

UNIVERSIDADE SÃO PAULO

CINTIA MARQUES RIBEIRO

*Estudo de genes candidatos aos
Transtornos do Espectro Autista*

*Study of candidate genes to Autism
Spectrum Disorders*

São Paulo
2013

CINTIA MARQUES RIBEIRO

Estudo de genes candidatos aos Transtornos
do Espectro Autista

Study of candidate genes to Autism Spectrum
Disorders

Tese apresentada ao
Instituto de Biociências da
Universidade de São
Paulo para obtenção do
título de Doutor em
Ciências, Área de
Concentração:
Biologia/Genética

Orientador: Prof. Dr.
Maria Rita Passos Bueno

São Paulo
2013

Ficha Catalográfica

Ribeiro, Cintia
Estudo de genes candidatos aos Transtornos
do Espectro Autista.
100 páginas

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

1. Transtornos do Espectro Autista
2. Variações no número de cópias
3. Microarray-CGH

I. Universidade de São Paulo. Instituto de
Biociências. Departamento de Genética e Biologia
Evolutiva.

Comissão Julgadora:

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

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

Dedicatórias

Ao meu marido, Alfredo Tavares, pela amizade, amor, companheirismo e apoio incondicional durante os anos de realização deste trabalho.

A minha amada família pelo carinho, apoio e dedicação

*"Eu acredito demais na sorte.
E tenho constatado que, quanto mais duro eu trabalho,
mais sorte eu tenho".*

Thomas Jefferson

Agradecimentos

Agradeço ao meu marido e grande amigo Alfredo, quem me incentiva a batalhar por meus sonhos sem nunca desistir, pela compreensão, dedicação, amor e cumplicidade.

Agradeço a minha mãe Rosely, quem me ensinou a gostar de estudar, ser esforçada e determinada, ao meu pai Sérgio e a minha avó Anita pela imensa dedicação, amor, amizade e apoio.

Agradeço meus irmãos Alexandre e Marcelo pelos exemplos de perseverança, determinação e força de vontade.

Agradeço a Sílvia, Danielle, Patrícia, Érika, Maurício, Jefferson e as crianças mais lindas desse mundo: Ana Vitória, Eduardo, Marcelo Júnior, Pedro Henrique, Emilly e Joaquim.

Agradeço aos meus sogros pelo carinho.

Família, amo vocês mais do que tudo!

Agradeço a professora Maria Rita pela oportunidade, orientação, ensinamentos, amizade e por despertar em mim o senso crítico, tão importante para um pesquisador.

Agradeço a todos meus amigos queridos do laboratório por todas as risadas que demos juntos. As lições foram muitas mas talvez a mais importante tenha sido aprender a sorrir sempre, o sorriso deixa tudo mais leve. Ana, Andressa, Melina, Francine, Simone, Larissa e Joana, para vocês obrigada é pouco. São mais do que amigas, são irmãzinhas especiais, companheiras nas horas de diversão e trabalho. Cada almoço, cafezinho, *happy hours*, jogos de poker e festinhas foram momentos dos quais me lembrei com carinho sempre. Ana, você acrescentou muito a este estudo, Si e Andressa, vocês também trabalharam duro e me ajudaram demais.

Agradeço de coração por tudo, inclusive pelos apelidos criativos que inventou para mim Ana. Mel obrigada por ser tão companheira e me ajudar quando eu estava totalmente perdida.

Agradeço a todos do grupo de Autismo, em especial Eloísa e Karina.

As minhas amigas de hoje e sempre, por deixarem minha vida tão mais alegre e divertida: Meire, Francine Ribeiro, Flávia, Mariane e Juliana. Apesar da distância carrego vocês no coração.

Agradeço toda a equipe do Centro de Estudos do Genoma Humano, aos psiquiatras e médicos, em especial a Dra Débora Bertola, que colaboraram com este estudo.

Agradeço com muito carinho a Silvia Costa por me ensinar a técnica e a análise de microarray e a Dra Carla Rosenberg por permitir o uso dos equipamentos de seu laboratório.

