex Collection</i>
SNPs	<i>Single Nucleotide Polymorphism</i>
SNVs	<i>Single Nucleotide Variants</i>
TEA	Transtorno do Espectro do Autismo
VABS	<i>Vineland Adaptive Behavior Scales</i>

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>3</b>
1.1	Fatores genéticos relacionados ao TEA	3
1.2	Relações Genética-Fenótipo no Transtorno do Espectro do Autismo	6
1.3	Heterogeneidade Fenotípica no TEA	7
1.4	Uso de modelos normativos para o estudo de transtornos complexos	9
1.5	Justificativa	10
<b>2</b>	<b>OBJETIVOS</b>	<b>12</b>
<b>3</b>	<b>MATERIAL E MÉTODOS</b>	<b>13</b>
3.1	Amostra	13
3.2	Instrumentos de avaliação comportamental nos pacientes	13
3.3	Análise de Componentes Principais	14
3.4	Modelagem Normativa	14
3.5	Construção do Escore de Gravidade Multidimensional	19
3.6	Análise de associações genéticas	21
<b>4</b>	<b>RESULTADOS</b>	<b>23</b>
4.1	Caracterização da amostra	23
4.2	Análise de componentes principais	25
4.3	Mapa de heterogeneidade de apresentação clínica	29
4.4	Escore de Gravidade Multidimensional	37
4.5	Associação entre escore de gravidade multidimensional e variantes genéticas	38
<b>5</b>	<b>DISCUSSÃO</b>	<b>45</b>
<b>6</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>49</b>
	<b>REFERÊNCIAS</b>	<b>50</b>



# 1 Introdução

O Transtorno do Espectro do Autismo (TEA) é um transtorno do neurodesenvolvimento caracterizado por início precoce e por alterações na sociabilização, assim como presença de padrões restritivos e repetitivos de comportamento, de interesses ou de atividades, conforme definição no Manual de Diagnóstico e Estatístico de Transtornos Mentais (DSM-5) (1). Estima-se que a prevalência de TEA em crianças de 8 anos de idade seja de aproximadamente 1 a cada 59, com uma menina diagnosticada para cada 4 meninos (2), mostrando um desequilíbrio entre sexos na prevalência do transtorno (3). A apresentação fenotípica do TEA possui uma grande heterogeneidade, de modo que raramente duas crianças afetadas apresentam o mesmo nível de alteração nos sintomas centrais de diagnóstico definidos pelo DSM-5, assim como em outras esferas de funcionalidade (4). Desse modo, o DSM-5 adota uma visão dimensional em que o TEA é tido como um *continuum* de gravidade (1).

A etiologia do transtorno ainda não está completamente esclarecida, mas a hipótese mais aceita é que tanto fatores ambientais quanto genéticos contribuem para o risco do TEA (5, 6, 7). Entre os fatores ambientais associados, podemos citar a idade mais avançada dos pais (8), condições nutricionais (9) e infecções maternas (10) pré-natais, assim como a exposição à substâncias tóxicas (11). Os fatores genéticos também possuem um papel importante, sendo que o estudo mais recente avaliando o papel desses fatores, estimou a herdabilidade genética em cerca de 87% para gêmeos ( $n = 37.570$  pares) e de 69% para os casos gerais ( $n = 1,6$  milhões de famílias) (12).

## 1.1 Fatores genéticos relacionados ao TEA

Há diferentes componentes genéticos que contribuem para o TEA. Entre eles estão as variantes raras estruturais do tipo *copy number variations* (CNVs), variantes muito raras de nucleotídeo único (*Single Nucleotide Variantion*, SNVs), que possuem frequência alélica mínima (*Minor Allele Frequency*, MAF) abaixo de 1% em populações humanas no geral, e as variantes de nucleotídeo únicas comuns (*Single Nucleotide Polymorphism*, SNPs) com MAF maior do que 1% (13), que possuem pequeno efeito individual mas quando somadas, contribuem para compor o risco poligênico do TEA (14).

As variantes comuns (SNPs) são alvos dos trabalhos de *Genome-Wide Association Studies* (GWAS). Os primeiros estudos com GWAS em autismo focavam em encontrar SNPs individuais que explicassem o risco para o transtorno, em desenhos experimentais do tipo caso-controle (15, 16, 17, 18), mas para encontrar SNPs que tenham efeito significante em doenças complexas, é necessário um enorme número amostral (muitas vezes acima de 10.000 indivíduos) (19) e mesmo nos casos em que esse requisito é atendido, os SNPs encontrados não são concordantes e nem replicáveis entre diferentes coortes populacionais (20, 21).

É provável que as variantes comuns individualmente possuam pouco efeito sob um determinado traço observado, mas que vários SNPs de pequeno efeito agindo em conjunto possam explicar melhor o risco para esse traço. Por consequência, as abordagens envolvendo GWAS começaram a sumarizar a contribuição de todos os SNPs para um determinado traço em um escore de risco poligênico, que permite avaliar o efeito de *locus* gênicos que de forma individual não atingem significância estatística (22). Um estudo conduzido por Klei *et al.* (23) demonstrou que esse paradigma é válido para TEA, de modo que variantes genéticas comuns agem de forma aditiva e provavelmente são responsáveis por uma grande porção do risco. Outro estudo evidenciou que o uso de escores de risco poligênico em GWAS é mais replicável do que a presença de SNPs de forma individual ao se analisar diferentes amostras da mesma etnia. (24, 21). Nesse estudo, os autores utilizaram mais de 18 mil crianças afetadas com TEA e mais de 27 mil casos controles provenientes de importantes coortes psiquiátricas como a iPSYCH, Psychiatric Genomic Consortium (PGC), Finnish autism case-control study e deCODE. Por meio dessas abordagens, foi possível estimar que a contribuição de diferentes SNPs para a herdabilidade do TEA varia entre 17% (25, 26) a 52% (27), conforme o número amostral e o método de cálculo utilizado em cada estudo.

Outro importante componente que contribui para a arquitetura genética do TEA são as variantes raras. As variantes raras (*Copy Number Variations* e *Single Nucleotide Variations*, herdadas ou *de novo*) são geralmente alvos dos trabalhos de Sequenciamento Completo do Exoma (SCE). O primeiro estudo que associou a presença de CNVs não herdadas com o TEA, antes mesmo da popularização do sequenciamento completo do exoma, foi realizado por Sebat *et al.* (28). Neste trabalho, os autores executaram uma técnica de hibridização genômica comparativa em 118 casos de TEA esporádico (ou *simplex*, em que não há histórico familiar do transtorno) e 196 controles, e apontaram que CNVs *de novo* estão significativamente associadas ao TEA, sendo que a maioria dessas variantes estruturais são pequenas, afetando regiões genômicas altamente heterogêneas e muitas vezes afetando genes únicos. Esse trabalho estabeleceu as CNVs *de novo* como um fator de risco para TEA e abriu campo para o estudo do papel de variantes não herdadas no transtorno.

O’Roak *et al* (29) realizaram o primeiro trabalho que utilizou o SCE para o estudo de TEA. Ao sequenciarem o exoma de 20 casos de TEA simplex assim como o dos seus pais, os autores descreveram 21 variantes não herdadas (*de novo*) das quais metade potencialmente alteravam as proteínas dos genes atingidos, em resíduos altamente conservados. Esse trabalho mostrou o potencial do SCE para a identificação de novos genes candidatos para TEA, além de sugerir que as SNVs *de novo* contribuem consideravelmente para a etiologia genética do TEA. Posteriormente, os mesmos autores aumentaram a amostra para o SCE de 189 trios (pais-filho) de casos esporádicos, assim como 50 irmãos não afetados e encontraram cerca de 120 SNVs não herdadas e de alto impacto nos pacientes afetados (30). Essas variantes afetam genes que participam de redes de interação de proteínas para remodelação da cromatina, além de potencialmente causarem o truncamento de proteínas como CHD8, NTNG1, mGRIN2B, LAMC3 e CSN1A, genes candidatos associados ao TEA.

Na mesma época, outros estudos começaram a avaliar a importância de variantes raras *de novo* para o transtorno. Neale (31) realizaram o SCE de 175 trios e identificaram que cerca de metade dos casos analisados possuíam variantes *de novo* do tipo *missense* e sem sentido, principalmente em genes que codificam para proteínas que participam de redes de interação proteína-proteína. Além disso, as variantes encontradas afetam genes relacionados com funções sinápticas e genes previamente associados ao TEA como CHD8 e KATNAL2. Deste modo, foi proposto um modelo poligênico em que a ocorrência de variantes disruptivas em um número elevado de genes aumentaria o risco de TEA em 5 a 20 vezes. Iossifov *et al.* (32), com o uso de um maior número amostral ( $n = 343$  trios) também determinaram que variantes não herdadas do tipo likely gene-disruptive (LGD) são duas vezes mais frequentes em crianças afetadas do que em crianças não afetadas, e atingem cerca de 400 genes de susceptibilidade ao autismo. Muitos desses genes estão associados com a proteína da síndrome do X frágil (FMRP), evidenciando novamente a associação de genes com funções sinápticas e atingidos por variantes *de novo* com TEA.

Além do SCE, o surgimento de coortes estruturadas que unem dados genéticos e fenotípicos de pacientes com TEA assim como seus familiares (pais e irmãos não afetados), permitiram grande avanço nos estudos da área. Entre eles, podemos citar a *Simons Simplex Collection* (SSC) da *Simons Foundation Autism Research Initiative* (SFARI) (33). Essa coorte populacional consiste em um repositório de dados de exoma e genoma completo de mais de 2.600 famílias com casos de autismo esporádico, sendo que cada amostra genética é acompanhada de uma caracterização precisa do indivíduo, incluindo dados fenotípicos e comportamentais. Um dos primeiros estudos a utilizar os dados genéticos da SSC demonstrou que 13% de variantes *de novo missense* e 43% de variantes *de novo* LGD explicam respectivamente 12% e 9% dos casos de TEA, e que a presença dessas variantes com CNVs não herdadas explicariam 30% de todos os casos simplex e 45% dos casos *simplex* nos quais os pacientes são meninas (34).

Em suma, o uso do SCE para o estudo do TEA e o surgimento de iniciativas que permitiram a estruturação de coortes populacionais com dados de exoma e fenótipos, permitiram um grande avanço na elucidação do papel de variantes genéticas muito raras para o TEA. Atualmente, sabemos que as variantes *de novo* do tipo LGD e *missense* possuem um papel importante para a composição genética do TEA (34), atingindo genes associados com proteínas sinápticas (especialmente sinapses glutamatérgicas) e com reguladores da cromatina que controlam a expressão gênica durante o neurodesenvolvimento embrionário (35, 36), além de genes que codificam para proteínas e que agem em conjunto em redes de co-expressão e de interação proteína-proteína (37, 38). Além disso, as variantes *de novo* geralmente atingem genes intolerantes à mutações, ou seja, genes com significativamente menor taxa mutacional na população geral e que possivelmente estão sob forte restrição seletiva (39, 40).

As SNVs raras herdadas também são outro componente genético importante no TEA. Assim como as SNVs *de novo*, essas variantes são mais presentes em pacientes do que em seus irmãos não afetados, impactando de forma bialélica a porção codificadora de diversos genes intolerantes e genes alvos da proteína CHD8 (41, 42, 43). Ainda, essas variantes afetam elementos

*cis*-regulatórios de genes que contribuem para o TEA e não apenas a porção codificadora dos mesmos, além de serem preferencialmente transmitidas para crianças afetadas e não para seus irmãos (44). Embora diversos estudos tenham elucidado o papel de variantes raras herdadas e *de novo* para o risco do TEA e proposto centenas de genes importantes para o transtorno, ainda há uma lacuna em entender de forma sistemática os mecanismos por trás da biologia do autismo. Isso ocorre pois as variantes por si só são raramente identificadas de forma replicativa em diferentes coortes populacionais em estudos independentes, dificultando o reconhecimento de quais variantes ocorrem de forma aleatória e quais realmente possuem um papel de causalidade (45, 46).

Um estudo realizado por Luo *et al.* (47) buscou lidar com esse desafio ao analisar mais de 7 mil variantes anteriormente relacionadas ao TEA em dois estudos de exoma completo em larga escala. Ao analisarem esses dados em nível de SNVs, genes afetados e módulos moleculares aos quais os genes afetados estavam relacionados, os autores demonstraram que nas diferentes coortes populacionais, o impacto observado não se replicava a nível de variante mas sim a nível de gene e via biológica. Dessa forma, foi proposto um modelo gene + *pathway dual-hit*, no qual a carga mutacional de pacientes se torna menos importante e a análise das vias biológicas afetadas melhor definiriam a biologia do TEA.

Na mesma abordagem, o maior trabalho analisando o SCE de TEA até o momento utilizou mais de 11.000 indivíduos afetados e evidenciaram a existência de alterações funcionais e de desenvolvimento na neurobiologia do transtorno (48). Ao integrar informações de diferentes níveis biológicos, Satterstrom, F. Kyle, *et al.* indicaram 102 genes de importância para o TEA. Esses genes mostram alta frequência de variantes *de novo* disruptivas e são expressos no cérebro nos estágios iniciais do neurodesenvolvimento, tendo papel importante na regulação da expressão gênica e em mecanismos de comunicação neural. Além disso, a nível celular, a expressão dos genes de risco indicados é enriquecida para neurônios excitatórios (glutamatérgicos) e inibitórios (gabaérgicos), indicando a existência de distúrbios nas vias biológicas excitatórias-inibitórias no TEA.

## 1.2 Relações Genética-Fenótipo no Transtorno do Espectro do Autismo

No contexto das relações genética-fenótipo no TEA, ainda existem grandes avanços a serem realizados. Como citado anteriormente, diferentes classes de variantes contribuem para a arquitetura genética do TEA. No que diz respeito às influências no fenótipo observado, pacientes com deficiência intelectual (tidos como mais graves), dismorfismos, sem história familiar têm maior chance de apresentar variantes muito raras de grande efeito (49, 50) e meninas também possuem maior carga de variantes deletérias (40, 51).

O trabalho de Chang *et al.* (52), foi um dos primeiros a estabelecer as relações entre a carga mutacional de variantes raras e o QI de indivíduos com TEA. Os autores observaram que os pacientes com QI abaixo de 100 carregam o dobro de SNVs *de novo* do tipo LGD em

relação aos pacientes classificados como TEA de alto funcionamento (QI acima de 100), além de possuírem uma menor taxa de SNVs não sinônimas. Além disso, foi demonstrado que os pacientes com TEA de alto funcionamento possuem um número médio de SNVs não sinônimas similar ao número médio dessas variantes nos irmãos não afetados, já os pacientes com deficiência intelectual (QI abaixo de 70) possuem uma taxa muito menor de SNVs não sinônimas em relação aos pacientes com QI maior do que 70. Esses achados também foram verificados no trabalho de Iossifov *et al.* (34) que constataram que a presença de variantes *de novo* do tipo LGD em cerca de 400 genes afetam mais os pacientes com QI mais baixo do que os pacientes com o QI mais elevado. Assim sendo, foi estabelecida uma importante relação entre SNVs *de novo* e o QI de indivíduos com TEA.

O estudo de Buja *et al.* (53) reanalisou a lista de variantes *de novo* encontradas por Iossifov *et al.* (34) em 2.500 pacientes da coorte Simon Simplex Collection. Nessa reanálise, os pesquisadores conseguiram correlacionar uma maior taxa de SNVs *de novo* de alto impacto com um déficit no controle motor, sendo que tanto esse déficit quanto o QI estão associados de modo independente com a carga mutacional. Eles também evidenciaram que essa nova correlação varia conforme o tipo da variante (LGD ou *missense*), o set de genes analisado (alvos da FMRP, reguladores da cromatina, alvos da CHD8) e o instrumento escolhido para medir o traço comportamental (Developmental Coordination Disorder Questionnaire ou Vineland Adaptive Behavior Scales).

Além de demonstrar a relação entre o número SNVs raras deletérias e QI, Chang *et al.* também buscaram analisar como a expressão gênica e as propriedades funcionais de genes associados ao risco de desenvolver TEA poderiam influenciar no fenótipo. Ao comparar indivíduos TEA com pontuações abaixo e acima da mediana nas escalas de QI e nos sub escores de socialização e comportamentos restritos e repetitivos da ADI-R os autores verificaram que os genes afetados associados com um QI mais baixo ou escores da ADI-R mais altas (que indicam fenótipos mais graves) possuem uma alta expressão no cérebro (52).

A investigação do papel de genes essenciais para o desenvolvimento embrionário em mamíferos (*Essential Genes*, EGs), no TEA também ajudou a esclarecer as relações entre variantes genéticas e traços comportamentais (54). Analisando o exoma de 2.500 pacientes da Simon Simplex Collection, os autores verificaram que uma maior quantidade de variantes raras herdadas e *de novo* em EGs está positivamente correlacionada com o escore da SRS em pacientes do sexo masculino, mas o mesmo não ocorre para as meninas. E novamente, foi demonstrada a relação entre o número de variantes prejudiciais e um menor QI verbal e não verbal em ambos os sexos.

## 1.3 Heterogeneidade Fenotípica no TEA

Conforme citado anteriormente, o TEA é um transtorno complexo que possui muitas facetas do ponto de vista genético, ambiental e clínico. Além dos sintomas centrais definidos pelo

DSM-5 como essenciais para o diagnóstico, outros fenótipos adjacentes como o sexo, a idade, quocientes de inteligência (QI) e nível da linguagem podem interferir na apresentação clínica principal, o que contribui para uma maior heterogeneidade de apresentação do quadro (55). Essa heterogeneidade comportamental traz dificuldades para o diagnóstico, prognóstico e resposta terapêutica tanto na prática clínica como em estudos científicos, e também no entendimento das bases biológicas que levam a determinados traços (56, 57).

Os estudos que buscam compreender a heterogeneidade clínica utilizam-se de duas abordagens principais. Uma abordagem é a busca de subgrupos de pacientes que compartilham as mesmas características, a fim de encontrar agrupamentos que refletem a heterogeneidade do conjunto de dados assim como identificar características comportamentais de relevância clínica que contribuem para a formação dos agrupamentos (57). Outra, é investigar as relações existentes entre as características clínicas por meio de metodologias como as de análise fatorial e outras técnicas de redução de dimensionalidade (58). Por meio de ambas abordagens, cada vez mais o TEA foi entendido como um gradiente de gravidade fenotípica (59, 60, 61).

Muitos trabalhos buscaram avaliar a contribuição de outras características modificadoras além do sexo, QI e nível de linguagem, para a variabilidade nos sintomas core de TEA. Bitsika *et al.* (2008) investigaram o papel da funcionalidade como possível modificador do espectro ao realizarem um estudo com 53 crianças acometidas com TEA em idade pré-escolar e escolar. A partir de dados da pontuação total de cada participante em escalas fenotípicas como o QI verbal e não verbal, o total e três subescalas da Vineland Adaptive Behavior Scales (VABS), a Childhood Autism Rating Scales (CARS), três sub escores da Autism Spectrum Disorder Behaviour Checklist (ASDBC) e a idade, foi possível a identificação de subgrupos homogêneos. O estudo reconheceu a presença de um 'espectro do autismo' e o importante papel dos vários níveis de funcionamento no *continuum* de gravidade do TEA (62). Na mesma época, Munson *et al.* (2008) realizou análise de classes latentes baseando-se nas pontuações obtidas por 100 adolescentes entre 14 e 16 anos com TEA em uma escala de QI e na VABS. Quatro classes latentes foram identificadas, que representam níveis de funcionamento intelectual diferentes, bem como diferentes padrões de habilidade verbal e não-verbal. Os grupos diferiam em seu comportamento geral e nas habilidades adaptativas (63). Hoje, as habilidades adaptativas e o funcionamento executivo já são reconhecidos como importantes fatores que contribuem para heterogeneidade da apresentação clínica (64).

Outra característica importante e comumente observada na prática clínica são os problemas comportamentais e emocionais, entendidos como comportamentos disruptivos. Os comportamentos disruptivos em crianças com TEA são apresentados por um quarto a um terço delas, incluindo explosões de raiva, irritabilidade, e comportamento opositor e agressivo (65). A presença destes sintomas pode manter os pacientes acometidos mais afastados das demandas sociais, podendo interferir na gravidade da apresentação clínica e portanto heterogeneidade da apresentação. Para analisar tal relação, um estudo examinou a sobreposição fenotípica entre as principais características diagnósticas e os problemas emocionais/comportamentais em uma amostra de 335 crianças em idade pré-escolar com TEA por meio de análise de componentes

principais. Os resultados sugeriram uma sobreposição fenotípica substancial entre as principais características do diagnóstico e os problemas emocionais/comportamentais, indicando que estes contribuem para a compreensão do fenótipo de TEA e evidenciando que a inclusão da análise de problemas emocionais/comportamentais nos procedimentos diagnósticos tradicionais pode ter um papel importante na caracterização clínica das crianças (66). Na mesma linha de hipótese, Hus e Lord (2012), usando técnicas de análise multivariada em uma amostra de 2.334 pacientes mostraram que uso de variáveis fenotípicas como as que refletem problemas de comportamento, além da idade, nível de linguagem expressiva e QI podem afetar os escores dos algoritmos de diagnóstico tradicionais (67). Os resultados do estudo de Pandolfi *et al.*, 2012 também apoiam o uso da Child Behavior Checklist (CBCL) em conjunto com outros dados clínicos ao avaliar desordens emocionais e comportamentais em jovens com TEA (66).

