Universidade de São Paulo

Eduarda Morgana da Silva Montenegro Malaguti de Souza

# Investigação de variantes *de novo* e alterações no número de cópias na etiologia do transtorno do espectro autista

# *Investigation of de novo variants and copy number variations in autism spectrum disorder etiology*

São Paulo, 2020.

Eduarda Morgana da Silva Montenegro Malaguti de Souza

# Investigação de variantes *de novo* e alterações no número de cópias na etiologia do transtorno do espectro autista

# *Investigation of de novo variants and copy number variations in autism spectrum disorder etiology*

Tese apresentada ao Instituto de Biociências da Universidade de São Paulo, para obtenção do Título de Doutorado em Ciências, na Área de Biologia (Genética). Orientadora: Profa. Dra. Maria Rita dos Santos e Passos-Bueno

São Paulo, 2020.

# FICHA CATALOGRÁFICA

Souza, Eduarda Morgana da Silva Montenegro Malaguti
Investigação de variantes *de novo* e alterações no número de cópias na etiologia do transtorno do espectro autista/ Eduarda Morgana da Silva Montenegro Malaguti de Souza; orientadora Maria Rita dos Santos e Passos Bueno. -- São Paulo, 2019.

125 f.

Tese (Doutorado) - Instituto de Biociências da Universidade de São Paulo, Departamento de Genética e Biologia Evolutiva.

1. Transtorno do Espectro Autista 2. Meta-análise de alterações *de novo* 3. Alterações no número de cópias 4. Variantes *de novo* em pais e probandos I. Passos Bueno, Maria Rita dos Santos, orient. II. Título.

## Comissão Julgadora

_____ _____

Prof(a). Dr(a). Prof(a). Dr(a).

_____ _____

Prof(a). Dr(a). Prof(a). Dr(a).

_____

Profa. Dra. Maria Rita dos Santos e Passos Bueno
Orientadora

Baghi, Fariba. *Became free. [* Mixed Media on Canvas]. Agora Gallery, New York.

"Love is the best thing in the world!"

Fariba Baghi

# AGRADECIMENTOS

Ao universo pela oportunidade de descobrir o amor como base primordial de tudo que faço na vida, pois é Ele quem conduz minhas ações.

A todos os professores do Brasil, minha imensa admiração e respeito.

Aos meus mentores: Verônica Roncelli, Dr. André Lima, Dr. Victor Ferraz, Dr Michael Simpson, Dra Maria Rita, Dr Stephen Scherer, e Dra Simone Lima.

Jamais haverá palavras para descrever a imensidão do amor e gratidão que sinto pelos meus pais Josi e Jackson, vocês são e para sempre serão meu alicerce. Meu amor incondicional à minha irmã Vitória, com você aprendi o amor de irmãe!

Às minhas melhores amigas da vida Bárbara e Taciana, minha vida não seria a mesma sem vocês! Obrigada por tudo!

Aos meus antepassados, muito obrigada por essas variantes tão positivamente selecionadas! Às minhas avós Maria, Marilu e Odélia, e ao meu avô Gustavo. É tanto amor, conselhos e dicas de como ser feliz, que para sempre manterei vocês em meu coração.

Aos meus tios-irmãos Meri, Rodrigo e Michele e às suas crianças amadas, Matheus, Miguel, André, Gustavinho e Júlia.

Ao Marcos e à Marcela, repito meu agradecimento, pois por mais imprevisível que seja a vida, ela sempre nos dá uma segunda chance. Ao meu primo Edu pela amizade sempre. Ao meu tio Marcelo, a Tati e meus primos Gustavo, Guilherme e Valentina, por me acolherem sempre tão bem.

À grande família de amigos que construí durante esses 30 anos de vida, que guardo sempre no coração: Rafa, Luciana, Pêra, *Gaviota*, Anão, Rá, Mari, Amandinha, Taisa, Bru, Nika, Jô, Bá, Jenny, Tata, Nane, Karina, Jéssica, Ádamo, Clarissa, Charles

Henri, Lucía, Juliana, Magno, Kamila, Greice, Jake, Inês, Christos, Vanuza, Luiza Mendes, Karina, Pontinho, Luquinhas, Cher, Leticia, Elisa, Claudinha, Belinha, Vanzinha, Sofia, Juninha, Luciano, Mari on fire, May, Dani, Naila, Simone, Gabi, Tati, GLY, Ada, Darwin, Mehdi, Wilson, Sergio, Bank, Gio, Joe, Lia, Mat, Fariba, Glau, Isa, Pri, Aninha, Mar, Mau, Dani, Rô, Bi, Angelita, Vivi, Bea, Claudinho, Jaque, Rô, Lili, Rogerinho, Gigi, Viking, Si, Débora, Deli, e Neto.

Aqueles amigos atemporais e íntimos: Simone, Ayling, Luiza, Natinha, Melina, Jojó, Carol & Léo, Clarice & Renatão, Ágatha & Paulo, Cami e Pri.

Ao meu companheiro Renan e ao seu novo mundo (Mariana, Flávia, Luan, Priscila e amigos), que com gentileza me acolheu, trazendo muitas alegrias!

Aos funcionários do departamento de genética do IB e do Genoma da USP.

Ao CNPq (bolsa de doutorado 140140/2016-6) e à FAPESP (bolsa de doutorado 2017/05824-2; BEPE 2018/13743-5), pelo apoio financeiro.

Às famílias deste estudo, que permitiram o conhecimento gerado.

# NOTAS

Esta tese de doutorado é resultante de um trabalho inédito desenvolvido entre os anos de 2016 e 2020, sob orientação da Profa. Dra. Maria Rita Passos-Bueno no Laboratório de Genética do Desenvolvimento da Universidade de São Paulo. Parte dessa tese foi também desenvolvida no laboratório The Center for Applied Genomics (TCAG), da Universidade de Toronto e do Hospital Sickkids, em colaboração com o Prof. Dr. Stephen W. Scherer, durante o período de doutorado-sanduíche.

A tese foi redigida em inglês, no formato de capítulos, que compreendem: introdução geral (capítulo I), artigo de meta-análise publicado (capítulo II), artigo não publicado sobre alterações no número de cópias (capítulo III), artigo não publicado sobre variantes *de novo* em três gerações (capítulo IV), seguidos de considerações gerais (capítulos V). Em apêndices encontram-se duas publicações em co-autoria.

O projeto que resultou na presente tese foi cadastrado na Plataforma Brasil e contou com o parecer consubstanciado do Comitê de Ética em Pesquisa do Instituto de Biociências da Universidade de São Paulo (número 1.133.486/2016).

Este trabalho recebeu apoio financeiro FAPESP (bolsa de doutorado 2017/05824-2; BEPE 2018/13743-5), e CNPq (bolsa de doutorado 140140/2016-6).

# NOTES

This PhD thesis is the result of an original work performed during 2016 and 2019. Professor Maria Rita Passos-Bueno, from the Human Genome and Stem Cell Research Center, Institute of Biosciences, University of São Paulo, São Paulo, Brazil, mentored the study. Also, Dr Stephen Scherer mentored part of this work during a sandwich-PhD, at the Center for Applied Genomics (TCAG), from the University of Toronto and Sickkids Hospital.

The thesis was organized into scientific articles that are presented as core chapters, being composed by the main Introduction (Chapter I), a published meta-analysis article (Chapter II), an unpublished article about copy number variations (Chapter III), an unpublished article about *de novo* variants analysis in three generations (Chapter IV), and general considerations (Chapter V). Additional co-authored publications were assigned to an appendix section.

The project that resulted in this thesis was registered on the Plataforma Brasil and approved by the Ethics Committee of the Institute of Biosciences at University of São Paulo (number 1.133.486/2016).

FAPESP (PhD scholarship 2017/05824-2; BEPE 2018/13743-5), and CNPq (PhD scholarship 140140/2016-6), financed this work.

