VIVIANE NERI DE SOUZA REIS

Variações *de novo* e raras no genoma de pacientes com transtornos do espectro do autismo verbais e não verbais

Dissertação apresentada à Faculdade de Medicina
da Universidade de São Paulo para obtenção do
título de Mestre em Ciências

Programa de Psiquiatria

Orientadora: Prof. 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

2014

VIVIANE NERI DE SOUZA REIS

Variações *de novo* e raras no genoma de pacientes com transtornos do espectro do autismo verbais e não verbais

Dissertação apresentada à Faculdade de Medicina
da Universidade de São Paulo para obtenção do
título de Mestre em Ciências

Programa de Psiquiatria
Orientadora: Prof. 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

2014

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

Variações *de novo* e raras no genoma de pacientes com transtornos do espectro do autismo verbais e não verbais / Viviane Neri de Souza Reis. -- São Paulo, 2014.

Dissertação(mestrado)--Faculdade de Medicina da Universidade de São Paulo.
Programa de Psiquiatria.

Orientadora: Helena Paula Brentani.

Descritores: 1.Transtorno autístico 2.Exoma 3.Sequenciamento de nucleotídeos em larga escala 4.Variação genética 5.Variações do número de cópias do DNA 6.Linguagem 7.Psiquiatria

USP/FM/DBD-279/14

*Dedico esse trabalho a minha querida família e amigos, que
me ajudaram e apoiaram em todos os momentos*

AGRADECIMENTOS

À Dr^a Helena Brentani, minha orientadora, por todos os ensinamentos, apoio e inspiração. Espero poder retribuir sempre a confiança e empenho.

À minha mãe, Rinalda, por sempre me apoiar e estimular em todos os momentos da minha vida, nos estudos e decisões. E ao meu pai querido, Brivaldo, que sempre se orgulhou das minhas realizações.

À todos integrantes da minha família e amigos queridos, os que estão perto e os que estão longe, os pequeninhos e os grandões, por sempre me incentivarem e torcerem pelo meu sucesso.

À toda equipe da genética, do ambulatório, do laboratório e de bioinformática: Cecília, Andressa, Lígia, Ana Paula, Elaine, Joana, Renato, Alessandra, Débora, Luciana, Aderbal, Henrique, Daniel, Gisele, Thais, Carol e todos que diretamente e indiretamente contribuíram nesse projeto. E agradeço especialmente a Bianca e Ana Tahira, por todo conhecimento, análises e ajuda, sem os quais esse trabalho não seria possível.

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.

Ao Prof^o Dr^o Euripedes, pela orientação inicial, e todas as pessoas do CEAPESQ, LIM23 e da secretaria de pós-graduação, pelo apoio e auxílio na realização do projeto. Agradeço a Prof^a Iscia Lopes Cendes, a Prof^a Maria Rita Passos Bueno e ao Joao Paulo Kitajima, além de todos de suas respectivas equipes, pela colaboração em momentos cruciais do projeto.

E finalmente, agradeço as agências de apoio financeiro para a realização do projeto: Fundação de Amparo a Pesquisa do Estado de São Paulo (Bolsa: de Mestrado: 2011/04956-6, Auxílio Pesquisa: 2011/14658-2) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

SUMÁRIO

RESUMO

1.	INTRODUÇÃO	01
1.1.	Caracterização dos transtornos do espectro do autismo	01
1.2.	Genética em TEA	04
1.3.	Variações comuns	06
1.4.	Variações raras	07
1.5.	Estudos de variações no número de cópias	08
1.6.	Variantes raras na era do sequenciamento	11
1.7.	Estudos de exoma em TEA	12
1.8.	Estudos de genética de TEA e linguagem	15
1.9.	Modelo amostral e proposta de estudo	17
2.	OBJETIVOS	19
3.	MÉTODOS	20
3.1.	Seleção da Amostra	20
3.2.	Escalas e instrumentos aplicados aos sujeitos da pesquisa	21
3.2.1.	Instrumentos de triagem	21
3.2.2.	Avaliação neuropsicológica	22
3.2.3.	Avaliação fonaudiológica	23
3.3.	Coleta e armazenamento de material biológico	26
3.4.	Captura e sequenciamento do exoma	27
3.5.	Análise de bioinformática	33
3.5.1.	Controle de qualidade e mapeamento das sequencias geradas	33
3.5.2.	Teste de erro mendeliano	34
3.6.	Análise de CNV	35
3.7.	Análise de SNV e Indels	35
3.8.	Construção e análise das redes	37
3.9.	Análise de vias e processos biológicos hiper-representados	39

3.10.	Considerações éticas	40
4.	RESULTADOS A - TEA SINDRÔMICO	41
4.1.	Caracterização amostral	41
4.2.	Qualidade do sequenciamento	42
4.3.	Análise de CNV	46
4.4.	Análise de SNV e Indels	50
4.5.	Análise de vias e processos biológicos hiper-representados	52
5.	RESULTADOS B - TEA ESPORÁDICO	54
5.1.	Caracterização amostral	54
5.2.	Qualidade de sequenciamento	55
5.2.1.	Análise qualitativa dos dados	55
5.2.2.	Análise do conteúdo de CG	58
5.2.3.	Chamada de variantes (GATK)	59
5.2.4.	Taxa de Erros Mendelianos	59
5.3.	Análise de CNV	60
5.4.	Análise de SNV e Indels	63
5.5.	Análise de redes	65
5.6.	Análise de vias e processos biológicos hiper-representados	68
5.6.1.	Grupo 1 – Não verbais	69
5.6.2.	Grupo 2 – Verbais	71
6.	DISCUSSÃO	74
7.	CONCLUSÃO	86
8.	REALIZAÇÕES FUTURAS	86
9.	ANEXOS	87
10.	REFERÊNCIAS BIBLIOGRÁFICAS	96

RESUMO

Estudos de gêmeos e famílias demonstram que os transtornos do espectro do autismo (TEA) apresentam um grande componente genético (~50%), porém sua etiologia ainda é desconhecida, possivelmente devido aos TEA serem caracterizados como doenças complexas, poligênicas e multifatoriais. Recentemente, variações no número de cópias (CNVs, do inglês *Copy Number Variations*) e mutações pontuais (SNV, do inglês *Single Nucleotide Variant*) raras, *de novo* e herdadas foram associadas com TEA, sugerindo novos *loci* e genes candidatos. No entanto, a grande maioria das alterações descritas são individuais, de forma que análises por agrupamento das mesmas em genes, e busca de funções biológicas ou vias hiper-representadas tem sido uma abordagem para a compreensão dos possíveis mecanismos etiopatológicos dos TEA. Como os TEA são muito heterogêneos clinicamente o uso de endofenótipos específicos para agrupamento das alterações gênicas pode auxiliar a discriminação de vias e processos biológicos relacionados a dimensões fenotípicas. Considerando os estudos realizados em autismo, e a natureza das variações comuns e raras, nesse trabalho foi realizado o sequenciamento do exoma de 1 família de dois irmãos com TEA sindrômico (sequenciamento piloto) e 18 trios de casos esporádicos de TEA, em busca alterações muito raras e/ou de novo com provável impacto funcional nos pacientes; Além disso, foi analisado se existe diferença entre as vias biológicas hiper-representadas de redes gênicas crescidas a partir dos genes que apresentavam variações raras e *de novo*, comparando pacientes de TEA com: (1) pouca ou nenhuma comunicação, chamados de não verbais e (2) média a boa comunicação, chamados de verbais. No sequenciamento piloto da família dos irmãos com TEA sindrômico, encontramos 1 duplicação em 4p16.3 e 1 deleção em 8p23.3, em ambos os irmãos; alterações estas encontradas em estudos previos em pacientes com características sindrômicas e TEA; na análise de SNVs e Indels foi encontrada 1 variação *de novo* e 117 variações não-sinônimas raras herdadas de um dos pais na irmã e 150 variações não-sinônimas raras herdadas de um dos pais no irmão; a análise de vias revelou que os genes com as mutações pontuais raras estavam hiper-representados em regiões cromossômicas diferentes em cada irmão (no cromossomo 1 na paciente do sexo feminino e no cromossomo 16 no paciente do sexo masculino), o que pode estar relacionado às diferenças fenotípicas por eles apresentadas. No sequenciamento do exoma dos trios foram encontradas alterações *de novo* em 9 dos pacientes: 1 CNV *de novo* (deleção) de 1,5Mb na região 3q29, região previamente associada com síndrome e transtornos do desenvolvimento; e 8 genes alterados por mutações pontuais de novo, dos quais um dele é o GABBR2, que apresenta evidência de associação com TEA. A análise de vias e redes das variantes herdadas raras, mostrou que muitos dos genes relacionados aos dois grupos verbais e não verbais são genes já associados com TEA ou que apresentam interação com aqueles genes associados ao TEA. As análises de vias e redes precisam ser replicadas em amostras maiores, mas com nossos resultados preliminares podemos perceber que nosso estudo contribui com alterações em genes de vias relacionadas a neurogênese e sinaptogênese, independentemente do fenótipo, que possam refletir um conjunto de genes específicos e ou numero de alterações relacionadas a gravidade do TEA.

ABSTRACT

Studies of twins and families have shown that autism spectrum disorders (ASD) are highly heritable (~50%), but its etiology is still unknown, possibly because it is a very heterogeneous phenotype and have multiple genes involved in its development, what characterizes a complex disease such as ASD. Recently, copy number variations (CNVs) and point mutations (SNVs) rare, inherited e *de novo*, were associated with ASD, suggesting new candidate genes and loci. Because they are very rare, the vast majority of the changes described are individual, so the analysis of different variations grouped by genes and searching for biological functions or hyper represented pathways has been an approach for understanding possible pathogenic mechanisms of ASD. As ASD is clinically very heterogeneous, the use of endophenotypes, specific grouping of genomic changes can help discriminating pathways and biological processes related to phenotypic dimensions. Considering the studies in autism, and the nature of common and rare variants, we sequenced all exons (exome) of 1 family with syndromic ASD (pilot sequencing) and 18 trios of sporadic ASD cases to search for *de novo* and rare variations with probable functional impact on Brazilian patients; Also, we analyzed whether there is a difference in the enrichment of biological pathways of gene networks from the list of genes affected with *de novo* and rare deleterious variants in two groups of ASD patients: (1) cases with little or no communication, called nonverbal and (2) cases with average to good communication, called verbal. In the pilot exome sequencing (ASD syndromic family), we found a duplication in 4p16.3 and a deletion in 8p23.3 in both siblings, alterations that were found in patients with syndromes and ASD in previously studies; the analysis of SNVs showed 1 variation *de novo* and 117 nonsynonymous rare variations inherited from only 1 of the parents in the female sibling, and 150 nonsynonymous rare variations inherited from only 1 of the parents in the male sibling; Pathway analysis revealed enrichment differences of chromosomal regions for each sibling (chromosome 1 for the female patient and chromosome 16 for the male patient), what may be related to their phenotypic differences. In the exome sequencing of trios, as expected, it was found *de novo* variation in 9 of the patients: 1 *de novo* CNV (deletion) of 1.5 Mb in the region 29 of the long arm of chromosome 3, a region previously associated with syndrome and developmental disorders; and 8 genes altered by *de novo* variations, one of those is in the GABBR2, gene with previous evidence of association with ASD. The pathways and networks analysis of rare inherited variants showed that many of the genes related to the two groups verbal and nonverbal are already associated with ASD or interacts with those genes associated with ASD. This pathway and gene network analyses need to be replicated in larger samples, but our preliminary results shows that our study contributes with variations in genes related to neurogenesis and synaptogenesis pathways, regardless of phenotype, with probable impact to specific genes that may be related to severity of clinical presentation.

1. INTRODUÇÃO

1.1 Caracterização dos transtornos do espectro do autismo

Os transtornos do espectro do autismo (TEA) são transtornos do desenvolvimento caracterizados por prejuízos precoces na socialização e comunicação, bem como comportamentos e interesses restritos e estereotipados (DSM-V, 2013). Esses transtornos mostram uma grande heterogeneidade clínica e os déficits na interação social/comunicação são observados em múltiplos contextos, caracterizados no **Quadro 1**. Na maioria dos casos a etiologia do TEA é desconhecida e um fator complicador é o fato de não haver um sistema de classificação com base biológica, sendo seu diagnóstico feito a partir de observação comportamental de acordo com prejuízos nos domínios principais (Piggot *et al.*, 2009).

Quadro 1 – Caracterização dos critérios para diagnóstico de TEA (DSM-V)

A) Comunicação e interação social	
Déficit na reciprocidade sócio emocional	<ul style="list-style-type: none"> → Contato social anormal; → Falha na conversação com alternância anormal entre os interlocutores; → Compartilhamento reduzido dos interesses, emoções e afetos; → Falha para iniciar ou responder às interações sociais
Déficit na comunicação não verbal usada para interação social	<ul style="list-style-type: none"> → Integração pobre da comunicação verbal e não-verbal; → Contato visual ou linguagem corporal anormal ou déficits no entendimento e uso de gestos; → Completa ausência de expressão facial e comunicação não-verbal
Déficits no desenvolvimento, manutenção e entendimento das relações	<ul style="list-style-type: none"> → Dificuldade em ajustar o comportamento frente às variações sutis do contexto social; → Dificuldade em compartilhar jogos imaginativos ou fazer amigos; → Ausência de interesse nos pares
B) Comportamentos, interesses ou atividades, repetitivos e restritos (pelo menos 2)	
Movimentos, uso de objetos ou fala repetitivos ou estereotipados	<ul style="list-style-type: none"> → Estereotipia motora ; → Alinhamento dos brinquedos ou sacolejo/arremesso de objetos;

	→ Simples ecolalia; → Frases idiossincráticas
Insistência em uniformidade/monotonia, aderência inflexível às rotinas, ou padrões ritualizados de comportamento verbal e não-verbal	→ Estresse extremo frente a pequenas mudanças; → Dificuldade com transições; → Padrões rígidos de pensamento; → Rituais de cumprimento; → Necessidade de fazer o mesmo percurso ou de comer a mesma comida todo dia
Interesses fixos, altamente restritos que são anormais em intensidade ou foco	→ Forte apego ou preocupação excessiva com objetos incomuns; → Interesses excessivamente circunscritos ou perseverativos
Hiper ou hipo reatividade a estímulos sensoriais ou interesses incomuns a aspectos sensoriais do ambiente	→ Indiferença aparente a dor/temperatura; → Respostas adversas a sons ou texturas específicas; → Cheirar ou tocar objetos excessivamente; → Fascinação visual com luzes ou movimentos

A razão de diagnósticos em TEA é de 4 meninos para cada menina (Duchan, Patel, 2012) e existe uma variabilidade na idade em que as crianças podem apresentar as características essenciais para o diagnóstico. Em geral, crianças em idade pré-escolar apresentam uma falta acentuada de interesse pelos outros, falhas na empatia, expressão e comunicação ausentes ou severamente atrasadas, marcada resistência à mudança, interesses restritos e movimentos estereotipados. Na idade escolar, as habilidades sociais e de comunicação em geral aumentam; entretanto, resistência à mudança e comportamentos de autoestimulação, como a autolesão, podem tornar-se mais proeminentes neste período. Durante a adolescência, alguns indivíduos com autismo apresentam ganhos acentuados no desenvolvimento, enquanto outros podem piorar (acessos de raiva, autolesão ou agressão). Preditores de evolução clínica incluem a presença de linguagem comunicativa aos 5 anos de idade e habilidade cognitiva geral (QI). A detecção precoce e a prestação de serviços melhoram o prognóstico a longo prazo, o que torna o diagnóstico precoce particularmente importante (Volkmar *et al.*, 2014).

Comorbidades são muito comuns em crianças com TEA. Simonoff *et al.* (2008) relataram que cerca de 70% dessas crianças apresentam pelo menos uma comorbidade e 48% duas ou mais. Deficiência Intelectual (DI) é muito comum no

TEA, sendo observada em até 80% dos casos (Benvenuto et al. 2009). Sintomas como comprometimento no desenvolvimento da linguagem, estereotípias e automutilação aumentam com a severidade da DI (Wing, Gould, 1979). A epilepsia pode ocorrer em 7 a 46% dos casos de TEA (Lo-Castro, Curatolo, 2013) e, segundo meta-análise realizada por Amiet *et al.* (2008), é mais frequentemente associada aos pacientes com DI (21.4%) em relação aos que não apresentam DI (8%).

Outras comorbidades psiquiátricas incluem transtorno de déficit de atenção e hiperatividade (TDAH), encontrado em cerca de 30% dos casos de TEA (Simonoff *et al.*, 2008); depressão em 2%, podendo atingir 30% quando considerado apenas a Síndrome de Asperger (Leyfer *et al.*, 2006; Matson, Nebel-Schwalm 2007); transtornos ansiosos num intervalo de aproximadamente 2% a 45% (Leyfer *et al.*, 2006; Simonoff *et al.*, 2008); transtornos do sono em 40% a 86% dos casos (Maski *et al.*, 2011); tiques em 20% a 36% dos casos (Bodfish *et al.*, 2000); e transtorno obsessivo compulsivo (TOC) em 37% das crianças com autismo (Leyfer *et al.*, 2006).

Wegiel *et al.* (2010) relataram não só o crescimento cerebral anormal no início da infância em pacientes com TEA, mas também outras anomalias de desenvolvimento, sugerindo alterações de várias regiões, entre elas, a região cerebelar e do neocórtex. Nesse estudo, foi encontrado displasia cortical em apenas 1 controle (n=14), enquanto que 92% (n=13) das crianças dentro do espectro apresentaram anormalidades focais no desenvolvimento cerebral (hiperplasia, hipoplasia, heterotopia), o que reflete problemas na neurogênese, migração neuronal e maturação cerebral multirregional irregular no TEA.

O envolvimento de áreas anatômicas relacionadas ao cérebro social em estudos *post-mortem* e de ressonância magnética estrutural mostraram alterações nos lobos frontais, amígdala e cerebelo em autismo, porém, a replicabilidade de tais achados ainda não é consistente em pesquisas de neuroanatomia, o que sugere a necessidade de estudos em pacientes com fenótipos bem definidos, maior número de amostras e tecidos melhor caracterizados (Amaral *et al.*, 2008).

Exames funcionais de imagem cerebral, como tomografia e ressonância magnética funcional, abriram uma nova perspectiva para o estudo do funcionamento cerebral normal e patológico. Esses estudos demonstraram hipoativação da maior parte do sulco temporal superior (STS) do lobo temporal em pacientes autistas, região anatômica de grande importância para a percepção de estímulos sociais essenciais. Tais estímulos estão envolvidos na percepção social de faces, voz e cognição social (Zilbovicius, 2006). Estudos de neuroimagem demonstram o envolvimento do STS no TEA, assim como alterações de conectividade de longa distância (Travers *et al.*, 2012).

A mudança proposta no DSM-V reflete estudos de classes latentes assim como análises fatoriais (Lord, Jones, 2012; Volkmar *et al.*, 2014), evidenciando que mesmo diante da grande variabilidade fenotípica, da presença de diferentes comorbidades e de diferentes tipos de evolução e prognóstico, podemos considerar as alterações relacionadas ao cérebro social e, mais especificamente, as dimensões da comunicação como um dos pilares dos diferentes fenótipos alocados dentro do transtorno do espectro autista, sendo o segundo pilar os comportamentos e interesses restritos e estereotipados.

1.2 Genética do TEA

O envolvimento de fatores genéticos na etiologia do TEA é reconhecido a muito tempo. (Muhle *et al.*, 2004). Os estudos de gêmeos demonstraram uma concordância no diagnóstico de TEA em gêmeos monozigóticos (MZ) entre 60% e 95%, e em dizigóticos (DZ) de até 31%, dependendo do estudo (Ronald, Hoekstra, 2011). Um trabalho recente com coorte populacional realizado na Suécia mostrou taxas de herdabilidade de TEA de ~52,4% (Gaugler *et al.*, 2014).

Os primeiros estudos de genes candidatos em TEA foram embasados em estudos das síndromes genéticas onde os portadores apresentavam sintomas de TEA em uma frequência maior do que o esperado na população (Abrahams, Geschwind, 2010). Entre as síndromes mais associadas com TEA estão Síndrome do X Frágil (1–2% dos casos de TEA), esclerose tuberosa (1%),

Síndrome de Rett (0.5%) e neurofibromatose (<1%) (Devlin, Scherer, 2012). Muhle *et al.* (2004) já em 2004 argumentam que apesar de existir uma associação destas síndromes com TEA, elas não explicam a complexidade fenotípica da doença, o que indica a necessidade de se investigar outros mecanismos e genes associados.

As investigações de variações genéticas relacionadas à susceptibilidade ao TEA focadas em estudos de ligação e genes candidatos resultaram em grande número de cromossomos e *loci* genéticos associados com TEA, indicando que vários genes podem estar envolvidos com sua etiologia (Freitag *et al.*, 2010). Apesar de não haver alta concordância entre diferentes estudos, regiões dos cromossomos 2, 7, 11, 16, 22 e X foram replicadas em mais de um estudo (Piggot *et al.*, 2009, Holt *et al.*, 2010). Entre os numerosos estudos de ligação realizados com TEA, as seguintes regiões foram positivas em pelo menos dois estudos independentes: 2q21-33, 3q25-27, 4q32, 6q14-21, 7q22, 7q31-36, 11p12-13, e 17q11-21 (Freitag *et al.*, 2010). Estes resultados sugeriram um número grande de variações genômicas relacionadas a fala e comunicação (Alarcón *et al.*, 2008; Arking *et al.*, 2008; Bakkaloglu *et al.*, 2008; Bishop, 2009), assim como genes associados com neurodesenvolvimento e sinaptogênese (Lagemaat, Grant, 2010).

O sequenciamento do genoma humano (Lander *et al.*, 2001; Venter *et al.*, 2001), o mapeamento de variações de nucleotídeos de base única conhecidos do inglês como *single nucleotide polymorphism* (SNPs) (Hinds *et al.*, 2005) e a conclusão do projeto de desenvolvimento do mapa de haplótipos do genoma humano, o HapMap (The International Hap Map Consortium, 2005; Frazer *et al.*, 2007) trouxeram um significativo progresso para o conhecimento da variabilidade do genoma humano (Stankiewicz, Lupski, 2010) e as bases moleculares de muitas doenças (Beckmann *et al.*, 2007). Como consequência dessas inovações tecnológicas, ocorreu uma primeira onda de estudos genéticos que avaliaram milhares de pacientes e controles e encontraram a associação de centenas de milhares de SNPs com muitas doenças comuns (Manolio *et al.*, 2009; Glessner, Hakonarson, 2011).

1.3 Variações comuns

Variações comuns são aquelas encontradas na população onde a frequência do menor alelo (MAF) é maior que 1% ($MAF \geq 0,01$). Elas são consideradas muito comuns quando $MAF \geq 0,05$ e menos comuns quando $0,01 \leq MAF < 0,05$ (Cirulli, Goldstein, 2010). Estes tipos de variantes, possivelmente são mais antigas e foram sujeitas a forças seletivas ao longo do tempo (Schork *et al.*, 2009). Isso pode significar que sua influência no risco de doenças provavelmente tenha efeito funcional modesto, o que é compatível com sua maior frequência na população (Raychaudhuri, 2011).

Estudos de associação representando o genoma inteiro (GWAS, do inglês *Genome Wide Association Studies*) apareceram como uma técnica potencialmente poderosa para o avanço nos estudos genéticos, uma vez que permitem a análise de milhares de variações de nucleotídeo único (SNPs), em um número grande de pacientes sem a necessidade de uma hipótese prévia como acontecia nos estudos de genes candidatos (Cichon *et al.*, 2009; Glessner, Hakonarson, 2011; Patnala *et al.*, 2013).

Dos GWAS realizados até hoje em TEA, variantes comuns foram identificados nas seguintes regiões: i) 5p14.1 entre os genes que codificam as caderinas 9 (*CDH9*) e 10 (*CDH10*), moléculas de adesão neuronal (Ma *et al.*, 2009; Wang *et al.*, 2009); ii) 5p15 próxima ao gene da semaforina 5A (*SEMA5A*), gene implicado no crescimento axonal e com expressão reduzida em células linfoblastóides em casos de autismo quando comparados com controles saudáveis (Weiss *et al.*, 2009); e iii) 20p12.1, intron do gene *MACROD2*, cuja proteína contém um domínio MACRO de alta afinidade para ligação ribose-ADP e é importante em diversos processos biológicos (Anney *et al.*, 2010). Segundo Devlin *et al.* (2011), ao combinarem os dados desses GWAS, a evidência de associação diminuiu para todos os alelos de risco identificados, que se tornariam estatisticamente não significantes.

Em estudo em busca de endofenótipos, Ronald *et al.* (2010), investigou a presença de QTL (lócus de traço quantitativo, do inglês *quantitative trait locus*) em indivíduos da população geral com traços “autísticos” relacionados a sociabilidade usando GWAS. Um SNP (rs11894053) localizado na região 21 do braço curto do cromossomo 2 (2p21) foi associado aos traços da sociabilidade ($P < 0.05$, não corrigido para testes múltiplos).

Ao longo do tempo ficou claro que embora o HapMap seja um instrumento poderoso para estudos genéticos de associação, o conjunto de SNPs usados como marcadores moleculares não cobre totalmente a variabilidade do genoma (Cichon *et al.*, 2009). Segundo Manolio *et al.* (2009), variações comuns analisadas pelos GWAS conferem apenas um aumento relativamente pequeno do risco de desenvolver um transtorno e explicam apenas uma pequena porção da herdabilidade. Em doenças complexas essas variantes comuns em geral tem efeito pequeno e parece ser necessário mais de 10.000 amostras para detecção das mesmas (Kiezun *et al.*, 2012). No entanto, vale lembrar que dada a complexidade fenotípica dos transtornos psiquiátricos, quanto maior a amostra maior o risco de diluição do fenótipo.

1.4 Variações raras

Por mais que os estudos de GWAS tenham contribuído para um maior entendimento sobre a complexidade da arquitetura genética e etiologia molecular do TEA, o pequeno efeito associado às variações comuns, a baixa replicabilidade de achados e a necessidade de amostras enormes com consequente diluição do fenótipo levou a uma busca por outros métodos que contribuíssem para a elucidação do componente genético ainda desconhecido desse transtorno, como a retomada dos estudos de variantes raras (Manolio *et al.*, 2009; Gibson, 2011; State, Levitt, 2011).

As variações raras são aquelas encontradas na população em uma frequência menor; são consideradas raras quando $MAF < 0,01$ (Cirulli, Goldstein, 2010) e muito raras quando $MAF \leq 0,005$ (Li *et al.*, 2011).

Alelos raros parecem ser mais recentes e podem ainda não terem sido submetidos à mesma pressão seletiva negativa que variações comuns, incluindo mutações relativamente mais deletérias. Portanto, é provável que variantes raras associados a doenças complexas sejam mais facilmente mapeadas e avaliadas quanto ao impacto funcional (Raychaudhuri, 2011). A proporção de variantes não sinônimas é maior entre os alelos raros do que entre os alelos comuns e, ainda entre os alelos raros, variantes missense (que causam alterações de aminoácidos) preditas tendem a ser mais deletérias (Kiezun *et al.*, 2012).

Segundo resultados obtidos por Tennesen *et al.* (2012), a grande maioria das variações em regiões codificadoras de proteínas é evolutivamente recente, rara e enriquecida com alelos deletérios, e possivelmente contribui de forma importante para a variação fenotípica e suscetibilidade às doenças.

1.5 Estudos de variações no número de cópias

Os primeiros estudos de variações raras no genoma inteiro em psiquiatria foram os de variações estruturais do número de cópias (CNVs), mecanismos de rearranjos cromossômicos no genoma como a recombinação homóloga não alélica (*Non Allelic Homologous Recombination* – NAHR) e a junção de extremidades não homólogas (*Non-Homologous End Joining* – NHEJ) poderiam explicar a formação de CNVs (Stankiewicz, Lupski, 2010).

Apesar da maioria das CNVs serem raras na população quando comparadas com os SNPs, em conjunto elas são uma fração significativa das variações genômicas do genoma humano (Alkan *et al.*, 2011). Essas variações cobrem de 12% a 15% do genoma humano (Redon *et al.*, 2006) e acredita-se que contribuem significativamente para variabilidade, uma vez que genes com CNVs codificam produtos que de alguma forma parecem influenciar a interação com o

ambiente como sensores ambientais: ex. receptores olfativos e de resposta inflamatória, sinalização celular de regiões genômicas correspondentes a RNAs não codificadores que regulam expressão gênica pós-transcricional em processos fisiológicos ao longo da vida e durante o neurodesenvolvimento (Rodriguez-Revenega *et al.*, 2007; Ionita-Laza *et al.*, 2009).

As CNVs podem ser herdadas ou *de novo* (esporádicas, não herdadas), podem conter partes de genes, um ou mais genes e regiões regulatórias do genoma, ou mesmo não conter qualquer elemento conhecido, e sua ocorrência pode interromper regiões funcionais do genoma e gerar novas combinações de sequências, o que está associado com a variabilidade fenotípica, como também, a ocorrência de doenças (Conrad *et al.*, 2010; Lee, Scherer, 2010).

O foco nas CNVs raras é em parte baseado nos resultados prévios de estudos citogenéticos, e várias regiões cromossômicas que apresentam CNVs são associadas com síndromes de microdeleção como, por exemplo, a síndrome da deleção da região cromossômica 16p11.2, síndrome de Williams (7q11.23); síndrome de Prader-Willi/Angelman (15q11-13) e síndrome VCF Di George (22q11.2) (Malhotra, Sebat, 2012). Os primeiros transtornos neuropsiquiátricos nos quais as CNVs parecem exercer papel no risco para o desenvolvimento da doença, foram os estudos de esquizofrenia (Roos, 2011) e autismo (Sebat *et al.*, 2007; Marshall *et al.*, 2008; Glessner *et al.*, 2009; Pinto *et al.*, 2010; Gilman *et al.*, 2011; Nord *et al.*, 2011; Sanders *et al.*, 2011).