Diferentes estudos na área de fenótipos do TEA até o momento, apoiam a existência de subtipos que provavelmente se organizam dentro de um gradiente de gravidade e confirmam a importância da funcionalidade e alterações do comportamento, além dos sintomas core do TEA e medidas de QI, para explicar a heterogeneidade fenotípica (68, 69, 70). Mas é importante notar que os estudos até aqui discutidos, não têm como foco um paciente individual na compreensão da heterogeneidade fenotípica do TEA.

## 1.4 Uso de modelos normativos para o estudo de transtornos complexos

Recentemente, pesquisadores têm utilizado um método conhecido como “modelagem normativa” para caracterizar a heterogeneidade dos transtornos psiquiátricos (71). De modo análogo aos gráficos de crescimento infantil amplamente utilizados em medicina pediátrica, a modelagem normativa permite que variáveis de interesse sejam medidas em um indivíduo e comparadas com faixas de variação esperadas em uma população referência. Esse tipo de abordagem leva em conta a individualidade do paciente ao estratificar coortes clínicas, dando poder para a realização de inferências estatísticas a nível do indivíduo em relação à uma variação normativa da população, uma vez que possibilita a quantificação e a caracterização de como cada indivíduo difere de um padrão esperado dentro da amostra analisada (71, 72).

Em termos práticos, a abordagem envolve a adoção de um modelo estatístico capaz de modelar a variância de uma variável resposta de interesse, de acordo com um conjunto de covariáveis clinicamente relevantes em toda uma coorte de referência. Por exemplo, pode-se estimar um modelo normativo para o grau de gravidade de um indivíduo em função da idade, sexo e medidas neuropsiquiátricas usando uma coorte de referência baseada na população afetada. A partir da estimativa do modelo normativo, é possível especificar uma função geral para o espaço amostral que quantifica quais relações entre a variável de interesse em função das covariáveis são mais ou menos prováveis (71).

Podemos então quantificar o grau em que outros indivíduos que não fazem parte da população normativa original se desviam das faixas centrais populacionais previstas para as

variáveis de interesse. Ou seja, o desvio refere-se ao grau em que os indivíduos em um conjunto ou amostra diferem entre si ao longo de uma ou mais características (71, 72). Portanto, não é mais necessário escolher apenas uma medida fenotípica específica para refletir os traços comportamentais dos pacientes em análise, uma vez que é possível considerar e quantificar as relações entre diferentes instrumentos como uma representação desses traços.

A modelagem normativa vem sendo utilizada para mapear padrões de variação entre traços de esquizofrenia (73, 74), transtorno de déficit de atenção e hiperatividade (75, 76) e TEA (77, 78). A maioria desses estudos, incluindo os trabalhos envolvendo autismo, buscaram utilizar técnicas estatísticas como regressão polinomial e linear para modelar a espessura cortical, o volume cerebral e sinais encefálicos medidos por eletroencefalograma e ressonância magnética funcional como biomarcadores para os traços comportamentais dos transtornos em questão.

Uma vez que a modelagem normativa se apresenta como um método estatístico flexível para entender a heterogeneidade de transtornos complexos como função de covariáveis preditoras, é possível utilizar essa abordagem para o entendimento do transtorno do espectro do autismo. Os trabalhos que utilizaram a abordagem de modelagem normativa para melhor caracterizar transtornos psiquiátricos utilizaram indivíduos saudáveis para criação de faixas normativas populacionais. No entanto, podemos criar um mapa de faixas normativas para compreender a heterogeneidade de uma apresentação clínica específica, sem considerar o que é esperado na população saudável. Nesse contexto, o paciente pode ser alocado no mapa normativo clínico, nos dando uma ideia da sua posição e facilitando não só a comparação entre pacientes como também as mudanças individuais ao longo do tempo.

## 1.5 Justificativa

A literatura tem evidenciado a associação entre variantes genéticas comuns e muito raras à susceptibilidade do TEA. No que se refere às variantes *de novo* e muito raras, aquelas que levam à perda de função em genes expressos no começo da gestação tem um papel mais importante no risco de desenvolver TEA (48, 52). A quantidade destas variantes também correlaciona-se com QI e comportamentos motores alterados, dando um indício de maior gravidade da apresentação clínica (53, 52). Entretanto, ainda há uma lacuna na literatura para entender a influência destas variantes na heterogeneidade fenotípica do autismo, considerando os seus sintomas principais e os modificadores como problemas de linguagem, problemas comportamentais e de funcionalidade (79). Isso pode ocorrer pois muitos estudos buscam associações entre o número total de variantes de diferentes classes e instrumentos de avaliação fenotípica específicos, sem necessariamente levar em conta que a melhor representação do TEA é um *continuum* de gravidade influenciado por diferentes fatores que agem em conjunto, ou seja, uma visão multidimensional.

Nossa hipótese é de que o número de variantes de alto impacto estariam associadas com um escore de gravidade dimensional e multivariado. Neste trabalho, propomos uma abordagem combinada de avaliação de medidas de variabilidade dos fenótipos que contribuem para a hete-

rogeneidade da apresentação clínica no TEA e de modelagem normativa. No lugar de comparar diferentes instrumentos de avaliação utilizando as suas pontuações brutas, média e desvio padrão, buscamos componentes principais a partir de medidas que contribuem para heterogeneidade fenotípica: sintomas core de TEA, QI, funcionalidade e alterações comportamentais. Ao aplicar a modelagem normativa nos componentes principais encontrados, buscamos propor um escore de gravidade dependente da heterogeneidade fenotípica. O desenvolvimento deste escore foi realizado em conjunto com a aluna de mestrado e neuropsicóloga especialista em autismo, Joana Portolese como parte de sua tese de mestrado, e sob orientação do Professor Dr. Ricardo Vêncio.

Consideramos também que variantes em genes importantes para a vulnerabilidade ao transtorno não são necessariamente os mesmos associadas ao espectro da heterogeneidade. Assim, nossa segunda hipótese é que não apenas variantes com perda de função em genes essenciais ou expressos no começo da gestação estão associados com o escore de gravidade.

## 2 Objetivos

### **Objetivo Principal:**

Testar a associação entre o escore de gravidade multidimensional com o número de variantes muito raras deletérias em indivíduos acometidos com TEA

### **Objetivos Específicos:**

Explorar as relações existentes entre QI, funcionalidade, comportamentos alterados e os sintomas principais de TEA a fim de identificar um constructo fenotípico que refletia a heterogeneidade de apresentação clínica.

Utilizar a abordagem de modelagem normativa para criar um escore de gravidade multidimensional.

Averiguar se há alguma relação entre o escore de gravidade e o número de variantes muito raras deletérias considerando a essencialidade do gene acometido e o tempo de expressão cerebral intra-útero.

### 3 Material e Métodos

#### 3.1 Amostra

Utilizamos os dados fenotípicos e genéticos de 2.744 pacientes diagnosticados com TEA, provenientes da coorte Simons Simplex Collection (SSC; versão 15.0), da base de dados Simons Foundation Autism Research Initiative (SFARI base). Como instrumentos de avaliação fenotípicas, utilizamos dados das escalas Autism Diagnostic Interview-Revised (ADI-R), Vineland Adaptive Behavior Scales (VABS), Child Behavior Checklist (CBCL), e Quociente de Inteligência (QI).

#### 3.2 Instrumentos de avaliação comportamental nos pacientes

A Autism Diagnostic Interview-Revised (ADI-R) é um instrumento amplamente utilizado para o diagnóstico de TEA. É uma entrevista aplicada aos pais e cuidadores de indivíduos possivelmente afetados e o objetivo é medir os principais sintomas de TEA (80). De acordo com os comportamentos descritos pelos responsáveis dos indivíduos, uma pontuação é gerada, abrangendo os domínios de Comunicação, Interação Social e Padrões Restritos e Repetitivos de Comportamento. O indivíduo é diagnosticado com TEA se atinge 10 pontos ou mais para o total, 10 pontos ou mais para o item Interação Social, 7/8 pontos ou mais para o item Comunicação, sendo 8 para os pacientes verbais e 7 para os não verbais e 3 pontos ou mais para o item Padrões Restritos e Repetitivos de Comportamento, de modo que quanto maior a pontuação, mais dificuldades o indivíduo apresenta em determinado aspecto comportamental.

A Vineland Adaptive Behavior Scales (VABS) é um instrumento utilizado para auxiliar no diagnóstico de deficiências intelectuais e de desenvolvimento em geral (81) O instrumento serve como avaliação do comportamento adaptativo, que diz respeito ao conjunto de habilidades necessárias para a realização de atividades de vida diárias como as de socialização, comunicação, hábitos de vida e auto-suficiência. A escala avalia o comportamento adaptativo geral e em 4 domínios principais: Comunicação, Habilidades de Vida Diária, Socialização e Habilidades Motoras. A VABS é respondida por pais e cuidadores dos indivíduos em avaliação e fornece escores entre 20 e 60 para a apresentação do comportamento adaptativo geral e escores referentes aos 4 domínios em um escala com média 100, de modo que quanto maior a pontuação, mais habilidades o indivíduo apresenta em cada aspecto avaliado. Embora esse instrumento não seja necessário para o diagnóstico de TEA, é amplamente recomendado que uma avaliação do comportamento adaptativo ocorra durante o processo diagnóstico a fim de auxiliar no diagnóstico final e no planejamento das intervenções futuras (82).

O perfil comportamental foi obtido a partir do Child Behavior Checklist (83). Esse instrumento é amplamente utilizado para avaliar problemas emocionais e comportamentais

em crianças no geral. A CBCL é também respondida por pais e cuidadores das crianças em avaliação e gera escores em oito categorias: Comportamento Agressivo, Ansiedade/Depressão, Problemas Atencionais, Comportamento de quebra de regras, Queixas Somáticas, Problemas Sociais, Problemas de Pensamento e problemas de Retirada/Depressão. Há também duas escalas abrangentes que combinam essas categorias: Problemas Internalizantes e Problemas Externalizantes. A pontuação bruta do instrumento é transformada em t-scores e gera valores até 100 pontos sendo que cada categoria comportamental possui o seu valor de corte para se considerar na faixa clínica e quanto maior a pontuação, maior é o agravamento do indivíduo nos aspectos analisados. Quanto aos problemas de comportamento internalizantes/externalizantes o ponto de corte é de 64.

### 3.3 Análise de Componentes Principais

A fim de identificar as principais fontes de variabilidade no conjunto de dados analisados e de explorar as relações entre QI, funcionalidade, comportamentos alterados e os sintomas principais de TEA, realizamos uma Análise de Componentes Principais (do inglês, Principal Component Analysis - PCA). A Análise de Componentes Principais é uma técnica estatística multivariada de redução de dimensionalidade que busca encontrar Componentes Principais não correlacionados que explicam a maior parte da variabilidade dos dados originais (84). Dessa forma, é possível identificar componentes que melhor refletem a variabilidade fenotípica dos pacientes, assim como identificar qual a contribuição de diferentes instrumentos de avaliação comportamental para a formação desses componentes. Realizamos a PCA utilizando como variáveis de entrada a VABS em seus domínios de Socialização, Comunicação e Habilidades de Vida Diárias, a ADI-R em seus domínios de Socialização, Comunicação e Comportamentos Repetitivos e Estereotipados, o QI total e os valores t-score da CBCL Internalizante e CBCL Externalizante. Foram selecionados os 3 componentes principais que explicam 70% da variabilidade da amostra. As variáveis com correlação maior ou igual a 0,5 com cada componente principal foram consideradas como as de maior contribuição para o componente específico (85). A análise foi realizada utilizando a função prcomp da linguagem R (versão 4.0.3) (86).

### 3.4 Modelagem Normativa

A Figura 1 mostra uma visão geral do método. A partir das coordenadas dos indivíduos em três componentes principais, definimos um mapa de coordenadas tridimensional que possui relação com os instrumentos de avaliação fenotípica e que reflete a heterogeneidade da apresentação clínica dos pacientes. Uma vez que os valores de coordenadas de todos os pacientes nos três componentes principais (normalizadas para Z-Score) possuem uma distribuição normal, utilizamos a abordagem de modelagem gaussiana para estimar o modelo normativo que descreve as relações de variação fenotípica nos pacientes.

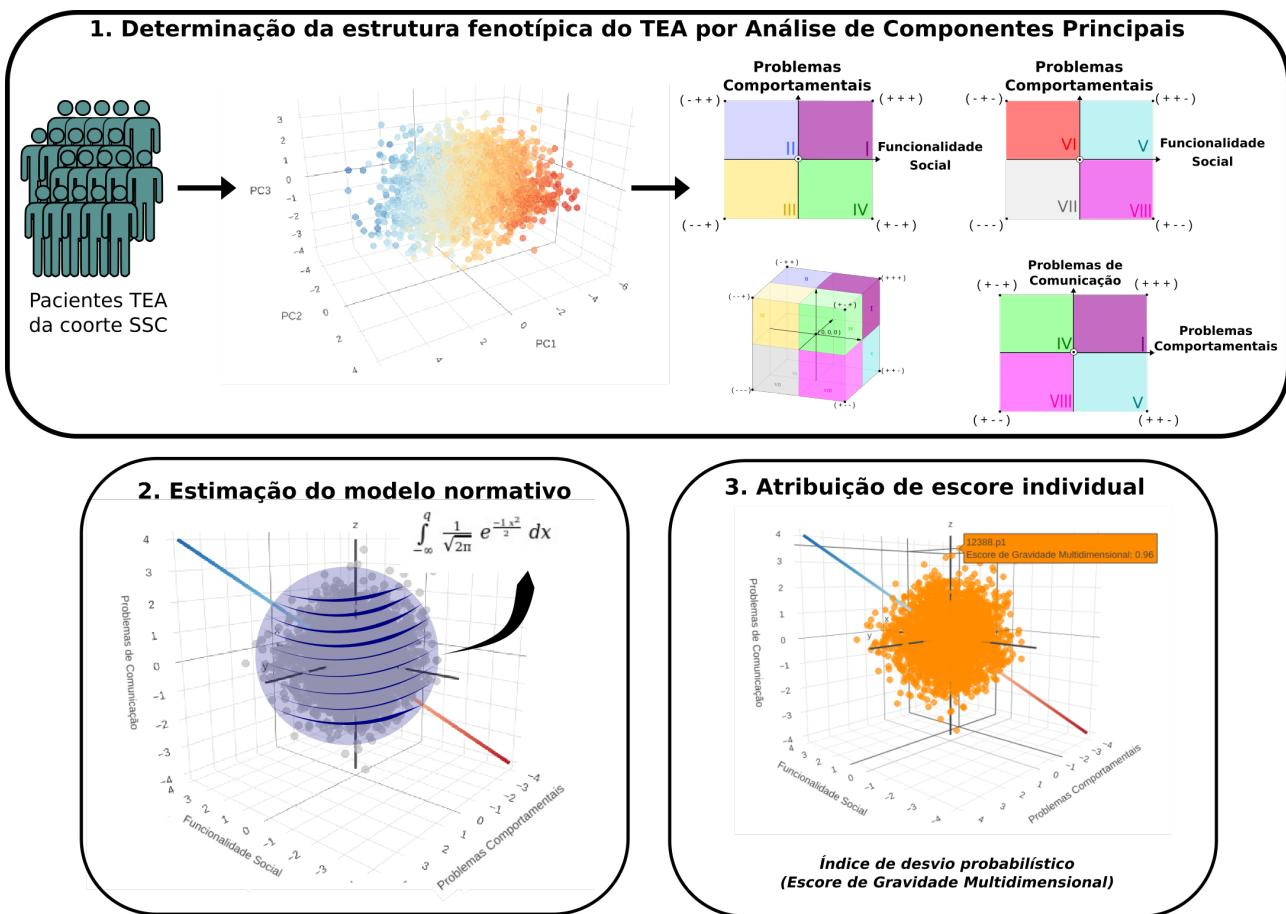


Figura 1 – Resumo do método - Versões detalhadas dos gráficos e esquemas estão apresentados na seção de Resultados.

Por meio desta abordagem são calculadas distribuições de probabilidade para cada região no espaço de coordenadas dos componentes principais, especificando dessa forma uma função geral que descreve as relações fenotípicas na amostra e que reflete quais são as relações fenotípicas mais ou menos prováveis (observadas), de acordo com a função de distribuição de probabilidades:

$$f(x) = (2\pi)^{\frac{-k}{2}} \det(\Sigma)^{\frac{-1}{2}} e^{\frac{-1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

(1)

Em que  $k$  representa o número de dimensões em que a população está distribuída (nesse caso,  $k = 3$ ),  $x$  é um vetor de coordenadas de um ponto no espaço (coordenadas de cada paciente em três componentes principais),  $\mu$  é um vetor de médias da população na qual o ponto  $x$  pertence, para cada dimensão, e  $\Sigma$  é a matriz de covariância entre as dimensões analisadas. Logo, é possível atribuir um valor de probabilidade para uma faixa de variação que vai de uma região de interesse (por exemplo,  $-\infty$  até o vetor de coordenadas  $x$  no espaço  $k$ -dimensional em questão (um paciente no mapa de heterogeneidade da apresentação clínica definido), desde que conhecido os valores de média ( $\mu$ ) e variância (determinante da matriz de covariância  $\Sigma$  entre as

dimensões) da amostra utilizada nas k dimensões. Para o caso bivariado (k = 2 dimensões) a distribuição de probabilidades o espaço de coordenadas dos componentes principais poderia ser visualmente representada por:

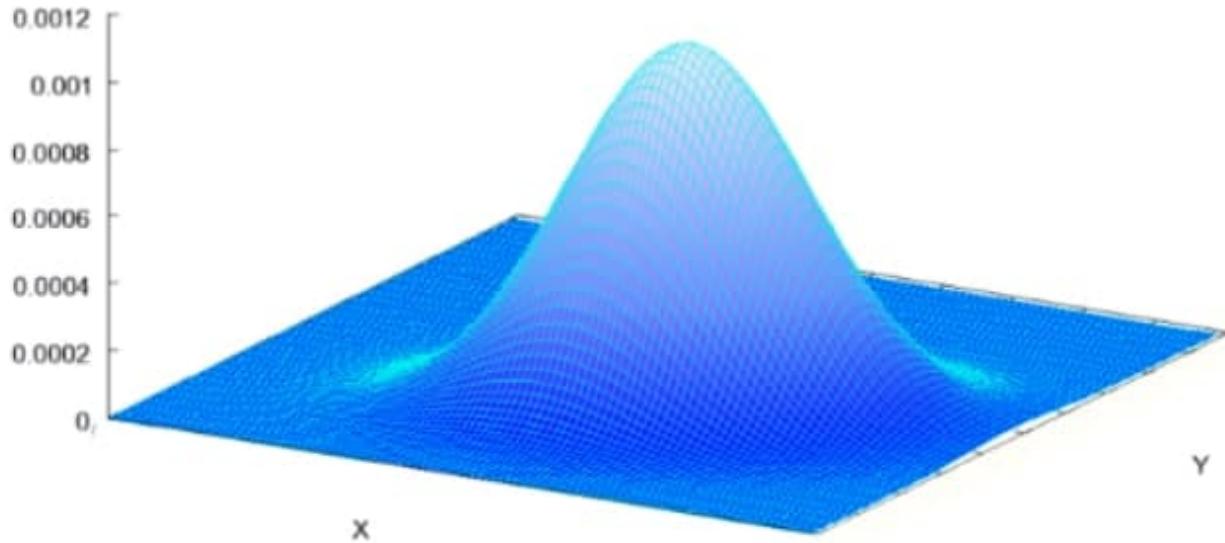


Figura 2 – Representação gráfica de uma distribuição normal multivariada, em duas dimensões  
Retirado de CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=1260349>

Nesse caso, os eixos x e y são as dimensões em que mapeamos as probabilidades para cada região de coordenadas em que um paciente se encontra e a altura da curva de densidade representa a probabilidade de ocorrência de um dado intervalo de coordenadas. No nosso caso, estamos lidando com distribuições de probabilidade no espaço tri-variado (k = 3 dimensões, três primeiros componentes principais) e como as coordenadas dos pacientes da coorte clínica de referência nos três primeiros componentes principais foram normalizados para Z-Score, a distribuição é parametrizada por  $\mu = 0$  e  $\sigma^2 = 1$ . Assim, a probabilidade de ocorrência de relações fenotípicas inferiores àquela observada em um determinado paciente (de acordo com suas coordenadas nos componentes principais) é dada por:

$$f(x) = (2\pi)^{-\frac{3}{2}} \det(\Sigma)^{\frac{-1}{2}} e^{\frac{-1}{2}(x-0)^T \Sigma^{-1} (x-0)}$$

(1.1)

$$f(x) = (2\pi)^{-\frac{3}{2}} \det(\Sigma)^{\frac{-1}{2}} e^{\frac{-1}{2}(x)^T \Sigma^{-1} (x)}$$

(1.2)

Após normalizado para Z-Score, a matriz de covariância entre as coordenadas das observações nos três componentes principais é dada por:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Portanto, a determinante da matriz assume o valor 1:

$$f(x) = (2\pi)^{-\frac{3}{2}} 1^{\frac{-1}{2}} e^{\frac{-1}{2}(x)^T \Sigma^{-1}(x)}$$

(1.3)

$$f(x) = (2\pi)^{-\frac{3}{2}} e^{\frac{-1}{2}(x)^T \Sigma^{-1}(x)}$$

(1.4)

A variável  $\mathbf{x}$  é um vetor de coordenadas para um ponto  $\mathbf{x}$  do espaço, dado por:  $\mathbf{x} = [x_1, x_2, x_3]$  em que  $x_1$  é a coordenada do paciente no primeiro componente principal,  $x_2$  é a coordenada do paciente no segundo componente principal e  $x_3$  é a coordenada do paciente no terceiro componente principal. Assim, o produto  $(x)^T \Sigma^{-1}$  se torna:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} x_{1.1} + x_{2.0} + x_{3.0} \\ x_{1.0} + x_{2.1} + x_{3.0} \\ x_{1.0} + x_{2.0} + x_{3.1} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

(1.5)

Por fim, a função densidade de probabilidade que atribui uma probabilidade para qualquer região em que o ponto  $\mathbf{x}$  se encontra, no espaço tridimensional formado pelas coordenadas da amostra nos três componentes principais analisados é dada por:

$$f(x) = (2\pi)^{-\frac{3}{2}} e^{\frac{-1}{2}(x)(x)}$$

(1.6)

$$f(x) = (2\pi)^{-\frac{3}{2}} e^{\frac{-1}{2}(x_1^2 + x_2^2 + x_3^2)}$$

(1.7)

Em que  $x_1$ ,  $x_2$  e  $x_3$  são as coordenadas do ponto  $\mathbf{x}$  nas dimensões 1, 2 e 3. A densidade de probabilidade para o espaço tridimensional com  $\mu = 0$  e  $\det(\Sigma) = 1$  é graficamente representada conforme mostrado na Figura 3 abaixo.