# LIST OF ABBREVIATIONS

| | |
|---|---|
| **AbraOM** | Arquivo Brasileiro Online de Mutações |
| **aCGH** | Array Comparative Genome Hybridization |
| **ACMG** | American College of Medical Genetics |
| **AD** | Autosomal Dominant |
| **ADHD** | Attention Deficit Hyperactivity Disorder |
| **ADI-R** | Autism Diagnostic Interview Revised |
| **ADO-S** | Autism Diagnostic Observation Schedule |
| **ASD** | Autism Spectrum Disorder |
| **Bp** | Base pair |
| **CARS** | Childhood Autism Rating Scale |
| **CDC** | Centers for Disease Control and Prevention |
| **CEGH-CEL** | Centro de Estudos do Genoma Humano e Células Tronco |
| **CEPID** | Centro de Pesquisa, Inovação e Difusão |
| **CHD** | Congenital Heart Disorder |
| **Chr** | Chromosomal |
| **Class.** | Classification |
| **CMA** | Chromosomal Microarray Analysis |
| **CNPq** | Conselho Nacional de Desenvolvimento Científico e Tecnológico |
| **CNVs** | Copy Number Variations |
| **DDD** | Deciphering Developmental Disorders study |
| **Del** | Deletion |
| **DNA** | Deoxyribonucleic Acid |
| **DNMs** | *De Novo* Mutations |
| **DNVs** | *De Novo* Variants |
| **DSM-V** | Diagnostic and Statistical Manual of Mental Disorders to 5th edition |
| **Dup** | Duplication |
| **ExAC** | Exome Aggregation Consortium |
| **F** | Female |
| **FAPESP** | Fundação de Amparo à Pesquisa do Estado de São Paulo |
| **Fig.** | Figure |
| **FSHD** | Facioscapulohumeral Muscular Dystrophy |
| **G1** | Generation 1 |
| **G2** | Generation 2 |
| **GATK** | Genome Analysis Tools kit |
| **gnomAD** | Genome Aggregation Database |
| **HALD4** | Hyperaldosteronism Familial Type IV |
| **Hg19** | Human genome 19 |
| **IB** | Instituto de Biociências |
| **ID** | Intellectual Disability |
| **IGV** | Integrative Genomics Viewer |
| **Indels** | Insertions and Deletions |
| **Kb** | Kilobase |
| **LoF** | Loss-of-Function |
| **M** | Male |
| **MAF** | Minor Allele Frequency |
| **Mat** | Maternal |
| **Mb** | Mega base |

| | |
|---|---|
| **MLPA** | Multiplex Ligation-dependent Probe Amplification |
| **mRNA** | Messenger Ribonucleic Acid |
| **NA** | Not Available |
| **ND** | Neurodevelopmental |
| **NDD** | Neurodevelopmental Disorders |
| **NGS** | Next Generation Sequencing |
| **NPMD** | Neuropsychomotor Delay |
| **NT** | Not Tested |
| **OMIM** | Online mendelian inheritance in man |
| **ON** | Ontario |
| **OR** | Odds Ratio |
| **Pat** | Paternal |
| **PCR** | Polymerase Chain Reaction |
| **PDD** | Pervasive Developmental Disorder |
| **PDD-NOS** | Pervasive Developmental Disorder - Not Otherwise Specified |
| **PHA2E** | Pseudo-hypoaldosteronism Type IIE |
| **PhD** | Philosophy Doctor |
| **PI** | Patient Identification |
| **PVT** | Protein truncating variants |
| **ROS** | Reactive Oxygen Species |
| **RT-qPCR** | Real-time quantitative polymerase chain reaction |
| **Score S** | Score Syndromic |
| **SCZ** | Schizophrenia |
| **SFARI** | Simons Foundation Autism Research Initiative |
| **SNVs** | Single nucleotide variants |
| **SSC** | Simons Simplex Collection |
| **TCAG** | The Centre for Applied Genomics |
| **UCSC** | University of California, Santa Cruz |
| **USP** | Universidade de São Paulo |
| **VUS** | Variant of Unknown Significance |
| **WES** | Whole Exome Sequencing |
| **WHS** | Wolf-Hirschhorn syndrome |

# INDEX

# CHAPTER ONE

# GENERAL INTRODUCTION

## 1. Autism Spectrum Disorder

### 1.1. Definition and assessment instruments

Autism spectrum disorder (ASD) includes a group of heterogeneous neurodevelopmental conditions with clinical manifestation in early childhood, being mainly characterized by impairment in social and communication skills, as well as restricted and repetitive behavior (AMERICAN PSYCHIATRIC ASSOCIATION, 2013). The update of the Diagnostic and Statistical Manual of Mental Disorders to 5th edition (DSM-V) recognized a spectrum of social impairments that could be grouped, as classic autism, Asperger's syndrome and Pervasive Developmental Disorder - Not Otherwise Specified (PDD-NOS). Since DSM-V update, it has been widespread, although it is discussed that the novel classification excluded individuals with mild traits, females, and older individuals (ZELDOVICH, 2018).

ASD symptoms are ample, being common the use of assessment scales by healthcare professionals, especially psychologists. The gold standard instruments to the diagnostic assessment, recommended by NIH, are the autism diagnostic interview revised (ADI-R) (LORD; RUTTER; LE COUTEUR, 1994), and the autism diagnostic observation schedule (ADO-S) (LORD et al., 2000). However, its high costs and extensive time to application turns them more feasible in high-income countries compared to low to moderate-income countries (SAMMS-VAUGHAN et al., 2017). Taking it into account, the childhood autism rating scale (CARS) is an effective alternative tool, widely used in Brazil. This tool is based in a diagnostic behavioral

assessment scale that categorizes autistic features in: normal, moderate and severe (SCHOPLER et al., 2010).

Despite the many efforts on patient's clinical evaluation, time to diagnosis and intervention remains a significant issue in ASD (WIGGINS; BAIO; RICE, 2006; ZWAIGENBAUM; PENNER, 2018). In this context, telehealth tools are emerging as novel approaches to support diagnostic. *e.g.* the use of clinically relevant home-recorded videos and behavioral probes by parents. These telehealth approaches are showing feasibility, yet large-scale analyses are needed to validate them (NAZNEEN et al., 2015; TALBOTT et al., 2019).

### 1.2. Prevalence and sexual bias

ASD prevalence has increased significantly over the years, the estimative in 2000 was one affected in 150 individuals, rising to one in 59 in the last review of the Centers for Disease Control and Prevention (CDC) (BAIO et al., 2018). Of note, the CDC data shows a lower prevalence of ASD in Hispanic Americans kids of one affected in 71 individuals, compared to one in 58 for non-Hispanic whites. Brazilian ASD prevalence is not well defined, the unique estimative is based in a pilot study with pervasive developmental disorders (PPD), which defined the prevalence in one affected in 370 individuals (PAULA et al., 2011).

Studies of ASD prevalence evidenced that there is an unequal sexual ratio of ASD affected individuals, males are estimated to be 3 to 4 folds more affected than females (LOOMES; HULL; MANDY, 2017). These differential sexual ratio is not completely elucidated, but two main models are evidenced as possible explanations to this difference: the presence of a female protective model, that guarantee to females to harbor more disrupting genetic alterations than males due to a protective mechanism

not yet elucidated (JACQUEMONT et al., 2014; ROBINSON et al., 2013); and the males susceptibility to perturbations in genes involved in neuronal plasticity due to empirical differences in male and female brains (MOTTRON et al., 2015). The most accepted model to explain sexual ratio difference is the "female protective effect", yet a recent Swedish study that analyzed national data of 847,732 children and multigenerational families, pointed that ASD relative risk did not differ between males and females in the second generation. Then, suggesting that the female protective model may not be the principal mechanism underlying the male sex bias in ASD (BAI et al., 2020).

### 1.3. ASD and clinically relevant classifications

The use of subgroup classifications in ASD is widespread, especially considering its complex etiology and ample range of symptoms. These subgroup classifications frequently aim to better delineate patient's clinical management. Some common clinically relevant concerns are regarding a) familial recurrence, b) comorbidities, and c) associated syndromes.

