Isabela Mayá Wayhs Silva

InDels e CNVs pequenas em pacientes com Transtorno do Espectro Autista

InDels and small CNVs in patients with Autism Spectrum Disorder

São Paulo

2017

Isabela Mayá Wayhs Silva

InDels e CNVs pequenas em pacientes com Transtorno do Espectro Autista

InDels and small CNVs in patients with Autism Spectrum Disorder

Dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo, para obtenção do Título de Mestre em Ciências, na Área de Biologia/Genética.

Orientadora: Maria Rita dos Santos e Passos Bueno

São Paulo

2017

FICHA CATALOGRÁFICA

Silva, Isabela Wayhs
InDels e CNVs pequenas em pacientes com Transtorno do Espectro Autista / Isabela Wayhs Silva; orientadora Maria Rita dos Santos Passos Bueno. -- São Paulo, 2017.

123 f.

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

1. Transtorno do Espectro Autista. 2. Variação no número de cópias. 3. array-CGH. 4. Sequenciamento de nova geração. I. Passos Bueno, Maria Rita dos Santos, orient. II. Título.

Catalogação da Publicação
Serviço de Biblioteca do Instituto de Biociências

Comissão Julgadora:

Prof(a) Dr.(a) :

Prof(a) Dr.(a) :

Prof(a) Dr.(a) :

DEDICATÓRIA

Aos meus pais

EPÍGRAFE

“La utopia se encuentra en el horizonte, sé muy bien que nunca la alcanzaré, que si yo camino diez passos ella se alejará diez passos, cuanto más la busque menos la encontraré porque ella se va alejando a medida que yo me acerco. Entonces ¿ Para que sirve la utopia? Pues para eso, para caminar”

Fernando Birri

AGRADECIMENTOS

À professora Dra. Maria Rita dos Santos e Passos-Bueno pelo acolhimento, ensinamentos e orientações essenciais para o desenvolvimento deste projeto e para a construção do meu conhecimento sobre pesquisa e ciência.

À Eloisa de Sá Moreira pelos ensinamentos e parceria que estiveram presentes do início ao fim do mestrado.

À Carla Rosenberg e sua equipe de laboratório (em especial à Silvia Costa) pela disponibilização de equipamentos e ajuda irrestrita com dúvidas e dificuldades que surgiram ao longo do desenvolvimento deste projeto.

Aos funcionários da pós-graduação do IB-USP por estarem sempre dispostos a ajudar e esclarecer dúvidas.

À professora Dra. Lauren Weiss por ter me recebido em seu laboratório e me auxiliado com uma parte importante do projeto.

Ao “team array” (Eduarda Morgana e Cláudia Ismania) pelas seguradas de mão, conselhos, amizade, pelas ajudas sem fim e pela parceira. Trabalhar com vocês não poderia ter sido melhor e sem vocês ficaria infinitas vezes mais difícil.

Aos amigos *muy* amigos do laboratório 200, que foram amigos e família desde que me mudei para São Paulo para iniciar o mestrado, com um adendo especial as meninas “do autismo” (Danielle Moreira, Tatiana Almeida, Eduarda Morgana, Cláudia Ismania) pela amizade, risadas sem fim e companheirismo.

Aos amigos da vida, Letícia Gabrielle, Letícia Rochael, Isadora Terra, Marina Drehmer, Rafael Kawano, Rafaela Ruthes, Anieli Maraschi por terem atravessado todos esses anos ao meu lado e permanecerem.

À Marcela Fontes pelo amor, pelo companheirismo, pela compreensão e pelos abraços.

Aos meus tios (Ivo e Salete) que foram também pais, abrigo e companhia para os dias bons e ruins.

À minha família (Ivan, Eliane e Dani), origem, motivo e fim de todas as coisas. Sem vocês não teria/seria nada, vocês são tudo.

À vida.

Este trabalho contou com suporte financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; processo nº 2014/13056-7), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

ÍNDICE

RESUMO..... 11

ABSTRACT..... 13

LISTA DE TABELAS..... 15

LISTA DE FIGURAS..... 16

LISTA DE GRÁFICOS..... 18

LISTA DE SIGLAS..... 19

1. INTRODUÇÃO..... 20

1.1 OBJETIVOS..... 21

1.1.1 Objetivo Geral..... 21

1.1.2 Objetivos Específicos..... 21

2. REFERENCIAL TEÓRICO..... 22

2.1 ASPECTOS CLÍNICOS E EPIDEMIOLÓGICOS DO TRANSTORNO DO ESPECTRO AUTISTA..... 22

2.2 CONTRIBUIÇÃO DOS FATORES GENÉTICOS PARA A ETIOLOGIA DO TEA..... 25

2.2.1 Variantes Comuns e o TEA..... 26

2.2.2 Variantes Raras e o TEA..... 26

2.3 VARIAÇÃO NO NÚMERO DE CÓPIAS E O TEA..... 28

2.3.1 Caracterização de CNVs em Pacientes com TEA..... 30

2.3.2 Detecção de CNVs..... 31

3. MATERIAL E MÉTODO..... 33

3.1 AMOSTRA BRASILEIRA..... 33

3.2 AMOSTRA NORTE AMERICANA..... 34

3.3.HIBRIDIZAÇÃO GENÔMICA COMPARATIVA POR MICROARRANJO PARA A DETECÇÃO DAS VARIAÇÕES NO NÚMERO DE CÓPIAS E PEQUENAS INSERÇÕES E DELEÇÕES..... 35

3.3.1 Realização do aCGH..... 37

3.3.2 Análise dos Dados de aCGH..... 39

3.4. MLPA SNP-ARRAY e aCGH COMERCIAL..... 40

3.5 SNP-ARRAY e aCGH COMERCIAL..... 41

3.6. REAÇÃO em CADEIA da POLIMERASE em TEMPO REAL (qPCR)..... 42

3.7 SEQUENCIMENTO do EXOMA por *NEXT GENERATION SEQUENCING* (*NGS*)..... 43

3.7.1 Análise dos Dados de *NGS* para CNVs..... 44

3.7.2 Chamada de CNVs..... 44

3.8 ANÁLISES ESTATÍSTICAS..... 45

4. RESULTADOS 46

4.1 ANÁLISE DOS DADOS DE aCGH 60K..... 46

4.1.1 Validação Preliminar da Lâmina Customizada 46

4.1.2 Análise dos Dados de aCGH dos Primeiros Vinte Pacientes..... 46

4.1.3 Análise dos Dados de aCGH de Noventa e Oito Pacientes..... 48

4.2 ANÁLISE DOS DADOS DE ARRAY CGH 180K..... 52

4.2.1 Validação Preliminar da Lâmina 180K Customizada..... 52

4.2.2 Parâmetros de Controle de Qualidade e Eficiência da Hibridição na Lâmina de aCGH 180K..... 58
4.2.3 Análise dos Dados de aCGH 180K de Sessenta e Três Pacientes. 62
4.3. ANÁLISE DOS DADOS DE *NEXT GENERATION SEQUENCING* PARA DETECÇÃO DE CNVS..... 65
4.4 ANÁLISE POR *NGS* DE CNVS NO GENE *SLC17A6*..... 69
5. DISCUSSÃO..... 73
5.1 aCGH CUSTOMIZADO..... 73
5.1.1 aCGH 60K..... 73
5.2 aCGH 180K..... 77
5.3 SEQUENCIAMENTO DE NOVA GERAÇÃO DE EXOMA..... 82
6. CONCLUSÕES..... 87
7. RECOMENDAÇÕES..... 88
REFERÊNCIAS BIBLIOGRÁFICAS..... 89
ANEXO I..... 105
ANEXO II..... 107
ANEXO III..... 112
ANEXO IV..... 114
ANEXO V..... 116

RESUMO

SILVA, I. M. W. **InDels e CNVs pequenas em pacientes com Transtorno do Espectro Autista**. 2017. 123 f. Dissertação (Mestrado) – Instituto de Biociências, Universidade de São Paulo, São Paulo, 2017.

O Transtorno do Espectro Autista (TEA) é uma doença do neurodesenvolvimento. É caracterizado por déficits significativos e persistentes na comunicação e na interação social e por padrões restritos e repetitivos de comportamento, interesses ou atividades. O TEA é considerado uma doença comum, afetando 1 a cada 68 crianças e com uma proporção de 4,2 meninos afetados para cada menina (.A etiologia do autismo apresenta um forte componente genético. Neste contexto, as metodologias genômicas de larga escala (Sequenciamento de nova geração, microarray) contribuíram para o conhecimento sobre a genética do TEA. No entanto, em aproximadamente 70% dos pacientes, o transtorno permanece com etiologia não identificada. Com base nisso, para o presente trabalho, elaborou-se a hipótese de que pequenas CNVs (entre 1 e 50 Kb), que se encontram abaixo da resolução da maioria dos microarrays comerciais, e cuja detecção ainda apresenta limitações para a sua detecção através de sequenciamento de nova geração, poderiam contribuir para o fenótipo em uma proporção significativa dos casos. Como primeira etapa para abordar essa questão, realizou-se a metodologia de aCGH customizado 60K cobrindo um total de 269 genes candidatos ao TEA, a fim de selecionar CNVs potencialmente patogênicas entre 98 pacientes brasileiros com TEA idiopático. Com esta triagem inicial, a prevalência de CNVs potencialmente patogênicas obtida foi de 9%, com 20% delas caracterizadas como pequenas. A análise subsequente foi realizada com a metodologia de aCGH customizado 180K, o qual cobriu um total de 1527 genes candidatos ao TEA. Um total de 63 pacientes com autismo foram analisados com este novo array. A partir destas hibridações, a prevalência de CNVs potencialmente patogênicas obtida foi de 12,7%, com 62,5 % delas classificadas como pequenas. Esta taxa de detecção é bastante expressiva, particularmente se considerarmos que a amostra de pacientes utilizada foi submetida a uma pré-triagem, com a finalidade de excluir os pacientes com as CNVs mais prevalentes no TEA, nas regiões 15q11-

q13, 16p11.2 e 22q13.3. A última abordagem utilizada neste trabalho foi comparar a detecção de CNVs pela metodologia de aCGH, referência padrão ouro para detecção de CNVs, com a de sequenciamento de nova geração (*NGS)*. Os dados de 9 pacientes obtidos por *NGS* foram analisados através dos softwares NextGene e XHMM. Os softwares, no entanto, apresentaram resultados discrepantes entre si e pouca sobreposição com os dados de aCGH 180K, de 38,9% e 50%, considerando o NextGene e o XHMM respectivamente. Os resultados obtidos sugerem que o aCGH customizado é promissor para a detecção de CNVs pequenas e que essas, por sua vez, podem contribuir para o risco de TEA em pelo menos 6,3 %, dos casos.

Palavras-chave: Transtorno do Espectro Autista, variação no número de cópias, array-CGH, sequenciamento de nova geração.

ABSTRACT

SILVA, I. M. W. **InDels and small CNVs in patients with Autism Spectrum Disorder**. 2017. 123 f. Dissertação (Mestrado) – Instituto de Biociências, Universidade de São Paulo, São Paulo, 2017.

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder. It is characterized by significant and persistent deficits in communication and social interaction, and by restricted and repetitive patterns of behavior, interests or activities. ASD is considered a common disease, affecting 1 in 68 children and a proportion of 4.2 boys affected for each girl.The etiology of autism has a strong genetic component. In this context, genomic methodologies of high-throughput (new generation sequencing, microarray) contributed to the knowledge about the genetics of ASD. However, in approximately 70% of patients with ASD, the disorder remains with unidentified etiology. Therefore, foi this work, it was hypothesized that small CNVs (between 1 and 50 Kb), which are below the resolution of most commercial microarrays and and whose detection still has limitations for its detection detection through NGS, could contribute to the phenotype in a proportion of cases. As a first step to address this hypothesis, it was performed the methodology of custom aCGH 60K, covering a total of 269 ASD candidate genes, in order to select potentially pathogenic CNVs among 98 Brazilian patients with idiopathic ASD. With this initial screening, the prevalence of potentially pathogenic CNVs obtained was 9%, with 20% of them characterized as small. The subsequent analysis was performed using the 180K custom aCGH methodology, which covered a total of 1527 TEA candidate genes. A total of 63 patients with autism were analyzed with this new array. From these hybridizations, the prevalence of potentially pathogenic CNVs obtained was 12.7%, with 62.5% of them classified as small. This detection rate is quite significant, particularly considering that the sample of patients used was pre-screened, in order to exclude patients with the most prevalent CNVs in ASD in the regions 15q11-q13, 16p11.2 and 22q13.3. The last approach used in this work was to compare the detection of CNVs by the methodology of aCGH, gold standard reference for CNVs detection, with the next generation sequencing (NGS).Data from 9 patients obtained by NGS were analyzed using NextGene and XHMM software. The softwares,

however, presented discrepant results among themselves and little overlap with the data of aCGH 180K, of 38.9% and 50%, considering NextGene and XHMM respectively. These results suggest that the customized aCGH represents a promising approach for the detection of small CNVs and that these, in turn, can contribute to the risk of ASD in at least 6,3 % of cases.

Keywords: Autism Spectrum Disorder, copy number variation, array-CGH, next generation sequencing.

LISTA DE TABELAS

TABELA 1. DELEÇÕES E DUPLICAÇÕES DETECTADAS EM 20 PACIENTES COM TEA APÓS ANÁLISE DO aCGH CUSTOMIZADO..... 47

TABELA 2. RESULTADO DA ANÁLISE DO aCGH 60K DE 98 PACIENTES AFETADOS PELO TEA..... 49

TABELA 3. RESULTADO DA ANÁLISE DO aCGH 180K DE 31 PACIENTES AFETADOS PELO TRANSTORNO DO ESPECTRO AUTISTA.... 56

TABELA 4. PARÂMETROS DE CONTROLE DE QUALIDADE DO aCGH..... 59

TABELA 5. RESULTADO DA ANÁLISE DO ACGH 180K DE 63 PACIENTES AFETADOS PELO TRANSTORNO DO ESPECTRO AUTISTA.. 63

TABELA 6. COMPARAÇÃO ENTRE O aCGH E OS DADOS DE *NGS* ANALISADOS COM OS SOFTWARES COM O XHMM E O NEXTGENE PARA 9 PACIENTES COM TEA 112

TABELA 7. RESULTADO DA COMPARAÇÃO DAS CNVS OBTIDAS PELOS SOFTWARES XHMM E NEXTGENE PARA 9 PACIENTES COM TEA..... 114

TABELA 8. RESULTADO DA ANÁLISE DE NGS PARA O GENE *SLC17A6* EM 5084 PACIENTES AFETADOS PELO TEA..... 70

LISTA DE FIGURAS

FIGURA 1. MARCAÇÃO EM DUPLICATA DE UM MESMO PACIENTE. A) EXEMPLO DE COMO É FEITA A HIBRIDAÇÃO DE PACIENTE *VERSUS* PACIENTE, COM CADA UM DELES MARCADO ALTERNADAMENTE DE CY3 E CY5. B) REGIÃO HACHURADA EM VERMELHO CORRESPONDE A REGIÃO NA QUAL O DNA DO PACIENTE ESTÁ MARCADO COM *CY3*, CARACTERIZANDO A ALTERAÇÃO COMO DUPLICAÇÃO, E REGIÃO HACHURADA EM VERDE CORRESPONDE A REGIÃO NA QUAL O DNA DO PACIENTE ESTÁ MARCADO COM *CY5*... 38

FIGURA 2. ELETROFORESE EM GEL DE AGAROSE (2,5 %) MOSTRANDO AS ALTERAÇÕES SELECIONADAS NOS PACIENTES 1722-1, 2452-1, 2697-1, 3569-1, 4209-1, 4922-1 E 5074-1. CTRL+*: CONTROLE; CTRL-*: CONTROLE NEGATIVO (ÁGUA). A FLECHA VERMELHA INDICANDO O FRAGMENTO DE REFERÊNCIA DO PESO MOLECULAR DE 100 PB (INVITROGEN)..... 48

FIGURA 3. REPRESENTAÇÃO DA CNV ENCONTRADA NO GENE *SLC17A6* NO PACIENTE 1 USANDO O SOFTWARE *NEXTGENE*..... 51

FIGURA 4. REPRESENTAÇÃO DE 7 CNVS CONFIRMADAS E UTILIZADAS PARA A VALIDAÇÃO DA LÂMINA CUSTOMIZADA 180K (A) PACIENTE 1665-1, MARCADO COM CY5, COM DUPLICAÇÃO NA REGIÃO 18q21.2, (B) PACIENTE 3555-1, MARCADO COM CY5, COM DELEÇÃO NA REGIÃO 22q13.3, (C) PACIENTE 4181-1, MARCADO COM CY5, COM DUPLICAÇÃO NA REGIÃO 11p14.3, (D) PACIENTE 5240-1, MARCADO COM CY5, COM DELEÇÃO NA REGIÃO 15q11-q13, (E) PACIENTE 4290-1, MARCADO COM CY5, COM DELEÇÃO NA REGIÃO 15q11-q13, (F) PACIENTE 4296-1, COM DELEÇÃO NO CROMOSSOMO 22q13.3 E (G) PACIENTE 6507-1, MARCADO COM CY3, COM DELEÇÃO NO CROMOSSOMO 16p11.2..... 53

FIGURA 5. REPRESENTAÇÃO DE HIBRIDAÇÃO COM BAIXA EFICIÊNCIA EXCLUÍDA DA ANÁLISE. (A) ÁREA DO ARRAY COM *RAIN DROP,* INDICADO POR SETA VERMELHA E (B) PARÂMETROS ALTERADOS DE CONTROLE DE QUALIDADE, CLASSIFICADOS EM SUA MAIORIA COMO "AVALIAR"……. 59

FIGURA 6. ELETROFORESE EM GEL DE AGAROSE (2,5 %) MOSTRANDO 4 AMOSTRAS DISTINTAS DE PACIENTES COM TEA FRAGMENTADAS POR 7 MINUTOS, 7 + 3 MINUTOS, 10 MINUTOS DIRETO E ATRAVÉS DE ENZIMA. A FLECHA VERMELHA INDICA O FRAGMENTO DE REFERÊNCIA DO PESO MOLECULAR DE 100 pb (INVITROGEN). OBS.: EN = FRAGMENTAÇÃO ENZIMÁTICA…… 61

FIGURA 7. REPRESENTAÇÃO DA CNV ENCONTRADA NO PACIENTE 8205-5, NA REGIÃO Xp11.23, USANDO O SOFTWARE NEXTGENE…… 64

FIGURA 8. DUPLICAÇÃO NOS EXONS 9 A 12 DO GENE *SLC17A6*……. 72

LISTA DE GRÁFICOS

GRÁFICO 1. 5.RAZÃO ENTRE O NÚMERO DE CÓPIAS DO GENE *DPP6* E DO *GAPDH* DO PACIENTE 10291-1, EM COMPARAÇÃO COM A MESMA RAZÃO OBTIDA EM UM INDIVÍDUO CONTROLE. A RAZÃO ESPERADA PARA UMA DELEÇÃO É DE 0,5, APROXIMADAMENTE..... 57

GRÁFICO 2. RAZÃO ENTRE O NÚMERO DE CÓPIAS DA DUPLICAÇÃO DETECTADA NO PACIENTE 1389-1, REGIÃO 1p13.3, E DO *GAPDH*, EM COMPARAÇÃO COM A MESMA RAZÃO OBTIDA EM UM INDIVÍDUO CONTROLE. A RAZÃO ESPERADA PARA UMA DUPLICAÇÃO É DE 2 APROXIMADAMENTE..... 58

GRÁFICO 3. BOXPLOT DA QUANTIDADE DE DUPLICAÇÕES E DELEÇÕES DETECTADAS EM CADA UM DOS NOVE PACIENTES ATRAVÉS DO (A) aCGH (B) *XHMM* E (C) *NEXTGENE*..... 67

GRÁFICO 4. RAZÃO ENTRE O NÚMERO DE CÓPIAS DO GENE *SLC17A6* E DO *TERT* DO PACIENTE 1, EM COMPARAÇÃO COM A MESMA RAZÃO OBTIDA PARA SUA MÃE, PAI E IRMÃ..... 71

LISTA DE SIGLAS

aCGH	*Microarray-based comparative genomic hybridization*
BG	*Background*
BWA	Burrows-Wheeler Aligner
CARS	Escala de Avaliação do Autismo na Infância
CDC	*Center for Disease Control and Prevention*
CEGH-CEL	Centro de Pesquisas sobre o Genoma Humano e Células-Tronco
CID	Classificação Internacional de Doenças
CNVRs	Variação do número de cópias
CNVs	Variação no número de cópias
DGV	*Database of Genomic Variants*
DLRS	Derivative *log ratio spread*
DSM	Manual Diagnóstico e Estatístico de Transtornos Mentais
FMUSP	Faculdade de Medicina da Universidade de São Paulo
FoSTeS	*Fork Stalling and Template Switching*
GWAS	*G*enome-wide *association study*
IB-USP	Instituto de Biociências da Universidade de São Paulo
InDels	Pequenas inserções e deleções
LCR	*Low copy repeats*
MLPA	*Multiplex ligation-dependent probe amplification*
NAHR	*Nonalelic Homologous Recombination*
NEHJ	*Non-homologous End Joining*
NGS	Sequenciamento de nova geração
qPCR	Reação em cadeia da polimerase quantitativa
SCQ	Questionário de Comunicação Social
SFARI	*Simons Foundation* Autism *Research Initiative*
SNP-array	*Single nucleotide polymorphism array-based*
SNVs	Variações de nucleotídeo único
TCLE	Termo de Consentimento Livre e Esclarecido
TEA	Transtorno do espectro autista
UNESP	Universidade Estadual de São Paulo
UTR	Regiões não traduzidas

1. INTRODUÇÃO

O Transtorno do Espectro Autista (TEA) é uma doença clinicamente heterogênea, caracterizada por déficits significativos e persistentes na comunicação/interação social e por padrões repetitivos e restritos de comportamento, interesses ou atividades.

Estima-se que 1 a cada 68 crianças apresente o Transtorno do Espectro Autista, de acordo com prevalência estimada nos Estados Unidos (CENTERS FOR DISEASE CONTROL AND PREVENTION – CDC, 2014), o que caracteriza o TEA como uma doença comum.

Devido à alta herdabilidade estimada para o TEA, de aproximadamente 50%, o foco principal de trabalhos envolvendo a etiologia deste transtorno tem sido as causas genéticas (GAUGLER *et al.*, 2014; CHASTE e LEBOYER, 2012).

A arquitetura genética do autismo é complexa, na qual diversas formas de variação genética, diferindo em frequência, modelos de herança, tipo de variação e modelos de ação, contribuem para a sua etiologia (DE RUBEIS e BUXBAUM, 2015).

Os fatores genéticos associados ao TEA correspondem a, aproximadamente, 35 % dos casos de autismo, número este inferior aos valores mais recentes estimados de herdabilidade, que é de aproximadamente 50% (GAUGLER *et al*., 2014; SANDIN *et al*., 2014).

Acredita-se que o efeito da interação gene-ambiente (ainda pouco compreendido quando se trata do TEA), as variações em regiões não codificantes, a contribuição das variantes comuns, a contribuição das variações no número de cópias (CNVs) pequenas e de pequenas inserções e deleções (InDels) ainda subcaracterizadas devido às limitações das tecnologias atuais (MANOLIO *et al*., 2009; MANOLIO, 2010), possam estar correlacionadas a uma parte dos casos de TEA, classificados como casos idiopáticos, ou de etiologia desconhecida.

Nesse contexto, o presente trabalho tem como objetivo detectar CNVs (duplicações e deleções maiores que 1 Kb), principalmente as pequenas, menores

que 50 Kb, potencialmente patogênicas e InDels (pequenas inserções e deleções menores que 1Kb) potencialmente patogênicas em pacientes com TEA idiopático.

1.1 OBJETIVOS

1.1.1 Geral

Avaliar a importância de pequenas CNVs/InDels raras na etiologia do Transtorno do Espectro Autista.

1.1.2 Específicos

- Validar as lâminas customizadas de aCGH 60K e 180K, através de outras metodologias como qPCR e outras plataformas comerciais de array.
- Caso o array customizado seja validado, realizar a análise de um número amostral maior de pacientes, a fim de caracterizar e avaliar a frequência de alterações potencialmente patogênicas grandes e pequenas na amostra.
- Verificar a possibilidade de se utilizar dados de sequenciamento de última geração (*NGS*) para a detecção de CNVs.

2. REFERENCIAL TEÓRICO

2.1 ASPECTOS CLÍNICOS E EPIDEMIOLÓGICOS DO TRANSTORNO DO ESPECTRO AUTISTA

O Transtorno do Espectro Autista (TEA) foi descrito pela primeira vez por Kanner, um psiquiatra austríaco, em 1943, após a observação de onze crianças, do sexo masculino em sua maioria, cujo quadro clínico era distinto de retardo mental por apresentarem isolamento social. No caso, os pacientes obeservados foram caracterizados como apresentando distúrbio autístico de contato afetivo.

Logo após essa observação, outros casos similares no que se refere ao isolamento social foram diagnosticados até se chegar a uma definição mais concreta das características que definem um indivíduo como afetado pelo Transtorno do Espectro Autista (FOLSTEIN, 2006).

Na última edição do Manual Diagnóstico e Estatístico de Transtornos Mentais (DSM-V), publicada em maio de 2013, definiu-se que o TEA é uma doença neuropsiquiátrica caracterizada por déficits significativos e persistentes na comunicação/interação social e por padrões repetitivos e restritos de comportamento, interesses ou atividades.

Os primeiros sintomas do TEA surgem durante o período inicial do desenvolvimento, porém são geralmente diagnosticáveis a partir dos 18 meses (MATSON e KOZLOWSKI, 2011). De acordo com o Manual DSM-V (2013), para que seja diagnosticado o paciente deve apresentar:

a) Três alterações na comunicação/interação social:

1. Falta de reciprocidade social;
2. Problemas com a comunicação não verbal;
3. Déficits no desenvolvimento, manutenção e compreensão das relações interpessoais;

b) Pelo menos duas manifestações de padrões repetitivos e restritos de comportamento, interesses ou atividades entre as quatro seguintes:

1. Movimentos motores, uso de objetos e fala repetitivos e estereotipados;
2. Aderência excessiva a rotinas e padrões comportamentais ritualizados;
3. Interesses fixos e restritos;
4. Hipo ou hiper-reatividade a estímulos sensoriais;

O Manual Diagnóstico e Estatístico de Transtornos Mentais (2013), eliminou as subcategorias Transtorno Autista, Síndrome de Asperger e Transtorno Global do Desenvolvimento Não Especificado, presente nas edições anteriores, e difíceis de serem distinguidas devido à sobreposição clínica existente entre elas. Adotou-se apenas um termo, mais amplo, denominado Transtorno do Espectro Autista, que é classificado em três níveis de gravidade baseado no nível de prejuízo na comunicação/ interação social e nos padrões comportamentais.

O nível 1, ou autismo de alto funcionamento, classifica um indivíduo que apresenta um QI na faixa normal ou superior, porém com habilidades sociais e linguagem inadequada, enquanto o 3, ou autismo de baixo funcionamento, mais grave, caracteriza um paciente muitas vezes não verbal com um grave prejuízo cognitivo (DSM-V; MUHLE *et al*., 2013).

O TEA é clinicamente heterogêneo, ou seja, além dos sintomas típicos, outras comorbidades também podem estar presentes. Assim, 30% a 50% dos pacientes apresentam também Transtorno de Déficit de Atenção e Hiperatividade e/ou ansiedade, 40% a 60% possuem déficit intelectual e 25% a 40% manifestam outras comorbidades como epilepsia, obesidade, problemas gastrointestinais, dismorfias e alergias (FOMBONNE, 1999; AMERICAN PSYCHIATRIC ASSOCIATION, 2000; TUCHMAN *et al.*, 2002; LEYFER *et al.*, 2006; MATSON *et al.*, 2011; DEVLIN e SCHERER, 2012).

Antes considerado raro, o TEA tem sido identificado em um número cada vez maior de indivíduos. Fombonne (1999), compilando vinte e três estudos publicados entre 1989 e 1998, conduzidos em doze países e somando um total de 4 milhões de

indivíduos com idade média de 8,5 anos, obtiveram uma prevalência do TEA de 7.2/10.000. Dez anos depois, novamente Fombonne (2009), com base em outra compilação, desta vez contemplando 43 estudos publicados entre 1966 e 2008, em dezessete países diferentes, encontraram prevalência de TEA na magnitude de 63,7/10.000.

Entre as estimativas mais recentes de prevalência está a do Center *for Disease Control and Prevention* (CDC, 2014), com um valor de 147/10.000, estimado a partir de levantamento realizado em onze comunidades diferentes dos Estados Unidos. Este último dado torna o TEA uma doença comum, tendo sido, inclusive, caracterizado como questão emergente de saúde pública pela Organização Mundial de Saúde (PANG *et al*., 2012).

Por outro lado, a crescente evidência dos casos de TEA não vem acompanhada de unanimidade, há controvérsias que apontam este fato como decorrente, entre outros aspectos, das mudanças no critério de diagnóstico, das novas políticas para educação especial, do aumento da disponibilidade de serviços nessa área e da diminuição da idade de diagnóstico. No entanto, não se pode excluir a possibilidade de ter ocorrido um aumento real na sua incidência (FOMBONNE *et al*., 1999; FOMBONNE *et al*., 2009, MATSON e KOZLOWSKI, 2011).