O primeiro estudo de CNVs em autismo foi realizado por Sebat *et al.* (2007), onde CNVs *de novo* foram associadas com autismo ($P = 0,0005$), presentes em 12 (10%) de 118 dos pacientes com autismo esporádico (único probando afetado), em 2 (3%) de 77 dos pacientes com parente de primeiro grau afetado, e apenas 2 (1%) de 196 dos controles. Christian *et al.* (2008), identificaram 51 CNVs exclusivas em pacientes com TEA, sendo 3 duplicações no cromossomo 15q11-13. Kumar *et al.* (2008) e Marshall *et al.* (2008) relataram associação de CNVs *de novo* de aproximadamente 600Kb no cromossomo 16p11.2 com TEA, estes CNVs apresentaram uma frequência de aproximadamente 1%, nos indivíduos com TEA.

Glessner *et al.* (2009) estudaram CNVs em 859 casos e 1409 controles (ascendência europeia), usando plataforma Illumina HumanHap550 BeadChip de aproximadamente 550.000 marcadores (SNPs) na tentativa de identificar CNVs que conferissem susceptibilidade ao TEA. Os resultados positivos foram avaliados em estudo independente com 1336 casos e 1110 controles. Além dos genes candidatos associados ao TEA, *NRXN1* e *CNTN4*, foram identificados novos genes envolvidos na codificação de moléculas de adesão de células neuronais, incluindo *NLGN1* e *ASTN2*. Além disso, somente em pacientes, foram identificadas CNVs dentro ou entre genes envolvidos na via da ubiquitina, incluindo *UBE3A* (gene E3A da ubiquitina ligase, frequentemente estudado em TEA). As ubiquitinas são responsáveis por modificações pós-transcricionais, que alteram funções proteicas, e operam em compartimentos pré e pós-sinápticos, regulando, entre outras funções, a liberação de neurotransmissores, reciclagem de vesículas sinápticas e modificações das espinhas dendríticas (Glessner *et al.*, 2009).

Pinto *et al.* (2010) caracterizaram CNVs raras presentes em TEA, ao estudar 996 indivíduos com TEA (ascendência europeia), pareados com 1287 controles. Nos pacientes foi observado um maior número global de CNVs (1,19 vezes, $P=0,012$) especialmente nos loci previamente associados a TEA e/ou deficiência intelectual (1,69 vezes, $P=3,43 \times 10^{-4}$). Dentre as diversas CNVs encontradas, os genes de *SHANK2*, *SYNGAP1* e *DLGAP2* (proteínas presentes na formação sináptica) aparecem como novos loci associados ao TEA. Além disso, através do estudo do mapa funcional de TEA (mapeamento de enriquecimento de CNVs em grupos de genes que compartilham uma função comum ou operam numa mesma via), descobriram enriquecimento de CNVs em grupos de genes envolvidos nas vias de desenvolvimento celular e neuronal, incluindo proliferação, motilidade e projeção celular.

Nord *et al.* (2011) estudaram crianças com autismo ($N=41$) e controles saudáveis ($N=367$) em busca de CNVs raras, definidas como tendo menos de 30% de sobreposição com CNVs identificadas no grupo controle ou no banco de variações genômicas DGV (*Database of Genomic Variants*). O estudo mostrou

que os indivíduos com autismo são mais afetados por CNVs raras com aproximadamente 10 Kb ou maiores, e que os casos apresentavam CNVs afetando desproporcionalmente genes envolvidos em transcrição, neurodesenvolvimento e atividade de receptores.

Outro estudo também reporta que vias funcionais comuns, associadas com função sináptica e neurotransmissão, aparecem enriquecidas com CNVs em pacientes com TEA em relação aos controles (Gai *et al.*, 2012).

CNVs têm sido associadas cada vez mais a uma variedade de distúrbios neuropsiquiátricos além do autismo e esquizofrenia, como deficiência intelectual, transtorno de déficit de atenção e hiperatividade (TDAH), transtorno bipolar, epilepsia entre outros (Guilmatre *et al.*, 2009; Malhotra, Sebat, 2012). Interessante notar que diferentes transtornos compartilham muitos dos mesmos CNVs. Kearney (2012) sugeriu que os efeitos pleiotrópicos das CNVs sugerem que as mesmas possam estar envolvidas nos processos fundamentais do neurodesenvolvimento, e que a manifestação clínica de cada fenótipo dependeria de outros fatores de risco genéticos e ambientais.

1.6 Variantes raras na era do sequenciamento

As técnicas de sequenciamento tornaram possível analisar não somente as variações comuns, como as raras, muito raras e aquelas nunca antes descritas (*novel*) (Cirulli, Goldstein, 2010; Goldstein *et al.*, 2013).

O sequenciamento de exoma é uma alternativa para o estudo de variações em regiões codificantes do genoma completo, e emergem nesse contexto como importante ferramenta para geração de dados novos e mais completos, para assim melhor mapear e caracterizar regiões genômicas, distribuições alélicas e contribuir na associação das variações de nucleotídeos únicos (do inglês *single nucleotide variation*, SNVs) e CNVs com um fenótipo particular (Bansal *et al.*, 2010; Kiezun *et al.*, 2012; Rehm, 2013).

O exoma permite o estudo de todos os éxons do genoma humano (~ 1-2 %), ou seja, as regiões gênicas codificadoras de proteínas (Tennessen *et al.*, 2012). O forte interesse no sequenciamento do exoma advém de 3 fatores: (i) a potencial identificação de muitos genes relacionados a transtornos complexos, (ii) a anotação funcional simples de variações na região codificadora e (iii) um custo substancialmente menor (aproximadamente 5x menor) que o do sequenciamento de genoma completo (Kiezun *et al.*, 2012). Ainda de acordo com Kiezun *et al.* (2012), é esperado que quase todos os genes apresentem variantes funcionalmente importantes que podem ser analisadas por meio de sequenciamento, mesmo sendo raras ou muito raras.

1.7 Estudos de exoma em TEA

Em autismo, O’Roak *et al.* (2011) realizaram o primeiro estudo seguindo essa linha ao sequenciarem o exoma de 20 trios de casos esporádicos de TEA, e encontraram 21 mutações *de novo*, das quais 11 alteravam proteínas, demonstrando o poder do sequenciamento do exoma de trios para identificar novos genes candidatos e melhorar o entendimento do papel das mutações *de novo* na etiologia do TEA.

Sanders *et al.* (2012), mostraram que SNVs *de novo* e não-sinônimas, são significativamente mais comuns em afetados (14% dos SNVs são preditas como associadas ao risco para TEA) em relação a irmãos saudáveis, e que essa diferença aumenta quando considerados SNVs presentes em genes expressos no cérebro (41%). Neste trabalho o grupo identificou 279 mutações *de novo* e refere que SNVs não-sinônimas podem apresentar *odds ratios* similares a aqueles previamente associados a CNVs *de novo* grandes (>500Kb).

Em estudo com maior número de trios (n=189), O’Roak *et al.* (2012), confirmaram seus resultados anteriores e observaram também a tendência de um aumento da taxa de variações não-sinônimas nos pacientes. Iossifov *et al.* (2012), em estudo com 343 quads (paciente, pais e irmão não afetado), não encontraram

diferenças nas taxas quando analisadas todas as SNVs encontradas, somente as sinônimas, ou somente as missense; porém encontram diferenças ao analisar número de SNVs nonsense e aquelas que alteravam sítios de splicing. Neale *et al.* (2012) estudaram 175 trios e seus resultados mostraram que 46,3% dos casos carregam uma variante missense ou nonsense *de novo*. Estes achados trazem evidências concordantes com o modelo poligênico para o aumento do risco do TEA e a importância de variantes raras *de novo* associadas a genes envolvidos com plasticidade sináptica (Neale *et al.*, 2012; O’Roak *et al.*, 2012; Sanders *et al.*, 2012).

Vale notar que mutações potencialmente patogênicas também podem ser herdadas e não apenas *de novo*. Um estudo de exoma em 16 famílias com compartilhando uma ancestralidade distante, revelou mutações recessivas potencialmente patogênicas em 4 genes. Os genes encontrados (*UBE3B*, *CLTCL1*, *NCKAP5L*, *ZNF18*) codificam para proteínas envolvidas em proteólise, sinalização mediada por GTP, organização do citoesqueleto e despolarização neuronal regulada por transcrição, o que confirma o papel de defeitos de transmissão e plasticidade neuronal na patogênese de TEA (Chahrou *et al.*, 2012).

Uma análise de rede de proteínas, com dados do sequenciamento de exoma, em uma coorte de famílias australianas com TEA, mostrou que variantes herdadas de pais com fenótipos associados a TEA e variantes *de novo* apresentam uma associação significativa com esses transtornos. Além disso, a análise de ontologia gênica revelou que variantes raras, possivelmente causais, se agrupam em processos neurobiológicos chaves, hiper-representados em funções envolvendo desenvolvimento neuronal, transdução de sinal e desenvolvimento de sinapse (An *et al.*, 2014).

Assim, modelos conceituais únicos como o de “variante comum, doença comum” (CVCD do inglês: *common variant common disease*), que segue a hipótese de que as variações genômicas associadas com doenças complexas são as variações comuns, ou o modelo de “variante rara, doença comum” (RVCD do inglês como: *rare variant common disease*), que presume que as variações raras

no genoma, cada qual com penetrância relativamente alta, é que estão associadas com a susceptibilidade das doenças comuns complexas (Schork *et al.*, 2009), isolados, não explicam todos os achados do TEA. Os resultados gerados suportam um modelo complexo da arquitetura genética do TEA, sendo o mesmo explicado pela junção dos dois modelos anteriores (Gibson, 2011; Sullivan *et al.*, 2012).

Essa heterogeneidade genética é associada com uma heterogeneidade fenotípica. Em revisão de 2011, Eapen propôs um modelo de patogenia do TEA, resumido no gráfico adaptado abaixo (**Figura 1**), onde uma combinação de alelos de risco resultaria nos traços patológicos. Os indivíduos apresentam algumas das variações comuns de pequeno efeito mas quando ultrapassam certo limite resultam nos fenótipos mais amplos de TEA com uma sobreposição com outros transtornos de neurodesenvolvimento e transtornos psiquiátricos (T1). Aqueles que atingem um limiar maior (T2), resultam nos casos mais graves. O TEA Síndrômico, ou seja, associado a síndromes genéticas em geral causadas por 1 gene de grande efeito e o TEA esporádico causado por mutações *de novo*, ou mutações muito raras herdadas de maior efeito, representam as extremidades da curva.

Figura 1 – Gráfico adaptado que resume o modelo de patogenia do TEA proposto por Eapen (2011)

1.8 Estudos de genética de TEA e linguagem

O prejuízo de linguagem era considerado uma característica essencial do autismo no DSM-IV, e o atraso de linguagem, definido como uma demora na fala relatado em aproximadamente metade das crianças com autismo (Alarcón *et al.*, 2008). As deficiências de linguagem apresentam-se heterogêneas dentro do espectro, podendo ser leves ou extremas, com muitos indivíduos permanecendo não verbais a vida toda (Herbert, Kenet, 2007; Kasari *et al.*, 2008). Uma vez que genes não codificam uma psicopatologia e, portanto, que a associação do efeito de um ou mais genes pode ser melhor identificada em um nível mais simples (baseado em fenótipos biológicos) e, levando-se em conta a importância da comunicação no TEA, podemos entender estudos genéticos realizados usando a capacidade do uso da linguagem como endofenótipo. Endofenótipos são subcomponentes do fenótipo mensuráveis e herdados envolvidos no processo entre uma variação genética até a expressão final da doença e podem ser neurofisiológicos, bioquímicos, endocrinológicos, neuroanatômicos, cognitivos, ou neuropsicológicos (Gottesman, Gould, 2003).

Seguindo esse contexto, Liu *et al.* (2008), selecionaram famílias com indivíduos com TEA e dividiram os mesmos em 6 subfenótipos baseados em: QI estimado, escore de interação social recíproca (SOC); escore de comportamentos repetitivos, restritos e estereotipados (BEH); idade da primeira palavra; idade da primeira frase; e condição verbal ou não verbal. Apesar de não encontrar associação significativa, foi encontrada uma tendência de associação nas regiões 15p13.3-q14 e 11p15.4-p15.3 com os grupos "QI > 70" e "atraso na fala da primeira frase" respectivamente. Esses resultados também levaram seus autores a argumentar a importância da seleção de subfenótipos bem caracterizados de TEA a fim de detectar *loci* de susceptibilidade.

Um estudo com famílias de pacientes com TEA, demonstrou que pacientes com deleção em *SHANK3* apresentavam entre outros sintomas retardo ou ausência de linguagem (Durand *et al.*, 2007). Em outros estudos, CNVs nesse

gene foram identificadas em indivíduos afetados não-verbais, com outros transtornos de linguagem e deficiências sociais (Gauthier *et al.*, 2009; Marshall *et al.*, 2008; Moessner *et al.*, 2007). A proteína de *scaffolding* SHANK3 é abundante na região pós-sináptica, onde interage com neuroglinas direta ou indiretamente, com múltiplas interações proteicas e está na interface de complexos formados por receptores de glutamato e proteínas regulatórias de actina, tendo papel ativo na morfogênese de espinhas dendríticas e na plasticidade sináptica. Quando expressa em neurônios do hipocampo, promove a maturação e alargamento de espinhas dendríticas (Betancur *et al.*, 2009).

Em estudo de GWAS realizado por Wang *et al.* (2009), um alelo de risco do SNP rs4307059 localizado no 5p14.1 entre os genes de *CDH9* e *CDH10*, foi associado com linguagem estereotipada e diminuição da comunicação pragmática. *CDH9* e *CDH10* são caderinas, grupo de proteínas que estão envolvidas nas junções célula-célula cálcio-dependentes presentes no sistema nervoso, responsáveis pela estrutura e conectividade neural do cérebro (Betancur *et al.*, 2009).

Alterações comuns e raras no gene de *CNTNAP2* aparecem diversas vezes associadas à pacientes com TEA (Alarcón *et al.*, 2008; Arking *et al.*, 2008; Wang *et al.*, 2009; Pinto *et al.*, 2010; Whitehouse *et al.*, 2011). *CNTNAP2* aparece expressa em cérebro fetal humano nos circuitos fronto-temporal e subcortical, que são críticos para a função executiva, e especula-se que sejam importantes na atenção compartilhada, precursor da comunicação verbal (Alarcón *et al.*, 2008). Segundo Betancur *et al.* (2009), *CNTNAP2* está presente na membrana plasmática de neurônios na região das sinapses e localizada em junções célula-célula de nervos mielinizados. Ela se liga diretamente e é regulada pela proteína do fator de transcrição FOX (*FOXP2*), relacionado à fala e desenvolvimento de linguagem.

St. Pourcain *et al.* (2010) realizaram um estudo com 7313 crianças com autismo e verificaram que o SNP rs4307059 da região 5p14.1 é um QTL relacionado a linguagem e comunicação social. As crianças que apresentavam

uma frequência maior do alelo mutado tinham uma diminuição nos escores de comunicação.

Um recente estudo de meta-análise com dados de diversos GWAS mostrou que SNPs específicos estão associados com um leque de transtornos psiquiátricos de início precoce e tardio (Smoller *et al.*, 2013). Segundo Smoller *et al.* (2013), esses resultados são evidências da importância de superar o modelo de diagnóstico categorial baseado em descrições de sintomas para uma nosologia dimensional e informativa da causa biológica da doença, sendo assim este estudo abordara a busca de alterações muito raras e *de novo* em trios de TEA, comparando dimensões da comunicação.

1.9 Modelo amostral e proposta de estudo

Sanders *et al.* (2012), propõe um modelo, no qual são necessários 4 ou mais diferentes alterações de base única em um mesmo gene, em diferentes casos, para associar o gene ao risco da doença. Como a probabilidade de encontrarmos a mesma alteração em dois probandos diferentes é muito pequena, a ideia de estudar um grande número de casos no intuito de se encontrar a mesma variação em todos eles, parece praticamente impossível, uma vez que o número amostral para tanto se torna pouco viável (Cirulli, Goldstein, 2010; Kiezun *et al.*, 2012). Ainda,, à medida que o tamanho da amostra aumenta e, por consequência, uma maior quantidade de dados é assimilada para análise combinada, o risco de associações espúrias devido à subestruturação populacional aumenta (Devlin *et al.*, 2001; Devlin, Roeder, 1999). Devido a forças demográficas, biológicas e aleatórias, variantes genéticas diferem em frequência alélica em populações em todo o mundo, criando a estrutura da população ou estratificação refletida pela ancestralidade (Hoggart *et al.*, 2003). Essa estratificação ainda é mais evidente em populações com história de miscigenação recente, como a população brasileira. Esta população é geneticamente formada por três distintas populações parentais em diversos graus de mistura (Europeu, Africano, e nativo americano) (Sans, 2000; Salzano, 2004), evidenciada por uma diversidade de marcadores genéticos: Y-DNA, mtDNA e marcadores

autossômicos (Abe-Sandes *et al.*, 2004; Callegari-Jacques *et al.*, 2003; Carvalho-Silva *et al.*, 2001; Ferreira *et al.*, 2006; Guerreiro-Junior *et al.*, 2009; Lins *et al.*, 2010, 2011; Marrero *et al.*, 2005; Parra *et al.*, 2003).

Portanto, estudos caso-controle são muito suscetíveis a associações espúrias entre as variantes genéticas e o status de doença, principalmente em populações híbridas (Lander *et al.*, 1994). E ainda, se estes estudos não forem corretamente delineados, mesmo correlações sutis, entre indivíduos de ascendência comum, afetam a distribuição dos testes de associação (Devlin *et al.*, 2001; Devlin, Roeder, 1999).

Em contrapartida, projetos baseados em estudos de famílias são mais robustos em relação à estratificação populacional. A grande vantagem é a existência de um background genético comum entre os membros da família. Além disso, as famílias tendem a ser mais homogêneas em relação à exposição a fatores ambientais possivelmente associados à etiologia da doença (Evangelou *et al.*, 2006). Nestes estudos, como os de trios, os pais são genotipados junto com a prole investigada, a análise envolve avaliação do “caso” (a descendência afetada) comparada com “pseudo-controles” (pais) e é inferida com base nos alelos transmitidos e não transmitidos pelos pais (Frayling *et al.*, 1999; Crossett *et al.*, 2010).

O estudo do exoma em trios permite, mesmo com um número amostral modesto, que seja possível encontrar variações de interesse, principalmente considerando a taxa de aproximadamente pelo menos uma alteração não sinônima e *de novo* por caso (O’Roak *et al.*, 2011, 2012; Iossifov *et al.*, 2012; Neale *et al.*, 2012; Sanders *et al.*, 2012). As variações observadas podem ser mapeadas e analisadas em redes gênicas, e levar ao descobrimento de novos *loci* de interesse, como observado por O’Roak *et al.* (2011, 2012), em estudo piloto realizado e corroborado em estudo subsequente do mesmo grupo e nos demais estudos de exoma em TEA detalhados anteriormente.

A formação de grupos de fenótipo bem caracterizados e extremos também se mostra uma estratégia para potencializar resultados em estudos exploratórios com número amostral pequeno, que podem ser seguidos de estudos maiores de

genotipagem das regiões variantes enriquecidas (Cirulli, Goldstein, 2010). Considerando os estudos realizados em autismo, a natureza das variações comuns e raras, o presente trabalho procura buscar CNVs e SNVs no exoma de trios de TEA, para avaliar as possíveis vias biológicas implicadas em diferentes dimensões da comunicação social.

2. OBJETIVOS

- Buscar alterações muito raras e/ou *de novo* com provável impacto funcional em pacientes com TEA usando trios da população brasileira.
- Verificar se existe diferença entre as vias biológicas hiper-representadas em redes gênicas crescidas a partir de genes que apresentam variações muito raras e/ou *de novo* comparando pacientes de TEA com : (1) pouca ou nenhuma comunicação, chamados de não verbais e (2) casos com média a boa comunicação, chamados de verbais.

3. MÉTODOS

3.1 Seleção da amostra

Os pacientes são provenientes do **Projeto Transtornos do espectro do autismo (PROTEA) do IPQ HC – USP**. O PROTEA contempla assistência, ensino e pesquisa. Este ambulatório funciona há 13 anos fazendo parte do Serviço de Psiquiatria da Infância e da Adolescência (SEPIA). Atualmente contam 445 casos com TEA no ambulatório, com idades de 3 a 40 anos.

Este estudo foi realizado em dois módulos:

(1) o sequenciamento piloto de um quad, de dois irmãos afetados, sendo um verbal e o outro não verbal visando o aprendizado da técnica.

(2) o sequenciamento de 18 trios de TEA esporádico divididos em verbais e não verbais visando, maximizar as chances de encontrarmos alterações *de novo* e portanto de maior impacto, justificando uma análise redes e de vias em número amostral tão pequeno.

Para o sequenciamento piloto foi escolhida uma família com dois irmãos afetados com TEA sindrômico sem causa genética conhecida (resultados dos exames citogenéticos e X-Frágil negativos). Para o projeto principal, foram selecionados 18 trios, ou seja, pacientes e seus pais biológicos. Critérios de inclusão para o probando dos trios: ter diagnóstico de TEA, confirmado por entrevista com psiquiatra segundo critérios do DSM-V; e ser considerado um caso esporádico, ou seja, ser caso único de TEA na família. Critério de exclusão: os pacientes não podem ter síndrome genética conhecida (e/ou bem caracterizada molecularmente) diagnosticada em prontuário.

Para coleta de dados fenotípicos foram utilizados os diversos instrumentos e escalas para uma melhor caracterização do diagnóstico, determinação do QI, avaliação fonoaudiológica e histórico familiar, realizados na rotina do ambulatório

por psiquiatras da infância e adolescência, psicólogos, terapeuta ocupacional, fonoaudiólogos e biólogos, que fazem parte do PROTEA.

3.2 Escalas e instrumentos aplicados aos sujeitos da pesquisa

Realizados na rotina do ambulatório por profissionais qualificados que fazem parte do ambulatório, são utilizadas para triagem (QAA e M-CHAT), confirmação diagnóstica (CARS) e para a melhor caracterização fenotípica avaliação Neuropsicológica, fonoaudiológica e de histórico familiar.

3.2.1 Instrumentos de triagem

Questionário de Avaliação de Autismo (QAA): É um questionário de triagem traduzido do Autism Screening Questionnaire (ASQ), composto por 40 itens de resposta sim ou não, que abrangem os domínios constantes no DSM-IV TR e CID-10 para diagnóstico dos transtornos invasivos do desenvolvimento. A validação para o Brasil foi realizada em trabalho colaborativo que contou com pesquisadores da Universidade Presbiteriana Mackenzie e do Instituto de Psiquiatria do Hospital de Clínicas de São Paulo (Sato *et al.*, 2009). Para as crianças menores que 4 anos é utilizado no lugar do QAA a escala Modified Checklist for Autism in Toddlers (M-CHAT), um instrumento de rastreamento precoce de autismo, ambos realizados por psicólogas.

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.

Entrevista de Histórico Familiar (adaptada da FIGS, do inglês *Family Interview for Genetics Studies*): para coletar informações sobre a estrutura e histórico psiquiátrico dos familiares do probando (Maxwell, 1992).

3.2.2. Avaliação neuropsicológica

Do ponto de vista dos critérios acadêmicos, funcionamento intelectual geral é definido pelo Quociente de Inteligência (QI ou equivalente). Segundo este parâmetro adotado pelo CID10 podemos definir 4 grupos de deficiência intelectual (DI): (1) DI leve, QI entre 50 e 69; (2) DI moderado, QI entre 35 e 49; (3) DI grave, QI entre 20 e 40 e (4) DI profundo, QI menor que 20. Na deficiência intelectual, como nas demais questões da psiquiatria, a capacidade de adaptação do sujeito ao objeto, ou da pessoa ao mundo, é o elemento mais fortemente ligado à noção de normal. Teoricamente, já que a unidade de observação é a capacidade de adaptação, deveriam ficar em segundo plano as questões mensuráveis de QI. Academicamente, é possível diagnosticar DI em indivíduos com QI entre 70 e 75, porém, que exibam déficits significativos no comportamento adaptativo. Cautelosamente o DSM-IV recomenda que o DI não deve ser diagnosticado em um indivíduo com um QI inferior a 70, se não existirem déficits ou prejuízos significativos no funcionamento adaptativo. Assim avaliamos tanto o QI como capacidades funcionais dos pacientes.

Escala de Inteligência Wechsler para Crianças – WISC-III: Este instrumento possibilita a avaliação da capacidade intelectual de crianças com idades entre 6 anos e 0 meses a 16 anos e 11 meses. Além do coeficiente intelectual (QI total) é possível avaliar o QI verbal e o QI de execução separadamente, a partir das treze provas propostas, sendo que cada uma avalia um aspecto diferente da inteligência. Para avaliação de adultos usamos a mesma escala versão adultos (Figueiredo *et al.*, 1998).

Vineland: Entrevista dirigida realizada com o cuidador que avalia o nível de independência pessoal e social dos indivíduos (Sparrow *et al.*, 1984).

3.2.3 Avaliação fonaudiológica

Para que os grupos verbais e não-verbais sejam melhor caracterizados e diferenciados, os sujeitos foram avaliados no que se refere as várias habilidades fonaudiológicas: fonologia, sintática, semântica, vocabulário, pragmática, competências metalinguísticas (incluindo leitura e escrita, dentre outras) e maturidade simbólica. Para tal, foram utilizados os instrumentos:

ABC – Autism Behavior Checklist: adaptado e traduzido para o Português Brasileiro como Inventário de Comportamentos Autísticos (ICA) (Martelleto, Pedromônico, 2005). O ABC/ICA é um questionário aplicado aos pais e constitui uma lista de 57 comportamentos associados ao autismo, os quais foram divididos em cinco áreas: estímulo sensorial; relacionamento; corpo e uso de objeto; linguagem; e, ajuda pessoal e social. Cada um dos comportamentos relacionados a estas subáreas recebe uma pontuação de 1 a 4, conforme sua associação com a patologia (4 pontos indicam maior associação), sendo possível pontuar por área e também geral. Quando a pontuação geral atinge o valor maior ou igual a 68, o indivíduo é classificado como autista.

Filmagens com a mãe/criança e examinadora/ criança em brincadeira livre (sem a interferência do interlocutor) e semi-dirigida (com a interferência do interlocutor). A partir das filmagens analisamos:

a) Maturidade simbólica - A avaliação da maturidade simbólica (Befi-lobes *et al.*, 2000) é utilizada para que possamos obter dados objetivos sobre o desenvolvimento cognitivo das crianças com alteração no desenvolvimento da linguagem. Esta avaliação consiste em duas situações brincadeira Livre e, imitação de esquemas gestuais simples e esquemas gestuais sequenciais em

rotinas familiares. Com o objetivo de observar as possíveis mudanças de comportamento a partir da intervenção de diferentes interlocutores, no caso o terapeuta fonoaudiólogo e o cuidador, acrescentamos outra situação denominada semi-dirigida.

b) Pragmática (Andrade *et al.*, 2000) - Para análise dos aspectos funcionais da linguagem e determinar o perfil comunicativo do sujeito utilizamos o teste de linguagem infantil ABFW na área pragmática. Neste teste, a unidade mínima de análise é o ato comunicativo e leva em conta os aspectos não linguísticos da comunicação e todos os meios utilizados (verbais, vocais e gestuais). Os dados obtidos permitem a análise do espaço comunicativo ocupado pela criança numa situação interacional e dos recursos comunicativos de que ela dispõe para tanto.

c) Atenção compartilhada (Menezes, Perissinoto, 2008) - Na avaliação da atenção compartilhada observamos os comportamentos de atenção compartilhada alternar, apontar, mostrar (por iniciativa ou resposta da criança) e olhar para a ação do adulto (sempre por resposta da criança). Este último incluiu a diferenciação do meio utilizado pelo adulto para chamar a atenção da criança (fala; gestos; ou, ambos).

Em sujeitos verbais acrescentamos as seguintes provas:

a) Vocabulário receptivo - Picture Vocabulary Test / Peabody (teste do vocabulário por imagens) (Capovilla *et al.*, 1997). Este teste avalia o desenvolvimento lexical ao domínio receptivo, ou seja, as habilidades de compreensão de vocabulário, de crianças entre dois anos e seis meses até 18 anos de idade. Ele fornece uma avaliação objetiva, rápida e precisa do vocabulário receptivo-auditivo em uma ampla variedade de áreas, incluindo pessoas, ações, qualidades, partes do corpo, tempo natureza, lugares, objetos, animais, termos matemáticos e ferramentas.

b) Vocabulário expressivo (Andrade *et al.*, 2000) - Utilizamos para avaliação do vocabulário expressivo a área de vocabulário do teste de linguagem ABFW. Tal instrumento consiste na verificação do conhecimento vocabular de 9 campos

conceituais: vestuário, animais, alimentos, meios de transporte, móveis e utensílios, profissões, locais, formas e cores e brinquedos e instrumentos musicais. Para cada um deles há um percentual de respostas que deve ser considerado como adequado para cada faixa etária.

c) Elaboração de narrativa autônoma e com tutela - Avaliada a partir da história “A pedra no caminho” que consiste em uma sequência de 5 figuras. Em um primeiro momento o paciente é apresentado as figuras por tempo suficiente para a observação. As figuras são recolhidas e o examinador estimula o paciente a fazer uma narrativa autônoma. Após a narração, o examinador cria uma situação diálogo sobre os aspectos centrais dos eventos propostos na sequência das figuras apresentadas (tutela). Em seguida é solicitado ao paciente nova narrativa sobre o mesmo tema, sem a presença das figuras.