Nesse caso, os eixos x, y e z são as coordenadas em Z-score da coorte clínica de referência no primeiro, segundo e terceiro componente principal. Após atribuir uma distribuição de probabilidade para a região no espaço tridimensional, podemos determinar quais intervalos de relações fenotípicas são mais ou menos prováveis (observadas) quando comparadas ao restante do

observado na população ao calcular quantis de probabilidade para faixas específicas de variação, de acordo com a posição do paciente no mapa.

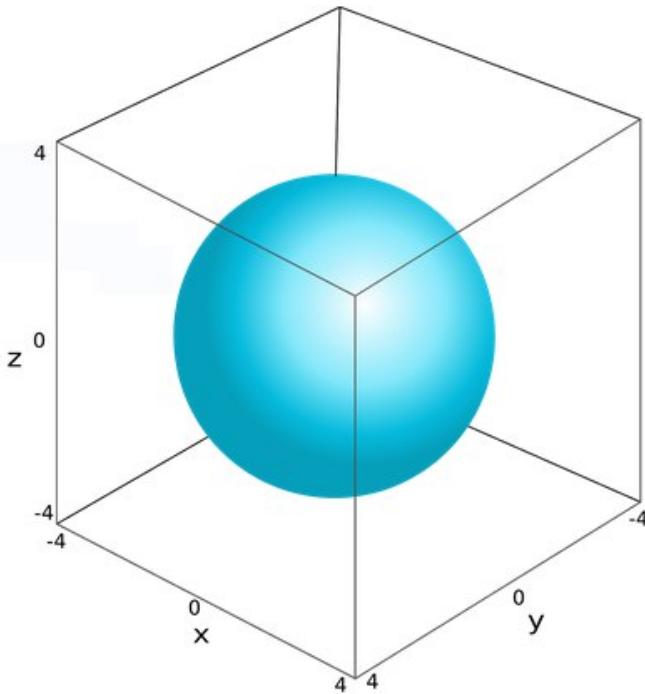


Figura 3 – Representação gráfica de uma distribuição normal multivariada, em três dimensões.

Para um determinado paciente representado no espaço tridimensional formado pelas coordenadas dos três componentes principais, a probabilidade de o seu perfil fenotípico estar acima de uma faixa de variação particular é dada pela probabilidade acumulada da área abaixo da posição no mapa em que o paciente se encontra, ou seja, a área abaixo da função densidade que descreve o perfil fenotípico daquele paciente. Para todas as 3 dimensões em que os pacientes estão distribuídos, o quantil em que um paciente  $\mathbf{x}$  se encontra é dado por:

$$(2) \quad \int_{-\infty}^q \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (2\pi)^{-\frac{3}{2}} e^{-\frac{1}{2}(x_1^2 + x_2^2 + x_3^2)} dx_1 dx_2 dx_3$$

$$(2.1) \quad \int_{-\infty}^q \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(x_1^2 + x_2^2 + x_3^2)} dx_1 dx_2 dx_3$$

$$(2.2) \quad \int_{-\infty}^q \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(x_1^2 + x_2^2 + x_3^2)} dx_1 dx_2 dx_3$$

$$\int_{-\infty}^q \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_1^2)} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_2^2)} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_3^2)} dx_1 dx_2 dx_3$$

(2.3)

A equação acima determina a probabilidade acumulada considerando as três direções de variação da distribuição gaussiana. Em nossa análise, utilizamos apenas uma direção para determinar o eixo central de variação que reflete a gravidade de apresentação clínica. Podemos então considerar que para determinar em qual quantil de probabilidade se encontra o perfil fenotípico observado em qualquer paciente, necessitamos da integral em apenas uma das direções, dada por:

$$p(q) = \int_{-\infty}^q \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

(2.4)

A equação nos retorna a probabilidade acumulada de todas as observações até o ponto q, em que q representa um Z-Score associado às coordenadas de um paciente x no espaço, para o qual desejamos determinar o grau de gravidade de apresentação clínica global.

### 3.5 Construção do Escore de Gravidade Multidimensional

É possível usar o espaço de coordenadas dos componentes principais normalizados para Z-Score para criar um mapa que nos permite compreender a heterogeneidade de apresentações clínicas específicas e que nos permite analisar as relações cognitivo-comportamentais e emocionais que ocorrem no indivíduo, com base em sua posição no mapa. Portanto, unimos a abordagem de representação da heterogeneidade fenotípica no espaço de componentes principais com o cálculo da densidade de probabilidade acumulada para a faixa de variação fenotípica em que cada paciente se encontra, sob o modelo normativo a fim de construir um mapa de probabilidade normativo que quantifica o grau de gravidade de apresentação clínica global de qualquer paciente individualmente em relação ao modelo normativo construído.

No mapa de heterogeneidade de apresentação clínica, determinamos um eixo central de variação de acordo com as relações entre a posição no mapa de coordenadas e os instrumentos de avaliação fenotípica. Esse eixo reflete o espectro de gravidade fenotípica global observado na amostra. Para determinar o escore de gravidade multidimensional, calculamos a distância entre as coordenadas do paciente nos três primeiros componentes principais e o eixo central de variação que reflete a gravidade. Essa distância é a projeção do paciente no eixo de variação fenotípica definido e é utilizada na variável q da Equação 2.4.

Nesse procedimento, a primeira etapa é considerar que há um hiperplano no espaço de coordenadas de componentes principais associado à cada paciente. Esse hiperplano é dado por um vetor normal N e por um ponto Q. O vetor N indica qual é a direção do hiperplano no espaço e o ponto Q indica as coordenadas no espaço do paciente ao qual o hiperplano está associado. A equação do hiperplano associado às coordenadas do paciente em uma determinada direção no espaço é dada por:

$$(3) \quad A(x - x_0) + B(y - y_0) + C(z - z_0) = 0$$

Em que x, y e z representam as coordenadas da extremidade final do espaço em que o paciente está contido e  $x_0$ ,  $y_0$  e  $z_0$  representam as coordenadas do paciente no espaço. O vetor N é dado pelos coeficientes A, B e C. A equação do plano também pode ser escrita como:

$$(3.1) \quad Ax + By + Cz + D = 0$$

Em que D é dado por:

$$(3.2) \quad D = -Ax_0 - By_0 - Cz_0$$

Os coeficientes A, B e C do vetor N são determinados pelas coordenadas de posição final e inicial do vetor no espaço. Quando consideramos o eixo central de variação que reflete a gravidade de apresentação clínica, os seus extremos finais x, y e z são dados pelos valores máximos observados nas coordenadas dos pacientes nos três componentes principais (vetor de coordenadas 4, 4, 4) e os seus extremos iniciais  $x_0$ ,  $y_0$  e  $z_0$  são dados valores mínimos observados nas coordenadas dos pacientes nos três componentes principais (vetor de coordenadas -4, -4, -4). Os extremos finais indicam uma região no mapa em que os escores brutos de todos os instrumentos de avaliação fenotípica refletem melhores apresentações clínicas de forma geral e os extremos iniciais indicam uma região no mapa em que os escores brutos nos instrumentos refletem piores apresentações clínicas de forma geral. Portanto, o vetor N do plano deve seguir a direção do eixo central de variação que reflete a gravidade fenotípica e as suas coordenadas serão dadas por:

$$(4) \quad N = (A, B, C); A = (x - x_0); B = (y - y_0); C = (z - z_0)$$

$$(4.1) \quad A = (4 - (-4)); B = (4 - (-4)); C = (4 - (-4))$$

$$(4.2) \quad N = (A, B, C) \rightarrow N = (8, 8, 8)$$

Após definir o plano que contém o paciente no espaço tridimensional para o qual desejamos determinar o grau de gravidade de apresentação clínica global, podemos projetar esse paciente

ao eixo central de variação que reflete a gravidade de apresentação clínica. Essa projeção é dada pela distância entre um ponto e um plano e nesse caso, o ponto é localizado no eixo central e o plano foi determinado de acordo com as coordenadas do paciente no espaço tridimensional. A distância é dada por:

$$q = d = \frac{D}{A^2 + B^2 + C^2}$$

(5)

Essa distância mapeia qualquer paciente no espaço de coordenadas de componentes principais ao eixo central de variação que reflete a gravidade de apresentação clínica e é utilizada na equação 2.4. Dessa forma, a probabilidade acumulada desde o início do eixo até a faixa de variação em que o paciente se encontra (sua posição em relação ao eixo central de gravidade) é o seu escore de gravidade multidimensional. Os cálculos de distâncias assim como o modelo normativo foram realizados utilizando scripts próprios implementados na linguagem de programação R (versão 4.0.3)

## 3.6 Análise de associações genéticas

Como conjunto de dados genéticos, utilizamos uma lista de variantes *de novo* e raras herdadas em genes codificadores de proteína, disponibilizada por Ji, X., Kember, R. L., Brown, C. D., Bućan, M. (2016) (54). O conjunto de dados consiste em uma lista de 79.730 variantes em 8.834 genes nos indivíduos afetados com autismo da coorte Simons Simplex Collection. As classes de variantes são divididas entre: (i) 372 variantes *de novo* de perda de função (SNVs *frameshift*, *stop-loss*, *stop-gain*, *start-loss*, doador ou receptor de splicing e *indels frameshift*); (ii) 1.497 variantes *de novo* não sinônimas deletérias (SNVs *missense* e *indels non-frameshift*) e (iii) 77.861 variantes raras herdadas deletérias. Para todas as classes de variantes foram mantidas apenas aquelas com CADD phred-scale maior do que 10 e a frequência do menor alelo (MAF) menor do que 0.01 no banco de dados Exome Variant Server (European ancestry), portanto, SNVs deletérias e muito raras.

Os 8.834 genes são classificados, de acordo com os autores, entre 3.915 Genes Essenciais e 4.919 Genes Não Essenciais. Os Genes Não Essenciais (do inglês, Non Essential genes - NEG) são genericamente classificados como ortólogos humanos de genes com fenótipos não letais relatados em camundongos. Os Genes Essenciais (do inglês, Essential genes - EG), são definidos como genes necessários para a conclusão bem-sucedida do desenvolvimento pré e pós-natal e apresentam intolerância para variantes do tipo perda de função devido à sua haploinsuficiência (87, 88). Esse grupo também é altamente enriquecido para genes associados à diversas doenças humanas e sofrem forte pressão evolutiva (89, 90).

Para testar a associação entre escore de gravidade multidimensional e variantes *de novo* e raras herdadas em genes codificadores de proteína nos indivíduos analisados, foram utilizados modelos de regressão beta por meio do pacote *betareg* no programa R (versão 4.0.3) (91). Como

variável dependente foi utilizado o escore de gravidade multidimensional para cada paciente e como variáveis independentes foram utilizados em modelos separados: o número total de variantes por indivíduos, o número total de variantes por indivíduos discriminando por classe de variante, o número total de variantes por indivíduos discriminando o grupo de genes afetado (EGs ou NEGs) e o número de variantes para cada classe e para cada grupo de gene afetado. Para todos os modelos, o sexo do indivíduo foi avaliado como variável de interação. As análises foram realizadas utilizando um subconjunto de 2.301 pacientes da coorte SSC que tinham as informações genéticas avaliadas no conjunto de dados analisados.

Buscamos também explorar possíveis associações entre o escore de gravidade multidimensional e genes expressos no cérebro que agem conjuntamente em vias biológicas nos processos de desenvolvimento ao longo do tempo. Para tanto, utilizamos os dados de grupos de genes disponibilizada por Ji, X., Kember, R. L., Brown, C. D., Bućan, M. (2016) Os autores analisaram dados de RNA-Seq de 15.952 genes codificadores de proteína em 16 tipos de tecido cerebral ao longo de 31 estágios de desenvolvimento pré e pós natal, avaliados em 495 indivíduos. Os autores então calcularam a co-expressão entre pares de genes e utilizaram a ferramenta WGCNA para identificar os módulos de co-expressão na amostra. A partir da lista dos genes que fazem parte de cada um dos 41 módulos de co-expressão identificados, buscamos testar associações entre o escore de gravidade multidimensional e variantes *de novo* e raras herdadas nos genes de cada módulo por meio de análises de regressão beta.

## 4 Resultados

### 4.1 Caracterização da amostra

Os indivíduos utilizados na análise são provenientes da coorte Simons Simplex Collection, da base de dados da Simons Foundation for Autism Research Initiative (SFARI, versão 15.0). Dos 2.857 pacientes disponíveis na base dados, 2.744 foram utilizados na Análise de Componentes Principais e na construção do escore de gravidade multidimensional e 2.301 foram utilizados para as análises de associação genética, de acordo com a disponibilidade dos dados.

A Tabela 1 mostra os valores de média e desvio padrão obtidos pelos pacientes em cada um dos instrumentos de avaliação fenotípicas utilizados. Todos os pacientes atingiram o valor de corte clínico para os subitens da ADI-R, mas não para os subitens da CBCL. A média dos valores de QI para essa amostra é relativamente elevada, considerando que muitos indivíduos acometidos com TEA também apresentam deficiência intelectual (92).

Tabela 1 – Médias e desvios padrão dos escores obtidos pelos pacientes com TEA provenientes da coorte SSC nos testes, escalas e questionários utilizados como entrada para a PCA.

Instrumento	Média (Desvio Padrão)
ADI-R Socialização	20.32 (5.70)
ADI-R Comunicação	15.95 (4.31)
ADI-R Comportamentos Restritos e Repetitivos	6.52 (2.49)
VABS Comunicação	77.12 (14.49)
VABS Socialização	70.98 (12.55)
VABS Hábitos de Vida Diários	76.45 (13.83)
CBCL Problemas Internalizantes	60.31 (9.55)
CBCL Problemas Externalizantes	56.57 (10.6)
QI total	81.23 (27.9)

A Tabela 2 e a Tabela 3 mostram, respectivamente, a prevalência das variantes genéticas do tipo *de novo* e do tipo rara herdada no subconjunto de 2.301 indivíduos utilizados para as análises de associação genética. É interessante notar que 90.48% dos pacientes analisados não apresentam nenhuma variante do tipo *de novo* de perda de função (média de 0.1 por indivíduo e desvio padrão de 0.32) e 69.5% dos pacientes analisados não apresentam nenhuma variante do tipo *de novo* não sinônima deletéria (média de 0.37 por indivíduo e desvio padrão de 0.61).

Tabela 2 – Números de variantes do tipo *de novo* nos 2.301 indivíduos utilizados nas análises de associação genética

Classe de variantes	No. de pacientes com 0 SNVs	No. de pacientes com 1 SNV	No. de pacientes com 2 SNVs	No. de pacientes com 3 SNVs	No. de pacientes com 4 SNVs
<i>de novo</i> de perda de função	2.082	207	12	0	0
<i>de novo</i> não sinônima deletéria	1.601	574	110	15	1

Tabela 3 – Números de variantes do tipo rara herdada os 2.301 indivíduos utilizados nas análises de associação genética.

	1º Quartil	Mediana	3º Quartil
Número de variantes raras herdadas	16	23	29

## 4.2 Análise de componentes principais

A fim de identificar e explorar as relações existentes entre QI, funcionalidade, comportamentos alterados e os sintomas principais de TEA por meio de uma visão multidimensional, realizamos uma Análise de Componentes Principais. Utilizamos como variáveis de entrada os escores obtidos por 2.744 indivíduos TEA nas escalas de QI, ADI-R nos subitens Comunicação, Socialização e Comportamentos Restritos e Repetitivos, VABS nos subitens de Comunicação, Socialização e Atividades de Vida Diária e CBCL nos domínios Internalizante e Externalizante.

A Figura 4 mostra as porcentagens de variância explicadas por cada componente principal. Foram obtidos 9 componentes principais sendo que os 3 primeiros componentes principais explicam 72.47% da variabilidade total dos indivíduos analisados.

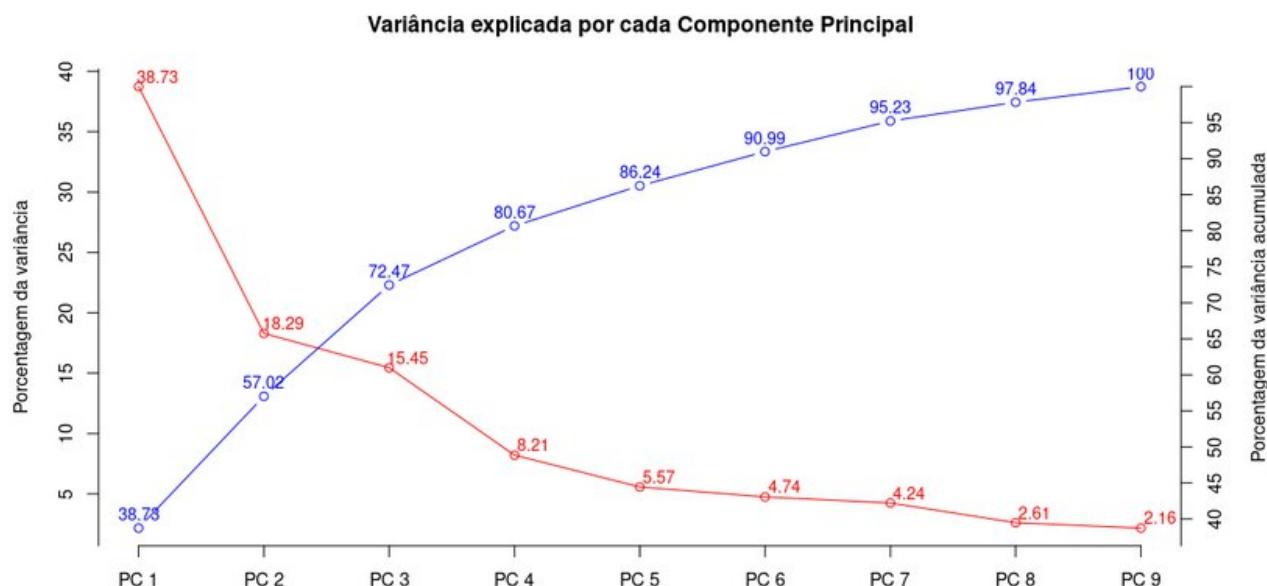


Figura 4 – Variância explicada por cada componente principal. - O eixo da esquerda e a linha em vermelho indicam as porcentagens de variância explicadas por cada um dos componentes principais. O eixo da direita e a linha em azul indicam a porcentagem da variância acumulada explicada pelo conjunto de componentes principais.

A Tabela 4 mostra as cargas das variáveis originais em cada componente principal obtido. A análise das contribuições das variáveis originais em cada componente principal mostra que: (i) a Componente 1 possui correlação maior ( $>0.65$ ) com os subitens da VABS, QI total e ADI-R Socialização; (ii) a Componente 2 possui correlação maior ( $>0.50$ ) com escores da CBCL Internalizante e Externalizante e com a ADI-R Comportamentos Restritos e Repetitivos; (iii) a Componente 3 possui correlação maior com a ADI-R Comunicação.

Tabela 4 – *Eigenvalues* e variância explicados por cada componente principal ao usar os subitens da VABS, ADI-R, QI total e CBCL internalizante e externalizante avaliados em 2.744 pacientes como medidas de entrada.