A prevalência do TEA no Brasil foi estimada em 27.2/10.000, em um estudo conduzido em Atibaia, em 2011, com um grupo amostral de 1470 crianças. Este valor está abaixo do que é encontrado para a população mundial, o que pode estar associado a fatores como: (a) pequena amostra local e (b) pouca conscientização da população e dos profissionais de saúde (PAULA *et al*., 2011).

O TEA é uma doença que atinge, no geral, 4,2 vezes mais homens do que mulheres, diferença esta que pode subir para 7:1 quando se leva em consideração apenas os quadros de autismo de alto funcionamento (DE RUBEIS e BUXBAUM, 2015).

2.2 CONTRIBUIÇÃO DOS FATORES GENÉTICOS PARA A ETIOLOGIA DO TEA

Não há consenso sobre a contribuição relativa dos fatores ambientais, genéticos e epigenéticos para o risco de TEA. A influência dos fatores ambientais, como exposição a teratógenos, complicações perinatais e infecções pré-natais, é difícil de ser mensurada. Neste contexto de dificuldades, os fatores ambientais podem explicar alguns casos do transtorno, associados ou não aos fatores genéticos (MUHLE *et al.*, 2004), que estão em conformidade com indicações de estudos epidemiológicos quanto a sua importância (NEWSCHAFFER *et al*., 2012; KIM e LEVENTHAL, 2015).

Aspectos como alta taxa de concordância encontrada para o TEA entre gêmeos monozigóticos, que varia, geralmente, entre 60 e 90% enquanto que para os dizigóticos não ultrapassa 30% (BAILEY *et al.*, 1995; FOLSTEIN e ROSEN-SHEIDLEY , 2001; BOURGERON, 2015), aumento de características subclínicas como dificuldades na comunicação e socialização, preferência por rotinas e dificuldade de lidar com mudanças, detectadas em parentes de primeiro grau de pacientes com TEA, bem como o risco aumentado de outros casos de TEA na família após um probando, reforçam a importância dos fatores genéticos para a etiologia do autismo (HOLLANDER *et al*., 2012; DEVLIN e SCHERER, 2012; FOLSTEIN e ROSEN-SHEIDLEY, 2001).

Estudos recentes, que extrapolam o modelo de pesquisa com gêmeos, estimam que os fatores genéticos estão associados a aproximadamente 50% dos casos de TEA. Um deles, elaborado por Gaugler *et al.* (2014), considerou variações genéticas raras e comuns para a estimativa da herdabilidade, mostrando um valor de herdabilidade de 52,4%. Outro, elaborado por Sandin *et al*. (2014), utilizando informações sobre diagnósticos de TEA em gêmeos monozigóticos e dizigóticos, irmãos completos e meio irmãos, considerou a contribuição para o TEA de fatores genéticos e ambientais, mostrando uma herdabilidade de 50%.

2.2.1 Variantes Comuns e o TEA

A abordagem doença comum/variante comum, que pressupõe que variantes genéticas mais amplamente distribuídas na população (polimórficas) estão associadas a um aumento no risco para as doenças de prevalência alta, também trouxe avanços no entendimento da genética do TEA. A partir dela iniciaram-se estudos de ligação e de associação (GWAS), os quais, apesar de promissores, apresentaram resultados de baixa replicabilidade e limitações quanto à identificação de regiões/alelos que contribuíssem para o TEA (MOLLOY *et al.*, 2005; SZATMARI *et al.*, 2007; ARKING *et al.*, 2008; ALÁRCON *et al.*, 2008; WANG *et al.*, 2009; GAI *et al.*, 2012; DE RUBEIS e BUXBAUM, 2015).

Devlin e Scherer (2012), com o objetivo de verificar a força de associação entre um suposto alelo de risco e o autismo, estimaram o *odds ratio* de variantes comuns individuais encontradas em alguns estudos de GWAS. Estes autores (*Ibidem*), obtiveram valor de *odds ratio de* 1.2 aproximadamente, o que indica que um indivíduo que carrega umas dessas variantes apresenta um risco 1.2 maior de desenvolver autismo do que um outro que não carrega esta variante.

Em base a esse contexto de conhecimentos se conclui que variantes comuns apresentam individualmente um efeito muito moderado para o risco de TEA, podendo apresentar, por outro lado, um efeito aditivo e contribuir de forma significativa para a susceptibilidade ao autismo em conjunto com as variações raras (BOURGERON, 2015).

2.2.2 Variantes Raras e o TEA

A relação entre as variantes raras e o TEA foi inicialmente verificada através do cariótipo com bandeamento G, a partir do qual verificou-se alterações estruturais microscópicas raras associadas a diversas síndromes, entre as quais algumas com sobreposição clínica com o TEA (EL FISHAWY e STATE, 2010; BOURGERON, 2016).

Alterações estruturais microscópicas são ganhos ou perdas de material genético, inversões, translocações, passíveis de serem vistas através da resolução

de >3Mb que o cariótipo apresenta. Estas alterações estão presentes em até 7,4% dos casos de TEA (EL FISHAWY e STATE, 2010; DEVELIN e SCHERER, 2012).

Entre as síndromes genéticas mais comuns com sobreposição ao TEA, estão: esclerose tuberosa (correspondente a cerca de 1 % dos casos de TEA), síndrome do X-frágil (1-2% dos casos), neurofibromatose (<1% dos casos), Síndrome de Rett (0.5%), entre outras (KUSENDA *et al.*, 2008; MILES *et al.*, 2010; DEVLIN e SCHERER, 2012).

A contribuição de variantes raras para a etiologia do TEA tem seguido duas linhas principais de investigação: análise de variações de nucleotídeo único (SNVs) e de pequenas inserções e deleções (InDels) e análise de variações no número de cópias (CNVs) em todo o genoma ou exoma (SBACCHI *et al.*, 2010; IOSSIFOV *et al*., 2012; NEALE *et al*., 2012; O'ROAK *et al*., 2012; SANDERS *et al.,* 2012).

A análise de variações de nucleotídeo único - SNVs (troca de um único nucleotídeo na sequência de DNA) e de pequenas inserções e deleções - InDels (inserções ou deleções com tamanho inferior a 1Kb), tem sido realizada principalmente com a metodologia de sequenciamento de nova geração-*NGS,* de início apenas com sequenciamento de exoma (IOSSIFOV *et al*., 2012; NEALE *et al*., 2012; O'ROAK *et al*., 2012; SANDERS *et al.,* 2012; IOSSIFOV *et al*., 2014) e mais recentemente também com o de genoma (YUEN *et al*., 2015; TURNER *et al*., 2016).

As informações existentes atualmente sobre SNVs e InDels (<10pb) raras de perda de função e *de novo*, indicam que estas alterações contribuem para cerca de 7-13,8 % dos casos *simplex* de autismo (DE RUBEIS e BUXBAUM, 2015).

Estudos sobre o impacto de SNVs/InDels de perda de função raras e herdadas para o risco de TEA são recentes, no entanto, estas alterações podem significar uma importante influência quando agem de forma dominante, podendo ou não serem penetrantes e/ou com expressividade variável, ou recessiva. Suspeita-se que variantes herdadas raras, assim como variantes comuns, também atuem de forma aditiva para o risco de TEA (DE RUBEIS e BUXBAUM, 2014; DE RUBEIS e BUXBAUM, 2015).

A disponibilidade de metodologias de *high-throughput (microarray, NGS),* com resolução superior à que se encontra, por exemplo, com o cariótipo de banda G*,* permitiu a identificação de variações no número de cópias - CNVs (duplicações ou deleções maiores que 1Kb) associadas ao TEA. As CNVs raras (>100 Kb), *de novo* e algumas herdadas, estão associadas a aproximadamente 10 % dos casos *simplex* de autismo (SEBAT *et al.*, 2007; MARSHALL *et al.*, 2008; PINTO *et al.*, 2010; BETANCUR *et al.*, 2011, LI *et al.*, 2012; DEVLIN e SCHERER, 2012).

As alterações SNVs, InDels, CNVs, classificadas em comuns e raras, herdadas e *de novo,* trouxeram contribuições para a identificação de genes associados ao TEA, conforme mostrado no site do *Simons Foundation Autism Research Initiative (SFARI).* Na última atualização do *SFARI*, feita em setembro de 2016, a lista de genes associados ao TEA exibia 845 genes. Paradoxalmente, de uma maneira geral, nenhum desses genes contribui para mais do que 1% dos casos de TEA (BOURGERON, 2015).

A distribuição funcional dos genes associados ao TEA engloba principalmente as funções moleculares de desenvolvimento neuronal e orientação axonal, sinalização de MAPK e outras sinalizações, modificação da cromatina e regulação da transcrição, tradução e degradação de proteínas e a neurotransmissão (SBACCHI *et al.*, 2009; GLESSNER *et al.*, 2009; SWANWICK *et al.*, 2011).

Esses aspectos, apesar de largamente difundidos entre as funções neurobiológicas, contribuem de maneira conjunta com a excitabilidade neuronal, acarretando em um mesmo fenótipo (GESCHWIND *et al.*, 2008; SBACCHI *et al.*, 2010; SWANWICK *et al.*, 2011).

2.3 VARIAÇÃO NO NÚMERO DE CÓPIAS E O TEA

CNVs são os tipos mais frequentes de variações genômicas estruturais, constituídas por duplicações ou deleções submicroscópicas de segmentos de DNA maiores que 1 Kb, podendo ser *de novo* ou herdadas (MERIKANGAS *et al*., 2009; SBACCHI *et al*., 2010). Estas alterações tendem a ocorrer com maior frequência em sequências de DNA que contêm ou são flanqueadas por regiões com baixo número de cópias (*low copy repeats* – LCR - fragmentos repetitivos de DNA com extensão entre 1 Kb a 400 Kb, que compartilham mais de 90 % de identidade), e que causam

instabilidade genômica, podendo mediar ou estimular a formação de CNVs (FREEMAN *et al.*, 2006; STANKIEWICZ *et al.*, 2010).

As CNVs podem ocorrer durante a meiose nas células germinativas, ou em células somáticas, levando ao mosaicismo[1]. Originam-se a partir de diferentes mecanismos, como: recombinação homóloga não alélica (*Nonalelic Homologous Recombination* – NAHR), junção de extremidades não homólogas (*Non-homologous End Joining* – NHEJ), erros na replicação do DNA (*Fork Stalling and Template Switching* – FoSTeS) e através de retrotransposição (SMITH *et al.*, 2008; STANKIEWICZ e LUPSKI, 2010).

Pode haver uma correlação entre o tamanho da CNV e o mecanismo de sua origem. CNVs grandes são, com uma maior frequência, associadas a NAHR, enquanto as CNVs pequenas tendem a estar associadas a NHEJ ou FoSTes (FREEMAN *et al*., 2006).

As CNVs podem afetar o fenótipo de diferentes maneiras tais como: efeitos de dosagem, devido à perda ou ganho de um gene inteiro, rompimento de sequências codificantes de um dado gene, alteração de regiões regulatórias. O efeito final pode tanto ser alteração da expressão de um ou mais genes quanto à formação de proteínas variantes. No caso das deleções, pode ocorrer a expressão de mutação recessiva em loco(s) correspondente no cromossomo homólogo (SMITH *et al.*, 2008; SWANWIK *et al.*, 2011).

Uma das grandes dificuldades encontradas quando se estuda a relação CNVs *versus* doença, é a averiguação da patogenicidade das CNVs, já que 12% a 15% do genoma humano apresenta variações no número de cópias, o que explica, em grande parte, a variação fenotípica normal encontrada na população (SENER, 2014).

As CNVs *de novo*, no geral, apresentam uma maior probabilidade de patogenicidade, já que são submetidas a uma menor pressão seletiva. No entanto, tendo em mente mecanismos como penetrância incompleta e expressividade variável, recorrentes no TEA, além do fator efeito aditivo, tem se discutido mais sobre a patogenicidade de alterações herdadas para o TEA (SWANWICK e LARSEN, 2011; DE RUBEIS e BUXBAUM, 2015).

[1] Coexistência de células geneticamente diferentes provenientes de um mesmo zigoto

A fim de verificar a patogenicidade das CNVs, é necessário levar em conta a diferença genética entre grupos étnicos, a frequência com que a variação aparece nos indivíduos saudáveis e nos afetados, a localização da alteração, a função dos genes envolvidos na alteração e saber que muitas vezes elas podem estar associadas à penetrância incompleta e/ou expressividade variável no TEA (SWANWICK e LARSEN, 2011).

2.3.1 Caracterização de CNVs em Pacientes com TEA

Algumas regiões genômicas, com variação do número de cópias (CNVRs), flanqueadas, em sua maioria, por regiões LCR, são alvos mais frequentes de diferentes CNVs em pacientes com TEA. As CNVRs para o TEA correspondem às regiões: 1q21.1, 2p16.3, 3p26.3, 6q26, 7q11.22, 15q11-q13, 15q13.3, 16p11.2, 22q11.21 e 22q13.33.

Entre essas CNVRs, as que apresentam maior frequência entre os casos de TEA de diferentes populações, correspondendo a 1-3% e reconhecidas como *hotspot*, são aquelas localizadas no 15q11-q13, 16p11.2 e 22q13.3. No contexto de *hotspot*, foi identificada no Brasil uma prevalência de CNVs de 2,7% (KUSENDA *et al.*, 2008; SWANWICK *et al.*, 2011; MOREIRA *et al*., 2014).

No total, 2163 CNVs loci diferentes já foram descritos como associadas ao TEA no SFARI, de 2016.

Pinto *et al*. (2014) identificaram 84 CNVs patogênicas, em conformidade com indicações do *American College of Medical Genetics,* em 82 (3,3%) de 2446 pacientes com TEA. Desses 82 pacientes, 83% foram classificados como TEA não sindrômico.

Estes autores (*Ibidem*) caracterizaram as 84 CNVs patogênicas como sendo 54 delas (64%) *de novo* (59% deleção e 41% duplicação) e 30 (36%) herdadas. Dentre as alterações herdadas, 20 são de origem materna e 10 de origem paterna.

Estas alterações patogênicas se distribuíram entre os tamanhos <30Kb até >10Mb. As deleções, em média, estão em maior proporção entre as alterações até 5Mb, enquanto na faixa de tamanho de maiores que 5Mb a maior proporção é de

duplicações patogênicas. O tamanho médio das CNVs de novo foi maior do que as herdadas, de 3,1 Mb e 1,4 Mb respectivamente. As CNVs menores que 30Kb representaram a menor porcentagem de CNVs patogências, inferior a 10%, o que se deve, provavelmente, a limitação, neste sentido, da metodologia utilizada, Illumina 1 M array (PINTO *et al*.,2014).

Entre os estudos que caracterizaram alterações pequenas, o do Krumm *et al.* (2013), através de sequenciamento de exoma, mostra uma frequência de cerca de 13% a mais de CNVs herdadas pequenas e raras (<50 Kb) em probandos quando comparados aos seus irmãos, sendo essas alterações transmitidas preferencialmente pela mãe. Extrapolando os dados de literatura existentes a respeito de CNVs, evidencia-se que as alterações pequenas (<20 Kb) sejam muito mais frequentes que as grandes (REDON *et al*., 2006).

Um aumento significativo de pequenas deleções raras (<30 Kb) também foi observado, através de sequenciamento de exoma, em uma amostra de pacientes com TEA quando comparada a um grupo controle, 28% versus 21%, respectivamente (POULTNEY *et al*., 2013). Esses dados fortalecem a ideia de que as pequenas CNVs correspondem a um importante fator de risco ao TEA.

As mulheres com TEA, em geral, apresentam ou CNVs patogênicas maiores ou uma maior quantidade de mutação do que os homens afetados, caracterizando o modelo de efeito de proteção feminina, que sugere que as mulheres precisam de uma carga etiológica maior para manifestar o fenótipo de TEA. Este é um dos fatores que se acredita contribuir para a diferença de prevalência do TEA em homens *versus* mulheres (JACQUEMONT *et al*., 2014; DE RUBEIS e BUXBAUM, 2015).

2.3.2 Detecção de CNVs

A detecção das CNVs pode ser feita por meio do uso de diversas técnicas, entre elas a hibridização genômica comparativa por microarranjo (aCGH), microarranjo de polimorfismos de base única (SNP-array), sequenciamento de última geração.

Atualmente a metodologia que é considerada *gold-standard* para a detecção dessas alterações é o aCGH. No entanto, visto que a metodologia de *NGS* apresenta uma maior cobertura e resolução do genoma, e pode possibilitar a

integração tanto de informações de mutações de base única como de CNV, espera-se que em breve essa seja a tecnologia que venha substituir o aCGH.

O aCGH, apesar de vantajoso em termos de custo e análise dos dados gerados, apresenta algumas limitações que incluem ruído de hibridização, dificuldades na detecção de pontos de interrupção, pouca cobertura do genoma.

A tecnologia de *NGS*, por sua vez, ainda apresenta desafios computacionais para a detecção de CNVs que precisam ser superados – associado ao conteúdo de GC (guanina-citosina), a regiões repetitivas do genoma, ao tamanho reduzido das *reads*, a outras regiões que também amplificam menos na etapa de PCR como regiões reguladoras - estes fatores dificultam a normalização e posteriormente o cálculo estatístico para a chamada de CNVs (ZHAO *et al*., 2013; ALKAN *et al*., 2011; MEI TEO *et al*., 2012).

Neste contexto, a solução imediata mais eficaz para ultrapassar as limitações de cada tecnologia é a integração das diferentes técnicas tais como aCGH e NGS, e a combinação de diferentes softwares para a análise dos dados de sequenciamento, principalmente quando se trata de detecção de CNVs pequenas e InDels, que representam um desafio ainda maior (ZHAO *et al*., 2013; ALKAN *et al*., 2011).

Em 2015 foi lançado pela *Agilent Technologies*® um novo *kit* para sequenciamento, o *OneSeq Target Enrichment*, que foi projetado para detectar e analisar variações de número de cópias, perda de heterozigosidade e variantes de base única em uma única reação. Este *kit* se caracteriza pelo enriquecimento de regiões exônicas de interesse associado a *targets*. Essa conformação do kit permite a detecção de CNVs grandes (>300Kb) que tenham pontos de quebra em regiões intrônicas.

A expectativa é que com o contínuo aperfeiçoamento dessas técnicas seja possível detectar CNVs cada vez menores e InDels (Inserções ou deleções menores que 1 Kb), que por sua vez, permitem a identificação de novos genes de risco de forma mais fácil e direta já que essas alterações incluem, geralmente, um menor número de genes.

3. MATERIAL E MÉTODO

3.1 AMOSTRA BRASILEIRA

No presente trabalho foi utilizada uma amostra de 585 pacientes de origem brasileira com TEA. Entre estes, 178 foram analisados por aCGH 60K e 180K e 407 por MLPA.

Entre os 178 pacientes brasileiros analisados por ambos os arrays, 136 são do sexo masculino (76,4%) e 42 são do sexo feminino (23,6%). Do total de 178 pacientes, 25 foram classificados como sindrômicos (14%). O termo sindrômico aqui é utilizado para designar pacientes nos quais estão presentes, além do quadro de TEA, outros sinais clínicos como carcterísticas dismórficas e/ou malformações congênitas.

Entre os 407 pacientes brasileiros restantes analisados por MLPA especificamente para a triagem de CNVs no gene *SLC17A6*, 322 são do sexo masculino (79,1%) e 85 do sexo feminino (20,9%). Para esta amostra não há informação sobre porcentagem de sindrômicos e não sindrômicos.

Estes 585 pacientes foram selecionados a partir do banco de DNA de pacientes com TEA do Centro de Pesquisas sobre o Genoma Humano e Células-Tronco (CEGH-CEL, Instituto de Biociências da Universidade de São Paulo - IBUSP). Os pacientes atendidos no CEGH-CEL foram em sua grande maioria diagnosticados com TEA por psiquiatras do Instituto de Psiquiatria, do Hospital das Clínicas de São Paulo – FMUSP (equipe liderada pelo psiquiatra Dr. Estevão Vasdaz) e da Faculdade de Medicina de São José do Rio Preto (equipe liderada pela Dra. Agnes Cristina Fett Conte). A idade destes pacientes variou de 1 ano a 47 anos.

Através da análise do histórico familiar, da ficha de anamnese e das informações clínicas a respeito dos pacientes brasileiros, foram excluídos da amostra os indivíduos cujos pais reportaram consanguinidade, histórico de uso de medicamentos teratogênicos, drogas ilícitas, álcool, doenças infecciosas durante a gravidez, complicações durante o parto.

Excluiu-se também pacientes com resultado positivo para X-Frágil ou com alterações nas regiões 15q11-13, 16p11.2 e 22q13.3 (testes realizados pelo Centro

de Estudos do Genoma Humano), relatadas com frequência em indivíduos com TEA e já caracterizadas.

Todos os diagnósticos para o TEA se basearam no Manual Diagnóstico e Estatístico de Transtornos Mentais 4ª Ed. (DSM-IV), na 10ª Revisão da Classificação Internacional de Doenças (CID-10) e/ou no Questionário de Comunicação Social (SCQ). A gravidade dos sintomas dos pacientes brasileiros foi caracterizada utilizando-se a Escala de Avaliação do Autismo na Infância (CARS).

O grupo controle, em sua totalidade, foi composto de 714 indivíduos brasileiros. Com 200 servindo de referência para os dados obtidos com o aCGH 60K, 350 analisados por MLPA para a triagem adicional de CNVs no gene *SLC17A6* e 100 (provenientes do projeto 80+ coordenado pela Dra. Mayana Zatz[2]), analisados a partir de dados de sequenciamento de nova geração, foram utilizados também para a triagem adicional de CNVs no gene *SLC17A6*. Os 64 indivíduos restantes são pacientes do Centro de Pesquisas sobre o Genoma Humano e Células-Tronco (CEGH-CEL, IBUSP), apresentando doenças sem etiologia com sobreposição do Transtorno do Espectro Autista.

3.2 AMOSTRA NORTE AMERICANA

Os pacientes de origem norte americana, utilizados para a triagem de CNVs no gene *SLC17A6*, somaram um total de 5084. Para estes não foi possível fazer a distinção entre a porcentagem de pacientes do sexo feminino e masculino, sindrômicos e não sindrômicos. Os dados destes pacientes, obtidos através de sequenciamento de nova geração realizado de forma prévia, foram disponibilizados por quatro bancos de dados diferentes, *The Autism Genetic Resource Exchange, Simons Simplex Collection e Early Markers for Autism Study* (dados da pesquisadora Dalila Pinto proveniente da *Icahn School of Medicine- Mount Sinai)* para o laboratório da Dra. Weiss[3].

[2] Dra. Profa. Titular de Genética do Instituto de Biociências da Universidade de São Paulo (USP)
[3] PhD., Professora assistente, departamento de psiquiatria – Universidade da Califórnia, San Francisco - EUA

A presente pesquisa foi aprovada pela Comissão de Ética do Instituto de Biociências da Universidade de São Paulo (USP) e os pacientes ou responsáveis legais dos mesmos firmaram consentimento de participação através da assinatura do Termo de Consentimento Livre e Esclarecido (TCLE).

3.3 HIBRIDIZAÇÃO GENÔMICA COMPARATIVA POR MICROARRANJO PARA A DETECÇÃO DAS VARIAÇÕES NO NÚMERO DE CÓPIAS E PEQUENAS INSERÇÕES E DELEÇÕES.

Inicialmente, os DNAs obtidos a partir do sangue venoso periférico foram testados quanto a sua concentração e qualidade com o uso do *NanoDrop ND 1000 (Espectofotômetro da Peqlab Biotechnologie™)* em gel de agarose 1,2%. O controle de qualidade foi feito com a técnica de Hibridização Genômica Comparativa por Microarranjo (aCGH) utilizando o *array* customizado.

A técnica de aCGH consiste na comparação entre a intensidade de fluorescência emitida através da hibridação do DNA teste *versus* DNA controle com as sondas dispostas na superfície de uma lâmina de vidro.

A customização da lâmina de aCGH 60K foi realizada previamente pela pesquisadora Dra. Cintia Ribeiro[4] e incluiu um total de 269 genes, (ANEXO I) abrangendo:

- aqueles já associados ao TEA e a outras doenças neuropsiquiátricas;
- os envolvidos em vias biológicas alteradas em autistas;
- os que possuem função similar aos já associados ao transtorno;
- aqueles que são expressos preferencialmente no cérebro e/ou que codificam proteínas com funções cerebrais.

A customização se deu de forma que se pudesse analisar todos os exons e junções exons/introns, incluindo as regiões não traduzidas (UTR) de cada um dos genes. As sondas de oligonucleotídeos da lâmina foram selecionadas através do programa eArray (versão 6.2), de modo a garantir tanto a qualidade da sonda quanto

[4] MSc., PhD., pesquisadora na diagnósticos moleculares de doenças genéticas – DLE

a sobreposição de no mínimo 3 e no máximo 6 sondas por região de interesse. Este design incluiu mais de 33.000 sondas com sobreposição cobrindo um total de 4229 exons referentes aos 269 genes, com uma média de 8 sondas por exon. Para o restante do genoma ("backbone") foram desenhadas cerca de 22.000 sondas com espaçamento médio de 130 Kb. As sondas possuem tamanho aproximado de 60 pares de base. A sobreposição aumenta a capacidade de detecção de alterações menores que 10 Kb.

Quanto a customização da lâmina de aCGH 180K, esta foi realizada pela pesquisadora Dra. Eloísa de Sá Moreira[5] e incluiu um total de 1527 genes (ANEXO II), abrangendo:

- aqueles já associados ao TEA e a outras doenças neuropsiquiátricas;
- os envolvidos em vias biológicas alteradas em autistas ou que possuem função similar aos já associados ao transtorno;
- aqueles que são expressos preferencialmente no cérebro e/ou que codificam proteínas com funções cerebrais;
- 667 genes candidatos para TEA compilados no banco de dados da "Simons Foundation Autism Research Initiative" (SFARI) (BASU, KOLLU e BANERJEE-BASU, 2009), assim como os genes que foram considerados como mais relevantes para TEA em estudos recentes utilizando sequenciamento de exoma (HE *et al.*, 2013; IOSSIFOV *et al.*, 2014; DE RUBEIS *et al.* 2014; KRUMM *et al.*, 2015) e genes possivelmente envolvidos com a diferença da prevalência do autismo entre os sexos (KANG *et al.,* 2011; NAVA *et al.,* 2012; KIM *et al.,* 2014; LAI *et al.,*2015).

A customização do 180K também se deu de forma que se pudesse analisar todos os éxons e junções éxons/íntrons, incluindo as regiões não traduzidas (UTR) de cada um dos genes.

As sondas de oligonucleotídeos dessa lâmina foram selecionadas com a utilização do programa *SureDesign* da Agilent, e possuem tamanho aproximado de

[5] PhD., Gerente de pesquisa clínica no Evidencias - Kantar Health, KH, Brasil e pesquisadora colaboradora do laboratório de genética do desenvolvimento - USP

60 pares de base. A mediana de espaçamento entre as sondas é de 131 pb para cerca de 25.000 segmentos, sendo o segmento formado por um exon mais 200 pb de cada um dos introns que o flanqueiam. Este espaçamento corresponde a uma resolução de cerca de 400 pb, utilizando-se o critério de alteração em três sondas consecutivas para detecção de CNVs. Cada segmento está coberto em média por quatro sondas (mediana=5). Também foram selecionadas sondas nos 10 Kb adjacentes aos genes candidatos, com mediana de 3 sondas nessas regiões e desenho de sondas adicionais em áreas pouco cobertas. O *backbone* encontra-se representado por 71.006 sondas, sendo 36 Kb a mediana de espaçamento entre elas (o que corresponde a uma resolução de 108 Kb, considerando o critério de três sondas consecutivas para detecção de CNVs).

3.3.1 Realização do aCGH

A metodologia de aCGH é composta de quatro etapas principais, que são:

1) Denaturação do DNA

Para a realização da denaturação e fragmentação do DNA foi usado tanto o protocolo por temperatura, quanto o protocolo enzimático .

a) Temperatura - (800 ng de DNA em 26 µl de volume) foi realizada a incubação a 98ºC durante 7 minutos, em seguida adicionou-se 5 µl de *random primers*, seguido de mais 3 minutos de incubação à 98º. Após a denaturação e fragmentação as amostras foram transferidas para o gelo durante 5 minutos.
b) Enzimático – (700 ng de DNA em 20,2 µl de volume) preparou-se o mix de digestão, composto por *Nuclease-Free Water, 10x Restriction Enzyme Buffer, BSA, Alu I e Rsa I.* O volume final de 5,8 µl foi adicionado a cada uma das amostras, que em seguida foram levadas para um bloco aquecido a 37ºC por 2 horas e logo depois transferidas para outro a 65º C por 20 minutos. Após esse processo, foi adicionado 5 µl de *random primers*, seguido de mais 3 minutos de incubação à 98º. Após a denaturação e fragmentação as amostras foram transferidas para o gelo durante 5 minutos.

2) Marcação do DNA

Neste trabalho optou-se por hibridar paciente *versus* paciente em duas áreas diferentes, com marcação alternada de Cy3 e Cy5 em cada uma das áreas, ao invés de hibridar o DNA de cada paciente contra um controle.