Agradeço a Deus, quem me amparou nos momentos difíceis, me deu a força e a sabedoria necessária para conquistar mais este objetivo.

Agradeço o apoio financeiro da Fundação de Amparo a Pesquisa do estado de São Paulo – **FAPESP**.

Índice

Capítulo 1 –	Introdução Geral, Justificativa e Objetivos	----- 01
	1_1 Características Clínicas, Prevalência dos TEA e Aspectos Genéticos	----- 01
	1_2 Variações no número de cópias (CNVs) e TEA	----- 06
	1_3 Sequenciamento de última geração e TEA	----- 08
	1_4 Justificativa	----- 09
	1_5 Objetivos	----- 10
Capítulo 2 -	Materiais e métodos	----- 11
	2_1 Fluxograma do projeto	----- 11
	2_2 Pacientes e Controles	----- 12
	2_3 Customização da lâmina de aCGH	----- 13
	2_4 aCGH	----- 16
	2_5 Análises in silico	----- 19
	2_6 SNP-array	----- 20
	2_7 PCR quantitativo	----- 21
	2_8 MLPA e PCR	----- 23
	2_9 Sequenciamento de última geração	----- 25
Capítulo 3 -	Use of a custom-designed array-CGH to identify small CNVs in nonsyndromic Autism Spectrum Disorders.	----- 35
	Abstract	----- 35
	Resumo	----- 36
	Introduction	----- 37
	Materials and methods	----- 38
	Results	----- 43
	Discussion	----- 48
	Reference	----- 51
Capítulo 4 -	Characterization of microchromosomal rearrangements in syndromic autistic patients	----- 59
	Abstract	----- 59
	Resumo	----- 59
	Introduction	----- 60
	Materials and methods	----- 61
	Results	----- 63
	Discussion	----- 67
	Reference	----- 71
Capítulo 5 -	Search for genetic factors that modulate the severity of Autism Spectrum Disorders	----- 75
	Abstract	----- 75
	Resumo	----- 75
	Introduction	----- 76
	Materials and methods	----- 77

Results	-----	79
Discussion	-----	84
Reference	-----	86
Capítulo 6 -	Discussão Geral e Conclusões	-----
6_1 Discussão Geral	-----	91
6_2 Conclusões	-----	94
Referências	-----	
Bibliográficas	-----	95

Lista de Tabelas e Figuras

Capítulo 2 -	Figura 2.1	-----	11
	Figura 2.2	-----	15
	Figura 2.3	-----	18
	Figura 2.4	-----	19
	Figura 2.5	-----	24
	Tabela 2.1	-----	27
	Tabela 2. 2	-----	32
	Tabela 2.3	-----	33
Capítulo 3 -	Tabela 3.1	-----	45
Capítulo 4 -	Tabela 4.1	-----	65
Capítulo 5 -	Figura 5.1	-----	80
	Tabela 5.1	-----	82