<b>Componente</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<i>Eigenvalue</i>	3.49	1.65	1.39	0.74	0.5	0.43	0.38	0.23	0.19
<b>Cargas das variáveis originais</b>									
VABS Comunicação	<b>0.89</b>	-0.16	0.17	-0.03	0.06	0	-0.09	0.16	-0.34
VABS Hábitos de vida diários	<b>0.85</b>	-0.13	0.19	-0.05	0	0.31	-0.05	-0.35	0.02
VABS Socialização	<b>0.87</b>	0.04	0.12	0.05	0.15	0.2	0.27	0.24	0.2
ADI-R Socialização	<b>-0.69</b>	-0.18	0.43	-0.28	-0.11	0.37	-0.24	0.15	0.03
ADI-R Comunicação	-0.31	-0.34	<b>0.73</b>	-0.32	0.21	-0.21	0.24	-0.08	-0.02
ADI-R Comportamentos Restritos e Repetitivos	-0.17	<b>-0.52</b>	0.45	0.7	-0.07	-0.02	-0.04	0.01	0.02
CBCL Problemas Externalizantes	-0.14	<b>-0.71</b>	-0.48	-0.03	0.47	0.03	-0.16	0	0.04
CBCL Problemas Internalizantes	-0.04	<b>-0.78</b>	-0.39	-0.15	-0.38	0.06	0.26	0.01	-0.04
QI total	<b>0.78</b>	-0.29	0.12	-0.2	-0.2	-0.32	-0.29	0.04	0.18

Para verificar se a estrutura das relações fenotípicas no TEA apontada pelos componentes principais obtidos poderia ser observada em uma amostra menor, a mesma análise foi realizada utilizando um subconjunto de pacientes. Ao utilizar 30% a 80% dos indivíduos da coorte, os resultados da PCA permanecem os mesmos (Tabela 5 - Tabela 10).

Tabela 5 – *Eigenvalues* e variância explicados por cada componente principal ao usar os subitens da VABS, ADI-R, QI total e CBCL internalizante e externalizante avaliados em 823 pacientes como medidas de entrada. - Na parte superior, estão em destaque os componentes principais com eigenvalue maior do que 1 e os componentes principais que em conjunto explicam cerca de 70% da variância da amostra. Na parte inferior, estão em destaque os valores de carga das variáveis originais em cada componente principal com valor superior a 0.5

<b>Componente</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<i>Eigenvalue</i>	<b>3.53</b>	<b>1.74</b>	<b>1.34</b>	0.76	0.48	0.37	0.37	0.22	0.19
Porcentagem da variância	39.21	19.32	14.85	8.4	5.36	4.15	4.10	2.46	2.15
Porcentagem da variância acumulada	<b>39.21</b>	<b>58.53</b>	<b>73.38</b>	81.78	87.14	91.29	95.39	97.85	100
<b>Cargas das variáveis originais</b>									
VABS Comunicação	<b>0.89</b>	-0.21	0.15	-0.03	0.05	-0.04	-0.02	0.18	-0.33
VABS Hábitos de vida diários	<b>0.87</b>	-0.15	0.16	-0.04	-0.06	-0.23	0.09	-0.36	-0.01
VABS Socialização	<b>0.88</b>	0.02	0.10	0.04	0.12	-0.03	0.35	0.19	0.22
ADI-R Socialização	<b>-0.69</b>	-0.27	0.41	-0.27	-0.15	-0.43	0.02	0.13	0.03
ADI-R Comunicação	-0.31	-0.43	<b>0.69</b>	-0.30	0.24	0.29	0.09	-0.08	-0.01
ADI-R Comportamentos Restritos e Repetitivos	-0.13	<b>-0.57</b>	0.37	0.72	-0.09	0.01	-0.05	0.01	0.03
CBCL Problemas Externalizantes	-0.16	<b>-0.66</b>	-0.54	0.00	0.47	-0.15	-0.05	-0.01	0.02
CBCL Problemas Internalizantes	-0.08	<b>-0.73</b>	-0.46	-0.16	-0.39	0.15	0.21	0.00	-0.02
QI total	<b>0.78</b>	-0.33	0.07	-0.22	-0.11	0.06	-0.42	0.06	0.18

Tabela 6 – *Eigenvalues* e variância explicados por cada componente principal ao usar os subitens da VABS, ADI-R, QI total e CBCL internalizante e externalizante avaliados em 1.098 pacientes como medidas de entrada. - Na parte superior, estão em destaque os componentes principais com eigenvalue maior do que 1 e os componentes principais que em conjunto explicam cerca de 70% da variância da amostra. Na parte inferior, estão em destaque os valores de carga das variáveis originais em cada componente principal com valor superior a 0.5

<b>Componente</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<i>Eigenvalue</i>	<b>3.45</b>	<b>1.68</b>	<b>1.38</b>	0.72	0.50	0.46	0.38	0.24	0.20
Porcentagem da variância	38.35	18.63	15.36	7.97	5.51	5.11	4.22	2.64	2.21
Porcentagem da variância acumulada	<b>38.35</b>	<b>56.98</b>	<b>72.34</b>	80.30	85.82	90.92	95.15	97.79	100.00
<b>Cargas das variáveis originais</b>									
VABS Comunicação	<b>-0.88</b>	0.18	-0.17	0.03	-0.05	-0.02	0.09	-0.20	0.33
VABS Hábitos de vida diários	<b>-0.84</b>	0.16	-0.21	0.00	-0.03	0.30	0.14	0.35	0.01
VABS Socialização	<b>-0.87</b>	-0.03	-0.12	-0.04	-0.19	0.24	-0.19	-0.22	-0.23
ADI-R Socialização	<b>0.70</b>	0.22	-0.42	0.21	0.11	0.30	0.34	-0.15	-0.05
ADI-R Comunicação	0.33	0.36	<b>-0.70</b>	0.38	-0.18	-0.12	-0.29	0.07	0.03
ADI-R Comportamentos Restritos e Repetitivos	0.20	<b>0.55</b>	-0.41	-0.69	-0.02	-0.09	0.01	-0.01	-0.03
CBCL Problemas Externalizantes	0.11	<b>0.68</b>	0.52	0.13	-0.44	-0.08	0.17	0.00	-0.04
CBCL Problemas Internalizantes	0.03	<b>0.77</b>	0.41	0.07	0.36	0.20	-0.24	-0.02	0.04
QI total	<b>-0.76</b>	0.28	-0.14	0.19	0.29	-0.39	0.17	-0.01	-0.17

Tabela 7 – *Eigenvalues* e variância explicados por cada componente principal ao usar os subitens da VABS, ADI-R, QI total e CBCL internalizante e externalizante avaliados em 1.372 pacientes como medidas de entrada. - Na parte superior, estão em destaque os componentes principais com eigenvalue maior do que 1 e os componentes principais que em conjunto explicam cerca de 70% da variância da amostra. Na parte inferior, estão em destaque os valores de carga das variáveis originais em cada componente principal com valor superior a 0.5

<b>Componente</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<i>Eigenvalue</i>	<b>3.46</b>	<b>1.65</b>	<b>1.38</b>	0.75	0.50	0.43	0.38	0.25	0.20
Porcentagem da variância	38.48	18.33	15.39	8.37	5.51	4.74	4.25	2.73	2.21
Porcentagem da variância acumulada	<b>38.48</b>	<b>56.80</b>	<b>72.19</b>	80.56	86.08	90.81	95.07	97.79	100.00
<b>Cargas das variáveis originais</b>									
VABS Comunicação	<b>-0.88</b>	0.17	-0.18	0.02	-0.06	0.04	-0.09	0.25	0.30
VABS Hábitos de vida diários	<b>-0.85</b>	0.14	-0.19	0.04	-0.01	-0.28	-0.15	-0.35	0.06
VABS Socialização	<b>-0.86</b>	-0.03	-0.12	-0.04	-0.20	-0.27	0.17	0.18	-0.25
ADI-R Socialização	<b>0.69</b>	0.15	-0.44	0.27	0.15	-0.28	-0.33	0.15	-0.06
ADI-R Comunicação	0.32	0.29	<b>-0.74</b>	0.33	-0.25	0.14	0.26	-0.08	0.03
ADI-R Comportamentos Restritos e Repetitivos	0.22	0.46	-0.47	-0.72	0.05	0.02	-0.02	0.00	-0.02
CBCL Problemas Externalizantes	0.16	<b>0.73</b>	0.46	0.04	-0.42	0.07	-0.22	0.00	-0.06
CBCL Problemas Internalizantes	0.10	<b>0.81</b>	0.33	0.12	0.30	-0.19	0.28	0.02	0.05
QI total	<b>-0.75</b>	0.32	-0.15	0.18	0.30	0.37	-0.13	0.00	-0.18

Tabela 8 – *Eigenvalues* e variância explicados por cada componente principal ao usar os subitens da VABS, ADI-R, QI total e CBCL internalizante e externalizante avaliados em 1.646 pacientes como medidas de entrada. - Na parte superior, estão em destaque os componentes principais com eigenvalue maior do que 1 e os componentes principais que em conjunto explicam cerca de 70% da variância da amostra. Na parte inferior, estão em destaque os valores de carga das variáveis originais em cada componente principal com valor superior a 0.5

<b>Componente</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<i>Eigenvalue</i>	<b>3.47</b>	<b>1.65</b>	<b>1.37</b>	0.75	0.50	0.44	0.39	0.24	0.19
Porcentagem da variância	38.56	18.36	15.18	8.36	5.60	4.87	4.30	2.66	2.11
Porcentagem da variância acumulada	<b>38.56</b>	<b>56.92</b>	<b>72.10</b>	80.47	86.06	90.93	95.23	97.89	100.00
<b>Cargas das variáveis originais</b>									
VABS Comunicação	<b>0.89</b>	-0.15	0.16	-0.04	0.07	0.00	0.10	0.19	0.33
VABS Hábitos de vida diários	<b>0.85</b>	-0.11	0.20	-0.06	-0.02	0.29	0.08	-0.36	0.00
VABS Socialização	<b>0.86</b>	0.03	0.12	0.04	0.14	0.26	-0.25	0.22	-0.21
ADI-R Socialização	<b>-0.68</b>	-0.18	0.43	-0.29	-0.15	0.34	0.28	0.14	-0.04
ADI-R Comunicação	-0.30	-0.33	<b>0.74</b>	-0.31	0.25	-0.19	-0.25	-0.07	0.02
ADI-R Comportamentos Restritos e Repetitivos	-0.13	<b>-0.53</b>	0.44	0.71	-0.09	-0.01	0.03	0.01	-0.02
CBCL Problemas Externalizantes	-0.13	<b>-0.72</b>	-0.47	-0.02	0.46	0.04	0.18	0.00	-0.05
CBCL Problemas Internalizantes	-0.02	<b>-0.78</b>	-0.37	-0.18	-0.37	0.07	-0.27	0.00	0.05
QI total	<b>0.78</b>	-0.26	0.12	-0.19	-0.20	-0.37	0.25	0.04	-0.18

Tabela 9 – *Eigenvalues* e variância explicados por cada componente principal ao usar os subitens da VABS, ADI-R, QI total e CBCL internalizante e externalizante avaliados em 1.912 pacientes como medidas de entrada. - Na parte superior, estão em destaque os componentes principais com eigenvalue maior do que 1 e os componentes principais que em conjunto explicam cerca de 70% da variância da amostra. Na parte inferior, estão em destaque os valores de carga das variáveis originais em cada componente principal com valor superior a 0.5

<b>Componente</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<i>Eigenvalue</i>	<b>3.54</b>	<b>1.63</b>	<b>1.35</b>	0.76	0.51	0.41	0.38	0.22	0.19
Porcentagem da variância	39.34	18.14	15.02	8.47	5.63	4.57	4.22	2.49	2.12
Porcentagem da variância acumulada	<b>39.34</b>	<b>57.48</b>	<b>72.50</b>	80.97	86.60	91.16	95.39	97.88	100.00
<b>Cargas das variáveis originais</b>									
VABS Comunicação	<b>0.89</b>	-0.16	0.19	-0.02	0.05	0.00	-0.09	-0.13	0.35
VABS Hábitos de vida diários	<b>0.86</b>	-0.13	0.18	-0.05	0.04	0.30	-0.01	0.34	-0.03
VABS Socialização	<b>0.87</b>	0.04	0.12	0.05	0.18	0.16	0.28	-0.25	-0.18
ADI-R Socialização	<b>-0.70</b>	-0.16	0.40	-0.31	-0.06	0.40	-0.20	-0.14	-0.02
ADI-R Comunicação	-0.34	-0.30	<b>0.74</b>	-0.30	0.19	-0.25	0.23	0.08	0.01
ADI-R Comportamentos Restritos e Repetitivos	-0.20	-0.48	0.47	0.71	-0.11	0.02	-0.04	-0.01	-0.03
CBCL Problemas Externalizantes	-0.14	<b>-0.74</b>	-0.42	0.01	0.48	0.00	-0.16	0.00	-0.04
CBCL Problemas Internalizantes	-0.04	<b>-0.78</b>	-0.37	-0.16	-0.37	0.06	0.28	-0.01	0.04
QI total	<b>0.77</b>	-0.29	0.16	-0.21	-0.23	-0.25	-0.32	-0.05	-0.18

Tabela 10 – *Eigenvalues* e variância explicados por cada componente principal ao usar os subitens da VABS, ADI-R, QI total e CBCL internalizante e externalizante avaliados em 2.195 pacientes como medidas de entrada. - Na parte superior, estão em destaque os componentes principais com eigenvalue maior do que 1 e os componentes principais que em conjunto explicam cerca de 70% da variância da amostra. Na parte inferior, estão em destaque os valores de carga das variáveis originais em cada componente principal com valor superior a 0.5

<b>Componente</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<i>Eigenvalue</i>	<b>3.45</b>	<b>1.68</b>	<b>1.40</b>	0.73	0.50	0.43	0.38	0.23	0.20
Porcentagem da variância	38.28	18.63	15.59	8.10	5.59	4.74	4.25	2.60	2.21
Porcentagem da variância acumulada	<b>38.28</b>	<b>56.91</b>	<b>72.51</b>	80.61	86.20	90.94	95.19	97.79	100
<b>Cargas das variáveis originais</b>									
VABS Comunicação	<b>0.89</b>	-0.18	0.15	-0.04	0.06	-0.01	-0.10	-0.13	0.35
VABS Hábitos de vida diários	<b>0.85</b>	-0.13	0.19	-0.04	0.00	-0.30	-0.03	0.35	-0.04
VABS Socialização	<b>0.87</b>	0.04	0.12	0.08	0.14	-0.20	0.26	-0.25	-0.19
ADI-R Socialização	<b>-0.67</b>	-0.23	0.42	-0.28	-0.11	-0.38	-0.24	-0.14	-0.03
ADI-R Comunicação	-0.30	-0.42	<b>0.69</b>	-0.30	0.25	0.21	0.25	0.07	0.01
ADI-R Comportamentos Restritos e Repetitivos	-0.15	<b>-0.57</b>	0.41	0.69	-0.12	0.03	-0.05	-0.01	-0.02
CBCL Problemas Externalizantes	-0.11	<b>-0.67</b>	-0.54	0.01	0.47	-0.05	-0.17	0.00	-0.05
CBCL Problemas Internalizantes	-0.02	<b>-0.74</b>	-0.46	-0.18	-0.37	-0.06	0.27	-0.01	0.04
QI total	<b>0.77</b>	-0.29	0.11	-0.23	-0.19	0.32	-0.28	-0.05	-0.18

É possível verificar que mesmo utilizando um variável número de pacientes na análise de componentes principais, na grande maioria dos casos, os resultados se mantiveram os mesmos. De todos os componentes principais obtidos, os primeiros três explicam cerca de 70% da variabilidade fenotípica observada na amostra e possuem autovalor (*eigenvalue*) maior do que 1, indicando que estes componentes contém a mesma quantidade de informação do que uma única variável original. Dessa forma, os três primeiros componentes principais foram mantidos na análise.

Quando analisamos a correlação das variáveis originais com cada um dos componentes principais, fica claro que a Componente 1 melhor sumariza as informações relativas à Funcionalidade Social, pois possui uma alta correlação com os subitens do instrumento que avalia o comportamento adaptativo dos indivíduos, com o QI, que é responsável por avaliar capacidades cognitivas, e com comportamentos relativos à socialização medidos pela ADI-R. Já a Componente 2 melhor sumariza as informações relativas à Problemas Comportamentais, uma vez que possui correlação com instrumentos que medem a presença de comportamentos internalizantes (comportamentos ansiosos/depressivos, queixas somáticas e isolamento social), comportamentos externalizantes (problemas sociais, comportamento social e comportamento de quebra de regras) e comportamentos e interesses restritos e repetitivos. A Componente 3 melhor sumariza as informações relativas à problemas de comunicação, medidos pelo subitem da ADI-R.

## 4.3 Mapa de heterogeneidade de apresentação clínica

Além de analisar quais instrumentos de avaliação fenotípica mais contribuem para cada componente principal, buscamos verificar se as coordenadas dos 2.744 pacientes nos componentes

principais selecionados possui alguma relação com as variáveis originais. Da Figura 5 até a Figura 13, é possível verificar a graduação das variáveis originais que possuem altas correlações com cada um dos três primeiros componentes principais.

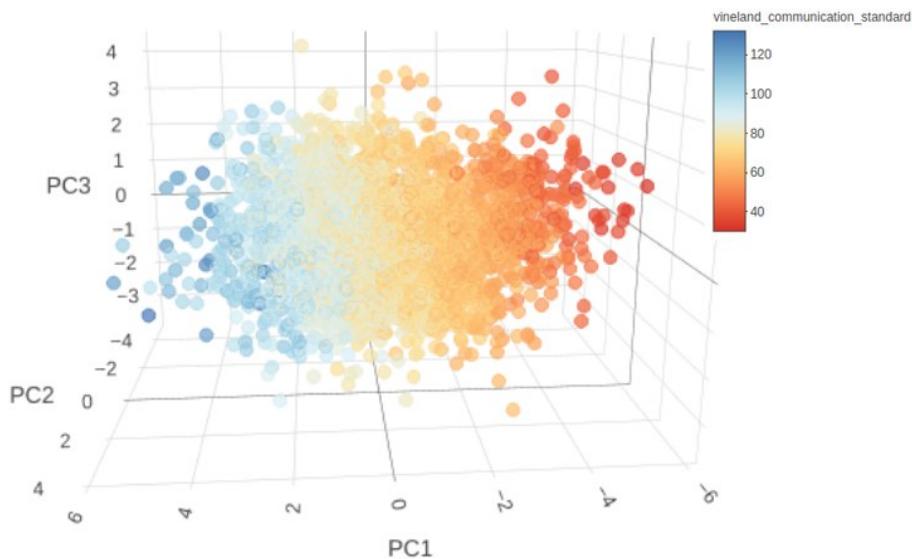


Figura 5 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e a VABS Comunicação. - Cada ponto no gráfico representa a coordenada de um paciente em três componentes principais. Cada paciente está colorido pela sua pontuação no subitem de Comunicação da VABS. As cores vermelhas representam as piores pontuações, portanto, menos habilidades comunicativas e as cores azuis representam as melhores pontuações, portanto, mais habilidades comunicativas.

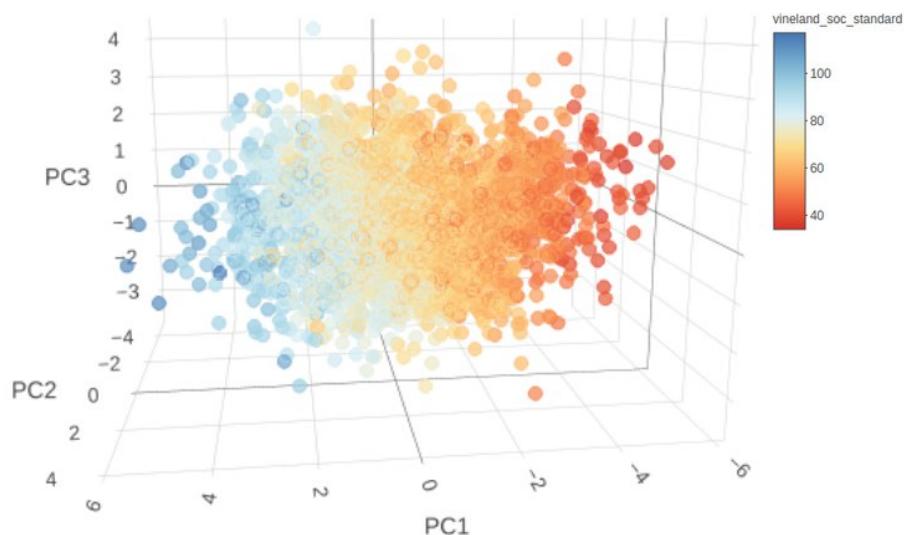


Figura 6 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e a VABS Socialização. - Cada ponto no gráfico representa a coordenada de um paciente em três componentes principais. Cada paciente está colorido pela sua pontuação no subitem de Socialização da VABS. As cores vermelhas representam as piores pontuações, portanto, menos habilidades de socialização e as cores azuis representam as melhores pontuações, portanto, mais habilidades de socialização.