A área aonde o paciente está marcado com Cy3 é utilizada como a de referência para caracterizar a alteração detectada como duplicação ou deleção (FIGURA 1).

A

Área da lâmina	Marcação com Cy3	Marcação com Cy5
1_1	Paciente A	Paciente B
1_2	Paciente B	Paciente C
1_3	Paciente C	Paciente D
1_4	Paciente D	Paciente A

B

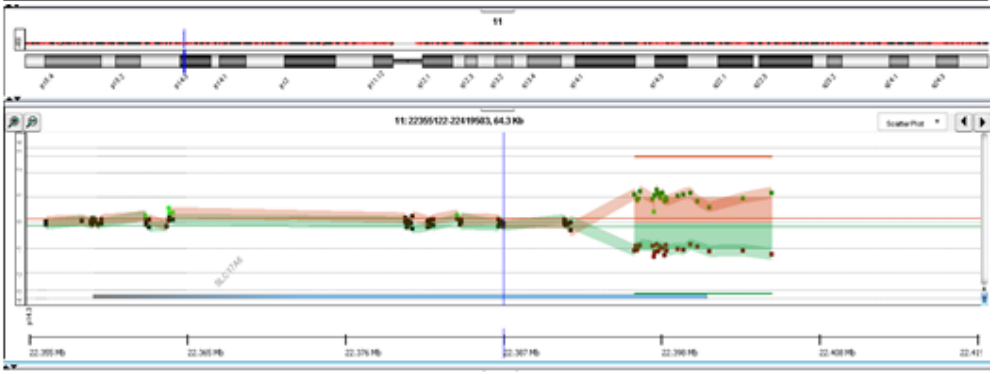


FIGURA 1. MARCAÇÃO EM DUPLICATA DE UM MESMO PACIENTE. A) EXEMPLO DE COMO É FEITA A HIBRIDAÇÃO DE PACIENTE *VERSUS* PACIENTE, COM CADA UM DELES MARCADO ALTERNADAMENTE DE CY3 E CY5. B) REGIÃO HACHURADA EM VERMELHO CORRESPONDE A REGIÃO NA QUAL O DNA DO PACIENTE ESTÁ MARCADO COM *CY3*, CARACTERIZANDO A ALTERAÇÃO COMO DUPLICAÇÃO, E REGIÃO HACHURADA EM VERDE CORRESPONDE A REGIÃO NA QUAL O DNA DO PACIENTE ESTÁ MARCADO COM *CY5.*

Para a marcação das amostras utilizou-se o *Labeling Kit, Agilent Technologies*®. Após a aplicação do mix de marcação em cada uma das amostras, essas foram incubadas durante 2 horas a 37ºC e logo em seguida a 65ºC por 10 minutos.

3) <u>Purificação do DNA</u>

A etapa de purificação é realizada para que se removam os fluoróforos que não foram incorporados. O sistema de purificação usado foi o de colunas (filtros *Amicon*) e lavagens com TE (pH 8.0).

4) <u>Hibridação</u>

Nessa etapa as amostras misturadas foram pareadas, seguindo-se a preparação do *Hybridization Master Mix* (Cot-1 DNA, Agilent 10x blocking agent, Agilent 2x Hi-RPM buffer) e a adição desse mix às amostras. Após esse processo as amostras foram incubadas a 98ºC durante 3 minutos e, imediatamente depois, a 37ºC por 30 minutos. Após esse processo, as amostras foram aplicadas na lâmina de base, o *gasket*, com a lâmina de vidro contendo as sondas colocadas por cima.

Posteriormente, as lâminas foram colocadas no forno estabilizado a 67ºC, com rotação a 20 rpm, durante 24 horas.

Após o período de 24 horas de hibridação, as lâminas foram submetidas a várias lavagens subseqüentes (com *Agilent Oligo aCGH Wash Buffer 1* por 5 minutos *e Agilent Oligo aCGH Wash Buffer 2* por 1 minuto), e a secagem final (com *Stabilization and Drying Solution, 500 ML, Agilent Technologies*®).

As hibridações realizadas com a lâmina customizada de aCGH 60K foram feitas durante o doutorado da Dra. Cintia Ribeiro, cuja tese foi entitulada “Estudo de genes candidatos aos transtornos do espectro autista” e apresentada no Instituto de Biociências da Universidade de São Paulo em 2013.

3.3.2 Análise dos Dados de aCGH

As imagens das hibridizações foram obtidas com o uso do *scanner Agilent* e do programa *Feature Extraction v10.5.* e analisadas com o programa *Agilent Genomic WorKbench Lite Edition v7.*

Para este trabalho os resultados obtidos com o aCGH 60K pela Dra. Cintia Ribeiro foram reanalisados. A reanálise foi feita visto que a customização do array apresentou limitações como: (a) excesso de sondas sobrepostas, o que pode acarretar em um aumento do ruído de hibridação e em resultados falso positivos e (b) espaçamento médio entre as sondas de *backbone* (130 Kb), que dificulta a delimitação do tamanho mínimo e máximo das alterações.

O algoritmo estatístico utilizado foi o ADM-2 e o limiar 7.0. Os critérios utilizados para a seleção das alterações foram: (a) valor do log2 da razão de DNA marcado com cy3 dividido por DNA marcado com cy5 acima de 0,3 ou abaixo de –0,3 e (b) alteração em duplicata, ou seja, presente nas duas áreas onde o paciente teve seu DNA hibridado, e (c) apresentar tamanho mínimo de 1Kb.

Para a classificação de CNVs como potencialmente patogênicas, foram usados critérios baseados no *American College of Medical Genetics* (KEARNY *et al.*, 2011), conforme a seguir :

- Apresentar frequência menor que 1% no nosso grupo de 200 controles hibridados com a lâmina customizada de aCGH 60K (Apenas para as CNVs obtidas com o aCGH 60K)
- Apresentar frequência menor que 1% no *Database of Genomic Variants* (DGV)
- Apresentar função relevante para o neurodesenvolvimento e expressão no tecido cerebral e/ou já ter sido associado ao TEA no banco de dados *Simons Foundation Autism Research Initiative* (SFARI) e/ou DECIPHER.

3.4. MLPA SNP-ARRAY e aCGH COMERCIAL

A metodologia de MLPA foi utilizada para a triagem inicial dos pacientes, a fim de excluir aqueles com alterações nas regiões 15q11-q13, 16p11.2 e 22q13.3 (*hotspot* em indivíduos com TEA), e para a triagem de novas alterações no *SLC17A6* em um número maior de pacientes. Para ambos os fins, os experimentos foram feitos previamente.

Para as regiões consideradas *hotspot* em indivíduos com TEA, os experimentos foram realizados no Centro de Pesquisa sobre o Genoma Humano – USP, enquanto que para a triagem de alterações no *SLC17A6*, os experimentos foram realizados no Laboratório de Genética do Desenvolvimento da USP, em colaboração com a doutoranda Ana Luiza Bossolani[6], da Universidade Estadual de São Paulo (UNESP), orientada pela Profa. Dra. Agnes Cristina Fett Conte[7].

As sondas foram desenhadas através do programa *Human MLPA Probe Design* (H-MAPD). O gene *SLC17A6* possui 12 *exons* e não foi possível desenhar sonda apenas para o *exon* 1.

[6] Phd., Professora pesquisado da Universidade Federal de Mato Grosso do Sul
[7] Phd., Professora pesquisadora da Universidade Estadual de Campinas

A técnica de MLPA foi realizada segundo descrito por Schouten *et al*. (2002), com adaptações para permitir a leitura de resultados no sequenciador automático *ABI-3730 DNA Analyzer*® (Life Technologies). Para as reações de hibridação e amplificação utilizou-se o kit EK1 (SALSA MLPA EK1 reagent kits - MRC-Holland) e para controle dos experimentos o kit de referência P200 (SALSA MLPA KIT P200-A1 Reference-1 - MRC-Holland), de acordo com as especificações do fabricante.

As análises dos resultados foram realizadas pelos programas *Genemapper*®, versão 4.0 (*Life Technologies*) e *Genemarker*®, versão 1.97 (*Softgenetics*), considerando-se razões de referência abaixo de 0.75 e acima 1.30 como deleção e duplicação, respectivamente.

3.5 SNP-ARRAY e aCGH COMERCIAL

Ambas as metodologias, aCGH comercial e SNP-array, foram utilizadas para a confirmação e delimitação das alterações identificadas com o aCGH customizado 60K.

A plataforma de SNP-array utilizada foi a *GeneChip® Human Mapping 500K Array Set, Affymetrix*® (Santa Clara, CA, EUA), de acordo com o protocolo do fabricante. Essa plataforma compreende dois *arrays*, cada um capaz de genotipar em média 250 mil polimorfismos de base única (SNPs) – aproximadamente 262 mil para a enzima de restrição N*sp* e 238 mil para a S*ty*. A distância física mediana entre as sondas é de 2,5Kb e a média de 5,8Kb.

Após a hibridação, lavagem e marcação, as lâminas foram escaneadas no *GeneChip® Scanner 3000 7G* (*Affymetrix*® - Santa Clara, CA, EUA). Os dados de intensidade foram gerados pelo programa *GeneChip Command Console Software* (*Affymetrix*® (Santa Clara, CA, EUA).

A qualidade da hibridização das amostras foi avaliada pelo índice de *QC Call Rate*; o limite mínimo escolhido para este índice foi 95,0. A análise desses dados foi realizada através dos programas *Genotyping Console, Affymetrix* (Santa Clara, CA, EUA), *PennCNV* (WANG *et al*., 2007; WANG, 2008) e dChip (LI E HUNG-WONG 2001; LIN *et al*., 2004).

A plataforma do aCGH comercial utilizada foi a *SurePrint G3 Human CGH Microarrays 180 K DA Agilent Technologies* (Santa Clara, CA, USA), seguindo o protocolo descrito no item 2.2.

Ambas as metodologias foram realizadas no laboratório de citoquímica e citometria, sob orientação da Dra. Carla Rosenberg[8].

3.6. REAÇÃO em CADEIA da POLIMERASE em TEMPO REAL (qPCR)

A técnica de Reação em Cadeia da Polimerase (PCR) em tempo real também foi utilizada para validação das CNVs através da quantificação do número de cópias do DNA. Os primers utilizados na qPCR foram desenhados com o uso do programa Ensembl, versão 77, e NCBI/Primer-Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/).

As reações foram realizadas utilizando 2 μL de DNA com concentração de 2,5 ng/ μL, 0,5 μL de cada um dos primers, F e R e 10 μL de *Syber Green PCR Master Mix* (*Applied Biosystems*, EUA) em um volume final de reação igual a 20L. As reações de qPCR foram feitas no sistema de detecção Applied Biosystems 7500 Fast Real-Time PCR System.

Os *primers* foram inicialmente testados por meio da análise da curva de *melting* (derivada da observação da temperatura na qual a fita dupla de DNA desnatura, fato dependente do tamanho da fita e da sua composição de nucleotídeos), para avaliar a formação de dímeros de *primer* ou à amplificação de produtos inespecíficos.

Para permitir comparação de número de cópias entre afetado e controle, cada conjunto de *primers* (referência e teste) deve ter uma eficiência de amplificação semelhante. Uma eficiência de 100% significa que o produto de amplificação dobra a cada ciclo da qPCR.

Os experimentos foram realizados em duplicata para cada um dos *primers*, e com uso de um controle negativo. As curvas foram analisadas através do programa 7500 Software Download v. 2.0.6 (Applied Biosystems).

[8] Phd., Professora pesquisadora do Instituto de Biologia da Universidade de São Paulo.

A quantificação do número de cópias foi realizada através da comparação com os valores obtidos de um gene envolvido em funções básicas necessárias para a manutenção da célula, os chamados *housekeeping* genes, os quais são usados como controles endógenos. Nesse experimento foi utilizado como *housekeeping gene o GAPDH, ACTB1, TRFC* e *TERT*.

O cálculo usado para a checagem da presença da deleção foi feito em conformidade com metodologia de Pfaffl *et al.* (2001), que leva em consideração as eficiências de amplificação de cada um dos amplicons (gene alvo e gene de referência) que estão sendo analisados.

3.7 SEQUENCIMENTO do EXOMA por *NEXT GENERATION SEQUENCING* (*NGS*)

A metodologia de *NGS* foi utilizada para verificar a possibilidade do seu uso para corroborar e complementar os resultados obtidos com o aCGH.

Essa técnica é realizada pela equipe especializada da Central de Serviços do CEGH-CEL, IBUSP. O *kit* para sequenciamento utilizado foi o *TruSeq Exome Enrichment Kit* (Illumina, Inc), seguindo o protocolo sugerido pelo fabricante, que engloba exons codificantes anotados pelo projeto GENCODE bem como o banco de dados do CCDS e RefSeq. Esse kit incorpora também regiões exônicas e RNAs não-codificantes, totalizando um tamanho de captura de aproximamdamente 50 Mb com, em média, apenas 3 µg de DNA genômico (gDNA). A captura das sequências de DNA das amostras é feita com o equipamento Illumina HiSeq 2000 (Illumina, Inc.) - *paired-end* 100x100 pares de base (bp).

3.7.1 Análise dos Dados de *NGS* para CNVs

A primeira e segunda etapa da análise dos dados foram realizadas no Centro de Estudos do Genoma Humano da USP, pelo Dr. Guilherme Yamamoto[9].

1º Pré alinhamento dos dados:

As sequencias em formato FASTQ, geradas a partir da plataforma HiSeq-Illumina, são inicialmente analisadas quanto a sua qualidade utilizado o programa FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

2º Alinhamento dos dados:

Com o arquivo FASTQ de boa qualidade é realizado o alinhamento com o uso do programa Burrows-Wheeler Aligner (BWA). As sequencias referências utilizadas para esse processo foram obtidas do GRCh37 (hg19) *Genome Reference Consortium Human Reference* (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/chromosomes/). O *output* gerado após o alinhamento possui formato BAM (*Sequence Alignment/Map*).

3.7.2 Chamada de CNVs

A análise de CNVs a partir do BAM. foi realizada com o uso do *NextGene Software (SoftGenetics*, LLC - State College, PA) e do XHMM (eXome-Hidden Markov Model, FROMER *et al*., 2012)

Ambos os softwares utilizam informações a respeito da cobertura, medida em RPKM (10^9 x C / N x L; C: Número de *reads* mapeados na região, equivalente ao exon; N: Número total de *reads*; L: Tamanho da região), para fazer a chamada de CNVs através do modelo de *Hidden Markov.*

Para a normalização dos dados, o NextGene utiliza os dados de cobertura de um paciente *versus* um controle, através do cálculo (Amostra / Amostra + Controle) para cada exon.

O *XHMM*, por sua vez, faz uso de componente principal de análise que remove variância de grande dimensão, como porcentagem de GC, variações de cobertura associadas a corrida de exoma do paciente, entre outras não associadas a CNVs,

[9] Técnico do Centro de Estudos do Genoma Humano da Universidade de São Paulo.

usando a cobertura como componente principal. Essa forma de normalizar os dados descarta a necessidade do uso de controles, e requer apenas um *pool* de pacientes.

Após a normalização dos dados, ambos os softwares convertem a taxa média de cobertura para cada região em z *score*, que por sua vez é utilizado para a chamada de CNVs através do *Hidden Markov Model*. Utilizamos para ambos os softwares os parâmetros *default.*

Para verificar o potencial do *NextGene* e do *XHMM* em complementar os resultados obtidos através do aCGH 180K, os dados de exoma foram sobrepostos aos do array, que foi utilizada como metodologia de referência.

O critério utilizado de sobreposição de CNVs entre o aCGH e o XHMM ou o NextGene, considerou sobreposta qualquer CNV que compartilhasse pelo menos um par de base entre si. Com isso, pretendeu-se diminuir a chance de excluirmos CNVs que são idênticas, porém, devido a possíveis limitações na capacidade de delimitação de ponto de quebra do aCGH e também do *NGS*, quando esse está localizado em região intrônica, acabam sendo chamadas em posições relativamente diferentes por cada uma das metodologias (TATTINI, 2015).

3.8 ANÁLISES ESTATÍSTICAS

A comparação entre a proporção de alterações encontradas em pacientes com TEA e nos controles foi realizada com o teste exato de Fisher e x^2. Para comparar médias entre os grupos relacionados foi usado o teste Mann-Whitney. O teste Spearman de correlação foi aplicado para correlacionar tamanho de CNV com um determinado grupo de pacientes. Todas as análises estatísticas foram realizadas com o programa *GraphPad Prism* versão 6.01 (*GraphPad Software*, Inc).

4. RESULTADOS

4.1 ANÁLISE DOS DADOS DE aCGH 60K

4.1.1 Validação Preliminar da Lâmina Customizada

Uma validação preliminar da lâmina customizada de aCGH 60K foi realizada previamente pela Dra. Cintia Ribeiro com uma amostra de dez pacientes, todos diagnosticados com Transtorno do Espectro Autista e com resultado positivo por MLPA para uma das regiões consideradas *hotspot* em indivíduos com TEA. Todas as CNVs foram detectadas pelo array customizado[10] .

4.1.2 Análise dos Dados de aCGH dos Primeiros Vinte Pacientes

A reanálise dos dados de aCGH 60K visou, em um primeiro momento, caracterizar a eficácia do array customizado em identificar InDels considerados grandes, isto é, entre 70pb-1Kb. Para esta primeira análise foi utilizada uma amostra de 20 pacientes, 13 do sexo masculino e 7 do sexo feminino, 1 sindrômico e o restante não sindrômico.

Foram encontradas 35 alterações em 15 pacientes conforme apresentado na Tabela 1, tendo sido escolhidas para validação por PCR apenas deleções com tamanho variando entre 70pb e 900pb. Estas somaram um total de 20, e estão identificadas com coloração cinza na Tabela 1.

[10] Tese "Estudo de genes candidatos aos transtornos do espectro autista", Instituto de Biociências da Universidade de São Paulo, 2013.

TABELA 1. DELEÇÕES E DUPLICAÇÕES DETECTADAS EM 20 PACIENTES COM TEA APÓS ANÁLISE DO aCGH C

Paciente	Sexo	Fenótipo	Tipo	Tamanho (pb) Min (Max)	Posição (Hg 19)	Genes envolvidos
F1359-1	F	ñ sindrômico	dup	726 (741)	4q32.2	*NPY1R*
F1385-1	M	ñ sindrômico	dup	102 (229)	1q25.2	*ASTN1*
			del	308 (320)	13q13.3	*TRPC4*
F1697-1	M	ñ sindrômico	del	95 (110)	17q21.32	*NSF*
F1722-1	M	ñ sindrômico	del	241 (634)	3p14.3	*CACNA2D3*
F2452-1	F	ñ sindrômico	del	673 (809)	1q25.2	*ASTN1*
			del	122 (130)	14q12	STXBP6
			del	93 (998)	2q34	*ERBB4*
F2697 -1	M	ñ sindrômico	del	673 (809)	1q25.2	*ASTN1*
			del	88 (103)	3p26.3	*CHL1*
			del	308 (320)	13q13.3	*TRPC4*
F3569-1	M	ñ sindrômico	sem alteração			
4173-1	M	ñ sindrômico	del	77 (535)	3p14.3	*CACNA2D3*
			del	76 (120)	13q13.3	*TRPC4*
4181-1	M	ñ sindrômico	sem alteração			
4204 - 1	F	ñ sindrômico	sem alteração			
4209-1	M	ñ sindrômico	del	76 (120)	13q13.3	*TRPC4*
4236-1	M	ñ sindrômico	del	131 (203)	7q36.2	*HTR5A*
4632-1	F	ñ sindrômico	del	97 (112)	12q21.1	TPH2
			del	76 (120)	13q13.3	*TRPC4*
			del	229 (257)	14q12	*STXBP6*
			del	133 (153)	17q11	*NF1/EVI2B*
			del	94 (128)	22q11	*MICAL3*
4922-1	M	ñ sindrômico	sem alteração			
5074-1	M	ñ sindrômico	del	131 (203)	7q36.2	*HTR5A*
5241-1	M	ñ sindrômico	del	76 (97)	4q34.1	*GLRA3*
			del	102 (128)	7q31.1	*NRCAM*
			del	81 (100)	9q33.1	*TRIM32*
5249-1	F	ñ sindrômico	dup	107 (124)	4p12	*GABRG1*
			dup	110 (133)	4p12	*GABRG1*
			del	81 (129)	7q31.1	*LAMB1*
			del	76 (120)	13q13.3	*TRPC4*
5252-1	F	ñ sindrômico	del	750 (768)	4p12	*GABRG1*
			del	76 (120)	13q13.3	*TRPC4*
			dup	496 (558)	17q21.32	*NSF*
5367-1	M	ñ sindrômico	sem alteração			
5368-1	F	sindrômico	del	673 (809)	1q25.2	*ASTN1*
			dup	673 (682)	4q32.2	*NPY1R*
5469-1	M	ñ sindrômico	del	85 (137)	1q41	*MARK1*
			del	397 (412)	2q31.1	*RAPGEF4*

Deleções selecionadas para serem validadas por PCR

M = masculino; F = feminino; ñ = não; dup = duplicação; del = deleção

O resultado esperado para a PCR, na presença de deleções, é o de 2 fragmentos, um de tamanho menor confirmando a alteração, e outro de tamanho normal equivalente ao alelo normal e observado nos controles. A eletroforese em gel de agarose, no entanto, não revelou fragmentos de tamanhos menores nos pacientes com as alterações selecionadas, quando comparados com os controles, como exemplificado na Figura 2. Desta forma, nenhuma das deleções selecionadas com tamanho entre 70 pb e 1 Kb foram validadas por este método.

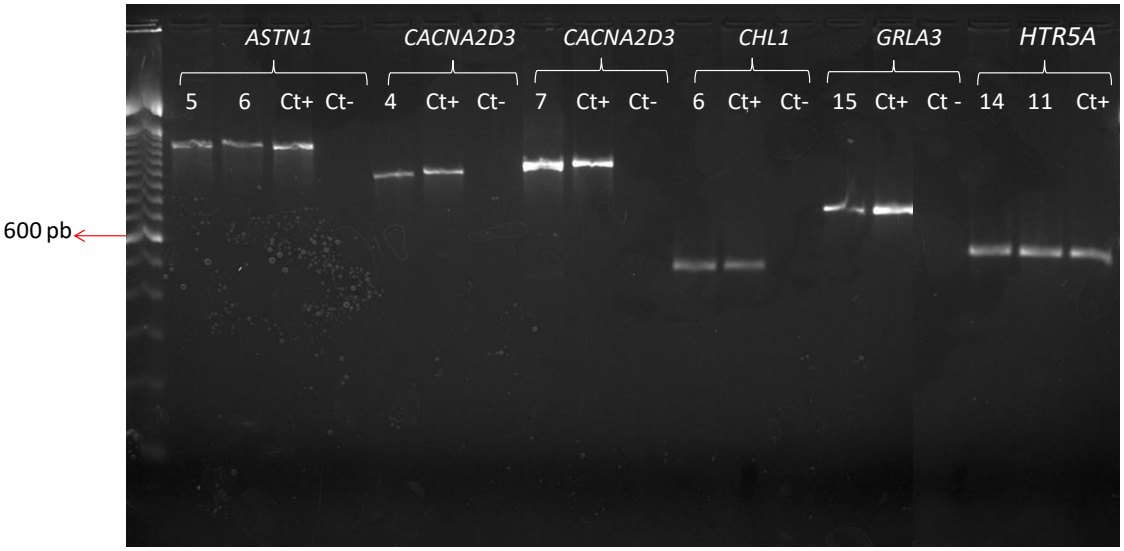


FIGURA 2. ELETROFORESE EM GEL DE AGAROSE (2,5 %) MOSTRANDO AS ALTERAÇÕES SELECIONADAS NOS PACIENTES 1722-1 (4), 2452-1 (5), 2697-1 (6), 4173-1 (7), 4236-1 (11) E 5074-1 (14) E 5241-1 (15). CTRL+*: CONTROLE; CTRL-*: CONTROLE NEGATIVO (ÁGUA). A FLECHA VERMELHA INDICANDO O FRAGMENTO DE REFERÊNCIA DO PESO MOLECULAR DE 100 PB (INVITROGEN).

Esses resultados sugerem que o array de 60K customizado não apresenta resolução suficiente para detectar InDels de 70pb a 1Kb, caracterizando os resultados descritos na Tabela 1 como falso-positivos.

Esses dados direcionaram a presente pesquisa no sentido de detectar e validar as CNVs com tamanho mínimo de 1Kb e que se enquadrassem no critério adotado de potencialmente patogênica.

4.1.3 Análise dos Dados de aCGH de Noventa e Oito Pacientes

Esta etapa teve como foco a identificação de CNVs maiores que 1Kb. Para isto, foram reanalisados dados de aCGH de 98 pacientes (94 casos esporádicos e 4 casos familiais), 71 homens e 27 mulheres, sendo 77 deles não sindrômicos e 21 sindrômicos. Entre os 98 pacientes estão incluídos os 20 indivíduos que foram analisados na sessão 4.1.2.

Dessa amostra, 9 casos esporádicos apresentaram 10 CNVs potencialmente patogênicas, correspondendo a uma prevalência de 9% (9/98) (TABELA 2).

A prevalência obtida difere quando se considera o grupo de sindrômicos e não sindrômicos isoladamente, sendo equivalente a 19% (4/21) e 7,8 % (5/77), respectivamente ($p = 0,007$; $x^2 = 3,11$).

Entre as 10 alterações, 2 são pequenas, menores que 50 Kb, e equivalem a 20 % das alterações.

Não foram identificadas sobreposições em termos de regiões genômicas entre as 10 CNVs.

TABELA 2. RESULTADO DA ANÁLISE DO aCGH 60k DE 98 PACIENTES COM TEA

Paciente	Sexo	Fenótipo	Posição da CNV	Tipo	Origem	Tamanho (kb) Min	Posição (Hg19)*	Gene(s) envolvidos
4181-1	M	ñ sindrômico	11p14.3	dup	ñ materna, DNA do pai ñ disponível	9 (31)	22391688-22401068	*SLC17A6*
6098-1	M	ñ sindrômico	7q31.1	del	pais ñ disponíveis	480 (533)	112644855-113125338	*GPR85, LOC401397*
6034-1	M	ñ sindrômico	13q21.2	del	paterna	306 (376)	60420512-	*DIAPH3*
5502-1	M	ñ sindrômico	1q21.1-21.2	dup	materna	1322 (3722)	146506310-	> 15
1665-1	M	ñ sindrômico	18q21.2	dup	materna	12 (85)	51680552-	*MBD2*
2701-1	M	sindrômico	Xq24-25	del	*de novo*	142 (2,779)	122318291-	*GRIA3* (> 15)
3536-1	M	sindrômico	11q13.3-13.4	del	ñ paterna, DNA da mãe ñ disponível	1260 (2677)	69482657-70742731	> 15
4174-1	F	sindrômico	1p31.3-22.3	dup	pais ñ disponíveis	21697 (25415)	64708662-86405527	> 15
			14q21.2-21.3	del		1620 (3458)	46375799-47995925	*LINC0087, RPL10L, MDGA2* (> 15)
6509-1	F	sindrômico	3p14.3-14.1	del	pais ñ disponíveis	10225 (14469)	58196379-68421182	> 15

M = masculino; F = feminino; ñ = não; dup = duplicação; del = deleção
*a posição das CNVs e os genes envolvidos nelas são referentes ao seu tamanho mínimo

A prevalência de homens *versus* mulheres com CNVs patogênicas foi de 9,9% (7 em um total de 71 homens) e 7,4% (2 em um total de 27 mulheres) respectivamente, diferença esta não significativa estatisticamente com $p = 0,70$ ($x^2 = 0,141$) .

A diferença entre a média de tamanho das alterações detectadas nas mulheres e nos homens foi de 11.180,7 Kb (d.p. =10073) e 504,4 Kb (d.p. = 562,5).

Cabe ressaltar também que o único paciente que apresentou 2 CNVs é do sexo feminino.

Em ambos os grupos, de homens e de mulheres, a maior parte das alterações foi caracterizada como deleção (6 do total de 10 CNVs), perfazendo um total de 60% das CNVs.

Foi possível determinar a origem apenas de 4 das 10 CNVs, por ausência de um ou ambos os pais. Destas, 3 são herdadas, 2 de origem materna e 1 de origem paterna, e 1 é *de novo*.

Observamos uma tendência de correlação positiva entre o tamanho da alteração e quadro clínico (sindrômico e não sindrômico) (r = 0,52; P = 0,19), apesar de não significativa estatisticamente, ou seja, pacientes não sindrômicos parecem ter alterações com tamanho médio menor do que o obtido para o grupo de pacientes sindrômicos, que apresentam, no geral, quadro clínico mais complexo e grave do que os pacientes não sindrômicos. A média de tamanho das CNVs identificadas em pacientes sindrômicos foi de 6.988,8 Kb, enquanto que a média obtida para o grupo de não sindrômicos foi de 425,8 Kb, apesar da aparente diferença entre as médias, esta não é significativa (p = 0,09).

A média do total das alterações foi de 3.707,3 Kb (d.p.=7025), com a média das duplicações 2,46 vezes maior do que a das deleções, 5.760 Kb (d.p.=10643) e 2.338,8 Kb (d.p.=3906), respectivamente, apesar desta diferença não apresentar significância estatística (p = 0,7).

