VIVIANE NERI DE SOUZA REIS

Estudo dos componentes de vulnerabilidade genética no Transtorno do Espectro Autista

Tese apresentada à Faculdade de Medicina
da Universidade de São Paulo para obtenção do
título de Doutor em Ciências

Programa de Psiquiatria

Orientadora: Profa. Dra. Helena Paula Brentani

(Versão corrigida. Resolução CoPGr 6018/11, de 1 de novembro de 2011. A versão original está disponível na biblioteca da FMUSP)

SÃO PAULO
2019

VIVIANE NERI DE SOUZA REIS

**Estudo dos componentes de vulnerabilidade genética
no Transtorno do Espectro Autista**

Tese apresentada à Faculdade de Medicina
da Universidade de São Paulo para obtenção do
título de Doutor em Ciências

Programa de Psiquiatria

Orientadora: Profa. Dra. Helena Paula Brentani

(Versão corrigida. Resolução CoPGr 6018/11, de 1 de novembro de 2011. A versão original está disponível na biblioteca da FMUSP)

SÃO PAULO
2019

Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da
Faculdade de Medicina da Universidade de São Paulo

©reprodução autorizada pelo autor

Reis, Viviane Neri de Souza
Estudo dos componentes de vulnerabilidade
genética no Transtorno do Espectro Autista /
Viviane Neri de Souza Reis. -- São Paulo, 2019.
Tese(doutorado)--Faculdade de Medicina da
Universidade de São Paulo.
Programa de Psiquiatria.
Orientadora: Helena Paula Brentani.

Descritores: 1.Transtorno autístico 2.Fatores de
risco 3.Metilação 4.Exoma 5.Biologia computacional
6.Epigenômica 7.Genética 8.Psiquiatria

USP/FM/DBD-091/19

Responsável: Erinalva da Conceição Batista, CRB-8 6755

“The distance is nothing when one has a motive.”

—Jane Austen

AGRADECIMENTOS

À minha orientadora Dr^a Helena Brentani pelos ensinamentos, apoio e toda inspiração. À minha mãe Rinalda, meu marido Leandro e todos integrantes da minha família, e amigos queridos por sempre me incentivarem e torcerem pelo meu sucesso.

À toda equipe do ambulatório, do laboratório e do grupo PsysBio, especialmente a Ana Carolina Tahira, Vinicius Daguano Gastaldi e Caroline Camilo, por todo o apoio e ajuda durante as análises desse projeto, e escrita da tese e artigos. Ao pessoal do LIM23 e da secretaria de pós-graduação, Eliza e Isabel pelo apoio e auxílio na realização do projeto.

À toda equipe responsável pelo planejamento, organização e coleta dos dados e material biológico dos pacientes das clínicas da Universidade de São Paulo (PROTEA – HCFMUSP), Universidade Federal de São Paulo (Transtorno do Espectro do Autismo Marcos Mercadante – TEAMM) e Universidade Mackenzie (Clínica TEA - MACK).

À toda equipe do grupo liderado pelo Professor John Quackenbush, no laboratório “Computational Biology and Functional Genomics, no Dana-Farber Cancer Institute e associado à Harvard TH Chan School of Public Health, localizado em Boston, Massachusetts (USA) por me receber e auxiliar durante meu estágio de doutorado-sanduiche.

Aos pacientes e responsáveis que aceitaram participar do nosso estudo e toda à equipe que cuida diretamente dos mesmos e também à equipe que sempre esteve presente nos bastidores armazenando, avaliando, analisando e ajudando de todas as formas possíveis.

E finalmente, agradeço as agências de apoio financeiro para a realização do projeto: CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES-PROEX – Processo 1503048 e CAPES-PDSE – Processo 88881.135898/2016-01) e FAPESP - Fundação de Amparo a Pesquisa do Estado de São Paulo (Auxílio Pesquisa Processo 2011/14658-2, e Processo 2012/51584-0 em acordo especial com a Fundação Maria Cecília Souto Vidigal).

SUMÁRIO

Lista de Figuras

Lista de Tabelas

Lista de Abreviações

Resumo

Abstract

1.	INTRODUÇÃO	01
1.1.	Transtornos do espectro do autismo	01
1.2.	Variantes raras e comuns no TEA	03
1.3.	Fatores Risco Ambiental associados a TEA e epigenética	07
1.4.	Metilação enquanto medida de relação genética e ambiente	10
1.5.	Proposta de estudo	12
2.	OBJETIVOS	14
2.1.	Objetivo Principal	14
2.2.	Objetivos Específicos	14
3.	MÉTODOS	15
3.1.	Seleção da Amostra	15
3.1.1	Escalas e Instrumentos aplicados aos sujeitos da pesquisa	16
3.1.2	Variáveis e medidas do fenótipo de risco	17
3.2.	Considerações éticas	20
3.3.	Análise Estatística, de Componentes e Clusterização	20
3.4.	Análise do Material Biológico	22
3.4.1	Coleta de Sangue Periférico	22
3.4.2	Metiloma	23

3.4.3	Exoma	26
4.	RESULTADOS	29
4.1.	Caracterização amostral	29
4.2.	Análise de Componentes e Clusterização	30
4.2.1	Análise de Componentes Principais	30
4.2.2	Análise de Clusterização	32
4.3.	Análise do Metiloma	36
4.3.1	Sítios Diferencialmente Metilados (DMPs)	36
4.3.2	Enriquecimento de VMRs e Regiões Regulatórias	37
4.3.3	Módulos Epigenético Funcionais	40
4.3.4	Idade Epigenética	44
4.4.	Exoma	44
5.	DISCUSSÃO	46
6.	CONCLUSÃO	51
7.	ANEXOS	52
8.	REFERÊNCIAS BIBLIOGRÁFICAS	63

Lista de Figuras

Figura 1 – Análise de PCA.....	31
Figura 2 – Biplot dos componentes principais.....	32
Figura 3 – Análise de Clusterização usando o método Ward D em conjunto com o algoritmo pvclust.....	33
Figura 4 – Distribuição dos indivíduos dos clusters 1 e cluster 2 no biplot de PCA representando os PC1 e PC2	34
Figura 5 – Gráfico de Vulcão da distribuição dos valores de delta-beta.....	37
Figura 6 – Heatmap de DMPs enriquecidos em VMRs.....	39
Figura 7 – Módulo Funcional do gene semente ROBO3.....	41
Figura 8 – Módulo Funcional do gene semente BCL2A1.....	41
Figura 9 – Módulo Funcional do gene semente UNC119.....	42
Figura 10 – Análise da Aceleração Metilômica.....	44
Figura 11 – Esquema dos Variantes do Exoma selecionados.....	45
Figura 12 – Modelo do papel dos componentes de vulnerabilidade genéticos comuns e ambientais na fisiopatologia de TEA.....	50
Figura Anexa 1 – PCA e Clusterização com variável Exposição Tóxica.....	56
Figura Anexa 2 – Distribuição de beta values produzidos pelas sondas Infinium I e Infinium II.....	57

Lista de Tabelas

Tabela 1 – Comparação de valores dos escores de risco e das entrevistas de fenótipo entre os agrupamentos encontrados.....	35
Tabela 2 – Comparação das DMPs dos pacientes com TEA com as VMRs de 4 bancos de dados.....	38
Tabela 3 – Análise de DMPs localizadas nas regiões de elementos regulatórios do banco de dados rVarbase.....	40
Tabela 4 – Resumo dos resultados da análise de Enriquecimento da lista de genes dos Módulos Funcionais.....	43
Tabela Anexa 1 – Caracterização e escores de fenótipo dos pacientes selecionados.....	52
Tabela Anexa 2 – Escores dos fatores de risco usados no PCA	54
Tabela Anexa 3 – Módulos Funcionais diferentemente metilados.....	58
Tabela Anexa 4 – Análise de Enriquecimento da lista de genes dos módulos Funcionais diferentemente metilados.....	60

Lista de Abreviações

AA	Aceleração de Idade Epigenética
ADI-R	Entrevista diagnóstica do autismo-Revisada
AU	Aproximadamente Imparcial
BP	Probabilidade Bootstrap
CARS	<i>Childhood Autism Rating Scale</i>
CBCL	<i>Child Behavior Checklist</i>
CpG	Sítio Citosina-Guanina
DI	Deficiência Intelectual
DMP	Sítios Diferencialmente Metilados
DMR	Regiões Diferencialmente Metilados
DP	Desvio Padrão
DSM-5	Manual Diagnóstico e Estatístico de Transtornos Mentais – 5
EEG	Eletroencefalografia
EWAS	Estudo de Associação do Epigenoma Completo
GxA	Interação Gene-Ambientes
LGD	Mutações provavelmente disruptivas
LoF	Mutações com Perda de Função
MAF	Frequência do Menor Alelo
NA	Informação Ausente
PC	<i>Componente Principal</i>
PCA	Análise de componentes principais
PROTEA	Programa do Transtorno do Espectro do Autismo
QI	Quociente de Inteligência)
RVIS	Escore de tolerância a variação do gene
SNP	Polimorfismo de Nucleotídeo Único
SNV	Variação de Nucleotídeo Único
SON-R	<i>Snijders-Oomen Nonverbal Intelligence Test</i>
TCLE	Termos de Consentimento Livre e Esclarecido
TDAH	Transtorno do Déficit de Atenção e Hiperatividade
TEA	Transtorno do Espectro do Autismo
TEA-MACK	Clínica de Transtornos do Espectro do Autismo
TEAMM	Transtorno do Espectro do Autismo Marcos Mercadante
VMR	Regiões Variavelmente Metiladas
WES	Sequenciamento do Exoma Completo

Reis VNS. Estudo dos componentes de vulnerabilidade genética no Transtorno do Espectro Autista [tese]. São Paulo: Faculdade de Medicina da USP; 2019.

Resumo

O transtorno do espectro do autismo (TEA) é um transtorno do neurodesenvolvimento com apresentação clínica heterogênea. A nova classificação dimensional do DSM-5 permitiu a inclusão de toda a variabilidade fenotípica sob o mesmo guarda-chuva, criando a oportunidade de entender melhor os subgrupos de TEA de acordo com seus mecanismos fisiopatológicos heterogêneos. O objetivo deste estudo foi buscar componentes de vulnerabilidade a partir de fatores de risco (escolaridade materna, classe social, estresse e exposição tóxica ambiental durante a gestação, complicações na gravidez e história psiquiátrica familiar) e caracterizar subgrupos de TEA a partir destes componentes. Para evitar qualquer possível agrupamento baseado em parâmetros fenotípicos estabelecidos, como QI e gravidade, analisamos especificamente um grupo de pacientes homogêneos com TEA grave. A análise de componentes principais (PCA) foi realizada em dados de 68 crianças com TEA entre 3 e 7 anos de idade, e encontramos dois componentes principais: PC1, componente de vulnerabilidade genética e metabólica e PC2 componente de vulnerabilidade psicossocial. Com os escores do PCA, realizamos uma análise de *clusters*. Os resultados mostraram um *cluster* representando uma dimensão com maior vulnerabilidade genética, e outro com maior exposição a ambiente desfavorável e estressante durante a gestação. A análise de metiloma foi realizada para validar e explorar melhor a diferença entre os subgrupos. Encontramos 11.879 probes ($p < 0.05$) diferencialmente metiladas (DMPs). Os sítios CpG das DMPs estavam enriquecidos para regiões de metilação variáveis (VMRs). Sondas hipermetiladas apresentaram taxas mais altas nas características rVarBase associadas a SNPs funcionais, indicando maior risco de doença explicado por variações comuns (SNPs). A análise do módulo funcional dos promotores de genes encontrou diferenças relacionadas à resposta imune, processos metabólicos e estresse. A análise do relógio de metilação do DNA mostrou uma tendência de aumento do DNAm Age para ambos o *clusters*, mas sem diferença estatística. Por fim, a análise do exoma de 33 pacientes representantes dos dois *clusters* mostrou como esperado que ambos os subgrupos têm variantes raras deletérias, mas sem diferenças entre eles no número de variantes em genes intolerantes à variância de acordo com o escore RVIS. Nossos resultados mostram que estes grupos apresentam diferenças quanto aos componentes de vulnerabilidade uma relacionada com antecedentes genéticos hereditários comuns e, outra mais relacionada à resposta ambiental ao estresse. Este estudo corrobora que variações comuns e raras são importantes, mas influências ambientais devem ser consideradas para melhor encontrar subgrupos de TEA.

Descritores: Transtorno autístico; Fatores de risco; Metilação; Exoma; Biologia computacional; Epigenômica; Genética; Psiquiatria

Reis VNS. [*Study of the components of genetic vulnerability in Autism Spectrum Disorder*] [*Thesis*] São Paulo: Faculdade de Medicina da USP; 2019. Portuguese.

Abstract

Autism spectrum disorder (ASD) is a neurodevelopmental disorder with highly heterogeneous clinical presentation. The new dimensional DSM-5 classification allowed the inclusion of all phenotypic variability under the same umbrella, creating the opportunity to better understand ASD subgroups according to its heterogeneous pathophysiology mechanisms. The aim of this study was to search components of vulnerability from risk factors during gestation (mother schooling, social class, stress and environmental toxic exposition during gestation, pregnancy complications and familial psychiatric history) e characterize ASD subgroups from those components. To avoid any possible grouping based on phenotypic established parameters such as IQ and severity, we analysed a group of homogeneous patients with severe ASD specifically. Principal component analysis (PCA) was performed on data from 68 children with ASD between 3 and 7 years of age, and we found two principal components: PC1, component of genetic and metabolic vulnerability e PC2 component of unfavorable social environment vulnerability. With the PCA scores we performed clustering analysis. The results showed one cluster representing a dimension with stronger genetic vulnerability, and the other with more exposure to unfavorable and stressful environment during gestation. Methylome analysis has been performed to better explore subgroup difference. We found 11.879 ($p < 0.05$) differentially methylated probes (DMPs). CpG sites from those DMPs were found to be enriched in variable methylated regions (VMRs). The clusters have hypermethylated probes presented higher rates in different rVarBase regulatory regions associated to functional SNPs, indicating they may have different affected regulatory regions and more liability to disease explained by common variations (SNPs). Functional module analysis on gene promoters found differences related to immune response, metabolic processes and stress. DNA methylation clock analysis showed a tendency of higher DNAm Age for both Clusters, but here was no statistical DNAm Age acceleration difference. Lastly exome analysis of 33 patients representing both clusters showed as expected that both subgroups have deleterious rare variants, but without differences between them in the number of variants in genes intolerant to variance according to RVIS score. Our results show that this groups presents differences of vulnerability components, one related to common hereditary genetic antecedents, and another more related to the environmental response to stress. This study corroborates that common and rare variants are important, but environmental influences should be considered to better find subgroups of ASD.

Descriptors: Autistic disorder; Risk factors; Methylation; Exome; Computational biology; Epigenomics; Genetics; Psychiatry.

1. INTRODUÇÃO

1.1 Transtornos do espectro do autismo

O transtorno do espectro do autismo (TEA) é um transtorno do neurodesenvolvimento caracterizado pelo prejuízo da comunicação social recíproca / interação social e comportamentos repetitivos, estereotipados ou interesses restritos (American Psychiatric Association 2013). Existe uma grande heterogeneidade na apresentação fenotípica (Jeste e Geschwind 2014) e, em geral, há um claro desequilíbrio na prevalência de TEA entre meninos e meninas (4:1) (Werling e Geschwind 2013).

Raramente dois pacientes com TEA têm sintomas idênticos e fatores como a severidade dos sintomas, a habilidade de linguagem e o QI (Quociente de Inteligência) complicam ainda mais o diagnóstico. Sendo assim, diferentes apresentações clínicas e níveis de incapacidades são vistos no TEA (Lai *et al.* 2013; Ousley e Cermak 2014). Cerca de 70% das crianças com TEA têm alguma comorbidade e 48% têm duas ou mais (Simonoff *et al.* 2008). A deficiência intelectual (DI) é observada em 70% dos casos (Fombonne 2003), epilepsia em 7% a 46% (Lo-Castro e Curatolo 2013) e o TEA é mais frequente em pacientes com DI (21,4%) do que em pacientes sem DI (8%) (Amiet *et al.* 2008). Transtorno do déficit de atenção e hiperatividade (TDAH) se apresentam em cerca de 30% dos casos de TEA (Simonoff *et al.* 2008); depressão em 2% a 30% dos casos (Leyfer *et al.* 2006; Matson e Nebel-Schwalm 2007); e ansiedade em 2% a 45% (Leyfer *et al.* 2006; Simonoff *et al.* 2008).

Muitos estudos tentaram identificar subgrupos de TEA, com o objetivo de compreender os diferentes mecanismos biológicos, desfechos clínicos, respostas ao tratamento e o desenvolvimento de uma abordagem personalizada para o tratamento de TEA (Miles *et al.* 2005; Munson *et al.* 2008; Grzadzinski *et al.* 2013; Cholemkery *et al.* 2016). Alguns desses estudos usaram fenótipos para agregar indivíduos, tais como os escores de QI (Munson *et al.* 2008; Ousley e Cermak 2014) e ADI-R (do inglês *Autism diagnostic interview–Revised*, Entrevista diagnóstica do autismo-Revisada) (Grzadzinski *et al.* 2013), confirmando o conceito bem conhecido de TEA de alto e baixo funcionamento. Seguindo essa linha de pensamento, Miles *et al.* (2005) definiram como TEA complexo um subgrupo de indivíduos com evidência de uma anormalidade ocorrida no início da morfogênese, manifestada por anomalias menores (dismórficas) ou microcefalia. Esse subgrupo compreende cerca de 20-30% do total da população com TEA estudada. Indivíduos com autismo complexo apresentam piores resultados como baixo QI, mais convulsões, mais alterações em Eletroencefalografia (EEG) (46% vs. 30%) e mais anormalidades cerebrais encontradas em ressonância magnética (28% vs.13%).

As diferenças fenotípicas de TEA e a complexidade genética parecem representar múltiplos caminhos suficientemente ligados para garantia de um diagnóstico comum, mas diferentes o suficiente para resultar em uma ampla gama de prognóstico e resposta ao tratamento (Iossifov *et al.* 2008).

Recentemente, a literatura médica concordou que o autismo é um espectro com diferentes apresentações clínicas e níveis de deficiência (Lai *et*

al. 2013; Ousley e Cermak 2014). O DSM-5 (Manual Diagnóstico e Estatístico de Transtornos Mentais – 5) substituiu o sistema multi-categórico por uma única dimensão diagnóstica, denominada transtorno do espectro do autismo, que representa a ideia de que suas características caem ao longo de um *continuum de* gravidade (American Psychiatric Association 2013; Grzadzinski *et al.* 2013; Cholemkery *et al.* 2016). Essa classificação contempla a inclusão de toda a variabilidade fenotípica e estimula os pesquisadores a buscarem por subgrupos de acordo com diferentes mecanismos biológicos associados ao TEA (Grzadzinski *et al.* 2013; Loke *et al.* 2015).