Resumo

Os transtornos do espectro autista (TEA) são condições neuropsiquiátricas caracterizadas por padrões comportamentais estereotipados, ausência ou limitação de comunicação verbal e de interação social recíproca. Diversos estudos têm mostrado que esses transtornos possuem etiologia genética complexa e heterogênea, o que dificulta a identificação dos fatores causais. Estima-se que cerca de 70% dos casos de TEA são idiopáticos. Portanto, com o objetivo de identificar mecanismos etiológicos associados aos TEA, utilizamos as seguintes estratégias: customização de uma lâmina de microarray CGH que possibilite a detecção não só de grandes CNVs, mas também de alterações menores do que 10 kbp, em *exons* e regiões UTR de genes potencialmente candidatos; a comparação entre os tipos de rearranjos detectados em pacientes sindrômicos e em não sindrômicos e, ainda, a investigação mais detalhada de uma família com indivíduos portadores de transtorno autista e síndrome de Asperger. Foram avaliados 103 portadores de TEA não sindrômicos e 18 sindrômicos, sendo as taxas de detecção de alterações potencialmente patogênicas, respectivamente, de 11,6% e 38,9%. Dentre as alterações detectadas 44,4% são menores do que 10 Kbp. Portanto, a estratégia de usar uma lâmina customizada, com alta densidade de sondas complementares aos *exons* e regiões não codificantes de genes potencialmente envolvidos na etiologia dos TEA, capaz de detectar tanto alterações grandes quanto pequenas, parece ser relevante na tentativa de elucidar o maior número de casos possíveis e melhor compreender esses transtornos. Além disso, essa lâmina também pode ser utilizada como uma ferramenta para auxiliar o diagnóstico clínico e o aconselhamento genético com um custo mais acessível em comparação a outras comerciais ou ao sequenciamento de última geração. Cerca de 33,3% das CNVs observadas afetam região UTR, sugerindo que mutações nessas regiões podem explicar uma proporção significativa dos casos nos quais não são detectadas alterações através de outros testes genômicos, visto que a maioria desses ainda não exploram adequadamente regiões não codificantes. Entre os pacientes autistas não sindrômicos verificou-se que a maioria dos genes afetados por CNVs estão envolvidos em duas principais funções biológicas - sinapses glutamatérgicas e orientação axonal, sugerindo que TEA não sindrômico pode ser causado por disfunção em genes diferentes envolvidos em processos fisiológicos comuns.

Diferente do que observamos entre pacientes não sindrômicos, detectamos mais de uma alteração em um mesmo indivíduo ou alterações que englobam mais de um gene entre os pacientes sindrômicos, reforçando o modelo oligogênico para alguns casos de TEA. Por fim, os dados obtidos no estudo da família com portadores de síndrome de Asperger e transtorno autista sugere que a gravidade do quadro clínico possa estar relacionada ao número de mutações e possivelmente por duas mutações diferentes em ambos os alelos de um mesmo gene. Nossos resultados, além de apoiar o envolvimento dos genes *MDGA2*, *FHIT*, *HTR2A*, *SHANK2*, *GRIA3*, *ZNF778*, *PRKCa*, *CDH15*, *DIAPH3*, *GCH1*, *GRM5*, *MARK1*, *SLC17A6*, *IMMP2L*, *BZR API*, *SYNGAP1*, *ANK3*, *MAP1A*, *GABRR2* e *LAMC3* nos TEA também sugere novos genes candidatos: *LRRC7*, *LRRIQ3*, *CADPS1*, *NUFIP*, *SEMA3A*, *SNAP29*, *MBD2*, *GAD2*, *DGKH* e *PARD3*.

Abstract

The autism spectrum disorders (ASD) are neuropsychiatric conditions typically characterized by social deficits, communication difficulties, stereotyped or repetitive behaviors and interests. Several studies have shown that these disorders have a complex and heterogeneous genetic etiology, which makes difficult to identify the causal factors. Approximately 70% of cases are idiopathic. In order to identify etiological mechanisms associated with ASD, we have used the following strategies: customized a microarray CGH platform that allows detection not only of large CNVs, but also alterations smaller than 10 kbp in exons and UTR regions of potential candidate genes, the comparison between the types of rearrangements detected in syndromic and non-syndromic patients and further, more detailed investigation of a family segregating both autistic disorder and Asperger syndrome. We evaluated 103 nonsyndromic and 18 syndromic patients by the custom-designed array and the detection rate of possibly pathogenic alterations were, respectively, 11.6% and 38.9%. Among these CNVs, 44.4% are smaller than 10 kbp. Therefore, the strategy of using a custom-designed array, enriched with probes targeted to genes potentially involved in the ASD etiology and able to detect both large and small CNVs, seems to be relevant in an attempt to elucidate the largest number of cases and to better understand these disorders. Furthermore, this platform can also be used as a tool to support the clinical diagnosis and genetic counseling with a more affordable cost compared to conventional other or next-generation sequencing. Approximately 33.3% of the observed CNVs affect UTR region, suggesting that mutations in non-coding regions might explain a significant proportion of ASD cases negative for most genomic screenings, which still do not explore adequately these regions. Among nonsyndromic autistic patients we found that most of the genes affected by CNVs are involved in two main biological functions – glutamatergic synapses and axonal guidance, suggesting that nonsyndromic ASD can be caused by dysfunction in different genes of a few common physiological processes. In contrast to our findings in nonsyndromic patients, we detected more than one alteration in a single individual or alterations that involve more than one gene among the syndromic patients, reinforcing the oligogenic model for some cases of ASD. Finally, the data obtained in the study of the family segregating both Asperger syndrome and autistic disorder suggests that the