#### 1.3.1. ASD recurrence: sporadic *versus* familial

ASD classification into familial or sporadic cases is extremely relevant for genetic counseling, since it may determine recurrence risks, consequently impacting familial management. Couples without previous ASD kids, present a populational risk of having affected children, meaning that sporadic cases risk is equivalent to ASD prevalence, estimated in 1,68% (1:59) (BAIO et al., 2018). However, for familial cases of ASD, wherein a couple already have an affected child, ASD recurrence risk for a second kid is estimated up to 18,6%, a 10-fold higher risk (OZONOFF et al., 2011).

The recurrence risk can vary accordingly to the gender of the first affected child (Figure 1), a couple with a previous ASD affected female have the recurrence risk of 6,7% of having a second female affected kid, and the risk of 16,7% of having a second male affected son. Meanwhile, if the first ASD affected kid is a male, the recurrence risk is 4,2% for a second female kid, and 12,9% for a second male kid (PALMER et al., 2017). Of note, for couples with two or more ASD affected children, the recurrence risk rises from 33 to 50% (OZONOFF et al., 2011; SANDIN et al., 2014) The increased risk for male siblings is expected considering the sexual ratio observed in ASD prevalence. Also, recurrence risk varies with familial distance of ASD affected individual, e.g. the presence of a first-degree cousin affected increases ASD risk in two-fold (OZONOFF et al., 2011; RISCH et al., 2014; SANDIN et al., 2014).

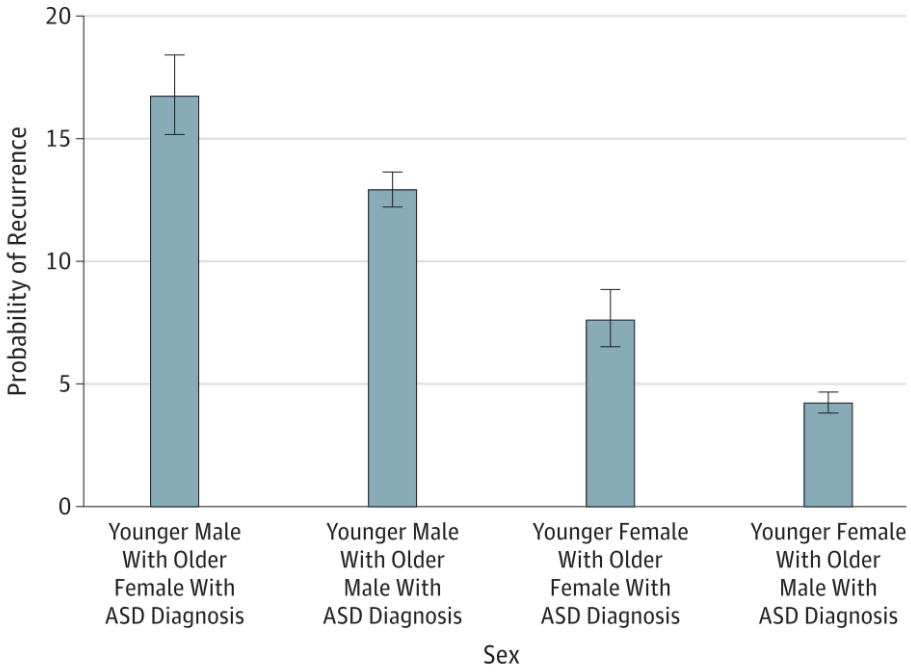


**Figure 1.** Sibling recurrence risk in Autism Spectrum Disorder. Adapted from PALMER et al. Association of Sex With Recurrence of Autism Spectrum Disorder Among Siblings. *JAMA Pediatr*. 2017;171(11):1107-1112.

### 1.3.2. ASD comorbidities: essential *versus* complex

ASD can be classified as complex, whenever some early morphological abnormality is present (as dysmorphology or microcephaly, and essential, when no abnormal morphogenesis is identified (MILES et al., 2005). More recently, dysmorphology quantification based on large control and affected cohorts have been used to create a novel approach to ASD phenotypic classification (SHAPIRA et al., 2019). It has been suggested that essential and complex subtypes may have distinct developmental and medical correlations, which may impact in prognosis, identifying medical comorbidities, directing diagnostic evaluations and treatment interventions (FLOR et al., 2017).

ASD symptoms often co-exist with other general medical condition and neuropsychiatric/behavioral alteration, being estimated that these comorbidities are present in approximately 80% of affected individuals. Frequent medical conditions associated to ASD are motor abnormalities (79%), gastrointestinal problems (70%), epilepsy (30%) (MANNION; LEADER, 2013). A recent case-control study also showed marked differences in hearing impairments (OR = 4.73), auricular disorders (OR = 5.04), neurological (OR = 8.20) and ophthalmological conditions (OR = 3.38) (DIZITZER et al., 2020). Common psychiatric comorbidities in ASD includes sleep disorders (up to 80%), schizophrenia spectrum (up to 67%), suicidal behavior disorders (up to 66%), attention-deficit/hyperactivity disorder (up to 65%), anxiety disorders (up to 54%), depressive disorders (up to 47%), intellectual disability (45%), and obsessive-compulsive disorder (up to 24%) (HOSSAIN et al., 2020; MANNION; LEADER, 2013).

### 1.3.3. ASD associated syndromes

As previously mentioned, ASD clinical manifestation includes a wide spectrum of symptoms and comorbidities, and they are often classified as non-syndromic and syndromic cases. In a small amount of ASD cases, it is possible to define a pattern in the clinical abnormalities' presentation, and these recognizable features can be classified into ASD syndromic. However, the ASD syndromic cases are estimated to represent only ~5% of all ASD cases (MILES et al., 2005; SZTAINBERG; ZOGHBI, 2016).

ASD syndromic cases are generally recognized by clinician's experience and confirmed with a targeted molecular genetic test. Some relevant syndromes that should be considered during clinical evaluation includes: X-fragile syndrome (KAUFMANN et al., 2004), Rett syndrome (NEUL, 2020) Phelan-McDermid syndrome (MUKADDES; HERGUNER, 2007), 15q11-13 deletion associated to Angelman (maternal) or Prader-Willi (paternal) syndromes (VELTMAN; CRAIG; BOLTON, 2005), 16p11.2 deletion syndrome (KUMAR et al., 2009), Cornelia de Lange syndrome (SRIVASTAVA; SCHWARTZ, 2014), Neurofibromatosis syndrome type 1 (GILLBERG; FORSELL, 1984) and Tuberous Sclerosis (VIGNOLI et al., 2015).

## 2. Genetics aspects of ASD

### 2.1. ASD multifactorial model

ASD is considered a heterogeneous complex disorder and as such, the multifactorial model is associated to its etiology, yet not explain all cases, meaning that environmental and genetic factors may play a role into clinical manifestations (SCHAEFER, 2016). Environmental factors associated to ASD usually occur during

prenatal period, and are related to mothers' exposition to drugs, infections, medicines, and toxic chemistries. More controversially, there are also some reports of birth traumas and parental care during first years of life (MANDY; LAI, 2016).

Although these environmental factors are reported associated to ASD etiology, the genetic component of ASD is considered high. A great review performed by Bourgeron (2016) analyzed studies from 1977 to 2015, aiming to clarify the genetic contribution to ASD, based on twin and heritability studies. The average concordance in monozygotic twins was estimated in 45%, reaching up 94.5%, just as average heritability was estimated in 50%, reaching up 95% (BOURGERON, 2016).

### 2.2. ASD complex genetic architecture

In the last decade, advances were made on ASD genomic field, yet much remains to be explained due the complex genetic architecture of this condition. Studies across populations evidenced different inheritance patterns and the role of both rare and common genetic variants in ASD etiology and phenotype heterogeneity (BOURGERON, 2015; GRIESI-OLIVEIRA; SERTIÉ, 2017; LEBLOND et al., 2018; TOMA, 2020; WEINER et al., 2017).