Fernandez e Scherer (2017), ao revisar a classificação clínica em autismo essencial e complexo baseado no padrão de anormalidades somáticas, proposta em 2005 por Miles *et al.* (Miles *et al.* 2005), recomendam que a designação *sindrômica versus* não *sindrômica* seja substituída pela classificação do TEA de acordo com sua etiologia genética. Isso reforça a importância da busca por uma melhor compreensão do papel de padrões genômicos considerando diferentes fenótipos.

1.2 Variantes raras e comuns no TEA

O histórico de estudos genéticos vêm confirmando que TEA é um transtorno multifatorial, poligênico e com arquitetura genética complexa (State e Levitt 2011; Devlin e Scherer 2012; Ronemus *et al.* 2014; Iossifov *et al.* 2015; Krumm *et al.* 2015). Variantes raras preditas a alterar a função de proteínas (não sinônimas e deletérias) foram observados mais frequentemente em casos

de TEA do que em seus irmãos não afetados ou controles saudáveis, e vários estudos sugerem que mutações *de novo* que deletam regiões gênicas de *splicing* ou truncam proteínas podem contribuir para o desenvolvimento de TEA em 15% a 30% de casos (Klei *et al.* 2012; Fernandez e Scherer 2017)

Os primeiros estudos de variações raras foram realizados buscando associar variações no número de cópias (CNVs) ao TEA. Esses estudos revelaram que CNVs raras estão enriquecidas em TEA quando comparados com controles saudáveis ou membros da família não afetados (Sebat *et al.* 2007; Marshall *et al.* 2008; Glessner *et al.* 2009; Pinto *et al.* 2010; Sanders *et al.* 2011). Vale notar que várias das mesmas CNVs também foram associadas com o aumento do risco de outros transtornos do neurodesenvolvimento, como esquizofrenia e epilepsia (Grayton *et al.* 2012) e que CNVs interrompendo um mesmo agrupamento de genes apresentam fenótipos mais similares do que o esperado (Andrews *et al.* 2015).

Ainda na busca por entender melhor o papel de variações raras, estudos de sequenciamento do exoma completo (WES) revelaram que SNVs (do inglês *Single Nucleotide Variations*, variantes de nucleotídeo único) *de novo*, não-sinônimas e *indels* (pequenas deleções e duplicações) são associados ao TEA (O’Roak *et al.* 2011; Iossifov *et al.* 2012; Sanders *et al.* 2012). Existe um aumento da taxa de mutações *de novo* com perda de função LoF (do inglês *loss of function*) e mutações provavelmente disruptivas tipo LGD (do inglês *likely gene disruption*) nos casos de TEA simplex (incidência familiar única). Mesmo entre os genes no qual há forte evidência de aumento da taxa de mutações *de novo*, uma mutação LGD pode não ser suficiente para

desenvolver autismo, sugerindo penetrância reduzida ou variabilidade de expressividade (Guo *et al.* 2018). Outro ponto importante é que, a presença de variações iguais em diferentes portadores é muito rara, uma vez que a frequência populacional desse tipo de alteração é baixa. Genes que carregam múltiplas mutações tipo LoF *de novo* são funcionalmente heterogêneos e é estimada a existência de 400 a 1.000 genes que causam algum risco para TEA devido às mutações desse tipo (O’Roak *et al.* 2012; Sanders *et al.* 2012; Iossifov *et al.* 2014, 2015).

Guo *et al.* (2018), mostraram que entre os genes de risco para TEA, principalmente os associados a variantes *de novo*, pais portadores de mutações deletérias em famílias de pacientes com TEA apresentam mais frequentemente fenótipos mais brandos. Em famílias onde os pacientes carregavam variantes *de novo* em dois ou mais genes de risco, eles parecem apresentar sintomas mais graves. Essas observações fornecem suporte para um modelo multifatorial do risco em TEA e reforça que o modelo monogênico de doença é muito simplista, mesmo nos casos causados por mutações de maior efeito. Eles ainda afirmam que o modelo apresentado é distinto do poligênico, pois propõe que um número relativamente pequeno de variantes raras ou *de novo* de grande efeito podem ser os principais responsáveis pela etiologia da doença e gravidade fenotípica. Porém, o resultado final ainda pode ser influenciado por outros fatores, como variantes comuns, ambiente ou estocasticidade durante o desenvolvimento.

Recentemente está cada vez mais estabelecido que variações comuns (Frequência do Menor Alelo, MAF > 1%) representam um importante papel

para vulnerabilidade ao TEA (Hallmayer *et al.* 2011; Klei *et al.* 2012), apesar do efeito individual das variações comuns ser pequeno (*odds-ratio* estimado < 1.2) (Anney *et al.* 2012). Segundo Gaugler *et al.* (2014), o maior componente de risco genético deriva de variantes genéticas comuns (a hereditariedade é de aproximadamente 54%) e uma contribuição menor da variação hereditária é devida a variantes *de novo* e raras (contribuição de 2,6% de variação no passivo). Weiner *et al.* (2017), mostraram que a variação poligênica contribui de forma aditiva para o risco nos casos de TEA que carregam uma variante *de novo* de alta penetrância, mas essa contribuição ocorre em um subconjunto limitado de casos. Nos últimos anos, as influências poligênicas e *de novo* comuns sobre o risco de TEA foram cada vez mais bem caracterizadas, particularmente em termos da distribuição de seus efeitos fenotípicos (Weiner *et al.* 2017; Hannon *et al.* 2018).

O modelo poligênico assume que existem muitas variantes genéticas herdadas que contribuem para o TEA, cada uma com um pequeno efeito que, de forma aditiva, resulta em um indivíduo desenvolver o TEA (Loke *et al.* 2015; de la Torre-Ubieta *et al.* 2016a). Esse modelo é apoiado por diversos aspectos, como o aumento do risco de autismo em mais de dez vezes em um irmão *versus* o risco populacional, e a observação de traços autísticos em parentes de primeiro grau daqueles afetados por TEA (Geschwind e State 2015). Não apenas o TEA, mas muitos transtornos psiquiátricos são mais comuns entre familiares de crianças com TEA (Daniels *et al.* 2008; Jokiranta *et al.* 2013).

Além disso, estudos de família apóiam a hipótese de um fator genético ‘*p*’ em que as influências genéticas na psicopatologia parecem ser gerais entre

os transtornos psiquiátricos e não específicas de cada um. Pesquisas recentes encontraram correlações genéticas substanciais entre transtornos, como Esquizofrenia e Transtorno Bipolar. A sobreposição genética entre os construtos internalizante e externalizante também foi observada, consistente com a hipótese de um fator 'p' geral (Selzam *et al.* 2018).

1.3 Fatores de Risco Ambiental associados ao TEA e epigenética

Uma grande quantidade da heterogeneidade do TEA deve-se, provavelmente, às suas múltiplas fisiopatologias, relacionadas à interação entre genética e ambiente (Loke *et al.* 2015; Kubota e Mochizuki 2016; Sealey *et al.* 2016; Li *et al.* 2018), ou seja, um processo complexo no qual a carga genética do indivíduo interage com a carga epigenética durante o ambiente pré-natal. Segundo revisão por Ciernia e LaSalle (2016), a metilação do DNA têm mostrado um papel regulatório para manter a estabilidade genômica, e considerando as grandes mudanças metilômicas que ocorrem durante o desenvolvimento fetal, essas alterações podem ter um grande potencial de afetar o desenvolvimento cerebral e de todo organismo. Eles completam que uma das formas que a metilação do DNA pode afetar a função cerebral está ligada à variação genética em fatores de transcrição, proteínas de ligação a metil e outras proteínas reguladoras e regiões de DNA, e que a combinação de variantes genéticas hereditárias específicas pode tornar alguns indivíduos mais suscetíveis a fatores de risco ambientais.

O estresse materno pode ser considerado um fator de risco importante para o prejuízo do desenvolvimento fetal, por exemplo, o estresse psicossocial

da mãe (Palma-Gudiel *et al.* 2015), e a depressão materna durante a gravidez pode causar alteração da metilação no gene *BDNF* (*Brain-derived neurotrophic factor*) em meninos (Braithwaite *et al.* 2015). A exposição ao estresse e o ambiente social desfavorável durante a gestação são fatores importantes não somente no início, mas também no final da gestação e no período pós-natal. Kenney *et al.* (2008) demonstraram que crianças que foram expostas a tempestades tropicais durante 5-6 meses ou no último mês de gestação apresentaram risco 3,83 vezes maior de desenvolver autismo do que aquelas com tempo de exposição diferente. O eixo hipotalâmico-pituitário-adrenal fetal humano passa a funcionar na 22ª semana de gestação (próximo ao 5º mês de gestação). Assim, a partir deste ponto em desenvolvimento, as experiências maternas de estresse têm potencial para alterar permanentemente a fisiologia de seus filhos (Palma-Gudiel *et al.* 2015).

Fatores ambientais de risco para TEA podem ser: (i) complicações obstétricas relacionadas com hipóxia ou estresse respiratório (Gardener *et al.* 2009; Froehlich-Santino *et al.* 2014); (ii) hospitalização materna no primeiro trimestre gestacional, indicando exposição intraútero a eventos inflamatórios devidos a infecção viral (Atladóttir *et al.* 2007; Patterson 2011; Atladottir *et al.* 2012; Lee *et al.* 2015), sendo que aproximadamente 10% das mães de autistas apresentam anticorpos que se ligam ao cérebro fetal (Croen *et al.* 2008; Singer *et al.* 2008); (iii) exposição à drogas, poluentes e xenobióticos (Bescoby-Chambers *et al.* 2001; Lyall *et al.* 2012; Keil e Lein 2016). Além disso, o ambiente socio-econômico desfavorável e baixa escolaridade materna também

aparecem como uma fonte de risco às crianças com TEA (Delobel-Ayoub *et al.* 2015), e pode estar relacionado com resposta inflamatória (Miller *et al.* 2009).

Muitos fatores ambientais, conhecidos por estarem envolvidos na patogênese de TEA, podem alterar o *status* epigenético e a expressão gênica (Loke *et al.* 2015; Sealey *et al.* 2016) e estudos têm demonstrado que fatores de risco ambientais, como exposição a produtos químicos tóxicos e estresse no início da vida, podem alterar a metilação do DNA (DNAm), o que pode afetar criticamente a expressão gênica do mecanismo regulador no desenvolvimento do cérebro e TEA (Ciernia e LaSalle 2016; Keil e Lein 2016; Kubota e Mochizuki 2016). Outro fator ambiental provavelmente envolvido com TEA é a obesidade materna (Moss e Chugani 2014; Reynolds *et al.* 2014; Walker *et al.* 2015a), embora não esteja claro se o peso materno antes da gravidez ou o ganho de peso excessivo durante a gravidez é que está associado a maior risco para o desenvolvimento da doença (Bilder *et al.* 2013a). E, também foram encontradas associações entre distúrbios metabólicos maternos (diabetes gestacional, hipertensão e pré-eclâmpsia) e o risco para TEA (Dodds *et al.* 2011; Krakowiak *et al.* 2012; Lyall *et al.* 2012).

A combinação de dados de metilação e de análises do transcriptoma do cérebro de casos de TEA corroboram com a proposta da mediação epigenética das influências ambientais no TEA (Newschaffer *et al.* 2002; Schanen 2006; Sullivan *et al.* 2012; Grayson e Guidotti 2016). Esses estudos mostram uma sobreposição entre genes, sendo que os genes caracterizados por hipometilação do DNA têm sua expressão aumentada (Voineagu e Eapen 2013; Nardone *et al.* 2014; Stoner *et al.* 2014), evidenciando que a modulação

epigenética pode ser particularmente importante para as mudanças na formação laminar do córtex cerebral e aumento da resposta imune observada nos TEA (Mor *et al.* 2015).

Muitos dos genes relacionados à resposta imune estão envolvidos tanto na especificação de destino de células da microglia quanto na poda sináptica durante o desenvolvimento cerebral (Zhan *et al.* 2014; Sanders *et al.* 2015; Habela *et al.* 2016; Li *et al.* 2018). Nesse cenário, os mecanismos epigenéticos poderiam ser a "ponte" entre as influências ambientais e o fenótipo por meio da regulação da expressão de genes (Shulha 2012; Stein 2012; Geaghan e Cairns 2014; Berko e Greally 2015). As alterações epigenéticas apontam não somente para os processos biológicos já associados com as variações genéticas de TEA, mas também incluem a "desregulação imune" no desenvolvimento de TEA.

1.4 Metilação do DNA como medida da relação genética e ambiente

Alterações da sequência primária de DNA não são o único tipo de variações genômicas que ocorrem entre os seres humanos. Em particular, existem agora exemplos bem documentados de marcas epigenéticas, tais como metilação do DNA e modificações de histonas, que mostram significativa variação interindividual. As assinaturas de DNAm podem ser um biomarcador útil para identificar indivíduos em risco (Schübeler 2015; Hachiya *et al.* 2017).

Estudos recentes mostraram que os níveis de metilação podem estar sob controle genético local (Teh *et al.* 2014; Hannon *et al.* 2016; Andrews *et al.*

2017; Hachiya *et al.* 2017; Schröder *et al.* 2017; Garg *et al.* 2018; Islam *et al.* 2019). Hachiya *et al.* (2017), propuseram que o foco em locais de CpG com alta variabilidade no nível de metilação do DNA (polimorfismos epigenéticos) aumentaria o poder em estudos de associação do epigenoma completo (EWAS). Garg *et al.* (2018), mostraram centenas de regiões variavelmente metiladas (VMR, do inglês *Variable methylated regions*), que são regiões com níveis de metilação do DNA altamente polimórficos localizadas próximas a outras variações genômicas funcionais e enriquecidas com sítios CpGs que influenciam a expressão gênica. Esses níveis variáveis de metilação influenciados por variantes genéticas e alterados pelo ambiente estão ligados à vulnerabilidade da doença, e parecem ser um importante objeto de estudo quando se busca a compreensão de fatores genéticos e ambientais sobre a etiologia molecular de doenças complexas (Hachiya *et al.* 2017).

Além disso, a metilação do DNA tem a capacidade de avaliar o impacto de fatores ambientais que medem o trabalho cumulativo realizado pelo sistema de manutenção epigenética (Horvath e Raj 2018). Este mecanismo, conhecido como "idade epigenética", tem o potencial para avaliar a idade biológica dos indivíduos com base nos níveis de metilação do DNA. A diferença entre a idade epigenética e cronológica (conhecida como aceleração da idade, denotada AA) tem demonstrado ter significância biológica, influenciando também fenótipos (Horvath 2013; Simpkin *et al.* 2016).

Sendo assim, o quadro atual em TEA mostra fatores genéticos de novo e muito raras fortemente associados à TEA e diferenças no fenótipicas entre indivíduos. E que fatores ambiental alteram a epigenética dos indivíduos e

essas alterações levam ao aumento do risco do desenvolvimento de TEA. Nesse contexto, a hipótese vigente propõe a existência de duas extremidades de uma fisiopatologia complexa de TEA, na qual alguns indivíduos têm um risco genético mais forte e, portanto, requerem menos alterações ambientais e epigenéticas para superar o limiar necessário e alterar o neurodesenvolvimento, e de que outros foram expostos durante a gestação à fatores ambientais que eram fortes o suficiente para alterar o epigenoma da prole, sem a presença de um risco genético (Vaiserman 2015).

Porém, ainda não foi investigado como fatores ambientais interagem com fatores genéticos comuns, alterando a epigenética. E se essa interação está relacionada a variância fenotípica.

1.5 Proposta de estudo

A hipótese do presente estudo é que componentes de vulnerabilidade composto por fatores de risco genético comum e ambiental durante a gestação podem separar subgrupos de TEA, independente de fenotipo. Diante desta hipótese, nosso objetivo foi verificar se existem componentes de vulnerabilidade que formem subgrupos de pacientes com TEA a partir de dados de risco genético e exposição ambiental.

Para tal, realizamos uma análise de componentes principais para redução de dados buscando identificar componentes que explicavam a maior variabilidade dos dados. A partir destes componentes poderíamos encontrar subgrupos de pacientes com cargas de vulnerabilidade diferente. Para evitar

qualquer possível agrupamento baseado em parâmetros fenotípicos estabelecidos, como QI e gravidade, analisamos especificamente um grupo de pacientes homogêneos com deficiências graves, QI mais baixo (QI <75) e sem síndrome genética clínica.

Quanto à essa população, como é esperado encontrar um componente genético representado por variações *de novo* e muito raras, de grande efeito, realizamos uma análise de exoma em uma sub-amostra para podermos verificar se subgrupos encontrados não seriam uma consequência desse componente genético. Levando em conta os possíveis efeitos de variação comum atuando nas regiões reguladoras do genoma e fatores de risco ambientais mediados por fatores epigenéticos, realizamos análise de metilação do DNA do genoma para validar e explorar melhor as diferenças entre os grupos.

2. OBJETIVOS

2.1 Objetivo Principal

Verificar se existem componentes de vulnerabilidade que formem subgrupos de pacientes com TEA a partir de dados de risco genético e exposição ambiental.

2.2 Objetivos Específicos

- a) Identificar componentes de vulnerabilidade a partir de fatores de risco e verificar se é possível formar subgrupos independentes do fenótipo;
- b) Verificar se os subgrupos formados não são consequência da presença de variantes de novo e ou muito raras de grande impacto;
- c) Identificar se há diferenças epigenéticas entre os grupos formados através da análise diferencial de metilação de DNA; E se as possíveis diferenças são enriquecidas para VMR;
- d) Verificar se existe diferença da aceleração da idade entre os subgrupos;

3. MÉTODOS

3.1 Seleção da amostra

Os pacientes foram selecionados entre três clínicas especializadas em diagnóstico e tratamento de TEA em São Paulo, Brasil: O Programa do Transtorno do Espectro do Autismo (PROTEA) do Instituto de Psiquiatria do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, o Transtorno do Espectro do Autismo Marcos Mercadante (TEAMM) da Universidade Federal de São Paulo; e a Clínica de Transtornos do Espectro do Autismo (TEA-MACK) da Universidade Presbiteriana Mackenzie.

Os indivíduos incluídos no estudo foram diagnosticados com TEA de acordo com a *Autism Diagnostic Interview* (ADI-R) (22). Eles também passaram por avaliação clínica por uma equipe multidisciplinar de especialistas com base no Manual Diagnóstico e Estatístico de Transtornos Mentais V (DSM-5) (American Psychiatric Association 2013). Todos tinham de 3 a 7 anos incompletos e o QI entre 50 e 75, avaliado pelo SON-R 2½-7, instrumento padronizado não verbal validado no Brasil (Karino *et al.* 2011). Também foi verificado sua funcionalidade utilizando-se a técnica de *Vineland* (Sparrow e Cicchetti 1985), gravidade de TEA com os instrumentos diagnósticos CARS (do inglês *Childhood Autism Rating Scale*) (Chlebowski *et al.* 2010) e ADI-R (Becker 2009), e os sintomas internalizantes e externalizantes utilizando o CBCL (do inglês *Child Behavior Checklist*) (Doyle *et al.* 2010; Bordin *et al.* 2013).