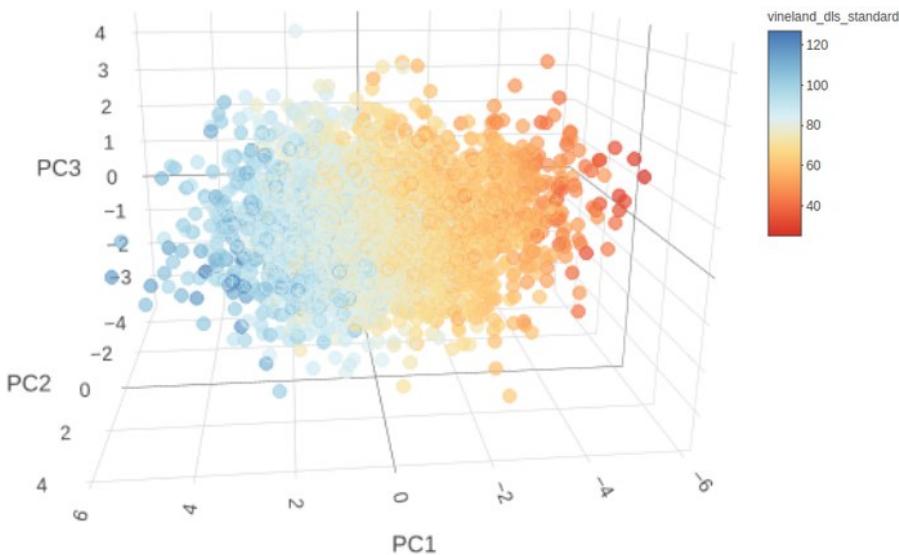


Figura 7 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e a VABS Habilidades de Vida Diárias. - Cada ponto no gráfico representa a coordenada de um paciente em três componentes principais. Cada paciente está colorido pela sua pontuação no subitem de Habilidades de Vida Diárias da VABS. As cores vermelhas representam as piores pontuações, portanto, piores habilidades de vida diárias e as cores azuis representam as melhores pontuações, portanto, melhores habilidades de vida diárias.

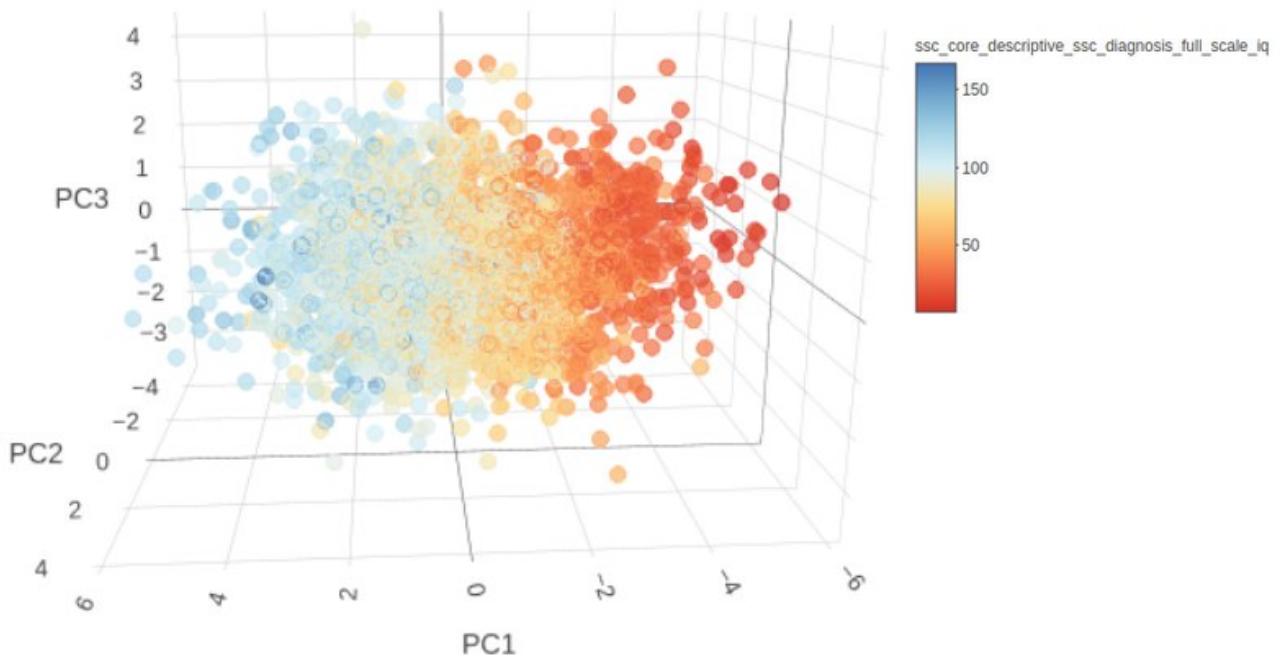


Figura 8 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e o QI. - Cada ponto no gráfico representa a coordenada de um paciente em três componentes principais. Cada paciente está colorido pela sua pontuação obtida no teste de QI. As cores vermelhas representam as piores pontuações, portanto, menor QI e as cores azuis representam as melhores pontuações, portanto, melhor QI.

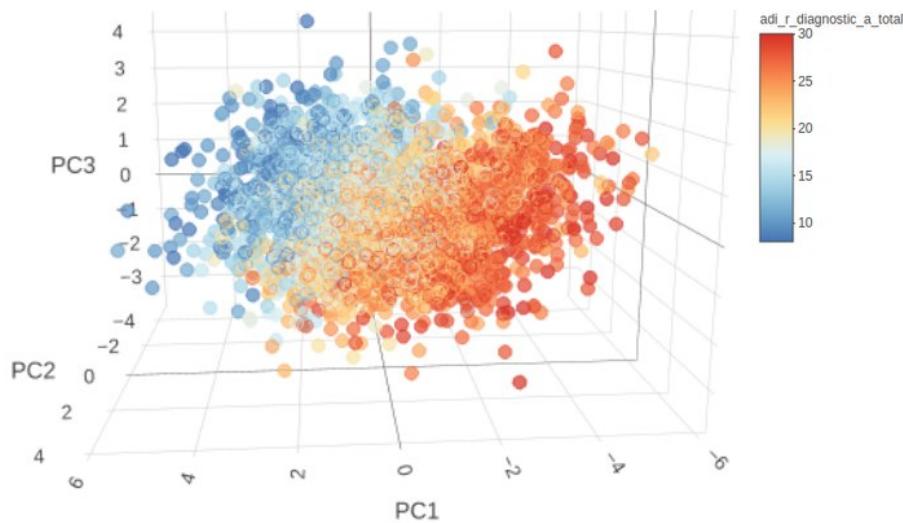


Figura 9 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e a ADI-R Socialização. - Cada ponto no gráfico representa a coordenada de um paciente em três componentes principais. Cada paciente está colorido pela sua pontuação obtida no subitem de Socialização da ADI-R. As cores vermelhas representam as piores pontuações, portanto, menor capacidade de socialização e as cores azuis representam as melhores pontuações, portanto, maior capacidade de socialização.

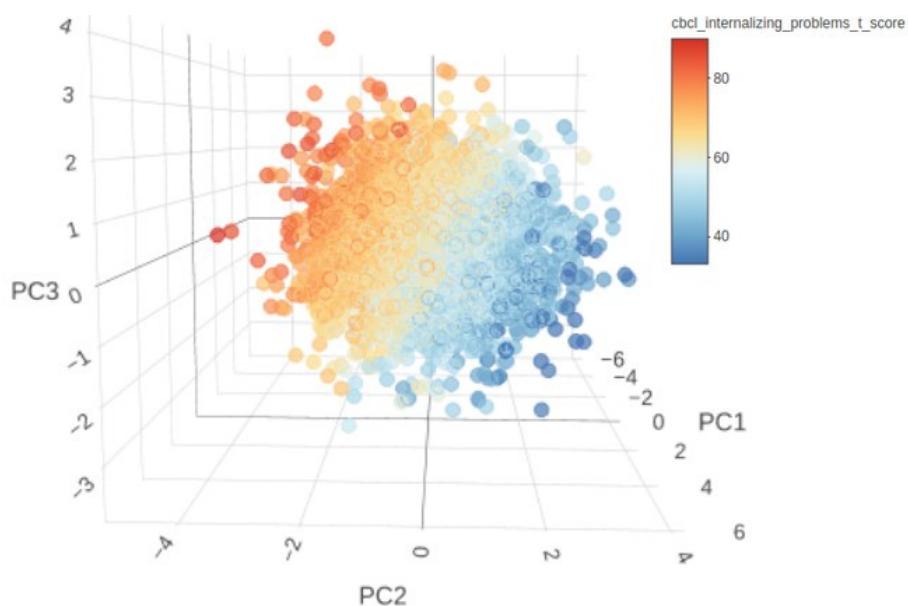


Figura 10 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e a CBCL Problemas Internalizantes. - Cada ponto no gráfico representa a coordenada de um paciente em três componentes principais. Cada paciente está colorido pela sua pontuação obtida no subitem de Problemas Internalizantes da CBCL. As cores vermelhas representam as piores pontuações, portanto, maior apresentação de problemas internalizantes e as cores azuis representam as melhores pontuações, portanto, menor apresentação de problemas internalizantes.

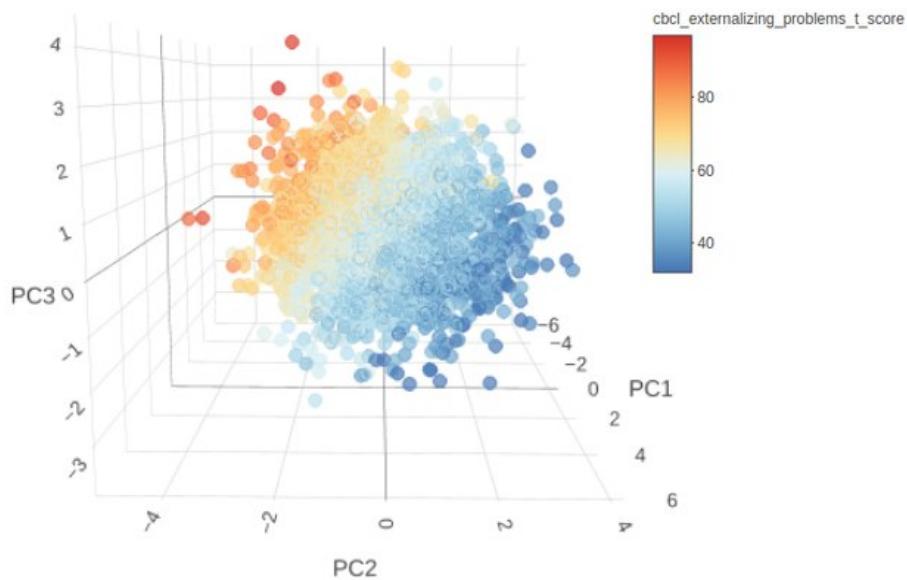


Figura 11 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e a CBCL Problemas Externalizantes. - Cada ponto no gráfico representa a coordenada de um paciente em três componentes principais. Cada paciente está colorido pela sua pontuação obtida no subitem de Problemas Externalizantes da CBCL. As cores vermelhas representam as piores pontuações, portanto, maior apresentação de problemas externalizantes e as cores azuis representam as melhores pontuações, portanto, menor apresentação de problemas externalizantes.

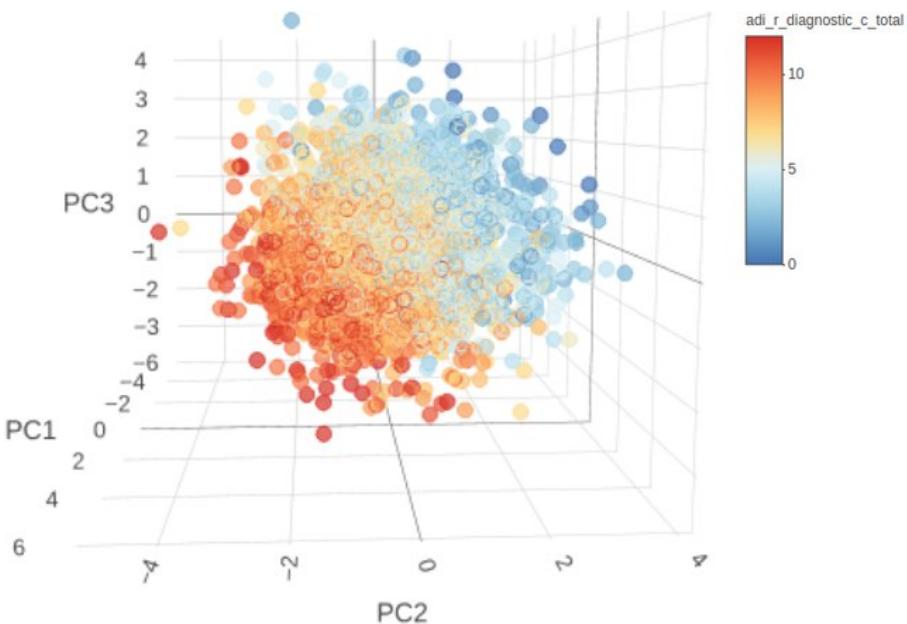


Figura 12 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e a ADI-R Comportamentos Restritivos e Repetitivos. - Cada ponto no gráfico representa a coordenada de um paciente em três componentes principais. Cada paciente está colorido pela sua pontuação obtida no subitem de Comportamentos Restritivos e Repetitivos da ADI-R. As cores vermelhas representam as piores pontuações, portanto, maior apresentação de comportamentos restritos e repetitivos e as cores azuis representam as melhores pontuações, portanto, menor apresentação de comportamentos restritos e repetitivos.

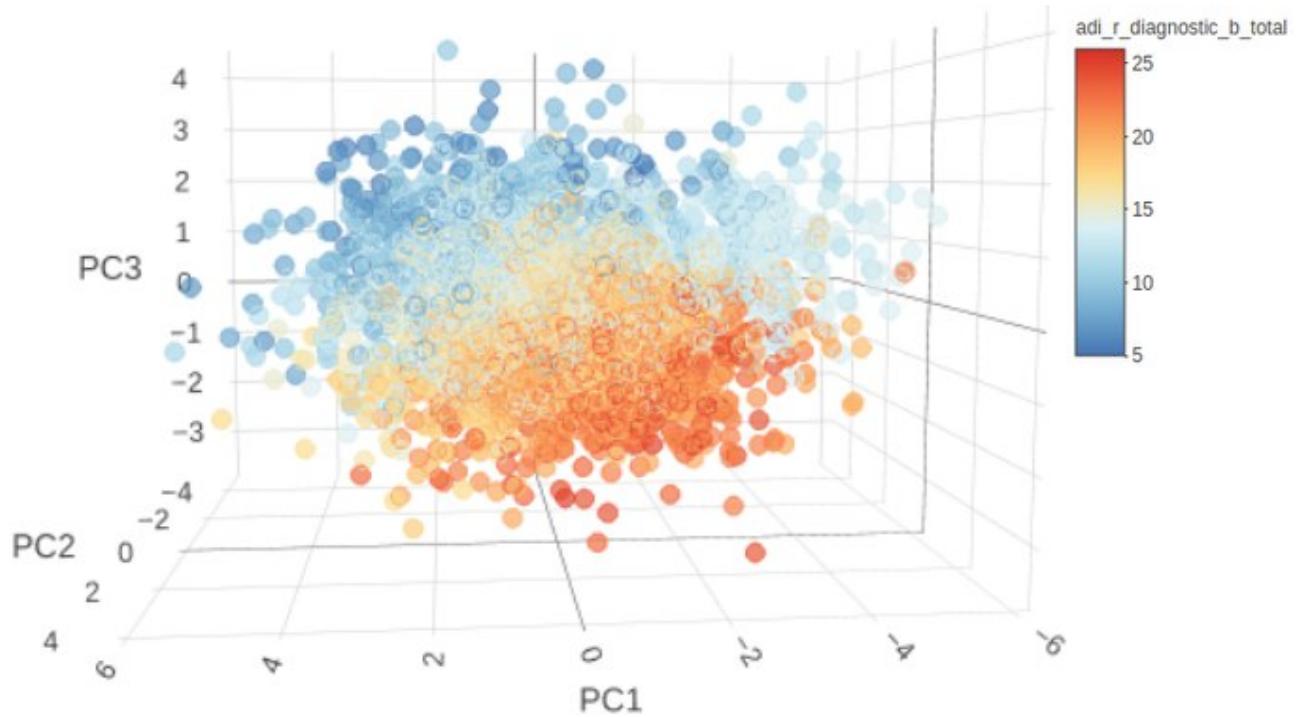


Figura 13 – Relação entre as coordenadas dos pacientes nos três primeiros componentes principais e a ADI-R Comunicação. - Cada ponto no gráfico representa a coordenada de um paciente em três componentes principais. Cada paciente está colorido pela sua pontuação obtida no subitem de Comunicação da ADI-R. As cores vermelhas representam as piores pontuações, portanto, menor capacidade de comunicação e as cores azuis representam as melhores pontuações, portanto, maior capacidade de comunicação.

A partir da visualização das figuras, é possível perceber que há uma relação entre as variáveis originais e as coordenadas dos pacientes nos componentes principais, e consequentemente entre o espaço de coordenadas 3D da PCA. Considerando todo o espaço 3D formado pelas coordenadas nos três componentes principais retidos na análise, verifica-se que os indivíduos localizados nos octantes de escore negativo na primeira componente principal (PC1) possuem piores pontuações nas subescalas dos instrumentos fenotípicos que refletem aspectos de Funcionalidade Social como a VABS, QI e ADI-R no subitem de Socialização. Os indivíduos localizados nos octantes de escore negativo na segunda componente principal (PC2) possuem piores pontuações nos subitens de instrumentos que refletem aspectos de alteração do comportamento como a CBCL Internalizante, CBCL Externalizante e ADI-R no subitem de Comportamentos Restritivos e Repetitivos. Os indivíduos localizados nos octantes de escore negativo no PC3 possuem pontuações piores no subitem de Comunicação da ADI-R.

Desta forma, ao setorizar o sistema de coordenadas de componentes principais em oito octantes, as relações entre a apresentação fenotípica dos pacientes e as suas coordenadas nos PCs pode ser generalizada como indicado na Figura 14.

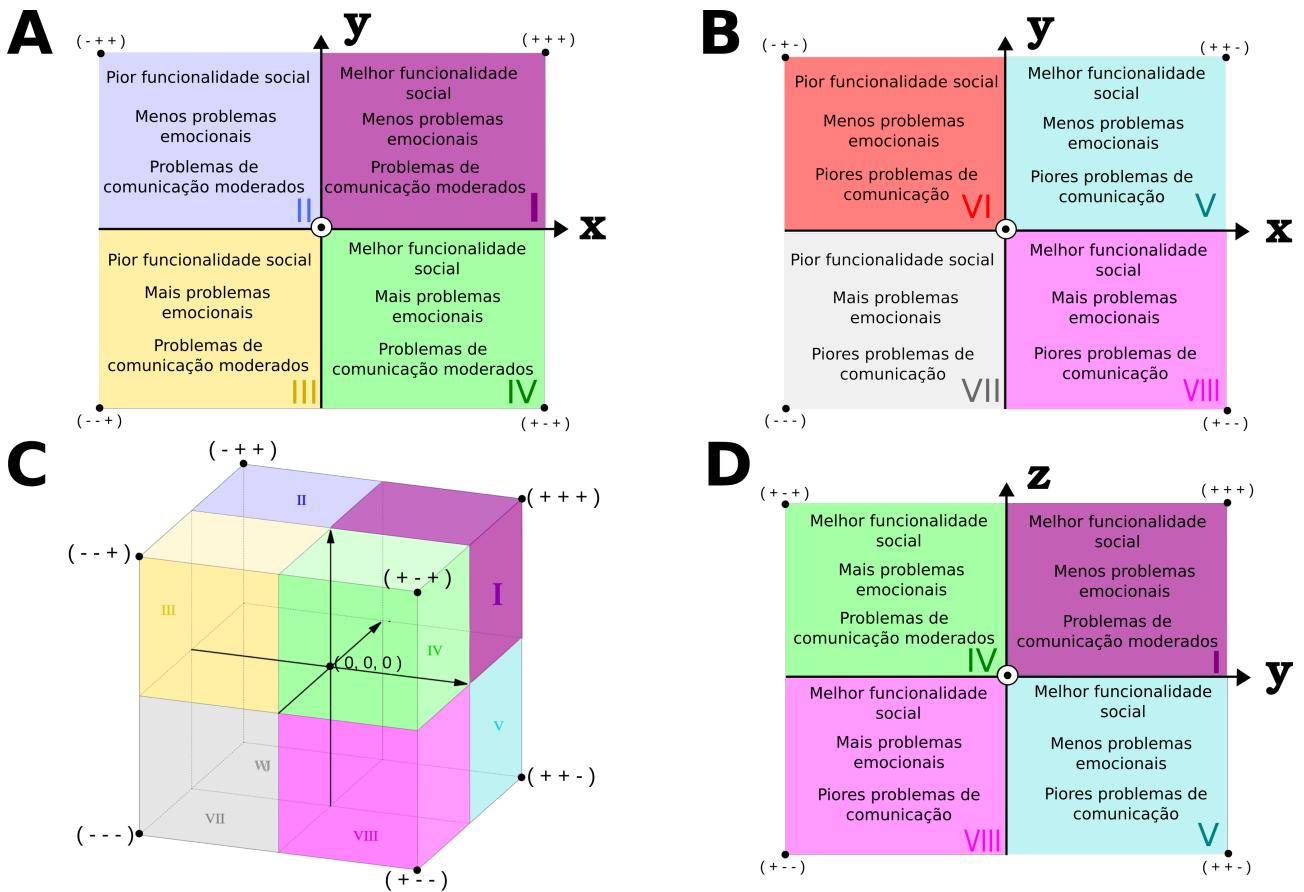


Figura 14 – Implicações clínicas e conceituais da representação dimensional de indivíduos autistas. - O espaço tridimensional proposto para representar os indivíduos é dividido em 8 octantes, rotulados de I a VIII (painel C). Esquematicamente, os eixos x, y e z representam os componentes principais 1, 2 e 3, respectivamente. O espaço formado pelas componentes principais estão representados em uma visão 2D para melhor entendimento (painéis A, B e D) em conjunto com a interpretação clínica de cada atribuída a octante (indicada no texto dentro de cada octante). Cada canto do octante indica com “+” ou “-” sinais qualitativos de melhor ou pior estado clínico para as 3 dimensões. O ponto circulado na origem representa um eixo direcionado para fora do plano mostrado e os octantes positivos e negativos do eixo z são mostrados separadamente no painel D.