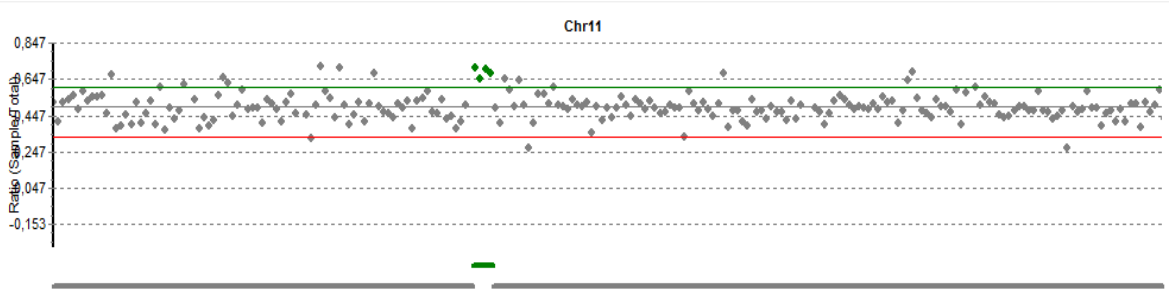
Das dez CNVs encontradas, 8 são maiores do que 50 Kb e, com exceção de uma deleção de 306 Kb, identificada no paciente 6034-1, que abrange apenas o gene *DIAPH3*, da sua região 5' UTR até o exon 22, as demais envolvem mais de um gene. Cinco destas 8 alterações grandes também foram detectadas pela Dra. Cintia Ribeiro e validadas por PCR quantitativa, SNP-array 500k ou array CGH comercial 180K[11].

As duas CNVs pequenas (<50 Kb), afetam apenas um único gene: a primeira é uma duplicação de 12 Kb, identificada no paciente 1665-1, envolvendo os exons 4-

[11] Tese "Estudo de genes candidatos aos transtornos do espectro autista", Instituto de Biociências da Universidade de São Paulo, 2013

7 do *MBD2* e a segunda é uma duplicação de 9 Kb, detectada no paciente 4181-1, envolvendo os exons 9-12 do *SLC17A6.*

A análise das CNVs pequenas nos genes *SLC17A6* e *MBD2* também foi feita por *NGS (*a realizada para o gene *SLC17A6* pode ser vista na FIGURA 3), PCR quantitativa e SNP-array (500 K), este último com o objetivo de identificar alterações nos genes contíguos. A análise de SNP-array não foi capaz de identificar as alterações pequenas no *SLC17A6* e no *MBD2* pois não apresenta sondas suficientes nesta região, no entanto não revelou alterações nos genes contíguos nem outras CNVs patogênicas.



Sample	F4181-1.m															
Control	F2685-1.m															
Index	Descriptio	Chr	Chr Start	Chr End	Gene	CDS	Length	Ratio	Total RPK	Dispersion	Normalize	Deletion S	Normal Sc	Duplicatio	HMM Calls	RPKM(Sar
3940		chr11	22359668	22360165	SLC17A6;	1	498	0,455	421,880	0,0254	-0,67;-0,19;	1,04	4,44	0,69	Normal	192,333;22
3941		chr11	22363075	22363326	SLC17A6;	2	252	0,537	1138,595	0,0070	-5,19;-0,16;	0,00	5,05	1,63	Normal	611,119;52
3942		chr11	22364794	22364911	SLC17A6;	3	118	0,425	243,136	0,0559	-0,43;-0,35;	2,02	2,54	0,89	Normal	103,453;13
3943		chr11	22380960	22381073	SLC17A6;	4	114	0,445	361,177	0,0317	-0,54;-0,25;	1,48	3,56	0,71	Normal	160,625;20
3944		chr11	22382444	22382530	SLC17A6;	5	87	0,375	273,659	0,0472	-0,24;-0,46;	3,73	1,83	0,36	Normal	102,538;17
3945		chr11	22384286	22384371	SLC17A6;	6	86	0,415	591,899	0,0157	-0,31;-0,30;	2,89	3,00	0,07	Normal	245,678;34
3946		chr11	22387094	22387235	SLC17A6;	7	142	0,507	371,941	0,0304	-1,10;-0,23;	0,36	3,87	1,75	Normal	188,469;18
3947		chr11	22391586	22391734	SLC17A6;	8	149	NA	NA	NA	NA;NA;NA	NA	NA	NA	Uncalled	0,000;0,000
3948		chr11	22396302	22396433	SLC17A6;	9	132	0,709	721,658	0,0118	-10,03;-2,52	0,00	0,01	25,16	Duplicatior	512,202;20
3949		chr11	22397529	22397638	SLC17A6;	10	110	0,651	360,433	0,0318	-2,81;-0,66;	0,01	1,08	6,56	Duplicatior	234,759;12
3950		chr11	22398092	22398218	SLC17A6;	11	127	0,701	674,171	0,0130	-8,76;-2,16;	0,00	0,03	21,56	Duplicatior	473,215;20
3951		chr11	22398952	22401046	SLC17A6;	12	2095	0,680	677,517	0,0129	-7,92;-1,84;	0,00	0,06	18,35	Duplicatior	460,778;21

FIGURA 3. REPRESENTAÇÃO DA CNV ENCONTRADA NO GENE *SLC17A6* NO PACIENTE 4181-1 USANDO O SOFTWARE *NEXTGENE*.

Dentre as 2 CNVs pequenas, a no *SLC17A6* é a única que contém um gene que não foi ainda associado como de susceptibilidade ao autismo. Este fato levou-nos a analisar o *SLC17A6* em outros 407 pacientes (322 homens e 85 mulheres) e 350 controles (202 homens e 148 mulheres) por MLPA, com sondas específicas para essa região. Encontrou-se duplicação neste gene envolvendo os exons 9-12 em outro probando, de origem materna, mas também em um controle.

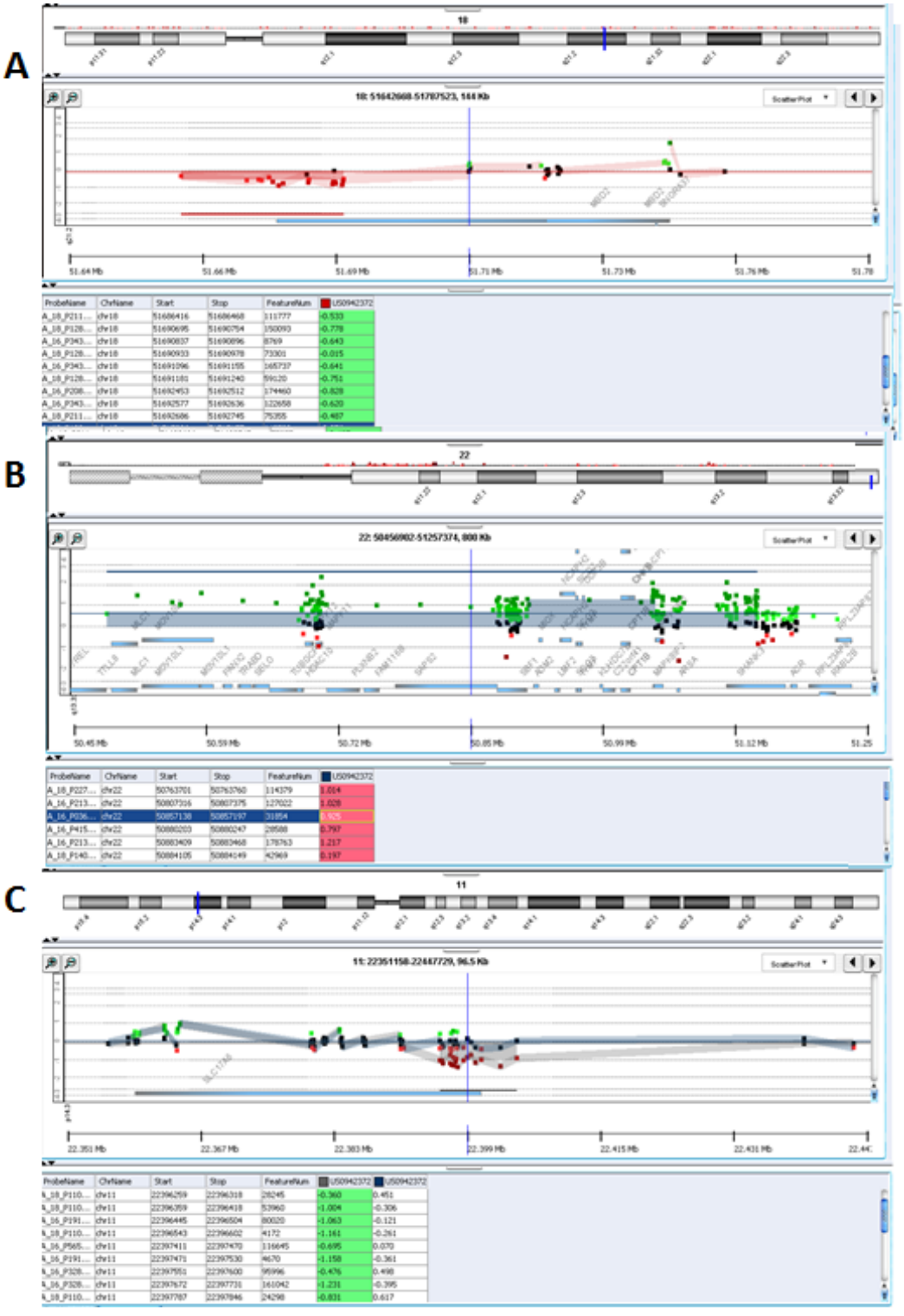
Com esta triagem adicional por MPLA obtivemos uma prevalência de CNVs no *SLC17A6* nos pacientes com TEA de 1/407 ou 0,002 e no grupo de controles de 1/350, ou 0,003.

4.2 ANÁLISE DOS DADOS DE ARRAY CGH 180K

Os resultados promissores obtidos com a lâmina customizada 60K em termos de sensibilidade para detectar grandes e pequenas CNVs, bem como a percepção de limitações desta customização, induziram ao estabelecimento de uma nova versão customizada de array (4 x 180K).

4.2.1 Validação Preliminar da Lâmina 180K Customizada

A validação preliminar da lâmina customizada de aCGH 180K foi realizada em duas etapas. A primeira foi conduzida com uma amostra de 7 pacientes com diagnostico de TEA, 5 deles com alterações confirmadas por MLPA nas regiões *hotspot*, 2 com CNV na região15q11-q13, 1 com CNV na região 16p11.2 e 2 com CNV na 22q13.3. Os 2 indivíduos restantes apresentaram alterações pequenas detectadas pelo aCGH customizado 60K e confirmadas na região 11p14.3 (*SLC17A6*) e 18q21.2 (*MBD2*), descritas na sessão 4.1 deste trabalho. Os tamanhos destas alterações se distribuíram entre 9 kb até 2,5 Mb. Todas as alterações foram detectadas pelo novo aCGH (FIGURA 4).


A
B
C

D
E
F
G

FIGURA 4. REPRESENTAÇÃO DE 7 CNVS CONFIRMADAS E UTILIZADAS PARA A VALIDAÇÃO DA LÂMINA CUSTOMIZADA 180K (A) PACIENTE 1665-1, MARCADO COM CY5, COM DUPLICAÇÃO NA REGIÃO 18q21.2, (B) PACIENTE 3555-1, MARCADO COM CY5, COM DELEÇÃO NA REGIÃO 22q13.3, (C) PACIENTE 4181-1, MARCADO COM CY5, COM DUPLICAÇÃO NA REGIÃO 11p14.3, (D) PACIENTE 5240-1, MARCADO COM CY3, COM DUPLICAÇÃO NA REGIÃO 15q11-q13, (E) PACIENTE 4290-1, MARCADO COM CY5, COM DELEÇÃO NA REGIÃO 15q11-q13, (F) PACIENTE 4296-1, COM DELEÇÃO NO CROMOSSOMO 22q13.3 E (G) PACIENTE 6507-1, MARCADO COM CY3, COM DELEÇÃO NO CROMOSSOMO 16p11.2.

A segunda etapa teve o seu foco na validação por PCR quantitativa de CNVs maiores que 1Kb selecionadas de acordo com os critérios utilizados durante a análise dos dados da lâmina de 60K, e classificadas como potencialmente patogênicas.

A amostra utilizada para essa etapa foi constituída por 31 pacientes independentes dos 7 indivíduos hibridados na primeira etapa de validação da lâmina 180K.

Foram identificadas 12 CNVs potencialmente patogênicas em 9 pacientes (TABELA 3). Dentre as 12 CNVs, 4 foram classificadas como pequenas (<50 Kb) e o restante como grande (>50Kb), tendo sido escolhidas para validação por qPCR um total de 9.

TABELA 3. RESULTADO DA ANÁLISE DO aCGH 180k DE 31 PACIENTES AFETADOS PELO TRANSTORNO DO ESPECTRO AUTISTA

Paciente	Sexo	Fenótipo	Posição da CNV	Tipo	Origem	Tamanho (Kb) Min (Max)	Posição (Hg19)*	Gene(s) envolvidos*
8205-6	**M**	ñ sindrômico	Xp11.23	del	ñ avaliada	66,2 (151,1)	47905478-47971637	*ZNF630, SSX6*
8205-5	**M**	ñ sindrômico	Xp11.23	del	ñ avaliada	66,2 (151,1)	47905478-47971637	*ZNF630, SSX6*
7935-5	M	ñ sindrômico	18q21.1	dup	ñ avaliada	4 (25,5)	47806537-47810550	*MBD1, CXXC1*
10291-1	M	ñ sindrômico	7q36.2	del	ñ avaliada	4,5 (392,9)	153998331-154002788	*DPP6*
6281-1	M	ñ sindrômico	4q31.23	del	ñ avaliada	254,8 (324,4)	149300414-149555247	*NR3C2*
1389-1	F	ñ sindrômico	1p13.3	dup	ñ avaliada	62,2 (105,1)	110177934-110240126	*GSTM1, GSTM2, GSTM4*
			7q11.23	del	ñ avaliada	475,9 (553,4)	76095829-76571698	*DTX2, FDPSP2, UPK3B, POM2P3*
5674-1	M	ñ sindrômico	2q14.3	del	ñ avaliada	59,5 (62,8)	127805569-127865041	*BIN1*
			6q13	del	ñ avaliada	82,6 (83,6)	72596189-72678775	*RIMS1*
			10q23.1	del	ñ avaliada	1,1 (4,6)	82116608-82117750	*DYDC2*
			11p11.2	del	ñ avaliada	152,7 (196,5)	46353215-46505917	*DGKZ, MIR4688, MDK, CHRM4, AMBRA1*
10636-1	F	ñ sindrômico	5q13.1-q13.2	dup	ñ avaliada	936,9 (2026,3)	68849594-69786489	> 15
4922-1	M	ñ sindrômico	Xp11.23	dup	ñ avaliada	2,2 (2,4)	49054299-49056452	*SYP*

Obs: Os pacientes 1 e 2, marcados em negrito, são irmãos
M = masculino; F = feminino; ñ = não; dup = duplicação; del = deleção
* as células sombreadas representam as CNVs selecionadas para serem validadas por qPCR
*a posição das CNVs e os genes envolvidos nelas são referentes ao seu tamanho mínimo

A validação da alteração por qPCR se dá em função da razão entre o número de cópias da região de interesse e de um gene *housekeeping* do paciente em comparação com a mesma razão obtida em um indivíduo controle.

O resultado esperado para uma deleção é uma razão de 0,5, aproximadamente, enquanto que o esperado para uma duplicação é uma razão de 2, aproximadamente.

Com exceção da deleção identificada no indivíduo 10291-1, envolvendo o gene *DPP6*, que foi validada (GRÁFICO 1), as 8 restantes não foram validadas por qPCR (exemplo no GRÁFICO 2), o que corresponde a 88,9 % das alterações selecionadas.

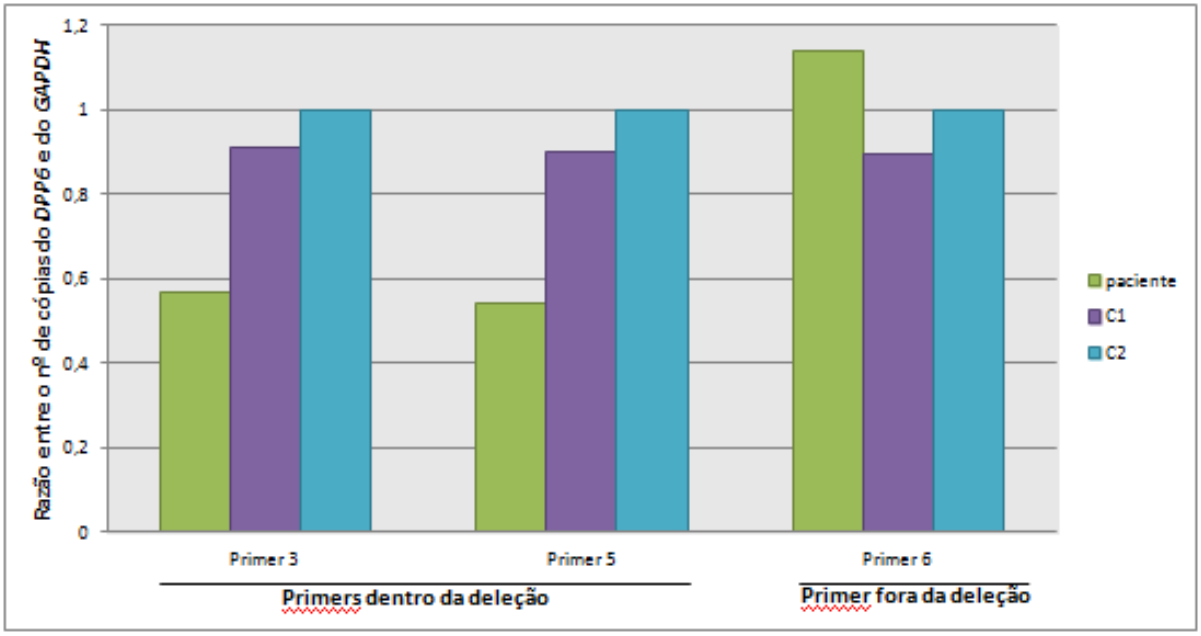


GRÁFICO 1. 5.RAZÃO ENTRE O NÚMERO DE CÓPIAS DO GENE *DPP6* E DO *GAPDH* DO PACIENTE 10291-1, EM COMPARAÇÃO COM A MESMA RAZÃO OBTIDA EM UM INDIVÍDUO CONTROLE.

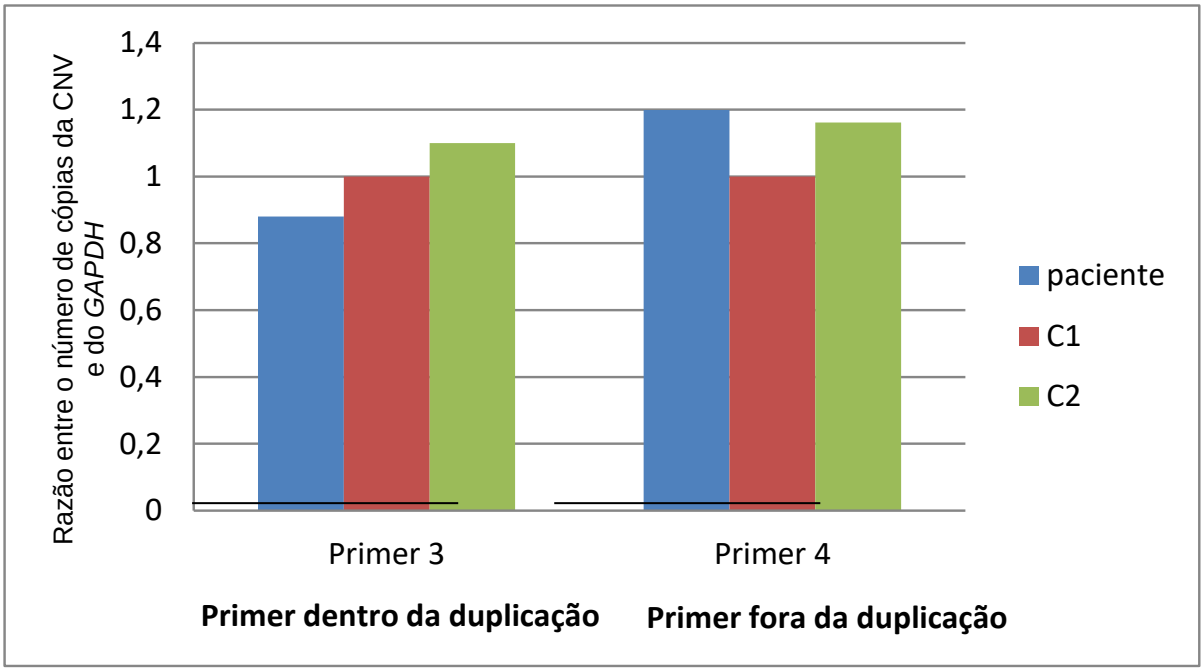


GRÁFICO 2. RAZÃO ENTRE O NÚMERO DE CÓPIAS DA DUPLICAÇÃO DETECTADA NO PACIENTE 1389-1, REGIÃO 1p13.3, E DO *GAPDH*, EM COMPARAÇÃO COM A MESMA RAZÃO OBTIDA EM UM INDIVÍDUO CONTROLE

Estes resultados falso positivos podem estar vinculados aos critérios adotados para a seleção das alterações, que talvez não tenham sido suficientemente rigorosos para a análise dos dados obtidos com o novo array customizado e/ou a qualidade da hibridação.

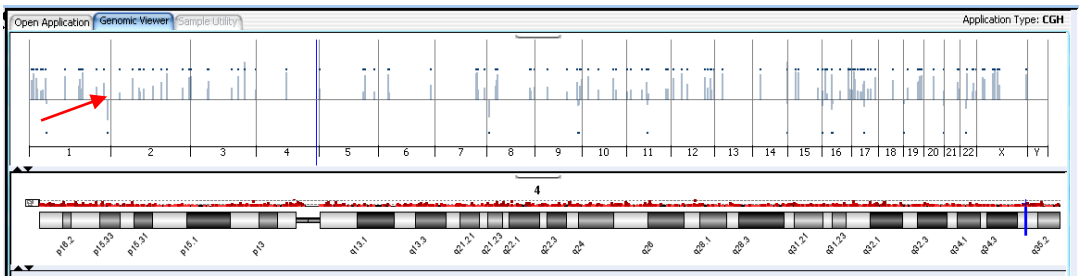
Assim, o critério de seleção de alterações foi reconsiderado, passando a serem selecionadas apenas alterações com o valor matemático de log2 para duplicação, equivalente a 0,58, somado ao valor de *DLRS* (Derivative Log Ratio Spread) (valor gerado para a área na qual o paciente foi hibridado), e valor matemático de log2 para deleção, equivalente a -1, somado ao *DLRS*. O *DLRS* é um importante parâmentro de qualidade do aCGH e indica a média da variância da razão log2 de sonda a sonda ao longo dos cromossomos.

4.2.2 Parâmetros de Controle de Qualidade e Eficiência da Hibridição na Lâmina de aCGH 180K

A hibridação inicial de 38 pacientes com TEA para que fosse realizada a validação preliminar da lâmina de aCGH de 180K apresentou uma eficiência reduzida.

A baixa eficiência se caracterizou por áreas do array com *rain drop* (quantidade excessiva de CNVs) e/ou parâmetros alterados de controle de qualidade, que tinham que ser excluídas da análise (Ex. na FIGURA 5).

A



B

IsGoodGrid	1.00	>1	NA	<1
AnyColorPrcntFeatNonU...	0.01	<1	1 to 5	>5
DerivativeLR_Spread	0.33	<0.20	0.20 to 0.30	>0.30
gRepro	0.27	0 to 0.05	0.05 to 0.20	<0 or >0.20
g_BGNoise	12.22	<5	5 to 15	>15
g_Signal2Noise	0.99	>100	30 to 100	<30
g_SignalIntensity	12.10	>150	50 to 150	<50
rRepro	0.26	0 to 0.05	0.05 to 0.20	<0 or >0.20
r_BGNoise	21.66	<5	5 to 15	>15
r_Signal2Noise	1.02	>100	30 to 100	<30
r_SignalIntensity	22.08	>150	50 to 150	<50
RestrictionControl	-1.00		0.80 to 1	<0.80 or >1

• Excellent • Good • Evaluate

FIGURA 5. REPRESENTAÇÃO DE HIBRIDAÇÃO COM BAIXA EFICIÊNCIA EXCLUÍDA DA ANÁLISE. (A) ÁREA DO ARRAY COM *RAIN DROP*, INDICADO POR SETA VERMELHA E (B) PARÂMETROS ALTERADOS DE CONTROLE DE QUALIDADE, CLASSIFICADOS EM SUA MAIORIA COMO "AVALIAR".

Os parâmetros de qualidade foram gerados pelo software *Feature Extraction* (*Agilent Techonologies* - v10.7) e representam medidas da qualidade geral dos dados obtidos, classificados como excelentes, bons e ruins e descritos em sua maioria na Tabela 4.

TABELA 4. PARÂMETROS DE CONTROLE DE QUALIDADE DO aCGH

Parâmetro	Excelente	Bom	Ruim
DLR_Spread	<0,20	0,2- 0,3	>0,3
Reproducibility	<0,05	0,05 - 0,2	>0,2
Background Noise	<5,00	05/out	>10
Signal to Noise	>100	30 - 100	<30
Signal Intensity	>150	50 - 150	<50

Fonte: Guia de Referência do Software *Feature Extraction* (v10.7).

Entre os parâmetros que se mostravam ruins com alguma frequência, estavam o *DLRS*, *background noise* (*BG noise*) e *signal to noise.* O *BG noise* é caracterizado pelo desvio padrão dos sinais de *background* das sondas calculado a

partir dos pontos de controle negativo da lâmina. Já o s*ignal to noise* é calculado dividindo a intensidade do sinal (*signal intensity* – medida média bruta de intensidade por canal) pelo *background noise* e indica o quão claramente a intensidade de fluorescência de cada canal pode ser detectada acima do sinal de *background*.

O primeiro teste realizado no protocolo de aCGH utilizado foi feito com o objetivo de otimizar a eficiência da lâmina, e se deu com a avaliação dos efeitos da fragmentação sobre o *DLRS*, que representa o parâmetro de qualidade geral mais importante do array.

A fragmentação inicialmente adotada tanto na lâmina 60K quanto para a validação da lâmina de 180K é a por temperatura, na qual as amostras são submetidas por 10 minutos a 98ºC, com um rápido intervalo aos 7 minutos para a adição do *random primer*.

Os padrões de fragmentação obtidos para cada amostra foram observados através de um gel de agarose. Com esse teste inicial foi possível observar que a fragmentação por temperatura apresenta grande variabilidade de amostra para amostra, bem como quando comparamos a mesma amostra fragmentada por 7 minutos, 7 + 3 minutos e 10 minutos direto (FIGURA 6). Além disso, algumas amostras não chegam a fragmentar mesmo submetidas a 98ºC por 10 minutos ininterruptos.

Com essa observação e sabendo da importância de se parear as amostras de acordo os tamanhos de fragmentos gerados para aumentar a eficiência de hibridação (CRAIG., *et al*, 2012) optou-se por avaliar o resultado, tanto no gel de agarose como no array, de se hibridar amostras fragmentadas com enzima.

Com o gel de agarose foi possível observar que amostras submetidas a fragmentação enzimática apresentam fragmentos de tamanhos muito similares entre si (FIGURA 6).

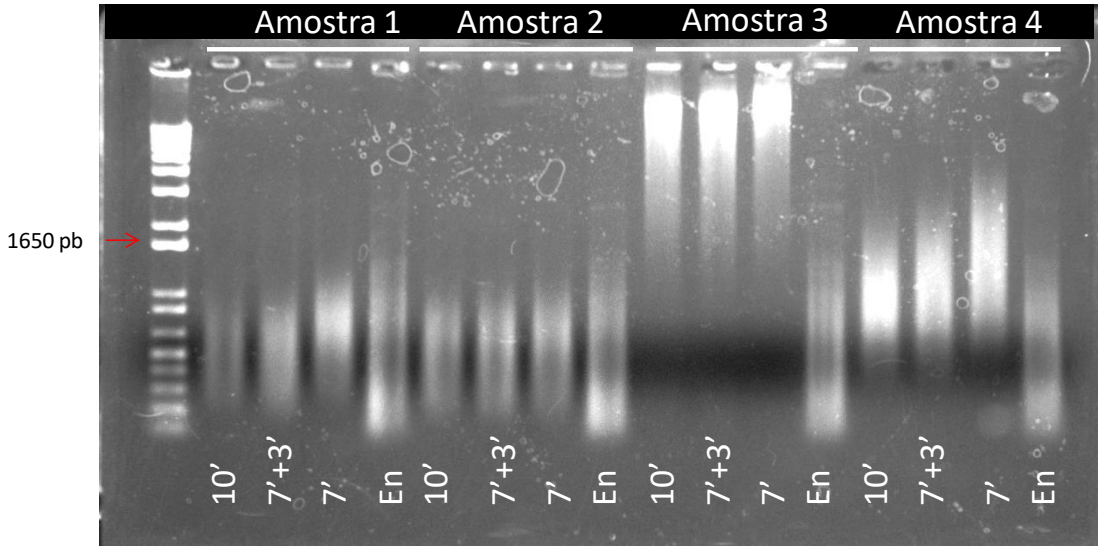


FIGURA 6. ELETROFORESE EM GEL DE AGAROSE (2,5 %) MOSTRANDO 4 AMOSTRAS DISTINTAS DE PACIENTES COM TEA FRAGMENTADAS POR 7 MINUTOS, 7 + 3 MINUTOS, 10 MINUTOS DIRETO E ATRAVÉS DE ENZIMA. A FLECHA VERMELHA INDICA O FRAGMENTO DE REFERÊNCIA DO PESO MOLECULAR DE 100 PB (INVITROGEN). OBS.: EN = FRAGMENTAÇÃO ENZIMÁTICA

Com isso, a fragmentação enzimática nos pareceu mais adequada, tendo sido adotada para as hibridações que se seguiram.