As crianças e suas famílias foram divididas em 4 grupos que foram avaliados em dias diferentes, sempre com a mesma equipe. Cada paciente, acompanhado por seu cuidador principal, passou de 5 a 6 horas de avaliação clínica completa. O cuidador respondeu a um questionário sociodemográfico que incluía histórias pessoais, familiares e gestacionais detalhadas, com descrição completa dos medicamentos e uso de medicamentos e outras exposições durante a gestação.

3.1.1 Escalas e instrumentos aplicados aos sujeitos da pesquisa

Childhood Autism Rating Scale (CARS): É uma escala de avaliação para observação comportamental formada por 15 itens. A versão atual foi traduzida e validada para o português pelo grupo de Neuropediatria do Hospital de Clínicas de Porto Alegre (Pereira *et al.*, 2008) denominada CARS-BR, utiliza os critérios do DSM-IV. A CARS classifica os paciente segundo a gravidade, sendo assim um instrumento adequado para avaliação dimensional.

Vineland: Entrevista dirigida realizada com o cuidador que avalia o nível de independência pessoal e social dos indivíduos (Sparrow e Cicchetti 1985).

Child Behavior Checklist (CBCL): Questionário que avalia habilidades sociais e problemas de comportamento, utilizando informações fornecidas pelos pais (Achenbach e Dumenci 2001; Bordin *et al.* 2013) e foi elaborado para indicar perfis que possam sugerir a situação de desenvolvimento da criança, não sendo válido automaticamente como diagnóstico. Tradução e validação no Brasil foram feita por Bordin *et al.* (2015). Apresenta dimensões do

comportamento infantil que podem ser aglutinadas em comportamentos internalizantes e externalizantes, assim como escores segundo o DSM IV que caracterizam perfis associados a problemas afetivos, problemas de ansiedade, problemas pervasivos do desenvolvimento, problemas relacionados a transtorno de atenção e hiperatividade e oposição e conduta.

Snijders-Oomen Nonverbal Intelligence Test - Revised (SON-R 2½-7): consiste em seis subtestes: Categorias, Analogias, Situações, Histórias, Mosaicos e Padrões. Avalia um espectro de habilidades cognitivas que não exigem o uso da fala e da linguagem escrita. Este teste vem para suprir uma dificuldade que enfrentamos com crianças que não são verbais e precisam de avaliação cognitiva (Ribeiro de Jesus 2009).

3.1.2 Variáveis e medidas do fenótipo de risco

De acordo com o descrito na literatura, os seguintes fatores de risco para o TEA foram coletados e utilizados para as análises formando os seguintes escores:

- a) *Classe Social*: A classe social familiar (Miller *et al.* 2009; Delobel-Ayoub *et al.* 2015), mensurada pelo Critério Brasil (ABEP 2016), questionário validado no Brasil. A pontuação da classe social familiar foi: nenhum ponto para as mães das classes A e B, 1 ponto para C e 2 pontos para D.
- b) *Escolaridade materna*: A pontuação da escolaridade materna (Delobel-Ayoub *et al.* 2015) foi 1 para o ensino superior completo, 2 para o ensino

incompleto, 3 para o ensino médio completo, 4 para o ensino médio incompleto, 5 para o ensino básico completo, 6 para o ensino fundamental incompleto e 7 para nenhum tipo de educação.

- c) Estresse materno: Referidos como estresse durante a gestação (Kinney *et al.* 2010; Babenko *et al.* 2015) foram pontuados: Sintomas psiquiátricos na gestação, agressão física na gestação, agressão verbal na gestação, estresse durante a gestação, conflitos familiares durante a gestação, doença de familiar durante a gestação e/ou morte de familiar durante a gestação (Braithwaite *et al.* 2015; Palma-Gudiel *et al.* 2015). A presença de cada um deles representou 1 ponto no escore final que variou de 0 (sem estresse) a 7 (presença de todos os itens acima).
- d) Exposições ambientais tóxicas: Tabaco, uso de álcool e drogas, bem como exposição a agrotóxicos e uso de medicamentos deletérios conhecidos durante a gravidez, por exemplo, diazepam, ácido valpróico, fluoxetina e sertralina (Kalkbrenner *et al.* 2014; Ornoy *et al.* 2016; Sealey *et al.* 2016), foram agregados em um escore. A presença de cada um deles representou 0-2 pontos, dependendo da frequência e quantidade de exposição, e a soma total representou uma pontuação final de 0 a 10.
- e) Risco genético familiar: Todos os transtornos psiquiátricos relatados das famílias do probando foram combinados em um único escore (Daniels *et al.* 2008; Frans *et al.* 2013; Jokiranta *et al.* 2013; Agerbo *et al.* 2015). As patologias listadas foram ASD, esquizofrenia, depressão, transtorno bipolar, abuso de álcool e / ou drogas, transtorno obsessivo-compulsivo, ansiedade, déficit de atenção e hiperatividade. Seguindo a teoria poligênica (Agerbo *et al.* 2015; Maher 2015; de la Torre-Ubieta *et al.*

2016b; Hannon *et al.* 2018), utilizamos a combinação de todos os transtornos para representar o risco poligênico do transtorno psiquiátrico. Cada pontuação foi balanceada de acordo com o grau de proximidade familiar ao probando. Os familiares de primeiro grau tinham peso 3, segundo grau 2 e terceiro grau 1. Também utilizamos a confiabilidade das informações para ponderar o escore: se um membro da família tivesse sido hospitalizado por transtorno psiquiátrico (peso 3), se tivesse usado medicações ou se tivesse do transtorno psiquiátrico (peso 2), e se não tivesse nenhum dos itens acima (peso 1). Um exemplo: o sujeito 18 tinha uma mãe com depressão e um tio com esquizofrenia. A mãe não usou medicação e nunca teve tratamento para a condição, por isso foi considerada de baixa confiabilidade, porém o tio havia sido internado por causa de sua desordem. Assim, a pontuação final para ID 18 era 3 (parente de 1º grau = mãe com depressão) $\times$ 1 (baixa confiabilidade) $+ 2$ (parente de 2º grau = tio com esquizofrenia) $\times$ 3 (alta confiabilidade) $= 9$. A pontuação final variou de 0 a 18 .

- f) Complicações gestacionais: Eclâmpsia / pré-eclâmpsia, diabetes gestacional e excesso de ganho de peso durante a gestação (mais de 12 kg) foram usados para formar a pontuação (Bilder *et al.* 2013b; Walker *et al.* 2015; Ornoy *et al.* 2016; Krakowiak *et al.* 2017). A presença de cada um deles representou 1 ponto no escore final que variou de 0 a 3.

3.2 Considerações Éticas

Este projeto envolve coleta de sangue para análises genéticas em pacientes com TEA e foi submetido e aprovado na comissão de ética em pesquisa do Hospital das Clínicas da Faculdade de Medicina (CAPPesq), juntamente com os Termos de Consentimento Livre e Esclarecido (TCLE) (CAAE: 57067016.2.0000.0068, Pacerer nº 1.637.312). A participação no estudo ocorreu somente mediante autorização dos responsáveis das crianças, através do TCLE no qual é garantido plena liberdade para interromper a participação quando o indivíduo desejar, sem implicar perdas ou prejuízos no atendimento prestado aos pacientes no serviço. Nesse documento estão explicitados objetivos, justificativas, riscos e desconfortos desta pesquisa.

Todas as informações e dados colhidos serão mantidos sob sigilo, os indivíduos serão identificados apenas por um código numérico, usados apenas para fins desta pesquisa. Caso os dados da literatura não sejam reproduzidos nesta amostra, será feita tentativa de publicar esses dados negativos para informar os pesquisadores e a população sobre os achados.

3.3 Análise Estatística, de Componentes e Clusterização

O *Softwares R Studio* foi utilizados para todas as análises estatísticas, análises de componentes e clusterização. Para obter comparabilidade entre as variáveis, os escores foram padronizados usando escores z .

A análise de componentes principais (PCA, do inglês “*Principal Component Analysis*”) por correlação usando a função “prcomp” do R usando as seguintes variáveis: classe social, escolaridade materna, estresse materno, exposição ambiental tóxica, risco genético familiar e complicações gestacionais. PCA é um algoritmo matemático que reduz a dimensionalidade dos dados mantendo a maior parte da variação do conjunto de dados e identificando componentes principais (vetores) ao longo das quais a variação nos dados é máxima. Usando esses componentes principais representantes da variabilidade dos dados é possível que as amostras sejam plotadas num gráfico e permite avaliar visualmente semelhanças e diferenças entre amostras, e determinar se as amostras podem ser agrupadas (Ringnér 2008). Apenas componentes que contabilizaram variâncias maiores que 1 (*eigenvalue* > 1 ou 70% da variância) foram incluídos.

Foi aplicado o método não supervisionado de clusterização hierárquico usando método Ward D (função “*ward.D2*”) e a distância euclidiana ao quadrado como medida de proximidade. Para realizar essa clusterização foi usado o algoritmo “*pvclust*”, que é um pacote R para avaliar a incerteza na análise hierárquica de clusters 1 . Para cada cluster é calculado p-valores através de reamostragem de bootstrap multiescala. O valor P de um cluster é um valor entre 0 e 1, que indica a intensidade com que o cluster é suportado pelos dados. O *pvclust* fornece dois tipos de valores p: valor p (AU) (Aproximadamente Imparcial, do inglês *Approximately Unbiased*) e valor BP (Probabilidade Bootstrap, do inglês *Bootstrap Probability*). O valor p de AU é calculado por reamostragem de bootstrap multiescala, uma aproximação

melhor do valor de p não enviesado do que o valor de BP, calculado por reamostragem de bootstrap normal.

Para comparar os clusters encontrados, foram realizados os testes estatísticos qui-quadrado para variáveis categóricas, e Wilcoxon Mann-Whitney para idade, escores de vulnerabilidade e nos escores fenotípicos (QI, Vineland, CARS, CBCL e ADI-R).

3.4 Análise do Material Biológico

3.4.1 Coleta de Sangue Periférico

Após concordância da família e assinatura do TCLE, foi realizada a extração automatizada de DNA e RNA, a partir de sangue periférico, com o uso do Instrumento *Maxwell*® 1 (que usa partículas magnéticas para captura, lavagem e purificação das moléculas de DNA genômico), seguindo o protocolo do fabricante (Promega Corporation, Madison, Wisconsin, EUA). A qualidade do DNA extraído foi verificado por espectrofotometria (NanoDrop® ND-1000, ThermoScientific, Waltham, Massachusetts, EUA), afim de garantir o uso de amostras com razão de pureza de, no mínimo 1,8. A quantificação do DNA foi realizado pelo instrumento *Qubit*® com o kit *dsDNA BR Assay* (Life Technologies Corporation, Carlsbad, Califórnia, EUA), de acordo com as instruções do fabricante.

3.4.2 Metiloma

Para a análise de metilação, o DNA coletado foi tratado usando o kit *EZ DNA Methylation* (Zymo Research Corp., Irvine, California, EUA) e as amostras de DNA convertidas com bissulfito foram hibridizadas nos *microarrays BeadChip Human Methylation 450 (HM450K)* (Illumina, San Diego, Califórnia, EUA) seguindo o protocolo de metilação *Illumina Infinium HD*.

Os dados brutos foram extraídos pelo scanner iScan SQ (Illumina) usando o *software GenomeStudio* (v.2011.1), com o módulo de metilação v.1.9.0 (Illumina), em arquivos IDAT. As sondas foram anotadas de acordo com seus genes mais próximos usando o pacote *FDb. InfiniumMethylation.hg19* com uma anotação adicional para sondas localizadas em mais de um gene (coordenadas de UCSC Hg19). Os níveis de metilação para cada sonda CpG foram fornecidos em valores beta (0 indicando CpGs não metilados e 1 CpGs totalmente metilados).

A análise dos dados foi realizada no ambiente R utilizando o pacote *minfi* (Aryee *et al.* 2014) e outros pacotes *Bioconductor* (<http://www.bioconductor.org>). Intensidades de sinal das sondas *Infinium I* e *Infinium II* foram normalizadas usando a normalização do quantil (Touleimat e Tost 2012) e o background foi corrigido usando o método '*noob*' (convolução exponencial normal usando sondas *out-of-band*) (Triche *et al.* 2013). Os efeitos em lote (*array/slide*) foram corrigidos usando a função *ComBat* (Johnson *et al.* 2007) do pacote ChAMP (do inglês *Chip Analysis Methylation Pipeline*) (Tian *et al.* 2017). A composição celular foi estimada usando a função

calculateCellCounts do *minfi*, que utiliza o método de Houseman (Tian *et al.* 2017). Não houve diferença nas estimativas de composição celular para os grupos. As etapas de controle de qualidade removeram sondas que tiveram valor de p de detecção $> 0,01$, sondas localizadas em cromossomos sexuais, sondas localizadas em SNPs e localizadas em locais de reação cruzada (Chen *et al.* 2013).

A comparação dos níveis de metilação do DNA da posição CpG entre casos e controles foi realizada com valores M (*logit* de valores B) empregando um modelo linear bayesiano empírico do pacote *limma* (Du *et al.* 2010; Ritchie *et al.* 2015). As posições com um valor de p ajustado (*adjP*) $< 0,05$ como determinado pelo modo de *Benjamini e Hochberg* foram consideradas como Sítios diferencialmente metiladas (DMPs, do inglês *Differently Methylated Probes*). Regiões Diferencialmente Metiladas (DMRs, do inglês *Differently Methylated Regions*) foram identificadas com dois métodos: *DMRcate* (Peters *et al.* 2015) com FDR $< 0,05$ e 0,05 delta beta médio (DB) como pontos de corte, e (Jaffe *et al.* 2012), considerando como significativo o *FWER* $< 0,05$ e selecionando a opção *pickCutoff*. Para hotspots interativos de metilação diferencial, o pacote FEM (do inglês *Functional Epigenetic Modules*) foi usado (Jiao *et al.* 2014).

A idade de metilação do DNA e a aceleração da idade (AA) foram calculadas usando o método do relógio epigenético descrito por Horvath (Horvath 2013). Para esta análise, os valores beta foram utilizados como entrada para o algoritmo sem correção por idade.

Para a análise de enriquecimento de regiões regulatórias, grupos aleatórios foram criados usando o programa de função *PERL* e coordenadas dos sítios CpG da lâmina 450K Illumina usados como pano de fundo (*background*). Para cada comparação, foram gerados 1.000 grupos aleatórios de acordo com CpGs hipermetilados em cada grupo (cl1 e cl2). A ferramenta *BEDTools* (Quinlan e Hall 2010) foi usada para comparar as coordenadas do banco de dados CpGs e *rVarBase*. O número de elementos sobrepostos em *rVarBase* (Guo *et al.* 2016) foi comparado entre os grupos aleatórios e os grupos estudados. O valor de p foi calculado de acordo com o número de sobreposições em grupos aleatórios igual ou superior ao número no grupo estudado dividido por 1.000 comparações.

Para a análise de enriquecimento de VMRs, foi utilizado o algoritmo MSET (Eisinger *et al.* 2013) com 10.000 permutações. Usamos as listas de VMRs de diferentes estudos: (i) Islam *et al.* (2019), estudo com comparação sistemática de padrões de DNAm do genoma entre células epiteliais bucais pediátricas (BECs) e células mononucleares do sangue periférico (PBMCs). Eles identificaram os CpGs como a) informativos (isto é, variáveis entre indivíduos e correlacionados entre BECs e PBMCs) (8140), b) diferencialmente metilados entre tecidos combinados (139,662), ou c) sob influência genética (4980; (ii) Teh *et al.* (2014), estudo das influências relativas dos efeitos interativos genotípicos, ambientais e genéticos sobre o metiloma neonatal, bem como a variação nos padrões de metilação do DNA genômico em amostras de cordão umbilical, genotipagem e medidas de condições ambientais *in utero*. Identificaram 1423 VMRs interindividuais entre os 237 indivíduos. Os níveis de

metilação em 25% dos 1423 VMR-CpGs foram melhor explicados pelo genótipo sozinho, enquanto os demais foram melhor explicados por modelos de interação Gene x Ambiente (GxA); (iii) Garg *et al.* (2018), que realizou uma triagem para identificar regiões de variação epigenética comum usando dados populacionais derivados de cinco diferentes tipos de células humanas. Eles procuraram por conjuntos de sondas com alta variabilidade inter-individual (VMRs) e exploraram os potenciais fatores subjacentes associados à regulação de VMRs usando diferentes estratégias (VMRs influenciadas por efeitos genético, ambientais e de GxA); e (iv) Hachiya *et al.* (2017) catálogo de VMRs em EWAS.

3.4.3 Exoma

Para a construção das bibliotecas das regiões dos *éxons* foi utilizado o kit *SureSelectXT Reagent Kit* (Agilent Technologies Inc, Santa Clara, California, EUA) e o sequenciamento das bibliotecas foi realizado na plataforma *HiSeq 2500 TM* (Illumina), de acordo com a técnica de sequenciamento por síntese *paired end*.

A análise qualitativa nos dados do sequenciamento foi feito com o *FastQC* (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) para verificar se as sequências obtidas possuem qualidade (nota *Phred*) dentro dos parâmetros robustos para continuar a análise. As sequências lidas (*reads*) dos fragmentos das bibliotecas foram trimadas com o programa *NGSToolkit* (Patel e Jain 2012), ou seja, 70% das bases de cada read deve possuir nota *Phred* $\geq$

20, caso contrário serão descartadas. Os *reads* foram alinhados com o algoritmo BWA (Li e Durbin 2010), utilizando como referência a montagem do *National Center for Biotechnology Information (NCBI) reference human genome build 37 (GRCh37/hg19)*. Após esta etapa, foi usado o *SAMtools* (Li *et al.* 2009) para a remoção da duplicata de PCR utilizando a opção *rmdup*. Para a chamada de variantes foi utilizado o programa GATK (*Genome Analysis ToolKit*) (McKenna *et al.* 2010; DePristo *et al.* 2011; Van der Auwera *et al.* 2013) com a opção *Haplotype Caller* com critério de pelo menos *Phred Score* ≥ 30 (Brockman *et al.* 2008; Altmann *et al.* 2012).

As variantes identificadas foram recalibradas segundo o *GATK guidelines*, utilizando os bancos de dados de variantes bem estabelecidas (*HapMap3*, *1000 Genome Project* (Altshuler *et al.* 2010), OMIM (“*Online Mendelian Inheritance in Man*, OMIM (TM),” n.d.) e dbSNP (Sherry 2001) para identificar o perfil dos parâmetros medidos. A anotação funcional foi feita com o ANNOVAR (Wang *et al.* 2010). Para a filtragem das variantes, foi seguido os seguintes passos:

- a) Variações de novo (mutações não herdadas): variações exclusivas do probando, ou seja, os pais devem ser homozigotos para a referência, sem nenhum read do alelo alternativo (Neale *et al.* 2012).
- b) Variações raras e herdadas dos genitores (mutações herdadas com MAF $\leq 0,1$): Para essa seleção, usamos a anotação de banco de dados: projeto 1000 genomas (<http://www.1000genomes.org/>), o banco de dados do *National Heart, Lung and Blood Institute GO Exome Sequencing Project*, que

sequenciou 6512 indivíduos (<http://evs.gs.washington.edu/>) e *Exome Aggregation Consortium* - *ExAC* (<http://exac.broadinstitute.org/>). Consideramos $MAF \geq 0,05$: Variação Comum; $MAF < 0,05$ e $\geq 0,01$: menos comum; $MAF < 0,01$: Raro e; $MAF < 0,005$: Muito raro (Cirulli e Goldstein 2010).