Em sujeitos leitores utilizamos o protocolo para compreensão leitora de Textos Expositivos (Saraiva *et al.*, 2005) que analisa a velocidade e acurácia de leitura e a compreensão dos textos.

No **Quadro 2** podemos observar os critérios considerados para a classificação quanto a comunicação: uso de palavras, uso de frases com 3 ou mais palavras possuindo ou não conectivos e funcionalidade da linguagem (capacidade de nomeação, capacidade de narrativa, presença de ecolalias).

Como o intuito desta diferenciação refere-se a grupos com mais ou menos comunicação social, uma vez que todos os pacientes apresentavam déficits de sociabilidade, apenas os pacientes que faziam uso de frases com três ou mais palavras, capazes de estabelecer um diálogo, mesmo que podendo apresentar dificuldade na elaboração de narrativa (repetitiva e descontextualizada), presença de vocabulário restrito e podendo apresentar comunicação expressiva/receptiva abaixo do esperado foram classificados como verbais. Aqueles classificados como “pouco verbais”, mesmo apresentando a capacidade de nomeação e uso de palavras isoladas, mas sem comunicação funcional foram classificados junto com os não verbais.

Quadro 2 – Detalhamento da caracterização da linguagem de acordo com a avaliação fonoaudiológica.

Fenótipo	Resumo da caracterização da linguagem
Não verbal	Uso de vocalizações não articuladas e articuladas com entonação de língua e gestos não simbólicos
Pouco Verbal	Uso de palavras isoladas ou frases de 2 a 3 palavras sem uso de conectivos, linguagem imediata e concreta, não funcional, nomeação (repertório restrito) , não narra fatos, pode apresentar ecolalia
Verbal	Uso de frases com três ou mais palavras, capaz estabelecer um diálogo, podendo apresentar dificuldade na elaboração de narrativa (repetitiva e descontextualizada), vocabulário restrito podendo apresentar comunicação expressiva/receptiva abaixo do esperado

3.3 Coleta e armazenamento de material biológico

Após concordância da família e assinatura do TCLE, foi realizado punção venosa por profissional capacitado, para coleta de 10 ml de sangue venoso em tubos de EDTA (2 tubos de 5 ml). Após a coleta, os tubos foram devidamente armazenados a 4° C até o momento da extração. As amostras de DNA genômico foram obtidas a partir da camada de leucócitos do sangue periférico dos pacientes.

Extração de DNA: A partir do sangue coletado, foi realizado o método de extração utilizando sal (*salting out*) de acordo com (Miller *et al.*, 1988). Para alguns indivíduos em que a punção venosa não foi possível, o DNA foi obtido a partir da saliva, cuja coleta foi feita em kit saliva Oragene® do fabricante Genotek® de acordo com as instruções do fabricante (site: <http://www.dnagenotek.com/>).

A qualidade do DNA foi verificada por espectrofotometria (NanoDrop® ND-1000, ThermoScientific), foram usadas as amostras com razão de pureza de, no mínimo 1,8. A integridade do DNA das amostras foi avaliada por eletroforese em gel de agarose.

Para utilização das amostras de DNA nos experimentos de exoma, o mesmo foi quantificado pelo método de fluorometria utilizando o kit dsDNA BR Assay (Invitrogen), de acordo com as instruções do fabricante. Este kit permite a

leitura da fita dupla de DNA por utilizar um reagente intercalante, oferecendo especificidade e sensibilidade.

3.4 Captura e sequenciamento do exoma

Depois da extração de DNA, foi realizada no Laboratório de Genética Molecular da FCM-Unicamp, da equipe da Prof.^a Iscia Lopes Cendes, a construção das bibliotecas e enriquecimento com sequências exclusivamente de regiões exônicas utilizando kit TruSeq® Exome Enrichment Kit (Illumina Inc., San Diego, USA) segundo protocolo do fabricante. O sequenciamento com esse kit, segue o fluxo representado pela **Figura 2**.

Figura 2 – Fluxo para sequenciamento de exoma usando o kit TruSeq da Illumina. Em roxo estão indicadas os 3 passos do preparo da biblioteca.

Antes do primeiro passo da **Figura 2**, realizamos a fragmentação mecânica do DNA genômico com o equipamento Bioruptor® (Diagenode S.A.) seguindo protocolo previamente otimizado para obtenção dos fragmentos entre 200 a 300 pares de bases (100 ul de amostra com concentração de 10 ng/ul de amostras, em 13 ciclos na potencia máxima, 30s *on*, 30s *off*). A qualidade da fragmentação foi analisada através do Bioanalyzer (Agilent Technologies, Inc.), cujo resultado

exemplo está apresentado na **Figura 3**. Mesmo considerando a variabilidade esperada, é importante que o pico de concentração fique localizado entre o intervalo de 150 e 400pb, para garantir que durante os passos posteriores do preparo da biblioteca sejam selecionados quantidade suficiente de fragmentos no tamanho desejado de 200-300pb (**Figura 3B**).

Figura 3 - A) Imagem da corrida de eletroforese do Bioanalyzer demonstrando o intervalo de fragmentação das amostras. B) Curva de concentração (pg/ulx1000) de fragmentos de acordo com tamanhos observados (em pares de base).

Após fragmentação, foi realizado o preparo das bibliotecas individuais segundo o protocolo da Illumina (resumido na **Figura 4**). Em seguida foram combinadas 500ng das bibliotecas individuais para formar um *pool* com 6 amostras, que são hibridizadas (enriquecidas) e amplificadas por PCR (**Figura 5**).

Figura 4 – Fluxograma da preparao das bibliotecas individuais. A) Fragmentao do DNA genmico. B) Reparao das pontas 3' dos fragmentos em fita dupla e fosforilao. C) Adio de adeninas nas pontas. D) Ligao dos adaptadores Rd1 e Rd2 e índices s pontas adeniladas. Os adaptadores so necessrios para a ligao s *probes* das lâminas de sequenciamento. E) Desnaturao e amplificao do produto final.

Figura 5 – Fluxograma do enriquecimento das bibliotecas. A) Desnaturao da biblioteca de DNA dupla fita. B) Hibridao de *probes* biotiniladas s regies alvo (xons). C) Adio de *beads* magnticas de streptavidina para ligao s *probes* hibridizadas. D) Eluio dos fragmentos selecionados por captura magntica e amplificao dos fragmentos hibridizados.

A validação da biblioteca pronta é feita via Bioanalyser (**Figura 6B**) e a quantificação da biblioteca via PCR em tempo real (**Figura 6A e 6C**) .

Figura 6 - A) Imagem da corrida de eletroforese do Bioanalyser demonstrando o intervalo de fragmentação de 1 biblioteca de 6 amostras (destacada pelo retângulo vermelho). B) Curva de concentração (pg/ulx1000) dos fragmentos de acordo com tamanhos observados. C) Informações referentes ao pico de interesse (selecionado entre linhas roxas), representando 75% do total de fragmentos, tamanho médio (*avarage size*) de 409 pares de base.

A partir da informação de tamanho médio de fragmentos (para normalização) e a concentração dos fragmentos com adaptadores RD1 e RD2 (e portanto aqueles que serão sequenciados) por PCR em tempo real, foi possível calcular a concentração das bibliotecas. A quantidade necessária para realização do sequenciamento é de 4nM e o tamanho médio da biblioteca maior que 450 pb (o que leva em conta o tamanho dos fragmentos de éxons de 200-300 mais adaptadores e indexes).

O sequenciamento das bibliotecas foi realizado usando a plataforma HiSeq 1000 TM (Illumina, Inc.) realizado pela equipe do Centro de Genômica Funcional Aplicado à Agropecuária e Agroenergia da Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ/USP), do Prof. Dr. Luiz Coutinho (*facility* da FAPESP). Por meio da plataforma Illumina os fragmentos das bibliotecas são sequenciados de acordo com a técnica de sequenciamento por síntese (SBS) de acordo com a **Figura 7** e

Figura 8, pelo método *paired end*, isto é, são lidos 100 bases de cada extremidade do fragmento.

Figura 7 – Detalhamento da técnica de sequenciamento por síntese realizado na plataforma Illumina, parte A. (Fonte: <http://www.illumina.com/>)

Figura 8 – Detalhamento da técnica de sequenciamento por síntese realizado na plataforma Illumina, parte B. (Fonte: <http://www.illumina.com/>)

3.5 Análise de bioinformática

3.5.1 Controle de qualidade e mapeamento das sequencias geradas

Uma análise qualitativa da corrida foi realizada com o FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), afim de verificar se as sequências obtidas possuíam qualidade dada pela nota *Phred* dentro dos parâmetros para continuar a análise. A nota *Phred* aponta a probabilidade da base na sequência ser realmente uma base verdadeira (Ewing *et al.*, 1998), (Brockman *et al.*, 2008) de acordo com a equação (1):

$$Q = -10 \log_{10} P_{(1)}$$

onde Q é a nota *Phred* dada pela equação e P é a probabilidade do erro da base.

As sequencias lidas (*reads*) dos fragmentos das bibliotecas foram trimadas com o programa NGSToolkit (Patel, Jain, 2012), ou seja, seguindo os parâmetros padrões, 70% das bases de cada *read* deveria possuir nota *Phred* ≥ 20 , caso contrário seriam descartas. Em seguida, os *reads* foram alinhados com o algoritmo BWA (Li, Durbin, 2009) utilizando como referência a montagem do National Center for Biotechnology Information (NCBI) reference human genome build 37 (GRCh37/hg19).

Após esta etapa, o SAMtools (Li *et al.*, 2009) foi utilizado para a remoção da duplicata de PCR utilizando a opção *rmdup*. Esta etapa é recomendada, uma vez que a reação de PCR usada para amplificação da biblioteca pode introduzir artefatos nas análises posteriores (Altmann *et al.*, 2012).

Após as filtragens, foi realizada a chamada de variantes utilizando o programa GATK (Genome Analysis ToolKit) (DePristo *et al.*, 2011; McKenna *et al.*, 2010; Auwera *et al.*, 2013) com a opção *Haplotype Caller*, e critério de qualidade de *Phred Score* ≥ 30 (Altmann *et al.*, 2012; Brockman *et al.*, 2008). 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 *et al.*, 2001) para identificar o perfil dos parâmetros medidos.

A anotação funcional foi realizada com o ANNOVAR (Wang *et al.*, 2010), para isso os bancos de dados utilizados disponíveis no programa foram: genomicSuperDup, PhastConElements (Siepel *et al.*, 2005), SIFT (Kumar *et al.*, 2009), PolyPhen2 HDIV and HVAR Score (Adzhubei *et al.*, 2010; Adzhubei *et al.*, 2013), LTR Scores, MutationTaster Scores (“MutationTaster,” n.d.), FATHMM Scores (Shihab *et al.*, 2013), GERP++ Scores (Cooper *et al.*, 2010), e PhyloP Scores (Siepel *et al.*, 2006). Os bancos de dados dos 1000 genomas (1000g2012apr_all) (Altshuler *et al.*, 2010) e 6500 exomas (esp65000si_all - (“Exome Variant Server,” n.d.)) foram utilizados para calcular a frequência alélica alternativa.

3.5.2 Teste de erro mendeliano

Para identificar se o núcleo familiar pertencia a uma família biológica, calculamos os erros mendelianos utilizando o Plink (Purcell *et al.* 2007). A taxa de erro mendeliano pode ser proveniente de 3 principais fontes: mutações de novo, erro de genotipagem ou a família não biológica. Assim, o Plink retorna o número que representa este erro verificado em cada família (N_e). Esse tipo de erro só pode ser diagnosticado quando pelo menos um pai é homozigoto para o alelo em questão. Nesses casos, se diz que o locus é informativo para o trio (i). Para calcular a frequência do erro mendeliano dividimos a quantidade de erros verificadas (N_e) pelo total de de alelos informativos (N_i) em cada família.

$$\text{Erro mendeliano ()} = \frac{N_e}{N_i} * 100 \quad (2)$$

3.6 Análise de CNV

A identificação de CNVs foi realizada com o algoritmo ExomeDepth (Plagnol *et al.*, 2012), pacote do R. O algoritmo faz uma comparação utilizando a contagem de reads observado pelo esperado. O cálculo de reads esperado foi realizado utilizando uma população referência (pais dos probandos), e somente as referências que atingiram uma correlação de contagem de reads por alvo maior do que 0,978 foram utilizadas. Após esta etapa, para escolher os melhores CNVs foram feitas as seguintes filtrações: i-) percentil 99 do valor BF (Bayes Factor), ii-) não estar anotado no banco de dados de variantes comuns descrito por Conrad *et al.* (2010) e iii-) possuir tamanho $\geq 500\text{kb}$. O valor BF mostra a probabilidade da identificação do CNV dividida pela hipótese nula (número de cópia normal), este valor está expresso como $\log_{10}$ (razão).

3.7 Análise de SNV

Para a filtração das variantes *de novo* e raras deletérias foi realizado a filtração dos resultados de acordo com os seguintes passos:

Variações de novo (mutações não herdadas): São consideradas *de novo* variações exclusivas do probando (ou seja, não estar presente nos pais). Para isso, os pais deveriam ser homozigotos para a referência sem apresentar nenhum *read* de cobertura no alelo alternativo (Neale *et al.*, 2012).

Variações raras e herdadas dos genitores (mutações herdadas com $MAF < 0,1$): Para essa seleção, usamos a anotação dos banco de dados do 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 dbNSFP (<http://www.ncbi.nlm.nih.gov/SNP/>): $MAF \geq 0,05$: Variação Comum; $MAF < 0,05$ e $\geq 0,01$: menos comum; $MAF < 0,01$: Raro e; $MAF < 0,005$: Muito raro (Cirulli, Goldstein, 2010). Somente variações raras e muito raras foram consideradas para as análises posteriores.

Cobertura (DP): Foram filtradas as variantes que apresentaram cobertura menor do que 20X (ou seja, deve ter pelo menos 20 *reads* cobrindo o alelo). Esta taxa de cobertura é similar ou maior do que as obtidas em outros estudos (Iossifov *et al.*, 2012; Neale *et al.*, 2012; Sanders *et al.*, 2012; Wang *et al.*, 2013). Nos trios, para priorizar a seleção de SNVs e Indels raras herdadas foi utilizada uma cobertura de no mínimo 40X.

Qualidade do genótipo (GQ): a qualidade do genótipo representa a probabilidade (PL) do genótipo dado pelo GT ser homozigoto para referência (0/0), heterozigoto (0/1) e homozigoto para o alelo alternativo (1/1) maior ou igual a 20 (Neale *et al.*, 2012)□.

Strand bias (FS): calculado por teste de Fisher, indica se há algum desbalanço das fitas *Foward* e *Reverse* na chamada de variantes. Um desbalanço de fitas pode ocasionar a presença de variantes falso positivas (Guo *et al.*, 2012). O GATK computa o FS de acordo com D'Antonio *et al.* (2013) $FS \leq 60$ para SNVs e $FS \leq 200$ para INDELS.

Cobertura do alelo alternativo: Selecionadas somente as variantes que possuíam no mínimo 20% da cobertura total de *reads*, ou seja, número de *reads* no alelo da variante dividido pelo número de *reads* total precisava ser maior ou igual a 0,2.

Regiões gênicas: Selecionadas somente as variantes que estavam mapeadas em regiões exônicas codificadoras de proteína e em sítios de *splicing* (UTRs, regiões de RNAs não codificadores e intergênicas foram descartadas).

Tipos de variantes: Todos os SNVs não sinônimos e INDELS foram selecionados para busca e análise de mutações *de novo*. Nos trios, para as análises com a lista de genes com variantes raras herdadas de um dos pais ou em homozigose, somente as variantes *nonsense*, ou seja, identificadas como *stop loss*, *stop gain*, *splicing site* e *frameshift* foram selecionadas (Iossifov *et al.*, 2012)

Patogenicidade das variações não sinônimas: Para as variações presentes em bancos, usamos os resultados algoritmos de predição funcional do LJB2 (<https://sites.google.com/site/jpopgen/dbNSFP>) PolyPhen-2 (Adzhubei *et*

al., 2010; Adzhubei *et al.*, 2013) e GERP++ (Cooper *et al.*, 2010; Liu *et al.*, 2011, 2013). Seleccionamos as variações preditas deletérias ou provavelmente deletérias, seguindo os padrões detalhados no **Quadro 3**.

Quadro 3 – Patogenicidade da variação de acordo com score obtido por algoritmo.

Algoritmo	Patogenicidade
PolyPhen2	Provavelmente causa dano > 0.85 (D), possivelmente causa dano entre 0.85-0.15(P), benigno < 0.15(B)
GERP++	Quanto maior o número mais conservado (> 0). Valores > 2 são considerados potencialmente funcionais

3.8 Construção e análise das redes

Para construir as redes de interação proteína-proteína, foi usada a compilação de bancos de dados do site iRefIndex. Esta compilação une interações dos principais bancos com esse tipo de interações: BIND (Alfarano *et al.*, 2005), BioGRID (Stark *et al.*, 2006), CORUM (Ruepp *et al.*, 2008; Ruepp *et al.*, 2010), DIP (Salwinski *et al.*, 2004), HPRD (Keshava *et al.*, 2009), IntAct (Kerrien *et al.*, 2012), MINT (Licata *et al.*, 2012), MPPI (Pagel *et al.*, 2005) e OPHID (Brown, Jurisica, 2005). Usamos somente ligações com pelo menos duas fontes de evidência de interação física entre as proteínas do par correspondente.

A rede foi tratada como um grafo não-direcionado. Obtivemos assim 12464 genes ao buscarmos as interações proteína-proteína em humanos, no banco iRefIndex, com pelo menos duas fontes de confirmação. Ao buscarmos somente genes expressos em cérebro durante a infância (amostras com até 7 anos), de acordo com o Brain Atlas, obtivemos um conjunto de 10440 genes. Para considerarmos uma amostra como expressa em cérebro de crianças menores que sete anos, todas as bibliotecas presentes no Brain Atlas foram separadas por idade do paciente. Usamos todas as áreas cerebrais existentes, sem fazer nenhuma divisão por área cerebral. Genes que apresentavam expressão positiva

segundo os critérios do Brain Atlas em pelo menos uma biblioteca de qualquer região do cérebro em indivíduos menores que sete anos, incluindo a gestação, foram considerados como expressos em cérebros de criança.

Medidas de centralidade:

Várias medidas topológicas dos nós da rede foram calculadas para analisarmos quais genes têm papéis importantes na rede. A seguir serão listadas as medidas usadas.

- O **grau** (k) de um nó é a quantidade de ligações (ou vizinho que ele tem)
- A **medida de intermediação** (ou *betweenness centrality*) é uma fração de quantos caminhos mínimos na rede passam por um nó i :

$$C_i^{Btw} = \sum_{j=1}^N \sum_{k=1}^{j-1} \frac{g_{jk}(i)}{g_{jk}},$$

em que $g_{jk}(i)$ é o número de caminhos mínimos do nó j ao nó k passando por i e g_{jk} é o total de caminhos mínimos entre j e k .

A **medida de ponte** para um dado nó i , é definida como:

$$C_i^{Bdg} = C_i^{Btw} \times BC_i,$$

em que C_i^{Btw} é a centralidade entre-nós do nó i , e BC_i é o coeficiente que avalia as características locais de ponte na vizinhança do nó i , que é definida como

$$BC_i = \frac{d(i)^{-1}}{\sum_{v \in N(i)} \frac{1}{d(v)}}$$

em que $d(i)$ é o grau do nó i e $N(i)$ é o conjunto de vizinhos do nó i .

O coeficiente de agrupamento (*clustering coefficient*) é uma razão do número de arestas que existem entre os vizinhos de i pelo total que poderiam existir:

$$C_i^{clu} = \frac{2n}{k_i(k_i-1)},$$

em que n é o número de arestas entre os vizinhos de i e k_i é o grau do nó i . Essa medida também varia entre 0 e 1. O zero indicando que os vizinhos de i não se ligam e 1 indicando que todos se ligam.

A **medida do nó corretor** (*brokering*) (Cai *et al.*, 2010) usa o grau e o coeficiente de agrupamento para ser calculada:

$$\text{Brokering}(i) = (1 - C_i^{\text{Clu}}) \times k_{\text{norm}}(i)$$

em que k_{norm} é o grau do vértice i dividido pelo maior grau na rede.

3.9 Análise de vias e processos biológicos hiper-representados

A análise das vias e processos biológicos hiper-representadas foi feita através do site *WebGestalt* (<http://bioinfo.vanderbilt.edu/webgestalt/>) (Kanehisa *et al.*, 2012; Wang *et al.*, 2013), que para sua análise coleta as informações dos bancos Gene Ontology (versão 1.2, 11/11/2012) para ontologia gênica; KEGG Pathways (03/21/2011) e WikiPathways (11/11/2012) para vias; predição GLAD4U (1/26/2013) para associação de genes a doenças; NCBI Gene (10/26/2012) para bandas citogenéticas; e Mammalian Phenotype Ontology (04/10/2013) and Human Phenotype Ontology (04/10/2013) para fenótipos. Em todas as análises foram usados o genoma humano como referência e teste hipergeométrico com $p \leq 0,05$ com correção por Benjamini-Hochberg.

3.10 Considerações éticas

Este projeto envolve coleta de sangue ou saliva para análises genéticas em pacientes com TEA. Este projeto 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) específico para ele.

A participação no estudo ocorreu somente mediante autorização dos responsáveis das crianças entrevistadas, através do TCLE em que é garantida 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 os objetivos, as justificativas e os 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

4. RESULTADOS A – TEA SINDRÔMICO

4.1 Caracterização amostral

Foi coletado sangue dos pais e dois irmãos, **Paciente 1** e **Paciente 2**, (**Figura 9**), que apresentam o diagnóstico de TEA e características sindrômicas (**Quadro 4**), com alguns traços fenotípicos diferentes entre os irmãos (**Quadro 4**), porém nenhuma síndrome conhecida associada (resultados de cariótipo, teste de erros inatos do metabolismo e exame de X-Frágil negativos).

Figura 9 – Heredograma da família selecionada para o sequenciamento piloto

Quadro 4 – Caracterização fenotípica dos irmãos com TEA Sindrômico selecionados para o sequenciamento de exoma piloto.

P	1	2
Sexo e Idade	Masculino, 33	Feminino, 29
Sintomas (Primeira consulta)	(10 anos) Não verbal (falou até os 4 anos e depois parou de falar), atraso intelectual, enurese noturna, agitação e agressividade	(7 anos) Verbal ; atraso intelectual; enurese noturna, bronquite, dificuldade com mudanças de rotinas, sociável, ansiedade presente o que a torna agitada, porém não agressiva
Sintomas Atuais	crises de agressividade e agitação .	Crises de agitação psicomotora; Agitação, ansiedade e comportamento inadequado no período pré-menstrual;
Exames realizados	EEG, tomografia, cariótipo, erros inatos do metabolismo, x-frágil e RX para Idade óssea sem alterações	
Medidas Morfológicas	Peso: 97kg pP - 95-97 Altura: 1,90m pE= 97 (alta estatura) Perímetro cefálico: 60cm PC> p97 Relação envergadura/estatura: +1,01	peso: 91,45 pP (peso) > p97 Altura: 1,83 pE (estatura)> p97 Perímetro cefálico: 59cm pC > p97 Relação envergadura/estatura: +1,02

	<p>Padrão da implantação dos cabelos: implantação posterior baixa</p> <p>Estrutura, formato e implantação de orelhas: orelhas pequenas ($p < 3$), baixa implantação</p> <p>Tamanho do nariz: base alargada</p> <p>Tamanho e estrutura da face: fâcies alongada</p> <p>Distância intercantal externa: 13,5cm ($P > 97$)</p> <p>Distância intercantal interna: 4cm ($P > 97$)</p> <p>Boca e dentes: hipertrofia gengival, dentes apinhados</p> <p>Dedos e polegares: dedos alongados, deformidade em 3º e 4º quirodáctilos (falange medial)</p> <p>Tamanho e formato dos pés: entre P75-97</p>	<p>Padrão da implantação dos cabelos: implantação posterior baixa</p> <p>Estrutura, formato e implantação de orelhas: Orelhas pequenas ($p = 3$), baixa implantação</p> <p>Tamanho do nariz: base alargada</p> <p>Tamanho e estrutura da face: fâcies alongada</p> <p>Distância intercantal externa: 14,5cm ($P > 97$)</p> <p>Distância intercantal interna: 4cm ($P > 97$)</p> <p>Boca e dentes: hipertrofia gengival, dentes apinhados</p> <p>Dedos e polegares: dedos alongados, deformidade em 2º e 3º quirodáctilos (falange medial)</p> <p>Tamanho e formato dos pés: $p = 97$ (pés grandes)</p>
Linguagem	Não fala, apenas emite sons guturais e gritos.	Fala palavras e frases simples com dislalia de troca (principalmente "r" pelo "l").
Exame Psíquico	Calmo, risos imotivados , compreende comandos simples, dificuldade de obedecer alguns comandos (parar, não fazer) e ausência de imitação. Ausência de comunicação funcional e contato social adequado .	Calma, colaborativa e extremamente risonha (risos imotivados e facilmente motivados). Interage bem com examinador , sabe dizer o mês do seu aniversário, sua idade; Elogia examinador, pede beijo e abraço, contato social inadequado . Compreende comandos simples e alguns mais complexos utilizando-se de imitação.
Exame Neurológico	Sifoescoliose. Deambulação com passos largos sem tendência a quedas. Reflexos profundos hipoativos globalmente.	Cifose; Romberg negativo, mantém ortostase em um membro por pouco tempo; equilíbrio dinâmico: marcha lenta com tronco pouco inclinado para frente, sem tendências a queda.
Medicação	propranolol 40mg (1/2 cp ao dia); risperidona 1mg (3x ao dia); carbamazepina 200mg (3x ao dia); clorpromazina 25mg (3x ao dia); fenitoína 100mg (3x ao dia)	risperidona 1mg(2x/dia); carbamazepina 200mg(3x/dia); Fluoxetina 20mg (1x/dia)

Obs: Caracterização do histórico clínico foi realizado a partir da análise do prontuário dos pacientes e a partir de entrevistas de rotina durante consulta, o exame físico foi realizado pela psiquiatra Ana Paula Martins, com supervisão da Dra. Helena Brentani e observação da aluna Viviane Neri de Souza Reis. O caso foi discutido com Dr^a Kette Valente, especialista em epilepsia, ficando estabelecido que mesmo estando descrito no prontuário o paciente não apresenta epilepsia.

4.2 Qualidade do sequenciamento

Os *reads* foram sequenciados *paired-end* na plataforma Hi-Scan da Illumina, produzindo em média 45 milhões de *reads* (intervalo de 42-47 milhões de *reads*) com tamanho de 100 bases, resultando em média 4,5Gb por amostra (Tabela 1).

Tabela 1 – Sumário do rendimento do sequenciamento. Valores totais gerados, removidos pelo controle de qualidade (duplicatas de PCR) e valor final mapeado.

	Mãe	Pai	Paciente 1	Paciente 2
Dados obtidos no sequenciamento				
Total de <i>reads</i> gerados (nº de reads)	44552415	42228273	47090732	46562751
Cobertura calculada (reads/alvo) ¹	71,76	68,02	75,85	75,00
Remoção de duplicatas de PCR				
Nº de <i>reads</i> duplicados	12472845	11389155	13374238	13175835
% de <i>reads</i> duplicados	28,00	26,97	28,40	28,30
Nº de <i>reads</i> restantes	32079570	30839118	33716494	33386916
% de <i>reads</i> restantes	72,00	73,03	71,60	71,70
Mapeamento e cobertura no alvo				
Nº de <i>reads</i> mapeados	31923730	30685694	33539194	33205438
% de <i>reads</i> mapeados	99,51	99,50	99,47	99,46
Cobertura média no alvo (reads/alvo)	34,84	33,34	36,36	36,03
% de alvos cobertos 1X	59,14	59,18	59,25	59,15
% de alvos cobertos 20X	45,27	44,27	44,81	44,17

¹ Quantidade de reads geradas x 100 / valor do alvo

A análise qualitativa dos dados realizada com o FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) mostrou que a maior parte dos nucleotídeos sequenciados possuíam nota *Phred* média igual ou acima de 30, a qual é considerada de boa qualidade (Nielsen *et al.*, 2011). Somente, a região mais 3' do *read* possui uma nota *Phred* menor do que 30 (**Figura 10**).

Figura 10 - Boxplots das notas *Phreds* obtidas na corrida das amostras (A) Pai, (B) Mãe, (C) Paciente 1 e (D) Paciente 2 gerados pelo *FastQC*. No painel esquerdo estão representados os read 1 e no direito os reads 2. No eixo Y estão representadas as notas *Phreds*. No eixo X estão descritas as posições das bases variando de 1-100 (o tamanho de cada *read* é 100 bases). A linha azul representa as médias obtidas em cada posição. As regiões demarcadas de verde, laranja e vermelho representam notas *Phred* de alta, razoável e baixa qualidade, respectivamente.

Análise do conteúdo de GC

A análise de conteúdo CG é realizada para verificar se há algum viés de representação em relação ao conteúdo de GC do alvo. É possível notar pela **Figura 11a**, que os *reads* se encontram mais concentrados entre as regiões de 0,2 a 0,8 do conteúdo de GC. Isto é esperado pois os alvos possuem essa mesma distribuição no conteúdo de GC como mostrado da **Figura 11b**.