Foram realizadas 32 hibridações com amostras fragmentadas através da ação enzimática e em todas elas o *DLRS* passou na classificação indicada como boa ou excelente pela *Agilent*.

A melhora dos parâmetros de *background noise* e *signal to noise* por sua vez foi obtida com a utilização de um procedimento mais longo de lavagem das lâminas após a hibridação.

No protocolo original as lâminas eram submetidas a uma primeira lavagem com *Agilent Oligo aCGH Wash Buffer 1* por cinco minutos *e* a uma segunda com *Agilent Oligo aCGH Wash Buffer 2* por um minuto. Esse protocolo foi alterado de forma que as lâminas fossem lavadas duas vezes com *Wash Buffer* 1, ambas por cinco minutos e só depois passassem para o *Wash Buffer* 2 por um minuto.

4.2.3 Análise dos Dados de aCGH 180K de Sessenta e Três Pacientes

Esta etapa teve como foco a análise de 63 pacientes, incluindo os 31 utilizados na segunda etapa de validação da lâmina de 180K e que foram reanalisados. Neste contexto, os critérios de seleção de CNVs utilizados foram os definidos após a validação. A seleção conservadora apenas de alterações maiores que 1Kb foi mantida.

Destes 63 pacientes (52 casos esporádicos e 11 casos familiais), 49 são homens e 14 mulheres, sendo 59 não sindrômicos e 4 sindrômicos. Entre o número total de indivíduos 7 se encontram entre os que também foram analisados com o array 60K, porém sem resultados positivos.

Foram identificados 8 pacientes com 7 CNVs potencialmente patogênicas entre os 63 analisados (TABELA 5), o que corresponde a uma prevalência de 12,7% (8/63). Destes, 5 são casos esporádicos e 3 são familiais.

Com a análise sob o novo critério, apenas 2 das 12 CNVs descritas na Tabela 3 foram mantidas, a envolvendo o *DPP6* (7q36.2) e a envolvendo o *ZNF630* e *SSX6* (Xp11.23). A CNV envolvendo o *DPP6* foi validada por qPCR (FIGURA 4, disponível na sessão 4.2.1 deste trabalho) e a envolvendo o *ZNF630* e *SSX6* foi detectada através do software NextGene de análise de sequenciamento de exoma, (FIGURA 7).

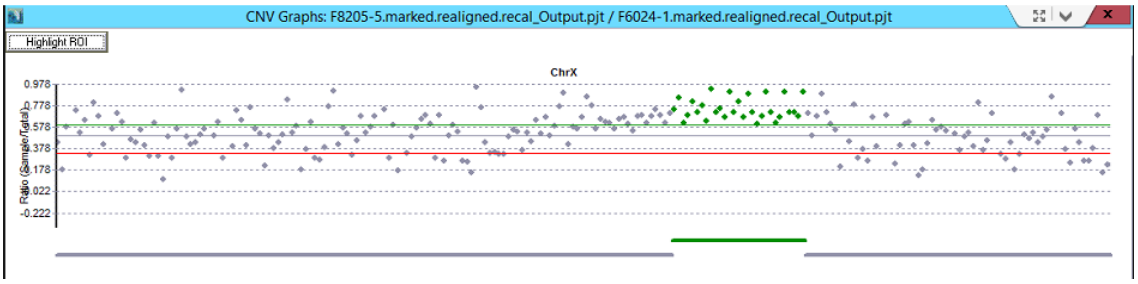
TABELA 5. RESULTADO DA ANALISE DO aCGH 180K DE 63 PACIENTES AFETADOS PELO TRANSTORNO DO ESPECTRO AUTISTA

Paciente	Sexo	Fenótipo	Posição da CNV	Tipo	Origem	Tamanho (Kb) Min (Max)	Posição (Hg19)*	Gene(s) envolvidos*
10301-1	F	ñ sindrômico	10q21.3-q22.1	Dup	ñ avaliada	105,8(377,3)	68275890-68381688	*CTNNA3*
7936-1	M	ñ sindrômico	15q13.2	Dup	ñ avaliada	70,2(807,4)	30871368-30941572	*ULK4P1, ULK4P2, GOLGA8H,* ARHGAP11B, LOC100288637
8205-5	**M**	ñ sindrômico	Xp11.23	Del	ñ avaliada	66(151,1)	47905537-47971578	*ZNF630, SSX6*
8205-6	**M**	ñ sindrômico	Xp11.23	Del	ñ avaliada	66(151,1)	47905537-47971578	*ZNF630, SSX6*
5616-1	M	ñ sindrômico	16p12.1-p11.2	Dup	ñ avaliada	46,9(95)	28699761-28746736	*EIF3CL, EIF3C*
5979-1	M	ñ sindrômico	2q31.1	Dup	ñ avaliada	2,7(2,9)	173420594-173423407	*PDK1*
c26025	M	sindrômico	12q24.31	Dup	ñ avaliada	1,7 (5)	124809219-124807481	*NCOR2*
10291-1	M	ñ sindrômico	7q36.2	Del	ñ avaliada	4,5 (392,9)	153998331-154002788	*DPP6*

Obs: Os pacientes 8205-5 e 8205-6, marcados em negrito, são irmãos

M = masculino; F = feminino; ñ = não; dup = duplicação; del = deleção

*a posição da CNVs e os genes envolvidos nelas são referentes ao seu tamanho mínimo



Sample	F8205-5.marked.realigned.recal_Output.pjt								
Control(s)	F6024-1.marked.realigned.recal_Output.pjt								
Comparison Type:Single Control									
Index	**Chr**	**Chr Start**	**Chr End**	**Gene**	**Length**	**Median Ratio**	**Median Total RPKM**	**Median Del Score**	**HMM Calls**
295	chrX	7870030	47867345	PNPLA4 ; SPACA5	39947316	0.4809	8.313	1.59	Normal
296	chrX	47917857	47970739	ZNF630 ; LOC100509575	52883	0	5.89	11.19	Deletion
297	chrX	47990070	49230924	SPACA5B ; GAGE2C	1240855	0.5364	13.573	1.37	Normal

FIGURA 7. REPRESENTAÇÃO DA CNV ENCONTRADA NO PACIENTE 8205-5, NA REGIÃO XP11.23, USANDO O SOFTWARE NEXTGENE

Dentre as 7 CNVs, 4 foram classificadas como pequenas o que corresponde a 57,1% das CNVs. Não foi realizado o cálculo de prevalência para o grupo de sindrômicos e não sindrômicos isoladamente devido ao número reduzido de pacientes sindrômicos nesta amostra.

Identificou-se uma sobreposição entre as regiões genômicas de duas alterações, que são idênticas, localizadas na posição Xp11.23, e foram identificadas em dois irmãos, pacientes 8205-5 e 8205-6.

Identificou-se uma taxa de detecção de CNVs em homens *versus* mulheres de 14,3 % e 7,1 %, equivalentes a 7/49 e 1/14, respectivamente ($p = 0,48$, $x^2 = 0,50$).

No que diz respeito à comparação entre o tamanho das CNVs encontradas em homens *versus* mulheres, a média de tamanho das CNVs nos homens equivaleu a 31,9 Kb (d.p.=32,8) enquanto a única CNV detectada na única mulher apresentou um tamanho de 105,8 Kb. Nenhum paciente apresentou mais de uma CNV.

A maior parte das alterações foi caracterizada como duplicação, somando um total de 71,4% das alterações (5 entre o total de 7 CNVs). Para essas análises não foram realizadas hibridações dos pais dos pacientes, o que impossibilitou saber a origem das alterações.

A média de tamanho das alterações encontradas foi de 42,5 Kb (d.p.=40,9), com as duplicações apresentando uma média de 45,5 Kb (d.p. = 44,7) e as deleções uma média de 35,3 Kb (distribuídas entre 4,5 Kb e 66 Kb).

Entre as 3 CNVs grandes encontradas, 2 envolvem mais de um gene, enquanto a terceira, presente no paciente 10301-1, de 105,8 Kb envolve apenas o exon 1 do gene *CTNNA3.*

As demais CNVs, totalizando 4, são todas pequenas. Três afetam um único gene, a primeira, no paciente 5979-1, de 2,7 Kb, afeta apenas o exon 1 do gene *PDK1*, a segunda, no paciente c26025, de 1,7 Kb, afeta apenas o exon 49 do gene *NCOR2* e a terceira, presente no paciente 10291-1, afeta o exon 1 do gene *DPP6*.

A CNV pequena restante, identificada no paciente 5616-1, está localizada na região 16p12.1-p11.2, apresenta 46,9 Kb e envolve os genes *EIF3CL, EIF3C.*

4.3. ANÁLISE DOS DADOS DE *NEXT GENERATION SEQUENCING* PARA DETECÇÃO DE CNVS

A detecção de 2 CNVs através dos dados de *NGS*, uma pequena envolvendo o gene *SLC17A6*, detectada inicialmente pelo aCGH 60K e outra grande envolvendo o gene *ZNF630* e *SSX6*, detectada inicialmente pelo aCGH 180K, fez com que a metodologia de *NGS* fosse utilizada neste trabalho para fazer a triagem de CNVs a fim de testar o seu potencial em ser complementar aos resultados obtidos com o array.

Verificou-se também a possibilidade do *NGS* em melhorar a delimitação de pontos de quebra das CNVs detectadas pelo aCGH, quando localizadas em regiões exônicas, bem como de auxiliar e corroborar os resultados obtidos a partir do aCGH.

Como objetivo principal buscou-se verificar a possibilidade de migração de tecnologia, do aCGH para o *NGS,* que apresenta dados mais robustos.

Com esse pressuposto, para a integração dos dados de aCGH e *NGS* foi usado como referência a metodologia de aCGH, já bem estabelecida para CNV, e os softwares *NextGene (SoftGenetics*, LLC - State College, PA) e XHMM (eXome-

Hidden Markov Model, Fromer *et al*., 2012 - Mount Sinai School of Medicine, NY) para análise dos dados de exoma.

Ambos os softwares utilizam informações a respeito da cobertura, medida em RPKM (10^9 x C / N x L; C: Número de *reads* mapeados na região, equivalente ao exon; N: Número total de *reads*; L: Tamanho da região), para fazer a chamada de CNVs através do modelo de *Hidden Markov.*

A amostra utilizada nessas análises foi composta por 9 indivíduos propósitos, todos escolhidos entre os sessenta e três hibridados na lâmina de aCGH 180K e que também possuíam dados de exoma. Entre estes, oito são casos esporádicos e um familial, sendo seis homens e três mulheres, todos não sindrômicos.

A sobreposição dos dados de *NGS* com os de aCGH foi feita levando em consideração o tamanho máximo de todas as CNVs que envolvessem regiões exônicas detectadas pelo array para os nove indivíduos, incluindo todas as polimórficas e raras. Com isso pretendeu-se obter uma informação mais completa a respeito do potencial de detecção de CNVs por ambos os softwares de análise de *NGS*.

No total foram detectadas pelo aCGH 21 CNVs que envolvem regiões exônicas nos 9 pacientes, 9 duplicações e 12 deleções. O tamanho médio destas alterações foi de 308,7 Kb (d.p. = 427,5) , sendo 17 delas polimórficas e 4 raras. Entre as raras, apenas 1 foi classificada como potencialmente patogênica.

Em comparação com estas 21 CNVs, a taxa de sobreposição do software XHMM foi de 52,4 % (sobreposição de 11 CNVs) e a do NextGene de 42,9% (sobreposição de 9 CNVs) - resultados estão disponíveis na Tabela 6, Anexo III.

O XHMM detectou 4 duplicações e 7 deleções, enquanto o NextGene detectou 7 deleções e 2 duplicações. Apesar da pequena amostra, observou-se uma proporção semelhante à do aCGH de deleções *versus* duplicações apenas no XHMM. Ambos os softwares chamaram alteração de tipo contrário ao encontrado no aCGH. Enquanto o aCGH chamou uma deleção, o XHMM detectou uma CNV sobreposta a essa do tipo duplicação, assim como o NextGene.

Com exceção de 1, todas as CNVs chamadas pelo aCGH nestes 9 pacientes foram grandes, > 50 Kb, distribuídas entre 48,7 Kb até 1876,7 Kb. A menor CNV foi detectada apenas pelo NextGene, enquanto a maior foi detectada apenas pelo

XHMM. O XHMM se sobrepôs a alterações com uma média de tamanho de 412,4 Kb (d.p.=542,6)e o NextGene a alterações com uma média de 255,7 Kb (d.p.=327,7), sem diferença significativa (p=0,22) (TABELA 6). No entanto, estes resultados sugerem uma possível correlação entre tamanho da CNV e o software que o detecta, embora não significativa estatisticamente, com o NextGene parecendo ser mais eficiente para detecção de CNVs pequenas (r = -0,29; p = 0,22)

A taxa de deleção-duplicação por paciente também foi comparada entre o aCGH e ambos os softwares, conforme descrito no Gráfico 3. Para o aCGH não houve um padrão entre quantidade de duplicações e deleções por paciente. Para o XHMM pôde ser observado uma maioria de pacientes com duplicações, equivalente a seis (66,7%), e uma minoria com mais deleções, equivalente a três (33,3%). Com o *NextGene* por sua vez, os resultados foram enviesados para deleção, todos os 9 apresentaram mais deleções que duplicações.

A taxa de deleção-duplicação não foi congruente entre o aCGH e *NGS* (NextGene e o XHMM) para nenhum dos pacientes

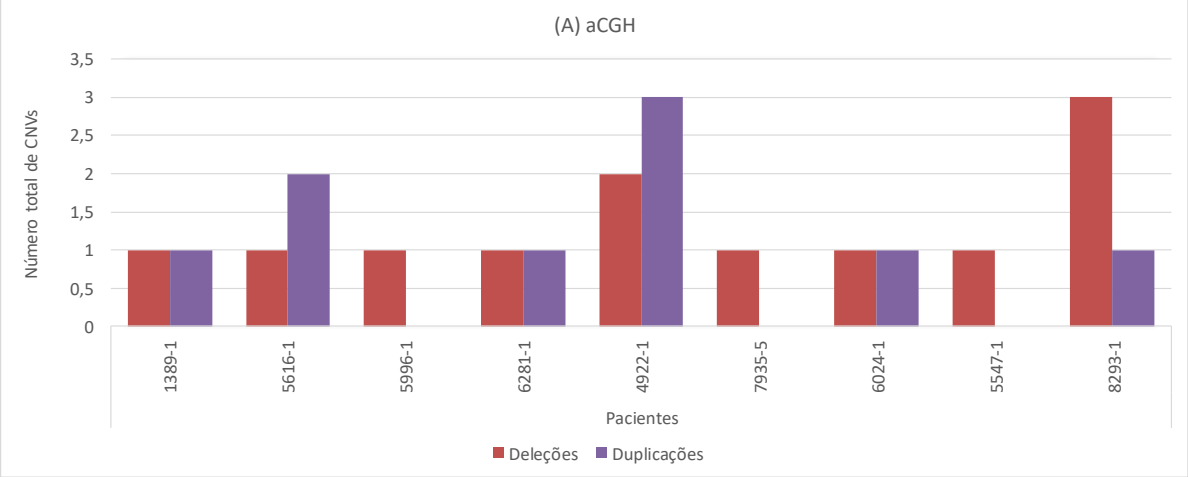


GRÁFICO 3. BOXPLOT DA QUANTIDADE DE DUPLICAÇÕES E DELEÇÕES DETECTADAS EM CADA UM DOS 9 PACIENTES ATRAVÉS DO (A) aCGH (B) *XHMM* E (C) *NEXTGENE.*

Foi realizada também a comparação apenas entre o XHMM e o NextGene para esses 9 pacientes, descrita na Tabela 7 (ANEXO IV), com a finalidade de verificar se haveria um número considerável de CNVs em comum, visto que a quantidade de alterações chamadas através da análise dos dados de sequenciamento de exoma é superior ao que encontramos por aCGH.

O XHMM gerou em média 95,7 CNVs por paciente, com aproximadamente 56,3 % delas sendo duplicações, com tamanho médio de 185,7 Kb (d.p.=81,4), e 43,7 % deleções, com tamanho médio de 192,1 Kb (d.p.=88,9). O NextGene gerou em média 749,8 CNVs por paciente, valor significativamente maior do que o gerado pelo XHMM (p = 0,0001) e, ao contrário do XHMM, 64,8 % foram deleções e o restante, 35,1 %, duplicações, com tamanho médio de 66,1 Kb (d.p.=103,3) e 22,3 Kb (d.p.=20,04), respectivamente.

O tamanho médio das duplicações detectadas pelo NextGene *versus* XHMM e das deleções detectadas pelo NextGene *versus* XHMM foi significativamente diferente, com p = 0,0001 e p = 0,0028, respectivamente.

Entre o total de alterações detectadas pelo XHMM, em média 25 % também é detectada pelo NextGene; já na análise contrária, entre o total das alterações chamadas pelo NextGene, em média apenas 8 % também é chamada pelo XHMM. As deleções apresentaram, em ambas as comparações, XHMM *versus* NextGene e NextGene *versus* XHMM, uma taxa um pouco maior de sobreposição do que as duplicações, conforme pode ser observado na Tabela 7.

4.4 ANÁLISE POR *NGS* DE CNVS NO GENE *SLC17A6*

Os resultados obtidos por aCGH 60K e MLPA para o *SLC17A6*, ainda não associado ao TEA, suscitaram dúvidas quanto a patogenicidade das CNVs neste gene, visto que a prevalência obtida em pacientes *versus* controles não foi muito diferente, de 0,003% e 0,002%, respectivamente. Por isso, realizamos análises adicionais nesse gene através da metodologia de *NGS*.

Nesse sentido, foram realizadas duas análises diferentes para a identificação de CNVs no *SLC17A6.* A primeira realizada em uma amostra de pacientes norte americanos com TEA e a segunda em uma amostra de 164 controles brasileiros.

A primeira análise, feita durante o período do BEPE, foi realizada em uma amostra independente de 5084 indivíduos com TEA, de quatro bancos de dados diferentes (*The Autism Genetic Resource Exchange, Simons Simplex Collection, Early Markers for Autism Study*, dados da pesquisadora Dalila Pinto proveniente da *Icahn School of Medicine- Mount Sinai*), disponíveis para o laboratório da Dra. Weiss. Os dados disponíveis desses pacientes são todos de sequenciamento de exoma.

Nessa análise, 2 CNVs grandes foram encontradas envolvendo, entre outros genes, o *SLC17A6*, cada uma em um paciente não aparentado entre si, ambas de origem materna e também presentes nas irmãs, todas descritas na Tabela 8. Com exceção dos probandos, todos os outros familiares não apresentam TEA. Os 2 probandos apresentam outras alterações raras, o primeiro outra duplicação herdada

da mãe na região 10p15.3-p15.2, de 1,1 Mb e o segundo uma duplicação herdada do pai, na região 12p11.21 de 52,3 Kb (TABELA 8).

TABELA 8. RESULTADO DA ANÁLISE DE *NGS* PARA O GENE *SLC17A* 6 EM 5084 PACIENTES AFETADOS PELO TEA

Paciente	Sexo	Fenótipo	Posição da CNV	Tipo	Origem	Tamanho (kb)	Posição (Hg19)	Gene(s) envolvidos
1	M	ñ sindrômico	11p14.3	dup	materna	4932,1	21802890-26734993	*ANO5,LUZP2,GAS2,SLC5A12, MUC15,ANO3,FANCF,CCDC17 9,SVIP,SLC17A6*
			10p15.3-p15.2	dup	materna	1000,1	1972527 - 3076486	*LOC282980;LOC399708*
2	M	ñ sindrômico	11p14.3	del	materna	1714,4	21870794 - 23585198	*ANO5,GAS2,FANCF,CCDC179 ,SVIP,SLC17A6*
			12p11.21	dup	paterna	52,3	31892782 - 31954269	*H3F3C*

M = masculino; ñ = não; dup = duplicação; del = deleção

A primeira CNV detectada, presente no paciente 1, proveniente dos dados analisados durante o BEPE, foi validada na UCSF pela Dra. Erika Yeh através de PCR quantitativa (GRÁFICO 4).

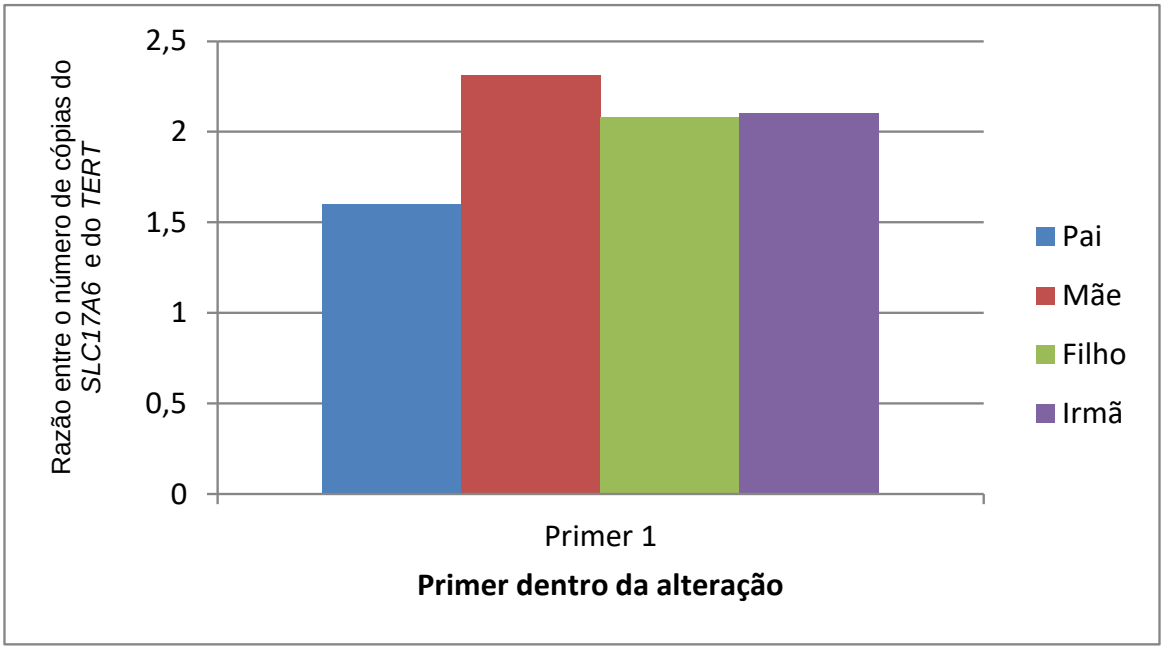


GRÁFICO 4. RAZÃO ENTRE O NÚMERO DE CÓPIAS DO GENE *SLC17A6* E DO *TERT* DO PACIENTE 1, EM COMPARAÇÃO COM A MESMA RAZÃO OBTIDA PARA SUA MÃE, PAI E IRMÃ.

A segunda análise foi realizada em um grupo inicial de 100 controles brasileiros provenientes do projeto 60+ e em uma amostra de 64 pacientes com diferentes quadros clínicos, provenientes do CEGH-CEL, e passíveis de serem usados como controles, pois apresentam doenças Mendelianas sem sobreposição clinica ou etiológica ao TEA. Os dados disponíveis desse grupo amostral são de *NGS.*

Nesta análise foi encontrado 1 paciente de 9 anos, do sexo masculino, diagnosticado com disostose espôndilo-costal e sem comprometimentos neurológicos, com uma duplicação nos exons 9 a 12 do *SLC17A6.* Esta alteração foi validada pelo aCGH customizado 180K (FIGURA 8).

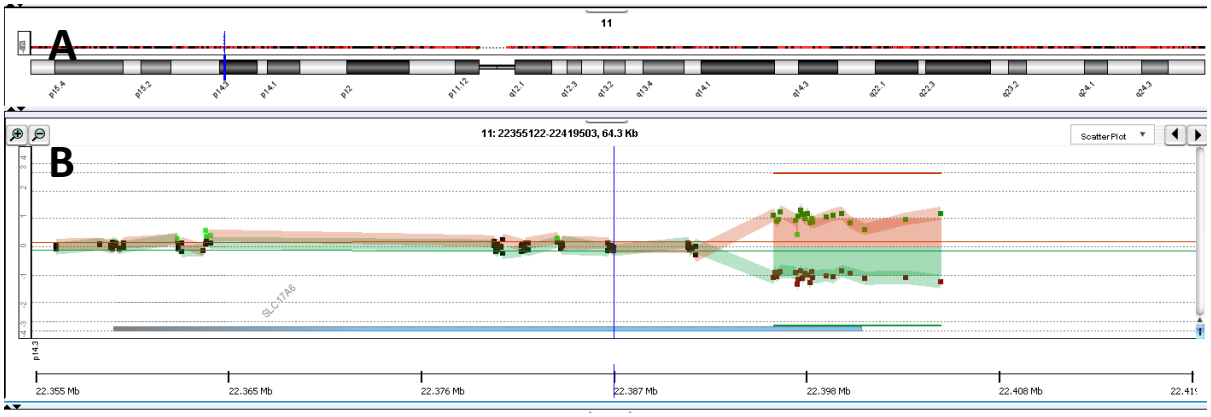


FIGURA 8. DUPLICAÇÃO NOS EXONS 9 A 12 DO GENE *SLC17A6*. A) POSIÇÃO DA MUTAÇÃO NO CROMOSSOMO 11 E B) EIXO X = SONDAS DE ACORDO COM A POSIÇÃO GENÔMICA, EIXO Y = VALOR DO log2 *cy3/cy5*. FAIXA EM VERMELHO INDICA A ÁREA ONDE O DNA DO PACIENTE ESTÁ MARCADO COM *cy3*, E FAIXA EM VERDE INDICA A ÁREA ONDE O DNA DO PACIENTE ESTÁ MARCADO COM *cy5.*

Agrupando todos os dados obtidos referentes ao *SLC17A6*, ficou-se com um total de 2 homens com TEA em um grupo amostral total de 505 pacientes brasileiros e 2 homens sem TEA, um proveniente do grupo de 350 controles brasileiros e o segundo proveniente do grupo de 64 pacientes com doenças que não apresentam sobreposição ao TEA.

Agrupando os dados obtidos nos Estados Unidos, referentes ao *SLC17A6*, ficou-se com um total de 2 homens com TEA em um grupo amostral total de 5084 pacientes.

Com isso a prevalência de CNVs envolvendo o *SLC17A6* em pacientes com TEA é de 0,0007%, ou quatro em 5589 pacientes (isolando o grupo de pacientes brasileiros e norte americanos a prevalência se modifica para 0,004%, considerando 2 afetados em um grupo amostral de 505 pacientes e 0,0004%, considerando 2 afetados em um grupo amostral de 5084 pacientes, respectivamente) *versus* 0,003%, dois controles em um total de 714 controles brasileiros. Ambas às comparações, 0,0007% *versus* 0,003% ou 0,004% *versus* 0,003% ($p = 0,12$; $x^2 = 2,3$ e $p = 1$, respectivamente) não apresentam diferença significativa estatisticamente.

5. DISCUSSÃO

Este trabalho foi realizado primordialmente em um grupo amostral de 178 pacientes brasileiros com TEA provenientes do Centro de Pesquisa sobre o Genoma Humano da USP. Como objetivo principal este trabalho pretendeu avaliar a importância de pequenas CNVs/InDels raras na etiologia do Transtorno do Espectro Autista.

Para essa finalidade foram utilizadas a metodologia de aCGH customizado 60 e 180K e de sequenciamento de nova geração de exoma.

5.1 aCGH CUSTOMIZADO

5.1.1 aCGH 60K

A hibridação inicial de um número reduzido de pacientes com este array teve como objetivo verificar o seu potencial em detectar InDels grandes (70 pb – 1Kb). Neste sentido, as deleções selecionadas para validação foram classificadas como falso-positivas, provavelmente em decorrência do excesso de sobreposição entre as sondas, muitas vezes sobrepostas quase que por completo, o que acaba gerando ruído. Em decorrência desta limitação da customização do aCGH 60K, acabamos sendo mais conservadoras nas análises dos dados, descartando todas as alterações menores que 1Kb.

As análises seguiram com a hibridação de 98 pacientes com TEA. Dez CNVs potencialmente patogênicas foram detectadas com o aCGH 60K customizado. Destas, 20 % se caracterizaram como pequenas, menores que 50 Kb (ou 40 % considerando apenas os casos não sindrômicos), sugerindo que pequenas CNVs desempenham um papel importante na etiologia do TEA (MOREIRA *et al.*, 2016 – ANEXO V).