- c) Variações deletérias: Não-sinonimas *missense* (troca de 1 nucleotídeo resultando num aminoácido diferente), Não-sinonimas *nosense* (troca de 1 nucleotídeo resultando num códon de parada), *splicing* (troca de 1 nucleotídeo num sítio de *splicing*) e *Indels frameshift* (inserções ou deleções pequenas resultando na mudança da ordem de leitura).
- d) Cobertura: foram excluídas variantes com cobertura menor do que 20X (20 fragmentos cobrindo o alelo), taxa de cobertura similar ou maior do que as obtidas em outros estudos de exoma em TEA (Iossifov *et al.* 2012; Neale *et al.* 2012; Sanders *et al.* 2012; Wang *et al.* 2013).

Após seleção dos variantes deletérios, para cada paciente foi computado um escore de tolerância gênica: um escore RVIS, ou seja, quanto existe de variação na tolerância a variação. Os genes presente na lista dos 25% genes mais intolerantes a variação de acordo com o escore RVIS (Petrovski *et al.* 2013).

4. RESULTADOS

4.1 Caracterização amostral

Foram selecionadas 69 crianças com TEA, sendo 57 do sexo masculino (82,6%) e 12 do sexo feminino (17,4%), com idade entre 3 e 7 anos (média de 4,75, DP 1,3). Os níveis de QI variaram entre 49-75 (média de 58,4; DP 8,9) e ADI-R de 25 a 60 (média de 45,9; DP 7,4) (**Anexo 1**).

A presença de dimorfismos foi avaliada por uma equipe de geneticistas em 65 indivíduos, e não foi encontrado nenhum paciente que apresentasse síndrome genética clinicamente descrita. Um paciente que não completou o questionário usado para calcular as pontuações de risco foi excluído da análise de PCA, totalizando 68 amostras

A faixa de escolaridade materna foi de ensino fundamental incompleto até ensino superior completo: 16% com ensino fundamental incompleto até ensino médio incompleto, 52% com ensino médio completo e 32% com ensino superior. A maioria das famílias era das classes sociais B (n= 39, 57.4%), C (n= 18, 26.4%) e D (n= 10, 14.7%), mas apenas uma família era da classe social A (1.5%).

Durante a gestação, 37 mulheres (54%) não relataram nenhum tipo de estresse; no entanto, 27 (39%) tiveram 1 a 3 eventos estressantes. Complicações médicas (diabetes gestacional, pré-eclâmpsia e ganho superior a 12 kg durante a gestação) estiveram presentes em 38 mulheres (55%), a maioria teve apenas uma das três complicações (28 mulheres, 41% do total).

Apenas 20 (29%) mulheres revelaram exposição a álcool, tabaco, drogas, medicamentos deletérios ou agrotóxicos durante a gestação. Antecedentes familiares de transtornos psiquiátricos estavam presentes em 43 famílias (62%). Havia 7 famílias (10%) que relataram outros casos de TEA na família. Depressão e esquizofrenia foram os distúrbios mais prevalentes entre as famílias. Os primeiros foram relatados em 32 parentes e o segundo em 12.

4.2 Análise de Componentes e Clusterização

4.2.1 Análise de Componentes Principais

Uma análise de PCA preliminar com os escores de todas as variáveis resultou em 3 PCs representando 70% da variabilidade. Essa análise apresentou algumas comunalidades da variável Exposição Tóxica, que também apresentou uma qualidade mais baixa que as demais e a menor contribuição para os componentes principais (PC) (**Anexo 2**).

Realizamos uma segunda análise de PCA, excluindo a variável de exposição tóxica, e obtivemos melhores resultados, que mostrou 2 PCs com uma boa qualidade de representação das variáveis de acordo com os valores de Cos^2 e variância acumulada de 60% (**Figura 1**). PC1 apresenta maior correlação com as variáveis de estresse psicossocial (estresse materno, escolaridade maternal e classe social), e o PC2 apresenta maior correlação com as variáveis Histórico Familiar Psiquiátrico e de Complicações Metabólicas. Nessa análise, podemos ver também que o PC3 aparece com o eigenvalue de quase 1 e apesar de somar-se aos demais para uma variância

acumulada de 78%, sua qualidade de representação é menor e apresenta comunalidade com o PC1 e PC2.

A Eigenvalues e variância

	eigenvalue	Variância (%)	Variância Acumulada (%)
PC1	1.75	35.03	35.03
PC2	1.25	24.93	59.96
PC3	0.93	18.60	78.56
PC4	0.62	12.44	91.00
PC5	0.45	9.00	100.00

B Correlação entre variáveis e componentes principais

	PC1	PC2	PC3	PC4	PC5
EstressMat	0.71	-0.06	-0.57	-0.15	-0.38
EscolarMat	0.74	-0.10	0.27	0.61	-0.01
ClassSocial	0.82	0.08	0.10	-0.37	0.41
HistFamPsi	-0.15	-0.79	-0.49	0.15	0.29
CompMetab	-0.09	0.77	-0.53	0.26	0.22

C Qualidade da Representação das Variáveis

Figura 1 - Análise de PCA. (A) Eigenvalues e Variância encontrada; (B) Valores de correlação dos componentes com as variáveis; (C) Valores de Cos2, medida de qualidade de variáveis dos componentes. Legenda – Variáveis EstressMat: Estresse Materno; EscolarMat: Escolaridade Materna; ClassSocial: Classe social; HistFamPsi: História Familiar Psiquiátrica; CompMetab: Complicações Metabólicas Gestacionais.

Na **Figura 2** podemos observar as projeções de PC1 e PC2. O primeiro PC está associado a uma dimensão de vulnerabilidade de estresse psicossocial, enquanto o segundo PC está associado a uma vulnerabilidade genética e de estresse metabólico.

Figura 2 – Biplot dos componentes principais. O eixo x representa os valores do PC1 (Dim1) e o eixo y os valores do PC2 (Dim2). As setas coloridas representam os dos fatores de riscos associados aos respectivos componentes e o vetor representando o seu respectivo *loading*. CompGest: Complicações Gestacionais; ClassSocial: Classe Social; MatStress: Estresse Materno; EscolMat: Escolaridade Materna; e PsiquHfam: História Psiquiátrica Familiar.

4.2.2 Análise de Clusterização

Para as análises de comparação de pacientes utilizamos um método de clusterização com os escores dos PCs para verificarmos como esses pacientes seriam divididos levando em conta esses componentes de vulnerabilidade encontrados. Realizamos a análise de 2 conjuntos dados, primeiramente usando os escores do PCA usando todas as variáveis, e depois analisamos a clusterização usando os escores do PCA excluindo a variável Exposição

Tóxica. A clusterização com os escores do PCA usando todas as variáveis resultou em uma árvore hierárquica com clusters pouco estáveis (**Anexo 3**).

Em seguida então rodamos uma clusterização com os escores excluindo a variável exposição (**Figura 3**), e o cluster formado usando 2 PCs (**Figura 3B**) aparece mais estável que o usando 3 PCs (**Figura 3A**) após bootstrap.

Figura 3 – Análise de Clusterização usando o método Ward D em conjunto com o algoritmo pvclust. Valores em vermelho representam o p-valor AU, calculado por reamostragem de bootstrap multiescala; os valores em verde representam o p-valor BP, calculado por reamostragem de bootstrap normal (A) Árvore hierárquica resultante da clusterização utilizando escores individuais

dos PCs 1, 2 e 3 da análise de PCA; (B) Árvore hierárquica resultante da clusterização utilizando escores individuais dos PCs 1 e 2.

Separando as amostras nos dois maiores agrupamentos, as amostras encontradas em cada agrupamento (*cluster 1* e *cluster 2*) foram identificadas de acordo com o posicionamento no biplot do PCA de forma a representar a distribuição dessas amostras de acordo com os PCs 1 e 2 (**Figura 4**).

Figura 4 - Distribuição dos indivíduos dos clusters 1 (Pontos azuis) e cluster 2 (pontos vermelhos) no biplot de PCA representando os PC1 (eixo X) e PC2 (eixo Y). E os vetores das variáveis. Legenda – Variáveis MatStress: Estresse Materno; MatSchool: Escolaridade Materna; SocialClass: Classe social; PsyFamHist: História Familiar Psiquiátrica; GestCompI: Complicações Gestacionais; Dim1: Componente Principal 1 (PC1); Dim2: Componente Principal 2 (PC2); Contrib: Contribuição da variável.

A **Figura 4** mostra que os pacientes do Cluster 1 apresentam valores de PC1 (vulnerabilidade psicossocial) mais negativos e o Cluster 2 apresentam valores de PC1 mais positivos. Ambos os clusters apresentam diferentes valores negativos e positivos do PC2 (vulnerabilidade genética-metabólica), sendo que quando os pacientes apresentam mais histórico familiar psiquiátrico eles apresentam menos complicações gestacionais e vice versa.

Análise estatística dos clusters (**Tabela 1**) confirma que os grupos possuem um fenótipo homogêneo quanto a gravidade de apresentação do TEA, mas apresentam diferenças relacionadas a exposição a fatores de estresse psicológico e sociais.

Tabela 1 – Comparação de valores dos escores de risco e das entrevistas de fenótipo (idade gravidade de TEA (CARS e ADI), sintomas extrenalizantes e internalizantes, funcionalidade (Vineland) e QI entre os agrupamentos encontrados.

Escores de Risco	Cluster 1			Cluster 2			p-valor*
	n	Média	DP	n	Média	DP	
Estress Materno	39	1.18	0.51	29	2.10	1.26	2.47E-04
Escolaridade Materna	39	2.05	1.07	29	3.72	1.16	2.33E-07
Classe Social	39	1.00	0.00	29	1.34	0.48	8.10E-05
Historia Familiar Psiquiatrica	39	5.95	4.70	29	3.72	3.53	0.047
Complicação Gestacional	39	1.13	0.41	29	1.24	0.51	0.242
Fenótipo							
Idade	39	4.64	1.27	29	4.90	1.40	0.473
CARS	39	38.26	6.25	29	34.52	6.27	0.055
CBCL Sintomas Internalizantes	39	10.44	5.24	29	12.03	4.75	0.131
CBCL Sintomas Externalizantes	39	13.51	5.78	29	14.59	8.32	0.780
ADI	38	46.68	7.03	26	47.42	6.95	0.537
QI	38	58.84	8.69	26	59.31	9.13	0.804
Vineland Total	38	49.97	9.26	26	48.85	7.96	0.951
Vineland Comunicação	38	49.18	9.74	27	44.81	7.30	0.203
Vineland Socialização	38	54.29	10.52	27	53.07	9.62	0.836
Vineland Atividade de Vida Diária	38	49.55	12.80	27	51.41	13.07	1.000

* teste Wilcoxon Mann-Whitney

Vale notar que apesar de não apresentarem diferença com relação a complicações gestacionais, é possível ver uma diferença com relação a histórico psiquiátrico familiar, com o grupo com maiores pontuações de estresse psicossocial apresentando menores pontuações de histórico familiar. Assim os indivíduos do Cluster 2 apresentam maiores valores de risco ao estresse psicossocial e menor valores de risco para genética enquanto os indivíduos do Cluster 1 apresentam maior valores de risco para genética e menor valores de risco para estresse psicossocial.

4.3 Análise do Metiloma

Os resultados de metilação foram normalizados de acordo com o *pipeline* descrito em materiais e métodos e mostram que as amostras são comparáveis e que podem ser utilizadas para as próximas etapas. Após normalizações (**Anexo 4**), seguindo-se a conduta de controle de qualidade estabelecida, das 485.512 sondas iniciais foram removidas as sondas com $detP < 0.01$ (2.784), sondas de cromossomos sexuais (11.214), sondas com SNP (16.474) e sondas reativas cruzadas (26.480). Os restantes 428,560 sítios CpG foram utilizados para a análise seguinte.

4.3.1 Sítios Diferencialmente Metilados (DMPs)

A análise de todas as probes diferentemente metiladas (DMPs) entre os dois clusters resultou em 11.879 probes (CpGs) ($p < 0,05$) (**Figura 5**). Com

essas DMPs, foi feita análise de enriquecimento dessas CpGs em bancos de regiões de metilação variáveis (VMRs) associadas a influências ambientais e genéticas, e também em elementos regulatórios do banco rvVarbase.

Figura 5 – Gráfico de Vulcão da distribuição dos valores de delta-beta. CpGs encontradas nos pacientes dos *clusters* 1 (pontos azuis) e *cluster* 2 (pontos vermelhos).

4.3.2 Enriquecimento de VMRs e Regiões Regulatórias

A comparação das DMPs com todos os conjuntos de dados de VMRs mostrou que nosso conjunto é enriquecido para VMRs associadas a influências

ambientais e genéticas observadas em dados da comparação metilômicas entre saliva e sangue (VMRs específicos desses tecidos), VMRs informativos, VMRs em estudos de EWAS, VMRs de Células B, T, Fibroblastos, Glia e Neurônio. **(Tabela 2).**

Tabela 2 – Comparação das DMPs dos pacientes com TEA com as VMRs de 4 bancos de dados.

	Database	Total	Overlap	P-value	FE
Islam et al, 2019	Tecidos Diferentes	139662	4549	< 0.0001	1.23
	Informativos	8140	295	< 0.0001	1.59
	meQTLs	4980	143	0.1617	1.09
Teh et al, 2014	In-utero Genético	362	14	0.0702	1.56
	In-utero Genético x Ambiente	1061	34	0.1296	1.23
Hachiya et al, 2017	VMRs EWAS	269	17	0.001	2.31
Garg et al, 2018	Células B	4367	214	< 0.0001	1.85
	Ambientais	804	27	0.1217	1.28
	Fibroblastos	4788	149	0.045	1.15
	Glia	6990	221	0.0021	1.21
	Neurônios	7075	230	1.00E-004	1.27
	Células T	8940	396	< 0.0001	1.72

Usando as CpGs das DMP presentes em VMRs enriquecidas nos nossos dados, rodamos uma análise de clusterização não supervisionada, que mostrou a separação dos dois *clusters* similares aos originais **(Figura 6).**

Figura 6 – Heatmap de DMPs enriquecidos em VMRs. Cada linha representa um sítio CpG. Na barra superior representando as amostras, cada cor representa o *cluster* daquela amostra (azul: cl1 e vermelho: cl2). A metilação do DNA é dimensionada por cores de acordo com o valor de metilação (maior metilação azul e menor metilação vermelha).

Com relação a análise de enriquecimento de DMPs em elementos regulatórios do banco rVabase, não foi encontrado enriquecimento em rVabase (2.285 CpGs). Dividindo as CpGs em CpGs hipermetilados no cluster 1 e os CpGs hipermetilados no cluster 2, 1.326 CpGs do Cluster 1 foram encontradas em elementos rVarBase, quando comparados a grupos aleatórios, e 959 no Cluster 2 ($p = 1$). Buscando por cada categoria (**Tabela 3**) a maior parte das CpGs hipermetiladas do cluster 1 estavam mais localizados em promotores ativos e enhancers gênicos enquanto que no cluster 2 estavam mais localizadas nos promotores e enhancers bivalentes (Qui-quadrado $<0,05$).

Tabela 3 - Análise de DMPs localizadas nas regiões de elementos regulatórios do banco de dados rVabase.

Tipo*	Cluster 1	Cluster 2	Esperado cl1 (cl1 vs cl2)	Esperado cl2 (cl1 vs cl2)	p-value (cl1 vs cl2)	Esperado cl1 (cl1 vs Bck)	p-valor (cl1 vs Bck)	Esperado cl2 (cl2 vs Bck)	p-valor (cl2 vs Bck)
TssA	604	532	622	514	2.80E-01	552	2.10E-02	457	2.40E-04
TssAFlnk	805	664	804	665	9.70E-01	735	6.40E-03	608	1.70E-02
TxFlnk	455	189	352	292	9.40E-17	365	1.10E-06	299	4.70E-11
Tx	533	338	477	394	7.70E-05	603	2.80E-03	498	4.80E-14
TxWk	1281	936	1213	1004	1.40E-03	1321	2.20E-01	1092	1.20E-07
EnhG	525	348	478	395	9.50E-04	553	2.20E-01	457	1.00E-07
Enh	1085	918	1096	907	6.00E-01	1097	7.10E-01	908	7.20E-01
ZNF/Rpts	194	140	183	151	2.30E-01	199	7.50E-01	164	5.80E-02
Inactive region	18	8	14	12	2.00E-01	19	9.20E-01	16	7.10E-02
TssBiv	326	368	380	314	2.80E-05	325	9.70E-01	270	8.50E-10
BivFlnk	244	363	332	275	2.10E-13	247	8.50E-01	207	1.00E-28
EnhBiv	379	532	499	412	1.60E-16	433	7.00E-03	361	1.10E-20

*Classificação de acordo com RoadMap Epigenomics

bck: background baseado na lamina 50K Illumina, cl: Cluster

TssA: Promotor TSS Ativo; TssAFlnk: promotor TSS ativo; TxFlnk: Transcrição no gene 5' e 3'; Tx: transcrição forte; TxWk: transcrição fraca; EnhG: Enhancers genicos; Enh: Enhancers; ZNF/Rpts: Genes zincfinger & repetições; TssBiv: promotor TSS bivalent; BivFlnk: Promotor Flanqueando TSS/Enh bivalente; EnhBiv: bivalente Enhancer.

A ausência de enriquecimento de DMPs como um todo em elementos regulatórios do rVabase, mas a presença de diferentes elementos nos dois grupos sugerem que essas regiões podem estar sendo afetadas por variações genômicas comuns em importantes regiões reguladoras de formas diferentes.

4.3.3 Módulos Epigenéticos Funcionais

A análise de diferença de metilação em módulos funcionais mostrou 3 módulos diferentemente metilados (**Figuras 7-9**, detalhes em **Anexo 5**).

Figura 7 – Módulo Funcional diferentemente metilado do gene semente ROBO3.

Figura 8 – Módulo Funcional diferentemente metilado do gene semente BCL2A1

Figura 9 – Módulo Funcional diferentemente metilado do gene semente UNC119

A análise de enriquecimento em bancos de vias (KEEG), funções biológicas (GO) e fenótipos (Phenotype Ontology), mostrou que dois dos 3 módulos encontrados apresentam enriquecimentos em vias e funções do sistema imune e resposta a stress (**Tabela 4, Anexo 6**). O módulo cujo gene semente é o UNC119, não apresentou enriquecimento porém seus genes estão relacionados com divisão celular e metabolismo.