Figura 11 – Análise do Conteúdo de CG dos dados de sequenciamento das amostras do quad. (A) *Scatterplot* com a média da contagem de *reads* das 4 amostras em relação ao conteúdo de CG dos alvos. (B) Histograma da frequência de alvos em relação ao conteúdo de GC. (C) Histograma da cobertura média dos alvos em relação ao conteúdo de GC. (D) Histograma do tamanho médio dos alvos em relação ao conteúdo de GC.

4.3 Análise de CNV

Para a identificar se havia algum CNV proveniente das amostras dos pacientes 1 e 2, ambos os pais foram utilizados como referência. Nesta análise foram identificados 64 CNVs para cada uma das amostras, totalizando 128 variações. Entretanto, somente CNVs que obtiveram *threshold* igual ou maior do que o percentil 99 de BF (*Bayes Factor*) foram considerados para a análise. Assim, somente CNVs que tiveram pontuação maior ou igual a 551,1 foram analisados posteriormente. Utilizando este valor de corte, somente a duplicação do cromossomo 4 de aproximadamente 2,8 Mb permaneceu na análise, sendo que ambas as amostras apresentaram a mesma duplicação (**Tabela 2**). Além disso esse CNV não está descrito como comum segundo o banco de dados de Conrad e colaboradores (Conrad *et al.*, 2010).

Tabela 2 – Detalhes dos CNVs encontrados na análise ExomeDepth dos dados de exoma do quad.

Amostra	Coordenadas cromossômicas	Tamanho (b)	Nº de exons	BF	Reads esperados	Reads observados
Paciente 1	chr4:264464-3101121	2836657	491	592	61863	82639
Paciente 2	chr4:264464-3101121	2836657	491	827	62347	82084

Na **Figura 12** estão plotados os gráficos que mostram a taxa de *reads* esperados que é definida pela referência utilizada (no caso os pais), e a taxa de *reads* observados nas amostras do paciente 1 e 2. Como é possível notar, a quantidade de *reads* observados (pontos vermelhos) está acima da taxa de *reads* esperados demarcados na faixa cinza. Desta forma, esta taxa está acima do esperado na maior parte da região cromossômica, indicando uma duplicação da região.

Esta região é composta por 56 genes únicos seguindo as anotações de GenelD e destes, 26 (46%) estão expressos no cérebro de crianças menores do

que 7 anos de idade (**Tabela 3**) utilizando os bancos Brain Span Atlas of the Developing Human Brain (Hawrylycz *et al.*, 2012) (<http://brainspan.org>) e Allen Human Brain Atlas (<http://human.brain-map.org>).

Figura 12 – O gráfico mostra a taxa de reads observados pelo esperado, tudo que está marcado de cinza é onde a quantidade de reads esperado é similar a do observado com intervalo de confiança de 95%. Tudo que está marcado fora deste intervalo é onde a quantidade de reads observado foi diferente da do esperado. Só estão plotadas regiões com pelo menos 20X de cobertura.

TABELA 3 - Genes localizados na região da duplicação do cromossomo 4 .

Gene	Nome do gene	BRAIN ATLAS
<i>ABCA11P</i>	ATP-binding cassette, sub-family A (ABC1), member 11, pseudogene	NA
<i>ADD1</i>	adducin 1 (alpha)	SIM
<i>ATP5I</i>	ATP synthase, H⁺ transporting, mitochondrial Fo complex, subunit E	SIM
<i>C4orf48</i>	chromosome 4 open reading frame 48	NA
<i>CPLX1</i>	complexin 1	SIM
<i>CRIPAK</i>	cysteine-rich PAK1 inhibitor	SIM
<i>CTBP1</i>	C-terminal binding protein 1	SIM
<i>CTBP1-AS2</i>	CTBP1 antisense RNA 2 (head to head)	NA
<i>DGKQ</i>	diacylglycerol kinase, theta 110kDa	SIM
<i>FAM193A</i>	family with sequence similarity 193, member A	SIM
<i>FAM53A</i>	family with sequence similarity 53, member A	NA
<i>FGFR3</i>	fibroblast growth factor receptor 3	SIM
<i>FGFRL1</i>	fibroblast growth factor receptor-like 1	SIM
<i>GAK</i>	cyclin G associated kinase	SIM
<i>GRK4</i>	G protein-coupled receptor kinase 4	SIM
<i>HAUS3</i>	HAUS augmin-like complex, subunit 3	SIM
<i>HTT</i>	huntingtin	SIM
<i>HTT-AS</i>	HTT antisense RNA (head to head)	NA
<i>IDUA</i>	iduronidase, alpha-L-	SIM
<i>LETM1</i>	leucine zipper-EF-hand containing transmembrane protein 1	SIM
<i>LOC100129917</i>	uncharacterized LOC100129917	NA
<i>LOC100130872</i>	uncharacterized LOC100130872	NA
<i>MAEA</i>	macrophage erythroblast attacher	SIM
<i>MFSD10</i>	major facilitator superfamily domain containing 10	NA
<i>MFSD7</i>	major facilitator superfamily domain containing 7	NA
<i>MIR4800</i>	microRNA 4800	NA
<i>MIR943</i>	microRNA 943	NA
<i>MXD4</i>	MAX dimerization protein 4	SIM
<i>MYL5</i>	myosin, light chain 5, regulatory	NA
<i>NAT8L</i>	N-acetyltransferase 8-like (GCN5-related, putative)	NA
<i>NELFA</i>	negative elongation factor complex member A	NA
<i>NOP14</i>	NOP14 nucleolar protein	NA
<i>NOP14-AS1</i>	NOP14 antisense RNA 1	NA
<i>PCGF3</i>	polycomb group ring finger 3	SIM
<i>PDE6B</i>	phosphodiesterase 6B, cGMP-specific, rod, beta	NA
<i>PIGG</i>	phosphatidylinositol glycan anchor biosynthesis, class G	NA
<i>POLN</i>	polymerase (DNA directed) nu	NA
<i>RNF212</i>	ring finger protein 212	SIM
<i>RNF4</i>	ring finger protein 4	SIM
<i>RP11-503N18.1</i>	uncharacterized LOC402160	NA
<i>SCARNA22</i>	small Cajal body-specific RNA 22	NA
<i>SH3BP2</i>	SH3-domain binding protein 2	SIM
<i>SLBP</i>	stem-loop binding protein	SIM
<i>SLC26A1</i>	solute carrier family 26 (anion exchanger), member 1	NA
<i>SPON2</i>	spondin 2, extracellular matrix protein	NA
<i>TACC3</i>	transforming, acidic coiled-coil containing protein 3	SIM

TMED11P	transmembrane emp24 protein transport domain containing 11, pseudogene	NA
TMEM129	transmembrane protein 129	NA
TMEM175	transmembrane protein 175	NA
TNIP2	TNFAIP3 interacting protein 2	SIM
UVSSA	UV-stimulated scaffold protein A	NA
WHSC1	Wolf-Hirschhorn syndrome candidate 1	SIM
ZFYVE28	zinc finger, FYVE domain containing 28	SIM
ZNF141	zinc finger protein 141	NA
ZNF721	zinc finger protein 721	NA
ZNF732	zinc finger protein 732	NA

Validação com MLPA

Para validar esse achado realizamos a técnica de MLPA (amplificação multiplex de sondas dependents de ligação, do ingles Multiplex ligation-dependent probe amplification). Usamos o SALSA® MLPA® KIT P036-E1 HUMAN TELOMERE-3 e P070-B2 HUMAN TELOMERE-5 (MRC-Holland) designado para detectar deleções e duplicações de 1 ou mais genes na região terminal dos telômeros da amostra de DNA, e que cobre a nossa região de interesse, 16.3 do braço curto no cromossomo 4.

A reação de PCR do MLPA não é realizada na molécula de DNA do paciente e sim de sondas que anelam à sequências alvo pesquisadas. A reação pode ser dividida em 5 grandes passos: 1) Denaturação do DNA do paciente e hibridação das sondas, 2) Reação de Ligação, onde há a união dos fragmentos da sondas, 3) reação de PCR tendo a sequência da sonda como alvo, 4) Separação dos fragmentos por eletroforese capilar, 5) análise dos dados, feita comparando os picos obtidos com as sequências referências e assim o número de cópias aberrante é detectada.

O teste de MLPA confirmou que ambos pacientes apresentam não somente a duplicação no 4p16.3, como também apresentam uma deleção no cromossomo 8, região p23.3 (**Figura 13, A e B**). Tanto o pai como a mãe não apresentaram nenhuma alteração (**Figura 13, C e D**). Na região da deleção no 8p23.3 estão presentes os genes *FBXO25*, *MIR596* e *MYOM2*.

Figura 13 – Gráfico dos resultados da técnica de MLPA. A) Paciente 1; B) Paciente 2; C) Pai; D) Mãe. Os pontos em A e B fora dos limites marcados pelas linhas, acima de 1,3 e abaixo de 0,7 representam as CNVs: a duplicação da região 4p16.3 e a deleção da região 8p23.3, respectivamente.

4.4 Análise de SNV e Indels

A chamada de variantes resultou em média uma quantidade 40706 SNVs e 2687 INDELs por amostra. A quantidade de transição e transversão foi em média 29566 e 11069, respectivamente. A taxa Ti/Tv (transição/transversão) ficou em média 2,67 (**Tabela 4**). Esta taxa está próxima da taxa esperada descrita para regiões codificadoras de proteína (Chahrour *et al.*, 2012; Heinrich *et al.*, 2013).

Tabela 4 - Sumário das quantidades de variações SNVs e INDELs encontradas nas amostras sequenciadas do quad após controles de qualidade.

	Mãe	Pai	Paicente1	Paciente 2
Nº de variantes	44503	40293	44803	43608
Nº de posições genômicas	44593	40373	44907	43699
SNVs	41845	37881	42083	41016
INDELs	2748	2492	2824	2683
Heterozigotos	31399	27663	31619	30754
Homozigotos	13104	12630	13184	12854
Ti	30500	27438	30565	29762
Tv	11270	10366	11435	11168
<i>ratio</i> (Ti/Tv)	2,71	2,65	2,67	2,66

Após as filtragens de qualidade, de cobertura e de patogenicidade, foi encontrada apenas 1 variação *de novo*, na paciente 2 (*TUBAL3*). Além disso, entre as variações raras herdadas (de um dos pais) foram encontradas 150 variantes exclusivas para o paciente 1 e 117 exclusivas da paciente 2 (**Tabela 5**). A quantidade de variantes raras exclusivas de cada paciente não foi muito diferente do número de SNVs compartilhados entre os pacientes sendo de 130 (120 SNVs não sinônimas, 9 *frameshifts* e 1 não conhecida).

Tabela 5 - Sumário das quantidades de variações SNVs e INDELs encontradas nas amostras sequenciadas do quad após controles de qualidade.

	Missense	Stopgain	Stoploss	Frameshift	Splicing	Total
Paciente1	131	5	1	10	3	150
Paciente2	105	4	0	5	3	117
Total	236	9	1	15	6	267

Para identificar outras potenciais mutações patogênicas recessivas, investigou-se o número de variações homozigóticas em relação ao alelo alternativo. Nesta análise foram identificados 1 variação raras homozigotas exclusivas do paciente 1, identificadas 5 para a paciente 2, e 3 compartilhadas entre os pacientes (**Tabela 6**).

Tabela 6 – Lista de genes afetados com as SNVs e Indels de novo e raras selecionas após filtragens.

Tipo de variação	Genes		
	Paciente 1	Paciente 2	Compartilhadas
De novo	0	<i>TUBAL3</i>	0
Homozigose (alelo alternativo)	PLK5	<i>RASSF10, NBPF8, OR11H2, SRA1*</i>	<i>GOLGA6L10*, HLAB</i>

* Genes com mais de uma mutação

Com a lista de genes gerada mostrada na tabela 5, somada as variantes raras herdadas de um dos pais, foram selecionados genes que eram expressos em cérebro de crianças menor do que 7 anos que estavam depositados no BRAIN Atlas. Desta forma, geramos 2 listas para realizar anotação funcional de vias de cada paciente (66 genes exclusivos do paciente 1 e 61 exclusivos da paciente 2).

4.5 Análise de vias e processos biológicos hiper-representados

A análise da lista de genes selecionadas do paciente 1, apresentou hiper-representação significativa da região cromossômica 16p, mais especificamente a 16p13 (**Tabela 7**). Enquanto que a mesma análise na Paciente 2 resultou em hiper-representação significativa em 2 regiões cromossômicas, 1p e na 17q (**Tabela 8**). Todas análises apresentam as seguintes informações do cálculo estatístico: Numero total de genes referência na categoria (C), numero de genes da lista estudada observada na categoria (O), numero de genes esperado na categoria (E), taxa de hiper-representação (R), p-valor do teste estatístico hipergeométrico (rawP), e p-valorajustado para múltiplas testagens.

Tabela 7 – Resultado da análise no Webgestalt de regiões citogenéticas hiper-representadas por mais de 5 genes alterados no Paciente 1.

Região Cromossômica	Nº de Genes	Gene	Estatística
16p13	7	<i>NLRC3, PPL, PDIA2, DNASE1, TRAP1, NMRAL1, E4F1</i>	C=351; O=7; E=0.53; R=13.23; rawP=1.03e-06; adjP=1.25e-05
16p	9	<i>PDIA2, DNASE1, TRAP1, ZNF768, NMRAL1, PPL, NLRC3, TGFB111, E4F1</i>	C=699; O=9; E=1.05; R=8.54; rawP=1.04e-06; adjP=1.25e-05

Tabela 8 – Resultado da análise do Webgestalt mostrando hiper-representação em regiões cromossômicas por genes alterados na Paciente 2.

Região Cromossômica	Nº de Gene	Genes	Estatística
1p	7	<i>SCMH1, CROCC, HCRTR1, KIF17, ENO1, KIAA0754, UBR4</i>	C=1520; O=7; E=2.11; R=3.31; rawP=0.0050; adjP=0.0400

Além disso a análise da lista de genes da Paciente 2, resultou na hiper-representação do módulo de interação proteína-proteína de formação de axônio (**Tabela 9**). e também na ontologia fenotípica de contratura congênitas de articulações (**Tabela 10**).

Tabela 9—Resultado da análise do Webgestalt mostrando hiper-representação em ontologia fenotípica por genes alterados na Paciente 2.

Banco	Human Phenotype Ontology
Nome do Fenótipo	Contracturas congénitas
Estatística	C=52; O=3; E=0.12; R=24.70; P=0.0001; adjP=0.0140
Genes	<i>HSPG2, ITGB4, TSEN54</i>

Tabela 10 - Resultado da análise do Webgestalt mostrando hiper-representação em módulos PPI por genes alterados na Paciente 2.

Banco	Modulos PPI
Nome do Módulo	Hsapiens_Module_289
Processo Biológico	Orientação do axônio (p=1.55e-15)
Componente celular	spectrina (p=5.11e-14)
Função celular	Constituinte estrutura do citoesqueleto (p=5.74e-09)
Estatística	C=14; O=2; E=0.05; R=39.38; rawP=0.0011; adjP=0.0242
Genes	<i>SPTBN4, ANK3</i>

5. RESULTADOS B – TRIOS TEA ESPORÁDICO

5.1 Caracterização amostral

Após triagem realizada por médicos do grupo PROTEA usando os critérios do DSM-V foram selecionados 18 pacientes portadores de TEA esporádico e sem síndrome genética caracterizada. (**Tabela 10**). Como parte do protocolo de pesquisa foi feita a CARS, avaliação neuropsicológica e fonoaudiológica. Não foi possível realizar a CARS do paciente do trio 53, mas análise detalhada do histórico médico prévio e avaliação fonoaudiológica confirmam o diagnóstico de TEA.

Não foi possível a realização da avaliação neuropsicológica em 4 pacientes e 2 tiveram apenas o QI estimado. Faziam parte de outro projeto de validação de diagnóstico em andamento no ambulatório 6 pacientes, assim também foram submetidos a ADOS. Como podemos notar 3 casos apresentaram pontuação na CARS abaixo de 30 (Trios 7, 20 e 30), no entanto todos tiveram confirmação diagnóstica baseados nos critérios de DSM-V por um outro médico, assim como diagnóstico compatível com TEA nas avaliações neuropsicológicas e fonoaudiológicas. O trio 7 também teve seu diagnóstico confirmado pela ADOS.

Todos os trios foram entrevistados buscando mais detalhamento do histórico familiar do paciente, assim o trio 31 foi excluído por além de apresentar histórico de sofrimento fetal (indicativo de transtorno mental orgânico), tinha um histórico familiar de transtorno mental em diversos membros da família (3 tios paternos com um transtorno mental grave não investigados, 2 falecidos na infância e 1 na adolescência; 1 tio paterno com deficiência mental leve e problemas de linguagem; e 1 tio materno com Síndrome de Down).

Na **Tabela 11** apresentamos tanto o QI como o escore da Vineland para os diferentes pacientes.

Tabela 11 – Resultados da caracterização fenotípica dos trios de pacientes com TEA esporádico, selecionados para o estudo após triagem diagnóstica.

Trio	Sexo	Linguagem	Comorbidade	ADOS	Idade	CARS	QI	Vineland
4	M	Pouco verbal			12	38,5	60	48
9	M	Não verbal		22	6	49	50	50
16	M	Não verbal			32	49,5	n/a	20
² 22	M	Não verbal	Epilepsia		19	32	n/a	20
23	M	Pouco verbal		20	10	47	50	56
¹ 31	M	Não Verbal	Epilepsia		27	n/a	n/a	20
² 33	M	Pouco verbal		16	5	37	64	n/a
37	M	Pouco verbal			5	35,5	86	68
54	M	Não Verbal			3	40	67	n/a
58	F	Pouco verbal			16	32	n/a	31
				Media	12,0	41,6	62,6	45,5
7	M	Verbal		14	9	27	58	61
17	M	Verbal			19	36	85	n/a
20	M	Verbal			18	24	50	n/a
24	M	Verbal			19	32,5	n/a	51
30	M	Verbal	TDAH		8	28,5	91	80
43	M	Verbal		15	14	30	82	60
50	F	Verbal		15	8	40	50	65
53	M	Verbal			21	n/a	n/a	24
				Media	14,5	31,1	69,3	56,8
				P-valor (0,05)	0,273845	0,004228*	0,266292	0,150749

¹Trio excluído por histórico de sofrimento fetal e histórico familiar de transtorno mental grave.

² Trios excluídos em passos subsequentes, relacionados ao controle de qualidade do sequenciamento.

Sendo assim, após confirmação diagnóstica e caracterização fenotípica, os trios restantes (n=15) foram separados em dois grupos amostrais, o grupo de não verbais e pouco verbais (Grupo 1 – G1) e o de verbais (Grupo 2 – G2), a partir da análise da avaliação fonoaudiológica .

5.2 Qualidade de Sequenciamento

5.2.1 Análise qualitativa dos dados

Após o sequenciamento dos fragmentos de DNA pela técnica *paired end* no *Illumina Hiseq*, a análise qualitativa das sequências mostrou que os dados gerados possuíam em média a nota *Phred* maior ou igual a 30 (**Figura 14**).

Figura 14 -: Boxplot das notas Phreds obtidas na corrida da amostra 8126 gerados pelo FastQC. No eixo Y estão representadas as notas Phreds. No eixo X estão descritas as posições das bases variando de 1-100 (tamanho do read 100 bases). A linha azul representa a média das bases.

Foram sequenciados em média uma quantidade de 135 milhões de *reads* (intervalo de 34-189 milhões) por amostra, com tamanho de 100 bases, resultando em 13,5Gb de bases sequenciadas por amostra. Após filtragem com NGS Toolkit a quantidade de sequências obtidas foi em média de 116 milhões de *reads* (11,6 Gb, representado cerca de 85% dos *reads* iniciais) por amostra (**Anexos, Tabela 1**).

Após a segunda filtragem para remoção de duplicatas, quantidade média de *reads* restantes foi em média de 72 milhões (7,2 Gb bases sequenciadas), resultando numa taxa de 47% de *reads* não duplicados. Desses *reads* não duplicados cerca de 98,7% foram mapeados no genoma humano (**Anexos, Tabela 2**).

Com os *reads* mapeados, foi obtida a cobertura nos alvos delimitados pelo exoma. A cobertura média foi de 62X e cerca de 78% das regiões alvos foram cobertas pelo menos 20X (**Anexos, Tabela 3**).

A comparação entre a quantidade de *reads* gerados com a porcentagem de alvos cobertos pelo menos 20X mostra que as amostras, exceto a T33-1 (probando do Trio 33), estão no mesmo patamar de porcentagem de alvos cobertos pelo menos 20X (**Figura 15a**). O mesmo padrão é visto pela cobertura média, é importante ressaltar que a partir de 40X de cobertura, as amostras mostram quantidades similares de alvos cobertos pelo menos 20X (**Figura 15b**). Considerando a discrepância da qualidade de cobertura as amostras do Trio 33, foram excluídas das análises de CNVs e SNVs.

Figura 15 – Gráficos de saturação dos alvos cobertos. Porcentagem de alvos do kit TrueSeq (62Mb) cobertos pelo menos 20X (eixo y) em relação aos (A) reads gerados em cada uma das corridas (eixo X) e (B) cobertura média nos alvos (eixo X). Os pontos circulados indicam os valores da amostra T33-01 (probando do trio 33).

5.2.2 Análise do conteúdo de CG

Nos gráficos da Figura 16 há a representação do conteúdo de GC das regiões alvos em relação à média de contagem de *reads* (**Figura16A**), frequência de alvos (**Figura16B**) e média de cobertura de *reads* (**Figura16C**).

A quantidade de *reads* localizadas entre conteúdo de CG de 0,3 - 0,7 é alta como mostrado na **Figura 16A**. Entretanto, a quantidade de alvos (True Seq) com este intervalo de CG se concentra nesta mesma faixa (**Figura16B**), assim é mais provável que as quantificações de contagem de *reads* acabem se concentrando neste mesmo intervalo. Contudo, ao olhar para o último gráfico que revela a cobertura média por região de conteúdo de CG, é possível verificar que esta se mantém equivalente entre 0,2 - 0,6 no conteúdo de CG e, somente a partir de conteúdo de CG 0,7, esta taxa de cobertura é afetada drasticamente.

Figura 16 –Análise do Conteúdo de CG. (A) Scatterplot mostrando a média da contagem de reads das 54 amostras em relação ao conteúdo de CG dos alvos. (B) Histograma da frequência de alvos em relação ao conteúdo de GC e (C) Histograma da cobertura média dos alvos em relação ao conteúdo CG.

5.2.3 Chamada de variantes (GATK)

Em média, foram identificadas 54490 SNVs e 6955 *indels* por amostra, em média a taxa de transição e transversão foi 38517 e 15873, respectivamente, resultando em um *ratio* (transição/transversão) médio de 2,43 (**Anexos, Tabela 4**). Para determinar quais seriam as variantes funcionais, a ferramenta ANNOVAR (Wang et al. 2010) foi utilizada para anotar as variantes e inserções e deleções para examinar o impacto funcional no gene.

5.2.4 Taxa de Erros Mendelianos

Na análise de controle de qualidade pelo cálculo da taxa de erros mendelianos (**Tabela 12**), o trio 22 apresentou um valor muito acima do permitido e foi excluída das análises seguintes.

Tabela 12 – Resultados do cálculo de erros mendelianos através do programa PLINK.

TRIO	Número de alelos informativos	Erros mendelianos	Taxa (%)
TRIO_16	15735	195	1.239
TRIO_17	16523	127	0.769
TRIO_20	15162	94	0.620
TRIO_22*	15821	4261	26.933
TRIO_23	15395	95	0.617
TRIO_24	18373	150	0.816
TRIO_30	14924	138	0.925
TRIO_31**	16757	111	0.662
TRIO_33**	9755	78	0.800
TRIO_37	16079	120	0.746
TRIO_43	15880	127	0.800
TRIO_4	11785	17	0.144
TRIO_50	16073	18	0.112
TRIO_53	16218	128	0.789
TRIO_54	15692	143	0.911
TRIO_58	15057	23	0.153
TRIO_7	14373	86	0.598
TRIO_9	15666	134	0.855

* Família retirada por possuir erro mendeliano $\geq 5\%$

** Famílias retiradas em etapas anteriores

5.3 Análise de CNV

Inicialmente foram identificados 4033 CNVs nas 15 famílias estudadas, pois três famílias foram descartadas previamente dos estudos como mostrados anteriormente (Trio 22, Trio 33 e Trio 31). Dentre os CNVs encontrados somente foram considerados aqueles que possuíam BF igual ou acima do percentil 99 (Figura 17), resultando em 26 CNVs restantes.

Figura 17 - Boxplot do BF (Bayes Factor) dos CNVs encontrados nos probandos. O valor BF mostra a probabilidade da identificação do CNV dividida pela hipótese nula (número de cópia normal), este valor está expresso como log10 (razão). Logo, quanto maior valor maior a chance do CNV identificado estar presente.

Dos 26 CNVs identificados, 20 (77%) estavam anotados como comuns no banco definido por Conrad *et al.* (2010) e 24 eram menores que 500kb, desta forma somente 2 foram selecionados para a análise restante (**Tabela 13**).

Tabela 13: Sumário dos CNVs encontrados na análise de 15 trios.

TRIO	Coordenadas cromossômicas	Tamanho (bp)	Nº de éxons	BF	Reads esperados	Reads observados	Taxa de reads
23	chr17:43502930-44375108	872178	75	140	4889	2755	564
30	chr3:195776156-197351068	1574912	184	734	24421	13133	538

Para identificar se as deleções encontradas nos trios 23 e 30 poderiam ser deleções *de novo*, a mesma análise foi feita com as amostras referentes aos pais. A **Figura 18a** mostra que o probando do trio 23 possui uma deleção de ~872kb, entretanto o pai mostra um perfil parecido nessa mesma região, indicando portanto que essa deleção provavelmente foi herdada do pai. Já para o trio do probando 30 (**Figura 18b**), há uma grande deleção de ~1,5M, que não aparece em nenhum dos pais. Logo, essa região é dada como um CNV *de novo*.

Figura 18: O gráfico mostra a taxa de reads observados pelo esperado, tudo que está marcado de cinza é onde a quantidade de reads esperado é similar a dos observados com intervalo de confiança de 95%. Tudo que está marcado fora deste intervalo é onde a quantidade de reads observado foi diferente da do observado. Só estão plotadas regiões com pelo menos 40X de cobertura.

Este CNV de novo foi encontrado no chr 3: 195776156-197351068. Foram mapeados 29 genes na região, na **Tabela 14** podemos observar quais destes

genes foram encontrados em pelo menos uma região do cérebro de indivíduos menores que sete anos e quais foram encontrados no banco do UniGene. Optamos por usar apenas os dados do Brain Atlas, pois no UniGene alguns transcritos eram representados por apenas uma etiqueta de sequência expressa do inglês *expressed sequence tag* (EST).

Tabela 14—Genes presentes na região chr 3: 195776156-197351068, localização da deleção *de novo* encontrada no paciente do trio 30.

Gene	Nome	¹ BRAIN ATLAS	UNIGENE (ESTS)
CEP19	centrosomal protein 19kDa	SIM	SIM
DLG1	discs, large homolog 1 (Drosophila)	SIM	SIM
FBXO45	F-box protein 45	SIM	SIM
MFI2	antigen p97	SIM	SIM
NCBP2	nuclear cap binding protein subunit 2, 20kDa	SIM	SIM
PAK2	p21 protein (Cdc42/Rac)-activated kinase 2	SIM	SIM
PCYT1A	phosphate cytidylyltransferase 1, choline, alpha	SIM	SIM
RNF168	ring finger protein 168, E3 ubiquitin protein ligase	SIM	SIM
SEN5	SUMO1/sentrin specific peptidase 5	SIM	SIM
TCTEX1D2	Tctex1 domain containing 2	SIM	SIM
TFRC	transferrin receptor (p90, CD71)	SIM	SIM
UBXN7	UBX domain protein 7	SIM	SIM
BDH1	3-hydroxybutyrate dehydrogenase, type 1	NÃO	FETO/ADULTO?
C3orf43	chromosome 3 open reading frame 43	NÃO	NÃO
LOC152217	uncharacterized LOC152217	NÃO	FETO?
LOC220729	succinate dehydrogenase complex, subunit A, flavoprotein (Fp) pseudogene	NÃO	FETO/ADULTO?
LOC401109	uncharacterized LOC401109	NÃO	NÃO
LRRC33	leucine rich repeat containing 33	NÃO	NÃO
PIGX	phosphatidylinositol glycan anchor biosynthesis, class X	NÃO	FETO/ADULTO?
PIGZ	phosphatidylinositol glycan anchor biosynthesis, class Z	NÃO	FETO/ADULTO?
SLC51A	solute carrier family 51, alpha subunit	NÃO	NÃO
TM4SF19	transmembrane 4 L six family member 19	NÃO	FETO/ADULTO?
TM4SF19-TCTEX1D2	TM4SF19-TCTEX1D2 readthrough	NÃO	FETO/ADULTO?
WDR53	WD repeat domain 53	NÃO	FETO/ADULTO?
ZDHHC19	zinc finger, DHHC-type containing 19	NÃO	FETO/ADULTO?
DLG1-AS1	DLG1 antisense RNA 1		NÃO/TESTIS ONLY
MFI2-AS1	MFI2 antisense RNA 1		FETO?
MIR4797	microRNA 4797		SIM

5.4 Análise de SNV e Indels

Após filtragem de qualidade de sequenciamento e patogenicidade, o número de variações *de novo* e as raras herdadas de um dos pais ou em homozigose foi 1327 SNVs não-sinônimas e 135 Indels (**Tabela 15**), alterando 1245 genes. Do total de variações, 8 são *de novo*, 1 é um Indel com alteração *frameshift* em homozigose herdadas de pais heterozigotos (**Tabela 16**) e 98 são *SNV nonsense* (stopgain/stoploss), regiões de *splicing* ou *frameshift*, herdados de um dos pais (**Quadro 5**).