The genetic background of ASD is particular to each individual, being possible to fit any inheritance pattern known as monogenic, oligogenic and polygenic (BOURGERON, 2016). Therefore, a single rare genetic alteration can be penetrant enough to cause ASD in a monogenic pattern, or it may be necessary the combination of few rare variants in an oligogenic pattern, or yet a combination rare variants and a burden of medium to high common genetic variants in polygenic pattern (Figure 2) (BOURGERON, 2015; GRIESI-OLIVEIRA; SERTIÉ, 2017; TOMA, 2020). Even though common variants are relevant to ASD risk, their contribution are challenging to

determine (ANNEY et al., 2012; DEVLIN; MELHEMA; ROEDER, 2011). Still, a recent large consortia study performed by Grove et al., (2019) is the first to point common risk variants robustly associated with ASD. Therefore, most of ASD genes discovery were identified on studies with rare genetic alterations.

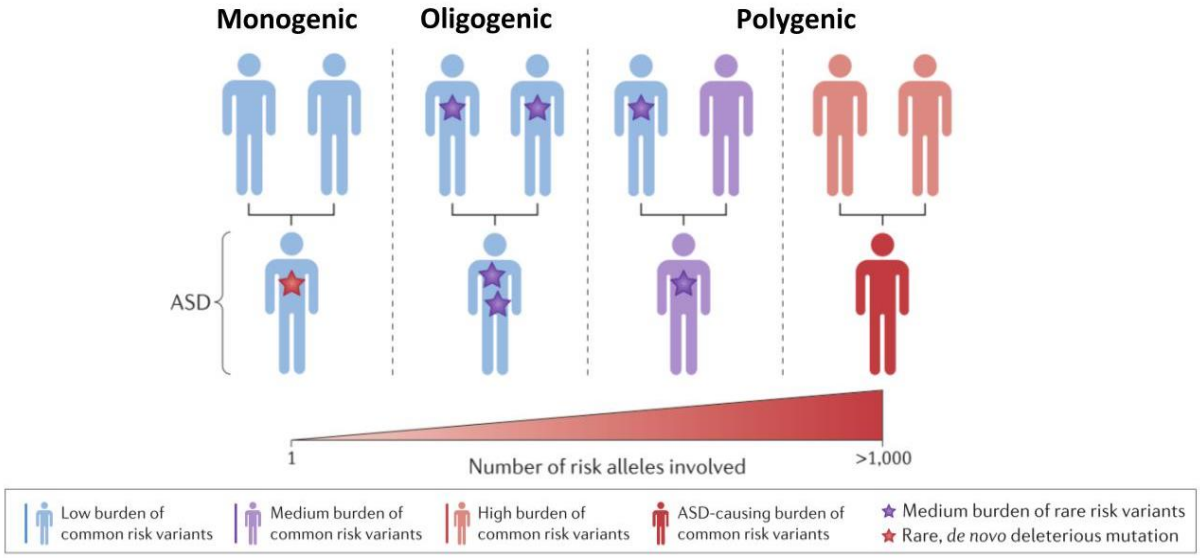


**Figure 2.** Inheritance patterns and genetics variants interplay on ASD. Adapted from Bourgeron, T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. Nat Rev Neurosci. 2015;16(9):551-63.

### 2.3. ASD genes and pathways

It is estimated that more than 1,000 genes are linked to ASD, the Simons Foundation Autism Research Initiative (SFARI) regularly compiles and evaluates studies reporting ASD candidate genes. The last SFARI update (April, 2020) presented 864 genes categorized with strong to suggestive evidence of ASD association (categories 1 to 3). Despite of the vast number of ASD linked genes, these genes convergence into common biological pathways, involved in chromatin remodeling, gene expression regulation, and synaptic function (Figure 3) (HUGUET; BENABOU; BOURGERON, 2016; LORD et al., 2020; PINTO et al., 2014; RUBEIS et al., 2014).

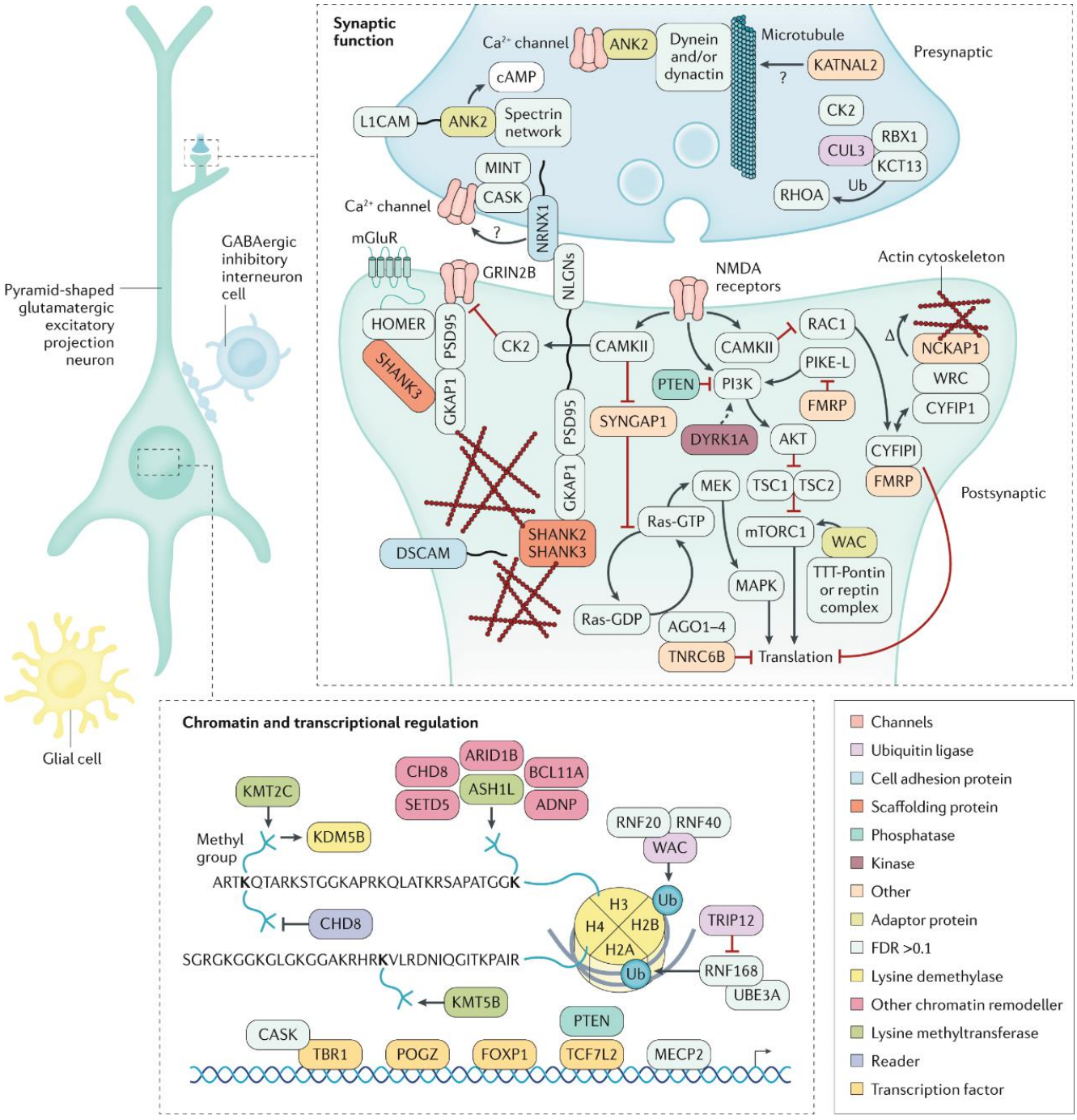


**Figure 3.** Genes and pathways associated to ASD risk. From Lord et al. Autism Spectrum Disorder. Nature Reviews. 2020; 5:6.