severity of ASD seems to be modulated by the number of hits and possibly by hits in both alleles of the same gene. Our results support the involvement of genes *MDGA2*, *FHIT*, *HTR2A*, *SHANK2*, *GRIA3*, *ZNF778*, *PRKCa*, *CDH15*, *DIAPH3*, *GCH1*, *GRM5*, *MARK1*, *SLC17A6*, *IMMP2L*, *BZRAP1*, *SYNGAP1*, *ANK3*, *MAP1A*, *GABRR2* and *LAMC3* in ASD etiology and also suggests new candidates: *LRRC7*, *LRRIQ3*, *CADPS1*, *NUFIP*, *SEMA3A*, *SNAP29*, *MBD2*, *GAD2*, *DGKH* and *PARD3*.

1_1 Características Clínicas, Prevalência dos TEA e Aspectos Genéticos

De acordo com o Manual Diagnóstico e Estatístico de Transtornos Mentais – DSM-IV – os Transtornos Globais do Desenvolvimento (TGD) compreendem: transtorno autista, transtorno desintegrativo da infância, síndrome de Asperger, transtorno global do desenvolvimento sem outra especificação (*Pervasive Developmental Disorders Not Otherwise Specified* – PDD-NOS) e a síndrome de Rett.

O transtorno autista se caracteriza por comprometimento do uso de múltiplos comportamentos não-verbais como por exemplo: contato visual direto, expressão facial, postura e linguagem corporal. Pode haver fracasso em desenvolver relacionamentos com seus pares e esse fracasso assume diferentes formas em diferentes idades. Os indivíduos mais jovens podem demonstrar pouco ou nenhum interesse em estabelecer amizades; os mais velhos podem ter interesse por amizades, mas não compreendem as convenções da interação social. Pode ocorrer uma ausência da busca espontânea pelo prazer compartilhado, interesses ou realizações com outras pessoas e falta de reciprocidade social ou emocional (DSM-IV).

Pode ocorrer atraso ou ausência total de desenvolvimento da linguagem falada. Em indivíduos que desenvolvem a fala, pode existir um acentuado comprometimento da capacidade de iniciar ou manter uma conversação, uso estereotipado e repetitivo da linguagem ou uma linguagem idiossincrática (linguagem que só pode ser entendida claramente por pessoas familiarizadas com o estilo de comunicação do indivíduo). Há um prejuízo acentuado no uso pragmático da linguagem evidenciado pela incapacidade de integrar palavras e gestos ou compreender os aspectos figurados da fala, tais como a ironia, duplo sentido e o humor (DSM-IV).

Quanto ao comportamento repetitivo, crianças com transtorno autista geralmente apresentam: preocupação com um ou mais padrões estereotipados e restritos de interesse, anormais em intensidade ou foco como, por exemplo, decorar número de telefones, datas, prefixos, estações de rádio, saber o nome de todas as espécies de dinossauros, saber itinerários de ônibus e trens, entre outros; adesão aparentemente inflexível a rotinas ou rituais específicos e não funcionais, por exemplo, percorrer exatamente o mesmo caminho para a escola todos os dias, enfileirar um número exato de brinquedos da mesma maneira repetidas vezes ou imitar a ação de um personagem de desenho inúmeras vezes, apresentar uma reação extremamente negativa a uma mudança mínima ocorrida no ambiente, tal como a mudança de uma cortina ou a mudança na posição de um móvel; maneirismos motores estereotipados e repetitivos (bater palmas, estalar os dedos, balançar-se); ou uma preocupação persistente com partes de objetos (rodinha dos carrinhos) (DSM-IV).