Tabela 4 – Resumo dos resultados da análise de Enriquecimento da lista de genes dos Módulos Funcionais diferentemente metilados (Tabela completa no **Anexo 6**).

Semente	Banco	Tam	Obs	Exp	Taxa	FDR	Descrição
ROBO3	GO	361	13	0.45	28.82	0	Regulação da resposta inflamatória
	GO	238	11	0.30	36.99	1.45E-13	Imunidade mediada por linfócitos
	GO	382	11	0.48	23.04	2.49E-11	Resposta imune adaptativa
	KEGG	79	11	0.15	74.28	0	Cascatas do sist. complemento e coagulação
	KEGG	133	10	0.25	40.11	0	Lúpus eritematoso sistêmico
	PO	23	10	0.09	113.77	0	Anormalidade do sistema do complemento
	PO	117	6	0.45	13.42	0	Autoimunidade
BCL2A1	GO	385	8	0.24	33.26	0	Regulação da via de sinalização apoptótica
	GO	116	5	0.07	68.99	1.46E-06	Alterações mitocondriais apoptóticas
	GO	237	4	0.15	27.01	1.09E-03	Transporte mitocondrial
	GO	31	2	0.02	103.26	0	Processo apoptótico envolvido no desenvolvimento
	GO	268	3	0.17	17.92	0	Resposta ao estresse do retículo endoplasmático
	GO	468	4	0.29	13.68	0	Atividade de heterodimerização de proteínas
	KEGG	33	3	0.04	75.44	0	Apoptose
UNC119*	GO	108	3	0.23	13.08	0	polimerização ou despolimerização de microtúbulos
	GO	210	3	0.45	6.72	1	dobramento de proteínas
	KEGG	60	2	0.11	17.78	1	Biossíntese de hormônios esteroides
	KEGG	133	2	0.25	8.02	1	Fosforilação oxidativa
	KEGG	43	1	0.08	12.41	1	Diabetes mellitus tipo I
	PO	33	2	0.07	29.73	1	Hiperatividade adrenal
	PO	91	2	0.19	10.78	1	Concentração anormal de cálcio no sangue
	PO	316	3	0.64	4.66	1	Anormalidade do nível de hormônio circulante

*Modulo sem enriquecimento significativo.

Legenda: Tam – Tamanho; Obs – Observado; Exp – Esperado; FDR – p-valor corrigido; PO – Banco de Ontologia Fenotípica; GO – Banco de Processos Biológicos; KEGG – Banco de Vias Biológicas.

4.3.4 Idade Epigenética

Seguindo o método de Horvath, calculamos a idade epigenética e pudemos observar que apesar de ambos os clusters apresentarem uma idade epigenética maior que a cronológica, eles não apresentaram diferença de aceleração de idade epigenética (AA) (**Figura 10**).

Figura 10 – Análise da Aceleração da Idade Epigenética. (A) Diferença de Idade cronológica entre os grupos; (B) Diferença de Idade epigenética entre os grupos; (C) Diferença de aceleração da idade epigenética entre grupos caiculada a partir da diferença de Idade cronológica pela idade epigenética.

4.4 Exoma

Análise do exoma foi realizado numa porção pequena dos pacientes dos clusters (cluster 1, n = 25 e Cluster 2, n = 8). Dos 6.625 (1966 presentes na lista de 25% mais intolerantes do RVIS) variantes deletérios selecionados,

5512 (RVIS: 1547, de novo: 18) estão nos pacientes do cluster 1 e 2578 (RVIS: 736, de novo: 8) nos pacientes do cluster 2 (**Figura 11**). Teste estatístico q-quadrado não encontrou diferença de quantidade de genes afetados entre os grupos ($p>0.05$).

Figura 11 – Esquema dos Variantes do Exoma selecionados.

5. DISCUSSÃO

Considerando nossa hipótese de estudo, os resultados confirmam que mesmo entre um grupo clinicamente homogêneo de pacientes com TEA é possível encontrar subgrupos com diferentes vulnerabilidades com relação a exposição ambiental (estresse psico-social materno e complicações gestacionais) e *background* genético (sugerido pela diferença de histórico psiquiátrico familiar). A análise de metilação mostrou que os sítios diferentemente metilados entre os subgrupos encontrados é influenciado por fatores genéticos, ambientais e da interação entre esses dois (GxA), e estes parecem participar em vias importantes para o TEA. Assim indivíduos com sintomas e diagnósticos semelhantes podem apresentar diferenças na sua base biológica.

O PCA mostrou dois componentes de vulnerabilidade e usando os escores do PCA foi possível formar dois grupos de pacientes. O *cluster* 1 tem um escore mais negativo no PC1 e apresenta mais história familiar psiquiátrica, o que implica um maior *background* genético poligênico (Agerbo *et al.* 2015; Maher 2015) . O *cluster* 2 parece mais associado a um escore positivo no PC1, o componente de vulnerabilidade desfavorável do ambiente psico-social. É importante notar que, de 7 indivíduos com história familiar positiva de TEA, 3 estavam no *cluster* 1 e 4 no *cluster* 2, portanto, não estamos formando subgrupos de TEA simplex e complex já descritos na literatura (Schaaf e Zoghbi 2011; Oerlemans *et al.* 2016).

É importante notar que os nossos subgrupos não têm diferenças em relação à carga de variantes *de novo* e muito raras com maior impacto uma vez

que os subgrupos baseados nesta característica específica já foram estabelecidos na literatura (Iossifov *et al.* 2014; Ronemus *et al.* 2014; Sanders *et al.* 2015). A análise estatística dos números totais das variantes de novo e raras, bem como genes intolerantes a variação entre os grupos não apresentou diferença estatística. Esse resultado indica que além de serem homogêneos quanto ao fenótipo, eles também são homogêneos quanto a carga de variantes raras e de maior efeito, e portanto não foi isso que resultou na separação dos subgrupos. Isso confirma que a variação comum age de forma aditiva em indivíduos com SNV *de novo* (Weiner *et al.* 2017), e sugere que considerar a exposição ambiental é importante para encontrar subgrupos.

Para validar os subgrupos em nível molecular investigamos as diferenças de metilação genômica comparando os *clusters* encontrados. A ausência de enriquecimento de DMPs como um todo em elementos regulatórios do rVArbase, mas a presença de diferentes elementos nos dois grupos sugerem que essas regiões podem estar sendo afetadas por variações genômicas comuns em importantes regiões reguladoras de formas diferentes. Como locais de metilação polimórficos se encontram aglomerados em regiões genômicas variáveis associadas a efeitos genéticos, ambientais e de interação GxA, esperávamos que nossas DMPs fossem enriquecidas por CpGs localizados dentro de VMRs. E encontramos enriquecimentos em bancos relacionados a VMRs específicos de 2 Tecidos diferentes (saliva e sangue) e VMRs VMRs de Células B, T, Fibroblastos, Glia e Neurônio, o que significa que esses VMRs não devem estar associados com a programação específica desses tecidos (Garg *et al.* 2018; Islam *et al.* 2019). O enriquecimento aparece em VMRs genéticas ou representando a interação GxA, ou seja, associadas a

SNPs (VMRs informativos, e VMRs em estudos de EWAS) (Teh *et al.* 2014; Islam *et al.* 2019). Além disso, o enriquecimento em VMRs em glia e neurônios sugerem que a variação destes CpGs pode gerar alterações também tecido cerebral (Garg *et al.* 2018). Eles não aparecem enriquecidos para VMRs do sangue do cordão umbilical neonatal, que teoricamente representa o ambiente intraútero, isto sugere que DMRs de cordão umbilical são tecido específicas. A maioria dos DMRs dos bancos de DMR utilizados foram explicada por efeitos genéticos ou efeitos da interação gene-ambiente, Como olhamos DMPs, não é possível dizer que um *cluster* tem mais VMRs G, E ou GxA do que o outro, mas sua importância pode ser observada na separação entre *clusters* fornecida pela análise de *cluster* a partir das CpGs.

Além de estarem mais localizados em regiões genômicas relacionadas a efeitos genéticos e ambientais também fizemos uma análise de enriquecimento de processos biológicos dos módulos funcionais epigenéticos. Esperaríamos que estes módulos estariam representados por processos biológicos já associados com TEA e relacionados com variantes genéticas e exposição ambiental. Como esperado, os genes dos módulos estavam relacionados a sistema imune, estresse e metabolismo (Legido *et al.* 2013; Voineagu e Eapen 2013; Walker *et al.* 2015; Connolly *et al.* 2016). A questão se os fatores de risco genéticos e ambientais perturbam diferentes vias moleculares subjacentes no desenvolvimento de TEA foi investigada por Voineagu *et al.* (2011). Eles sugeriram que a disfunção de sinalização sináptica e neuronal tem uma base genética forte, enquanto que as alterações imunológicas têm um componente genético menos pronunciado e, portanto, são provavelmente causadas por fatores ambientais. Distúrbios metabólicos como resistência à

insulina, obesidade e hipertensão crônica estão associados à inflamação sistêmica (Walker *et al.* 2015). A obesidade leva a um estado persistente de inflamação de baixo grau por meio do recrutamento e secreção de citocinas pró-inflamatórias por adipócitos hipertróficos. Além disso, excesso de gordura corporal e hiperglicemia são fontes de estresse oxidativo (Connolly *et al.* 2016). A análise de metilação diferencial não nos permitiu discriminar vias em subgrupos, mas corrobora que as vias associadas a fatores genéticos e ambientais validaram nossos subgrupos com base nos componentes de vulnerabilidade..

Este trabalho tem uma amostra relativamente pequena de apenas pacientes jovens brasileiros com TEA e pode não refletir a população total de TEA. Trata-se de um estudo transversal realizado com dados que se baseiam na qualidade das informações fornecidas pelos familiares dos pacientes. Entre os dados coletados, tivemos que excluir indivíduos devido à falta de informações nos questionários. Também não medimos o tempo específico de exposição ao estresse. Mesmo considerando essas limitações, sugerimos que mais do que uma dicotomia genética ou ambiental, ou a influência de diferentes vias biológicas específicas, a combinação de fatores de risco genéticos e ambientais provavelmente associados a efeitos dependentes de tempo podem representar melhor os subtipos de TEA mesmo entre os casos mais graves.

A análise de metilação diferencial corrobora que as vias associadas a fatores genéticos e ambientais validaram nossos subgrupos de componentes de vulnerabilidade, uma vez que encontramos diferenças em locais de metilação associados a herança genética comum herdada e resposta ambiental ao estresse. Assim como sítios diferentemente metilados em genes de vias

importantes para o neurodesenvolvimento e TEA, como resposta inflamatória e estresse.

Esse resultado é importante pois sugere que Indivíduos com sintomas e diagnósticos semelhantes podem apresentar diferenças na sua base biológica (fisiopatológica). A partir dessa observação é possível pensar num modelo de vulnerabilidade e fisiopatologia mostrando que variantes comuns e exposição de estresse ambiental durante a gestação são componentes de vulnerabilidade que contribuem para o risco do desenvolvimento do transtorno de desenvolvimento como o TEA e que a variabilidade fenotípica como diferença em gravidade e funcionalidade devem estar mais relacionadas com as variações raras e de novo (**Figura 12**).

Figura 12 – Modelo do papel dos componentes de vulnerabilidade genéticos comuns e ambientais na fisiopatologia de TEA.

6. CONCLUSÃO

Em nosso estudo, encontramos dois grupos diferentes entre os indivíduos mais gravemente comprometidos com autismo. Um *cluster* representa uma dimensão com maior vulnerabilidade genética e outro com maior exposição a ambiente desfavorável e estresse psico-social durante a gestação. A análise de metilação mostra que este grupo apresenta diferenças em locais de metilação associados a herança genética comum herdada e resposta ambiental ao estresse. Os mecanismos de etiologia do TEA determinados por fatores de risco genéticos e ambientais parecem convergir para vias de desenvolvimento comprometidas semelhantes e, sobretudo, uma resposta inflamatória e de estresse celular. Quanto à apresentação do fenótipo, ambos os grupos foram muito semelhantes e é plausível que diferentes componentes de vulnerabilidade contribuem para o risco do desenvolvimento do TEA, independente das diferenças de gravidade e funcionalidade.

7. ANEXOS

Anexo 1

Tabela Anexa 1 – Caracterização e escores de fenótipo dos pacientes selecionados

ID	CI	Idade	Sexo	CARS	QI	VCom	VHvd	VSocial	VTotat	ADI	CBCInt	CBCLExt	Exoma*
1	1	6	F	37	49	39	37	48	124	60	12	10	0
2	2	7	M	38.5	50	39	NA	32	71	56	19	34	0
4	2	6	F	24	49	34	32	48	114	51	8	10	0
5	1	3	M	40	66	54	53	51	158	49	13	14	1
6	1	5	M	38.5	na	67	56	59	97	na	13	15	0
7	1	3	M	52	67	52	59	54	165	53	8	16	1
8	1	4	M	32.5	71	49	54	61	164	36	5	18	1
9	1	3	M	27.5	72	50	55	51	156	44	3	16	1
10	2	4	F	43.5	72	51	63	54	168	53	12	23	0
11	2	6	M	39.5	51	39	51	42	132	43	12	14	0
12	1	5	M	35	73	52	60	66	178	38	16	14	0
14	2	4	M	27.5	56	49	56	54	159	NA	15	23	1
15	1	6	M	41.5	49	44	48	55	147	52	15	17	1
16	1	5	M	30	56	37	21	45	103	57	14	11	0
17	1	4	M	35.5	55	45	50	60	155	44	8	6	1
19	1	6	M	47	51	63	84	68	215	NA	5	16	1
20	2	6	M	42	50	47	42	61	150	54	11	8	0
21	1	6	M	36	69	40	54	65	159	49	14	22	1
22	2	3	M	40	69	67	64	79	210	25	5	17	0
23	1	4	M	44	69	48	59	53	160	34	14	16	1
24	2	4	M	40	55	50	54	51	155	43	5	11	1
25	2	7	M	25.5	49	33	31	49	113	52	16	28	0
26	2	5	M	31	NA	NA	NA	NA	NA	NA	19	16	0
27	1	5	F	49	49	47	35	54	136	54	6	9	1
28	1	6	M	49.5	49	45	67	61	173	47	5	17	0
29	1	5	F	38	49	42	44	51	137	48	6	3	0
30	1	4	M	52	54	50	52	63	165	48	3	12	1
31	2	7	F	36	49	28	20	46	94	43	11	4	0
32	1	6	M	36	71	NA	NA	NA	NA	51	7	8	0
33	1	6	M	32	50	50	52	54	156	41	8	3	1
34	2	5	M	32	75	65	56	85	206	49	11	25	0
35	2	4	M	33	52	49	53	74	176	32	15	14	1
36	1	3	M	36	69	58	55	80	193	31	8	9	1

37	1	5	F	36.5	66	49	49	61	159	57	21	27	0
38	1	5	M	33	49	41	35	44	120	42	17	14	1
39	1	3	F	42	71	51	55	53	159	46	12	10	1
42	2	6	M	31	49	35	25	50	110	37	13	17	0
43	1	4	M	37	54	48	45	56	149	48	13	10	0
44	na	5	M	41	63	51	59	53	163	28	11	13	0
46	1	5	M	35	49	23	21	31	75	NA	6	3	1
47	2	4	M	28.5	68	45	53	54	152	42	4	17	0
48	1	3	M	32	75	52	50	56	158	39	10	16	1
49	2	6	M	37	49	31	32	45	108	NA	10	19	1
50	1	3	M	25	62	54	70	61	185	49	12	9	1
51	1	3	M	43,5	67	49	52	51	152	45	9	7	1
52	2	7	M	49	55	38	53	37	128	49	11	4	0
53	1	7	M	35	49	32	20	45	97	49	19	17	0
54	2	4	M	37	56	49	47	52	148	52	14	15	1
56	1	4	M	26	55	54	68	66	188	32	14	12	1
57	1	4	M	36	60	47	52	52	151	44	11	4	1
58	1	4	M	33.5	52	45	42	51	138	45	6	16	1
59	1	6	M	35	49	42	36	48	126	46	8	5	0
60	2	7	M	39	54	28	46	21	95	55	14	10	0
61	2	3	F	35	72	50	68	56	174	50	6	9	1
62	1	7	M	38	50	32	48	66	146	47	6	4	0
63	1	7	M	39	49	34	23	48	105	50	7	8	0
64	2	3	M	37	75	NA	NA	NA	NA	NA	16	28	0
65	2	4	F	44	55	44	45	56	145	48	14	15	1
66	1	4	M	49	60	61	62	53	176	42	5	10	1
68	2	6	F	36	66	41	41	50	132	49	10	18	0
69	2	3	M	33	62	46	55	52	153	51	4	6	0
71	2	5	M	39	70	47	45	51	143	51	18	15	0
72	1	5	M	30	59	47	53	57	157	54	17	18	1
73	1	3	M	35	69	51	50	47	148	42	8	7	1
74	2	5	M	31	49	44	46	51	141	33	10	18	0
75	1	4	F	36	51	60	77	89	226	NA	10	15	0
77	2	4	M	37	48	52	53	51	156	44	6	4	0
78	2	4	M	37	65	60	56	65	181	31	6	3	1
80	2	3	M	38	62	54	60	81	195	48	9	6	0

Vcom: Escore de comunicação da Vinelad; VHvd: Escore de habilidade de vida diária da Vineland;
Vsocial: Escore de socialização da Vinelad; Vtotal: Escore Total da Vinelad; ADI: Escore da ADI-R; CARS:
Escore da CARS; CBCLInt: Escore dos Sintomas Internalizantes da CBCL; CBCLExt: Escore dos
Sintomas Externalizantes da CBCL. *Tem exoma? 0= Não , 1= Sim

Anexo 2

Tabela Anexa 2 – Escores (z escores) dos fatores de risco usados no PCA

ID	StressMat	EscolMat	ClassSocial	CompGest	Psiqufam
01	-0.759	-0.553	-0.776	0.337	1.140
02	1.421	2.339	-0.776	0.337	1.565
04	0.695	0.893	0.577	0.337	0.290
05	0.695	0.170	-0.776	-0.937	0.927
06	-0.759	0.170	-0.776	-0.937	-0.984
07	-0.032	0.170	-0.776	-0.937	0.715
08	-0.759	0.170	-0.776	0.337	-0.984
09	-0.759	0.170	-0.776	-0.937	0.290
10	2.875	0.170	0.577	-0.937	0.927
11	-0.032	2.339	-0.776	-0.937	-0.984
12	-0.759	0.170	-0.776	0.337	-0.984
14	0.695	0.170	0.577	2.887	-0.984
15	-0.759	0.170	-0.776	0.337	-0.984
16	-0.759	0.170	-0.776	0.337	-0.984
17	-0.759	0.170	-0.776	0.337	0.715
19	-0.759	-1.276	-0.776	0.337	-0.984
20	-0.032	2.339	1.929	0.337	-0.347
21	-0.032	-1.276	-0.776	-0.937	1.565
22	1.421	0.893	1.929	0.337	-0.347
23	1.421	-1.276	0.577	-0.937	2.839
24	-0.759	0.170	1.929	-0.937	0.290
25	-0.759	0.170	0.577	-0.937	-0.984
26	-0.759	0.170	0.577	0.337	-0.134
27	-0.759	0.170	-0.776	-0.937	-0.984
28	-0.759	0.170	-0.776	0.337	1.140
29	-0.759	-1.276	0.577	1.612	-0.984
30	-0.759	0.170	-0.776	-0.937	-0.984
31	-0.759	0.170	1.929	-0.937	-0.771
32	-0.759	0.170	-0.776	-0.937	1.352
33	-0.759	0.170	0.577	-0.937	0.290
34	1.421	2.339	1.929	-0.937	0.927
35	-0.032	0.893	-0.776	0.337	0.290
36	-0.759	-1.276	-0.776	-0.937	0.290
37	-0.759	-1.276	-0.776	0.337	0.290
38	-0.759	-1.276	-0.776	0.337	-0.134
39	-0.032	0.170	-0.776	0.337	2.414
42	-0.032	0.170	1.929	-0.937	0.290
43	-0.759	-1.276	0.577	0.337	1.565
46	-0.759	-0.553	-0.776	-0.937	1.565
47	-0.032	1.616	0.577	-0.937	1.777
48	-0.032	-1.276	-0.776	2.887	0.290