A prevalência de pacientes com CNVs potencialmente patogênicas obtida com esta análise foi de 9%.

Este resultado mostrou-se similar ao encontrado na literatura, o qual associa CNVs raras a aproximadamente 10% dos casos de TEA (DEVLIN e SCHERER, 2012; PINTO *et al.*, 2010). Neste estudo, contudo, a amostra foi submetida a uma

pré-triagem, com a finalidade de excluir os pacientes com as CNVs mais prevalentes no TEA, nas regiões 15q11-q13, 16p11.2 e 22q13.3.

A prevalência de CNVs nestas regiões *hotspot*, em pacientes brasileiros com TEA, foi estimada em 2,7%, no trabalho de Moreira em 2014. Soma-se a isto o fato do nosso array cobrir apenas 269 genes candidatos, contrastando com outros estudos que utilizaram arrays comerciais que contém sondas ao longo de todo o genoma e ao atual número de genes candidatos ao TEA disponíveis no *SFARI*, equivalente a 845.

Apesar da pequena amostra, a prevalência de CNVs obtida também foi analisada de forma a isolar diferentes grupos, tais como sindrômicos *versus* não-sindrômicos e homens *versus* mulheres.

Os pacientes sindrômicos, de uma forma geral, apresentaram uma maior prevalência de CNVs associada a CNVs maiores do que os não-sindrômicos. Esta tendência está de acordo com o esperado e descrito em outros trabalhos, já que um número maior de alterações, bem como alterações grandes tem o potencial de afetar múltiplos genes podendo, possivelmente, gerar um maior impacto ao longo do desenvolvimento embrionário (MARSHALL *et al*. 2008; BUCAN *et al*., 2009; GIRIRAJAM *et al*., 2013).

Os pacientes do sexo masculino e feminino apresentaram uma prevalência similar de CNVs. Entre estes, apenas uma mulher apresentou mais de uma CNV potencialmente patogênica, modelo oligogênico do TEA. Além disso, as pacientes mulheres também apresentaram uma média de tamanho das CNVs maior do que a dos homens.

Esses resultados parecem seguir uma tendência similar aos resultados de outros trabalhos, que mostram que as mulheres com TEA tendem a ter uma carga maior de mutações e/ou CNVs maiores que os homens com TEA, sugerindo que elas apresentam, de alguma forma, um “modelo protetivo”, necessitando uma carga etiológica maior do que os homens para manifestar o fenótipo de TEA (ITSARA *et al.*, 2010; SANDERS *et al.*, 2011; KRUM *et al*.,2013; JACQUEMONT *et al*., 2014; DE RUBEIS e BUXBAUM, 2015; DESACHY *et al.,* 2015;).

Porém, devido à composição do grupo amostral, que combina pacientes com TEA sindrômicos com não sindrômicos, que talvez apresentem mecanismos etiológicos diferentes, bem como o pequeno número amostral, não é possível fazer maiores inferências acerca da quantidade e tamanho médio da CNVs nos homens *versus* mulheres.

No geral mais de 50% das alterações foram classificadas como deleções, enquanto as duplicações mostraram-se maiores, em termos de tamanho, do que as deleções. Deleções, de uma maneira mais generalizada, apresentam uma maior probabilidade de serem patogênicas do que duplicações (CANLI, 2015; REVENGA *et al.*, 2007).

Para a maioria das alterações detectadas não foi possível determinar a sua origem. Entre as 4 nas quais foi possível, 3 eram herdadas e 1 *de novo*. A correlação entre alterações *de novo* e o TEA se dá de forma mais direta, com essas apresentando geralmente um maior potencial de patogenicidade, visto que essas são submetidas a uma seleção evolutiva menos rigorosa (MARSHALL *et al*., 2008; KRUMM *et al*, 2013; POULTNEY *et al*., 2013; DE RUBEIS e BUXBAUM, 2015). No entanto, as variações herdadas também estão implicadas nos modelos genéticos atuais de TEA (DE RUBEIS e BUXBAUM, 2015). Trabalhos recentes sugerem um papel importante para as pequenas CNVs raras e herdadas (KRUMM *et al* 2013; KRUMM *et al*, 2015).

Com relação as oito CNVs grandes que foram detectadas, 4 (identificadas nos pacientes 5502-1, 3536-1, 4174-1 e 6507-1) envolvem mais de 15 genes e estão em regiões já descritas como associadas ao TEA no Decipher (https://decipher.sanger.ac.uk/) e/ou no SFARI. Uma delas inclusive (no paciente 5502-1), está localizada em uma região que é um alvo frequente de CNVs (CNVRs) em pacientes com TEA.

Das 4 CNVs grandes restantes, 2 envolvem mais de um gene (pacientes 6098-1 e 4174-1), ambas localizadas em regiões já associadas ao TEA. Na primeira, paciente 6098-1 região 7q31.1, apenas o gene *GPR85* já foi associado ao TEA em um trabalho (PRASAD *et al*., 2012). Ele é responsável por codificar um receptor acoplado à proteína G, que pertence à família SREB (Receptor Super Conservado Expresso no Cérebro), apresentando ampla expressão em diferentes estruturas

cerebrais e predito por apresentar um importante papel na regulação da neurogênese na região hipocampal (CHEN *et al*., 2012).

Na segunda CNV grande que envolve mais de um gene, identificada no paciente 4174-1 região 14q21.2-21.3, é o gene *MDGA2* que já foi correlacionado ao TEA (BUCAN *et al*., 2009; SFARI), sendo responsável por codificar uma molécula de adesão celular da família IgCAM e aparentemente importante para a extensão axonal ao longo do eixo longitudinal (JOSET *et al*., 2011).

As duas outras alterações grandes, que envolvem apenas um gene, são uma deleção envolvendo a maioria dos exons do gene *DIAPH3*, paciente 6034-1, e a segunda, uma deleção envolvendo os exons 1 a 4 do *GRIA3,* paciente 2701-1.

O *DIAPH3* atua no remodelamento dos filamentos de actina, função essencial para a migração neuronal durante o desenvolvimento do cérebro. Quatro pacientes com TEA já foram descritos com deleções intragênicas no *DIAPH3,* no banco de dados SFARI. O *GRIA3*, por sua vez, ainda não foi associado ao TEA, mas codifica um receptor de glutamato da família AMPA, desempenhando um papel importante na transmissão sináptica excitatória, além de ter uma importante expressão no cérebro, inclusive durante os períodos iniciais do desenvolvimento (www.genecards.org).

Entre as 2 CNVs pequenas, a primeira, no paciente 4181-1, é uma duplicação envolvendo os exons 9 a 12 do gene *SLC17A6*, ou *VGLUT2*, que codifica uma das três isoformas conhecidas de transportadores vesiculares de glutamato (BERRY *et al*., 2012), responsável pela captação de glutamato nas vesículas sinápticas em terminais nervosos pré-sinápticos (MOECHARS *et al*., 2006; CHIOCCHETTI *et al*., 2014).

Devido à importância da sua função e por ainda não ter sido associado ao TEA, esse gene foi escolhido para análises subsequentes em um grupo maior de pacientes brasileiros e norte americanos e em outro de controles

Com isso chegou-se a uma prevalência de CNVs envolvendo o *SLC17A6* na amostra de pacientes brasileiros de 0,004% e de 0,0004% na de pacientes norte americanos. Enquanto a prevalência obtida para o grupo de controles foi de 0,003%. Portanto, estes dados não suportam a associação entre o TEA e mutações no *SLC17A6* e evidenciam a importância de analisar diferentes populações quando se

trata da averiguação da patogenicidade de mutações em diferentes genes, visto que a frequência da alteração pode variar de população para população.

Com relação ao aCGH 60K, a segunda alteração pequena, no paciente 1665-1, é uma duplicação e envolve os exons 4 a 7 do gene *MBD2*. Esse gene codifica uma proteína de ligação a metil-CpG e várias CNVs neste gene já foram descritas em pacientes com TEA.

Estas análises sugerem que a abordagem através do aCGH customizado 60K tem um potencial significativo em termos de melhorar a sensibilidade para a detecção de CNVs pequenas relacionadas ao TEA.

Com relação ao desenho da lâmina 60K, este apresentou limitações como o excesso de sondas sobrepostas em alguns exons, resultando em ruído e InDels falso positivos, o que limitou a nossa análise apenas para alterações maiores que 1Kb, e o excesso de espaçamento entre as sondas *backbone*.

Com base nessas descobertas, na viabilidade de custo e nas atuais diretrizes da *Agilent* para customizar a lâmina de aCGH, foi criada uma nova versão do array customizado (4x180K), com sondas abrangendo 1527 genes candidatos ao TEA.

5.2 aCGH 180K

A primeira etapa de validações apontou que 7 alterações nas regiões *hotspot*, detectadas incialmente por MLPA, foram identificadas pelo aCGH 180K. Enquanto na segunda etapa, apenas uma alteração potencialmente patogênica foi validada. A discrepância entre estes resultados possivelmente se deve aos critérios adotados para a seleção das alterações, que talvez não tenham sido rigorosos o suficiente para distinguir resultados com maior probabilidade de serem falso positivos, bem como a qualidade da hibridação, que muitas vezes apresentava parâmetros de qualidade alterados.

Com base nessas validações o critério de seleção de CNVs foi refinado. De forma que ao reanalisar os dados dos pacientes utilizados para a segunda etapa de validação, entre as CNVs selecionadas para validação por qPCR, apenas uma delas foi novamente selecionada com este novo critério. Esta também foi a única alteração validada pelo qPCR

No geral, obteve-se resultados com tendências semelhantes às da lâmina 60K. Ao total mais de 50% das CNVs potencialmente patogênicas detectadas são pequenas.

A prevalência de pacientes com CNVs potencialmente patogênicas obtidas com esta análise foi de 12,7%. Esta amostra também foi submetida a uma pré-triagem com a finalidade de se excluir os pacientes com as CNVs mais prevalentes no TEA, nas regiões 16p11.2, 15q11-q13, 16p11.2 e 22q13, o que caracteriza este resultado como acima ao que é encontrado na literatura (PINTO *et al.*, 2010; DEVLIN e SCHERER, 2012), reforçando o potencial da customização para a detecção de CNVs associadas ao TEA, principalmente as pequenas.

Da mesma forma como foi feito para a lâmina 60K, calculou-se a prevalência de CNVs de forma a isolar o grupo homens *versus* mulheres, porém não para sindrômicos *versus* não sindrômicos já que esta amostra foi composta basicamente de pacientes não sindrômicos.

A diferença entre a prevalência de CNVs obtida no grupo dos homens *versus* o das mulheres foi maior do que a obtida com os dados de 60K. Não foi possível a comparação da média de tamanho das CNVs identificadas nos homens com a das mulheres, visto que foi identificada apenas uma mulher com CNV. No entanto, pode-se observar que a média de tamanho de CNVs nos homens foi aproximadamente 16 vezes menor do que a que obteve-se com os dados de 60K.

Os tamanhos das CNVs observadas nos pacientes do sexo masculino podem estar correlacionados ao fato de que as CNVs pequenas, isoladamente, detectadas em maior proporção com o aCGH 180K, sejam mais patogênicas para pacientes do sexo masculino, enquanto que em mulheres a tendência é que se tenha ou CNVs grandes ou uma maior quantidade de alterações para que elas manifestem o fenótipo de TEA, o que vai ao encontro do “modelo protetivo” feminino (JACQUEMONT *et al*., 2014).

Do total de alterações patogênicas, mais de 50% foram caracterizadas como duplicações, com a média de tamanho das duplicações sutilmente maior do que a das deleções.

A maior parte das 7 CNVs potencialmente patogênicas encontradas, não possuem genes já descritos como associados ao TEA no Decipher (https://decipher.sanger.ac.uk/) e/ou no SFARI (http://sfari.org/).

Entre as 4 alterações grandes, duas são idênticas, encontradas em uma família *multiplex*, pacientes 8205-5 e 8205-6, sendo que os portadores da alteração são irmãos. Esta CNV é uma deleção na região Xp11.23, de 66,2 Kb, envolvendo os genes *ZNF630, SSX6.* Isto reacende uma discussão em torno do gene *ZNF630,* um *zinc finger*, no qual Lugtenberg *et al*.*(*2010), detectaram alterações em pacientes com deficiência intelectual, mas que, no entanto, não segregaram, em algumas famílias, com os casos de DI e também não apresentaram um aumento significativo em pacientes *versus* controles.

Com isso, esse achado pode trazer novos esclarecimentos com relação à correlação entre CNVs envolvendo este gene e o TEA. O outro gene envolvido na CNV, o *SXX6*, não apresenta expressão no cérebro e foi associado apenas a câncer, o que o torna pouco provável de estar associado ao quadro de autismo neste paciente.

A alteração identificada no paciente 10301-1, apesar de grande, com 105,8 Kb, afeta apena os exon 1 do gene *CTNNA3.* Este gene é um dos membros da família α-catenina, responsável por codificar αT-catenina, e apresenta um papel de grande importância na adesão celular.

Assim como os outros membros dessa família o *CTNNA3* tem participação na via canônica Wnt, importante para o desenvolvimento neuronal e função sináptica, e já caracterizada como associada ao TEA. Apesar de apresentar baixos níveis de expressão no cérebro já foi associado a casos de TEA em grupos amostrais de ancestralidade Europeia e replicado em outras amostras em estudos de associação (WEISS *et al*., 2009; WANG *et al*., 2009), CNVs raras nesses gene também foram encontradas em pacientes com TEA (LEVY *et al* ., 2011; GIRIRAJAN *et al*., 2012).

O paciente 7936-1 apresenta a quarta alteração grande, uma duplicação de 70,2 Kb que afeta cinco genes, são eles: *ULK4P1, ULK4P2, GOLGA8H, ARHGAP11B, LOC100288637.* O *ARHGAP11B* é o único já associado ao TEA em apenas um trabalho correlacionado ao modelo de múltiplos hits para o TEA (LEBLOND, *et al*.,

2012). No entanto não se deve excluir a possível participação dos outros genes presentes nesta CNV para a manifestação do fenótipo.

O gene *ARHGAP11B* codifica uma proteína ativadora de Rho-GTPase, que são cruciais na organização do citoesqueleto celular, crescimento, diferenciação, desenvolvimento neuronal e funções sinápticas. Acredita-se que esse tenha sido um gene chave no desenvolvimento e expansão evolutiva do neocórtex humano (FLORIO, *et al*., 2015).

Dentre as 4 alterações pequenas, o paciente 5616-1 apresentou uma duplicação, localizada na região 16p12.1-p11.2, de 46,9 Kb envolvendo os genes *EIF3CL, EIF3C*.

Tanto o *EIF3CL* quanto o *EIF3C* não foram associados ao TEA até o momento. Ambos os genes apresentam alta expressão no cérebro e codificam subunidades centrais do fator 3 de iniciação da tradução eucariótica (eIF-3), que é necessário para várias etapas na iniciação da síntese proteica. O complexo eIF-3 associa-se a subunidade ribossomal 40S e estimula o recrutamento de outro fatores para que se inicie a tradução, sendo também necessário para a reciclagem e desmontagem dos complexos ribossomais.

O paciente 5979-1 apresentou uma duplicação pequena de 2,7 Kb, envolvendo apenas o exon 1 do gene *PDK1*. Este gene codifica a quinase piruvato desidrogenase 1, que desempenha um papel chave na regulação do metabolismo de glicose e ácidos graxos através da fosforilação das subunidades de piruvato desidrogenase PDHA1 e PDHA2. Dessa forma a atividade da piruvato desidrogenase é inibida, acabando com a conversão de piruvato em acetil-CoA , o que acarreta na diminuição do consumo mitocondrial de oxigênio e regulação da respiração aeróbica.

Alterações no *PDK1* ainda não foram associadas ao TEA. No entanto redução na expressão desse gene foi observada em amostras da região cortical de pacientes com Alzheimer, enquanto a super-expressão de PDK1 em modelo animal com camundongo conferiu resistência a beta amiloide (Aβ) e a outras neurotoxinas tais como H2O2 (NEWINGTON *et al*., 2012).

A menor das alterações foi identificada no paciente c26025, é uma duplicação de 1,7 Kb envolvendo o exon 49 do gene *NCOR2.* Esse gene codifica um co-

repressor transcricional de certos genes alvo, que atua através de um complexo de subunidades modificando a estrutura da cromatina (https://www.ncbi.nlm.nih.gov/refseq). Assim como o *PDK1*, o *NCOR2* também não foi associado ao TEA até o momento. Tendo sido, no entanto, associado a outras doenças neuropsiquiátricas como Alzheimer e depressão.

No caso do Alzheimer, este gene foi apontado como um regulador essencial de genes diferencialmente expressos em neurônios piramidais do hipocampo de pacientes (ACQUAAH-MENSAH *et al*., 2016). Com relação a depressão o *NCOR2* foi apontado como participante na modulação do hormônio liberador de corticotropina (CRH), um fator de risco para depressão (AZEVEDO *et al*., 2016)

A última e CNV pequena, identificada no paciente 10291-1, é uma deleção de 4,5 Kb , envolvendo apenas o exon 1 do gene *DPP6* (chr7q36.2), responsável por codificar uma glicoproteína de membrana tipo II, membro da família de peptidase S9B, porém sem função de protease, que se liga a canais específicos de potássio dependente de voltagem, alterando suas atividades e propriedades biofísicas, sendo expresso predominantemente no sistema nervoso central.

Alterações estruturais em heterozigose e SNVs nesse gene já foram associadas ao TEA, à microcefalia e ao retardo mental, no Decipher. Uma deleção semelhante à encontrada em no paciente em observação, envolvendo apenas o primeiro exon do gene *DPP6*, também foi correlacionada à Síndrome de Tourette (PRONTERA *et al.*, 2014). Entretanto, é necessário que se tenha mais evidências de correlação entre esse gene e o TEA, para que de fato se possa compreender essa relação causal.

Apenas as duas CNVs identificadas no paciente 5616-1 estão em uma região considerada *hotspot* para o TEA, a 16p12.1-p11.2.

Com esses resultados finalizou-se as análises dos dados de aCGH customizado. Nenhuma das regiões com alteração, incluindo as detectadas pelo aCGH 60K e 180K, foi detectada mais de uma vez na amostra estudada, o que vai ao encontro do modelo proposto para o TEA no qual mutações em diferentes genes podem convergir em um mesmo fenótipo que caracteriza o TEA (SEBAT *et al*., 2007).

No total, somando os dados obtidos com o aCGH 60K e o 180K, foram obtidas 17 CNVs potencialmente patogênicas, que envolveram um total de 461 genes.

A pequena amostra considerada na pesquisa e a falta de acesso à origem de todas as alterações, foram fatores limitantes neste trabalho, não permitindo inferências mais robustas com poder estatístico, como correlação entre tamanho e origem das alterações, fenótipo (sindrômico *versus* não sindrômico) e tipo/tamanho das alterações, entre outras. Entretanto, o aCGH customizado mostrou-se uma estratégia promissora para elucidar mecanismos etiológicos do TEA idiopático, visto que se obteve uma prevalência relativa de CNVs maior do que a descrita na literatura com ambas as customizações.

Embora SNP-arrays com milhões de sondas estejam atualmente disponíveis a um custo similar ao do aCGH 180K customizado, estes SNP-arrays muitas vezes não conseguem detectar CNVs pequenas devido à distribuição limitada e desigual de SNPs no genoma (CELESTINO-SOPER *et al*., 2011; WISZNIEWSKA *et al*., 2014). Além disto um estudo de Haraksingh *et al*. (2011) comparando doze plataformas diferentes baseadas em quatro abordagens principais, SNP-array / CGH-array / SNP-array + CGH-array de todo o genoma e "aCGH customizado", sugeriram que este último foi o melhor para detectar CNVs conhecidas em uma amostra europeia bem caracterizada.

Com relação às CNVs pequenas, foco deste trabalho, o aCGH 180K se mostrou mais promissor em relação ao 60K, tendo detectado, no total, o equivalente a 51,7 % de CNVs pequenas, enquanto o aCGH 60K detectou 20%. Este resultado sugere que as CNVs pequenas podem contribuir para o risco de TEA em pelo menos 6,3 %, valor similar ao proposto por Poultney em 2013 de aproximadamente 7% para CNVs com tamanho variando entre 1 e 30 Kb.

A detecção de alterações pequenas facilita a identificação de genes possivelmente correlacionados ao TEA, visto que estes geralmente envolvem um número reduzido de genes.

5.3 SEQUENCIAMENTO DE NOVA GERAÇÃO DE EXOMA

O microarray, apesar de vantajoso em termos de custo e análise dos dados gerados, apresenta algumas limitações que incluem ruído de hibridação, dificuldades na detecção de pontos de quebra e cobertura reduzida do exoma. Já a tecnologia de

NGS apresenta uma maior e melhor cobertura do exoma, garantindo uma detecção mais precisa dos pontos de quebra, caso esses estejam mapeados nas regiões exônicas.

O *NGS* permite também o agrupamento de dados de variantes de bases únicas (SNVs) aos de CNVs, a partir da análise de um mesmo dado (ZHAO *et al.*, 2013; ALKAN *et al*., 2011; MEI TEO *et al*., 2012). A tendência é que se tenha a substituição da metodologia de aCGH para a de *NGS* à medida que os softwares para as chamadas das CNVs a partir de *NGS* se aprimorem.

Visto isso, e tendo em mente o fato de que a detecção de pequenas CNVs ainda representa um grande desafio, a solução imediata, e que parece a mais eficaz para ultrapassar as limitações e unir os pontos positivos de cada tecnologia, é a integração das técnicas de aCGH e *NGS* (ZHAO *et al*., 2013; ALKAN *et al*., 2011), utilizando como referência a metodologia de aCGH, já bem estabelecida para CNV.

Com o objetivo de avaliar o uso potencial de *NGS* para chamadas de CNVs, foram testados nove pacientes com resultados positivos para o array CGH 180K. Para a análise dos dados de sequenciamento utilizaram-se dois softwares diferentes, o NextGene e o XHMM, sendo que ambos se baseiam em cobertura, que é o método mais popular, para a chamada de CNV (DUAN *et al*., 2013).

Através de nove pacientes o aCGH 180K identificou um total de 21 CNVs polimórficas e raras, envolvendo regiões exônicas.

Tanto o NextGene quanto o XHMM foram ineficientes em detectar as alterações identificadas no array, mesmo utilizando um critério muito pouco restritivo para considerar as alterações como sobrepostas. A taxa de sobreposição do NextGene é em média inferior ao que tem sido visto na literatura, também não tão elevada, mas geralmente com um valor acima de 50% (NAM *et al*., 2015; HWANG *et al*., 2014; GUO *et al*., 2013)

O que se pode observar é mais da metade das sobreposições observadas para cada um dos softwares isoladamente *versus* o aCGH, estava em concordância. No entanto a sobreposição obtida com o NextGene com relação ao aCGH sugere que este software tenha sido menos eficiente para detectar as deleções. O aCGH e o XHMM apresentaram uma proporção similar de duplicação e deleção.

Ambos os softwares identificaram CNVs com direção oposta (duplicação ao invéz de deleção) em comparação com as do aCGH. Estes resultados discordantes obtidos entre os softwares já foram vistos em outros trabalhos e podem estar correlacionados ao viés associado ao algoritmo usado por cada um deles (GUO *et al*., 2013).

O NextGene se mostrou capaz de detectar alteração menor que 50 Kb, tendo sido o único a detectar a alteração pequena chamada pelo aCGH 180K, identificada no paciente 1389-1. Este mesmo software detectou em média CNVs relativamente menores do que o XHMM. A maior CNV detectada pelo aCGH no paciente 4922-1 foi detectada apenas pelo XHMM. Na literatura os softwares para chamada de CNV a partir de *NGS* apresentam, em geral, melhor desempenho para alterações que envolvem mais de três exons (RETTERER *et al*., 2014; FROMER *et al*., 2013).

Em comparação ao aCGH, ambos os softwares XHMM e NextGene apresentaram, para cada um dos pacientes, taxas diferentes de deleção-duplicação. Aparentemente nenhum paciente apresentou resultados congruentes entre as três metodologias. Esta variação observada na taxa de deleção-duplicação de cada paciente quando analisado através de metodologias diferentes também pode estar associada a um viés correlacionado ao algoritmo utilizado por cada um deles (GUO *et al.*, 2013).

Tanto o NextGene quanto o XHMM detectaram a única CNV classificada como potencialmente patogênica, identificada pelo aCGH no paciente 5616-1.

Com relação a comparação apenas entre o NextGene e o XHMM, o NextGene gerou uma quantidade de CNVs por paciente significativamente superior ao que obteve-se com o XHMM, com mais deleções do que duplicações, ao contrário do que foi obtido com o XHMM.

A quantidade diferente de CNVs gerada por cada software pode estar associada a métodos diferentes de concatenação das alterações e normalização dos dados. O NextGene, no geral, apresenta uma técnica de concatenação menos eficiente que à do XHMM, o que, provavelmente, correlaciona-se ao fato de que grande parte do excesso de CNVs obtidas com este software, em comparação ao XHMM, é de alterações pequenas (<50 Kb).

Geralmente é mais difícil a detecção de duplicações do que de deleções, visto que a variação da quantidade de *reads* não é tão grande quando uma região está duplicada, do que quando está deletada (TEO *et al*., 2012). Isto, porém, está em concordância apenas com os dados gerados pelo NextGene.

A sobreposição média entre esses dois softwares também foi bastante pequena, com o NextGene apresentando um forte viés para deleção. Entre os fatores cogitados para esta baixa sobreposição tanto na comparação entre os dados de *NGS* versus aCGH quanto apenas entre os dados de *NGS*, estão: a cobertura média baixa das amostras, equivalente a 32 *reads* por par de base, que talvez demande ajustes especiais nos parâmetros de cada um dos softwares; a normalização pouco eficaz, visto que o ideal seria utilizar sempre amostras e/ou controles da mesma corrida, mas nem sempre isso é possível, podendo gerar variação de cobertura entre pacientes que estavam em diferentes corridas que não associada à CNV; a diferença de algoritmo de cada um dos softwares e o tamanho reduzido de algumas CNVs detectadas pelo aCGH, muitas vezes com pontos de quebra fora da região exônica, entre outros.

Além disto, considerar cobertura como fator predominante para a chamada de CNVs traz possíveis distorções já que a correlação quantitativa entre variação no número de cópia e cobertura pode variar de uma região genômica para outra, bem como de paciente para paciente. Entre os fatores que podem contribuir para a variação da cobertura estão: quantidade de GC na região, tamanho das *reads*, proximidade de duplicações segmentares, presença de SNPs, concentração do DNA, temperatura de hibridação, entre outros (FROMER *et al*., 2012).

Os softwares escolhidos ainda não se mostram promissores por si só quando comparados com o array, bem como quando comparados entre si. Este fato impossibilita, no momento, o uso de ambos para fins de delimitação das alterações, visto que entre as CNVs sobrepostas, nos dois softwares, a posição de início e fim variou bastante, e para fins de substituição da metodologia de aCGH por *NGS* para a chamada de CNVs.

Uma possível alternativa seria utilizar softwares com metodologias diferentes e, muitas vezes, complementares, para a chamada de CNVs. Exemplificando, o

método de mapeamento de *pair end* é mais eficaz para detectar pontos de quebra do que o baseado em cobertura. Este, por sua vez, se sai melhor para detectar CNVs grandes. Esta alternativa talvez possa contribuir para redução do número de resultados falso positivos, principalmente quando se trata de alterações pequenas (ZHAO *et al*., 2013).

6. CONCLUSÕES

Os resultados obtidos neste trabalho permitiram concluir que:

1. As customizações de aCGH realizadas não foram eficazes para a detecção de InDels grandes (70 pb – 1kb). Portanto estas alterações permanecem subcaracterizadas quando se trata de TEA.
2. O aCGH customizado mostrou-se uma estratégia promissora para elucidar mecanismos etiológicos do TEA idiopático associado a CNVs, visto que se obteve uma prevalência relativa de CNVs maior do que a descrita na literatura com ambas as customizações, podendo se tornar uma possível ferramenta de diagnóstico.
3. A análise dos dados de *NGS* através dos softwares *NextGene* e *XHMM* não é eficaz para a delimitação dos pontos de quebra de alterações detectadas pelo aCGH e não permite a migração de tecnologia, do aCGH para o *NGS* (considerando apenas análises feitas através do *NextGene* e *XHMM*).
4. As CNVs pequenas contribuem para o risco de TEA em pelo menos 6,3 %, reforçando a sua importância para a etiologia do TEA.

7.RECOMENDAÇÕES

- É de grande importância analisar diferentes populações e grupos amostrais maiores quando se trata da averiguação da patogenicidade de mutações.
- Recomenda-se avaliar o critério de seleção de CNVs a ser utilizado em função do desenho de aCGH utilizado.

REFERÊNCIAS BIBLIOGRÁFICAS

ACQUAAH-MENSAH, G.K. e TAYLOR, R.C. Brain in situ hybridization maps as a source for reverse-engineering transcriptional regulatory networks: Alzheimer's disease insights. Gene, 586(1):77-86, 2016.