Podemos assim atribuir uma interpretação com significado clínico para todo o sistema de coordenadas de componentes principais. As coordenadas na PC1 representam a variação de indivíduos com menos habilidades cognitivas e sociais e pode ser interpretada como um eixo geral de “**Funcionalidade Social**”. Nas coordenadas da PC2, observamos uma graduação de indivíduos com mais problemas de comportamento para indivíduos com menos problemas de comportamento, ou seja, maior adequação da regulação emocional. Este eixo portanto pode ser interpretado como “**Problemas Comportamentais**”. Finalmente, na PC3 há uma graduação de indivíduos com mais problemas de comunicação para indivíduos com menos problemas de comunicação, portanto este componente principal pode ser interpretado como um eixo geral de “**Problemas de Comunicação**”.

A partir da generalização realizada, é possível utilizar o sistema de coordenadas dos componentes principais como um mapa de heterogeneidade de apresentação fenotípica. Ao invés de comparar individualmente 9 escores obtidos por um paciente TEA nos instrumentos de avaliação fenotípica tradicionais, podemos alocar qualquer paciente avaliado por estes instrumentos no mapa construído e observar a sua posição no sistema de coordenadas dos componentes principais como uma forma de entender as dinâmicas comportamentais neste indivíduo no que diz respeito aos aspectos de Funcionalidade Social, Problemas Comportamentais e Problemas de Comunicação.

Podemos resumir as relações entre a posição no sistema de coordenadas de componentes principais e os instrumentos de avaliação fenotípica originais em um único eixo de heterogeneidade fenotípica. Os pacientes que se encontram no octante VII, região de coordenadas negativas para todas as três componentes principais, possuem uma pior apresentação clínica de forma geral, pois apresentam menor nível de funcionalidade social, mais problemas comportamentais e mais problemas de comunicação. Já os pacientes que se encontram no octante I, região de coordenadas positivas para todas as três componentes principais, possuem uma melhor apresentação clínica de forma geral, pois apresentam melhor nível de funcionalidade social, menos problemas comportamentais e menos problemas de comunicação. Assim, se considerarmos a graduação de apresentação clínica dos pacientes partindo do octante VII até o octante I, é possível determinar uma direção central, que une transversalmente os vértices destes octantes, para criar um eixo central de variação fenotípica que reflete as piores apresentações clínicas de forma geral para melhores apresentações clínicas de forma geral (Figura 15).

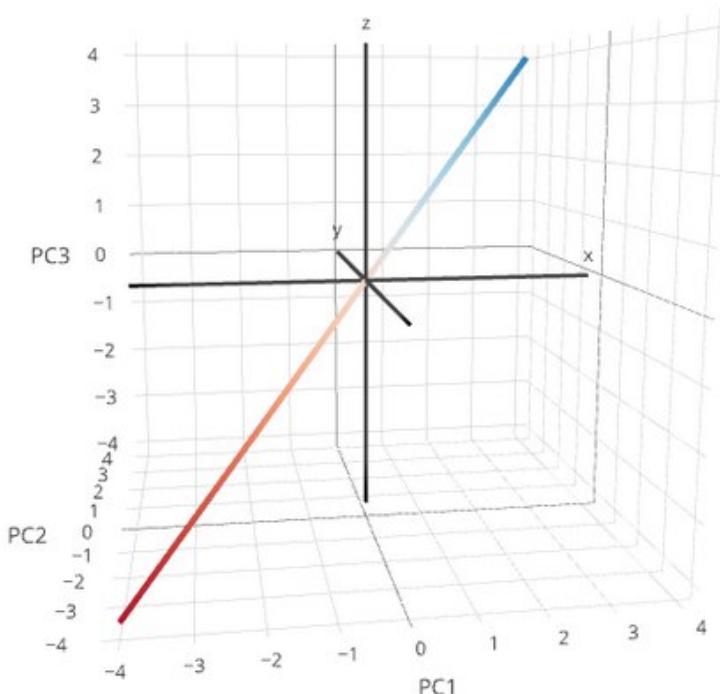


Figura 15 – Eixo de variação fenotípica que representa a gravidade da apresentação clínica de forma geral.

## 4.4 Escore de Gravidade Multidimensional

Além de visualizar os pacientes em um mapa de heterogeneidade de apresentação clínica e analisar empiricamente as suas forças e fraquezas em três domínios comportamentais, é possível quantificar essa representação por meio da modelagem normativa. A heterogeneidade fenotípica observada na coorte utilizada foi modelada por uma função de densidade de probabilidade que leva em consideração a posição relativa entre cada paciente e o eixo central de variação fenotípica.

A lógica por trás da modelagem é a mesma utilizada nos gráficos de crescimento infantil. Calculamos quantis de probabilidade em uma fração a partir do pior extremo de apresentação clínica no mapa de heterogeneidade até a posição de um determinado indivíduo. Os quantis de probabilidade obtidos fornecem uma maneira simples de avaliar a gravidade da apresentação clínica dos pacientes pois retorna uma pontuação individual em uma escala de 0 a 1, sendo que quanto mais próximo do 0, mais próximo o paciente está da região do mapa de heterogeneidade de apresentação clínica em que todos os escores nos PCs são negativos, portanto, o paciente tem uma pior apresentação clínica de forma geral. Essa pontuação foi denominada “Escore de Gravidade Multidimensional”.

Em contraste com os quantis padrão usados em gráficos de crescimento infantil, que são bilaterais, a pontuação Escore de Gravidade Multidimensional proposta é unilateral, uma vez que assume o octante em que todos os escores nos PCs são negativos e o octante em que todos os escores nos PCs são positivos como as piores e melhores apresentações clínicas possíveis, respectivamente, de modo que a apresentação clínica melhora ao longo dessa direção (Figura 14 e Figura 15). O Escore de Gravidade Multidimensional foi atribuído para todos os 2.744 pacientes utilizados na análise ao calcular a distância entre as coordenadas do paciente nas três componentes principais e o eixo central de variação fenotípica e utilizar essa distância como quantil de referência para o cálculo da probabilidade acumulada.

Os painéis A e B da Figura 16 mostram, respectivamente, a distribuição do Escore de Gravidade Multidimensional atribuído para os pacientes da coorte SSC e um boxplot dos valores. É possível perceber que a distribuição das pontuações é aproximadamente uniforme, o que é esperado já que a coorte SSC é um banco de dados de referência amplamente utilizado e que um dos intuios da iniciativa é estabelecer um repositório permanente e representativo de indivíduos TEA, portanto, é esperado que pacientes de todo o espectro de gravidade fenotípica sejam bem representados.

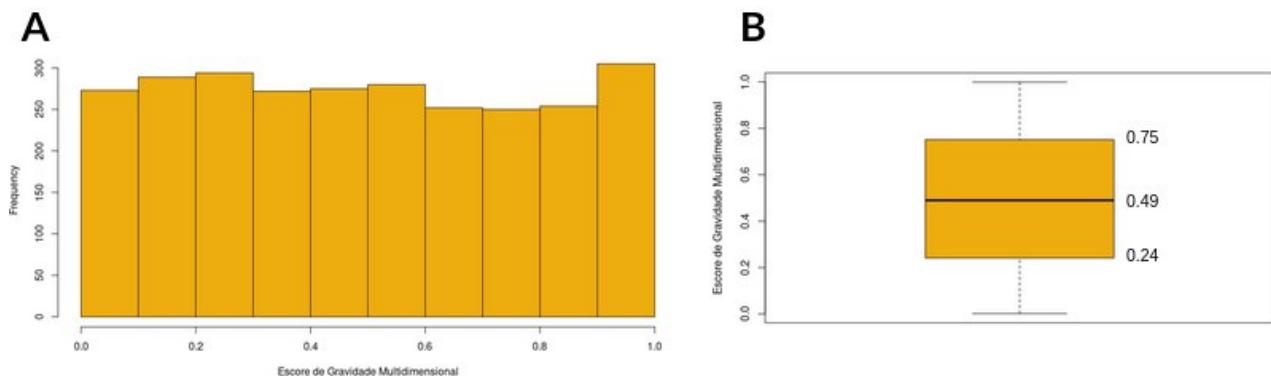


Figura 16 – Distribuição dos valores de Escore de Gravidade Multidimensional nos 2.301 pacientes da SSC.

## 4.5 Associação entre escore de gravidade multidimensional e variantes genéticas

Buscamos entender se o Escore de Gravidade Multidimensional proposto neste trabalho, que reflete a heterogeneidade da apresentação clínica no TEA, possui alguma associação com variantes muito raras herdadas ou *de novo* deletérias em genes categorizados como Essenciais para o desenvolvimento pré e pós natal e em Genes Não Essenciais. Nas Tabelas 11 - Tabela 20 abaixo, mostramos os resultados das regressões, tendo o sexo dos indivíduos como variável de interação.

Tabela 11 – Associação do Escore de Gravidade Multidimensional e o total de variantes. - O total de variantes é a soma de todas as variantes *de novo* de perda de função, *de novo* não sinônimas deletérias e raras herdadas que ocorrem em genes essenciais e não essenciais de um indivíduo. Pseudo R-quadrado: 0.0001985

	Coeficiente	Erro padrão	p-valor
Total de variantes	0.0008026	0.0045580	0.860
Sexo	-0.0121261	0.1406548	0.931
Total de variantes : sexo	-0.0011876	0.0050206	0.813

Tabela 12 – Associação do Escore de Gravidade Multidimensional e o total de variantes em Genes Não Essenciais. - O total de variantes é a soma de todas as variantes *de novo* de perda de função, *de novo* não sinônimas deletérias e raras herdadas que ocorrem em genes não essenciais de um indivíduo. Pseudo R-quadrado: 0.0006198

	Coeficiente	Erro padrão	p-valor
Total de variantes em NEGs	0.0038532	0.0074962	0.607
Sexo	-0.0303821	0.1298220	0.815
Total de variantes em NEGs : sexo	-0.0007489	0.0083144	0.928

Tabela 13 – Associação do Escore de Gravidade Multidimensional e o total de variantes em Genes Essenciais. - O total de variantes é a soma de todas as variantes *de novo* de perda de função, *de novo* não sinônimas deletérias e raras herdadas que ocorrem em genes essenciais de um indivíduo. Pseudo R-quadrado: 0.001068

	<b>Coeficiente</b>	<b>Erro padrão</b>	<b>p-valor</b>
Total de variantes em EGs	-0.002880	0.009788	0.769
Sexo	-0.007745	0.137525	0.955
Total de variantes em EGs : sexo	-0.003034	0.010671	0.776

Tabela 14 – Associação do Escore de Gravidade Multidimensional e o total de variantes raras herdadas. - O total de variantes é a soma de todas as variantes raras herdadas que ocorrem em genes não essenciais e genes essenciais de um indivíduo. Pseudo R-quadrado: 0.0001873

	<b>Coeficiente</b>	<b>Erro padrão</b>	<b>p-valor</b>
Total de variantes raras herdadas	0.0007058	0.0045854	0.878
Sexo	-0.0180327	0.1387424	0.897
Total de variantes raras herdadas : sexo	-0.0009678	0.0050453	0.848

Tabela 15 – Associação do Escore de Gravidade Multidimensional e o total de variantes *de novo* de perda de função (*Loss of Function*, LoF). - O total de variantes é a soma de todas as variantes *de novo* de perda de função que ocorrem em genes não essenciais e genes essenciais de um indivíduo. Pseudo R-quadrado: 0.0003049

	<b>Coeficiente</b>	<b>Erro padrão</b>	<b>p-valor</b>
Total de variantes <i>de novo</i> LoF	0.00227	0.16949	0.989
Sexo	-0.03681	0.07106	0.604
Total de variantes <i>de novo</i> LoF : sexo	-0.04653	0.18754	0.804

Tabela 16 – Associação do Escore de Gravidade Multidimensional e o total de variantes *de novo* não sinônimas deletérias. - O total de variantes é a soma de todas as variantes *de novo* não sinônimas deleteérias (*nonsense damaging*, NSD) que ocorrem em genes não essenciais e genes essenciais de um indivíduo. Pseudo R-quadrado: 0.0005982

	<b>Coeficiente</b>	<b>Erro padrão</b>	<b>p-valor</b>
Total de variantes <i>de novo</i> NSD	0.045356	0.095455	0.635
Sexo	-0.008808	0.080410	0.913
Total de variantes <i>de novo</i> NSD : sexo	-0.080547	0.103935	0.438

Tabela 17 – Associação do Escore de Gravidade Multidimensional e o total de variantes raras herdadas em Genes Essenciais. - Pseudo R-quadrado: 0.0009966

	Coeficiente	Erro padrão	p-valor
Variantes raras herdadas em EGs	-0.003822	0.009950	0.701
Sexo	-0.022188	0.136074	0.870
Variantes raras herdadas em EGs : sexo	-0.001765	0.010826	0.871

Tabela 18 – Associação do Escore de Gravidade Multidimensional e o total de variantes *de novo* de perda de função e *de novo* não sinônimas deletérias em Genes Essenciais. - Nessa análise consideramos os efeitos individuais de cada classe de variante e também os efeitos de interação. Pseudo R-quadrado: 0.001792

	Coeficiente	Erro padrão	p-valor
Variantes <i>de novo</i> LoF em EGs	0.32183	0.29383	0.273
Variantes <i>de novo</i> NSD em EGs	0.08145	0.13827	0.556
Sexo	0.00517	0.07693	0.946
Variantes <i>de novo</i> LoF : variantes <i>de novo</i> NSD	-0.28831	0.42161	0.494
Variantes <i>de novo</i> LoF : sexo	-0.43567	0.31564	0.168
Variantes <i>de novo</i> NSD : sexo	-0.09309	0.15005	0.535
Variantes <i>de novo</i> LoF : variantes <i>de novo</i> NSD : sexo	0.21044	0.465	0.651

Tabela 19 – Associação do Escore de Gravidade Multidimensional e o total de variantes raras herdadas em Genes Não Essenciais. - Pseudo R-quadrado: 0.0006528

	Coeficiente	Erro padrão	p-valor
Variantes raras herdadas em NEGs	0.0040793	0.0075275	0.588
Sexo	-0.0288919	0.1284337	0.822
Variantes raras herdadas em NEGs : sexo	-0.0008867	0.0083413	0.915

Tabela 20 – Associação do Escore de Gravidade Multidimensional e o total de variantes *de novo* de perda de função e *de novo* não sinônimas deletérias em Genes Não Essenciais. - Nessa análise consideramos os efeitos individuais de cada classe de variante e também os efeitos de interação. Pseudo R-quadrado: 0.004278

	Coeficiente	Erro padrão	p-valor
Variantes <i>de novo</i> LoF em NEGs	0.13394	0.37399	0.7203
Variantes <i>de novo</i> NSD em NEGs	0.10345	0.13845	0.4550
Sexo	-0.03423	0.07424	0.6447
Variantes <i>de novo</i> LoF : variantes <i>de novo</i> NSD	-1.40258	0.62970	<b>0.0259*</b>
Variantes <i>de novo</i> LoF : sexo	-0.01852	0.40078	0.9631
Variantes <i>de novo</i> NSD : sexo	-0.09309	0.15005	0.535
Variantes <i>de novo</i> LoF : variantes <i>de novo</i> NSD : sexo	1.42168	0.70473	<b>0.0437*</b>

Foi encontrada uma associação entre o Escore de Gravidade Multidimensional e as variantes *de novo* de perda de função e *de novo* não sinônimas deletérias em Genes Não Essenciais quando consideramos a interação entre essas variantes, mas não foi encontrada associação quando considerando os efeitos individuais. A Figura 17 mostra os efeitos da interação entre variantes *de novo* de perda de função e *de novo* não sinônimas deletérias em Genes Não Essenciais na média marginal estimada do escore de gravidade multidimensional, dependendo do sexo. É possível observar que para pacientes do sexo feminino, a ocorrência de apenas uma variante *de novo* de perda de função em conjunto com mais de uma variante *de novo* não sinônima deletéria é capaz de diminuir o Escore de Gravidade Multidimensional. Entretanto, é preciso ressaltar que apenas 17 pacientes de todo o conjunto de 2.301 analisados possuem pelo menos uma variante do tipo *de novo* de perda de função e mais de uma variante do tipo *de novo* não sinônima deletéria, e destes, apenas 5 são do sexo feminino.

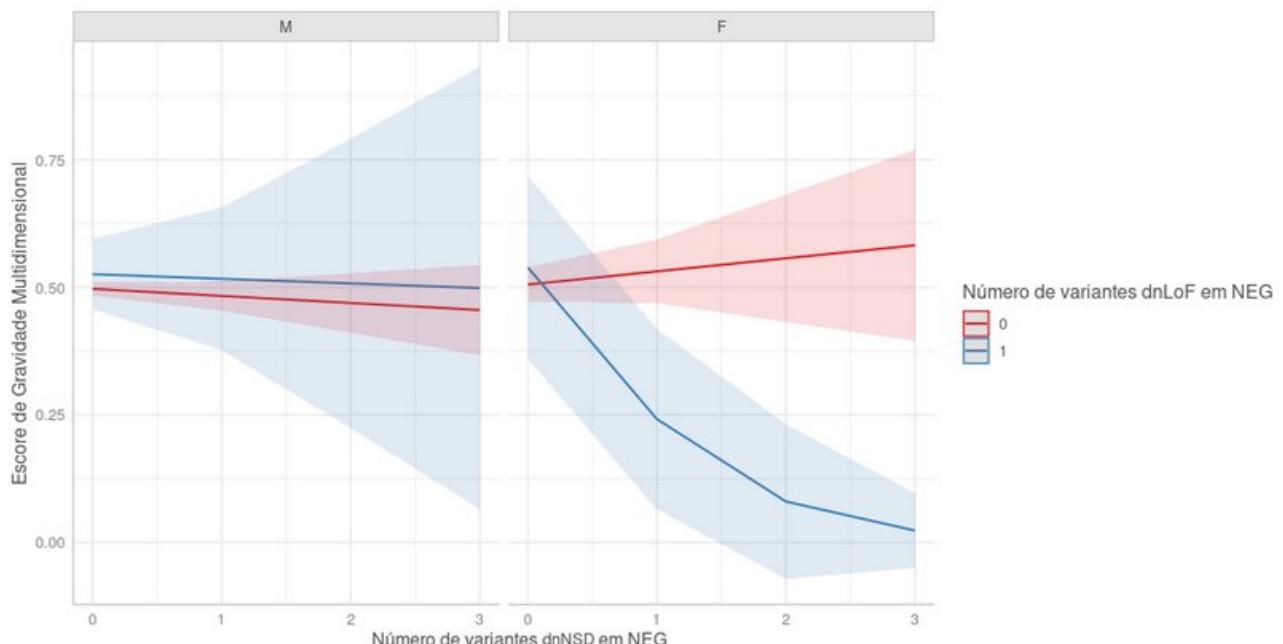


Figura 17 – Efeitos da interação entre variantes *de novo* de perda de função e *de novo* não sinônimas deletérias em Genes Não Essenciais na média marginal estimada do escore de gravidade multidimensional. - No eixo y são mostrados os valores de Média Marginal Estimada para o Escore de Gravidade Multidimensional. No eixo x são mostrados o número de variantes *de novo* não sinônimas deletérias (dnNSD) em Genes Não Essenciais (NEG). A linha vermelha representa a ocorrência de nenhuma variante *de novo* de perda de função (dnLoF) nos genes não essenciais e a linha azul representa a ocorrência de uma variante. O painel indicado por M representam os pacientes do sexo masculino e o painel indicado por F representam os pacientes do sexo feminino.

Buscamos em seguida investigar a existência de associações entre o escore de gravidade e genes que apresentam padrões de expressão temporal em módulos de co-expressão no cérebro em desenvolvimento. Foram realizadas regressões individualmente para os genes de cada um

dos 41 módulos de co-expressão apontados por Ji, X., *et al.* (2016). O escore de gravidade sempre foi considerado como variável dependente e o número de variantes de cada classe que afetam os genes de cada módulo foi considerado como variável independente, da mesma forma que as análises anteriores. Encontramos uma possível associação entre o escore de gravidade multidimensional e os genes afetados por variantes nos módulos de co-expressão M01, M11, M12, M13 e M31 (Tabelas 21 - 25).