49	2.148	0.170	1.929	-0.937	-0.984
50	-0.759	-1.276	-0.776	0.337	0.078
51	-0.759	-1.276	-0.776	-0.937	0.503
52	0.695	0.893	0.577	0.337	-0.984
53	-0.759	0.170	-0.776	1.612	-0.984
54	-0.759	0.170	0.577	0.337	-0.984
56	-0.759	-1.276	-0.776	0.337	-0.559
57	-0.759	-1.276	-0.776	0.337	-0.984
58	1.421	-1.276	-0.776	-0.937	0.927
59	-0.032	-1.276	-0.776	-0.937	1.565
60	1.421	0.170	0.577	-0.937	0.290
61	2.148	0.170	-0.776	1.612	0.290
62	-0.759	1.616	-0.776	0.337	1.140
63	0.695	-1.276	-0.776	0.337	-0.347
64	1.421	0.170	0.577	1.612	-0.984
65	-0.032	0.170	0.577	1.612	-0.984
66	0.695	-1.276	-0.776	0.337	-0.984
68	1.421	0.170	0.577	1.612	-0.984
69	2.148	2.339	1.929	-0.937	-0.984
71	-0.759	0.170	0.577	-0.937	-0.984
72	-0.759	-0.553	-0.776	1.612	0.290
73	-0.759	0.170	-0.776	-0.937	-0.559
74	-0.759	0.170	0.577	-0.937	-0.984
75	-0.759	-1.276	-0.776	0.337	-0.134
77	0.695	0.170	0.577	-0.937	-0.984
78	2.148	0.170	1.929	1.612	-0.559
80	-0.032	0.170	1.929	0.337	-0.347

Anexo 3

Figura Anexa 1 – PCA e Clusterização com variável Exposição Tóxica. (A) Eigenvalues e Variância encontrada; (B) Valores de correlação dos componentes com as variáveis; (C) Valores de Cos2, medida de qualidade de variáveis dos componentes; (D) Árvore hierárquica resultante da clusterização utilizando escores individuais dos PCs 1, 2 e 3. Legenda – Variáveis EstresseMat: Estresse Materno; EscolarMat: Escolaridade Materna; ClassSocial: Classe social; HistFamPsi: História Familiar Psiquiatrica; CompMetab: Complicações Metabólicas Gestacionais.

Anexo 4

A

B

Figura Anexa 2 – Distribuição de beta values produzidos pelas sondas Infinium I e Infinium II. (A) Valores pré-normalização; (B) valores pós normalização.

Anexo 5

Tabela Anexa 3 – Detalhes dos Modulos Funcionais diferentemente metilados.

Modulo (Gene Semente)	EntrezID	Gene	stat (DNAm)*	P*
UNC119	9094	UNC119	-5.09	3.00E-06
	306	ANXA3	1.17	0.25
	379	ARL4D	1.50	0.14
	2948	GSTM4	0.79	0.43
	81890	QTRT1	-0.45	0.65
	7572	ZNF24	-0.65	0.52
	1031	CDKN2C	-0.79	0.43
	6783	SULT1E1	1.84	0.07
	5464	PPA1	-1.93	0.06
	1787	TRDMT1	-1.40	0.17
	10063	COX17	0.06	0.95
	1164	CKS2	-0.58	0.56
	5863	RGL2	-1.91	0.06
	53826	FXVD6	-0.50	0.62
	3117	HLA-DQA1	1.45	0.15
	55274	PHF10	0.06	0.95
	402	ARL2	2.27	0.03
	1070	CETN3	-2.00	0.05
	54810	GIPC2	0.67	0.50
	27436	EML4	3.88	0.00
	7809	BSND	0.79	0.43
	5522	PPP2R2C	-0.42	0.68
	23568	ARL2BP	-0.77	0.44
	4809	SNU13	1.62	0.11
	6820	SULT2B1	-0.36	0.72
	60492	CCDC90B	0.85	0.40
	84056	KATNAL1	1.27	0.21
	6904	TBCD	1.32	0.19
	6903	TBCC	-3.33	0.00
	6905	TBCE	-0.52	0.61
	7976	FZD3	1.66	0.10
	5068	REG3A	0.13	0.90
	51282	SCAND1	2.24	0.03
	7579	ZSCAN20	-1.11	0.27
	10127	ZNF263	0.52	0.60
	90933	TRIM41	-1.54	0.13
	130540	ALS2CR12	-1.69	0.10
BCL2A1	597	BCL2A1	-4.46	3.00E-05
	10018	BCL2L11	-0.69	0.49
	578	BAK1	-0.55	0.58
	5366	PMAIP1	-0.77	0.45
	638	BIK	-2	0.05

	5452	POU2F2	-2.7	0.01
	599	BCL2L2	3.34	1.00E-03
	90427	BMF	-1.06	0.29
	7417	VDAC2	-1.51	0.14
	55669	MFN1	0.09	0.93
ROBO3	64221	ROBO3	3.97	2.00E-04
	11061	CNMD	2.01	0.05
	1191	CLU	-2.8	0.01
	712	C1QA	1.11	0.27
	10924	SMPDL3A	3.01	4.00E-03
	714	C1QC	1.57	0.12
	1356	CP	-0.8	0.43
	713	C1QB	-2.63	0.01
	715	C1R	3.15	2.00E-03
	64386	MMP25	1.92	0.06
	716	C1S	-0.44	0.67
	730	C7	3.65	5.00E-04
	5444	PON1	0.75	0.45
	3273	HRG	0.03	0.98
	732	C8B	-0.64	0.52
	727	C5	1.41	0.16
	735	C9	-0.05	0.96
	731	C8A	1.4	0.17
	1369	CPN1	-0.8	0.43
	1370	CPN2	-0.37	0.71

* stat (DNAm): estatística da metilação diferencial calculado pelo algoritmo FEM ; P: P-valor da metilação diferencial

Anexo 6

Tabela Anexa 4 – Análise de Enriquecimento da lista de genes dos módulos Funcionais diferentemente metilados.

Semente	Gene Set	Tam	Obs	Exp	Taxa	FDR	Genes (EntrezGene ID)	Descrição
BCL2A1	GO:2001233	385	8	0.24	33.3	2.94E-09	578;597;599;638;5366;7417;10018;90427	regulation of apoptotic signaling pathway
	GO:0097193	285	7	0.18	39.3	2.63E-08	578;597;599;5366;7417;10018;90427	intrinsic apoptotic signaling pathway
	GO:0008637	116	5	0.07	69.0	1.46E-06	578;638;5366;10018;90427	apoptotic mitochondrial changes
	GO:0007006	126	5	0.08	63.5	1.78E-06	578;5366;10018;55669;90427	mitochondrial membrane organization
	GO:0010821	176	5	0.11	45.5	7.96E-06	578;638;5366;10018;90427	regulation of mitochondrion organization
	GO:0038034	69	4	0.04	92.8	1.21E-05	578;597;599;10018	signal transduction in absence of ligand
	GO:0097191	220	5	0.14	36.4	1.64E-05	578;597;599;5366;10018	extrinsic apoptotic signaling pathway
	GO:0090559	79	4	0.05	81.0	1.64E-05	578;5366;10018;90427	regulation of membrane permeability
	GO:0043254	447	5	0.28	17.9	4.86E-04	638;5366;7417;10018;90427	regulation of protein complex assembly
	GO:0006839	237	4	0.15	27.0	1.09E-03	578;5366;10018;90427	mitochondrial transport
	GO:0007548	266	4	0.17	24.1	1.57E-03	578;599;638;10018	sex differentiation
	GO:0070585	137	3	0.09	35.0	0.01	5366;10018;90427	protein localization to mitochondrion
	GO:0061458	428	4	0.27	15.0	0.01	578;599;638;10018	reproductive system development
	GO:1902742	31	2	0.02	103.3	0.01	578;10018	apoptotic process involved in development
	GO:0048872	245	3	0.15	19.6	0.03	578;5366;10018	homeostasis of number of cells
	GO:0034976	268	3	0.17	17.9	0.04	578;5366;10018	response to endoplasmic reticulum stress
	GO:0019867	206	8	0.13	62.2	0.00	578;597;599;5366;7417;10018;55669;90427	outer membrane
	GO:0046982	468	4	0.29	13.7	0.01	578;597;599;10018	protein heterodimerization activity
	hsa04215	33	3	0.04	75.4	0.00	10018;5366;578	Apoptosis
	hsa04210	136	4	0.16	24.4	0.00	10018;5366;578;597	Apoptosis
	hsa05210	86	3	0.10	28.9	0.01	10018;5366;578	Colorectal cancer
	hsa05206	299	4	0.36	11.1	0.02	10018;578;599;90427	MicroRNAs in cancer
ROBO3	GO:0002697	381	13	0.48	27.3	0.00E+00	712;713;714;715;716;727;730;731;732;735;1191;1369;1370	regulation of immune effector process
	GO:0050727	361	13	0.45	28.8	0.00E+00	712;713;714;715;716;727;730;731;732;735;1191;1369;1370	regulation of inflammatory response
	GO:0051604	322	13	0.40	32.3	0.00E+00	712;713;714;715;716;727;730;731;732;735;1191;1369;1370	protein maturation

	GO:0006959	242	14	0.30	46.3	0.00E+00	712;713;714;715;716;727; 730;731;732;735;1191; 1369;1370;3273	humoral immune response
	GO:0002526	154	13	0.19	67.6	0.00E+00	712;713;714;715;716;727; 730;731;732;735;1191; 1369;1370	acute inflammatory response
	GO:0072376	92	13	0.11	113. 1	0.00E+00	712;713;714;715;716;727; 730;731;732;735;1191; 1369;1370	protein activation cascade
	GO:0002449	238	11	0.30	37.0	1.45E-13	712;713;714;715;716;727; 730;731;732;735;1191	lymphocyte mediated immunity
	GO:0002250	382	11	0.48	23.0	2.49E-11	712;713;714;715;716;727; 730;731;732;735;1191	adaptive immune response
	GO:0019835	40	6	0.05	120. 0	8.11E-10	727;730;731;732;735;3273	cytolysis
	GO:0072562	108	11	0.13	81.5	0.00E+00	713;714;715;716;731;735; 1191;1356;1370;3273; 5444	blood microparticle
	GO:0031012	496	7	0.62	11.3	1.70E-04	712;713;714;1191;1370; 3273;64386	extracellular matrix
	GO:0005581	87	3	0.11	27.6	0.02	712;713;714	collagen trimer
	GO:0017171	208	5	0.26	19.2	5.08E-04	712;713;714;715;716	serine hydrolase activity
	GO:0004175	436	6	0.54	11.0	1.11E-03	712;713;714;715;716; 64386	endopeptidase activity
	hsa04610	79	11	0.15	74.3	0.00E+00	1191;712;713;714;715;716; 727;730;731;732;735	Complement and coagulation cascades
	hsa05020	35	8	0.07	121. 9	3.30E-14	712;713;714;727;730;731; 732;735	Prion diseases
	hsa05322	133	10	0.25	40.1	2.20E-13	712;713;714;715;716;727; 730;731;732;735	Systemic lupus erythematosus
	hsa05150	56	6	0.10	57.2	2.87E-08	712;713;714;715;716;727	Staphylococcus aureus infection
	hsa05133	76	6	0.14	42.1	1.52E-07	712;713;714;715;716;727	Pertussis
	hsa05146	96	3	0.18	16.7	0.03	731;732;735	Amoebiasis
	hsa05142	102	3	0.19	15.7	0.03	712;713;714	Chagas disease (American trypanosomiasis)
	HP:0005339	23	10	0.09	113. 8	0.00E+00	712;713;714;715;716;727; 730;731;732;735	Abnormality of complement system
	HP:0004431	20	9	0.08	117. 8	0.00E+00	731;732;712;713;714;715; 727;730;735	Complement deficiency
	HP:0005368	182	10	0.70	14.4	6.92E-08	712;713;714;715;716;727; 730;731;732;735	Abnormality of humoral immunity
	HP:0002960	117	6	0.45	13.4	1.45E-03	712;713;714;715;716;731	Autoimmunity
	HP:0000123	33	4	0.13	31.7	2.49E-03	712;713;714;715	Nephritis
UNC119*	GO:0031109	108	3	0.23	13.1	0.33	402;1070;6904	microtubule polymerization or depolymerization
	GO:0006399	183	3	0.39	7.7	1.00	1787;5464;81890	tRNA metabolic process
	GO:0006457	210	3	0.45	6.7	1.00	6903;6904;6905	protein folding
	GO:0016328	56	3	0.12	25.2	0.29	402;6904;7976	lateral plasma membrane
	GO:0031970	87	3	0.18	16.2	0.33	402;10063;23568	organelle envelope lumen
	GO:0005874	402	5	0.85	5.9	0.33	6903;6904;6905;27436; 84056	microtubule

GO:0045171	59	2	0.13	16.0	1.00	2948;9094	intercellular bridge
GO:0015631	321	5	0.68	7.3	0.33	1070;6903;6904;27436;84056	tubulin binding
GO:0051087	100	3	0.21	14.1	0.33	6903;6904;6905	chaperone binding
GO:0016782	70	2	0.15	13.4	1.00	6783;6820	transferase activity, transferring sulfur-containing groups
hsa05225	168	3	0.31	9.5	0.81	2948;55274;7976	Hepatocellular carcinoma
hsa00140	60	2	0.11	17.8	0.81	6783;6820	Steroid hormone biosynthesis
hsa05166	255	3	0.48	6.3	1.00	1031;3117;7976	Human T-cell leukemia virus 1 infection
hsa00190	133	2	0.25	8.0	1.00	10063;5464	Oxidative phosphorylation
hsa04390	154	2	0.29	6.9	1.00	5522;7976	Hippo signaling pathway
hsa04934	154	2	0.29	6.9	1.00	1031;7976	Cushing syndrome
hsa05310	31	1	0.06	17.2	1.00	3117	Asthma
hsa05330	38	1	0.07	14.0	1.00	3117	Allograft rejection
hsa05332	41	1	0.08	13.0	1.00	3117	Graft-versus-host disease
hsa04940	43	1	0.08	12.4	1.00	3117	Type I diabetes mellitus
HP:0002717	33	2	0.07	29.7	1.00	7809;1031	Adrenal overactivity
HP:0006829	60	2	0.12	16.4	1.00	6904;6905	Severe muscular hypotonia
HP:0011767	60	2	0.12	16.4	1.00	1031;6905	Abnormality of the parathyroid physiology
HP:0002510	62	2	0.13	15.8	1.00	6904;6905	Spastic tetraplegia
HP:0000828	67	2	0.14	14.6	1.00	1031;6905	Abnormality of the parathyroid gland
HP:0010929	217	3	0.44	6.8	1.00	1031;6905;7809	Abnormality of cation homeostasis
HP:0003111	241	3	0.49	6.1	1.00	1031;6905;7809	Abnormality of ion homeostasis
HP:0040077	91	2	0.19	10.8	1.00	1031;6905	Abnormal concentration of calcium in blood
HP:0010783	101	2	0.21	9.7	1.00	1031;6820	Erythema
HP:0003117	316	3	0.64	4.7	1.00	1031;23568;6905	Abnormality of circulating hormone level

*Modulo sem enriquecimento, resultado dos top 10 para cada banco

8. REFERÊNCIAS BIBLIOGRÁFICAS

- ABEP. Critério Brasil 2015 e atualização da distribuição de classes para 2016. Critério Classif econômica Bras. 2016;
- Achenbach TM, Dumenci L. Advances in empirically based assessment: Revised cross-informant syndromes and new DSM-oriented scales for the CBCL, YSR, and TRF: Comment on Lengua, Sadowski, Friedrich, and Fisher (2001). *J Consult Clin Psychol* [Internet]. 2001 [cited 2014 Oct 5];69(4):699–702. Available from: <http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-006X.69.4.699>
- Agerbo E, Sullivan PF, Vilhjálmsdóttir BJ, Pedersen CB, Mors O, Børghlum AD, et al. Polygenic Risk Score, Parental Socioeconomic Status, Family History of Psychiatric Disorders, and the Risk for Schizophrenia. *JAMA Psychiatry* [Internet]. 2015 Jul 1 [cited 2019 Mar 5];72(7):635. Available from: <http://archpsyc.jamanetwork.com/article.aspx?doi=10.1001/jamapsychiatry.2015.0346>
- Altmann A, Weber P, Bader D, Preuss M, Binder EB, Müller-Myhsok B. A beginners guide to SNP calling from high-throughput DNA-sequencing data. *Hum Genet* [Internet]. 2012 Oct [cited 2014 Jul 9];131(10):1541–54. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22886560>
- Altshuler DM, Gibbs R a, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, et al. Integrating common and rare genetic variation in diverse human populations. *Nature* [Internet]. 2010 Sep 2 [cited 2014 Jul 10];467(7311):52–8. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3173859&tool=pmcentrez&rendertype=abstract>
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (5th Edition) [Internet]. 5th ed. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition. Washington D.C: American Psychiatric Association; 2013. Available from: <http://dsm.psychiatryonline.org//content.aspx?bookid=556§ionid=41101754>
- Amiet C, Gourfinkel-An I, Bouzamondo A, Tordjman S, Baulac M, Lechat P, et al. Epilepsy in Autism is Associated with Intellectual Disability and Gender: Evidence from a Meta-Analysis. *Biol Psychiatry* [Internet]. 2008;64(7):577–82. Available from: <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed8&AN=2008420101>
- Andrews S V., Ellis SE, Bakulski KM, Sheppard B, Croen LA, Hertz-Picciotto I, et al. Cross-tissue integration of genetic and epigenetic data offers insight into autism spectrum disorder. *Nat Commun* [Internet]. 2017 Dec 24 [cited 2019 Mar 5];8(1):1011. Available from: <http://www.nature.com/articles/s41467-017-00868-y>