ALARCÓN, M. *et al*. Linkage, Association, and Gene-Expression Analyses Identify CNTNAP2 as an Autism-Susceptibility Gene. **Am J Hum Genet**, 82(1):150-9, 2008.

ALKAN, C.; BRADLEY, P.C. e EICHLER, E.E. Genome structural variation discovery and genotyping, **Nature Reviews/Genetics**, 12:363-375, 2011.

KEARNEY, HM., *et al.,* American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants, **Genet Med.,** 13(7):680-5., 2011

AMERICAN PSYCHIATRIC ASSOCIATION. Diagnostic and statistical manual

of mental disorders: DSM-IV. 4th ed. Washington (DC): **American Psychiatric**

Association, 943 p., 2000.

AMERICAN PSYCHIATRIC ASSOCIATION. Diagnostic and statistical manual

of mental disorders: DSM-V. 5th ed. Washington (DC): **American Psychiatric**

Association, 943 p., 2013.

ANNEY, R. *et al.* A genome-wide scan for common alleles affecting risk for autism. **Human Molecular Genetics**, 1-11.2010.

ARKING, D.E. A Common Genetic Variant in the Neurexin Superfamily Membar CNTNAP2 Increases Familial Risk of Autism, **Am J Hum Genet**, 82:160-164, 2008.

ASADOLLAHI, R. *et al.* The clinical significance of small copy number variants in neurodevelopmental disorders. **J Med Genet**, 51:677-688, 2014.

AZEVEDO JA., *et al*. The microRNA Network Is Altered in Anterior Cingulate Cortex of Patients With Unipolar and Bipolar Depression. J Psychiatr Res, **82, 58-67, 2016.**

BAILEY, A. *et al.* Autism as a strongly genetic disorder: evidence from a British twin study. **Psychol Med**, 25(1):63-77, 1995.

BARLETT, C.W.; GOEDKEN, R. e VIELAND, J. Effects of auditing linkage evidence across subsets of data: reanalysis of the autism genetic resource exchange data set. **Am J Hum Genet**, 76:688-695, 2005.

BASU, S. N.; KOLLU, R. e BANERJEE-BASU, S. AutDB: a gene reference resource for autism research. **Nucleics Acids Research**, 37(Databaseissue), 832-836, 2009.

BERRY, C.T. *et al.* Developmental upregulation of vesicular glutamate transporter-1 promotes neocortical presynaptic terminal development. ***PLoSOne***, 7, e50911, 2012.

BETANCUR, C. Etiological Heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. **Brain Research**, 1380:42-77. 2011.

BOURGERON, T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. **Nat Rev Neurosci**. 16(9):551-63, 2015.

BOURGERON, T. Current knowledge on the genetics of autism and propositions for future research. **C R Biol**, 339(7-8):300-7, 2016.

BUCAN, M. *et al*. Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. **PLoS Genet.** Jun;5(6):e1000536, 2009.

BRITO, L.A. *et al*. Genetic contribution for non-syndromic cleft lip with or without cleft palate (NS CL/P) in different regions of Brazil and implications for association studies. **Am. J. Med. Genet** Part A, v. 155A, 7, 1581-7, 2011.

CANLI, T. Neurogenethics: An emerging discipline at the intersection of ethics, neuroscience, and genomics. **Appl Transl Genom**. 5:18-22, 2015.

CARREIRA, I. M. *et al*. Copy-number variants priorization after array-CGH analysis - a cohort of 1000 patients. **Molecular Cytogenetics**, 8:13, 2015.

CELESTINO-SOPER, P. B. *et al*. Use of array CGH to detect exonic copy number variants throughout the genome in autism families detects a novel deletion in TMLHE. **Human Molecular Genetics**, 20:4360-4370, 2011.

CENTERS FOR DISEASE CONTROL AND PREVENTION. Prevalence of autism spectrum disorder among children aged 8 years – autism and developmental disabilities monitoring network, 11 sites, United States, 2010. **MMWR Surveillance Summaries**, 63:1-21, 2014.

CHASTE, P. e LEBOYER, M. Autism risk factors: genes, environment, and gene-environment interactions. **Dialogues Clin Neurosci**. (3), 281-92, 2012.

CHEN Q. *et al*. SREB2/GPR85, a schizophrenia risk factor, negatively regulates hippocampal adult neurogenesis and neurogenesis-dependent learning and memory. Eur J Neurosci, 36(5): 2597–2608, 2012.

CHEN, L. *et al*. Genome Architecture and its roles in human copy number variation. **Genomics Inform**, 12(4):136-144, 2014.

CHIOCCHETTI, A.G.; BOUR, H.S. e FREITAG, C.M. Glutamatergic candidate genes in autismo spectrum disorder: an overview. **Journal of Neural Transmission**, 121, 1081-1106, 2014.

COHEN, J.C. *et al*. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. **Proc. Natl. Acad. Sci. USA**., Feb 7;103(6):1810-5, 2006.

CONRAD, D.F. *et al*. Origins and functional impact of copy number variation in the human genome. **Nature**, 464: 704-712, 2010.

CRAIG JM., *et al*. DNA Fragmentation Simulation Method (FSM) and Fragment Size Matching Improve aCGH Performance of FFPE Tissues. **Plos One**, 10.1371/0038881, 2012.

DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources. Firth, H.V. *et al* **Am.J.Hum.Genet** 84, 524-533, 2009 (DOI: dx.doi.org/10/1016/j.ajhg.2009.03.010)

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

DE RUBEIS, S. e BUXBAUM, J. D. Genetics and genomics of autism spectrum disorder: embracing complexity. **Human Molecular Genetics**, 24: R24-R31, 2015.

DESACHY, G. *et al.* Increased female autosomal burden of rare copy number variants in human populations and in autism families. **Molecular Psychiatry**, 1-6, 2015.

DEVLIN, B. e SCHERER, S. W. Genetic architecture in autism spectrum disorder. **Current Opinion in Genetics e Development**, 22, 229-237, 2012.

DUAN, J. *et al*. Comparatives studies of copy number variation detection methods for next-generation sequencing technologies. **PLoS ONE**, 8(3), 2013.

EL-FISHAWY, P. e STATE, M. W. The genetics of autism: key issues, recent

findings, and clinical implications. **Psychiatr. Clin. North. Am.**, 33(1):83-105, 2010.

ENGCHUAN, W. *et al.* Performance of case-control rare copy number variation annotation in classification of autism. **BMC Medical Genomics**, 8(Suppl 1):57, 2015**.**

FLORIO, M. *et al*. Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. **Science**, v. 347, n. 6229, p. 1465-1470, 2015.

FOLSTEIN, S.E. e ROSEN-SHEIDLEY, B. Genetics of autism: complex aetiology for a heterogeneous disorder. **Nature**, 2(12):943-55, 2001.

FOLSTEIN, S.E. The clinical spectrum of autism. **Clinical Neuroscience Resarch**. 6(3-4), 113-117, 2006.

FOMBONNE, E. The epidemiology of autism: a review. **Psychol Med**. 29(4), 769-86, 1999.

FOMBONNE, E. Epidemiology of pervasive developmental disorders. **Pediatric Research** 65, 591-8, 2009.

FREEMAN, J.L. *et al.* Copy number variation: New insights in genome diversity. **Genome Research**, 16:949-961, 2006.

FREITAG, C.M. The genetics of autistic disorders and its clinical relevance: a review of the literature. **Molecular Psychiatry**, 12:2-22, 2007.

FROMER, M. et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. **Am J Hum Genet,** 91(4), 597-607, 2012.

FROMER, M. et al. De novo mutations in schizophrenia implicate synaptic networks. **Nature**, 506(7487), 179-84, 2014.

GAI, X. *et al*. Rare structural variation of synapse and neurotransmission genes in autism. **Molecular Psychiatry**, 17:402-411, 2012.

GAUGLER, T. *et al*. Most genetic risk for autism resides with common variation. **Nat. Genet.** 46, 881–885, 2014.

GAZZELLONE, M. J. *et al*. Copy number variaton in Han Chinese individuals with Autism Spectrum Disorder. **Journal of Neurodevelopmental Disorders**, 6:34, 2014.

GESCHWIND, D.H. Autism: many genes, common pathways?. **National Institute of Health**, 135(3):391-395, 2008.

GHEBRANIOUS, N. *et al*. A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation. **Am. J.Medical Genetics**, 143A: 1462-1471, 2007.

GIRIRAJAN, S. *et al.* Phenotypic heterogeneity of genomic disorders and rare copy-number variants. **N. Engl. J. Med.** 367, 1321–1331, 2012.

GIRIRAJAN, S. *et al*. Refinement and discovery of new hotspots of copy-number variation associated with Autism Spectrum Disorder. **Am J Hum Genet**., 92,221-237, 2013.

GLESSNER, J. T. *et al*. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. **Nature**. 459, 569-573, 2009.

GUO, Y. *et al.* Comparative studies of exome copy number variation estimation tools using array comparative genomic hybridization as control. **BioMed Research International**, vol. 2013, 7p., 2013.

GUPTA, A.R. e STATE, M.W. Recent advances in the genetics of autism. **Biol. Psychiatry**, 61:429-437, 2007.

HARAKSINGH, R. R. *et al*. Genome-wide mapping of copy number variations in humans: comparative analysis of high resolution array platforms**. PLoS One**, 6, e27859, 2011.

HOLLANDER, E. *et al*. Obsessive-compulsive behaviors in parents of multiplex autism families. **Psychiatry Research**, 117:11-16, 2003.

HOLLANDER, E. *et al.* A double-blind placebo-controlled trial of fluoxetine for repetitive behaviors and global severity in adult autism spectrum disorders. **Am J Psychiatry**, 169(3):292-9, 2012.

HU, V.W. *et al.* Gene Expression Profiling of Lymphoblasts from Autistic and Nonaffected Sib Pairs: Altered Pathways in Neuronal Development and Steroid Biosynthesis. **PLoS One**., 3;4(6):e5775, 2009.

HURLES, M.E.; DERMITZAKIS, E.T. e TYLER-SMITH, C. The functional impact of structural variation in humans. **Trends Genet.**, 24(5):238-245, 2008.

HWANG, MY. *et al*. Combinatorial Approach to Estimate Copy Number Genotype Using Whole-Exome Sequencing Data. **Genomics**,105 (3), 145-149, 2014.

IOSSIFOV, I. *et al*. De novo gene disruptions in children on the autistic spectrum.**Neuron**. 26;74(2):285-99, 2012.

ITSARA, A. *et al.* De novo rates and selection of large copy number variation. **Genome Res.** 20(11):1469-81, 2010.

IYER, J. e GIRIRAJAN, S. Gene discovery and functional assessment of rare copy-number variants in neurodevelopmental disorders. **Briefings in Functional Genomics**, 1-14, 2015.

JACQUEMONT, M. L. *et al*. Array-based comparative genomic hybridization identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. **J. Med. Genet.**, 43, 843-849, 2006.

JACQUEMONT, S. *et al*. A higher mutational burden in females supports a “female protective model” in neurodevelopmental disorders. **Am J Hum Genet**, 94(3), 415-425, 2014.

JIANG, Y. *et al*. Detection of clinically relevant genetic variants in Autism Spectrum Disorder by whole-genome sequencing. **Am J Hum Genet**., 93, 249-263, 2013.

JOSET, P. *et al*. Rostral growth of commissural axons requires the cell adhesion molecule *MDGA2*.**Neural Dev**, 4;6:22, 2011

KANNER, L. Autistic disturbances of affective contact. Nerv. Child 2, 217–250, 1943.

KEARNEY *et al*. Recomendations for systematic evaluation and clinical interpretation of CNVs. **Genetics in Medicine,** vol XX,XXX, 2011.

KIM, Y. S. *et al*. A comparison of DSM-IV pervasive developmental disorder and DSM-5 autism spectrum disorder prevalence in an epidemiologic sample. **J Am Acad Child Adolesc Psychiatry**, 53(5), 500-8, 2014.

KIM, Y.S. e LEVENTHAL, B.L. Genetic epidemiology and insights into interactive genetic and environmental effects in autism spectrum disorders. **Biol Psychiatry**, 77(1), 66-74, 2015.

KLAUCK, S.M. Genetics of autism spectrum disorder. **European Journal of Human Genetics**, 14, 714-720, 2006.

KRUMM, R. *et al.* Transmission disequilibrium of small CNVs in simplex autism. **Am J Hum Genet**., 93, 595-606, 2013.

KUMAR, R.A. *et al*. Recurrent 16p11.2 microdeletions in autism. **Human Molecular Genetics**, 17(4):628-638, 2008.

KUSENDA, M. e SEBAT, J. The role of rare structural variants in the genetics of autism spectrum disorders. **Cytogenetic Genome Res.**, 123, 36-43, 2008.

LAI, M. C. *et al*. Sex/gender differences and autism: setting the scene for future research. **J Am Acad Child Adolesc Psychiatry**, 54(1), 11-24, 2015.

LEBLOND, C. S. *et al*. Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. **PLoS Genet**, 8(2), e1002521, 2012.

LEYFER, O.T. *et al.* Comorbid Psychiatric Disorders in Children with Autism: Interview Development and Rates of Disorders. **Journal of Autism and Developmental Disorders**, 36(37): 849 – 61, 2006.

LEVY D, *et al*. Rare *de novo* and transmitted copy-number variation in autistic spectrum disorders. **Neuron** 70(5):886-897, 2011.

LI, C. e HUNG-WONG, W. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. **Genome Biol.** 2(8): RESEARCH0032, 2001.

LI, X.; ZOU, H. e BROWN, W.T. Genes associated with autism spectrum disorder. **Brain Research Bulletin**, 88:543-552, 2012.

LIN, M. *et al.* dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. **Bioinformatics.** 20(8):1233-40, 2004.

LINTAS, C. e PERSICO, A.M. Autistic phenotypes and genetic testing: state-of-the-art for the clinical geneticist. **J Med Genet**, 46(1):1-8, 2009.

LIU, J. *et al*. A Genomewide Screen for Autism Susceptibility Loci. **Am. J. Hum. Genet**. 69:327-340, 2001.

LOKE, Y. J.; HANNAN, A.J. e CRAIG, J.M. The role of epigenetic change in autism spectrum disorders. Front Neurol. 6:107, 2015.

LUGNEGARD, T.; HALLERBÄCK, M.U. e GILLBERG, C. Psychiatric comorbidity in young adults with a clinical diagnosis of Asperger syndrome. **Research in Developmental Disabilities**. 32:1910-1917, 2011.

LUGTENBERG, D. *et al.* Recurrent deletion of ZNF630 at Xp11.23 is not associated with mental retardation. **Am J Med Genet A.** 152A(3):638-45, 2010.

LUPSKI, J.R. Genomic rearrangements and sporadic disease. **Nat Genet**., 39(7 Suppl):S43-7, 2007

MANOLIO, T.A. *et al.* Finding the missing heritability of complex diseases. **Nature** 461, 747-753, 2009.

MANOLIO, T.A. Genomewide association studies and assessment of the risk of

disease. **N Engl J Med**. 363(2): 166-76, 2010.

MARSHALL, C. R. *et al*. Structural variation of chromosomes in autism spectrum

disorder. **Am. J. Hum. Genet**. 82:477–488, 2008.

MATSON,J.L. e KOSLOWSKI, A.M. The increasing prevalence of autism spectrum disorders. **Research in Autism Spectrum Disorders**, 5:418-425, 2011.

MATSON, M. L.; MATSON, J. L. e BEIGHLEY, J. S. Comorbidity of physical and motor problems in children with autism. **Research in Developmental Disabilities**, 32 (6): 2304-2308, 2011.

MERIKANGAS, A. K.; CORVIN, A. P. e GALLAGHER, L. Copy-number variants in neurodevelopmental disorders: promises and challenges. **Trends Genet**. Dec;25(12):536-44, 2009.

MERIKANGAS, A.K. *et al.* The phenotypic manifestations of rare genic CNVs in Autism Spectrum Disorder. **Molecular Psychiatry**, 1-7, 2014.

MILES, J. H. *et al*. Essential versus complex autism: definition of fundamental prognostic subtypes. **Am. J. Med. Genet**. A 135, 171–180, 2005.

MILES, J.H. *et al.* **Autism Spectrum Disorders**. In: Pagon RA, Bird TD, Dolan CR, Stephens K. Adam MP, editors. GeneReviews™ [Internet]. Seattle (WA): University of Washington, Seattle; 2010.

MOECHARS, D. *et al.* Vesicular glutamate transporter VGLUT2 expression levels control quantal size and neuropathic pain. **Journal of Neuroscience**, 26, 12055-12066, 2006.

MOLLOY, C.A., KEDDACHE, M. e MARTIN, L.J. Evidence for linkage on 21q and 7q in a subset of autism characterized by developmental regression. **Molecular Psychiatry**, 10:741-746, 2005.

MOREIRA, D. P. *et al .* Investigation of 15q11-q13, 16p11.2 and 22q13 CNVs in autism spectrum disorder Brazilian individuals with and without epilepsy. **PLoS One**, 9, e107705, 2014.

MOREIRA, E. S. *et al .* Detection of small copy number variations (CNVs) in Autism Spectrum Disorder (ASD) by custom array comparative genomic hybridization (aCGH). **Res, Autism Spectr Disord.,** 23, 145–151, 2016.

MUHLE, R.; TRENTACOSTE, S.R. e RAPIN, I. The genetics of autism. **Pediatrics** 113(5):e472-86, 2004.

MUHLE, H. *et al*. The role of SLC2A1 in early onset and childhood absence epilepsies. **Epilepsy Res**. 105(1-2), 229-33, 2013.

MUÑOZ-AMATRIAÍN, M. *et al.* Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome. **Genome Biology**, 14:R58, 1-17, 2013.

NAM, JY.*et al*. Evaluation of Somatic Copy Number Estimation Tools for Whole-Exome Sequencing Data**. Brief Bioinform**, **17 (2), 185-192, 2015**

NAVA, C. et al. Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including TMLHE. **Transl Psychiatry**, 2(10), e179, 2012.

NEALE, B. M. *et al*. Patterns and rates of exonic de novo mutations in autism

spectrum disorders. **Nature**. 4;485(7397):242-5, 2012.

NEWINGTON, Jordan T. et al. Overexpression of pyruvate dehydrogenase kinase 1 and lactate dehydrogenase A in nerve cells confers resistance to amyloid β and other toxins by decreasing mitochondrial respiration and reactive oxygen species production. **J Biolog Chem**, 287(44), 37245-58, 2012.

NEWSCHAFFER, C.J. *et al.* Infant siblings and the investigation of autism risk factors. **J Neurodev Disord**, 4(1):7, 2012.

O'ROAK, B. J. *et al*. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. **Nature**. 4;485(7397):246-50, 2012.

PANG, T. The impact of genomics on global health. **American Journal of Public Health**, 92(7), 1077-1079, 2002.

PAULA, C. S. Autism in Brazil: perspectives from science and society. **Rev. Assoc.Med. Bras**., 57(1):2-5, 2011.

PAULA, C.S. *et al.* Brief Report: prevalence of Pervasive Developmental disorder in Brazil: a pilot study. **J. Autism Dev. Disord.**, 41:1738-1742, 2011.

PAULA, C.S. *et al.* Brief Report: Prevalence of Pervasive Developmental Disorder in Brazil: A Pilot Study. **Journal of autism and developmental disordes**. Epub ahead of print, 2011.

PFAFFL,M. W. A new mathematical model for relative quantification in real-time RT–PCR. **Nucleic acids research** 29.9, 2001.

PINTO, D. *et al*. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. **Am. J. Hum. Genet***.* V. 94, pp. 677–694, 2014.

PINTO, D. *et al*. Functional Impact of Global Rare Copy Number Variation in Autism

Spectrum Disorder. **Nature**. 15;466(7304):368-72, 2010.

PIVEN, J. *et al.* Broader Autism Phenotype: Evidence From a Family History Study of Multiple-Incidence Autism Families. **Am. J. Psychiatry**, 154:185-190, 1997.

POULTNEY, J. *et al.* Identification of small exonic CNV from whole-exome sequence data and application to Autism Spectrum Disorder. **Am J Hum Genet**., 93, 607-619, 2013.

PRASAD, A. *et al.* A discovery resource of rare copy number variations in individuals with autism spectrum disorder. **G3**, 2, 1665-1685, 2012.

PRONTERA, P. *et al*. DPP6 gene disruption in a family with Gilles de la Tourette syndrome. N**eurogenetics**, 15(4), 237-42, 2014.

QIAO, Y. *et al*. Clinical application of 2.7M cytogenetics array for CNV detection in subjects with idiopathic autism and/or intellectual disability. **Clin. Genet.**, 83:145-154, 2013.

REDON, R. *et al*. Global variation in copy number in the human genome. **Nature.** 23;444(7118): 444-454, 2006.

REDON R, *et al*. Global variation in copy number in the human genome. **Nature**,444:444–454,2006.

RETTERER, K. *et al.* Assessing copy number from exome sequencing and exome array CGH based on CNV spectrum in a large clinical cohort. **American College of Medical Genetics and Genomics**, 2014.

RODRIGUEZ-REVENGA, L. *et al*. Structural variation in the human genome: the impact of copy number variants on clinical diagnosis.**Genet Med**, 9(9), 600-6, 2007.

RIBEIRO, C.M. Estudo de genes candidatos aos Transtornos do Espectro Autista.Tese apresentada ao Instituto de Biociências da Universidade de São Paulo, 2013.

SANDERS, S. J. *et al*. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. **Nature**, 4;485(7397):237-41, 2012.

SANDERS, S. J. *et al*. Insights into Autism Spectrum Disorder genomic architecture and biology from 71 Risk Loci. **Neuron**, 87, 1215-1233, 2015.

SANDIN, **S**. *et al*. The familial risk of autism. **JAMA**, v. 311, pp. 1770–1777, 2014.

SBACCHI, S. *et al*. Functional Annotation of Genes Overlapping Copy Number

Variants in Autistic Patients: Focus on Axon Pathfinding. **Current Genomics**, 11,136-145, 2010.

SCHAAF, C.P. *et al*. Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. **Human Molecular Genetics**, 20(17):3366-3375, 2011.

SEBAT, J. *et al*. Strong association of de novo copy number mutations with autism.**Science**, 316:445–449, 2007.

SENER, E.F. Association of copy number variations in Autism Spectrum Disorders: a systematic review. **Chinese Journal of Biology,** vol. 2014, 9p., 2014.

SHU, M.T. *et al.* Statistical challenges associated with detecting copy number variations with next-generation sequencing. **Bioinfomatics Advance Access,** 2012.

SIU, W. K. *et al.* Diagnostic yield of array CGH in patients with Autism Spectrum Disorder in Hong Kong. **Clinical and Translational Medicine**, 5:18, 2016.

SMITH, A.J. *et al.* Human genes involved in copy number variation: mechanism of origin, functional effects and implications for disease. **Cytogenet. Genome Res.**, 123:17-26, 2008.

SMITH, C.L.; BOLTON, A. e NGUYEN, G. Genomic and Epigenomic Instability, Fragile Sites, Schizophrenia and Autism. **Current Genomics**, 11, 447-469, 2010.

STANKIEWICZ, P. e LUPSKI. J. R. Structural variation in the human genome and its role in disease. **Annu. Rev. Med**., 61:437-55, 2010.

STEFANSSON, H. *et al.* CNVs conferring risk of autism or schizophrenia affect cognition in controls. **Nature**, 505, 2014.

SWANWICK, C.C.; LARSEN,E.C. e BANERJEE-BASU, S. Genetic heterogeneity of autism spectrum disorders , Autism Spectrum Disorders: The Role of Genetics in Diagnosis and Treatment, Prof. Stephen Deutsch (Ed.), ISBN: 978-953-307-495-5, InTech, 2011.

SZATMARI, P. *et al*. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. **Nat Genet**, 39(3), 319-28, 2007.

TATTINI, L.; D'AURIZIO, R. e MAGI, A. Detection of genomic structural variants from next-generation sequencing data. **Front Bioeng Biotechnol**, 3, 92, 2015.

TEO, S. M. *et al*. Statistical challenges associated with detecting copy number variations with next-generation sequencing. **Bioinformatics**, 28(21), 2711-18, 2012.

TEO SM., et al. Statistical challenges associated with detecting copy number variations with next-generation sequencing.**Bioinformatics**, 28, 2711–2718, 2012.

TUCHMAN, R. e RAPIN, I. Epilepsy in autism. **Lancet Neurol,** 1(6), 352-8, 2002.

TURNER, T. N. *et al*. Genome sequencing of autism-affected families reveals disruption of putative noncoding regulatory DNA. **Am J Hum Genet**, 98(1), 58-74, 2016.

VASSON, A. *et al.* Custom oligonucleotide array-based CGH: a reliable diagnostic tool for detection of exonic copy-number changes in multiple targeted genes. **European Journal of Human Genetics**, 21, 977-987, 2013.

VEENSTRA-VANDERWEELE, J. e COOK JR, E.H. Molecular genetics of autism spectrum disorder. **Molecular Psychiatry**, 9:819-832, 2004.

VORSTMAN, J.A.S. *et al.* Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. **Molecular Psychiatry**, 11(18-28), 2006.

WANG, K. Common genetic variants on 5p14.1 associate with autism spectrum disorders. **Nature**, 459(7246):528-33, 2009.

WANG, K. *et al*. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. **Genome Research**, 17, 1665–1674, 2007.

WANG, K. e BUCAN,M. Copy Number Variation Detection via High-Density SNP Genotyping. **Cold Spring Harb. Protoc**.; DOI:10.1101/pdb.top46, 2008.

WEISS, L.A. *et al*. Association between Microdeletion and Microduplication at 16p11.2 and Autism. **N. Engl. J. Med.** 358(7):667-675, 2008.

WEISS, L.A.; ARKING, D.E. e THE GENE DISCOVERY PROJECT OF JOHNS HOPKINS THE AUTISM CONSORTIUM. A genome-wide linkage and association scan reveals novel loci for autism. **Nature**. 8;461(7265):802-808, 2009.

WISZNIEWSKA, J. *et al.* Combined array CGH plus SNP genome analyses in a single assay for optimized clinical testing. **European Journal of Human Genetics**, 22, 79-87, 2014.

XU, J. *et al*. Molecular Cytogenetics of Autism. **Current Genomics**, 5, 347-364, 2004.

YUEN, R.K.C *et al*. Whole-genome sequencing of quartet families with autism spectrum disorder. **Nat Med**, 21(2), 185-91, 2015.

ZHAO, X. *et al*. A unified genetic theory for sporadic and inherited autism. **PNAS**, 104(31):12831-12836, 2007.

ZHAO, M. *et al.* Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. **BMC Bioinformatics**, 14(Suppl 11):S1, 1-16, 2013.

ANEXO I – LISTA DE GENES

ANEXO 1. LISTA DOS 269 GENES COBERTOS PELO aCGH CUSTOMIZADO 60K.