## 2.4. ASD genomic variation: *de novo* and inherited variants

Large genomic studies with ASD affected individuals have revealed the role of single nucleotide variants (SNVs) and small insertions and deletions (Indels), as well as copy number variants (CNVs) to ASD etiology (RAMASWAMI; GESCHWIND, 2018; SATTERSTROM et al., 2020; WOODBURY-SMITH; SCHERER, 2018). Notably, *de novo* mutations (DNMs) are pointed as relevant to ASD etiology, wherein both SNVs/Indels and CNVs genomic were enhanced on ASD cases compared to control populations. For example, *de novo* CNVs are identified in 5 to 15% of ASD cases compared to 1 to 2% in control population (PINTO et al., 2014; SANDERS et al., 2015; WOODBURY-SMITH; SCHERER, 2018). Also, exonic *de novo* SNVs/Indels were shown to occur in a higher rate (2 to 3 folds) in affected individuals, when compared to their healthy siblings (IOSSIFOV et al., 2014); and exonic *de novo* protein truncating variants (PTVs) showed an enrichment of 3.5 folds, compared to 1.2-fold for exonic inherited PTVs ($p=0.07$) in ASD probands (SATTERSTROM et al., 2020). Of note, besides *de novo* PTVs are evidenced, *de novo* missense variants are also relevant to ASD, and probably predispose ASD risk through gain-of-function (CASTELLANI; ARKING, 2020; SATTERSTROM et al., 2020).

Although it is well established DNMs role in ASD etiology, these variations are identified in clinically relevant ASD genes only in approximately 10% of cases (BOURGERON, 2016). In this context, rare inherited variation are also relevant, indeed ASD individuals present more deleterious inherited variants than general population, evidencing the additional contribution of rare inherited variants to ASD clinical manifestation (KRUMM et al., 2015; RUZZO et al., 2019; WILFERT et al., 2020). In fact, both rare inherited and *de novo* variants have been shown to impact shared networks, indicating common genetic risk pathways (RUZZO et al., 2019; SÁNCHEZ-SÁNCHEZ et al., 2018; SATTERSTROM et al., 2020).

## 2.5. DNMs and ASD candidate genes

Remarkably, DNMs analysis are shown to be an important tool for gene discovery in ASD (C YUEN et al., 2017; IOSSIFOV et al., 2012; NEALE et al., 2012; O'ROAK et al., 2012; RUBEIS et al., 2014; SANDERS et al., 2011; SATTERSTROM et al., 2020; SEBAT et al., 2007; TAKATA et al., 2018). However, *de novo* CNVs analysis for gene discovery adds more complexity to infer high-risk single genes, especially for large CNVs that harbors multiple genes. In this context, SNVs/Indels analysis are more likely to infer high risk candidate genes in ASD (ALONSO-GONZALEZ; RODRIGUEZ-FONTENLA; CARRACEDO, 2018; SANDERS et al., 2015).

Different strategies are performed to prioritize candidate genes based on coding DNMs, but they generally include the analysis of integrated datasets of expression data, sequence information, functional annotation and the biomedical literature (ALONSO-GONZALEZ; RODRIGUEZ-FONTENLA; CARRACEDO, 2018; MOREAU; TRANCHEVENT, 2012). Coding DNMs that lead to loss-of-function are frequently considered more damaging than DNMs missense variants and consequently more directly highlighted on genes discovery analysis, although as previous mentioned missense variants were shown to contribute to ASD risk through gain-of-function (ALONSO-GONZALEZ; RODRIGUEZ-FONTENLA; CARRACEDO, 2018; SATTERSTROM et al., 2020). Indeed, to assess DNMs pathogenicity is challenging, and many tools are performed to evaluate variants impact based mainly on protein function, structure, and biochemical properties, such as Polyphen2, SIFT, Provean, Mutation Taster, and CADD (ADZHUBEI et al., 2010; CHOI; CHAN, 2015; KIRCHER et al., 2014; KUMAR; HENIKOFF; NG, 2009; SCHWARZ et al., 2014). Other relevant tool that has become widely considered is the constraint metric, that

evaluate gene tolerance or intolerance to mutations, mostly based in expected versus observed number of variants in a certain gene, which is represent for missense and loss-of-function variants by gnomAD database metrics: z-score and pLI, respectively (KARCZEWSKI et al., 2020; LEK, 2016). Finally, other relevant strategies to genes prioritization are regarding statistical data analysis from variants identified in previous large studies and databases, genes function and expression data information (ALONSO-GONZALEZ; RODRIGUEZ-FONTENLA; CARRACEDO, 2018; C YUEN et al., 2017; TAKATA et al., 2018). Therefore, to analyze DNMs has proven to be a powerful approach to identify novel candidate genes, as well to validate genes not well established in ASD, being particularly interesting in poorly explored cohorts, such as the Brazilian.

### 2.6. DNMs and parental age

DNMs are generated due to errors during DNA replication and repair, its rate is influenced by DNA repair efficiency, number of mitoses a cell has undergone, and the time between mitoses (GAO et al., 2016). DMVs were traditionally considered to occur only in germline cells and they indeed represent most of them (89%), yet they are also generated during parental embryogenesis (4%) and during postzygotic events (7%) (FRANCIOLI et al., 2015; GOLDMANN et al., 2016; KONG et al., 2012; RAHBARI et al., 2016).

The paternal rate of DNMs in males germline cells can increase from one to three per year due the constant mitosis process of spermatogenesis, while maternal DNMs rate increases 0.24 to 0.44 per year before conception (GOLDMANN et al., 2016; RAHBARI et al., 2016; SASANI et al., 2019). Not surprising, approximately 80% of all DNMs are originated in the paternal allele (FRANCIOLI et al., 2015;

GOLDMANN et al., 2016; JÓNSSON et al., 2017; KONG et al., 2012). Taking it into account, it is more clear how parental increased age is associated to higher rates of de novo variants in general populations (GAO et al., 2019; JÓNSSON et al., 2017; KONG et al., 2012).

ASD prevalence has been positively correlated with parental and grand paternal age (FRANS et al., 2013; HULTMAN et al., 2011; REICHENBERG et al., 2006; SANDIN et al., 2016), which may partially explain the high proportion of DNMs in ASD probands. Considering ASD risk association to advanced parental age, and the increased number of DNMs in ASD individuals, we speculated that three-generation analyses of DNMs, could provide new insights into relevance of variants that have arisen *de novo* across generations.

**REFERENCES**

ADZHUBEI, I. A. et al. A method and server for predicting damaging missense mutations. **Nature Methods**, v. 7, n. 4, p. 248–249, 2010.

ALONSO-GONZALEZ, A.; RODRIGUEZ-FONTENLA, C.; CARRACEDO, A. De novo mutations (DNMs) in autism spectrum disorder (ASD): Pathway and network analysis. **Frontiers in Genetics**, v. 9, n. SEP, 2018.

AMERICAN PSYCHIATRIC ASSOCIATION. **Diagnostic and Statistical Manual of Mental Disorders: DSM-V.** 5th. ed. Arlington, VA: American Psychiatric Publishing, 2013.

ANNEY, R. et al. Individual common variants exert weak effects on the risk for autism spectrum disorders. **Human Molecular Genetics**, v. 21, n. 21, p. 4781–4792, 2012.

BAIO, J. et al. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. **Surveillance Summaries**, v. 67, n. 6, 2018.

BOURGERON, T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. **Nature Reviews Neuroscience**, v. 16, n. 9, p. 551–563, 2015.

BOURGERON, T. Current knowledge on the genetics of autism and propositions for future research. **Comptes Rendus - Biologies**, v. 339, n. 7–8, p. 300–307, 2016.

C YUEN, R. K. et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. **Nature Neuroscience**, v. 20, p. 602, 6 mar. 2017.

CASTELLANI, C. A.; ARKING, D. E. High-Risk, High-Reward Genetics in ASD. **Neuron**, v. 105, n. 3, p. 407–410, 2020.

CHOI, Y.; CHAN, A. P. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. **Bioinformatics**, v. 31, n. 16, p. 2745–2747, 2015.

DEVLIN, B.; MELHEMA, N.; ROEDER, K. Do common variants play a role in risk for autism? Evidence and theoretical musings. **Brain Res.**, n. 1380, p. 78–84, 2011.