TABELA 15 – Número de variações SNV e Indels, *de novo* e herdadas raras por trio, separado de acordo com o grupo amostral 1 (pacientes Pouco ou não verbais) e 2 (pacientes verbais).

		Variações <i>de novo</i>			Variações Herdadas			Total Geral
		Indel	SNV	Total	Indel	SNV	Total	
Grupo 1								
TRIO 4	-	-	-	8	66	74	74	
TRIO 9	-	-	-	8	81	89	89	
TRIO 16	-	-	-	13	104	117	117	
TRIO 23	-	-	-	6	73	79	79	
TRIO 37	-	-	-	12	97	109	109	
TRIO 54	-	1	1	13	98	111	112	
TRIO 58	-	2	2	4	85	89	91	
Total Grupo 1	-	3	3	64	604	668	671	
Grupo 2								
TRIO 7	1	-	-	6	60	66	67	
TRIO 17	-	-	-	12	131	143	143	
TRIO 20	-	1	1	8	69	77	78	
TRIO 24	1	1	2	11	153	164	166	
TRIO 30	-	-	-	6	52	58	58	
TRIO 43	-	1	1	5	68	74	75	
TRIO 50	-	-	-	11	78	89	89	
TRIO 53	-	-	-	10	106	116	116	
Total Grupo 2	2	3	5	69	717	786	791	
Total Geral	2	6	8	133	1321	1454	1462	

Tabela 16 – Relação de SNVs e Indels *de novo* ou herdados homozigotos com alteração de frameshift dos pacientes com TEA após filtragem de qualidade e patogenicidade.

Trio	Cr.	Começo	Fim	Ref.	Var.	Gene	Tipo de Variação	Het / Hom	VAF	Cobertura										PP2	GERP
										Probando			Mãe			Pai					
										DP	Ref	Var	DP	Ref	Var	DP	Ref	Var			
54	3	135870098	135870098	C	A	MSL2	SNV missense <i>de novo</i>	Het	0,50	103	52	51	86	86	0	101	101	0	D	4,14	
58	12	4553370	4553370	C	T	FGF6	SNV missense <i>de novo</i>	Het	0,45	51	28	23	56	56	0	44	44	0	D	4,73	
58	12	56355517	56355517	C	T	PMEL	SNV splicing <i>de novo</i>	Het	0,55	76	34	42	72	72	0	48	48	0	n/a	4,89	
7	1	74507345	74507345	-	TGTG CCT	LRR1Q3	Ins. frameshift <i>de novo</i>	Het	0,27	49	36	13	86	86	0	66	66	0	n/a	n/a	
20	18	77893505	77893505	C	G	ADNP2	SNV missense <i>de novo</i>	Het	0,21	24	19	5	31	31	0	20	20	0	D	5,44	
24	19	50059805	50059805	C	T	NOSIP	SNV splicing <i>de novo</i>	Het	0,22	23	18	5	28	28	0	23	23	0	n/a	4,62	
24	15	43360099	43360100	TT	-	UBR1	Del. frameshift <i>de novo</i>	Het	0,55	66	30	36	92	92	0	69	0	0	n/a	n/a	
43	9	101068527	101068527	A	G	GABBR2	SNV misense <i>de novo</i>	Het	0,56	82	36	46	99	99	0	90	90	0	P	5	
20	1	54605318	54605318	-	G	CDCP2	Ins. frameshift Herdada	Hom	0,96	84	3	81	138	79	59	114	65	49	n/a	n/a	

G. Grupo; Cr. Cromossomo; Ref. Alelo Referência; Var. Alelo Variante; Ins. Inserção; Del. Deleção; Hom. Homozigoto; Het. Heterozigoto; VAF - Frequência do alelo variante; DP – Cobertura total do alelo (numero de fragmentos com o alelo); PP2 – Algoritmo de Patogenicidade Polyphen2; GERP - Algoritmo de Conservação.

Quadro 5 – Lista de genes afetados por variações herdadas raras, consideradas deletérias SNVs *nonsense*, de regiões de *splicing* ou *frameshift*, herdadas em heterozigose, separado por Grupo amostral (Símbolo do Gene_Tipo)

Grupo 1		
<p>AMZ1 stopgain SNV</p> <p>C17orf80 frameshift deletion</p> <p>C8orf44 frameshift deletion</p> <p>CARD6 frameshift insertion</p> <p>CD96 splicing</p> <p>CDCP2 frameshift insertion</p> <p>CHIT1 stopgain SNV</p> <p>CRELD2 frameshift deletion</p> <p>DNHD1 splicing</p> <p>DNHD1 stopgain SNV</p> <p>ELP2 frameshift insertion</p> <p>FAM111A frameshift insertion</p> <p>FAM120A splicing</p> <p>FAT2 frameshift deletion</p> <p>FCRL3 splicing</p> <p>FILIP1L stopgain SNV</p> <p>GMPR frameshift insertion</p>	<p>HGC6.3 frameshift insertion</p> <p>HMMR stopgain SNV</p> <p>HSPA4 stopgain SNV</p> <p>KCNAB3 stopgain SNV</p> <p>LRMP splicing</p> <p>MAML3 (2) frameshift deletion</p> <p>MBD4 frameshift deletion</p> <p>NOTCH2NL splicing</p> <p>OR5F1 frameshift insertion</p> <p>P2RX6 splicing</p> <p>PDE4DIP frameshift deletion</p> <p>PRR21 frameshift insertion</p> <p>RHBG splicing</p> <p>RPUSD4 splicing</p> <p>SDSL frameshift deletion</p> <p>SEC23A stopgain SNV</p> <p>SLC28A1 frameshift deletion</p>	<p>SPZ1 frameshift insertion</p> <p>SRA1 (2) frameshift insertion</p> <p>ST13 stoploss SNV</p> <p>SULT1C4 stopgain SNV</p> <p>SYT15 splicing</p> <p>TAF6 frameshift deletion</p> <p>TAS2R8 frameshift insertion</p> <p>TECTB splicing</p> <p>TIMD4 splicing</p> <p>TMPRSS7 stopgain SNV</p> <p>TNFRSF10 frameshift deletion</p> <p>C</p> <p>TRIM58 stoploss SNV</p> <p>TYRO3 splicing</p> <p>UBQLNL frameshift deletion</p> <p>ZNF492 splicing</p>

Grupo 2		
ALDH1B1 frameshift deletion	FHAD1 splicing	OPN4 frameshift deletion
AMICA1 splicing	GFRAL frameshift insertion	<u>PDE4DIP</u> frameshift deletion
<u>C17orf80</u> frameshift deletion	GJB7 frameshift deletion	PKD1L2 splicing
C2orf76 splicing	GMIP splicing	PKD1L2 (3) frameshift deletion
CCDC158 stopgain SNV	GPR113 stopgain SNV	PPP1R36 stopgain SNV
CELA1 frameshift deletion	GPRIN1 (2) frameshift deletion	PRSS53 frameshift insertion
CFTR stopgain SNV	GUF1 splicing	RBL1 stopgain SNV
<u>CHIT1</u> stopgain SNV	LGALS3BP stopgain SNV	RP1L1 (2) frameshift insertion
CLDN16 frameshift deletion	LNX1 splicing	SAMD3 frameshift insertion
COL11A1 splicing	LRGUK stopgain SNV	<u>SYT15</u> splicing
COL23A1 splicing	METTL7B stopgain SNV	<u>TAS2R8</u> stopgain SNV
COL28A1 stopgain SNV	MRE11A stopgain SNV	TEKT3 stopgain SNV
CYP3A5 frameshift insertion	MSR1 stopgain SNV	THBS1 stopgain SNV
DIS3 splicing	MYCBP2 splicing	UGGT1 splicing
DMBT1 frameshift insertion	MYOT splicing	VWDE splicing
EMR1 frameshift deletion	NMI frameshift insertion	WIF1 splicing
FCRL2 frameshift deletion	<u>NOTCH2NL</u> Splicing	ZNF527 frameshift insertion

5.5 Análise de Redes

A partir da lista de genes das variações *de novo* e herdadas raras deletérias descritos na **Tabela 15** e **Quadro 5**, selecionamos as encontradas no Brain Atlas em regiões cerebrais de indivíduos menores que sete anos, para realizar as análises de rede.

Para a rede do Grupo 1 existiam 55 genes (51 genes únicos), 20 são expressos em cérebro de criança, dos quais 19 estavam no interatoma. Usando esses 19 genes como semente, o crescimento da rede resultou em 206 genes (**Figura19**). Para a rede do Grupo 2, existiam 61 genes (56 genes únicos), 22 estão expressos em cérebro de criança, dos quais 20 estavam no interatoma. O crescimento da rede usando esses 20 como semente, resultou em 293 genes (**Figura 20**). O conjunto de intersecção entre as redes foi formado por 15 genes: *NOS3*, *POLR2G*, *RELA*, *SMAD2*, *SMAD3*, *SRC*, *STAT1*, *STAT3*, *STAT5B*, *SUV39H1*, *TAF1*, *TRAF2*, *UBC*, *UBQLN4* e *YWHAG*. Na **tabela 17** podemos observar a comparação de medidas topológicas entre as mesmas.

Tabela 17– Comparação dos atributos das redes do Grupo 1 e Grupo 2.

Atributo	Rede do G1	Rede do G2
Número de <i>nodes</i>	206	293
Número de <i>edges</i>	1015	1092
Clustering (média)	0.443230432	0.363778475
Grau de clustering	0.048070092	0.02552714
shortest path length (média)	2.441676533	2.652812193
Diameter	6	6
no. of leaves	16	32
Bridges	11	15
Brokers	11	15

Os resultados detalhados dos genes de cada rede, suas medidas e classe podem ser vistos nas **Tabelas 5 e Tabela 6**, dos **Anexos**. Genes *broquers* representam genes que se conectam com vários outros genes que não se conectariam se aquele gene *broquer* não estivesse ali, assim naturalmente podem ser considerados como genes fundamentais para a manutenção da topologia e estabilidade da mesma.

Os *broquers* da rede do Grupo 1 são: *HDAC1*, *HSPA8*, *HSPA8*, *BRCA1*, *STUB1*, *HSPA1A*, *HSP90AA1*, *HSPA4*, *SRC*, *TP53*, *UBC* e *RELA*. Podemos notar que 4 genes *BRCA1*, *HDAC1*, *RELA* e *UBC* aparecem como *broquers* nas duas redes. Os genes *broquers* da rede do grupo 2 são : ***BRCA1***, ***EP300***, ***HDAC1***, *LNK1*, *MDM2*, *NMI*, *RB1*, *RBL1*, ***RELA***, *SMAD2*, *SMAD3*, *SUMO2*, *THBS1*, ***UBC***, *UBR1*.

Os genes caracterizados como *bridges*, *bottleneck* não apresentaram nenhum *overlap* entre as redes. Vale notar que *SMAD2*, *SMAD3*, *SRC* apesar de não serem caracterizados como *broquers* nas duas redes, foram genes que apareceram em ambas. Dos 26 genes compartilhados pelas duas redes, 10 deles apresentam papéis diferentes (ver em Anexos, Tabelas 5 e Tabela 6).

Outras diferenças quanta a topologia da Rede são: i) Em G1 (Não verbal) o gene *HSPA4* alterado (semente) é broker, e os genes sementes *MSL2* (com variação de novo) e *NOTCH2NL2* são bridges; ii) Em G2 (Verbal) dos 5 genes

alterados (sementes), os genes *LNK1*, *THBS1*, *RBL1*, *NMI* e *UBR1* são brokers (*UBR1* com variação de novo).

Figura 19 – Rede de Interação Proteína-Proteína crescida a partir dos genes sementes selecionados do Grupo 1. Os nós com borda verde representam as proteínas dos genes sementes.

Figura 20 – Rede de Interação Proteína-Proteína crescida a partir dos genes sementes selecionados do Grupo 2. Os nós com borda rosa representam as proteínas dos genes sementes.

5.6 Análise de vias e processos Hiper-representados

A partir do conjunto de genes de cada rede fizemos análises funcionais usando a ferramenta Webgestalt buscando vias, ontologias, fenótipos e grupos de

genes associados com doenças hiper-representados. Todas análises apresentam as seguintes informações do cálculo estatístico: Numero total de genes referência na categoria (C), numero de genes da lista estudada observada na categoria (O), numero de genes esperado na categoria (E), taxa de hiper-representação (R), p-valor do teste estatístico hipergeometrico (rawP), e p-valorajustado para múltiplas testagens

5.6.1 Grupo 1 – Não verbais

A análise do conjunto de genes da rede do Grupo 1 resultou em 13 modulos de IPP hiper-representados, dos quais destacamos a estrutura do módulo 18 (**Figura 21**) a ontologia deste e de 3 outros (**Figura 22**).

Figura 21 – Rede IPP do módulo 758 protein fold hiper-representado com os 18 genes alterados no Grupo 1 (pouco ou não verbais).

Figura 22– Resultado da análise da hiper-representação de genes do Grupo 1, pouco ou não verbais, nos módulos de IPP

A análise de hiper-representação com o banco KEGG resultou em 23 vias canônicas (**Anexos, Tabela 7**).

E finalmente, finalizando com o **Grupo 1**, a análise no banco Human Phenotype Ontology resultou em 5 processos biológicos hiper-representados, na Figura DAG Phenotype.

Figura 23 – Resultado da análise da hiper-representação de genes do Grupo 1, pouco ou não verbais, no banco Human Phenotype Ontology

5.6.2 Grupo 2 – Verbais

A análise do conjunto de genes da rede do Grupo 2, resultou em 10 módulos de IPP hiper-representados, dos quais destacamos a ontologia gênica de 6 deles na **Figura 24**.

Figura 24 – Resultado da análise da hiper-representação de genes do Grupo 2, verbais, nos módulos de IPP

A análise de hiper-representação com o banco KEGG resultou em 21 vias (Anexos, Tabela 8). E a análise de região cromossômica revelou uma hiper-representação no braço longo do cromossomo 17 (Tabela 18)

Tabela 18 – Regiões cromossômicas com hiper-representação dos genes do grupo 2, verbais.

Cytogenetic Band	#Gene	Estatística
hsachr17q	18	C=1053;O=18;E=7.15;R=2.52;rawP=0.0004;adjP=0.0014
hsachr17q21	10	C=349;O=10;E=2.37;R=4.22;rawP=0.0002;adjP=0.0014

E finalmente, finalizando com o grupo 2, a análise no banco Human Phenotype Ontology resultou em 19 processos biológicos hiper-representados, representados na **Figura 25**.

Figura 25 – Resultado da análise da hiper-representação de genes do Grupo 2, verbais, no banco Human Phenotype Ontology. As categorias de fenótipos com hiper-representação significativa apresentam o adjP (p-valor ajustado).

6. DISCUSSÃO

Dominios do TEA

O TEA é classificado como transtorno do neurodesenvolvimento, o qual pode ser caracterizado por sintomas neurológicos e psiquiátricos (como déficit cognitivo; distúrbios de comportamento; deficiências sensoriais; convulsões e/ou disfunções neuromotoras; e dificuldades de fala e linguagem comprometendo a comunicação) observado ao longo do desenvolvimento cerebral desde a fase germinativa até a idade adulta (Lee, Lupski 2006). Avanços em neurobiologia, estudos de imagem e genética aumentaram a compreensão da fisiopatologia e da natureza multifatorial dos transtornos de neurodesenvolvimento, tais como epilepsia, síndrome de Tourette, transtorno obsessivo compulsivo (TOC), transtorno de déficit de atenção e hiperatividade (TDAH) e TEA. Além disso, conhecimentos moleculares revelaram potenciais sobreposições dos transtornos que pareciam ser totalmente diferentes, enfatizando a necessidade de compreensão das dimensões dos comportamentos alterados (Geschwind, 2011).

Com o DSM-V os domínios diagnósticos para TEA foram reduzidos de 3 para 2, focando na comunicação e interação social e padrões restritos e repetitivos de comportamento e interesses. Há uma grande variedade de sintomas que não são considerados como “core”, mas que são proeminentes, em uma boa parcela de indivíduos com TEA, como dificuldade atencional, crise convulsiva, alterações de sono, disfunções gastrointestinais (Silver, Rapin, 2012).

Transtornos no desenvolvimento da linguagem tem impacto na socialização e podem ser confundidos com TEA, uma distinção que é particularmente difícil em crianças pré-escolares (Taylor *et al.*, 2014). Contudo, 2 comportamentos são reportados de forma consistente para diferenciar crianças autistas daquelas com prejuízo na linguagem na idade de 20 a 42 meses de idade: apontar para mostrar interesse e uso de gestos convencionais (Bennet *et al.*, 2014).

Da mesma forma, diferenciar atraso no desenvolvimento leve ou moderado de TEA pode ser difícil, particularmente ao avaliar crianças menores (Kantzer *et al.*, 2013). Um estudo identificou alguns itens que diferenciam entre esses dois grupos aos 24 meses, especialmente direcionar a atenção (mostrar) e atenção a voz. Aos 36 meses, 4 itens classificam corretamente todos os sujeitos: uso do corpo de outro, atenção a voz, apontar e maneirismos com os dedos. Dos 38 aos 61 meses, crianças com TEA são mais propensas a mostrar alterações no comportamento não verbal, como contato visual, para regular interação social (Sappok *et al.*, 2013). Mesmo com diversas alterações do DSM-IV para o DSM-V, onde ocorreram mudanças nos critérios para o diagnóstico de TEA, ainda fica clara a importância da fala enquanto instrumento para comunicação social no TEA. Isso justifica a escolha da mesma como endofenótipo para o estudo aqui proposto. Uma das mais importantes modificações foi a junção das alterações da fala com as alterações de sociabilidade, caracterizando um contínuo de comunicação social, e assim, separando em grupo a parte os indivíduos com apenas transtornos de linguagem (Lord, Jones, 2012).

Avanços Genéticos em TEA

Atualmente estudos genéticos e genômicos mostraram que aproximadamente 30% dos casos de TEA podem apresentar alguma causa genética conhecida: (i) ~5-15% dos indivíduos com TEA possuem um transtorno mendeliano identificável, anormalidade cromossômica ou síndrome genética conhecida; (ii) ~5-10% dos casos possuem CNVs (variações no número de cópias) raras *de novo* ou herdadas; e (iii) ~5% dos casos podem apresentar mutações em genes penetrantes raros. Nos 75% dos casos restantes, a contribuição genética é ainda desconhecida, e salvo algumas exceções, o autismo é uma doença complexa, poligênica e multifatorial. (Abrahams, Geschwind, 2008; Pinto *et al.*, 2010; Devlin, Scherer, 2012; Carter, Scherer, 2013).

Desta forma o uso de tecnologias de sequenciamento de alta resolução podem ser utilizadas na compreensão da fisiopatologia do TEA de duas formas diferentes: (1) diante de casos clínicos individuais, para buscar o componente genético que explique o fenótipo; (2) associadas a abordagens de redes gênicas e interações proteicas, para buscar vias e processos biológicos alterados associados a dimensões fenotípicas.

Nosso estudo foi realizado através do exoma de uma família com dois irmãos afetados, buscando causas genéticas que pudessem explicar seus fenótipos compartilhados e individuais, e de trios de TEA esporádico tentando maximizar o achado de variações muito raras e de novo, potencialmente deletérias, com impacto em vias e processos biológicos associados aos fenótipos.

Qualidade do Exoma

Diante de dados de exoma, antes de qualquer discussão biológica é fundamental avaliar a qualidade do sequenciamento, uma vez que este é a base para qualquer outra avaliação. Segundo Heinrich *et al.* (2013), quando o sequenciamento tem cobertura baixa, há um risco de falsos negativos, bem como de falsos positivos, pois é mais provável que um erro de sequenciamento seja classificado como um variante heterozigota, caso a posição tiver cobertura de só alguns reads.

Além disso, as inserções e deleções (Indels) são uma das maiores causas de erros na descoberta de SNPs (Li, Homer, 2010). A taxa de alvos cobertos pelo menos 20x no estudo aqui apresentado está bem próxima das descritas em outros estudos de exoma relacionados ao autismo, que foi 78% (Iossifov *et al.*, 2012) e 83% (Sanders *et al.*, 2012), os quais utilizaram NimbleGen EZ Exome v2,0 como kit de enriquecimento (36Mb, aproximadamente 300.000 exons). A quantidade de *reads* mapeados nos alvos (*targets*) ficou em torno de 65%, esta

taxa não é muito diferente de 63% descrita por Cukier *et al.*, (2014) utilizando o kit SureSelect (38Mb), ou pelo estudo de Chesi *et al.*, (2013) que utiliza o mesmo kit de enriquecimento (TrueSeq) em outra doença ALS (*Amyotrophic lateral sclerosis*).

Quanto a análise do conteúdo GC, ela é realizada pelo fato de ser uma fonte de ineficiência da reação proveniente de regiões com baixo ou alto conteúdo de GC. Segundo Dohm *et al.*, (2008), foi observado que regiões com alto ou baixo conteúdo de CG tendem a ser menos cobertas. Além disso, a maior fonte de viés no conteúdo de CG em estudos de sequenciamento de larga escala é proveniente de etapas iniciais de PCR (Aird *et al.* 2011). No estudo aqui apresentado, foi observado apenas baixa cobertura nas regiões extremas de conteúdo de CG (baixo e alto), o que já foi relatado por Clark *et al.* (2011), sendo o padrão esperado no sequenciamento Illumina.

A taxa de transição e transversão do nosso trabalho foi um pouco abaixo do que é esperado para exomas, que é de aproximadamente 3 para regiões codificadoras de proteína (Chahrour *et al.*, 2012; DePristo *et al.*, 2011; Heinrich *et al.*, 2013). Entretanto o kit de enriquecimento que utilizamos foi o TrueSeq (62Mb), que possui alvos fora das regiões codificadoras de proteína como 5' e 3' UTR, microRNAs e outros ncRNAs, como descritos no manual do fabricante. Logo, a taxa de Ti/Tv resultante da nossa análise foi menor do que a estabelecida, e está próxima da descrita por Guo *et al.* (2012), que fizeram uma comparação de kits de enriquecimento incluindo o TrueSeq, e obtiveram taxa de 2,30.

Exoma de Irmãos com TEA Sindrômico

Diante da possibilidade de explicação do fenótipo associado ao TEA principalmente em casos sindrômicos, familiares, e/ou com deficiência intelectual associado, a sociedade canadense de pediatria preconiza a investigação por

array (aCGH) em todos os casos que apresentem história familiar, TEA com dimorfismos e deficiência intelectual (Duncan, Langlois, 2011; Heil, Schaaf, 2013). Malan *et al.* (2014) também sugerem que o aCGH deve ser considerado como análise genética essencial para pacientes com excesso de crescimento localizado ou generalizado, pois duplicações associadas a síndromes conhecidas com essas condições (ou seja, a síndrome de Sotos ou síndrome de Weaver) ou as não classificadas podem ser causadas por translocações genômicas equilibradas sutis.

Os irmãos aqui apresentados preenchem todos estes critérios e decidimos pela realização do Exoma, pois, além da possibilidade da busca por CNVs, a técnica também possibilitaria uma investigação de alguma outra variação ou um conjunto de alterações raras que poderiam explicar as diferenças fenotípicas relacionadas a dimensão da comunicação social no par de irmãos.

A análise de exoma revelou em ambos pacientes uma duplicação *de novo* na região p16.3 do cromossomo 4. Essas duplicações foram confirmadas pela técnica de MLPA, que também revelou que os irmãos possuíam uma deleção na região p23.3 do cromossomo 8. A duplicação no cromossomo 4 associada com deleção na mesma região do cromossomo 8 aparece descrita anteriormente em estudo realizado por Partington *et al.* (1997). Nesse estudo os afetados apresentavam fenótipos bem similares ao dos nossos pacientes (faces síndrômicas, alta estatura, deficiência intelectual e sintomas autísticos) e uma análise de FISH mostrou que um dos pais desses probandos apresentavam uma translocação equilibrada, que resultou nas translocações desequilibradas encontrada nos filhos. A duplicação do *FGFR3*, gene do receptor do fator de crescimento de fibroblasto presente no 4p16.3 pode ser um gene candidato para o crescimento anormal nesses probandos.

Segundo Feuk *et al.* (2006) translocações nos loci 4p16 e 8p23 são recorrentes, pois apresentam um aglomerado de genes de receptores olfativos polimórficos com duplicações segmentares que mediam translocações, e os pais de indivíduos com essas translocações normalmente são portadores. Esses pais não apresentam fenótipo associado, enquanto que descendentes apresentam

fenótipos que vão desde síndrome de Wolf-Hirschhorn (deleção do 4p) a um espectro mais brando de características dismórficas. Além disso, diversos estudos de CNVs em TEA reportaram deleções ou duplicações em ambas regiões em casos com atraso de linguagem, deficiência intelectual e características dismórficas (Pinto *et al.*, 2010; Kaminsky *et al.*, 2011; Sanders *et al.*, 2011; O'Roak *et al.*, 2012; Schönewolf-Greulich *et al.*, 2013)

Apesar desses achados suportarem que a duplicação do 4p16.3 e deleção do 8p23.3 sejam a causa do quadro de TEA sindrômico apresentado por ambos irmãos, ainda será necessário confirmar pela técnica de FISH se estas são produtos de uma translocação herdada, uma vez que não é possível verificar translocações equilibradas por MLPA.

Em relação as análises de SNVs, foi encontrada apenas uma variação *de novo*, na paciente 2 (do sexo feminino). Essa alteração ocorre no exon 4 do gene *TUBAL3*, que codifica para a tubulina homologa a cadeia alfa 3 (*TUBAL3*, do inglês *Tubulin alfa chain-like 3*), e é considerada deletéria pelo algoritmo Polyphen2 e conservada no GERP++, (promove a alteração do aminoácido glutamato na posição 174 por uma lisina). Segundo banco do Autworks (<http://autworks.hms.harvard.edu/>), site de pesquisa dedicado a achados em autismo (Nelson *et al.*, 2012), *TUBAL3* não aparece em nenhuma rede de doenças, não tem nenhum variante descrito, mas interage com um gene associado a autismo, *GJA1* que codifica para a conexina 43 e está alterada em cérebro de pacientes com autismo (Fatemi *et al.*, 2008).

O paciente 1, do sexo masculino, apresenta um quadro mais grave que a irmã, uma vez que não apresenta nenhum nível funcional de comunicação nem verbal ou gestual. Levando-se em conta a gravidade do quadro do irmão esperaríamos encontrar mais alterações deletérias no mesmo (150 nele, contra 117 encontrado na paciente 2), diferença não significativa (p-valor 0,8109).

Porém, quando levamos em conta as análises de grupo de genes hiper-representados, encontramos resultados que poderiam explicar essa diferença em gravidade. As regiões cromossômicas do paciente 1 apresentam uma hiper-

representação considerável do braço curto do cromossomo 16, especificamente sua região 13 (Taxa R=13.23, p-valor ajustado = 0,0000125). A região 16p13 é classicamente associada com TEA sindrômico, deficiência intelectual, atraso na linguagem, epilepsia, TDAH e diversas outras doenças neuropsiquiátricas (Persico, Napolioni, 2013; Ramalingam *et al.*, 2011; Smalley *et al.*, 2002; Stam *et al.*, 2013; Ullmann *et al.*, 2007; Zara *et al.*, 2000; Ingason *et al.*, 2011). E em doenças do neurodesenvolvimento essa região é reportada como enviesada para o sexo masculino (Tropeano *et al.*, 2013).

A mesma análise em sua irmã revelou uma hiper-representação bem mais discreta do braço curto do cromossomo 1 (Taxa R=3.31, p-valor ajustado =0,04), que em um estudo recente aparece associado com um caso de TEA e dismorfias leves (Piccione *et al.*, 2010), além de aparecer associado a anorexia nervosa, transtorno bipolar juvenil, dislexia (Doyle *et al.*, 2010; Genovese *et al.*, 2013; Grice *et al.*, 2002; Manor *et al.*, 2001; Ragnarsson *et al.*, 1999).

Exoma de trios com TEA esporádico

Sem dúvida não cabe mais a discussão do importante papel de CNVs em TEA, mas diante da sua pequena especificidade fenotípica Merikangas *et al.* (2009) argumentam que as CNVs estejam associadas ao aumento de vulnerabilidade, e que seu efeito em conjunto com outras alterações genéticas, contribui para o desenvolvimento de todas as alterações responsáveis pelo diagnóstico dos TEA. Pesquisas sobre o fatores de risco genético para o TEA, em geral, apresentam pouca especificidade para TEA (Betancur, 2011). Genes de risco parecem conferir vulnerabilidade para uma variedade de transtornos do neurodesenvolvimento (Sahoo *et al.*, 2011; Rapoport *et al.*, 2012).

Um recente estudo de meta-análise com dados de diversos GWAS mostrou que SNPs específicos estão associados com um leque de transtornos psiquiátricos de início precoce e tardio, e que especialmente variações em genes

de canal de cálcio parecem ter um efeito pleiotrópico em psicopatologias (Smoller *et al.*, 2013). Segundo Smoller *et al.* (2013), esses resultados são evidências da importância de superar o modelo de diagnóstico categorial baseado em descrições de sintomas para uma nosologia dimensional e informativa da causa biológica da doença. Dessa forma, atualmente, além das questões nosológicas, para o entendimento de todos os fatores de risco genético para TEA será necessário a identificação do fator que direciona o indivíduo com um gene de risco por uma via do transtorno de neurodesenvolvimento além de identificar apenas um único gene. (Moreno-De-Luca *et al.*, 2013).

Considerando estudos anteriores, era esperado que fosse encontrado CNVs *de novo* em 10% dos pacientes estudados (Sebat *et al.*, 2007) e mutações pontuais *de novo* não-sinônimas em cerca de 50% deles (Neale *et al.*, 2012). Dentro dessa expectativa, encontramos 1 CNV e 6 mutações pontuais *de novo*, em 7 dos 15 pacientes.