Tabela 21 – Associação do Escore de Gravidade Multidimensional e o total de variantes em genes do módulo M01. - No módulo, há 1.176 genes em 1.842 indivíduos afetados por 48 variantes *de novo* de perda de função, 147 variantes *de novo* não sinônima deletéria e 4.648 variantes raras herdadas. Pseudo R-quadrado: 0.004455

	<b>Coeficiente</b>	<b>Erro padrão</b>	<b>p-valor</b>
Variantes <i>de novo</i> de perda de função	-0.36126	0.16057	<b>0.0245*</b>
Variantes <i>de novo</i> não sinônima deletéria	-0.08547	0.09412	0.3638
Variantes raras herdadas	0.01799	0.01389	0.1953

As variantes *de novo* de perda de função que ocorrem nos genes do módulo M01 se mostraram associadas com o escore de gravidade multidimensional. Dos 1.848 pacientes utilizados nessa análise, 48 possuem uma variante *de novo* de perda de função em 44 genes diferentes, dos quais 20 são Genes Essenciais, 5 são Genes Não Essenciais e 19 genes não pertencem a nenhum desses dois grupos. Conforme reportado por Ji, X., *et al.* (2016) os genes pertencentes ao módulo M01 são mais expressos no início da gestação, com seus valores mais altos em 8 semanas e decaindo até o momento do nascimento.

Tabela 22 – Associação do Escore de Gravidade Multidimensional e o total de variantes em genes do módulo M11. - No módulo, há 359 genes em 1.227 indivíduos afetados por 5 variantes *de novo* de perda de função, 37 variantes *de novo* não sinônima deletéria e 1.968 variantes raras herdadas. Pseudo R-quadrado: 0.005038

	<b>Coeficiente</b>	<b>Erro padrão</b>	<b>p-valor</b>
Variantes <i>de novo</i> de perda de função	-0.18627	0.49134	0.7046
Variantes <i>de novo</i> não sinônima deletéria	0.22081	0.18547	0.2338
Variantes raras herdadas	0.06980	0.03080	<b>0.0234*</b>

As variantes raras herdadas que ocorrem nos genes do módulo M11 se mostraram associadas com o escore de gravidade multidimensional. Dos 1.227 pacientes utilizados nessa análise, 1.205 possuem uma variante rara herdada deletéria em 354 genes diferentes, dos quais 60 são Genes Essenciais, 143 são Genes Não Essenciais e 151 genes não pertencem a nenhum desses dois grupos. Conforme reportado por Ji, X., *et al.* (2016) os genes pertencentes ao módulo M11 são mais expressos no período pós-natal.

Tabela 23 – Associação do Escore de Gravidade Multidimensional e o total de variantes em genes do módulo M12. - No módulo, há 342 genes em 1.299 indivíduos afetados por 8 variantes *de novo* de perda de função, 35 variantes *de novo* não sinônima deletéria e 2.319 variantes raras herdadas. Pseudo R-quadrado: 0.005785

	<b>Coeficiente</b>	<b>Erro padrão</b>	<b>p-valor</b>
Variantes <i>de novo</i> de perda de função	-0.85836	0.39101	<b>0.0281*</b>
Variantes <i>de novo</i> não sinônima deletéria	0.15667	0.19103	0.4121
Variantes raras herdadas	0.02609	0.02683	0.3309

As variantes *de novo* de perda de função que ocorrem nos genes do módulo M12 se mostraram associadas com o escore de gravidade multidimensional. Dos 1.299 pacientes utilizados nessa análise, 15 possuem uma variante *de novo* de perda de função em 11 genes diferentes, dos quais 5 são Genes Essenciais, 4 são Genes Não Essenciais e 2 genes não pertencem a nenhum desses dois grupos. Os genes pertencentes ao módulo M12 possuem alto nível de expressão no período pós-natal (54).

Tabela 24 – Associação do Escore de Gravidade Multidimensional e o total de variantes em genes do módulo M13. - No módulo, há 307 genes em 1.108 indivíduos afetados por 7 variantes *de novo* de perda de função, 28 variantes *de novo* não sinônima deletéria e 1.681 variantes raras herdadas. Pseudo R-quadrado: 0.01043

	<b>Coeficiente</b>	<b>Erro padrão</b>	<b>p-valor</b>
Variantes <i>de novo</i> de perda de função	0.60848	0.42055	0.14793
Variantes <i>de novo</i> não sinônima deletéria	0.55418	0.21317	<b>0.00933**</b>
Variantes raras herdadas	0.07363	0.03692	<b>0.04613*</b>

As variantes *de novo* sem sentido deletérias e raras herdadas que ocorrem nos genes do módulo M13 se mostraram associadas com o escore de gravidade multidimensional. Dos 1.108 pacientes utilizados nessa análise, 1.103 possuem ao menos uma variante rara herdada em 303 genes diferentes, dos quais 75 são Genes Essenciais, 138 são Genes Não Essenciais e 90 genes não pertencem a nenhum desses dois grupos. Outros 32 pacientes possuem ao menos uma variante *de novo* não sinônima deletéria em 29 genes diferentes, dos quais 12 são Genes Essenciais, 10 são Genes Não Essenciais e 7 não pertencem a nenhum grupo. Os genes pertencentes ao módulo M13 possuem alto nível de expressão no período pós-natal (54).

Tabela 25 – Associação do Escore de Gravidade Multidimensional e o total de variantes em genes do módulo M31. - No módulo, há 145 genes em 586 indivíduos afetados por 8 variantes *de novo* de perda de função, 25 variantes *de novo* não sinônima deletéria e 701 variantes raras herdadas. Pseudo R-quadrado: 0.01268

	<b>Coeficiente</b>	<b>Erro padrão</b>	<b>p-valor</b>
Variantes <i>de novo</i> de perda de função	-0.91135	0.40476	<b>0.0243*</b>
Variantes <i>de novo</i> não sinônima deletéria	0.19363	0.24358	0.4266
Variantes raras herdadas	0.07345	0.08616	0.3940

As variantes *de novo* de perda de função que ocorrem nos genes do módulo M31 se mostraram associadas com o escore de gravidade multidimensional. Dos 586 pacientes utilizados nessa análise, 8 possuem ao menos uma variante rara herdada em 8 genes diferentes, dos quais 2 são Genes Essenciais, 6 são Genes Não Essenciais e 6 genes não pertencem a nenhum desses dois grupos. Os genes pertencentes ao módulo M31 possuem alto nível de expressão no período pré-natal, com o pico de expressão entre 19 e 24 semanas de gestação (54).

## 5 Discussão

Buscamos analisar as relações existentes entre os sintomas core do TEA a outros comportamentos apontados pela literatura como modificadores dos mesmos a fim de melhor compreender a heterogeneidade de apresentação clínica. Por meio da análise de componentes principais nos instrumentos que medem essas características, selecionamos três componentes principais que explicaram 70% da variabilidade da amostra de referência e construímos um mapa de heterogeneidade de apresentação clínica. A partir da modelagem normativa utilizando os componentes principais, criamos um escore de gravidade multidimensional que pode ser utilizado como uma forma sistematizada para analisar a heterogeneidade fenotípica no TEA em qualquer paciente avaliado pelas escalas fenotípicas utilizadas. Esse escore também possui alguma associação com variantes raras deletérias, e pode ser utilizado futuramente para o melhor entendimento das relações genética-fenótipo no transtorno.

Ao reduzir a dimensionalidade das características recorrentemente apontadas na literatura como associadas à heterogeneidade e gravidade da apresentação clínica do TEA por meio da PCA, obtivemos três componentes principais que possuem relação com três amplos aspectos comportamentais. Vale ressaltar que a estabilidade dos componentes foi verificada ao se utilizar 30-80% da amostra da SSC.

O primeiro componente principal obtido está associado a aspectos de Funcionalidade Social e reflete trabalhos na literatura que mostraram que o QI e o comportamento adaptativo possuem correlação com os escores de socialização e também interferem na gravidade clínica (63, 62, 64). O peso do VABS neste componente é corroborado por estudos longitudinais que mostraram um papel importante do funcionamento adaptativo no delineamento de subgrupos de TEA que possuem diferentes trajetórias de desenvolvimento (93). A presença do QI neste componente pode ser explicada por estudos que indicam que um QI mais alto pode ser necessário, embora não suficiente por si só, para a obtenção de resultados positivos em áreas de funcionamento, como relacionamentos, emprego e auto suficiência (94). Tillmann *et al.*, (2019) testaram preditores únicos de funcionamento adaptativo medido pelo VABS e a discrepância entre QI e funcionamento adaptativo no TEA, e mostraram que os comportamentos sociocomunicativos, mas não os comportamentos repetitivos ou sintomas psiquiátricos concomitantes (como ansiedade, depressão e TDAH), estão associados a um menor funcionamento adaptativo e maiores dificuldades na capacidade adaptativa (95).

O segundo componente principal obtido está associado a Problemas Comportamentais, quando levamos em conta características internalizantes e externalizantes e comportamentos restritos e repetitivos. Antes da nova classificação do DSM-5, existiam três dimensões para o diagnóstico do TEA: os comportamentos restritos e repetitivos eram considerados como uma dimensão comportamental para diagnóstico de TEA, alterações na sociabilidade eram consideradas uma segunda dimensão e problemas na comunicação eram considerados como

uma terceira dimensão. Na nova classificação, os comportamentos restritos e repetitivos são considerados uma dimensão separada da segunda dimensão que é a comunicação social, dessa forma era de se esperar que os comportamentos repetitivos e estereotipados estariam mais correlacionados com um Componente Principal diferente daquele que reflete a sociabilidade.

Comorbidades psiquiátricas e médicas estão presentes em cerca de 70% dos indivíduos com TEA (66), sendo que os transtornos de ansiedade afetam cerca de 40% das crianças e adolescentes com TEA (96) e a depressão é altamente prevalente ao longo da vida afetando entre 17% e 70% dos pacientes (97). As deficiências na regulação emocional já foram apontadas como um fator de risco para ansiedade em TEA (98). Georgiades *et al.* (2011) examinaram a sobreposição fenotípica entre as principais características diagnósticas e os problemas emocionais / comportamentais em TEA usando PCA (99). Assim como observado neste trabalho, o componente associado ao comportamento emocional era independente das habilidades intelectuais e adaptativas funcionais das crianças. Além disso, o comportamento restrito e repetitivo e a falta de flexibilidade foram considerados como contribuintes para a desregulação emocional em crianças com TEA (100). Outros estudos apontam a presença de hiper-foco em eventos inesperados e perturbadores em indivíduos TEA assim como persistência do sentimento de raiva, (101, 102), uma vez que a interação entre o diagnóstico de TEA e a presença de comportamentos restritos e repetitivos foi capaz de prever a persistência do sentimento de raiva (100).

O terceiro componente principal obtido está associado a Problemas de Comunicação, medidos pelo subitem da ADI-R. No DSM-5, os problemas de comunicação foram integrados em um modelo bidimensional (1). A ausência de expressão facial, e atitudes como apontar, balançar a cabeça, e a comunicação não verbal, foram considerados parte do construto de interação social enquanto a ecolalia, frases idiossincráticas, rituais verbais foram considerados parte de déficits em padrões repetitivos e restritos de comportamento. Mudanças de linguagem não são necessárias para o diagnóstico, mas 30% das crianças com TEA permanecem minimamente verbais ao entrar na escola (103, 104). Um estudo recente concluiu que resultados adaptativos mais baixos, QI mais alto e habilidade de linguagem na infância tendem a predizer resultados de autonomia na idade adulta, o que é influenciado pela capacidade de comunicação (105). Estudos posteriores ao DSM-5 reconhecem os problemas de linguagem como um fator prognóstico e para melhor representar a heterogeneidade clínica, e o DSM-5 propõe o uso da habilidade de linguagem como um modificador do TEA (106). Em nossa análise, o escore de comunicação ADI-R tem sua contribuição para PC1, PC2, mas foi separado como componente específico na PC3 dando suporte para a contribuição da linguagem dentro da heterogeneidade da apresentação clínica e do espectro de gravidade (107).

Portanto, por meio da análise de componentes principais, corroboramos diversos estudos publicados que enfatizam a importância de avaliar problemas emocionais e comportamentais, funcionalidade, QI e habilidade de linguagem em conjunto com os principais sintomas diagnósticos em crianças com TEA para melhor caracterizar a heterogeneidade clínica e avaliação da gravidade (62, 65, 64). Além disso, propomos uma forma sistematizada de avaliar essas informações em conjunto por meio de um mapa de heterogeneidade de apresentação clínica e um escore de

gravidade global, construído a partir de modelagem normativa. Além de utilizar o escore para melhor caracterização de um paciente de forma individual, é possível fazer comparações entre casos de uma forma padronizada.

Estudos na literatura apontam a relação entre variantes muito raras, *de novo* e de alto impacto com aspectos comportamentais individuais como o QI ou deficiências no controle motor (34, 53, 52). Aqui, consideramos que uma das formas de avaliar as relações entre essas variantes e a heterogeneidade fenotípica no TEA seria a partir da representação dos fenótipos como um *continuum* de gravidade multidimensional. Quando analisamos a associação entre o escore de gravidade gerado e variantes do tipo *de novo* de perda de função, *de novo* sem sentido deletéria e rara herdada considerando apenas o número total de variantes de alto impacto, não foi possível verificar nenhuma relação. O mesmo ocorreu ao analisar as variantes que ocorriam apenas em grupos de genes classificados como essenciais para o desenvolvimento. Já ao analisar a associação entre o escore e as variantes de alto impacto em genes não essenciais, encontramos uma associação quando consideramos a interação das variantes *de novo* de perda de função e não sinônimas deletérias e o sexo. Esse achado é interessante pois corrobora o papel do escore criado como medida de gravidade e sugere que apesar da carga mutacional de variantes *de novo* deletérias em genes essenciais apresentarem importante papel para a vulnerabilidade (54), talvez não seja o que mais contribui para a gravidade de apresentação clínica. No entanto, o quanto estas variantes contribuem para explicação da variabilidade do escore é pouco significativo.

Independentemente da essencialidade dos genes, estudos têm mostrado o importante papel de variantes *de novo* e muito raras associadas ao TEA quando afetam genes expressos no cérebro durante a gestação e primeiros anos de vida (48). Assim, esperaríamos que variantes nesses genes estivessem associados com o escore de gravidade. Ao refinar o nível de informação biológica considerada na análise e investigar o papel de genes que participam de módulos de co-expressão no desenvolvimento cerebral, foi possível identificar associações mais robustas com o escore de gravidade multidimensional. Apontamos que a ocorrência de variantes *de novo* de perda de função, *de novo* não sinônima deletéria e rara herdada deletéria em 697 genes com expressão predominantemente pós-natal e que participam de três módulos de co-expressão distintos estão associadas com o escore de gravidade multidimensional. Também verificamos essa associação em menor grau com variantes *de novo* de perda de função em 52 genes com expressão predominante no período pré-natal e que participam de dois módulos de co-expressão distintos. Dessa forma, a segunda hipótese do estudo de que não apenas variantes em genes essenciais, ou expressos no começo da gestação estariam associados com o índice de gravidade se confirmou.

Outros estudos na literatura apontam a importância de se levar em consideração os padrões de expressão gênica dos genes afetados na investigação das relações genética-fenótipo no TEA. CHANG, Jonathan *et al.* (2015) investigaram a relação entre os níveis médios de expressão de genes afetados por SNVs truncadoras em TEA e os fenótipos observados e verificaram que os genes que apresentavam maiores níveis médios de expressão no tecido cerebral (tanto no período pré e pós natal) estavam associados com piores níveis de QI e pontuações mais altas na ADI-R, portanto, um quadro mais grave. Da mesma forma, GAZESTANI, Vahid H. *et al.*

(2019) (108) verificaram a existência de uma rede de regulação gênica alterada que incluía genes com alta expressão cerebral durante o neurodesenvolvimento embrionário e indicaram que a alteração da atividade dos genes pertencentes à essa rede perturbada está correlacionada com as pontuações de déficit social apontados pela ADOS (instrumento de diagnóstico de TEA, semelhante à ADI-R) em crianças com TEA.

Em nossa análise, não foi possível encontrar associação do escore de gravidade com variantes deletérias em genes essenciais para o desenvolvimento expressos no cérebro no início da gestação. Esses grupos de genes já foram amplamente associados com o risco de desenvolvimento de TEA (109, 110) e a ausência de associação pode indicar que os mecanismos responsáveis pelos fatores que interferem na heterogeneidade fenotípica são distintos daqueles de maior impacto, responsáveis pelo risco.

Neste trabalho, consideramos apenas o papel de variantes muito raras e altamente deletérias como um possível indicador da gravidade fenotípica. Entretanto, aspectos como problemas emocionais e comportamentais, funcionalidade, QI e habilidades de linguagem são comportamentos presentes em indivíduos no geral, apesar de importantes modificadores dos sintomas principais do TEA. Dessa forma, é razoável pensar que fatores genéticos adicionais, como variantes herdadas comuns, fatores ambientais e epigenéticos também podem interferir na heterogeneidade de apresentação fenotípica e na gravidade global (111, 112).

## 6 Considerações Finais

O presente trabalho foi dividido em duas etapas. A primeira etapa foi realizada por meio de uma colaboração com a aluna de mestrado e neuropsicóloga especializada em autismo, Joana Portolese, como parte de sua tese de mestrado. Objetivando um projeto multidisciplinar, analisamos diferentes instrumentos de avaliação fenotípica utilizados na prática clínica em pacientes afetados pelo TEA por meio de técnicas de estatística multivariada e desenvolvemos um escore de gravidade com base nos componentes contribuintes para a gravidade que melhor caracteriza a heterogeneidade fenotípica observada como um *continuum* de gravidade. Esses resultados fazem parte de um artigo científico atualmente em preparação no qual dividimos a autoria.

Na segunda etapa, realizamos a análise da relação entre o escore desenvolvido e variantes exônicas deletérias. Verificamos a existência de uma associação entre o escore e a interação de variantes *de novo* de perda de função e variantes *de novo* não sinônimas deletérias em genes não essenciais para pacientes do sexo feminino. No entanto, a associação explica apenas uma pequena parte da amostra. Essa associação também foi verificada em variantes que afetam genes expressos no cérebro e não apenas no período pré-natal. Concluímos que a utilização de modelos normativos em conjunto com técnicas de análise multivariada permitem a análise do TEA como um transtorno multidimensional do ponto de vista fenotípico e que variantes do tipo e que a gravidade global possui relação com variantes exônicas deletérias em grupos de genes específicos. No entanto, consideramos que para aumentar o poder explicativo dessa associação, outros fatores genéticos, epigenéticos e ambientais precisam ser investigados mais a fundo.

# Referências

- 1 ASSOCIAÇÃO, A. P. et al. *DSM-5: Manual diagnóstico e estatístico de transtornos mentais*. [S.l.]: Artmed Editora, 2014.
- 2 BAIO, J. et al. Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2014. *MMWR Surveillance Summaries*, Centers for Disease Control and Prevention, v. 67, n. 6, p. 1, 2018.
- 3 WERLING, D. M.; GESCHWIND, D. H. Sex differences in autism spectrum disorders. *Current opinion in neurology*, NIH Public Access, v. 26, n. 2, p. 146, 2013.
- 4 MASİ, A. et al. An overview of autism spectrum disorder, heterogeneity and treatment options. *Neuroscience bulletin*, Springer, v. 33, n. 2, p. 183–193, 2017.
- 5 HALLMAYER, J. et al. Genetic heritability and shared environmental factors among twin pairs with autism. *Archives of general psychiatry*, American Medical Association, v. 68, n. 11, p. 1095–1102, 2011.
- 6 SEALEY, L. et al. Environmental factors in the development of autism spectrum disorders. *Environment international*, Elsevier, v. 88, p. 288–298, 2016.
- 7 GARDENER, H.; SPIEGELMAN, D.; BUKA, S. L. Prenatal risk factors for autism: comprehensive meta-analysis. *The British journal of psychiatry*, Cambridge University Press, v. 195, n. 1, p. 7–14, 2009.
- 8 REICHENBERG, A. et al. Advancing paternal age and autism. *Archives of general psychiatry*, American Medical Association, v. 63, n. 9, p. 1026–1032, 2006.
- 9 LI, M. et al. The association of maternal obesity and diabetes with autism and other developmental disabilities. *Pediatrics*, Am Acad Pediatrics, v. 137, n. 2, p. e20152206, 2016.
- 10 LEE, B. K. et al. Maternal hospitalization with infection during pregnancy and risk of autism spectrum disorders. *Brain, behavior, and immunity*, Elsevier, v. 44, p. 100–105, 2015.
- 11 CHRISTENSEN, J. et al. Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. *Jama*, American Medical Association, v. 309, n. 16, p. 1696–1703, 2013.
- 12 SANDIN, S. et al. The heritability of autism spectrum disorder. *Jama*, American Medical Association, v. 318, n. 12, p. 1182–1184, 2017.
- 13 CIRULLI, E. T.; GOLDSTEIN, D. B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews Genetics*, Nature Publishing Group, v. 11, n. 6, p. 415, 2010.
- 14 IAKOUCHeva, L. M.; MUOTRI, A. R.; SEBAT, J. Getting to the cores of autism. *Cell*, Elsevier, v. 178, n. 6, p. 1287–1298, 2019.
- 15 GLESSNER, J. T.; CONNOLLY, J. J.; HAKONARSON, H. Genome-wide association studies of autism. *Current Behavioral Neuroscience Reports*, v. 1, n. 4, p. 234–241, Dec 2014. ISSN 2196-2979. Disponível em: <<https://doi.org/10.1007/s40473-014-0023-0>>.