AKT3	*AHI1*	*A2BP1*	*ANXA5*
ARID4B	*ANXA1*	*AANAT*	*ANXA7*
ASTN1	*ASTN2*	*AKT1*	*ARHGEF9*
C3orf58	*CACNA2D1*	*AKT2*	*BDNF*
CACNA2D2	*CADPS2*	*ANKRD11*	*BRWD3*
CACNA2D3	*CNTNAP2*	*APBA2*	*CDKL5*
CACNB4	*DDC*	*AQP4*	*CUL4B*
CHL1	*DLGAP2*	*ARX*	*DGKZ*
CNTN3	*DNM1*	*BZRAP1*	*FMR1*
CNTN4	*DPP6*	*CACNA1G*	*GABRA3*
CNTNAP5	*EN2*	*CACNA1H*	*GABRB1*
DISC1	*FOXP2*	*CACNA1I*	*GABRE*
DPP10	*GABBR1*	*DLGAP4*	*GABRG1*
DPYD	*GABBR2*	*DOC2A*	*GABRQ*
ERBB4	*GABRA1*	*ERBB2*	*GAD2*
FARP2	*GABRA6*	*GABRA5*	*GDI1*
FBXO40	*GABRB2*	*GABRB3*	*GLRA3*
GABRA2	*GABRG2*	*GABRG3*	*GLRB*
GABRA4	*GABRP*	*GPHN*	*GRIA4*
GABRD	*GJA1*	*GRIA3*	*GRM5*
GAD1	*GPLD1*	*GRIN2A*	*GRM6*
GAP43	*GRIA1*	*GRIN2C*	*HTR1A*
GBX2	*GRIK2*	*GRIN2D*	*HTR3A*
GPC1	*GRIN1*	*GRIN3B*	*HTR3B*
GRIA2	*GRIN3A*	*GSS*	*HTR4*
GRM2	*GRM1*	*HOXB1*	*HUWE1*
GRM7	*GRM3*	*HTR2C*	*IL1RAPL1*
HDAC1	*GRM4*	*ITGB3*	*KDM5C*
HDLBP	*GRM8*	*MAPK1*	*MAOA*
HTR1D	*HDAC2*	*MAPK3*	*MeCP2*
HTR1F	*HGF*	*MBD1*	*NLGN3*
HTR2B	*HOXA1*	*MBD2*	*NLGN4X*
HTR3C	*HTR1B*	*MBD3*	*NPY1R*
HTR3D	*HTR1E*	*MDGA2*	*NPY5R*
HTR3E	*HTR5A*	*MLST8*	*NRG3*
HTR6	*HTR7*	*NF1*	*NRXN2*
INPP1	*KCND2*	*NLGN2*	*PAK3*
KCNAB2	*LAMB1*	*NPTXR*	*PCDH10*
KIF1A	*MAFK*	*NRG4*	*PTCHD1*
MARK1	*MET*	*NRXN3*	*PTEN*
MBD4	*NHLRC1*	*NSF*	*RbBP7*
MTOR	*NPTX2*	*ORAI3*	*SAP30*
NLGN1	*NrCAM*	*PLAUR*	*SLC17A6*
NOS1AP	*NRG1*	*PLD2*	*SLC1A2*
NRXN1	*NRG2*	*PRKCA*	*SLC9A6*
OXTR	*ORAI2*	*PRKCB*	*SYN1*
PASK	*PARK2*	*PRODH*	*SYP*
PDK1	*PIK3CG*	*RAB3A*	*TRPC3*
PIKFYVE	*PRKAA1*	*RPTOR*	*TRPC5*
PLD1	*RAC1*	*SCAMP5*	*UBE2A*
PRKAA2	*RELN*	*SEZ6L2*	
RAPGEF4	*RHEB*	*SHANK3*	
RbBP4	*Rictor*	*SIN3A*	
RFWD2	*RNF216L*	*SLC17A7*	
SAP130	*SEMA3A*	*SLC1A6*	
SCN1A	*SLC17A1*	*SLC32A1*	
SCN2A	*SLC17A2*	*SLC6A4*	
SLC1A4	*SLC18A1*	*SNAP25*	
SLC1A7	*SLC18A2*	*STXBP6*	
SLC25A12	*SLC1A1*	*SYN3*	
SLC6A1	*STX1A*	*SYT4*	
SLC6A11	*STXBP1*	*TSC2*	
SLC9A9	*TSC1*	*UBE3A*	
SYN2	*UBE2H*	*UBR*	
SYT11	*VGF*	*VAMP2*	
TRPC1	*WNT2*	*ZNF778*	

ANEXO II – LISTA DE GENES

ANEXO 2. LISTA DOS 1527 GENES COBERTOS PELO aCGH CUSTOMIZADO 180K.							
5S_rRNA	ASXL3	CD99	DGCR6L	DYDC1	GAD2	HIRA	KIAA0182
7SK	ATG7	CD99L2	DGCR8	DYDC2	GALNT13	HIRIP3	KIAA0232
AANAT	ATOH1	CDC42BPB	DGKZ	DYRK1A	GALNT14	HIST1H1A	KIAA0430
AB209061	ATP10A	CDC45	DHCR7	EEF1A2	GALNTL4	HIST1H2BB	KIAA1586
ABAT	ATP1B1	CDH10	DIAPH3	EFR3A	GAN	HIST1H2BC	KIAA1653
ABCA1	ATP2A1	CDH11	DIP2A	EGR2	GAP43	HIST1H2BK	KIAA2022
ABCA7	ATP2B2	CDH22	DISC1	EHMT1	GARNL1	HIST1H3F	KIF13B
ABCC1	ATRNL1	CDH8	DKFZP434L.	EIF1AY	GAS2	HIST1H4A	KIF1A
ABCC6	ATXN2L	CDH9	DLG4	EIF2S3	GATA4	HIST1H4B	KIF22
ABHD11	ATXN7	CDIPT	DLGAP1	EIF3C	GATM	HIST1H4C	KIF4A
ABI2	AUTS2	CDIPT-AS1	DLGAP2	EIF3CL	GBX2	HIST1H4D	KIF5C
ABL1	AVPR1A	CDK2	DLGAP3	EIF4E	GCH1	HIST1H4F	KIRREL3
abParts	AX747189	CDK4	DLGAP4	EIF4EBP2	GDA	HLA-A	KIT
ACP6	AX747846	CDKL5	DLL1	EIF4H	GDI1	HLA-DRB1	KLC1
ACR	AXL	CDKN1B	DLX1	ELN	GDPD3	HMGN1	KLC2
ACTN4	BAIAP2	CECR2	DLX2	EML1	GGNBP2	HNRNPH2	KLF13
ACY1	BANK1	CELF4	DLX3	EN2	GIGYF1	HOMER1	KLHL13
ADA	BAZ1B	CELF6	DLX6	ENO2	GJA1	HOXA1	KLHL22
ADAMTS18	BBS4	CEP290	DMD	EP300	GJA5	HOXB1	KMO
ADARB1	BC029255	CEP41	DMPK	EP400	GJA8	HRAS	KMT2A
ADCY5	BC033281	CERS4	DNAH10	EPC2	GLO1	HS3ST5	KMT2C
ADK	BC035867	CHAT	DNAJB1	EPHA6	GLRA2	HSD11B1	KMT2E
ADNP	BC039500	CHD1L	DNAJC19	EPHB2	GLRA3	HSD3B7	KPTN
ADORA2A	BC041003	CHD2	DNAJC30	EPHB6	GLRB	HSP90AA1	KRT34
ADORA3	BC041466	CHD7	DNER	EPS8	GLUD2	HSPH1	LAMA1
ADRB2	BC043570	CHD8	DNHD1	ERBB2	GNAS	HTR1A	LAMB1
ADSL	BC050343	CHL1	DNM1	ERBB2IP	GNB1L	HTR1B	LAMC3
AF072097	BC050599	CHRFAM7A	DNM1L	ERBB3	GOLGA6L1	HTR1D	LAT
AFF2	BC070376	CHRM3	DNMT1	ERBB4	GOLGA6L2	HTR1E	LAT2
AGAP1	BC073928	CHRNA7	DNMT3A	ERCC6	GOLGA6L7I	HTR1F	LEO1
AGAP2	BC127858	CHRNB3	DNMT3B	ERG	GOLGA8EP	HTR2A	LEP
AGBL4	BCAP31	CIB2	DOC2A	ESR1	GOLGA8F	HTR2B	LIMK1
AGMO	BCAS1	CLDN3	DOCK10	ESR2	GOLGA8G	HTR2C	LIN7B
AGTR2	BCKDK	CLDN4	DOCK11	ESRRB	GOLGA8I	HTR3A	LINC00035
AHI1	BCL11A	CLDN5	DOCK4	ETFB	GOLGA8S	HTR3B	LINC00624
AIFM3	BCL7B	CLIP2	DOLK	ETS1	GOLGA8T	HTR3C	LINC00895
AK023548	BCL7C	CLN3	DPP10	EXOC3	GPC1	HTR3D	LINC00896
AK023809	BCL9	CLSTN3	DPP3	EXOC5	GPC6	HTR3E	LINC00929
AK056973	BDNF	CLTCL1	DPP6	EXOC6	GPD2	HTR4	LMCD1
AK075019	BDNF-AS	CMIP	DPYD	EXOC6B	GPHN	HTR5A	LMX1B
AK091688	BGN	CNOT3	DQ571461	EXT1	GPLD1	HTR6	LOC100128
AK093758	BIN1	CNR1	DQ572979	F13A1	GPR139	HUWE1	LOC100288
AK097453	BIRC6	CNR2	DQ574263	FABP3	GPR37	HV593096	LOC100288
AK097472	BOLA2	CNTN3	DQ575242	FABP5	GPR6	HV593110	LOC100289
AK097527	BRAF	CNTN4	DQ575284	FAM135B	GPR85	HV593127	LOC100289
AK124131	BRCA2	CNTN5	DQ575576	FAM189A1	GPR89B	HV593134	LOC100862
AK124673	BRSK2	CNTN6	DQ575741	FAM190A	GPR89C	HV593135	LOC23117
AK125489	BRWD1	CNTNAP2	DQ576952	FAM57B	GPX1	HV593178	LOC283683
AK129567	BRWD3	CNTNAP3	DQ577333	FAM91A1	GRIA1	HV593183	LOC283685

AK302306	BX648073	CNTNAP4	DQ577902	FAN1	GRIA2	HYDIN	LOC283710
AK302545	BZRAP1	CNTNAP5	DQ578199	FARP2	GRIA3	ICA1	LOC283767
AK303607	C11orf30	COL25A1	DQ578369	FAT1	GRIA4	IKBKG	LOC284865
AK307870	C12orf57	COMT	DQ578700	FBRS	GRID1	IL16	LOC388849
AK309255	C15orf43	CORO1A	DQ578838	FBXL19	GRID2	IL1R2	LOC400891
AK310030	C15orf49	CREBBP	DQ579907	FBXL19-AS1	GRID2IP	IL1RAPL1	LOC440243
AK310228	C16orf45	CRKL	DQ582071	FBXO15	GRIK2	IL1RAPL2	LOC440354
AK310526	C16orf54	CSDE1	DQ582073	FBXO33	GRIK3	IL27	LOC595101
AK311660	C16orf92	CSMD1	DQ582448	FBXO40	GRIN1	IMMP2L	LOC606724
AKT1	C16orf93	CSNK1E	DQ582641	FCRL6	GRIN2A	INADL	LOC613037
AKT2	C16orf98	CSTF2T	DQ582939	FER	GRIN2B	INO80E	LOC613038
AKT3	C22orf29	CTCF	DQ583847	FEZF2	GRIN2C	INPP1	LOC643699
ALDH1A3	C22orf39	CTF1	DQ585847	FGA	GRIN2D	IPW	LOC646278
ALDH5A1	C3orf58	CTNNA3	DQ586415	FGD1	GRIN3A	IQGAP2	LOC729444
ALDOA	C4B	CTNNB1	DQ586919	FGFBP3	GRIN3B	IQSEC2	LOC730183
ALOX5AP	CA6	CTTNBP2	DQ587888	FHIT	GRIP1	ITGA4	LPL
AMELY	CACNA1B	CUBN	DQ588687	FKBP6	GRM1	ITGA5	LRFN5
AMPD1	CACNA1C	CUL2	DQ588973	FLJ00278	GRM2	ITGAL	LRP2
AMT	CACNA1D	CUL3	DQ589249	FLJ00285	GRM3	ITGAM	LRP2BP
ANK2	CACNA1F	CUL4B	DQ589803	FLJ00322	GRM4	ITGB3	LRPPRC
ANK3	CACNA1G	CX3CR1	DQ590322	FLNA	GRM5	ITGB7	LRRC1
ANKRD11	CACNA1H	CXCR3	DQ590616	FLT1	GRM6	JARID2	LRRC14
ANXA1	CACNA1I	CXorf36	DQ593032	FMO5	GRM7	JB175342	LRRC7
ANXA5	CACNA2D1	CYBB	DQ593239	FMR1	GRPR	JMJD1C	LSAMP
ANXA7	CACNA2D2	CYFIP1	DQ593342	FOLH1	GSC2	JPX	LZTR1
AP1S2	CACNA2D3	CYP11B1	DQ593983	FOPNL	GSDMC	JUP	LZTS2
APBA2	CACNB2	CYP19A1	DQ594309	FOXG1	GSK3B	JX456220	MACROD2
APC	CACNB4	DAB1	DQ594669	FOXP1	GSN	KAL1	MAFK
APH1A	CADM1	DAGLA	DQ595055	FOXP2	GSS	KANK1	MAGED1
APOBR	CADM2	DAPK1	DQ595631	FRK	GSTM1	KAT2B	MAGEL2
APP	CADPS2	DAPP1	DQ595648	FRMPD4	GTF2I	KAT8	MAL
AQP4	CAMK2G	DAZ1	DQ596229	FUS	GTF2IRD1	KATNAL2	MAOA
AR	CAMK4	DAZ2	DQ596303	FZD9	GUCY1A2	KCNAB2	MAOB
ARHGAP11E	CAMSAP2	DAZ3	DQ596319	GABBR1	HAUS7	KCND2	MAP2
ARHGAP24	CAMTA1	DAZ4	DQ596685	GABBR2	HBT8	KCNJ10	MAPK1
ARHGAP32	CAPN12	DCTN5	DQ596686	GABRA1	HCFC1	KCNJ12	MAPK11
ARHGDIA	CAPRIN1	DCTPP1	DQ597560	GABRA2	HCN1	KCNJ15	MAPK12
ARHGEF9	CAPS2	DCUN1D1	DQ597873	GABRA3	hCPE-R	KCNJ2	MAPK3
ARID1B	CARKD	DCX	DQ599532	GABRA4	HDAC1	KCNMA1	MAPK7
ARID4B	CASC4	DDC	DQ599733	GABRA5	HDAC2	KCNQ2	MAPK8
ARNT2	CASK	DDX11	DQ599785	GABRA6	HDAC4	KCNQ3	MAPK8IP2
ARSA	CBS	DDX3X	DQ600136	GABRB1	HDAC6	KCNT1	MARK1
ARVCF	CBX4	DDX3Y	DQ600342	GABRB2	HDAC9	KCTD13	MAX
ARX	CC2D1A	DDX53	DQ786262	GABRB3	HDHD1	KCTD4	MAZ
ASH1L	CC2D1B	DDX54	DQ786280	GABRD	HDLBP	KDM3A	MBD1
ASMT	CCDC101	DEAF1	DRD1	GABRE	HEPACAM	KDM4B	MBD2
ASMTL	CCDC64	DEPDC5	DRD2	GABRG1	HERC2	KDM5B	MBD3
ASPHD1	CD163L1	DGCR10	DRD3	GABRG2	HERC2P10	KDM5C	MBD3L1
ASS1	CD19	DGCR11	DRP2	GABRG3	HERC2P2	KDM5D	MBD3L2
ASTN1	CD2BP2	DGCR14	DST	GABRP	HERC2P7	KDM6B	MBD3L3
ASTN2	CD38	DGCR2	DTNA	GABRQ	HERC2P9	KHDRBS2	MBD3L4
	CD44	DGCR5	DUSP22	GAD1	HGF	KIAA0100	MBD3L5

MBD4	NOS1	PIK3R2	RFWD2	SLC6A12	SPNS1	TRPC3	ZNF646
MBD5	NOS1AP	PIKFYVE	RGS7	SLC6A13	SRCAP	TRPC4	ZNF668
MBD6	NOS2A	PINX1	RHEB	SLC6A2	SRD5A2	TRPC5	ZNF688
MC4R	NPAP1	PIR	RHOXF1	SLC6A3	SREBF1	TRPC6	ZNF689
MCC	NPAS2	PITX1	RIBC1	SLC6A4	SRPK2	TRPM1	ZNF713
MCF2	NPIP	PKD1P1	RIMBP3	SLC6A8	SRY	TSC1	ZNF74
MCPH1	NPIPL1	PLA1A	RIMS1	SLC7A4	ST7	TSC2	ZNF747
MDGA2	NPIPL3	PLAUR	RIMS3	SLC7A5P1	ST8SIA2	TSN	ZNF764
MDM2	NPY1R	PLCB1	RNF216P1	SLC9A6	STAU2	TSPAN6	ZNF768
MECP2	NPY5R	PLCD1	RNF40	SLC9A9	STK39	TSPAN7	ZNF771
MED13L	NR3C2	PLD1	RNPS1	SLCO1B1	STS	TSPYL2	ZNF774
MED15	NRCAM	PLD2	ROBO1	SLCO1B3	STX1A	TSSK2	ZNF778
MEF2C	NRG1	PLN	ROBO2	SLIT3	STX1B	TTI2	ZNF785
MET	NRG2	PLXNA3	RORA	SLITRK5	STX4	TTLL3	ZNF8
Metazoa_SR	NRG3	PLXNA4	RPL10	SLX1A	STXBP1	TTTY10	ZNF804A
MFRP	NRG4	PLXNB3	RPP25	SLX1B-SULT	STXBP5	TTTY14	ZNF827
MIB1	NRP2	PNPLA4	RPS4Y1	SMARCC2	STXBP6	TTTY15	
MIB2	NRXN1	POGZ	RPS6KA2	SMC2	SULT1A1	TUBA3FP	
Mir_548	NRXN2	POM121L4P	RPS6KA3	SMCX	SULT1A2	TUBGCP5	
Mir_649	NRXN3	POMGNT1	RPTOR	SMCY	SULT1A4	TUFM	
MIR1286	NSD1	PON1	RRN3	SMG6	SUV420H1	TXLNG	
MIR1306	NSF	POT1	RRN3P2	SMURF1	SVIL	TXLNG2P	
MIR137	NTAN1	PPFIA1	RTN4R	SNAP25	SYAP1	TXNRD2	
MIR185	NTNG1	PPM1D	S100G	SNAP29	SYN1	TYR	
MIR211	NTRK1	PPP1R1B	SAE1	SND1	SYN2	U84523	
MIR3179-2	NTRK3	PPP1R3F	SAP130	SNORA30	SYN3	UBA6	
MIR3179-3	NUPR1	PPP2R5D	SAP30	SNORD109I	SYNCRIP	UBE2A	
MIR3180-3	NXF5	PPP3CB	SATB2	SNORD115-	SYNE1	UBE2H	
MIR3180-4	NXPH1	PPP4C	SBF1	SNORD115-	SYNGAP1	UBE3A	
MIR3618	OCA2	PRICKLE1	SCAMP5	SNORD115-	SYNJ2	UBE3B	
MIR4284	ODF3L2	PRICKLE2	SCARA3	SNORD115-	SYP	UBE3C	
MIR4508	OFD1	PRKAA1	SCARF2	SNORD115-	SYT1	UBN2	
MIR4509-1	OGT	PRKAA2	SCFD2	SNORD115-	SYT11	UBR1	
MIR4517	OPHN1	PRKAB2	SCN1A	SNORD115-	SYT17	UBR3	
MIR4518	OPRM1	PRKCA	SCN2A	SNORD115-	SYT3	UBR5	
MIR4519	OR1C1	PRKCB	SCN5A	SNORD115-	SYT4	UBR7	
MIR4715	ORAI2	PRKD1	SCN7A	SNORD115-	TAAR2	UFD1L	
MIR4721	ORAI3	PRKY	SCN8A	SNORD115-	TAF1C	ULK4P1	
MIR4761	OTUD7A	PRODH	SCRIB	SNORD115-	TAF4	ULK4P3	
MIR484	OTX1	PRPF39	SDC2	SNORD115-	TANGO2	UPB1	
MIR5087	OXTR	PRR14	SDK1	SNORD115-	TAOK2	UPF2	
MIR590	P2RX4	PRRG4	SEMA3A	SNORD115-	TBC1D10B	UPF3B	
MIR762	P2RX6	PRRT2	SEMA5A	SNORD115-	TBC1D23	USP41	
MKL2	P2RX6P	PRSS36	SEPHS2	SNORD115-	TBC1D5	USP9X	
MKRN3	PAFAH1B1	PRSS38	SEPT1	SNORD115-	TBC1D7	USP9Y	
MLST8	PAGR1	PRSS53	SEPT5-GP1	SNORD115-	TBCK	UTP6	
MLXIPL	PAH	PRSS8	SERPIND1	SNORD115-	TBL1X	UTRN	
MPP6	PAK3	PRUNE2	SERPINE1	SNORD115-	TBL1XR1	UTY	
MPV17L	PAR-SN	PRY	SETBP1	SNORD115-	TBL2	VAMP1	

MRPL40	PAR1	PSD3	SETD1A	SNORD115-	TBP	VAMP2
MSN	PAR4	PSMB1	SETD2	SNORD115-	TBR1	VASH1
MSR1	PAR5	PSMC2	SETD5	SNORD115-	TBX1	VCP
MTF1	PARD3	PSMD10	SETDB1	SNORD115-	TBX6	VCX
MTHFR	PARD3B	PTCHD1	SETDB2	SNORD115-	TCF20	VCX2
MTMR10	PARK2	PTEN	SEZ6L2	SNORD115-	TCF3	VGF
MTMR12	PASK	PTGER3	SF1	SNORD115-	TCF4	VIL1
MTOR	PAX5	PTGS2	SGSH	SNORD115-	TCTE3	VIP
MTR	PCDH10	PTPN11	SGSM3	SNORD115-	TDO2	VKORC1
MTX2	PCDH11X	PTPRB	SH2B1	SNORD115-	TENM1	VPS13B
MVP	PCDH11Y	PTPRC	SH3GL3	SNORD115-	TGM1	VPS37D
MYH10	PCDH15	PTPRD	SH3KBP1	SNORD115-	TH	VSIG4
MYH11	PCDH19	PTPRM	SHANK1	SNORD115-	THAP7	WAC
MYH4	PCDH8	PTPRT	SHANK2	SNORD115-	THAP7-AS1	WBP5
MYLPF	PCDH9	PWRN1	SHANK2-AS	SNORD116-	THAP8	WBSCR22
MYO16	PCDHA1	PWRN2	SHANK2-AS	SNORD116-	THBS1	WBSCR27
MYO1A	PCDHA10	PXDN	SHANK3	SNORD116-	THRA	WBSCR28
MYO9B	PCDHA11	PYCARD	SIN3A	SNORD116-	TIGD4	WDFY3
MYOC	PCDHA12	PYDC1	SIX2	SNORD116-	TIMP1	WDR55
MYOZ1	PCDHA13	QPRT	SKI	SNORD116-	TJP1	WDR93
MYT1L	PCDHA2	QRICH1	SLC16A7	SNORD116-	TLK2	WHAMMP2
NAA15	PCDHA3	RAB11FIP5	SLC17A1	SNORD116-	TLS/FUS-ER	WHAMMP3
NBEA	PCDHA4	RAB19	SLC17A2	SNORD116-	TM4SF20	WHSC1
NBPF13P	PCDHA5	RAB2A	SLC17A6	SNORD116-	TMEM191A	WNK3
NBPF24	PCDHA6	RAB39B	SLC17A7	SNORD116-	TMEM191B	WNT1
NBPF8	PCDHA7	RAB3A	SLC17A8	SNORD116-	TMEM219	WNT2
NCKAP1	PCDHA8	RAB3IL1	SLC18A1	SNORD116-	TMEM231	WT1
NCKAP5	PCDHA9	RAB5A	SLC18A2	SNORD116-	TMLHE	XIST
NCKAP5L	PCDHAC1	RABEP2	SLC18A3	SNORD116-	TMLHE-AS1	XPC
NCOR1	PCDHAC2	RAC1	SLC1A1	SNORD116-	TMSB4Y	XPO1
NCOR2	PCDHGA11	RAC3	SLC1A2	SNORD116-	TMX2	Y_RNA
NDE1	PCOLCE	RAI1	SLC1A4	SNORD116-	TNIP2	YEATS2
NDN	PDE1C	RANBP1	SLC1A6	SNORD116-	TNK2	YPEL3
NDNL2	PDE4A	RANBP17	SLC1A7	SNORD116-	TNN	YTHDC1
NDUFA5	PDE4B	RAP1GDS1	SLC22A15	SNORD116-	TNRC6B	YTHDC2
NEFL	PDIA3P	RAPGEF4	SLC24A2	SNORD116-	TOMM20	YWHAE
NELL1	PDK1	RASD1	SLC25A1	SNRPN	TOP1	ZBTB16
NF1	PDXDC1	RASSF5	SLC25A12	SNTG2	TOP3B	ZBTB20
NFATC2IP	PDZD4	RB1CC1	SLC25A14	SNURF	TPO	ZDHHC8
NFIA	PDZK1P1	RbBP7	SLC25A24	SNURF-SNR	TPRX1	ZFX
NHLRC1	PECR	RBFOX1	SLC25A27	SNX19	TRIM17	ZFY
NIPA1	PER1	RBM8A	SLC29A4	SNX29P2	TRIM33	ZG16
NIPA2	PEX7	RBMS3	SLC30A5	SNX9	TRIM72	ZMYND11
NIPBL	PGR	RBMX	SLC32A1	SOD1	TRIO	ZNF18
NLGN1	PHF15	RBMX2	SLC35A3	SOD2	TRIP12	ZNF292
NLGN2	PHF2	RBMY1B	SLC38A1	SOX1	TRMT2A	ZNF385B
NLGN3	PHF8	RBMY1F	SLC38A10	SOX5	TRNA_Asn	ZNF407
NLGN4X	PHKG2	REEP3	SLC39A11	SOX9	TRNA_Gln	ZNF433
NLGN4Y	PI4KA	RELN	SLC4A10	SPARCL1	TRNA_Glu	ZNF48
NOMO1	PI4KAP1	RER1	SLC6A1	SPAST	TRNA_His	ZNF517
NOMO3	PIK3CB	RERE	SLC6A11	SPATA13	TRNA_Pseu	ZNF559
NONO	PIK3CG	RFC2		SPN	TRPC1	ZNF629

ANEXO III – TABELA 6

TABELA 6. COMPARAÇÃO ENTRE O aCGH E OS DADOS DE *NGS* ANALISADOS COM OS SOFTWARES COM O XHMM E O NEXTGENE PARA 9 PACIENTES COM TEA

Paciente	aCGH					XHMM				Nextgene			
	Chr	Posição de início	Posição de fim	Tamanho kb	Tipo	Chr	Posição de início	Posição de fim	Tipo	Chr	Posição de início	Posição de fim	Tipo
1389-1	3	195340900	195486396	145,5	DUP	3	195384610	195411036	DUP	3	195447879	195453443	DUP
	6	29808446	29857121	48,7	DEL					6	29795622	29912393	DUP
5616-1	11	55357067	55468560	111,5	DUP								
	16	28668145	28763145	95	DUP	16	28722482	28723408	DUP	16	28617377	28619924	DUP
	17	44277006	44788273	511,3	DEL	17	44247921	44657388	DEL	17	44409918	44627885	DEL
5996-1	11	55357067	55468560	111,5	DEL	11	55370530	55419615	DEL	11	55370530	55419615	DEL
6281-1	11	55357067	55468560	111,5	DUP								
	19	20574511	20718937	144,4	DEL								
4922-1	5	140222886	140239281	164	DEL	5	140174144	140234044	DEL	5	140165876	140812750	DEL
	11	55357067	55468560	111,5	DEL								
	15	20538064	20686671	148,6	DUP								
	15	20798419	22675094	1876,7	DUP	15	20711061	22278402	DUP				
	19	7018207	7063341	451,3	DUP								
7935-5	6	32430238	32489852	59,6	DEL					6	32489682	32632844	DEL
6024-1	16	34468047	34798082	330	DUP	16	34612643	34715267	DUP				
	X	151622408	151788761	166,4	DEL	X	151866945	151938540	DUP				
5547-1	8	39222426	39416498	194,1	DEL	8	39232898	39380770	DEL				
8293-1	3	162494699	162639153	144,5	DUP								
	8	7010253	8111027	1100,8	DEL	8	7194337	8090696	DEL	8	7170368	7718245	DEL
	8	39222426	39416498	194,1	DEL	8	39324842	39371974	DEL				
	22	18628078	18889039	261	DEL					22	18762555	18778742	DEL

Chr = cromossomo; DEL = deleção; DUP = duplicação

Obs.: a Posição de início e fim das CNVs é referente ao seu tamanho máximo

Comum

Rara

Potencialmente Patogênica

ANEXO IV – TABELA 7

TABELA 7. RESULTADO DA COMPARAÇÃO DAS CNVs OBTIDAS PELOS SOFTWARES XHMM E NEXTGENE PARA 9 PACIENTES

Software	XHMM					XHMM *versus* NextGene		
Alterações	**Total CNVs**	**Total DEL**	**Tamanho médio DEL**	**Total DUP**	**Tamanho médio DUP**	**Sobrep. DEL**	**Sobrep. DUP**	**Sobrep. Total %**
1389-1	124	36	122,7	88	83,2	9	15	0,19
5616-1	103	56	96,4	47	259,7	14	13	0,26
5996-1	136	72	99,3	64	253,4	23	23	0,34
6281-1	127	79	210,2	48	78	27	11	0,30
7935-5	77	37	161,4	40	129,3	8	9	0,22
6024-1	70	24	214,4	46	179,2	12	0	0,17
5547-1	129	34	381,9	95	290,5	25	0	0,19
8293-1	57	25	197,7	32	255,0	14	3	0,30
F4922-1	38	13	244,7	25	142,5	9	1	0,26
Total	861	376	1728,6	485	1670,9	141	75	2,24
Média	95,7	41,8	192,1	53,9	185,7	15,7	8,3	0,25
Software	**NEXTGENE**					**NextGene *versus* XHMM**		
Alterações	**Total CNVs**	**Total DEL**	**Tamanho médio DEL**	**Total DUP**	**Tamanho médio DUP**	**Sobrep. DEL**	**Sobrep. DUP**	**Sobrep. Total %**
1389-1	510	332	23,3	178	36,6	11	17	0,05
5616-1	479	280	3,8	199	25,7	20	19	0,08
5996-1	3.753	2.139	81,6	1.614	47,3	42	24	0,02
6281-1	573	361	15,0	212	10,1	48	13	0,11
7935-5	155	89	9,4	66	52,5	9	10	0,12
6024-1	211	211	77	0	0	16	0	0,08
5547-1	217	217	329	0	0	28	0	0,13
8293-1	709	612	0,7	97	6,6	25	5	0,04
F4922-1	141	134	54,6	7	23,9	10	1	0,08
Total	6.748	4.375	595	2373	202,7	209	73	0,71
Média	749,8	486,1	66,1	296,6	22,5	23,2	8,11	0,08

*del = deleção; dup = duplicação; sobrep. = sobreposição

ANEXO V - Detection of small copy number variations (CNVs) in autism spectrum disorder (ASD) by custom array comparative genomic hybridization (aCGH)