Um dos pacientes (trio 30, sexo masculino) apresentou uma deleção de aproximadamente 1,5Mb na região 29 do braço longo do cromossomo 3. Essa região aparece associada com a síndrome da deleção do 3p29 (OMIM#609425), e diversos casos apresentam deleções de tamanhos similares ao que encontramos no nosso estudo. A deleção apresenta 29 genes, incluindo *PAK2* and *DLG1*, que são homólogos autossômicos dos genes do cromossomo X associados a deficiência intelectual (*PAK3* e *DLG3*). Os pacientes podem apresentar transtorno do desenvolvimento, deficiência intelectual e microcefalia (Kaminsky *et al.*, 2011, Miller *et al.*, 2010).

As mutações pontuais *de novo* foram encontradas nos genes *MSL2*, *FGF6*, *PMEL*, *LRRIQ3*, *ADNP2*, *NOSIP*, *UBR1* e *GABBR2*. Destes genes o *GABBR2*, que codifica o receptor 2 do neurotransmissor GABA, apresenta evidência de associação com autismo por aparecer alterado em cerebelo de pacientes com TEA (Fatemi *et al.*, 2009; Oblak *et al.*, 2010). O gene *ADNP2* (*ADNP* homeobox 2) não foi diretamente associado a TEA, mas é um homólogo de *ADNP*, que aparece associado com TEA em meta-análise com dados de exoma (Ben-David,

Shifman, 2013). O desbalanço na expressão tanto de *ADNP* e *ADNP2* parece ter um impacto em esquizofrenia (Dresner *et al.*, 2011).

Vale notar que o *FGF6* mesmo não sendo diretamente ligado a TEA, ao ser analisado no banco autworks (<http://autworks.hms.harvard.edu/>), aparece interagindo com *MET*, gene previamente associado com TEA e fenótipos de comunicação (Campbell *et al.*, 2010). Em ratos *FGF6* atua no desenvolvimento da língua (Han *et al.*, 2012). *MSL2* (*male-specific lethal 2 homolog*) tem papel de ubiquitina ligase e atua na modificação de cromatina (acetilação de cromatina), regulação de transcrição e papel na resposta a dano no DNA (Smith *et al.*, 2005). E *NOSIP* (*nitric oxide synthase interacting protein*) aparece concentrado em sinapses, atua na proliferação celular e interage com *NOS1*, gene previamente associado com TEA em estudo de associação em trios da população coreana (Kim *et al.*, 2009).

Análise de redes

As abordagens integrativas de bioinformática tem sido uma solução para este problema. A união da teoria dos grafos com dados biológicos de bancos públicos, bem como resultados anteriores de estudos genéticos podem ser usados de maneira integrada para indicar funções biológicas e priorizar genes. Para encontrar tais efeitos e interações gênicas, abordagens integrativas e teoria de biologia sistêmica têm sido usadas com sucesso em várias doenças psiquiátricas como transtorno bipolar (Akula *et al.*, 2011), esquizofrenia (Jia *et al.*, 2012) e autismo (Lee *et al.*, 2012).

As diversas proteínas e outras macromoléculas são mediadoras de interações físicas que possibilitam o funcionamento da maquinaria celular, as interações entre proteínas, proteínas e DNA, e proteínas e RNA, (Vidal, 2005). Assim, é essencial considerar as proteínas como redes de interações que atuam

de forma integrada formando um sistema complexo que rege diversas funções celulares (Barabási *et al.*, 2011).

Atualmente, existem bancos de dados que contêm informações sobre interações proteína-proteína (ou PPI, do inglês *protein-protein interactions*), chamados de interatomas, que contêm dados sobre interações físicas entre elas (demonstradas em laboratório e/ou preditas). O conjunto de interações proteína-proteína pode ser modelado como uma rede complexa, ferramenta matemática usada para representar um grupo de elementos que interagem entre si (Bell, Lewitter, 2006). Nas redes as proteínas são representadas pelos nós, e a interação física entre as duplas de proteínas são representadas pelas arestas, que pode ser direcionada (com um ponto de partida e um ponto alvo final) ou não direcionada (interações mútuas) (Albert, 2005; Barabási, Oltvai, 2004).

De acordo com a hipótese da localidade, mostrada por Barabási no artigo “*Network Medicine: a network-based approach to human disease*”, em 2011, as proteínas relacionadas à mesma doença/fenótipo tendem a interagir ou estar próximas na rede. Um corolário desse princípio é que mutações em proteínas que interagem frequentemente levam a doenças/fenótipos parecidos. Além disso, no mesmo artigo por Barabási (2011) foi mostrado o princípio da parcimônia, que vias moleculares causais frequentemente coincidem com os caminhos mínimos entre componentes sabidamente associados a uma doença. Com base nesses conceitos, vários métodos foram propostos para crescimento da rede (usando medidas topológicas) e descoberta de novos genes relacionados àqueles que são a priori associados à doença em questão.

Usamos as medidas topológicas de centralidade para comparar as redes crescidas a partir dos genes com variações *de novo* e muito raras com alta probabilidade de serem deletérias nos dois grupos fenotípicos aqui estudados. Buscando priorizar genes em cada rede pudemos observar que 4 genes *BRCA1*, *HDAC1*, *RELA* e *UBC* aparecem como *broquers* nas duas redes. Os genes caracterizados como *bridges* e *bottleneck* não apresentaram nenhuma sobreposição entre as redes. Podemos entender que esses genes devem exercer papel fundamental no TEA independentemente das questões da fala.

Exposição a ácido valproico intra-útero, um inibidor de histona deacetilase (*HDAC*), foi proposto como um modelo animal para TEA (Argyropoulos *et al.*, 2013). O gene *BRCA1* nunca foi diretamente associado ao TEA no entanto várias estudos genômicos indicam regiões do cromossomo 17q ou 17p como loci de suscetibilidade para TEA (Risch *et al.*, 1999; Yonan *et al.*, 2003), além de descrição de alterações de quebra no 17p13 terem sido observadas em síndrome de Asperger (Tentler *et al.*, 2003). Um estudo mais recente mostra uma paciente portadora de neurofibromatose, retardo mental e autismo, que apresentava alteração desta região incluindo o gene *BRCA1* (Havlovicova *et al.*, 2007).

UBC é um gene de ubiquitina, mais especificamente a ubiquitina C. Estudos publicados nos últimos anos indicam que alterações do sistema de ubiquitina estão bastante relacionadas a doenças do neurodesenvolvimento, por exemplo, mutações no gene de parkin que codificam para uma ligase de ubiquitina E3 e ubiquitina C hidrolase estão associadas a Parkinson de início precoce. A falta da proteína *UBE3A* impede que circuitos cerebrais da região encefálica envolvida na regulação das percepções sensoriais (neocórtex sensorial) de camundongos se fortaleçam em resposta às experiências repetitivas (Sato *et al.*, 2010; Yashiro *et al.*, 2009). Alterações em *UBE3A* estão associadas a Angelman e TEA, assim como alterações em *HUWE/Mule/ARF-BP* estão associadas a *MRXST* (gene de retardo mental sindrômico ligado ao X tipo Turner do inglês mental retardation syndromic X-linked Turner type) (Kessler, 2013).

NF-kappa-B é um fator de transcrição envolvido em diferentes processos biológicos. Ele fica em uma forma inativa no citoplasma e quando ativado move-se para o núcleo ativando fatores de transcrição. NF-kappa-B é composto por *NFKB1* ou *NFKB2* ligados a *REL*, *RELA*, ou *RELB*. Experimentos de deleção de *RELA* em neurônios o associam a capacidade de regeneração de axônios (Haenold *et al.*, 2014). Recentemente foi demonstrado o papel de NF-Kappa-B durante a diferenciação de células tronco (ES), sendo que a inibição de NF-Kappa-B aumenta a expressão de marcadores de pluripotência. Demonstrou-se que miR-290 mantém a pluripotência e capacidade de auto renovação das ES,

reprimindo a sinalização de NF-Kappa-B. A expressão forçada de p65 causa a perda da pluripotência e promove diferenciação celular (Zhang, Hu, 2012)

Análise de vias e processos biológicos hiper-representadas

Muitos dos genes relacionados aos dois grupos são genes já associados com TEA ou que apresentam interação com genes associados ao TEA relacionados a vias de ubiquitinação e do NFK-B (Glessner, et al., 2009; Sbacchi *et al.*, 2010, Kessler, 2013). No entanto, vias muito parecidas apareceram nas redes específicas de cada grupo. As diferentes análises buscando genes e vias específicas relacionadas com os grupos estudados precisam ser replicadas em amostras maiores. Nos nossos resultados preliminares, podemos perceber que a distinção entre verbais e não verbais talvez não faça sentido, mas possam refletir um conjunto de genes específicos e ou número de alterações relacionadas a gravidade do TEA.

7. CONCLUSÃO

Com esse trabalho foi possível concluir que:

1. Em casos familiares com TEA Sindromico e DI é preconizado o uso de aCGH, e como esperado em caso síndrômico, encontramos duplicações no 4p16.3 e deleções 8p23.3 em ambos os irmãos, alterações previamente encontradas em pacientes com características síndromicas e TEA. Além disso, diferentes mutações pontuais raras hiper-representadas em regiões cromossômicas diferentes em cada irmão (cromossomo 16 no paciente do sexo masculino e cromossomo 1 na paciente feminina) podem estar relacionadas com as diferenças fenotípicas por eles apresentadas.

2. De acordo com os estudos anteriores de CNV e exoma em TEA, foram encontradas alterações *de novo* em 9 dos 15 pacientes: i) 1 CNV (deleção) de 1,5Mb na região 29 do braço longo do cromossomo 3, região previamente associada com síndrome e transtornos do desenvolvimento; ii) Dos 8 genes alterados por mutações pontuais *de novo*, o GABBR2 apresenta evidência de associação com TEA.

3. TEA é uma doença complexa, poligênica e multifatorial, e nosso estudo contribui com alterações em genes, de vias relacionadas a neurogenese e sinaptogenese, independentemente do fenótipo, mas com provável impacto em genes específicos que podem estar relacionados a gravidade da apresentação clínica.

8. REALIZAÇÕES FUTURAS

A fim de melhor separar e esclarecer o papel das variações dos diferentes grupos fenotípicos serão sequenciadas novas amostras para serem adicionadas na análise final. Após validação dos resultados obtidos, serão selecionadas as variações de maior interesse para estudo de validação funcional em modelo biológico.

9. ANEXOS

Tabela 1 - Quantidade e cobertura dos dados gerados antes e depois de filtragem por score *Phred*

¹ Amostra	Total Reads Gerados (x10 ⁶)	Cobertura Bruta (reads por base)	Total Reads filtrados (x10 ⁶)	% de reads filtrados
T04-1	112,5	181	99,4	88,4
T04-2	186,6	301	164,6	88,2
T04-3	160,8	259	142,6	88,7
T07-1	96,0	155	84,8	88,4
T07-2	174,5	281	155,1	88,9
T07-3	114,7	185	101,9	88,8
T09-1	96,3	155	88,1	91,5
T09-2	149,0	240	134,7	90,5
T09-3	91,4	147	83,0	90,8
T16-1	129,0	208	108,4	84,0
T16-2	130,3	210	107,3	82,4
T16-3	189,0	304	155,8	82,5
T17-1	185,9	299	167,9	90,3
T17-2	137,5	221	124,0	90,2
T17-3	127,4	205	114,7	90,0
T20-1	110,7	178	97,0	87,6
T20-2	160,2	258	138,1	86,2
T20-3	112,2	181	98,2	87,5
T22-1	135,1	218	116,6	86,3
T22-2	146,8	237	126,2	86,0
T22-3	148,4	239	127,3	85,8
T23-1	125,8	203	109,2	86,8
T23-2	114,7	185	99,0	86,3
T23-3	142,5	229	123,1	86,4
T24-1	133,0	214	116,1	87,3
T24-2	158,4	255	136,6	86,3
T24-3	147,6	238	127,3	86,3
T30-1	127,9	206	106,0	82,9
T30-2	149,8	241	123,1	82,2
T30-3	85,1	137	71,3	83,9
T31-1	160,9	259	139,1	86,4
T31-2	154,3	248	132,0	85,5
T31-3	166,6	268	143,3	86,0
T33-1	35,0	56	29,7	85,0
T33-2	184,4	297	158,8	86,1
T33-3	162,6	262	139,2	85,6
T37-1	117,9	190	99,5	84,4
T37-2	117,5	189	99,3	84,5
T37-3	141,0	227	118,7	84,1
T43-1	107,1	173	90,7	84,7
T43-2	102,5	165	86,4	84,2
T43-3	120,8	195	102,1	84,5

T50-1	132,5	213	109,1	82,4
T50-2	145,3	234	118,3	81,4
T50-3	141,6	228	116,8	82,5
T53-1	173,7	280	142,6	82,1
T53-2	139,1	224	114,7	82,4
T53-3	125,7	202	102,5	81,5
T54-1	143,7	231	121,9	84,8
T54-2	119,1	192	100,6	84,5
T54-3	153,0	246	129,4	84,5
T58-1	139,0	224	118,8	85,5
T58-2	139,2	224	118,3	85,0
T58-3	98,6	159	83,6	84,8
Média	135,2	218	116,0	85,8

1 O número da amostra representa o trio e o integrante da família, (1) Probando, (2)Mãe,(3)Pai

2 cobertura bruta: quantidade de reads gerados multiplicada pelo tamanho dos reads de acordo com as regiões alvo do kit True-Seq (62 Mb).

Tabela 2 – Contagem e mapeamento de reads por amostra.

Amostra	Total Reads gerados (x10⁶)	Quantidade de Reads finais (x10⁶)	Duplicações (%)	Total Reads Mapeados (x10⁶)	Reads finais mapeados (%)
T04-1	112,5	51,8	47,9	50,9	98,3
T04-2	186,6	92,0	44,1	91,2	99,1
T04-3	160,8	70,5	50,6	69,7	98,8
T07-1	96,0	44,4	47,7	42,7	96,1
T07-2	174,5	89,2	42,5	88,3	99,0
T07-3	114,7	54,1	46,9	53,2	98,4
T09-1	96,3	68,2	22,6	67,7	99,3
T09-2	149,0	104,3	22,6	103,9	99,6
T09-3	91,4	51,5	37,9	50,8	98,7
T16-1	129,0	65,0	40,0	63,8	98,2
T16-2	130,3	72,6	32,4	71,3	98,3
T16-3	189,0	96,6	38,0	94,8	98,2
T17-1	185,9	129,0	23,2	128,2	99,4
T17-2	137,5	97,6	21,3	97,2	99,6
T17-3	127,4	91,1	20,6	90,7	99,5
T20-1	110,7	63,3	34,7	62,7	99,0
T20-2	160,2	97,0	29,8	95,9	98,9
T20-3	112,2	67,7	31,0	66,9	98,8
T22-1	135,1	66,5	42,9	65,9	99,0
T22-2	146,8	70,9	43,8	70,2	99,0
T22-3	148,4	68,3	46,3	67,5	98,8
T23-1	125,8	55,9	48,8	55,2	98,7
T23-2	114,7	54,2	45,2	53,6	98,9
T23-3	142,5	71,7	41,8	71,0	98,9
T24-1	133,0	75,1	35,4	74,3	99,0
T24-2	158,4	93,4	31,6	92,4	98,9
T24-3	147,6	86,2	32,3	85,3	99,0
T30-1	127,9	69,7	34,2	68,5	98,2

T30-2	149,8	81,9	33,5	80,3	98,1
T30-3	85,1	42,3	40,7	41,4	97,8
T31-1	160,9	86,1	38,1	85,3	99,1
T31-2	154,3	79,4	39,8	78,0	98,2
T31-3	166,6	83,1	42,0	82,1	98,8
T33-1	35,0	18,8	36,8	16,7	88,7
T33-2	184,4	103,6	34,8	102,5	99,0
T33-3	162,6	88,8	36,2	87,9	99,0
T37-1	117,9	69,3	30,3	68,6	98,9
T37-2	117,5	63,6	35,9	62,9	98,8
T37-3	141,0	79,8	32,8	78,9	98,9
T43-1	107,1	56,7	37,5	56,1	98,9
T43-2	102,5	57,1	33,9	55,5	97,2
T43-3	120,8	61,5	39,8	60,6	98,6
T50-1	132,5	67,5	38,1	66,8	98,9
T50-2	145,3	74,9	36,6	74,0	98,7
T50-3	141,6	71,0	39,2	70,0	98,6
T53-1	173,7	80,1	43,9	79,1	98,8
T53-2	139,1	73,4	36,0	72,5	98,8
T53-3	125,7	65,4	36,2	64,6	98,7
T54-1	143,7	74,6	38,8	73,8	99,0
T54-2	119,1	61,5	38,9	60,7	98,7
T54-3	153,0	75,9	41,3	75,0	98,7
T58-1	139,0	71,8	39,6	71,1	99,0
T58-2	139,2	67,7	42,8	66,8	98,7
T58-3	98,6	44,7	46,5	43,8	97,9
Média	135,2	72,6	37,5	71,6	98,5

Tabela 3 – Cobertura e porcentagem de reads no alvo por amostra.

Amostra	Cobertura média (reads por alvo)	Regiões alvos cobertas 20X (%)	Numero Total de Reads nos alvos (x10 ⁶)	Reads nos alvos (%)
T04-3	61,9	77,1	47,2	67,8
T04-2	80,2	79,7	60,9	66,9
T04-1	44,4	71,7	33,9	66,6
T07-3	46,8	70,7	35,6	66,9
T07-2	77,3	78,9	58,6	66,4
T07-1	36,8	66,9	28,3	66,4
T09-3	43,7	72,6	33,3	65,6
T09-2	86,5	81,0	65,7	63,2
T09-1	57,6	74,7	43,7	64,5
T16-3	81,6	83,1	62,1	65,5
T16-2	60,7	79,9	46,2	64,7
T16-1	55,2	77,1	42,1	65,9
T17-3	74,6	79,6	56,6	62,4
T17-2	81,3	80,8	61,7	63,5
T17-1	108,9	83,7	83,0	64,7
T20-3	55,4	77,2	42,1	63,0
T20-2	78,5	83,8	59,7	62,2

T20-1	54,5	79,0	41,5	66,1
T22-3	58,0	80,8	44,1	65,4
T22-2	61,6	82,1	46,8	66,7
T22-1	56,1	79,0	42,7	64,8
T23-3	61,2	81,6	46,5	65,5
T23-2	46,1	75,8	35,1	65,4
T23-1	48,3	77,8	36,9	66,9
T24-3	72,3	84,5	55,1	64,6
T24-2	79,7	85,4	60,6	65,6
T24-1	63,9	82,0	48,7	65,6
T30-3	35,5	70,3	27,2	65,6
T30-2	66,4	81,4	50,6	63,0
T30-1	58,5	79,6	44,6	65,1
T31-3	72,3	84,7	55,3	67,4
T31-2	68,5	84,0	52,3	67,1
T31-1	74,1	83,7	56,5	66,2
T33-3	75,0	84,1	57,1	64,9
T33-2	87,9	85,9	66,9	65,2
T33-1	13,6	22,6	10,7	64,0
T37-3	69,4	84,2	53,1	67,2
T37-2	55,6	81,8	42,5	67,6
T37-1	58,6	81,1	44,7	65,1
T43-3	51,0	79,6	39,1	64,5
T43-2	46,6	74,8	35,6	64,2
T43-1	48,1	78,9	36,7	65,5
T50-3	61,8	81,8	47,4	67,6
T50-2	65,0	82,4	49,6	67,1
T50-1	58,0	79,7	44,3	66,3
T53-3	55,6	79,2	42,4	65,7
T53-2	62,9	82,1	48,0	66,2
T53-1	68,0	81,4	52,1	65,8
T54-3	66,0	79,7	50,4	67,2
T54-2	53,3	76,4	40,6	66,9
T54-1	63,8	79,5	48,6	65,8
T58-3	37,5	70,4	28,7	65,6
T58-2	56,8	78,5	43,3	64,9
T58-1	62,8	76,8	47,8	67,2
Média	61,6	78,3	46,9	65,6

Tabela 4 – Sumário das quantidades de variações encontradas e taxas cauculadas.

Amotras	Nº variantes	Nº de posições genômicas	SNVs	INDELS	Hetero- zigoto	Homo- zigoto	Ti	Tv	taxa (Ti/Tv)
T04-1	55889	56227	49774	6453	39056	16833	35114	14606	2,40
T04-2	69699	70240	61550	8690	48903	20796	43307	18164	2,38
T04-3	59934	60370	52949	7421	39542	20392	37137	15744	2,36
T07-1	47554	47763	42269	5494	32065	15489	29755	12470	2,39
T07-2	59552	59957	52550	7407	38643	20909	36993	15496	2,39
T07-3	54712	55026	48424	6602	37645	17067	34172	14193	2,41
T09-1	57641	57878	51864	6014	38724	18917	36781	15035	2,45

T09-2	64700	65011	58044	6967	43107	21593	41176	16808	2,45
T09-3	55801	55964	50636	5328	38473	17328	36014	14580	2,47
T16-1	58985	59303	52525	6778	39165	19820	37035	15439	2,40
T16-2	63508	63830	56897	6933	42979	20529	40190	16650	2,41
T16-3	64864	65319	57265	8054	42293	22571	40436	16769	2,41
T17-1	68582	68943	61173	7770	45947	22635	43533	17563	2,48
T17-2	64037	64284	57747	6537	42685	21352	41064	16626	2,47
T17-3	63050	63273	57134	6139	42019	21031	40677	16395	2,48
T20-1	58829	59142	52397	6745	38594	20235	37018	15329	2,41
T20-2	65468	65870	58005	7865	43120	22348	40958	16988	2,41
T20-2	57991	58286	51743	6543	38223	19768	36610	15089	2,43
T22-1	58345	58688	51908	6780	37613	20732	36634	15222	2,41
T22-2	61214	61578	54608	6970	39487	21727	38534	16022	2,41
T22-3	60486	60838	53825	7013	38811	21675	38070	15702	2,42
T23-1	57159	57431	51001	6430	37612	19547	36249	14700	2,47
T23-2	62555	62871	55722	7149	41484	21071	39553	16106	2,46
T23-3	54155	54407	48351	6056	34656	19499	34240	14070	2,43
T24-1	70223	70635	62666	7969	49388	20835	44276	18306	2,42
T24-2	70159	70585	62566	8019	47488	22671	44237	18244	2,42
T24-3	75159	75634	66909	8725	52284	22875	47254	19567	2,41
T30-1	59882	60167	53211	6956	38908	20974	37707	15455	2,44
T30-2	64148	64468	57017	7451	43048	21100	40399	16563	2,44
T30-3	46131	46254	41746	4508	29522	16609	29748	11962	2,49
T31-1	66460	66938	58764	8174	43827	22633	41434	17254	2,40
T31-2	69394	69795	61573	8222	48216	21178	43622	17871	2,44
T31-3	65313	65785	57948	7837	40952	24361	40836	17040	2,40
T33-1	13500	13519	12298	1221	10366	3134	8520	3764	2,26
T33-2	78418	78948	69607	9341	55172	23246	49345	20149	2,45
T33-3	73569	74068	65463	8605	51373	22196	46415	18955	2,45
T37-1	62629	62972	55861	7111	41391	21238	39533	16262	2,43
T37-2	62219	62533	55884	6649	41642	20577	39546	16272	2,43
T37-3	66719	67101	59551	7550	44170	22549	42372	17103	2,48
T43-1	60076	60388	53757	6631	39858	20218	38025	15671	2,43
T43-2	55582	55799	50259	5540	37642	17940	35752	14453	2,47
T43-3	59776	60064	53641	6423	39187	20589	38001	15578	2,44
T50-1	61501	61817	54985	6832	40987	20514	38940	15986	2,44
T50-2	63134	63470	56370	7100	41285	21849	40011	16303	2,45
T50-3	65023	65362	58277	7085	44137	20886	41398	16816	2,46
T53-1	64829	65240	57386	7854	42870	21959	40565	16765	2,42
T53-2	64136	64456	57326	7130	42604	21532	40562	16695	2,43
T53-3	60803	61127	54414	6713	40785	20018	38551	15810	2,44
T54-1	62455	62888	55376	7512	42125	20330	38993	16322	2,39
T54-2	57527	57860	51127	6733	38551	18976	36061	15006	2,40
T54-3	63206	63627	55837	7790	42705	20501	39271	16504	2,38
T58-1	60262	60674	53259	7415	40707	19555	37532	15672	2,39
T58-2	62306	62652	55445	7207	42592	19714	38999	16392	2,38
T58-3	48421	48612	43467	5145	31428	16993	30749	12677	2,43

TABELA 5– Lista deAtributos de todos os genes da rede IPP do Grupo 1

node	geneID	degree	Bridging centrality	clustering	betweenness	norm_degree	brokering	is_seed	class
ST13	6767	19	0,00041	0,12865	0,03851	0,13103	0,11418	Yes	bottleneck
TAF6	6878	25	0,00033	0,33333	0,02489	0,17241	0,11494	Yes	bottleneck
TAF1	6872	19	0,00055	0,4152	0,02189	0,13103	0,07663	No	bottleneck
TBP	6908	23	0,00044	0,36364	0,02183	0,15862	0,10094	No	bottleneck
SEC23A	10484	13	0,00042	0,11538	0,03409	0,08966	0,07931	Yes	bottleneck
UBQLN4	56893	4	0,01203	0,16667	0,0314	0,02759	0,02299	No	bottleneck_ bridge
HSPA4	3308	89	0,0001	0,08401	0,13381	0,61379	0,56223	Yes	bottleneck_ broker
UBC	7316	145	0,00017	0,06178	0,58552	1	0,93822	No	bottleneck_ broker
HSP90AA1	3320	41	0,00018	0,19756	0,03074	0,28276	0,2269	No	bottleneck_ broker
TP53	7157	43	0,00028	0,16058	0,0493	0,29655	0,24893	No	bottleneck_ broker
SRC	6714	24	0,00043	0,24275	0,02486	0,16552	0,12534	No	bottleneck_ broker
CDKN2A	1029	8	0,00303	0,5	0,01814	0,05517	0,02759	No	bridge
CDH5	1003	2	0,00468	0	0,00976	0,01379	0,01379	No	bridge
IGF2	3481	3	0,00285	0,33333	0,00719	0,02069	0,01379	No	bridge
MSL2	55167	4	0,00231	0,16667	0,01946	0,02759	0,02299	Yes	bridge
TTC1	7265	3	0,00227	0,66667	0,00065	0,02069	0,0069	No	bridge
POLR2G	5436	3	0,00667	0	0,01082	0,02069	0,02069	No	bridge
NOTCH2NL	388677	3	0,00208	0	0,00988	0,02069	0,02069	Yes	bridge
ALB	213	3	0,00177	0,66667	0,00165	0,02069	0,0069	No	bridge
ASF1A	25842	2	0,00197	0	0,00041	0,01379	0,01379	No	bridge
ERCC1	2067	4	0,00207	0,66667	0,00167	0,02759	0,0092	No	bridge
HSPA8	3312	39	0,0001	0,22942	0,01399	0,26897	0,20726	No	broker
STUB1	10273	27	0,00008	0,31909	0,00467	0,18621	0,12679	No	broker
RELA	5970	27	0,00025	0,29345	0,01288	0,18621	0,13156	No	broker
HDAC1	3065	28	0,0001	0,29365	0,00797	0,1931	0,1364	No	broker
BRCA1	672	35	0,00015	0,17479	0,01756	0,24138	0,19919	No	broker
HSPA1A	3303	34	0,00012	0,18004	0,01415	0,23448	0,19227	No	broker

Tabela 6 - Lista deAtributos de todos os genes da rede IPP do Grupo 1

node	geneID	degree	Bridging centrality	clustering	betweenness	norm_degree	brokering	is_seed	class
HDAC1	3065	28	0,00008	0,29894	0,00551	0,15217	0,10668	No	broker
SUMO2	6613	24	0,00008	0,15942	0,0052	0,13043	0,10964	No	broker
MDM2	4193	26	0,00008	0,2	0,00651	0,1413	0,11304	No	broker
PAICS	10606	2	0,02183	0	0,00096	0,01087	0,01087	No	bridge
BRCA1	672	39	0,00013	0,24831	0,01593	0,21196	0,15932	No	broker
KCTD13	253980	2	0,02183	0	0,00096	0,01087	0,01087	No	bridge
DHFR	1719	3	0,00741	0,33333	0,00165	0,0163	0,01087	No	bridge
LNK1	84708	60	0,00007	0,00791	0,11246	0,32609	0,32351	Yes	bottleneck_ broker
UBR1	197131	27	0,00009	0,06553	0,02382	0,14674	0,13712	Yes	bottleneck_ broker
PAFAH1B3	5050	3	0,00565	0,33333	0,00151	0,0163	0,01087	No	bridge
FN1	2335	20	0,00051	0,17368	0,03449	0,1087	0,08982	No	bottleneck
CEP72	55722	2	0,02183	0	0,00096	0,01087	0,01087	No	bridge
UBC	7316	184	0,0001	0,03445	0,68303	1	0,96555	No	bottleneck_ broker
TRAF2	7186	19	0,00032	0,12865	0,02492	0,10326	0,08998	No	bottleneck
THBS1	7057	33	0,0001	0,07955	0,03442	0,17935	0,16508	Yes	bottleneck_ broker
RHOC	389	2	0,02183	0	0,00096	0,01087	0,01087	No	bridge
SMAD2	4087	24	0,00015	0,27174	0,0095	0,13043	0,09499	No	broker
SMAD3	4088	27	0,00012	0,26781	0,0084	0,14674	0,10744	No	broker
UGGT1	56886	5	0,00127	0	0,01797	0,02717	0,02717	Yes	bottleneck
SRC	6714	17	0,00058	0,17647	0,01829	0,09239	0,07609	No	bottleneck
RBL1	5933	43	0,0001	0,14064	0,03554	0,2337	0,20083	Yes	bottleneck_ broker
ARVCF	421	2	0,02183	0	0,00096	0,01087	0,01087	No	bridge
EP300	2033	40	0,00016	0,17564	0,0282	0,21739	0,17921	No	bottleneck_ broker
RELA	5970	28	0,00013	0,2037	0,00989	0,15217	0,12118	No	broker
CXADR	1525	2	0,02183	0	0,00096	0,01087	0,01087	No	bridge
LRP5	4041	2	0,01818	0	0,0013	0,01087	0,01087	No	bridge
NMI	9111	32	0,00015	0,09476	0,04743	0,17391	0,15743	Yes	bottleneck_ broker
CUTC	51076	2	0,02183	0	0,00096	0,01087	0,01087	No	bridge
GABBR2	9568	7	0,00073	0,14286	0,02053	0,03804	0,03261	Yes	bottleneck
NSF	4905	3	0,01115	0,33333	0,01165	0,0163	0,01087	No	bridge
RB1	5925	40	0,0001	0,17051	0,02272	0,21739	0,18032	No	bottleneck_ broker
RPIA	22934	2	0,02183	0	0,00096	0,01087	0,01087	No	bridge
AURKB	9212	4	0,00613	0,33333	0,00218	0,02174	0,01449	No	bridge