DIZITZER, Y. et al. Comorbidity and health services' usage in children with autism spectrum disorder: A nested case-control study. **Epidemiology and Psychiatric Sciences**, 2020.

FLOR, J. et al. Developmental functioning and medical Co-morbidity profile of children with complex and essential autism. **Autism Research**, v. 10, n. 8, p. 1344–1352, 2017.

FRANCIOLI, L. C. et al. Genome-wide patterns and properties of de novo mutations in humans. **Nature Genetics**, v. 47, n. 7, p. 822–826, 2015.

FRANS, E. et al. Autism Risk Across Generations: A Population Based Study of Advancing Grandpaternal and Paternal Age. **JAMA psychiatry**, v. 70, n. 5, p. 516–521, 2013.

GAO, Z. et al. Interpreting the Dependence of Mutation Rates on Age and Time. **PLoS Biology**, v. 14, n. 1, p. 1–16, 2016.

GAO, Z. et al. Overlooked roles of DNA damage and maternal age in generating human germline mutations. **Proceedings of the National Academy of Sciences of the United States of America**, v. 116, n. 19, p. 9491–9500, 2019.

GILLBERG, C.; FORSELL, C. Childhood psychosis and neurofibromatosis-more than a coincidence? **Journal of Autism and Developmental Disorders**, v. 14, n. 1, p.

1–8, 1984.

GOLDMANN, J. M. et al. Parent-of-origin-specific signatures of de novo mutations. **Nature Genetics**, v. 48, n. 8, p. 935–939, 2016.

GRIESI-OLIVEIRA, K.; SERTIÉ, A. Autism spectrum disorders : an updated guide for genetic counseling. **Einstein**, v. 15, n. 55 11, p. 233–238, 2017.

HOSSAIN, M. M. et al. Prevalence of comorbid psychiatric disorders among people with autism spectrum disorder: An umbrella review of systematic reviews and meta-analyses. **Psychiatry Research**, v. 287, p. 1–30, 2020.

HUGUET, G.; BENABOU, M.; BOURGERON, T. The Genetics of Autism Spectrum Disorders. **Research and Perspectives in Endocrine Interactions**, p. i–xiii, 2016.

HULTMAN, C. M. et al. Advancing paternal age and risk of autism: new evidence from a population-based study and a meta-analysis of epidemiological studies. **Molecular psychiatry**, v. 16, n. 12, p. 1–10, 2011.

IOSSIFOV, I. et al. De novo gene disruption in children on the autistic spectrum. **Neuron**, v. 74, n. 2, p. 285–299, 2012.

IOSSIFOV, I. et al. The contribution of de novo coding mutations to autism spectrum disorder. **Nature**, v. 13, n. 515, p. 216–221, 2014.

JACQUEMONT, S. et al. A higher mutational burden in females supports a “female protective model” in neurodevelopmental disorders. **American Journal of Human Genetics**, v. 94, n. 3, p. 415–425, 2014.

JÓNSSON, H. et al. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. **Nature**, v. 549, n. 7673, p. 519–522, 2017.

KARCZEWSKI, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. **Nature**, v. 581, n. 7809, p. 434–443, 2020.

KAUFMANN, W. E. et al. Autism spectrum disorder in fragile X syndrome: Communication, social interaction, and specific behaviors. **American Journal of Medical Genetics**, v. 129 A, n. 3, p. 225–234, 2004.

KIRCHER, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. **Nature Genetics**, v. 46, n. 3, p. 310–315, 2014.

KONG, A. et al. Rate of de novo mutations and the importance of father-s age to disease risk. **Nature**, v. 488, n. 7412, p. 471–475, 2012.

KRUMM, N. et al. Excess of rare, inherited truncating mutations in autism. **Nature genetics**, v. 47, n. 6, p. 582–588, 2015.

KUMAR, P.; HENIKOFF, S.; NG, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. **Nature Protocols**, v. 4, n. 7, p. 1073–1082, 2009.

KUMAR, R. A. et al. Association and mutation analyses of 16p11.2 autism candidate genes. **PLoS ONE**, v. 4, n. 2, 2009.

LEBLOND, C. S. et al. Both rare and common genetic variants contribute to autism in the Faroe Islands. **npj Genomic Medicine**, n. December, 2018.

LEK, M. ET AL. Analysis of protein-coding genetic variation in 60,706 humans. **Nature**, v. 536, n. 7616, p. 285–291, 2016.

LOOMES, R.; HULL, L.; MANDY, W. P. L. What is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. **Journal of the American Academy of Child & Adolescent Psychiatry**, 2017.

LORD, C. et al. The Autism Diagnostic Observation Schedule - Generic: A standard mesure of social and communication deficits associated with the spectrum of autism. **Journal of Autism and Developmental Disorders**, v. 30, n. 3, p. 205–223, 2000.

LORD, C. et al. Autism spectrum disorder. **Nature Reviews Disease Primers**, v. 6, n. 1, 2020.

LORD, C.; RUTTER, M.; LE COUTEUR, A. Autism Diagnostic Interview - Revised: A Revised Version of a Diagnostic Interview for Caregivers of Indivduals with Possible Pervasive Developmental Disorders. **Journal of Autism and Developmental Disorders**, v. 24, n. 5, p. 659–685, 1994.

MANDY, W.; LAI, M.-C. Annual Research Review: The role of the environment in the developmental psychopathology of autism spectrum condition. **Journal of Child Psychology and Psychiatry**, v. 57, n. 3, p. 271–292, 2016.

MANNION, A.; LEADER, G. Comorbidity in autism spectrum disorder: A literature review. **Research in Autism Spectrum Disorders**, v. 7, n. 12, p. 1595–1616, 2013.

MILES, J. H. et al. Essential versus complex autism: Definition of fundamental prognostic subtypes. **American Journal of Medical Genetics**, v. 135 A, n. 2, p. 171–180, 2005.

MOREAU, Y.; TRANCHEVENT, L. C. Computational tools for prioritizing candidate genes: Boosting disease gene discovery. **Nature Reviews Genetics**, v. 13, n. 8, p. 523–536, 2012.

MOTTRON, L. et al. Sex differences in brain plasticity: A new hypothesis for sex ratio bias in autism. **Molecular Autism**, v. 6, n. 1, p. 1–19, 2015.

MUKADDES, N. M.; HERGUNER, S. Autistic disorder and 22q11.2 duplication.

**World Journal of Biological Psychiatry**, v. 8, n. 2, p. 127–130, 2007.

NAZNEEN, N. et al. A Novel System for Supporting Autism Diagnosis Using Home Videos: Iterative Development and Evaluation of System Design. **JMIR mHealth and uHealth**, v. 3, n. 2, p. e68, 2015.

NEALE, B. et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. **Nature**, v. 485, n. 7397, p. 242–245, 2012.

NEUL, J. L. **Rett Syndrome and MECP2-Related Disorders**. [s.l.] Elsevier Inc., 2020.

O'ROAK, B. J. et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. **Nature**, v. 485, n. 7397, p. 246–250, 2012.

OZONOFF, S. et al. Recurrence risk for autism spectrum disorders: A baby siblings research consortium study. **Pediatrics**, v. 128, n. 3, 2011.

PALMER, N. et al. Association of sex with recurrence of autism spectrum disorder among siblings. **JAMA Pediatrics**, v. 171, n. 11, p. 1107–1112, 2017.

PAULA, C. S. et al. Brief report: Prevalence of pervasive developmental disorder in Brazil: A pilot study. **Journal of Autism and Developmental Disorders**, v. 41, n. 12, p. 1738–1742, 2011.

PINTO, D. et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. **American Journal of Human Genetics**, v. 94, n. 5, p. 677–694, 2014.

RAHBARI, R. et al. Timing, rates and spectra of human germline mutation. **Nature Genetics**, v. 48, n. 2, p. 126–133, 2016.

RAMASWAMI, G.; GESCHWIND, D. H. **Genetics of autism spectrum disorder**.