- Andrews T, Honti F, Pfundt R, Leeuw N de, Hehir-Kwa J, Silfhout AV, et al. The clustering of functionally related genes contributes to CNV-mediated disease. *Genome Res* [Internet]. 2015;25(6):802–13. Available from: <http://genome.cshlp.org/content/25/6/802%5Cnhttp://genome.cshlp.org/content/25/6/802.full.pdf%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/25887030>
- Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, Baird G, et al. Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum Mol Genet* [Internet]. 2012 Nov 1 [cited 2013 May 27];21(21):4781–92. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3471395&tool=pmcentrez&rendertype=abstract>
- Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* [Internet]. 2014 May 15 [cited 2018 Mar 27];30(10):1363–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24478339>
- Atladóttir HO, Henriksen TB, Schendel DE, Parner ET. Autism After Infection, Febrile Episodes, and Antibiotic Use During Pregnancy: An Exploratory Study. *Pediatrics* [Internet]. 2012 Dec 1 [cited 2019 Mar 5];130(6):e1447–54. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23147969>
- Atladóttir HÓ, Parner ET, Schendel D, Dalsgaard S, Thomsen PH, Thorsen P. Time Trends in Reported Diagnoses of Childhood Neuropsychiatric Disorders. *Arch Pediatr Adolesc Med* [Internet]. 2007 Feb 1 [cited 2019 Mar 5];161(2):193. Available from: <http://archpedi.jamanetwork.com/article.aspx?doi=10.1001/archpedi.161.2.193>
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. In: *Current Protocols in Bioinformatics* [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2013. p. 11.10.1-11.10.33. Available from: <http://doi.wiley.com/10.1002/0471250953.bi1110s43>
- Babenko O, Kovalchuk I, Metz GAS. Stress-induced perinatal and transgenerational epigenetic programming of brain development and mental health. *Neurosci Biobehav Rev* [Internet]. 2015 Jan [cited 2019 Mar 5];48:70–91. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0149763414003091>
- Becker M. Tradução e validação da entrevista Autism diagnostic interview-revised (ADI-R) para diagnóstico de autismo no Brasil. 2009 [cited 2013 Mar 10]; Available from: <http://www.lume.ufrgs.br/handle/10183/16449>
- Berko ER, Greally JM. How might epigenetic dysregulation in early embryonic life contribute to autism spectrum disorder? *Epigenomics* [Internet]. 2015 Feb;7(1):1–4. Available from:

<http://www.futuremedicine.com/doi/abs/10.2217/epi.14.86>

- Bescoby-Chambers N, Forster P, Bates G. Foetal valproate syndrome and autism: additional evidence of an association. *Dev Med Child Neurol* [Internet]. 2001 Dec 18 [cited 2019 Mar 5];43(12):847. Available from: <http://doi.wiley.com/10.1017/S0012162201211542>
- Bilder DA, Bakian A V, Viskochil J, Clark EAS, Botts EL, Smith KR, et al. Maternal Prenatal Weight Gain and Autism Spectrum Disorders. *Pediatrics* [Internet]. 2013a Nov 1 [cited 2019 Mar 5];132(5):e1276–83. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24167172>
- Bilder DA, Bakian A V, Viskochil J, Clark EAS, Botts EL, Smith KR, et al. Maternal Prenatal Weight Gain and Autism Spectrum Disorders. *Pediatrics* [Internet]. 2013b Nov 1;132(5):e1276–83. Available from: <http://pediatrics.aappublications.org/cgi/doi/10.1542/peds.2013-1188>
- Bordin I a, Rocha MM, Paula CS, Teixeira MCT V, Achenbach TM, Rescorla L a, et al. Child Behavior Checklist (CBCL), Youth Self-Report (YSR) and Teacher's Report Form (TRF): an overview of the development of the original and Brazilian versions. *Cad Saude Publica* [Internet]. 2013;29(1):13–28. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23370021>
- Braithwaite E, Kundakovic M, Ramchandani P, Murphy S, Champagne F. Maternal prenatal depressive symptoms predict infant NR3C1 1F and BDNF IV DNA methylation. *Epigenetics* [Internet]. 2015 May 4;10(5):408–17. Available from: <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed17&NEWS=N&AN=71846221>
- Brockman W, Alvarez P, Young S, Garber M, Giannoukos G, Lee WL, et al. Quality scores and SNP detection in sequencing-by-synthesis systems. *Genome Res* [Internet]. 2008 May [cited 2014 Jul 16];18(5):763–70. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2336812&tool=pmcentrez&rendertype=abstract>
- Chen Y, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* [Internet]. 2013 Feb 27 [cited 2019 Mar 5];8(2):203–9. Available from: <http://www.tandfonline.com/doi/abs/10.4161/epi.23470>
- Chlebowski C, Green J a., Barton ML, Fein D. Using the Childhood Autism Rating Scale to Diagnose Autism Spectrum Disorders. *J Autism Dev Disord* [Internet]. 2010 Jul 7;40(7):787–99. Available from: <http://link.springer.com/10.1007/s10803-009-0926-x>
- Cholemkery H, Medda J, Lempp T, Freitag CM. Classifying Autism Spectrum Disorders by ADI-

- R: Subtypes or Severity Gradient? *J Autism Dev Disord* [Internet]. 2016 Jul 8 [cited 2019 Mar 5];46(7):2327–39. Available from: <http://link.springer.com/10.1007/s10803-016-2760-2>
- Ciernia AV, LaSalle J. The landscape of DNA methylation amid a perfect storm of autism aetiologies. *Nat Rev Neurosci* [Internet]. 2016;17(7):411–23. Available from: <http://dx.doi.org/10.1038/nrn.2016.41>
- Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* [Internet]. 2010 Jun [cited 2013 May 22];11(6):415–25. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20479773>
- Connolly N, Anixt J, Manning P, Ping-I Lin D, Marsolo KA, Bowers K. Maternal metabolic risk factors for autism spectrum disorder-An analysis of electronic medical records and linked birth data. *Autism Res* [Internet]. 2016 Aug 1 [cited 2019 Mar 5];9(8):829–37. Available from: <http://doi.wiley.com/10.1002/aur.1586>
- Croen LA, Braunschweig D, Haapanen L, Yoshida CK, Fireman B, Grether JK, et al. Maternal Mid-Pregnancy Autoantibodies to Fetal Brain Protein: The Early Markers for Autism Study. *Biol Psychiatry* [Internet]. 2008 Oct 1 [cited 2019 Mar 5];64(7):583–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18571628>
- Daniels JL, Forssen U, Hultman CM, Cnattingius S, Savitz D a, Feychting M, et al. Parental Psychiatric Disorders Associated With Autism Spectrum Disorders in the Offspring. *Pediatrics* [Internet]. 2008 May 1;121(5):e1357–62. Available from: <http://pediatrics.aappublications.org/cgi/doi/10.1542/peds.2007-2296>
- Delobel-Ayoub M, Ehlinger V, Klapouszczak D, Maffre T, Raynaud J-P, Delpierre C, et al. Socioeconomic Disparities and Prevalence of Autism Spectrum Disorders and Intellectual Disability. Triche EW, editor. *PLoS One* [Internet]. 2015 Nov 5;10(11):e0141964. Available from: <http://dx.plos.org/10.1371/journal.pone.0141964>
- DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* [Internet]. 2011 May [cited 2014 May 23];43(5):491–8. Available from: <http://www.nature.com/ng/journal/v43/n5/abs/ng.806.html>
- Devlin B, Scherer SW. Genetic architecture in autism spectrum disorder. *Curr Opin Genet Dev* [Internet]. 2012 Jun [cited 2012 Nov 2];22(3):229–37. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22463983>
- Dodds L, Fell DB, Shea S, Armson BA, Allen AC, Bryson S. The Role of Prenatal, Obstetric and Neonatal Factors in the Development of Autism. *J Autism Dev Disord* [Internet]. 2011 Jul 5 [cited 2019 Mar 5];41(7):891–902. Available from: <http://link.springer.com/10.1007/s10803-010-1114-8>

- Doyle AE, Biederman J, Ferreira MAR, Wong P, Smoller JW, Faraone S V. Suggestive linkage of the child behavior checklist juvenile bipolar disorder phenotype to 1p21, 6p21, and 8q21. *J Am Acad Child Adolesc Psychiatry*. 2010;49(4):378–87.
- Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* [Internet]. 2010 Dec 30 [cited 2019 Mar 5];11(1):587. Available from: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-587>
- Eisinger BE, Saul MC, Driessen TM, Gammie SC. Development of a versatile enrichment analysis tool reveals associations between the maternal brain and mental health disorders, including autism. *BMC Neurosci* [Internet]. 2013 Dec 19 [cited 2019 Mar 5];14(1):147. Available from: <https://bmcneurosci.biomedcentral.com/articles/10.1186/1471-2202-14-147>
- Fernandez BA, Scherer SW. Syndromic autism spectrum disorders: moving from a clinically defined to a molecularly defined approach. *Dialogues Clin Neurosci* [Internet]. 2017;19(4):353–71. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29398931>
- Fombonne E. The Prevalence of Autism. *JAMA* [Internet]. 2003 Jan 1;289(1):87–9. Available from: <http://dx.doi.org/10.1001/jama.289.1.87>
- Frans EM, Sandin S, Reichenberg A, Långström N, Lichtenstein P, McGrath JJ, et al. Autism risk across generations: a population-based study of advancing grandpaternal and paternal age. *JAMA psychiatry* [Internet]. 2013 May [cited 2014 Jun 4];70(5):516–21. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3701020&tool=pmcentrez&rendertype=abstract>
- Froehlich-Santino W, Londono Tobon A, Cleveland S, Torres A, Phillips J, Cohen B, et al. Prenatal and perinatal risk factors in a twin study of autism spectrum disorders. *J Psychiatr Res* [Internet]. 2014 Jul 1 [cited 2019 Mar 5];54:100–8. Available from: <https://www.sciencedirect.com/science/article/pii/S0022395614000922?via%3Dihub>
- Gardener H, Spiegelman D, Buka SL. Prenatal risk factors for autism: comprehensive meta-analysis. *Br J Psychiatry* [Internet]. 2009 Jul 2 [cited 2019 Mar 5];195(01):7–14. Available from: https://www.cambridge.org/core/product/identifier/S0007125000007285/type/journal_article
- Garg P, Joshi RS, Watson C, Sharp AJ. A survey of inter-individual variation in DNA methylation identifies environmentally responsive co-regulated networks of epigenetic variation in the human genome. Relton C, editor. *PLOS Genet* [Internet]. 2018 Oct 1 [cited 2019 Mar 5];14(10):e1007707. Available from: <http://dx.plos.org/10.1371/journal.pgen.1007707>

- Gaugler T, Klei L, Sanders SJ, Bodea C a, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variation. *Nat Genet* [Internet]. 2014 Jul 20 [cited 2014 Jul 21];46(November 2013):881–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25038753>
- Geaghan M, Cairns MJ. MicroRNA and Posttranscriptional Dysregulation in Psychiatry. *Biol Psychiatry* [Internet]. 2014;78(4):231–9. Available from: <http://dx.doi.org/10.1016/j.biopsych.2014.12.009>
- Geschwind DH, State MW. Gene hunting in autism spectrum disorder: On the path to precision medicine. *Lancet Neurol*. 2015;14(11):1109–20.
- Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* [Internet]. 2009 May 28 [cited 2013 Mar 15];459(7246):569–73. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19404257>
- Grayson DR, Guidotti A. Merging data from genetic and epigenetic approaches to better understand autistic spectrum disorder. *Epigenomics* [Internet]. 2016 Jan [cited 2019 Mar 5];8(1):85–104. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26551091>
- Grayton HM, Fernandes C, Rujescu D, Collier D a. Copy number variations in neurodevelopmental disorders. *Prog Neurobiol* [Internet]. 2012 Oct [cited 2013 Apr 10];99(1):81–91. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22813947>
- Grzadzinski R, Huerta M, Lord C. DSM-5 and autism spectrum disorders (ASDs): an opportunity for identifying ASD subtypes. *Mol Autism* [Internet]. 2013;4(1):12. Available from: <http://molecularautism.biomedcentral.com/articles/10.1186/2040-2392-4-12>
- Guo H, Wang T, Wu H, Long M, Coe BP, Li H, et al. Inherited and multiple de novo mutations in autism/developmental delay risk genes suggest a multifactorial model. *Mol Autism* [Internet]. 2018 Dec 13 [cited 2019 Mar 5];9(1):64. Available from: <https://molecularautism.biomedcentral.com/articles/10.1186/s13229-018-0247-z>
- Guo L, Du Y, Qu S, Wang J. rVarBase: an updated database for regulatory features of human variants. *Nucleic Acids Res* [Internet]. 2016 Jan 4 [cited 2019 Mar 6];44(D1):D888–93. Available from: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv1107>
- Habela CW, Song H, Ming G. Modeling synaptogenesis in schizophrenia and autism using human iPSC derived neurons. *Mol Cell Neurosci* [Internet]. 2016 Jun [cited 2019 Mar 5];73:52–62. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26655799>
- Hachiya T, Furukawa R, Shiwa Y, Ohmomo H, Ono K, Katsuoka F, et al. Genome-wide identification of inter-individually variable DNA methylation sites improves the efficacy of epigenetic association studies. *npj Genomic Med* [Internet]. 2017 Dec 13 [cited 2019 Mar

- 5];2(1):11. Available from: <http://www.nature.com/articles/s41525-017-0016-5>
- Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T, et al. Genetic heritability and shared environmental factors among twin pairs with autism. *Arch Gen Psychiatry* [Internet]. 2011 Nov [cited 2013 May 22];68(11):1095–102. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21727249>
- Hannon E, Schendel D, Ladd-Acosta C, Grove J, Hansen CS, Andrews S V., et al. Elevated polygenic burden for autism is associated with differential DNA methylation at birth. *Genome Med*. 2018;10(1):1–13.
- Hannon E, Spiers H, Viana J, Pidsley R, Burrage J, Murphy TM, et al. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat Neurosci* [Internet]. 2016 Jan 30 [cited 2019 Mar 5];19(1):48–54. Available from: <http://www.nature.com/articles/nn.4182>
- Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol* [Internet]. 2013 Dec 10 [cited 2019 Mar 5];14(10):R115. Available from: <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2013-14-10-r115>
- Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet* [Internet]. 2018 Jun 11 [cited 2019 Mar 5];19(6):371–84. Available from: <http://www.nature.com/articles/s41576-018-0004-3>
- Iossifov I, Levy D, Allen J, Ye K, Ronemus M, Lee Y-H, et al. Low load for disruptive mutations in autism genes and their biased transmission. *Proc Natl Acad Sci U S A* [Internet]. 2015;112(41):E5600–7. Available from: <http://www.pnas.org/content/112/41/E5600.abstract>
- Iossifov I, O’Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* [Internet]. 2014;515(7526):216–21. Available from: <http://dx.doi.org/10.1038/nature13908>
- Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De novo gene disruptions in children on the autistic spectrum. *Neuron* [Internet]. 2012 Apr 26 [cited 2012 Oct 30];74(2):285–99. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22542183>
- Iossifov I, Zheng T, Baron M, Gilliam TC, Rzhetsky A. Genetic-linkage mapping of complex hereditary disorders to a whole-genome molecular-interaction network. *Genome Res* [Internet]. 2008 Jul [cited 2013 May 5];18(7):1150–62. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2493404&tool=pmcentrez&rendertype=abstract>
- Islam SA, Goodman SJ, MacIsaac JL, Obradović J, Barr RG, Boyce WT, et al. Integration of DNA methylation patterns and genetic variation in human pediatric tissues help inform

- EWAS design and interpretation. *Epigenetics Chromatin* [Internet]. 2019 Dec 2 [cited 2019 Mar 5];12(1):1. Available from: <https://epigeneticsandchromatin.biomedcentral.com/articles/10.1186/s13072-018-0245-6>
- Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* [Internet]. 2012 Feb 1 [cited 2019 Mar 5];41(1):200–9. Available from: <https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dyr238>
- Jeste SS, Geschwind DH. Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat Rev Neurol* [Internet]. 2014 Feb [cited 2014 Jul 9];10(2):74–81. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24468882>
- Jiao Y, Widschwendter M, Teschendorff AE. A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. *Bioinformatics* [Internet]. 2014 Aug 15 [cited 2018 Mar 27];30(16):2360–6. Available from: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btu316>
- Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* [Internet]. 2007 Jan;8(1):118–27. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16632515>
- Jokiranta E, Brown AS, Heinimaa M, Cheslack-Postava K, Suominen A, Sourander A. Parental psychiatric disorders and autism spectrum disorders. *Psychiatry Res* [Internet]. 2013 May;207(3):203–11. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0165178113000097>
- Kalkbrenner AE, Schmidt RJ, Penlesky AC. Environmental Chemical Exposures and Autism Spectrum Disorders: A Review of the Epidemiological Evidence. *Curr Probl Pediatr Adolesc Health Care* [Internet]. 2014 Nov;44(10):277–318. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S1538544214000741>
- Karino CA, Laros JA, Jesus GR de. Evidências de validade convergente do SON-R 2½-7[a] com o WPPSI-III e WISC-III. *Psicol Reflexão e Crítica* [Internet]. 2011 [cited 2017 Sep 22];24(4):621–9. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-79722011000400001&lng=pt&nrm=iso&tlng=en
- Keil KP, Lein PJ. DNA methylation: a mechanism linking environmental chemical exposures to risk of autism spectrum disorders? *Environ Epigenetics* [Internet]. 2016 Jan 30;2(1):dvv012. Available from: <https://academic.oup.com/eep/article-lookup/doi/10.1093/eep/dvv012>