JAG1	182	2	0,01818	0	0,0013	0,01087	0,01087	No	bridge
LGALS3BP	3959	20	0,00026	0,13158	0,0261	0,1087	0,09439	Yes	bottleneck
NUMB	8650	3	0,00665	0,33333	0,00121	0,0163	0,01087	No	bridge
ATF4	468	8	0,00326	0,25	0,02075	0,04348	0,03261	No	bottleneck

Tabela7 – Resultado da análise da hiper-representação de genes da rede do Grupo 1 (não verbais), em vias e processos biológicos (KEGG)

Vias e Processos Biológicos KEGG	#Gene	Estatística
Pathways in cancer	32	C=326; O=32; E=1.55; R=20.65; rawP=2.40e-32; adjP=5.52e-31
Basal transcription factors	14	C=37; O=14; E=0.18; R=79.60; rawP=1.06e-23; adjP=1.22e-22
Pancreatic cancer	15	C=70; O=15; E=0.33; R=45.08; rawP=4.87e-21; adjP=3.73e-20
Neurotrophin signaling pathway	17	C=127; O=17; E=0.60; R=28.16; rawP=5.55e-20; adjP=3.19e-19
Protein processing in endoplasmic reticulum	17	C=165; O=17; E=0.78; R=21.68; rawP=5.30e-18; adjP=2.44e-17
Chronic myeloid leukemia	13	C=73; O=13; E=0.35; R=37.46; rawP=2.89e-17; adjP=1.11e-16
NOD-like receptor signaling pathway	12	C=58; O=12; E=0.28; R=43.53; rawP=7.08e-17; adjP=2.33e-16
Prostate cancer	13	C=89; O=13; E=0.42; R=30.73; rawP=4.41e-16; adjP=1.27e-15
MAPK signaling pathway	18	C=268; O=18; E=1.27; R=14.13; rawP=1.13e-15; adjP=2.89e-15
Toxoplasmosis	14	C=132; O=14; E=0.63; R=22.31; rawP=3.21e-15; adjP=7.38e-15
Hepatitis C	14	C=134; O=14; E=0.64; R=21.98; rawP=3.98e-15; adjP=8.32e-15
Apoptosis	11	C=87; O=11; E=0.41; R=26.60; rawP=4.36e-13; adjP=8.36e-13
Chemokine signaling pathway	14	C=189; O=14; E=0.90; R=15.58; rawP=4.82e-13; adjP=8.53e-13
Cell cycle	12	C=124; O=12; E=0.59; R=20.36; rawP=9.61e-13; adjP=1.58e-12
Leishmaniasis	10	C=72; O=10; E=0.34; R=29.22; rawP=1.96e-12; adjP=3.01e-12
Chagas disease (American trypanosomiasis)	11	C=104; O=11; E=0.49; R=22.25; rawP=3.24e-12; adjP=4.66e-12
Huntington's disease	13	C=183; O=13; E=0.87; R=14.94; rawP=5.69e-12; adjP=7.70e-12
ErbB signaling pathway	10	C=87; O=10; E=0.41; R=24.18; rawP=1.37e-11; adjP=1.75e-11
Osteoclast differentiation	11	C=128; O=11; E=0.61; R=18.08; rawP=3.20e-11; adjP=3.87e-11
Toll-like receptor signaling pathway	10	C=102; O=10; E=0.48; R=20.63; rawP=6.88e-11; adjP=7.91e-11

Ubiquitin mediated proteolysis	10	C=135; O=10; E=0.64; R=15.58; rawP=1.11e-09; adjP=1.22e-09
Focal adhesion	11	C=200; O=11; E=0.95; R=11.57; rawP=3.79e-09; adjP=3.96e-09
Alzheimer's disease	10	C=167; O=10; E=0.79; R=12.60; rawP=8.73e-09; adjP=8.73e-09

Tabela 8 – Resultado da análise da hiper-representação de genes da rede do Grupo 2, verbais, em vias e processos biológicos(KEGG)

Vias e Processos Biológicos KEGG	#Gene	Estatística
Pathways in cancer	40	C=326; O=40; E=2.21; R=18.06; rawP=7.91e-38; adjP=1.66e-36
Cell cycle	29	C=124; O=29; E=0.84; R=34.42; rawP=3.06e-36; adjP=3.21e-35
Small cell lung cancer	16	C=85; O=16; E=0.58; R=27.71; rawP=7.05e-19; adjP=2.96e-18
Focal adhesion	21	C=200; O=21; E=1.36; R=15.46; rawP=6.74e-19; adjP=2.96e-18
TGF-beta signaling pathway	16	C=84; O=16; E=0.57; R=28.04; rawP=5.76e-19; adjP=2.96e-18
Chronic myeloid leukemia	15	C=73; O=15; E=0.50; R=30.24; rawP=2.16e-18; adjP=7.56e-18
Pancreatic cancer	14	C=70; O=14; E=0.48; R=29.44; rawP=4.49e-17; adjP=1.35e-16
ECM-receptor interaction	14	C=85; O=14; E=0.58; R=24.24; rawP=8.06e-16; adjP=2.12e-15
Amoebiasis	14	C=106; O=14; E=0.72; R=19.44; rawP=1.98e-14; adjP=4.62e-14
MAPK signaling pathway	19	C=268; O=19; E=1.82; R=10.44; rawP=4.61e-14; adjP=9.68e-14
RNA degradation	12	C=71; O=12; E=0.48; R=24.88; rawP=6.91e-14; adjP=1.32e-13
Bladder cancer	10	C=42; O=10; E=0.29; R=35.05; rawP=2.18e-13; adjP=3.81e-13
Prostate cancer	11	C=89; O=11; E=0.60; R=18.19; rawP=2.69e-11; adjP=4.35e-11
Hepatitis C	11	C=134; O=11; E=0.91; R=12.08; rawP=2.31e-09; adjP=3.46e-09
Toxoplasmosis	10	C=132; O=10; E=0.90; R=11.15; rawP=2.70e-08; adjP=3.78e-08
Wnt signaling pathway	10	C=150; O=10; E=1.02; R=9.81; rawP=9.10e-08; adjP=1.19e-07
Phagosome	10	C=153; O=10; E=1.04; R=9.62; rawP=1.10e-07; adjP=1.36e-07
Jak-STAT signaling pathway	10	C=155; O=10; E=1.05; R=9.50; rawP=1.24e-07; adjP=1.45e-07
Regulation of actin cytoskeleton	11	C=213; O=11; E=1.45; R=7.60; rawP=2.77e-07; adjP=3.06e-07
Endocytosis	10	C=201; O=10; E=1.37; R=7.32; rawP=1.36e-06; adjP=1.43e-06
Metabolic pathways	16	C=1130; O=16; E=7.68; R=2.08; rawP=0.0049; adjP=0.0049

9. REFERÊNCIAS BIBLIOGRÁFICAS

- Abe-Sandes K, Silva WA, Zago MA. Heterogeneity of the Y chromosome in Afro-Brazilian populations. *Hum Biol an Int Rec Res*. 2004; 76(1):77–86. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15222681>
- Abrahams BS, Geschwind DH. Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet*. 2008 May; 9(5):341–55. Available from: <http://www.nature.com/nrg/journal/v9/n5/full/nrg2346.html>
- Abrahams BS, Geschwind DH. Connecting genes to brain in the autism spectrum disorders. *Arch Neurol*. 2010 Apr; 67(4):395–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20385903>
- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat. Methods*. 2010 Apr; 7(4):248–9. Available from: <http://www.nature.com/nmeth/journal/v7/n4/full/nmeth0410-248.html>
- Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet*. 2013 Jan; 7(SUPPL.76):Unit7.20. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23315928>
- Aird D, Ross MG, Chen W-S, Danielsson M, Fennell T, Russ C, et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol*. 2011 Jan; 12(2):R18. Available from: <http://genomebiology.com/2011/12/2/R18>
- Akula N, Baranova A, Seto D, Solka J, Nalls MA, Singleton A, et al. A network-based approach to prioritize results from genome-wide association studies. *PLoS One*. 2011;6(9).
- Alarcón M, Abrahams B, Stone J. Linkage, Association, and Gene-Expression Analyses Identify CNTNAP2 as an Autism-Susceptibility Gene. *Am J Hum Genet*. 2008; 82 (January):150–9. Available from: <http://www.sciencedirect.com/science/article/pii/S0002929707000110>
- Albert R. Scale-free networks in cell biology. *J Cell Sci*. 2005 Nov; 118 (Pt 21):4947–57. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16254242>
- Alfarano C, Andrade CE, Anthony K, Bahroos N, Bajec M, Bantoft K, et al. The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Res*. 2005 Jan 1; 33: D418–24. Available from: http://nar.oxfordjournals.org/content/33/suppl_1/D418
- Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet*. Nature Publishing Group; 2011 May; 12(5):363–76. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21358748>
- 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*. 2012 Oct;131(10):1541–54. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22886560>
- Altshuler DM, Gibbs R a, Peltonen L, Altshuler DM, Gibbs R a, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010; 467(7311):52–8. Available from: <http://www.nature.com/nature/journal/v467/n7311/full/nature09298.html>

Amaral DG, Schumann CM, Nordahl CW. Neuroanatomy of autism. *Trends Neurosci.* 2008 Mar; 31(3):137–45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18258309>

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.* 2008; 64(7):577–82. Available from: <http://www.sciencedirect.com/science/article/pii/S006322308005611>

Andrade CRF. 2000. Fluência. In: Andrade CRF, Béfi-Lopes DM, Fernandes FDM, Wertzner HF. ABFW: Teste de linguagem infantil nas áreas de fonologia, vocabulário, fluência e pragmática. Carapicuíba (S.P.): *Pró-Fono* 89: 61–69.

Anney R, Klei L, Pinto D, Regan R, Conroy J, Magalhaes TR, et al. A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet.* 2010 Oct 15;19(20):4072–82. Available from: <http://hmg.oxfordjournals.org/content/19/20/4072>

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 4th ed. Ver: DSM-IV-R. Washington, DC: *American Psychiatric Association*; 2000.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Am. J. Psychiatry. *American Psychiatric Publishing*; 2013.

Argyropoulos A, Gilby KL, Hill-Yardin EL. Studying autism in rodent models: reconciling endophenotypes with comorbidities. *Front Hum Neurosci.* 2013 Jan; 7(July):417. Available from: <http://journal.frontiersin.org/Journal/10.3389/fnhum.2013.00417/full>

Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, Ikeda M, et al. A Common Genetic Variant in the Neurexin Superfamily Member CNTNAP2 Increases Familial Risk of Autism. *Am J Hum Genet.* 2008; 82(1):160–4. Available from: <http://www.sciencedirect.com/science/article/pii/S0002929707000213>

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. *Current Protocols in Bioinformatics.* 2013; 11-10. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/0471250953.bi1110s43/full>

Bakkaloglu B, O’Roak BJ, Louvi A, Gupta AR, Abelson JF, Morgan TM, et al. Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. *Am J Hum Genet.* 2008 Jan; 82(1):165–73. Available from: <http://www.sciencedirect.com/science/article/pii/S0002929707000237>

Bansal V, Libiger O, Torkamani A, Schork NJ. Statistical analysis strategies for association studies involving rare variants. *Nat Rev Genet.* Nature Publishing Group; 2010 Nov; 11(11):773–85. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20940738>

Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet.* 2011 Jan;12(1):56–68. Available from: <http://www.nature.com/nrg/journal/v12/n1/full/nrg2918.html>

Barabási A-L, Oltvai ZN. Network biology: understanding the cell’s functional organization. Sukopp H, Et Al., editors. *Nat Rev Genet.* Nature Publishing Group; 2004;5(2):101–13. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14735121>

Beckmann JS, Estivill X, Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet.* 2007 Aug;8(8):639–46. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17637735>

Befi-Lopes DM, Takiuchi N, Araújo, K. Avaliação da maturidade simbólica nas alterações de desenvolvimento da linguagem. *JBF*, v. 1, n. 3, p. 6-15, 2000

Bell GW, Lewitter F. Visualizing networks. *Methods Enzymol.* 2006; 411(06):408–21. Available from: [http://dx.doi.org/10.1016/S0076-6879\(06\)11022-8](http://dx.doi.org/10.1016/S0076-6879(06)11022-8)

Ben-David E, Shifman S. Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. *Mol Psychiatry.* Nature Publishing Group; 2013; 18(10):1054–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23147383>

Bennett T a, Szatmari P, Georgiades K, Hanna S, Janus M, Georgiades S, et al. Language Impairment and Early Social Competence in Preschoolers with Autism Spectrum Disorders: A Comparison of DSM-5 Profiles. *J. Autism Dev. Disord.* 2014 May 28; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24865586>

Benvenuto A, Moavero R, Alessandrelli R, Manzi B, Curatolo P. Syndromic autism: causes and pathogenetic pathways. *World J Pediatr.* 2009 Aug; 5(3):169–76. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19693459>

Betancur C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res.* Elsevier B.V.; 2011 Mar 22; 1380:42–77. Available from: <http://dx.doi.org/10.1016/j.brainres.2010.11.078>

Betancur C, Sakurai T, Buxbaum JD. The emerging role of synaptic cell-adhesion pathways in the pathogenesis of autism spectrum disorders. *Trends Neurosci.* 2009 Jul; 32(7):402–12. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19541375>

Bishop DVM. Genes, cognition, and communication: insights from neurodevelopmental disorders. *Ann N Y Acad Sci.* 2009 Mar; 1156:1–18. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2805335/pdf/nyas1156-0001.pdf>

Bodfish JW, Symons FJ, Parker DE, Lewis MH. Varieties of repetitive behavior in autism: comparisons to mental retardation. *J Autism Dev Disord.* 2000 Jun; 30(3):237–43. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11055459>

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.* 2008 May;18(5):763–70. Available from: <http://genome.cshlp.org/content/18/5/763>

Brown KR, Jurisica I. Online predicted human interaction database. *Bioinformatics.* 2005; 21(9):2076–82. Available from: <http://bioinformatics.oxfordjournals.org/content/21/9/2076.full>

Cai JJ, Borenstein E, Petrov DA. Broker Genes in Human Disease. *Genome Biol. Evol. Oxford University Press*; 2010; 2(0):815–25. Available from: <http://gbe.oxfordjournals.org/content/2/815>

Callegari-Jacques SM, Grattapaglia D, Salzano FM, Salamoni SP, Crossetti SG, Ferreira ME, et al. Historical genetics: spatiotemporal analysis of the formation of the Brazilian population. *Am J Hum Biol Off J Hum Biol Counc.* 2003; 15(6):824–34. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14595874>

Campbell DB, Warren D, Sutcliffe JS, Lee EB, Levitt P. Association of MET with social and communication phenotypes in individuals with autism spectrum disorder. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 2010;153B(2):438–46.

Capovilla FC, Nunes LR, Capovilla AG, Raphael WD, Gonçalves MJ. Avaliação do desenvolvimento verbal da criança via Teste de Vocabulário por Imagens Peabody. *O Mundo da Saúde.*

Carter MT, Scherer SW. Autism spectrum disorder in the genetics clinic: a review. *Clin Genet.* 2013 May; 83(5):399–407. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23425232>

Carvalho-Silva DR, Santos FR, Rocha J, Pena SDJ. The phylogeography of Brazilian Y-chromosome lineages. *Am J Hum Genet.* 2001; 68(1):281–6. Available from: <http://www.science-direct.com/science/article/pii/S0002929707624987>

Chahrour MH, Yu TW, Lim ET, Ataman B, Coulter ME, Hill RS, et al. Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS Genet.* 2012 Jan ; 8(4):e1002635. Available from: <http://dx.doi.org/10.1371/journal.pgen.1002635>

Chesi A, Staahl BT, Jovičić A, Couthouis J, Fasolino M, Raphael AR, et al. Exome sequencing to identify de novo mutations in sporadic ALS trios. *Nat. Neurosci.* 2013;16(7):851–5. Available from: <http://www.nature.com/neuro/journal/v16/n7/full/nn.3412.html>

Christian SL, Brune CW, Sudi J, Kumar R a, Liu S, Karamohamed S, et al. Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. *Biol Psychiatry.* 2008 Jun; 63(12):1111–7. Available from: <http://dx.doi.org/10.1016/j.biopsych.2008.01.009>

Cichon S, Craddock N, Daly M, Faraone S V, Gejman P V, Kelsoe J, et al. Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry.* 2009 May; 166(5):540–56. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19339359>

Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet.* Nature Publishing Group; 2010 Jun; 11(6):415–25. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20479773>

Clark MJ, Chen R, Lam HYK, Karczewski KJ, Chen R, Euskirchen G, et al. Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol.* Nature Publishing Group; 2011 Oct; 29(10):908–14. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21947028>

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. *Nature.* Nature Publishing Group; 2010 Apr; 464(7289):704–12. Available from: <http://dx.doi.org/10.1038/nature08516>

Cooper GM, Goode DL, Ng SB, Sidow A, Bamshad MJ, Shendure J, et al. Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat. Methods.* 2010. p. 250–1. Available from: <http://www.nature.com/nmeth/journal/v7/n4/full/nmeth0410-250.html>

Crossett A, Kent BP, Klei L, Ringquist S, Trucco M, Roeder K, et al. Using ancestry matching to combine family-based and unrelated samples for genome-wide association studies. *Stat Med.* 2010; 29(28):2932–45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20862653>

Cukier HN, Dueker ND, Slifer SH, Lee JM, Whitehead PL, Lalanne E, et al. Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and

neuropsychiatric disorders. *Mol. Autism*. 2014; 5(1):1. Available from: [http://www. Molecularautism.com/content/5/1/1](http://www.Molecularautism.com/content/5/1/1)

D'Antonio M, D'Onorio De Meo P, Paoletti D, Elmi B, Pallocca M, Sanna N, et al. WEP: a high-performance analysis pipeline for whole-exome data. *BMC Bioinformatics*. BioMed Central Ltd; 2013 Jan. Available from: <http://www.biomedcentral.com/1471-2105/14/S7/S11>

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*. 2011 May; 43(5):491–8. Available from: <http://www.nature.com/ng/journal/v43/n5/abs/ng.806.html>

Devlin B, Melhem N, Roeder K. Do common variants play a role in risk for autism? Evidence and theoretical musings. *Brain Res*. Elsevier B.V.; 2011 Mar; 1380:78–84. Available from: <http://www.sciencedirect.com/science/article/pii/S0006899310025394>

Devlin B, Roeder K. Genomic control for association studies. *Int Biometric Soc*. 1999; 55(4):997–1004. Available from: <http://dx.plos.org/10.1111/j.0006-341X.1999.00997.x>

Devlin B, Roeder K, Wasserman L. Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol*. 2001 Nov; 60(3):155–66. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11855950>

Devlin B, Scherer SW. Genetic architecture in autism spectrum disorder. *Curr Opin Genet Dev*. Elsevier Ltd; 2012 Jun; 22(3):229–37. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22463983>

Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res*. 2008 Sep; 36(16):e105. Available from: <http://nar.oxfordjournals.org/content/36/16/e105>

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. Available from: <http://dx.plos.org/10.1016/j.jaac.2010.01.008>

Dresner E, Agam G, Gozes I. Activity-dependent neuroprotective protein (ADNP) expression level is correlated with the expression of the sister protein ADNP2: deregulation in schizophrenia. *Eur Neuropsychopharmacol*. Elsevier B.V.; 2011 May; 21(5):355–61. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20598862>

Duchan E, Patel DR. Epidemiology of autism spectrum disorders. *Pediatr Clin North Am*. Elsevier Inc; 2012 Feb; 59(1):27–43, ix–x. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22284791>

Duncan A, Langlois S. Use of array genomic hybridization technology in prenatal diagnosis in Canada. *J. Obstet. Gynaecol. Can*. 2011. p. 1256–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22166281>

Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, Fauchereau F, et al. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat Genet*. 2007 Jan; 39(1):25–7. Available from: <http://www.nature.com/ng/journal/v39/n1/full/ng1933.html>

Eapen V. Genetic basis of autism: is there a way forward? *Curr. Opin. Psychiatry*. 2011 May; 24(3):226–36. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21460645>

Evangelou E, Trikalinos T a, Salanti G, Ioannidis JPA. Family-based versus unrelated case-control designs for genetic associations. *PLoS Genet*. Public Library of Science; 2006 Aug; 2(8):e123. Available from: <http://dx.plos.org/10.1371/journal.pgen.0020123>

Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. *Error probabilities*. *Genome Res*. 1998; 8(3):186–94. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9521922>

Fatemi SH, Folsom TD, Reutiman TJ, Thuras PD. Expression of GABA(B) receptors is altered in brains of subjects with autism. *Cerebellum*. 2009 Mar; 8(1):64–9. Available from: <http://link.springer.com/article/10.1007/s12311-008-0075-3>

Ferreira LB, Mendes-Junior CT, Wiezel CEV, Luizon MR, Simões AL. Genomic ancestry of a sample population from the state of São Paulo, Brazil. *Am J Hum Biol Off J Hum Biol Counc*. Wiley Online Library; 2006; 18(5):702–5. Available from: <http://dx.doi.org/10.1002/ajhb.20474>

Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet*. 2006 Feb; 7(2):85–97. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16418744>

Figueiredo VLM, Pinheiro S, Nascimento E Do. Teste de Inteligência WISC-III: Adaptando para a População Brasileira. *Psicol Esc e Educ*. 1998; 2(2).

Foley AG, Gannon S, Rombach-Mullan N, Prendergast A, Barry C, Cassidy AW, et al. Class I histone deacetylase inhibition ameliorates social cognition and cell adhesion molecule plasticity deficits in a rodent model of autism spectrum disorder. *Neuropharmacology*. 2012 Sep; 63(4):750–60. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22683514>

Frayling TM, Walker M, McCarthy MI, Evans JC, Allen LI, Lynn S, et al. Parent-offspring trios: a resource to facilitate the identification of type 2 diabetes genes. *Diabetes*. 1999 Dec; 48(12):2475–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10580439>

Frazer K a, Ballinger DG, Cox DR, Hinds D a, Stuve LL, Gibbs R a, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007 Oct; 449(7164):851–61. Available from: <http://www.nature.com/nature/journal/v449/n7164/full/nature06258.html>

Freitag CM, Staal W, Klauck SM, Duketis E, Waltes R. Genetics of autistic disorders: review and clinical implications. *Eur Child Adolesc Psychiatry*. 2010; 19(3):169–78. Available from: <http://dx.doi.org/10.1007/2Fs00787-009-0076-x>

Gai X, Xie HM, Perin JC, Takahashi N, Murphy K, Wenocur a S, et al. Rare structural variation of synapse and neurotransmission genes in autism. *Mol Psychiatry*. 2012 Apr; 17(4):402–11. Available from: <http://www.nature.com/mp/journal/v17/n4/full/mp201110a.html>

Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variation. *Nat. Genet*. 2014 Jul 20. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25038753>

Gauthier J, Spiegelman D, Piton A, Lafrenière RG, Laurent S, St-Onge J, et al. Novel de novo SHANK3 mutation in autistic patients. *Am J Med Genet B Neuropsychiatr Genet*. 2009 Apr 5; 150B(3):421–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18615476>

Genovese G, Handsaker RE, Li H, Altemose N, Lindgren AM, Chambert K, et al. Using population admixture to help complete maps of the human genome. *Nat Genet.* Nature Publishing Group; 2013 Apr; 45(4):406–14, 414e1–2. Available from: <http://www.nature.com/ng/journal/v45/n4/full/ng.2565.html>

Geschwind DH. Neurodevelopmental disorders: hope for a new beginning. *Curr Opin Neurol.* 2011 Apr; 24(2):95–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21358544>

Gibson G. Rare and common variants: twenty arguments. *Nat Rev Genet.* Nature Publishing Group; 2011 Feb; 13(2):135–45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22251874>

Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron.* Elsevier Inc.; 2011 Jun 9; 70(5):898–907. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21658583>

Glessner JT, Hakonarson H. Genome-wide association: from confounded to confident. *Neuroscientist.* 2011 Apr; 17(2):174–84. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21252379>

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.* 2009 May 28; 459(7246):569–73. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19404257>

Goldstein DB, Allen A, Keebler J, Margulies EH, Petrou S, Petrovski S, et al. Sequencing studies in human genetics: design and interpretation. *Nature Reviews Genetics.* 2013 June; 14(7):460–470. Available from: <http://www.nature.com/doifinder/10.1038/nrg3455>

Gottesman I, Gould T. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry.* 2003; 160(4):636–45. Available from: <http://dx.doi.org/10.1176/appi.ajp.160.4.636>

Grice DE, Halmi KA, Fichter MM, Strober M, Woodside DB, Treasure JT, et al. Evidence for a susceptibility gene for anorexia nervosa on chromosome 1. *Am J Hum Genet.* 2002;70(3):787–92.