1. ed. [s.l.] Elsevier B.V., 2018. v. 147

REICHENBERG, A. et al. Advancing paternal age and autism. **Archives of General Psychiatry**, v. 63, n. 9, p. 1026–32, set. 2006.

RISCH, N. et al. Familial recurrence of autism spectrum disorder: Evaluating genetic and environmental contributions. **American Journal of Psychiatry**, v. 171, n. 11, p. 1206–1213, 2014.

ROBINSON, E. B. et al. Examining and interpreting the female protective effect against autistic behavior. **Proceedings of the National Academy of Sciences**, v. 110, n. 13, p. 5258–5262, 2013.

RUBEIS, S. DE et al. Synaptic, transcriptional, and chromatin genes disrupted in autism A. **Nature**, v. 515, n. 7526, p. 209–215, 2014.

RUZZO, E. K. et al. Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. **Cell**, v. 178, n. 4, p. 850- 866.e26, 2019.

SAMMS-VAUGHAN, M. et al. The diagnosis of autism and autism spectrum disorder in Low and Middle Income Countries: Experience from Jamaica. **Autism.**, v. 21, n. 5, p. 564–572, 2017.

SÁNCHEZ-SÁNCHEZ, S. et al. Rare RELN variants affect Reelin-DAB1 signal transduction in autism spectrum disorder. **Human Mutation**, v. 39, n. 10, p. 1372–1383, 2018.

SANDERS, S. J. et al. Multiple recurrent de novo copy number variations (CNVs), including duplications of the 7q11.23 Williams-Beuren syndrome region, are strongly associated with autism. **Neuron.**, v. 70, n. 5, p. 863–885, 2011.

SANDERS, S. J. et al. Insights into Autism Spectrum Disorder Genomic Architecture

and Biology from 71 Risk Loci. **Neuron**, v. 87, n. 6, p. 1215–1233, 2015.

SANDIN, S. et al. The Familial Risk of Autism. **The Journal of the American Medical Association**, v. 311, n. 7, p. 1770–1777, 2014.

SANDIN, S. et al. Autism risk associated with parental age and with increasing difference in age between the parents. **Molecular Psychiatry**, v. 21, n. 5, p. 693–700, 2016.

SASANI, T. A. et al. Large, three-generation human families reveal post-zygotic mosaicism and variability in germline mutation accumulation. **eLife**, v. 8, p. 1–24, 2019.

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

SCHAEFER, G. B. Clinical genetic aspects of ASD spectrum disorders. **International Journal of Molecular Sciences**, v. 17, n. 2, 2016.

SCHOPLER, E. et al. Childhood Autism Rating Scale. v. 1971, n. 1978, p. 1978–1980, 2010.

SCHWARZ, J. M. et al. Mutationtaster2: Mutation prediction for the deep-sequencing age. **Nature Methods**, v. 11, n. 4, p. 361–362, 2014.

SEBAT, J. et al. Strong Association of De Novo Copy Number Mutations with Autism. **Science**, v. 316, n. 5823, p. 445–449, 2007.

SHAPIRA, S. K. et al. A Novel Approach to Dysmorphology to Enhance the Phenotypic Classification of Autism Spectrum Disorder in the Study to Explore Early Development. **Journal of Autism and Developmental Disorders**, v. 49, n. 5, p.

2184–2202, 2019.

SRIVASTAVA, A. K.; SCHWARTZ, C. E. Intellectual disability and autism spectrum disorders : Causal genes and molecular mechanisms. **Neuroscience and Biobehavioral Reviews**, v. 46, p. 161–174, 2014.

SZTAINBERG, Y.; ZOGHBI, H. Y. Lessons learned from studying syndromic autism spectrum disorders. **Nature Neuroscience**, v. 19, n. 11, p. 1408–1418, 2016.

TAKATA, A. et al. Integrative Analyses of De Novo Mutations Provide Deeper Biological Insights into Autism Spectrum Article Integrative Analyses of De Novo Mutations Provide Deeper Biological Insights. **CellReports**, v. 22, n. 3, p. 734–747, 2018.

TALBOTT, M. R. et al. Brief Report : Preliminary Feasibility of the TEDI : A Novel Parent - Administered Telehealth Assessment for Autism Spectrum Disorder Symptoms in the First Year of Life. **Journal of Autism and Developmental Disorders**, n. 0123456789, 2019.

TOMA, C. Genetic Variation across Phenotypic Severity of Autism. **Trends in Genetics**, v. xx, n. xx, p. 1–4, 2020.

VELTMAN, M. W. M.; CRAIG, E. E.; BOLTON, P. F. Autism spectrum disorders in Prader-Willi and Angelman syndromes: A systematic review. **Psychiatric Genetics**, v. 15, n. 4, p. 243–254, 2005.

VIGNOLI, A. et al. Autism spectrum disorder in tuberous sclerosis complex: Searching for risk markers. **Orphanet Journal of Rare Diseases**, v. 10, n. 1, p. 1–9, 2015.

WEINER, D. J. et al. Polygenic transmission disequilibrium confirms that common

and rare variation act additively to create risk for autism spectrum disorders. **Nature Genetics**, v. 49, n. 7, p. 978–985, 2017.

WIGGINS, L. D.; BAIO, J.; RICE, C. Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample. **Journal of Developmental and Behavioral Pediatrics**, v. 27, n. 2 SUPPL. 2, 2006.

WILFERT, A. B. et al. Recent ultra-rare inherited mutations identify novel autism candidate risk genes. **bioRxiv**, p. 2020.02.10.932327, 2020.

WOODBURY-SMITH, M.; SCHERER, S. W. Progress in the genetics of autism spectrum disorder. **Developmental Medicine and Child Neurology**, v. 60, n. 5, p. 445–451, 2018.

ZWAIGENBAUM, L.; PENNER, M. Autism spectrum disorder: Advances in diagnosis and evaluation. **BMJ (Online)**, v. 361, p. 1–16, 2018.

# CHAPTER FIVE

## GENERAL CONSIDERATIONS

Fortunately, we accomplished our main goals by focusing our analyses into the identification of genetic alterations (DNVs and CNVs) and novel candidate loci, besides the investigation of DNVs roles across multigenerational families. Of note, this study was divided in three different ASD cohorts: one cohort of 30 ASD trios, composed by patients and parents (n = 90 individuals); a second cohort of 144 ASD individuals; and a third cohort of 33 multigenerational families composed by ASD probands, parents and grandparents (called septets, n=231 individuals).

The first trios-based study (Chapter II) allowed the analysis of CNVs and DNVs in the 30 ASD probands. Although we had a small cohort, we inferred power to our candidate genes findings by performing a meta-analysis in more than 20,000 individuals, from large genomic databases of neurodevelopmental disorders (NDDs). Happily, our analyses supported three pathogenic CNVs: 1q21.1, 4q35, and 17p13.3; and four pathogenic DNVs, being one of them identified in a novel ASD candidate gene: *PRPF8*. This cohort presented a good diagnostic yield, estimated in 23%, which may be partially explained by the high proportion of severe comorbidities in these patients (60%).

The CMA analysis (Chapter III) was performed in collaboration with the PhD student Claudia Costa. We screened 144 individuals and prioritized 46 rare CNVs, which were evaluated in larger CNVs cohorts of approximately 10,000 NDDs individuals. We highlighted four CNVs in ASD genes: *NR3C2*, *DPP6*, *CDH8/CDH11*, and *CSMD1*; and suggested two potential ASD candidate genes

identified in our CNVs: *FGF2* and *PTPRN2*. Finally, 18% (7/39) of individuals carrying a likely pathogenic or pathogenic CNV also carried a second rare CNV, reinforcing the oligogenic model relevance for ASD.

Through the septets analysis (Chapter IV) we were able to determine for the first time, DNVs rate in parents, which was lower than in probands and similar to controls. We observed a higher proportion of parents over 30 years old compared to grandparents, a positive correlation between DNVs and fathers age, and a predominantly paternal origin of DNVs in both generations. Intriguingly, besides the small cohort, we observed a segregation bias in the paternal transmission of pathogenic DNVs, which suggests the selection of these variants in paternal gametes. Finally, we evidenced *ZNF536, MSL2* and *HDAC9,* as possible candidate genes to ASD.