- Kinney DK, Barch DH, Chayka B, Napoleon S, Munir KM. Environmental risk factors for autism: do they help cause de novo genetic mutations that contribute to the disorder? *Med Hypotheses* [Internet]. 2010 Jan [cited 2013 Jun 9];74(1):102–6. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2788022&tool=pmcentrez&rendertype=abstract>
- Kinney DK, Miller AM, Crowley DJ, Huang E, Gerber E. Autism Prevalence Following Prenatal Exposure to Hurricanes and Tropical Storms in Louisiana. *J Autism Dev Disord* [Internet]. 2008 Mar 6;38(3):481–8. Available from: <http://link.springer.com/10.1007/s10803-007-0414-0>
- Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, Willsey AJ, et al. Common genetic variants, acting additively, are a major source of risk for autism. *Mol Autism* [Internet]. 2012 Jan 15 [cited 2012 Oct 25];3(1):9. Available from: <http://molecularautism.biomedcentral.com/articles/10.1186/2040-2392-3-9>
- Krakowiak P, Walker CK, Bremer AA, Baker AS, Ozonoff S, Hansen RL, et al. Maternal Metabolic Conditions and Risk for Autism and Other Neurodevelopmental Disorders. *Pediatrics* [Internet]. 2012 May 1 [cited 2019 Mar 5];129(5):e1121–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22492772>
- Krakowiak P, Walker CK, Tancredi D, Hertz-Picciotto I, Van de Water J. Autism-specific maternal anti-fetal brain autoantibodies are associated with metabolic conditions. *Autism Res* [Internet]. 2017 Jan;10(1):89–98. Available from: <http://doi.wiley.com/10.1002/aur.1657>
- Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, et al. Excess of rare, inherited truncating mutations in autism. *Nat Genet* [Internet]. 2015 May 11;47(6):582–8. Available from: <http://www.nature.com/doifinder/10.1038/ng.3303>
- Kubota T, Mochizuki K. Epigenetic Effect of Environmental Factors on Autism Spectrum Disorders. *Int J Environ Res Public Health* [Internet]. 2016 May 14;13(5):504. Available from: <http://www.mdpi.com/1660-4601/13/5/504>
- de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH. Advancing the understanding of autism disease mechanisms through genetics. *Nat Med* [Internet]. 2016a Apr 1 [cited 2019 Mar 5];22(4):345–61. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27050589>
- de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH. Advancing the understanding of autism disease mechanisms through genetics. *Nat Med* [Internet]. 2016b Apr 1 [cited 2019 Mar 5];22(4):345–61. Available from: <http://www.nature.com/articles/nm.4071>
- Lai M-C, Lombardo M V, Chakrabarti B, Baron-Cohen S. Subgrouping the Autism “Spectrum”: Reflections on DSM-5. Nestler E, editor. *PLoS Biol* [Internet]. 2013 Apr 23 [cited 2019 Mar

- 5];11(4):e1001544. Available from: <http://dx.plos.org/10.1371/journal.pbio.1001544>
- Lee BK, Magnusson C, Gardner RM, Blomström Å, Newschaffer CJ, Burstyn I, et al. Maternal hospitalization with infection during pregnancy and risk of autism spectrum disorders. *Brain Behav Immun* [Internet]. 2015 Feb 1 [cited 2019 Mar 5];44:100–5. Available from: <https://www.sciencedirect.com/science/article/pii/S0889159114004528?via%3Dihub>
- Legido A, Jethva R, Goldenthal MJ. Mitochondrial dysfunction in autism. *Semin Pediatr Neurol* [Internet]. 2013 Sep [cited 2014 Jan 10];20(3):163–75. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24331358>
- Leyfer OT, Folstein SE, Bacalman S, Davis NO, Dinh E, Morgan J, et al. Comorbid Psychiatric Disorders in Children with Autism: Interview Development and Rates of Disorders. *J Autism Dev Disord* [Internet]. 2006 Oct 15 [cited 2014 Sep 2];36(7):849–61. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16845581>
- Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* [Internet]. 2010;26(5):589–95. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20080505>
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* [Internet]. 2009 Aug 15 [cited 2014 Jul 9];25(16):2078–9. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2723002&tool=pmcentrez&rendertype=abstract>
- Li M, Santpere G, Imamura Kawasawa Y, Evgrafov O V., Gulden FO, Pochareddy S, et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* (80-) [Internet]. 2018 Dec 14;362(6420):eaat7615. Available from: <http://www.sciencemag.org/lookup/doi/10.1126/science.aat7615>
- Lo-Castro A, Curatolo P. Epilepsy associated with autism and attention deficit hyperactivity disorder: Is there a genetic link? *Brain Dev* [Internet]. 2013 May 29 [cited 2013 Jun 6]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23726375>
- Loke YJ, Hannan AJ, Craig JM. The Role of Epigenetic Change in Autism Spectrum Disorders. *Front Neurol* [Internet]. 2015 May 26;6(MAY):107. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26074864>
- Lyll K, Pauls DL, Spiegelman D, Ascherio A, Santangelo SL. Pregnancy complications and obstetric suboptimality in association with autism spectrum disorders in children of the nurses' health study II. *Autism Res* [Internet]. 2012 Feb 1 [cited 2019 Mar 5];5(1):21–30. Available from: <http://doi.wiley.com/10.1002/aur.228>
- Maher BS. Polygenic Scores in Epidemiology: Risk Prediction, Etiology, and Clinical Utility. *Curr*

- Epidemiol Reports [Internet]. 2015 Dec 28 [cited 2019 Mar 5];2(4):239–44. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26664818>
- Marshall C, Noor A, Vincent J. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* [Internet]. 2008 [cited 2012 Nov 30];82(February):477–88. Available from: <http://www.sciencedirect.com/science/article/pii/S0002929707000353>
- Matson JL, Nebel-Schwalm MS. Comorbid psychopathology with autism spectrum disorder in children: an overview. *Res Dev Disabil* [Internet]. 2007 [cited 2013 May 22];28(4):341–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16765022>
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* [Internet]. 2010 Sep [cited 2014 Jul 9];20(9):1297–303. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2928508&tool=pmcentrez&rendertype=abstract>
- Miles JH, Takahashi TN, Bagby S, Sahota PK, Vaslow DF, Wang CH, et al. Essential versus complex autism: definition of fundamental prognostic subtypes. *Am J Med Genet A* [Internet]. 2005 Jun 1 [cited 2014 Oct 13];135(2):171–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15887228>
- Miller GE, Chen E, Fok AK, Walker H, Lim A, Nicholls EF, et al. Low early-life social class leaves a biological residue manifested by decreased glucocorticoid and increased proinflammatory signaling. *Proc Natl Acad Sci* [Internet]. 2009 Aug 25;106(34):14716–21. Available from: <http://www.pnas.org/cgi/doi/10.1073/pnas.0902971106>
- Mor M, Nardone S, Sams DS, Elliott E. Hypomethylation of miR-142 promoter and upregulation of microRNAs that target the oxytocin receptor gene in the autism prefrontal cortex. *Mol Autism* [Internet]. 2015 Dec 14 [cited 2019 Mar 5];6(1):46. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26273428>
- Moss BG, Chugani DC. Increased Risk of Very Low Birth Weight, Rapid Postnatal Growth, and Autism in Underweight and Obese Mothers. *Am J Heal Promot* [Internet]. 2014 Jan 2 [cited 2019 Mar 5];28(3):181–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23875984>
- Munson J, Dawson G, Sterling L, Beauchaine T, Zhou A, Koehler E, et al. Evidence for Latent Classes of IQ in Young Children With Autism Spectrum Disorder. Frances Connors, editor. *Am J Ment Retard* [Internet]. 2008 Nov [cited 2019 Mar 5];113(6):439–52. Available from: <http://www.aaiddjournals.org/doi/10.1352/2008.113%3A439-452>
- Nardone S, Sams DS, Reuveni E, Getselter D, Oron O, Karpuj M, et al. DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. *Transl Psychiatry* [Internet]. 2014;4(9):e433. Available from:

<http://www.nature.com/doi/10.1038/tp.2014.70>
<http://www.ncbi.nlm.nih.gov/pubmed/25180572>
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4203003>

Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* [Internet]. 2012 May 10 [cited 2012 Nov 1];485(7397):242–5. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/22495311>

Newschaffer CJ, Fallin D, Lee NL. Heritable and nonheritable risk factors for autism spectrum disorders. *Epidemiol Rev*. 2002;24(2):137–53.

O’Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* [Internet]. 2011 Jun [cited 2012 Nov 1];43(6):585–9. Available from:
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3115696&tool=pmcentrez&rendertype=abstract>

O’Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* (80-) [Internet]. 2012 Nov 15 [cited 2012 Nov 17];338(6114):1619–22. Available from:
<http://www.sciencemag.org/cgi/doi/10.1126/science.1227764>

Oerlemans AM, Hartman CA, Franke B, Buitelaar JK, Rommelse NNJ. Does the cognitive architecture of simplex and multiplex ASD families differ? *J Autism Dev Disord* [Internet]. 2016 Feb 4 [cited 2019 Mar 5];46(2):489–501. Available from:
<http://link.springer.com/10.1007/s10803-015-2572-9>

Ornoy A, Weinstein- Fudim L, Ergaz Z. Genetic Syndromes, Maternal Diseases and Antenatal Factors Associated with Autism Spectrum Disorders (ASD). *Front Neurosci* [Internet]. 2016 Jul 6;10. Available from:
<http://journal.frontiersin.org/Article/10.3389/fnins.2016.00316/abstract>

Ousley O, Cermak T. Autism Spectrum Disorder: Defining Dimensions and Subgroups. *Curr Dev Disord Reports* [Internet]. 2014 Mar 20 [cited 2019 Mar 5];1(1):20–8. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/25072016>

Palma-Gudiel H, Córdova-Palomera A, Eixarch E, Deuschle M, Fañanás L. Maternal psychosocial stress during pregnancy alters the epigenetic signature of the glucocorticoid receptor gene promoter in their offspring: a meta-analysis. *Epigenetics* [Internet]. 2015 Oct 3;10(10):893–902. Available from:
<http://www.tandfonline.com/doi/full/10.1080/15592294.2015.1088630>

Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing

- data. PLoS One [Internet]. 2012 Jan [cited 2014 Jul 10];7(2):e30619. Available from: <http://dx.doi.org/10.1371/journal.pone.0030619>
- Patterson PH. Maternal infection and immune involvement in autism. Trends Mol Med [Internet]. 2011 Jul 1 [cited 2019 Mar 5];17(7):389–94. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21482187>
- Pereira A, Riesgo RS, Wagner MB. Childhood autism: translation and validation of the Childhood Autism Rating Scale for use in Brazil. J Pediatr (Rio J) [Internet]. 2008 Oct 14 [cited 2014 Oct 13];84(6):487–94. Available from: http://www.jped.com.br/conteudo/Ing_resumo.asp?varArtigo=1884&cod=&idSecao=1
- Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, V Lord R, et al. De novo identification of differentially methylated regions in the human genome. Epigenetics Chromatin [Internet]. 2015 Dec 27 [cited 2019 Mar 5];8(1):6. Available from: <https://epigeneticsandchromatin.biomedcentral.com/articles/10.1186/1756-8935-8-6>
- Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. PLoS Genet. 2013;9(8).
- Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature [Internet]. 2010 Jul 15 [cited 2014 Jul 15];466(7304):368–72. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3021798&tool=pmcentrez&rendertype=abstract>
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics [Internet]. 2010 Mar 15 [cited 2019 Mar 5];26(6):841–2. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20110278>
- Reynolds LC, Inder TE, Neil JJ, Pineda RG, Rogers CE. Maternal obesity and increased risk for autism and developmental delay among very preterm infants. J Perinatol [Internet]. 2014 Sep 8 [cited 2019 Mar 5];34(9):688–92. Available from: <http://www.nature.com/articles/jp201480>
- Ribeiro de Jesus G. Normatização e validação do teste não-verbal de inteligência SON-R 2½-7[a] para o Brasil. 2009.
- Ringnér M. What is principal component analysis? Nat Biotechnol [Internet]. 2008 Mar;26(3):303–4. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18327243
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res [Internet]. 2015

- Apr 20 [cited 2019 Mar 5];43(7):e47–e47. Available from:
<http://academic.oup.com/nar/article/43/7/e47/2414268/limma-powers-differential-expression-analyses-for>
- Ronemus M, Iossifov I, Levy D, Wigler M. The role of de novo mutations in the genetics of autism spectrum disorders. *Nat Rev Genet* [Internet]. 2014;15(2):133–41. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/24430941>
- Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* [Internet]. 2011 Jun 9 [cited 2013 Feb 28];70(5):863–85. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21658581>
- Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* [Internet]. 2015 Sep;87(6):1215–33. Available from:
<http://dx.doi.org/10.1016/j.neuron.2015.09.016>
- Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey a J, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* [Internet]. 2012 May 10 [cited 2012 Oct 26];485(7397):237–41. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/22495306>
- Schaaf CP, Zoghbi HY. Solving the autism puzzle a few pieces at a time. *Neuron* [Internet]. 2011 Jun 9 [cited 2012 Nov 1];70(5):806–8. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/21658575>
- Schanen NC. Epigenetics of autism spectrum disorders. *Hum Mol Genet* [Internet]. 2006 Oct 15 [cited 2019 Mar 5];15(suppl_2):R138–50. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/16987877>
- Schröder C, Leitão E, Wallner S, Schmitz G, Klein-Hitpass L, Sinha A, et al. Regions of common inter-individual DNA methylation differences in human monocytes: genetic basis and potential function. *Epigenetics Chromatin* [Internet]. 2017 Dec 26 [cited 2019 Mar 5];10(1):37. Available from:
<http://epigeneticsandchromatin.biomedcentral.com/articles/10.1186/s13072-017-0144-2>
- Schübeler D. Function and information content of DNA methylation. *Nature* [Internet]. 2015;517(7534):321–6. Available from:
<http://dx.doi.org/10.1038/nature14192>
- Sealey LA, Hughes BW, Sriskanda AN, Guest JR, Gibson AD, Johnson-Williams L, et al. Environmental factors in the development of autism spectrum disorders. *Environ Int* [Internet]. 2016 Mar [cited 2019 Mar 5];88:288–98. Available from:

<http://www.ncbi.nlm.nih.gov/pubmed/26826339>

- Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong association of de novo copy number mutations with autism. *Science* [Internet]. 2007 Apr 20 [cited 2013 Jan 29];316(5823):445–9. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2993504&tool=pmcentrez&rendertype=abstract>
- Selzam S, Coleman JRI, Caspi A, Moffitt TE, Plomin R. A polygenic p factor for major psychiatric disorders. *Transl Psychiatry* [Internet]. 2018 Dec 2 [cited 2019 Mar 6];8(1):205. Available from: <http://www.nature.com/articles/s41398-018-0217-4>
- Sherry ST. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* [Internet]. 2001 Jan 1 [cited 2014 Jul 27];29(1):308–11. Available from: <http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/29.1.308>
- Shulha HP. Epigenetic Signatures of Autism. *Arch Gen Psychiatry* [Internet]. 2012 Mar 1 [cited 2019 Mar 5];69(3):314. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22065254>
- Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G. Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *J Am Acad Child Adolesc Psychiatry* [Internet]. 2008;47(8):921–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18645422>
- Simpkin AJ, Hemani G, Suderman M, Gaunt TR, Lyttleton O, Mcardle WL, et al. Prenatal and early life influences on epigenetic age in children: a study of mother–offspring pairs from two cohort studies. *Hum Mol Genet* [Internet]. 2016 Jan 1 [cited 2019 Mar 5];25(1):191–201. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26546615>
- Singer HS, Morris CM, Gause CD, Gillin PK, Crawford S, Zimmerman AW. Antibodies against fetal brain in sera of mothers with autistic children. *J Neuroimmunol* [Internet]. 2008 Feb [cited 2019 Mar 5];194(1–2):165–72. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18093664>
- Sparrow SS, Cicchetti D V. Diagnostic Uses of the Vineland Adaptive Behavior Scales. *J Pediatr Psychol* [Internet]. 1985;10(2):215–25. Available from: <http://jpepsy.oxfordjournals.org/cgi/doi/10.1093/jpepsy/10.2.215>
- State MW, Levitt P. The conundrums of understanding genetic risks for autism spectrum disorders. *Nat Neurosci* [Internet]. 2011 Dec [cited 2012 Oct 26];14(12):1499–506. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22037497>
- Stein R a. Epigenetics and environmental exposures. *J Epidemiol Community Health* [Internet]. 2012 Jan [cited 2012 Nov 7];66(1):8–13. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22045849>

- Stoner R, Chow ML, Boyle MP, Sunkin SM, Mouton PR, Roy S, et al. Patches of Disorganization in the Neocortex of Children with Autism. *N Engl J Med* [Internet]. 2014 Mar 27 [cited 2019 Mar 5];370(13):1209–19. Available from: <http://www.nejm.org/doi/10.1056/NEJMoa1307491>
- Sullivan EL, Nousen EK, Chamblou KA, Grove KL. The impact of maternal high-fat diet consumption on neural development and behavior of offspring. *Int J Obes Suppl* [Internet]. 2012 Dec 11 [cited 2019 Mar 5];2(S2):S7–13. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26069734>
- Teh AL, Pan H, Chen L, Ong M-L, Dogra S, Wong J, et al. The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. *Genome Res* [Internet]. 2014 Jul 1 [cited 2019 Mar 5];24(7):1064–74. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24709820>
- Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, et al. ChAMP: Updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics*. 2017;
- Touleimat N, Tost J. Complete pipeline for Infinium ® Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics* [Internet]. 2012 Jun 12 [cited 2019 Mar 5];4(3):325–41. Available from: <https://www.futuremedicine.com/doi/10.2217/epi.12.21>
- Triche TJ, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res* [Internet]. 2013 Apr 1 [cited 2019 Mar 5];41(7):e90–e90. Available from: <https://academic.oup.com/nar/article/41/7/e90/1070878>
- Vaiserman A. Epidemiologic evidence for association between adverse environmental exposures in early life and epigenetic variation: a potential link to disease susceptibility? *Clin Epigenetics* [Internet]. 2015 Dec 11 [cited 2019 Mar 5];7(1):96. Available from: <http://www.clinicalepigeneticsjournal.com/content/7/1/96>
- Voineagu I, Eapen V. Converging Pathways in Autism Spectrum Disorders: Interplay between Synaptic Dysfunction and Immune Responses. *Front Hum Neurosci* [Internet]. 2013 Jan [cited 2014 Jan 17];7(November):738. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3819618&tool=pmcentrez&rendertype=abstract>
- Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* [Internet]. 2011 May 25;474(7351):380–4. Available from: <http://www.nature.com/doi/10.1038/nature10110>

- Walker CK, Krakowiak P, Baker A, Hansen RL, Ozonoff S, Hertz-Picciotto I. Preeclampsia, Placental Insufficiency, and Autism Spectrum Disorder or Developmental Delay. *JAMA Pediatr* [Internet]. 2015 Feb 1 [cited 2019 Mar 5];169(2):154. Available from: <http://archpedi.jamanetwork.com/article.aspx?doi=10.1001/jamapediatrics.2014.2645>
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* [Internet]. 2010 Sep [cited 2012 Nov 1];38(16):e164. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2938201&tool=pmcentrez&rendertype=abstract>
- Wang Z, Liu X, Yang B-Z, Gelernter J. The role and challenges of exome sequencing in studies of human diseases. *Front Genet* [Internet]. 2013 Jan [cited 2014 Nov 21];4(August):160. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3752524&tool=pmcentrez&rendertype=abstract>
- Weiner DJ, Wigdor EM, Ripke S, Walters RK, Kosmicki JA, Grove J, et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat Genet* [Internet]. 2017 May 15;49(7):978–85. Available from: <http://www.nature.com/doifinder/10.1038/ng.3863>
- Werling DM, Geschwind DH. Understanding sex bias in autism spectrum disorder. *Proc Natl Acad Sci U S A* [Internet]. 2013 Mar 8 [cited 2013 Mar 23];110(13):4868–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23476067>
- Zhan Y, Paolicelli RC, Sforazzini F, Weinhard L, Bolasco G, Pagani F, et al. Deficient neuron-microglia signaling results in impaired functional brain connectivity and social behavior. *Nat Neurosci* [Internet]. 2014 Mar 2 [cited 2019 Mar 5];17(3):400–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24487234>