Guerreiro-Junior V, Bisso-Machado R, Marrero A, Hünemeier T, Salzano FM, Bortolini MC. Genetic signatures of parental contribution in black and white populations in Brazil. *Genet Mol Biol.* SciELO Brasil; 2009;32(1):1–11. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3032968/>

Guilmatre A, Dubourg C, Mosca A-L, Legallic S, Goldenberg A, Drouin-Garraud V, et al. Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. *Arch Gen Psychiatry.* 2009 Sep; 66(9):947–56. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19736351>

Guo Y, Long J, He J, Li C-I, Cai Q, Shu X-O, et al. Exome sequencing generates high quality data in non-target regions. *BMC Genomics.* 2012 Jan; 13(1):194. Available from: <http://www.biomedcentral.com/1471-2164/13/194>

Haenold R, Weih F, Herrmann KH, Schmidt KF, Krempler K, Engelmann C, et al. NF- κ B controls axonal regeneration and degeneration through cell-specific balance of RelA and p50 in the adult CNS Development. *J Cell Sci.* 2014 Oct 1; 127(Pt 19):4329. Available from: <http://jcs.biologists.org/cgi/pmidlookup?view=long&pmid=25270873>

Han D, Zhao H, Parada C, Hacia JG, Bringas P, Chai Y. A TGF β -Smad4-Fgf6 signaling cascade controls myogenic differentiation and myoblast fusion during tongue development. *Development*. 2012 May; 139(9):1640–50. Available from: <http://dev.biologists.org/content/139/9/1640.long>

Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. 2012 Sep; 489(7416):391–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22996553>

Havlovicova M, Novotna D, Kocarek E, Novotna K, Bendova S, Petrak B, et al. A girl with neurofibromatosis type 1, atypical autism and mosaic ring chromosome 17. *Am. J. Med. Genet. A*. 2007 Jan 1; 143A(1):76–81. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17163520>

Heil KM, Schaaf CP. The genetics of autism spectrum disorders - a guide for clinicians. *Curr. Psychiatry Rep*. 2013; 15(1):334. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23250815>

Heinrich V, Kamphans T, Stange J, Parkhomchuk D, Hecht J, Dickhaus T, et al. Estimating exome genotyping accuracy by comparing to data from large scale sequencing projects. *Genome Med*. BioMed Central Ltd; 2013; 5(7):69. Available from: <http://genomemedicine.com/content/5/7/69>

Herbert MR, Kenet T. Brain abnormalities in language disorders and in autism. *Pediatr Clin North Am*. 2007 Jun; 54(3):563–83, vii. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17543910>

Hinds D a, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, et al. Whole-genome patterns of common DNA variation in three human populations. *Science*. 2005 Feb; 307(5712):1072–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15718463>

Hodgetts S, Zwaigenbaum L, Nicholas D. Profile and predictors of service needs for families of children with autism spectrum disorders. *Autism*. 2014 Jul 29; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25073749>

Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles R a, Clayton DG, et al. Control of confounding of genetic associations in stratified populations. *Am J Hum Genet*. 2003 Jun; 72(6):1492–504. Available from: <http://dx.doi.org/10.1086/375613>

Holt R, Barnby G, Maestrini E, Bacchelli E, Brocklebank D, Sousa I, et al. Linkage and candidate gene studies of autism spectrum disorders in European populations. *Eur J Hum Genet*. 2010 Sep; 18(9):1013–9. Available from: <http://www.nature.com/ejhg/journal/v18/n9/full/ejhg201069a.html>

Ingason A, Rujescu D, Cichon S, Sigurdsson E, Sigmundsson T, Pietiläinen OPH, et al. Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Mol. Psychiatry*. 2011; 16(1):17–25. Available from: <http://www.nature.com/mp/journal/v16/n1/full/mp2009101a.html>

The International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005 Oct 27; 437(7063):1299–320. Available from: <http://dx.doi.org/10.1038/nature04226>

Ionita-Laza I, Rogers AJ, Lange C, Raby B a, Lee C. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. *Genomics*. Elsevier Inc.; 2009 Jan; 93(1):22–6. Available from: <http://www.sciencedirect.com/science/article/pii/S0888754308002115>

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*. Elsevier Inc.; 2012 Apr 26; 74(2):285–99. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22542183>

Iqbal Z, Vandeweyer G, van der Voet M, Waryah AM, Zahoor MY, Besseling J a, et al. Homozygous and heterozygous disruptions of ANK3: at the crossroads of neurodevelopmental and psychiatric disorders. *Hum. Mol. Genet.* 2013 May 15; 22(10):1960–70. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23390136>

Jia P, Zhao Z. Searching joint association signals in CATIE schizophrenia genome-wide association studies through a refined integrative network approach. *BMC Genomics*. 2012 Jan; 13 Suppl 6(Suppl 6):S15. Available from: <http://www.biomedcentral.com/1471-2164/13/S6/S15>

Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, et al. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genet Med*. 2011 Sep; 13(9):777–84. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3661946/>

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*. 2012; 40:D109–14. Available from: <http://nar.oxfordjournals.org/content/40/D1/D109.full>

Kantzer A-K, Fernell E, Gillberg C, Miniscalco C. Autism in community pre-schoolers: developmental profiles. *Res. Dev. Disabil.* Elsevier Ltd; 2013 Sep; 34(9):2900–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23816626>

Karbowiczek M, Henske EP. The role of tuberlin in cellular differentiation: are B-Raf and MAPK involved? *Ann. N. Y. Acad. Sci.* 2005 Nov; 1059:168–73. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16382052>

Kasari C, Paparella T, Freeman S, Jahromi LB. Language outcome in autism: randomized comparison of joint attention and play interventions. *J Consult Clin Psychol*. 2008 Feb; 76(1):125–37. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18229990>

Kearney J. Advances in epilepsy genetics and genomics. *Epilepsy Curr*. 2012 Jul; 12(4):143–6. Available from: <http://dx.doi.org/10.5698/1535-7511-12.4.143>

Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, et al. Human Protein Reference Database – 2009 update. *Nucleic Acids Res*. 2009 Jan; 37(Database issue):D767–72. Available from: http://nar.oxfordjournals.org/content/37/suppl_1/D767.long

Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, et al. The IntAct molecular interaction database in 2012. *Nucleic Acids Res*. 2012 Jan; 40(Database issue):D841–6. Available from: <http://nar.oxfordjournals.org/content/40/D1/D841.long>

Kessler BM. Ubiquitin - omics reveals novel networks and associations with human disease. *Curr. Opin. Chem. Biol.* 2013 Jan 19; 17(1):59–65. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23339974>

Kiezun A, Garimella K, Do R, Stitzel NO, Neale BM, McLaren PJ, et al. Exome sequencing and the genetic basis of complex traits. *Nat Genet*. Nature Publishing Group; 2012 Jun; 44(6):623–30. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22641211>

Kim H-W, Cho S-C, Kim J-W, Cho IH, Kim SA, Park M, et al. Family-based association study between NOS-I and -IIA polymorphisms and autism spectrum disorders in Korean trios. *Am J Med*

Genet B Neuropsychiatr Genet. 2009 Mar 5; 150B(2):300–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18563708>

Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat. Methods.* 2009 May; 6(4):291–5. Available from: <http://www.nature.com/nmeth/journal/v6/n4/pdf/nmeth.1311.pdf>

Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 2009 Jan; 4(7):1073–81. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19561590>

Kumar R a, KaraMohamed S, Sudi J, Conrad DF, Brune C, Badner J a, et al. Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet.* 2008; 17(4):628–38. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18156158>

Lagemaat LN, Grant SGN. Genome variation and complexity in the autism spectrum. *Neuron.* Elsevier Inc.; 2010 Jul 15; 67(1):8–10. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20624587>

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Chun SA, Janssen M, Gil-Garcia JR, editors. *Nature.* Macmillan Magazines Ltd.; 2001; 409(6822):860–921. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11237011>

Lander ES, Schork NJ. Genetic dissection of complex traits. *Science.* 1994 Sep 30; 265(5181):2037–48 [published erratum appears in *Science* 1994 Oct 21; 266(5184):353]. Available from: <http://www.sciencemag.org/content/265/5181/2037.full.pdf>

Lee C, Scherer SW. The clinical context of copy number variation in the human genome. *Expert Rev Mol Med.* 2010; 12(March):e8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20211047>

Lee JA, Lupski JR. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron.* 2006; 52(1):103–21. Available from: <http://www.sciencedirect.com/science/article/pii/S0896627306007343>

Lee T-L, Raygada MJ, Rennert OM. Integrative gene network analysis provides novel regulatory relationships, genetic contributions and susceptible targets in autism spectrum disorders. *Gene.* 2012 Apr 1; 496(2):88–96. Available from: <http://dx.doi.org/10.1016/j.gene.2012.01.020>

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.* 2006 Oct; 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.* Oxford University Press; 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.* 2009; 25(16):2078–9. Available from: <http://bioinformatics.oxfordjournals.org/content/25/16/2078.long>

- Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. *Brief Bioinform.* 2010; 11(5):473–83. Available from: <http://bib.oxfordjournals.org/content/11/5/473.short>
- Li L, Li Y, Browning SR, Browning BL, Slater AJ, Kong X, et al. Performance of genotype imputation for rare variants identified in exons and flanking regions of genes. *PLoS One.* 2011 Jan; 6(9):e24945. Available from: <http://dx.plos.org/10.1371/journal.pone.0024945>
- Li R, Li Y, Fang X, Yang H, Wang J, Kristiansen K, et al. SNP detection for massively parallel whole-genome resequencing. *Genome Res.* 2009 Jun; 19(6):1124–32. Available from: <http://genome.cshlp.org/content/19/6/1124>
- Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, et al. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* 2012 Jan; 40(Database issue):D857–61. Available from: <http://nar.oxfordjournals.org/content/40/D1/D857.long>
- Lins TC, Vieira RG, Abreu BS, Gentil P, Moreno-Lima R, Oliveira RJ, et al. Genetic heterogeneity of self-reported ancestry groups in an admixed Brazilian population. *J Epidemiol Japan Epidemiol Assoc.* 2011;21(4):240–5. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3899415/>
- Lins TC, Vieira RG, Abreu BS, Grattapaglia D, Pereira RW. Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs. *Am J Hum Biol Off J Hum Biol Counc.* 2010; 22(2):187–92. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19639555>
- Liu X-Q, Paterson AD, Szatmari P. Genome-wide linkage analyses of quantitative and categorical autism subphenotypes. *Biol Psychiatry.* 2008 Oct 1; 64(7):561–70. Available from: <http://www.sciencedirect.com/science/article/pii/S0006322308006860>
- Lo-Castro A, Curatolo P. Epilepsy associated with autism and attention deficit hyperactivity disorder: Is there a genetic link? *Brain Dev.* The Japanese Society of Child Neurology; 2013 May 29; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23726375>
- Lord C, Jones RM. Annual research review: re-thinking the classification of autism spectrum disorders. *J. Child Psychol. Psychiatry.* 2012 May; 53(5):490–509. Available from: <http://online.library.wiley.com/doi/10.1111/j.1469-7610.2012.02547.x/full>
- Ma D, Salyakina D, Jaworski JM, Konidari I, Whitehead PL, Andersen AN, et al. A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Ann Hum Genet.* 2009 May; 73(Pt 3):263–73. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2918410/>
- Malan V, Chevallier S, Soler G, Coubes C, Lacombe D, Pasquier L, et al. Array-based comparative genomic hybridization identifies a high frequency of copy number variations in patients with syndromic overgrowth. *Eur. J. Hum. Genet.* 2010 Feb; 18(2):227–32. Available from: <http://dx.doi.org/10.1038/ejhg.2009.162>
- Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell.* Elsevier Inc.; 2012 Mar 16; 148(6):1223–41. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22424231>
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature.* Nature Publishing Group; 2009; 461(7265):747–53. Available from: <http://dx.doi.org/10.1038/nature08494>

Manor I, Eisenberg J, Tyano S, Sever Y, Cohen H, Ebstein RP, et al. Linkage studies suggest a possible locus for developmental dyslexia on chromosome 1p. *Am J Med Genet Neuropsychiatr Genet.* 2001;105(1):120–9.

Marrero AR, Das Neves Leite FP, De Almeida Carvalho B, Peres LM, Kommers TC, Da Cruz IM, et al. Heterogeneity of the genome ancestry of individuals classified as White in the state of Rio Grande do Sul, Brazil. *Am J Hum Biol.* 2005; 17(4):496–506. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15981186>

Marshall C, Noor A, Vincent J. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet.* 2008; 82(February):477–88. Available from: <http://www.sciencedirect.com/science/article/pii/S0002929707000353>

Marteletto MRF, Pedromônico MRM. Validity of Autism Behavior Checklist (ABC): preliminary study. *Rev. Bras. Psiquiatr.* 2005 Dec; 27(4): 295-301. Available from: <http://dx.doi.org/10.1590/S1516-44462005000400008>.

Maski KP, Jeste SS, Spence SJ. Common neurological co-morbidities in autism spectrum disorders. *Curr Opin Pediatr.* 2011 Dec; 23(6):609–15. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21970828>

Matson JL, Nebel-Schwalm MS. Comorbid psychopathology with autism spectrum disorder in children: an overview. *Res Dev Disabil.* 2007; 28(4):341–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16765022>

Maxwell ME. Family Interview for Genetic Studies (FIGS): a manual for FIGS. *Clinical Neurogenetics Branch*, Intramural Research Program, National Institute of Mental Health, Bethesda, MD; 1992.

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.* 2010 Sep; 20(9):1297–303. Available from: <http://dx.doi.org/10.1101/gr.107524.110>

Menezes CGL, Perissinoto J. Habilidade de atenção compartilhada em sujeitos com transtornos do espectro autístico. *Pró-Fono R. Atual. Cient.* 2008 Dec; 20(4): 273-278. Available from: <http://dx.doi.org/10.1590/S0104-56872008000400012>

Merikangas AK, Corvin AP, Gallagher L. Copy-number variants in neurodevelopmental disorders: promises and challenges. *Trends Genet.* 2009 Dec; 25(12):536–44. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19910074>

Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am. J. Hum. Genet.* 2010 May 14; 86(5):749–64. Available from: <http://www.sciencedirect.com/science/article/pii/S0002929710002089>

Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* Oxford University Press; 1988; 16(3):1215. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC334765/pdf/nar00145-0424.pdf>

Moessner R, Marshall CR, Sutcliffe JS, Skaug J, Pinto D, Vincent J, et al. Contribution of SHANK3 mutations to autism spectrum disorder. *Am J Hum Genet.* 2007 Dec; 81(6):1289–97. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2276348/pdf/AJHGv81p1289.pdf>

Moreno-De-Luca A, Myers SM, Challman TD, Moreno-De-Luca D, Evans DW, Ledbetter DH. Developmental brain dysfunction: revival and expansion of old concepts based on new genetic evidence. *Lancet. Neurol.* 2013 Apr; 12(4):406–14. Available from: <http://www.sciencedirect.com/science/article/pii/S1474442213700115>

Muhle R, Trentacoste S V., Rapin I. The Genetics of Autism. *Pediatrics.* 2004 Apr 30;113(5):e472–e486. Available from: <http://pediatrics.aappublications.org/cgi/doi/10.1542/peds.113.5.e472>

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.* Nature Publishing Group; 2012 May 10; 485(7397):242–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22495311>

Nelson TH, Jung J-Y, Deluca TF, Hinebaugh BK, St Gabriel KC, Wall DP. Autworks: a cross-disease network biology application for Autism and related disorders. *BMC Med. Genomics.* 2012 Jan; 5(1):56. Available from: <http://www.biomedcentral.com/1755-8794/5/56>

Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* 2011 Jun; 12(6):443–51. Available from: <http://www.nature.com/nrg/journal/v12/n6/full/nrg2986.html>

Nord AS, Roeb W, Dickel DE, Walsh T, Kusenda M, O'Connor KL, et al. Reduced transcript expression of genes affected by inherited and de novo CNVs in autism. *Eur J Hum Genet.* 2011 Jun; 19(6):727–31. Available from: <http://dx.doi.org/10.1038/ejhg.2011.24>

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.* Nature Publishing Group; 2011 Jun; 43(6):585–9. Available from: <http://www.nature.com/ng/journal/v43/n6/full/ng.835.html>

O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature.* Nature Publishing Group; 2012 May 10; 485(7397):246–50. Available from: <http://www.nature.com/nature/journal/v485/n7397/full/nature10989.html>

Oblak AL, Gibbs TT, Blatt GJ. Decreased GABA(B) receptors in the cingulate cortex and fusiform gyrus in autism. *J. Neurochem.* 2010 Sep 1; 114(5):1414–23. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/j.1471-4159.2010.06858.x/full>

Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, et al. The MIPS mammalian protein-protein interaction database. *Bioinformatics.* 2005 Mar; 21(6):832–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15531608>

Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, Pena SDJ. Color and genomic ancestry in Brazilians. *Proc Natl Acad Sci U S A.* 2003 Jan 7; 100(1):177–82. Available from: <http://www.pnas.org/content/100/1/177>

Partington MW, Fagan K, Soubjaki V, Turner G. Translocations involving 4p16.3 in three families: deletion causing the Pitt-Rogers-Danks syndrome and duplication resulting in a new overgrowth

syndrome. *J. Med. Genet.* 1997 Sep 1; 34(9):719–28. Available from: <http://jmg.bmj.com/cgi/doi/10.1136/jmg.34.9.719>

Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*. 2012 Jan; 7(2):e30619. Available from: <http://dx.doi.org/10.1371/journal.pone.0030619>

Patnala R, Clements J, Batra J. Candidate gene association studies: a comprehensive guide to useful in silico tools. *BMC Genet.* 2013 Jan;14:39. Available from: <http://www.biomedcentral.com/1471-2156/14/39>

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)*. 2008; 84(6):487–94. Available from: http://www.jped.com.br/conteudo/Ing_resumo.asp?varArtigo=1884&cod=&idSecao=1

Persico AM, Napolioni V. Autism genetics. *Behav Brain Res*. Elsevier B.V.; 2013 Jun; Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0166432813003562>

Piccione M, Antona V, Antona R, Gambino G, Pierluigi M, Malacarne M, et al. Array-CGH defined chromosome 1p duplication in a patient with autism spectrum disorder, mild mental deficiency, and minor dysmorphic features. *Am. J. Med. Genet. A*. 2010 Feb; 152A(2):486–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20101695>

Piggot J, Shirinyan D, Shemmassian S, Vazirian S, Alarcón M. Neural systems approaches to the neurogenetics of autism spectrum disorders. *Neuroscience*. Elsevier Inc.; 2009 Nov 24;164(1):247–56. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19482063>

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 SI. *Nature*. Nature Publishing Group; 2010 Jul 15; 466(7304):368–72. Available from: <http://dx.doi.org/10.1038/nature09146>

Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*. 2012 Nov 1; 28(21):2747–54. Available from: <http://bioinformatics.oxfordjournals.org/content/28/21/2747>

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007 Sep; 81(3):559–75. Available from: <http://dx.doi.org/10.1086/519795>

Ragnarsson G, Eiriksdottir G, Johannsdottir JT, Jonasson JG, Egilsson V, Ingvarsson S. Loss of heterozygosity at chromosome 1p in different solid human tumours: association with survival. *Br J Cancer*. 1999;79(9-10):1468–74. Available from: <http://www.nature.com/bjc/journal/v79/n9/abs/6690234a.html>

Ramalingam A, Zhou X-G, Fiedler SD, Brawner SJ, Joyce JM, Liu H-Y, et al. 16p13.11 duplication is a risk factor for a wide spectrum of neuropsychiatric disorders. *J Hum Genet.* 2011; 56(7):541–4. Available from: <http://www.nature.com/jhg/journal/v56/n7/full/jhg201142a.html>

Rapoport JL, Giedd JN, Gogtay N. Neurodevelopmental model of schizophrenia: update 2012. *Mol. Psychiatry*. 2012 Dec; 17(12):1228–38. Available from: <http://www.nature.com/mp/journal/v17/n12/full/mp201223a.html>

Raychaudhuri S. Mapping rare and common causal alleles for complex human diseases. *Cell*. Elsevier Inc.; 2011 Sep 30; 147(1):57–69. Available from: <http://www.sciencedirect.com/science/article/pii/S0092867411010695>

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature*. 2006 Nov 23; 444(7118):444–54. Available from: <http://www.nature.com/nature/journal/v444/n7118/full/nature05329.html>

Rehm HL. Disease-targeted sequencing: a cornerstone in the clinic. *Nat Rev Genet*. 2013 Apr; 14(4):295–300. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23478348>

Risch N, Spiker D, Lotspeich L, Nouri N, Hinds D, Hallmayer J, et al. A genomic screen of autism: evidence for a multilocus etiology. *Am. J. Hum. Genet*. 1999 Aug; 65(2):493–507. Available from: <http://www.sciencedirect.com/science/article/pii/S0002929707620667>

Rodriguez-Revena L, Mila M, Rosenberg C, Lamb A, Lee C. Structural variation in the human genome: the impact of copy number variants on clinical diagnosis. *Genet Med*. 2007 Sep; 9(9):600–6. Available from: <http://www.nature.com/doi/10.1097/GIM.0b013e318149e1e3>

Ronald A, Butcher LM, Docherty S, Davis OSP, Schalkwyk LC, Craig IW, et al. A genome-wide association study of social and non-social autistic-like traits in the general population using pooled DNA, 500 K SNP microarrays and both community and diagnosed autism replication samples. *Behav Genet*. 2010 Jan; 40(1):31–45. Available from: <http://dx.doi.org/10.1007/s10519-009-9308-6>

Ronald A, Hoekstra R a. Autism spectrum disorders and autistic traits: a decade of new twin studies. *Am J Med Genet B Neuropsychiatr Genet*. 2011 Apr; 156B(3):255–74. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21438136>

Roos J. Genetics of schizophrenia: communicating scientific findings in the clinical setting. *Afr J Psychiatry*. 2011 Jun 21; 14(2):105–11. Available from: <http://www.ajol.info/index.php/ajpsy/article/view/67294>

Ruepp A, Waegel B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, et al. CORUM: the comprehensive resource of mammalian protein complexes – 2009. *Nucleic Acids Res*. 2010 Jan; 38:D497–501. Available from: http://nar.oxfordjournals.org/content/38/suppl_1/D497

Sahoo T, Theisen A, Rosenfeld J a, Lamb AN, Ravnan JB, Schultz R a, et al. Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems. *Genet. Med*. 2011 Oct; 13(10):868–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21792059>

Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res*. 2004 Jan; 32:D449–51. Available from: http://nar.oxfordjournals.org/content/32/suppl_1/D449

Salzano FM. Interethnic variability and admixture in Latin America--social implications. *Rev Biol Trop*. 2004; 52(3):405–15. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17361535>

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*. 2011 Jun 9; 70(5):863–85. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21658581>

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*. 2012 May 10; 485(7397):237–41. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22495306>

Sans M. Admixture studies in Latin America: from the 20th to the 21st century. *Hum Biol an Int Rec Res*. 2000;72(1):155–77. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10721616>

Sappok T, Diefenbacher A, Budczies J, Schade C, Grubich C, Bergmann T, et al. Diagnosing autism in a clinical sample of adults with intellectual disabilities: how useful are the ADOS and the ADI-R? *Res. Dev. Disabil*. 2013 May; 34(5):1642–55. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23475013>

Sato FP, Paula CS, Lowenthal R, Nakano EY, Brunoni D, Schwartzman JS, et al. Instrument to screen cases of pervasive developmental disorder: a preliminary indication of validity. *Rev Bras Psiquiatr*. 2009 Mar; 31(1):30–3. Available from: <http://dx.doi.org/10.1590/S1516-44462009000100008>

Sato M, Stryker MP. Genomic imprinting of experience-dependent cortical plasticity by the ubiquitin ligase gene Ube3a. *Proc. Natl. Acad. Sci. U. S. A*. 2010 Mar 23;107(12):5611–6. Available from: <http://www.pnas.org/content/107/12/5611.short>

Saraiva RA, Moojen SMP, Munarski R. Avaliação da compreensão leitora de textos expositivos para fonoaudiólogos e psicopedagogos. São Paulo: Casa do Psicólogo, 2006.

Schönewolf-Greulich B, Ravn K, Hamborg-Petersen B, Brøndum-Nielsen K, Tümer Z. Segregation of a 4p16.3 duplication with a characteristic appearance, macrocephaly, speech delay and mild intellectual disability in a 3-generation family. *Am J Med Genet A*. 2013 Sep; 161(9):2358–62. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23894085>

Sbacchi S, Acquadro F, Calò I, Calì F, Romano V. Functional annotation of genes overlapping copy number variants in autistic patients: focus on axon pathfinding. *Curr Genomics*. 2010 Apr; 11(2):136–45. Available from: <http://dx.doi.org/10.2174/138920210790886880>

Schork NJ, Murray SS, Frazer K a, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev*. 2009 Jun;19(3):212–9. Available from: <http://dx.doi.org/10.1590/S1516-44462009000100008>

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*. 2007 Apr 20; 316(5823):445–9. Available from: <http://www.sciencemag.org/content/316/5823/445>

Sherry ST. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001 Jan 1; 29(1):308–11. Available from: <http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/29.1.308>

Shihab H a, Gough J, Cooper DN, Stenson PD, Barker GL a, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat*. 2013 Jan; 34(1):57–65. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/humu.22225/full>

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. 2005 Aug; 15(8):1034–50. Available from: <http://genome.cshlp.org/content/15/8/1034.long>

Siepel A, Pollard KS, Haussler D. New Methods for Detecting Lineage-Specific Selection. In: Research in Computational Molecular Biology. *Lect. Notes Comput. Sc.* Springer Berlin Heidelberg; 2006; 3909:190–205. Available from: <http://www.springerlink.com/index/10.1007/11732990>

Silver WG, Rapin I. Neurobiological basis of autism. *Pediatr. Clin. North Am.* 2012 Feb; 59(1):45–61, x. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22284792>

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.* 2008;47(8):921–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18645422>

Smalley SL, Kustanovich V, Minassian SL, Stone JL, Ogdie MN, McGough JJ, et al. Genetic linkage of attention-deficit/hyperactivity disorder on chromosome 16p13, in a region implicated in autism. *Am J Hum Genet.* 2002; 71(4):959–63. Available from: <http://www.sciencedirect.com/science/article/pii/S0002929707603802>

Smith ER, Cayrou C, Huang R, Lane WS, Côté J, Lucchesi JC. A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. *Mol Cell Biol.* 2005 Nov; 25(21):9175–88. Available from: <http://mcb.asm.org/content/25/21/9175.long>

Smoller JW, Craddock N, Kendler K, Lee PH, Neale BM, Nurnberger JI, et al. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet.* 2013 Apr 20; 381(9875):1371–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23453885>

Sparrow SS, Balla DA, Cicchetti DV. Vineland adaptive behavior scales: Interview edition, survey form manual. Circle Pines, MN: American Guidance Service, 1984.

St Pourcain B, Wang K, Glessner JT, Golding J, Steer C, Ring SM, et al. Association between a high-risk autism locus on 5p14 and social communication spectrum phenotypes in the general population. *Am J Psychiatry.* 2010 Nov; 167(11):1364–72. Available from: <http://ajp.psychiatryonline.org/doi/abs/10.1176/appi.ajp.2010.09121789>

Stam AJ, Schothorst PF, Vorstman JA, Staal WG. The genetic overlap of attention deficit hyperactivity disorder and autistic spectrum disorder. *Appl Clin Genet.* 2013; 2:7–13. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23935378>

Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med.* 2010 Jan; 61:437–55. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20059347>

State MW, Levitt P. The conundrums of understanding genetic risks for autism spectrum disorders. *Nat Neurosci. Nature Publishing Group;* 2011 Dec; 14(12):1499–506. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22037497>

Stark C, Breitkreutz B-J, Regulj T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 2006 Jan 1; 34:D535–9. Available from: http://nar.oxfordjournals.org/content/34/suppl_1/D535

Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat Rev Genet. Nature Publishing Group;* 2012 Aug; 13(8):537–51. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22777127>

Taylor LJ, Maybery MT, Grayndler L, Whitehouse AJO. Evidence for distinct cognitive profiles in autism spectrum disorders and specific language impairment. *J. Autism Dev. Disord.* 2014 Jan; 44(1):19–30. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23670577>

Tennessen J a, Bigam AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science.* 2012 Jul 6; 337(6090):64–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22604720>

Tentler D, Johannesson T, Johansson M, Råstam M, Gillberg C, Orsmark C, et al. A candidate region for Asperger syndrome defined by two 17p breakpoints. *Eur. J. Hum. Genet.* 2003 Feb; 11(2):189–95. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12634868>

Travers BG, Adluru N, Ennis C, Tromp DPM, Destiche D, Doran S, et al. Diffusion tensor imaging in autism spectrum disorder: a review. *Autism Res.* 2012 Oct; 5(5):289–313. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3474893/>

Tropeano M, Ahn JW, Dobson RJB, Breen G, Rucker J, Dixit A, et al. Male-biased autosomal effect of 16p13.11 copy number variation in neurodevelopmental disorders. *PLoS One.* 2013 Jan; 8(4):e61365. Available from: <http://dx.doi.org/10.1371/journal.pone.0061365>

Ullmann R, Turner G, Kirchhoff M, Chen W, Tonge B, Rosenberg C, et al. Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum Mutat.* 2007; 28(7):674–82. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/humu.20546/pdf>

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science.* 2001 Feb 16 ; 291(5507):1304–51. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11181995>

Vidal M. Interactome modeling. *FEBS Lett.* 2005; 579(8):1834–8. Available from: <http://www.sciencedirect.com/science/article/pii/S0014579305002334>

Volkmar F, Siegel M, Woodbury-Smith M, King B, McCracken J, State M. Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *J Am Acad Child Adolesc Psychiatry.* Elsevier Inc; 2014 Feb; 53(2):237–57. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24472258>

Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* 2013;1–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23703215>

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010 Sep; 38(16):e164. Available from: <http://nar.oxfordjournals.org/content/38/16/e164>

Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature.* Nature Publishing Group; 2009 May 28; 459(7246):528–33. Available from: <http://www.nature.com/nature/journal/v459/n7246/full/nature07999.html>

Wang Z, Liu X, Yang B-Z, Gelernter J. The role and challenges of exome sequencing in studies of human diseases. *Front. Genet.* 2013 Jan; 4(August):160. Available from: <http://journal.frontiersin.org/Journal/10.3389/fgene.2013.00160/full>

Wegiel J, Kuchna I, Nowicki K, Imaki H, Wegiel J, Marchi E, et al. The neuropathology of autism: defects of neurogenesis and neuronal migration, and dysplastic changes. *Acta Neuropathol.* 2010 Jun; 119(6):755–70. Available from: <http://link.springer.com/article/10.1007/s00401-010-0655-4>

Weiss LA, Arking DE, The Gene Discovery Project of Johns Hopkins the Autism Consortium. A genome-wide linkage and association scan reveals novel loci for autism. *Nature.* 2009; 461(7265):802–8. Available from: <http://www.nature.com/nature/journal/v461/n7265/abs/nature08490.html>

Whitehouse a JO, Bishop DVM, Ang QW, Pennell CE, Fisher SE. CNTNAP2 variants affect early language development in the general population. *Genes Brain Behav.* 2011 Jun; 10(4):451–6. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/j.1601-183X.2011.00684.x/pdf>

Willatt L, Cox J, Barber J, Cabanas ED, Collins A, Donnai D, et al. 3q29 Microdeletion Syndrome: Clinical and Molecular Characterization of a New Syndrome. *Am. J. Hum. Genet.* 2005; 77(1):154–60. Available from: <http://www.sciencedirect.com/science/article/pii/S0002929707609124>

Wing L, Gould J. Severe impairments of social interaction and associated abnormalities in children: epidemiology and classification. *J Autism Dev Disord.* 1979 Mar; 9(1):11–29. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/155684>

Yashiro K, Riday TT, Condon KH, Roberts AC, Bernardo DR, Prakash R, et al. Ube3a is required for experience-dependent maturation of the neocortex. *Nat Neurosci.* 2009 Jun; 12(6):777–83. Available from: <http://www.nature.com/neuro/journal/v12/n6/full/nn.2327.html>

Yonan AL, Alarcón M, Cheng R, Magnusson PKE, Spence SJ, Palmer A a, et al. A Genomewide Screen of 345 Families for Autism-Susceptibility Loci. *Am. J. Hum. Genet.* 2003 Oct; 73(4):886–97. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0002929707636362>

Zara F, Gennaro E, Stabile M, Carbone I, Malacarne M, Majello L, et al. Mapping of a locus for a familial autosomal recessive idiopathic myoclonic epilepsy of infancy to chromosome 16p13. *Am J Hum Genet.* 2000; 66(5):1552–7. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1378007/>

Zhang Y, Hu W. NFκB signaling regulates embryonic and adult neurogenesis. *Front. Biol.* 2012 Aug 12; 7(4):277–91. Available from: <http://link.springer.com/10.1007/s11515-012-1233-z>

Zilbovicius M. Autismo: neuroimagem. *Rev. Bras. Psiquiatr.* 2006; 28(Supl I):21–8. Available from: <http://dx.doi.org/10.1590/S1516-44462006000500004>