It is challenging to determine whether a genetic variant is a main cause of ASD, or a contributing factor to the clinical manifestation. In our analysis, even for ASD cases wherein DNVs CNVs were identified in ASD-genes, other rare variants were also identified and may contribute to phenotype, supporting the oligogenic model relevance. Our unexplored Brazilian population with tetra-hybrid composition was valuable to validate previous candidate genes as well as to identify novel ones. As we have shown, this was possible due the availability of genomic data from large cohorts, such as MSSNG, SSC, DDD and NDD big cohort's papers. Furthermore, the study delineation led us to support candidate loci, and to identify potential novel candidate genes to ASD: *PRPF8, FGF2, PTPRN2, ZNF536, MSL2,* and *HDAC9.*

# CHAPTER SIX

## ABSTRACT

Autism Spectrum Disorder (ASD) is a heterogeneous and complex genetic disorder that includes conditions with clinical manifestation in early childhood, characterized by persistent difficult in social/communication skills, besides restricted interests, and repetitive behavior. Large genomic studies inferred the relevance of single nucleotide variants (SNVs), small insertions and deletions (Indels), and copy number variants (CNVs) to ASD etiology. Additionally, ASD patients are enhanced for *de novo* mutations (DNMs) compared to control populations, for both SNVs/Indels and CNVs, being these alterations analyses powerful approaches to gene discovery and validation, together with analysis of large genomic databases of neurodevelopmental disorders (NDD). Therefore, our first main goal was to characterize rare copy number variations (CNVs) and exonic *de novo* variants (DNVs) in ASD Brazilian cohorts, the first composed by 30 trios (chapter II), and the second wherein only CNVs were analyzed, composed by 144 ASD individuals (chapter III). Also, aiming to elucidate these findings role to ASD etiology, we evaluated them in two large databases analysis of more than 20,000 and 10,000 NDD individuals, respectively. The first cohort analysis supported three pathogenic CNVs: 1q21.1, 4q35, and 17p13.3; and four pathogenic DNVS, being one of them identified in a novel ASD candidate gene: *PRPF8*. Regarding the second cohort of 144 individuals, we identified three deletions and one duplication in ASD candidate genes highly intolerant to loss of function: *NR3C2*, *DPP6, CDH8/CDH11,* and *CSMD1*. Besides CNVs affecting two potential ASD candidate genes: *FGF2* and *PTPRN2.* Finally, 18% of individuals carrying a likely pathogenic or pathogenic CNV, also carried a second rare CNV, reinforcing

the oligogenic model relevance for ASD. Our second main goal was to evaluate DNVs in 33 multigenerational families (231 individuals; chapter IV), to provide new insights into potential DNVs pathogenicity across generations. We determined for the first time, DNVs rate in parents, which was lower than probands and similar to controls. We observed a higher proportion of parents over 30 years old compared to grandparents, a positive correlation between DNVs and fathers age, and a predominantly paternal origin of DNVs in both generations. Intriguingly, besides the small cohort, we observed a segregation bias in the paternal transmission of pathogenic DNVs, which was not observed in the transmission of the maternal DNV, suggesting the selection of these variants in paternal gametes. Also, we evidenced *ZNF536*, *MSL2* and *HDAC9*, as possible candidate genes to ASD. Our unexplored Brazilian population with tetra-hybrid composition was valuable to validate previous candidate genes and to identify novel ones. As we have shown, this was possible due the availability of genomic data from large cohorts, such as MSSNG, SSC, DDD and NDD big cohort's papers. To our knowledge, this is the first study showing DNVs in parents, and despite our small sample, we were able to observe DNVs rates difference, as well as the segregation bias in the paternal transmission of pathogenic DNVs. Finally, the study delineation led us to support potential novel candidate genes to ASD: *PRPF8*, *FGF2*, *PTPRN2*, *ZNF536*, *MSL2* e *HDAC9*.

# RESUMO

O Transtorno do espectro autista (TEA) é uma condição genética complexa e heterogênea, que inclui manifestações clínicas durante a primeira infância, caracterizadas principalmente pela dificuldade persistente nas habilidades sociais e de comunicação, além de interesses restritos e comportamentos repetitivos. Grandes estudos genômicos demonstraram a importância das variantes de nucleotídeos único (do inglês, SNVs), das pequenas inserções e deleções (Indels), e das alterações no número de cópias (do inglês, CNVs), na etiologia do TEA. Além disso, os pacientes com TEA possuem aumento no número de mutações *de novo* (do inglês, DNMs), tanto para SNVs/Indels, quanto CNVs, sendo a análise dessas variantes poderosas abordagens para a identificação e validação de novos genes para TEA, juntamente com a análise de grandes bancos de dados genômicos de doenças de neurodesenvolvimento (do inglês, NDD). Assim, nosso primeiro objetivo foi caracterizar CNVs e variantes *de novo* (do inglês, DNVs) em coortes brasileiras de TEA, sendo a primeira composta por 30 trios (capítulo II), e a segunda composta por 144 probandos de TEA, onde apenas CNVs foram analisadas (capítulo III). Ademais, com o objetivo de elucidar o papel desses achados na etiologia do TEA, nós os avaliamos em duas grandes análises de bancos de dados genômicos, compostos por mais de 20.000 e 10.000 indivíduos, respectivamente. A análise da primeira coorte permitiu a identificação de três CNVs patogênicas: 1q21.1, 4q35, and 17p13.3; e quatro DNVs patogênicas, sendo uma delas em um novo gene candidato para o TEA: *PRPF8*. Em relação a segunda coorte de 144 indivíduos, nós identificamos três deleções e uma duplicação em genes associados ao TEA, altamente intolerantes a perda de função: *NR3C2*, *DPP6*, *CDH8/CDH11* e *CSMD1*. Além de CNVs afetando dois genes potenciais candidatos para TEA: *FGF2* e *PTPRN2*. Por fim, 18% dos

indivíduos com uma CNV patogênica ou provavelmente patogênica, também possuíam uma segunda CNV rara, reforçando a relevância do modelo oligogênico para o TEA. Nosso segundo objetivo maior foi avaliar as variantes *de novo* em 33 famílias multigeracionais (231 indivíduos; capítulo IV), para fornecer novas percepções sobre o potencial patogênico de variantes *de novo* entre gerações. Nós determinamos pela primeira vez a taxa de DNVs nos pais, que foi menor do que nos probandos, e similar a dos controles. Nós observamos uma maior proporção de pais acima dos 30 anos comparados aos avós, assim como uma correlação positiva entre a idade paterna e o número de DNVs, e uma origem predominantemente paterna de variantes DNVS nas duas gerações. Interessantemente, apesar da nossa amostra pequena, nós observamos um viés de transmissão de variantes *de novo* patogênicas nos pais, o que não foi observado nas mães, sugerindo que houve seleção dessas variantes nos gametas paternos. Nós também evidenciamos *ZNF536, MSL2* e *HDAC9,* como potenciais genes candidatos para TEA. Nossa população brasileira pouco explorada de composição tetra-hibrida, foi valiosa para a validação de genes já descritos para TEA, assim como para a identificação de novos genes. Conforme demonstramos, isso foi possível devido a disponibilidade de grandes coortes de bancos de dados genômicos, como o MSSNG, SSC, DDD e artigos com grandes coortes de indivíduos com NDD. Além disso, ao nosso conhecimento, esse foi o primeiro estudo demonstrando variantes *de novo* em pais de autistas, e apesar da nossa amostra pequena, nós pudemos observar diferenças entre as taxas de variantes de novo, assim como o viés de transmissão paterna de variantes *de novo* patogênicas. Por fim, a delineação deste estudo, nos permitiu identificar novos genes potenciais candidatos para TEA: *PRPF8, FGF2, PTPRN2, ZNF536, MSL2* e *HDAC9.*