- 16 WANG, K. et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature*, Springer Science and Business Media LLC, v. 459, n. 7246, p. 528–533, abr. 2009. Disponível em: <<https://doi.org/10.1038/nature07999>>.
- 17 MA, D. et al. A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Annals of Human Genetics*, Wiley, v. 73, n. 3, p. 263–273, maio 2009. Disponível em: <<https://doi.org/10.1111/j.1469-1809.2009.00523.x>>.
- 18 WEISS, L. A.; ; ARKING, D. E. A genome-wide linkage and association scan reveals novel loci for autism. *Nature*, Springer Science and Business Media LLC, v. 461, n. 7265, p. 802–808, out. 2009. Disponível em: <<https://doi.org/10.1038/nature08490>>.
- 19 KIEZUN, A. et al. Exome sequencing and the genetic basis of complex traits. *Nature Genetics*, Springer Nature, v. 44, n. 6, p. 623–630, maio 2012. Disponível em: <<https://doi.org/10.1038/ng.2303>>.
- 20 TORRICO, B. et al. Lack of replication of previous autism spectrum disorder GWAS hits in european populations. *Autism Research*, Wiley, v. 10, n. 2, p. 202–211, jul. 2016. Disponível em: <<https://doi.org/10.1002/aur.1662>>.
- 21 GROVE, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics*, Springer Science and Business Media LLC, v. 51, n. 3, p. 431–444, fev. 2019. Disponível em: <<https://doi.org/10.1038/s41588-019-0344-8>>.
- 22 DUDBRIDGE, F. Correction: Power and predictive accuracy of polygenic risk scores. *PLoS Genetics*, Public Library of Science (PLoS), v. 9, n. 4, abr. 2013. Disponível em: <<https://doi.org/10.1371/annotation/b91ba224-10be-409d-93f4-7423d502cba0>>.
- 23 KLEI, L. et al. Common genetic variants, acting additively, are a major source of risk for autism. *Molecular Autism*, Springer Nature, v. 3, n. 1, p. 9, 2012. Disponível em: <<https://doi.org/10.1186/2040-2392-3-9>>.
- 24 GROVE, J. et al. Common risk variants identified in autism spectrum disorder. *BioRxiv*, Cold Spring Harbor Laboratory, p. 224774, 2017.
- 25 META-ANALYSIS of GWAS of over 16, 000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Molecular Autism*, Springer Nature, v. 8, n. 1, maio 2017. Disponível em: <<https://doi.org/10.1186/s13229-017-0137-9>>.
- 26 GENETIC relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics*, Springer Science and Business Media LLC, v. 45, n. 9, p. 984–994, ago. 2013. Disponível em: <<https://doi.org/10.1038/ng.2711>>.
- 27 GAUGLER, T. et al. Most genetic risk for autism resides with common variation. *Nature Genetics*, Springer Science and Business Media LLC, v. 46, n. 8, p. 881–885, jul. 2014. Disponível em: <<https://doi.org/10.1038/ng.3039>>.
- 28 SEBAT, J. et al. Strong association of de novo copy number mutations with autism. *Science*, American Association for the Advancement of Science, v. 316, n. 5823, p. 445–449, 2007.
- 29 O'ROAK, B. J. et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature genetics*, Nature Publishing Group, v. 43, n. 6, p. 585, 2011.
- 30 O'ROAK, B. J. et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*, Nature Publishing Group, v. 485, n. 7397, p. 246, 2012.

- 31 NEALE, B. M. et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*, Nature Publishing Group, v. 485, n. 7397, p. 242, 2012.
- 32 IOSSIFOV, I. et al. De novo gene disruptions in children on the autistic spectrum. *Neuron*, Elsevier, v. 74, n. 2, p. 285–299, 2012.
- 33 FISCHBACH, G. D.; LORD, C. The simons simplex collection: a resource for identification of autism genetic risk factors. *Neuron*, Elsevier, v. 68, n. 2, p. 192–195, 2010.
- 34 IOSSIFOV, I. et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, Nature Publishing Group, v. 515, n. 7526, p. 216, 2014.
- 35 BEN-DAVID, E.; SHIFMAN, S. Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. *Molecular psychiatry*, Nature Publishing Group, v. 18, n. 10, p. 1054, 2013.
- 36 RUBEIS, S. D. et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, Nature Publishing Group, v. 515, n. 7526, p. 209, 2014.
- 37 WILLSEY, A. J. et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*, Elsevier, v. 155, n. 5, p. 997–1007, 2013.
- 38 HORMOZDIARI, F. et al. The discovery of integrated gene networks for autism and related disorders. *Genome research*, Cold Spring Harbor Lab, v. 25, n. 1, p. 142–154, 2015.
- 39 SHOHAT, S.; BEN-DAVID, E.; SHIFMAN, S. Varying intolerance of gene pathways to mutational classes explain genetic convergence across neuropsychiatric disorders. *Cell reports*, Elsevier, v. 18, n. 9, p. 2217–2227, 2017.
- 40 KOSMICKI, J. A. et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nature genetics*, Nature Publishing Group, v. 49, n. 4, p. 504, 2017.
- 41 STEIN, J. L.; PARIKSHAK, N. N.; GESCHWIND, D. H. Rare inherited variation in autism: beginning to see the forest and a few trees. *Neuron*, Elsevier, v. 77, n. 2, p. 209–211, 2013.
- 42 TIMOTHY, W. Y. et al. Using whole-exome sequencing to identify inherited causes of autism. *Neuron*, Elsevier, v. 77, n. 2, p. 259–273, 2013.
- 43 KRUMM, N. et al. Excess of rare, inherited truncating mutations in autism. *Nature genetics*, Nature Publishing Group, v. 47, n. 6, p. 582, 2015.
- 44 BRANDLER, W. M. et al. Paternally inherited cis-regulatory structural variants are associated with autism. *Science*, American Association for the Advancement of Science, v. 360, n. 6386, p. 327–331, 2018.
- 45 WILLSEY, A. J.; STATE, M. W. Autism spectrum disorders: from genes to neurobiology. *Current opinion in neurobiology*, Elsevier, v. 30, p. 92–99, 2015.
- 46 TORRE-UBIETA, L. de la et al. Advancing the understanding of autism disease mechanisms through genetics. *Nature medicine*, Nature Publishing Group, v. 22, n. 4, p. 345, 2016.
- 47 LUO, W. et al. Systematic reconstruction of autism biology from massive genetic mutation profiles. *Science advances*, American Association for the Advancement of Science, v. 4, n. 4, p. e1701799, 2018.

- 48 SATTERSTROM, F. K. et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*, Elsevier, v. 180, n. 3, p. 568–584, 2020.
- 49 GUO, H. et al. Genome sequencing identifies multiple deleterious variants in autism patients with more severe phenotypes. *Genetics in medicine: official journal of the American College of Medical Genetics*, v. 21, n. 7, p. 1611–1620, 2019.
- 50 ROBINSON, E. B. et al. Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 111, n. 42, p. 15161–15165, 2014.
- 51 TURNER, T. N. et al. Sex-based analysis of de novo variants in neurodevelopmental disorders. *The American Journal of Human Genetics*, Elsevier, v. 105, n. 6, p. 1274–1285, 2019.
- 52 CHANG, J. et al. Genotype to phenotype relationships in autism spectrum disorders. *Nature neuroscience*, Nature Publishing Group, v. 18, n. 2, p. 191, 2015.
- 53 BUJA, A. et al. Damaging de novo mutations diminish motor skills in children on the autism spectrum. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 115, n. 8, p. E1859–E1866, 2018.
- 54 JI, X. et al. Increased burden of deleterious variants in essential genes in autism spectrum disorder. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 113, n. 52, p. 15054–15059, 2016.
- 55 LORD, C.; RISI, S.; PICKLES, A. Trajectory of language development in autistic spectrum disorders. In: *Developmental language disorders*. [S.l.]: Psychology Press, 2004. p. 18–41.
- 56 MARIN, O. Developmental timing and critical windows for the treatment of psychiatric disorders. *Nature medicine*, Nature Publishing Group, v. 22, n. 11, p. 1229, 2016.
- 57 PALMER, C. J. et al. ‘subtypes’ in the presentation of autistic traits in the general adult population. *Journal of Autism and Developmental Disorders*, Springer, v. 45, n. 5, p. 1291–1301, 2015.
- 58 HUS, V. et al. Using the autism diagnostic interview—revised to increase phenotypic homogeneity in genetic studies of autism. *Biological psychiatry*, Elsevier, v. 61, n. 4, p. 438–448, 2007.
- 59 BEGLINGER, L. J.; SMITH, T. H. A review of subtyping in autism and proposed dimensional classification model. *Journal of Autism and Developmental Disorders*, Springer, v. 31, n. 4, p. 411–422, 2001.
- 60 CONSTANTINO, J. N. et al. The factor structure of autistic traits. *Journal of Child Psychology and Psychiatry*, Wiley Online Library, v. 45, n. 4, p. 719–726, 2004.
- 61 RING, H. et al. Clinical heterogeneity among people with high functioning autism spectrum conditions: evidence favouring a continuous severity gradient. *Behavioral and Brain Functions*, BioMed Central, v. 4, n. 1, p. 1–6, 2008.
- 62 BITSIKA, V.; SHARPLEY, C.; ORAPELENG, S. An exploratory analysis of the use of cognitive, adaptive and behavioural indices for cluster analysis of asd subgroups. *Journal of Intellectual Disability Research*, Wiley Online Library, v. 52, n. 11, p. 973–985, 2008.

- 63 MUNSON, J. et al. Evidence for latent classes of iq in young children with autism spectrum disorder. *American Journal on Mental Retardation*, v. 113, n. 6, p. 439–452, 2008.
- 64 KANNE, S. M. et al. The role of adaptive behavior in autism spectrum disorders: Implications for functional outcome. *Journal of autism and developmental disorders*, Springer, v. 41, n. 8, p. 1007–1018, 2011.
- 65 WHITE, S. W. et al. Social-cognitive, physiological, and neural mechanisms underlying emotion regulation impairments: Understanding anxiety in autism spectrum disorder. *International Journal of Developmental Neuroscience*, Elsevier, v. 39, p. 22–36, 2014.
- 66 PANDOLFI, V.; MAGYAR, C. I.; DILL, C. A. An initial psychometric evaluation of the cbcl 6–18 in a sample of youth with autism spectrum disorders. *Research in Autism Spectrum Disorders*, Elsevier, v. 6, n. 1, p. 96–108, 2012.
- 67 HUS, V.; LORD, C. Effects of child characteristics on the autism diagnostic interview-revised: Implications for use of scores as a measure of asd severity. *Journal of autism and developmental disorders*, Springer, v. 43, n. 2, p. 371–381, 2013.
- 68 OUSLEY, O.; CERMAK, T. Autism spectrum disorder: defining dimensions and subgroups. *Current developmental disorders reports*, Springer, v. 1, n. 1, p. 20–28, 2014.
- 69 CHOLEMKERY, H. et al. Classifying autism spectrum disorders by adi-r: subtypes or severity gradient? *Journal of autism and developmental disorders*, Springer, v. 46, n. 7, p. 2327–2339, 2016.
- 70 SYRIOPOULOU-DELLI, C. K.; PAPAEFSTATHIOU, E. Review of cluster analysis of phenotypic data in autism spectrum disorders: distinct subtypes or a severity gradient model? *International Journal of Developmental Disabilities*, Taylor & Francis, v. 66, n. 1, p. 13–21, 2020.
- 71 MARQUAND, A. F. et al. Conceptualizing mental disorders as deviations from normative functioning. *Molecular psychiatry*, Nature Publishing Group, v. 24, n. 10, p. 1415–1424, 2019.
- 72 MARQUAND, A. F. et al. Understanding heterogeneity in clinical cohorts using normative models: beyond case-control studies. *Biological psychiatry*, Elsevier, v. 80, n. 7, p. 552–561, 2016.
- 73 WOLFERS, T. et al. Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. *JAMA psychiatry*, American Medical Association, v. 75, n. 11, p. 1146–1155, 2018.
- 74 ALNÆS, D. et al. Brain heterogeneity in schizophrenia and its association with polygenic risk. *JAMA psychiatry*, 2019.
- 75 KESSLER, D.; ANGSTADT, M.; SRIPADA, C. Growth charting of brain connectivity networks and the identification of attention impairment in youth. *JAMA psychiatry*, American Medical Association, v. 73, n. 5, p. 481–489, 2016.
- 76 WOLFERS, T. et al. Individual differences v. the average patient: mapping the heterogeneity in adhd using normative models. *Psychological medicine*, Cambridge University Press, p. 1–10, 2019.
- 77 LEFEBVRE, A. et al. Alpha waves as a neuromarker of autism spectrum disorder: the challenge of heterogeneity and reproducibility. *Frontiers in neuroscience*, Frontiers, v. 12, p. 662, 2018.

- 78 ZABIHI, M. et al. Dissecting the heterogeneous cortical anatomy of autism spectrum disorder using normative models. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, Elsevier, v. 4, n. 6, p. 567–578, 2019.
- 79 TOMA, C. Genetic variation across phenotypic severity of autism. *Trends in Genetics*, Elsevier, v. 36, n. 4, p. 228–231, 2020.
- 80 LORD, C.; RUTTER, M.; COUTEUR, A. L. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of autism and developmental disorders*, Springer, v. 24, n. 5, p. 659–685, 1994.
- 81 SPARROW, S. S. et al. *Vineland adaptive behavior scales: Survey forms manual*. [S.l.]: American Guidance Service, 2005.
- 82 PERRY, A.; CONDILLAC, R. A.; FREEMAN, N. L. Best practices and practical strategies for assessment and diagnosis of autism. *Journal on Developmental Disabilities*, v. 9, n. 2, p. 61–75, 2002.
- 83 MAZEFSKY, C. A. et al. Child behavior checklist scores for school-aged children with autism: Preliminary evidence of patterns suggesting the need for referral. *Journal of psychopathology and behavioral assessment*, Springer, v. 33, n. 1, p. 31–37, 2011.
- 84 JOLLIFFE, I. T. Principal component analysis: a beginner’s guide—i. introduction and application. *Weather*, Blackwell Publishing Ltd Oxford, UK, v. 45, n. 10, p. 375–382, 1990.
- 85 KAISER, H. F. Coefficient alpha for a principal component and the kaiser-guttman rule. *Psychological reports*, SAGE Publications Sage CA: Los Angeles, CA, v. 68, n. 3, p. 855–858, 1991.
- 86 TEAM, R. C. et al. *R: A language and environment for statistical computing*. [S.l.]: Vienna, Austria, 2013.
- 87 GEORGI, B.; VOIGHT, B. F.; BUĆAN, M. From mouse to human: evolutionary genomics analysis of human orthologs of essential genes. *PLoS Genet*, Public Library of Science, v. 9, n. 5, p. e1003484, 2013.
- 88 DEUTSCHBAUER, A. M. et al. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics*, Genetics Soc America, v. 169, n. 4, p. 1915–1925, 2005.
- 89 ZHANG, M. et al. The orphan disease networks. *The American Journal of Human Genetics*, Elsevier, v. 88, n. 6, p. 755–766, 2011.
- 90 DICKINSON, M. et al. International mouse phenotyping consortium jackson laboratoryinfrastructure nationale phenomin, institut clinique de la souris (ics) charles river laboratories mrc harwell toronto centre for phenogenomics wellcome trust sanger institute riken bioresource center high-throughput discovery of novel developmental phenotypes. *Nature*, v. 537, p. 508–514, 2016.
- 91 ZEILEIS, A. et al. Beta regression in r. *Journal of statistical software*, v. 34, n. 2, p. 1–24, 2010.
- 92 DEMYER, M. K. et al. The measured intelligence of autistic children. *Journal of Autism and Childhood Schizophrenia*, Springer, v. 4, n. 1, p. 42–60, 1974.

- 93 FARMER, C. et al. Classifying and characterizing the development of adaptive behavior in a naturalistic longitudinal study of young children with autism. *Journal of neurodevelopmental disorders*, Springer, v. 10, n. 1, p. 1, 2018.
- 94 HOWLIN, P. et al. Social outcomes in mid-to later adulthood among individuals diagnosed with autism and average nonverbal iq as children. *Journal of the American Academy of Child & Adolescent Psychiatry*, Elsevier, v. 52, n. 6, p. 572–581, 2013.
- 95 TILLMANN, J. et al. Investigating the factors underlying adaptive functioning in autism in the eu-aims longitudinal european autism project. *Autism Research*, Wiley Online Library, v. 12, n. 4, p. 645–657, 2019.
- 96 GOTHAM, K. et al. Exploring the relationship between anxiety and insistence on sameness in autism spectrum disorders. *Autism Research*, Wiley Online Library, v. 6, n. 1, p. 33–41, 2013.
- 97 HOLLOCKS, M. J. et al. Anxiety and depression in adults with autism spectrum disorder: a systematic review and meta-analysis. *Psychological medicine*, Cambridge University Press, v. 49, n. 4, p. 559–572, 2019.
- 98 NEUHAUS, E.; WEBB, S. J.; BERNIER, R. A. Linking social motivation with social skill: The role of emotion dysregulation in autism spectrum disorder. *Development and psychopathology*, Cambridge University Press, v. 31, n. 3, p. 931–943, 2019.
- 99 GEORGIADES, S. et al. Phenotypic overlap between core diagnostic features and emotional/behavioral problems in preschool children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, Springer, v. 41, n. 10, p. 1321–1329, 2011.
- 100 IBRAHIM, K. et al. Anger rumination is associated with restricted and repetitive behaviors in children with autism spectrum disorder. *Journal of autism and developmental disorders*, Springer, v. 49, n. 9, p. 3656–3668, 2019.
- 101 PATEL, S. et al. Association between anger rumination and autism symptom severity, depression symptoms, aggression, and general dysregulation in adolescents with autism spectrum disorder. *Autism*, Sage Publications Sage UK: London, England, v. 21, n. 2, p. 181–189, 2017.
- 102 PUGLIESE, C. E. et al. Longitudinal examination of adaptive behavior in autism spectrum disorders: Influence of executive function. *Journal of Autism and Developmental Disorders*, Springer, v. 46, n. 2, p. 467–477, 2016.
- 103 NORRELGEN, F. et al. Children with autism spectrum disorders who do not develop phrase speech in the preschool years. *Autism*, SAGE Publications Sage UK: London, England, v. 19, n. 8, p. 934–943, 2015.
- 104 ROSE, V. et al. The proportion of minimally verbal children with autism spectrum disorder in a community-based early intervention programme. *Journal of Intellectual Disability Research*, Wiley Online Library, v. 60, n. 5, p. 464–477, 2016.
- 105 TOMASZEWSKI, B. et al. Developmental trajectories of adaptive behavior from toddlerhood to middle childhood in autism spectrum disorder. *American Journal on Intellectual and Developmental Disabilities*, The American Association on Intellectual and Developmental Disabilities, v. 125, n. 3, p. 155–169, 2020.
- 106 LORD, C.; BISHOP, S. L. Recent advances in autism research as reflected in dsm-5 criteria for autism spectrum disorder. *Annual review of clinical psychology*, Annual Reviews, v. 11, p. 53–70, 2015.

- 107 LOMBARDO, M. V.; LAI, M.-C.; BARON-COHEN, S. Big data approaches to decomposing heterogeneity across the autism spectrum. *Molecular psychiatry*, Nature Publishing Group, v. 24, n. 10, p. 1435–1450, 2019.
- 108 GAZESTANI, V. H. et al. A perturbed gene network containing pi3k–akt, ras–erk and wnt– $\beta$ -catenin pathways in leukocytes is linked toasd genetics and symptom severity. *Nature neuroscience*, Nature Publishing Group, v. 22, n. 10, p. 1624–1634, 2019.
- 109 ALONSO-GONZALEZ, A.; RODRIGUEZ-FONTENLA, C.; CARRACEDO, A. De novo mutations (dnms) in autism spectrum disorder (asd): pathway and network analysis. *Frontiers in genetics*, Frontiers, v. 9, p. 406, 2018.
- 110 COE, B. P. et al. Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nature genetics*, Nature Publishing Group, v. 51, n. 1, p. 106–116, 2019.
- 111 GANNA, A. et al. Quantifying the impact of rare and ultra-rare coding variation across the phenotypic spectrum. *The American Journal of Human Genetics*, Elsevier, v. 102, n. 6, p. 1204–1211, 2018.
- 112 WARRIER, V. et al. Social and non-social autism symptoms and trait domains are genetically dissociable. *Communications biology*, Nature Publishing Group, v. 2, n. 1, p. 1–13, 2019.