Outras condições médicas, tais como: anomalias no sono e alimentação, hiperatividade, ansiedade, impulsividade, comportamento hétero e auto agressivo, aumento do perímetrocefálico, respostas incomuns a estímulos sensoriais, epilepsia, hipotonía e distúrbios gastrointestinais, podem também estar presentes (AMERICAN PSYCHIATRIC ASSOCIATION, 2000).

Se a criança apresentar um período de pelo menos dois anos de desenvolvimento normal, refletido pela presença de comunicação verbal e não-verbal, relacionamentos sociais, jogos e comportamento adaptativo próprios da idade, seguido de regressão, ou seja, perda de habilidades já adquiridas, o diagnóstico adequado, segundo o DSM-IV, é de transtorno desintegrativo da infância. Essa condição geralmente está associada a grave retardamento mental.

A Síndrome de Asperger difere do transtorno autista devido à ausência de atraso na comunicação verbal e de prejuízo significativo de desenvolvimento cognitivo. Já portadores de PDD-NOS são chamados de “autistas atípicos” porque a manifestação clínica desses pacientes não satisfaz por completo os critérios para o diagnóstico de transtorno autista (VEENSTRA-VANDERWEELE; COOK, 2004).

A síndrome de Rett é geralmente causada por mutações *de novo* no gene *MeCP2*, localizado na região Xq28 e o padrão de herança dessa condição é dominante ligado ao X. Estas

mutações são geralmente letais em homens, o que explica o aparecimento quase exclusivo da síndrome no sexo feminino. A maioria das pacientes são aparentemente saudáveis ao nascimento e se desenvolvem normalmente até os 6-18 meses de vida, quando então ocorre um período de estagnação do desenvolvimento, seguido de rápida regressão da linguagem e controle motor. A característica mais marcante da síndrome é a perda de controle de movimentos voluntários das mãos e a sua substituição por movimentos automáticos repetitivos e estereotipados. Episódios de gritos e choro incontrolável são comuns no final do segundo ano de vida. Características adicionais incluem ataques de pânico, bruxismo, apnéia e ou hiperpnéia episódicas, ataxia da marcha e apraxia, tremores, convulsões (50% dos casos) e microcefalia adquirida (PENA, 2005). Durante os anos pré escolares, as crianças com Rett podem exibir dificuldades na interação social similares as observadas no transtorno autista mas essas tendem a ser temporárias (DSM-IV).

Devido a sobreposição clínica existente entre o transtorno autista, a síndrome de Asperger e o PDD-NOS, tornou-se comum o uso do termo Transtornos do Espectro Autista (TEA) para se referir a estas três categorias de TGD.

A nova edição do Manual de Classificação de Doenças Mentais da Associação Americana de Psiquiatria, DSM-V, a qual está prevista para ser lançada em maio de 2013, trará modificações no que diz respeito aos TEA. A principal mudança será a eliminação das categorias: transtorno autista, síndrome de Asperger e PDD-NOS. Existirá apenas uma denominação e essa também mudará, não será “transtornos do espectro autista” e sim “transtorno autista” (American Psychiatric Association DSM-5 Development, disponível *on-line*). O autismo era anteriormente definido por alterações na capacidade de comunicação, na capacidade de interagir socialmente e comportamento esteriotipado e repetitivo, já no DSM-V os dois primeiros sintomas citados se tornam apenas um. Essa mudança é justificada com o argumento de que a interação social depende da comunicação verbal. Ainda, segundo o DSM-IV, um desses sintomas deveria ocorrer pela primeira vez antes dos 3 anos. No DSM-V, esse limite com relação a idade desaparece e o que se coloca é que os sintomas devem surgir na infância (American Psychiatric Association DSM 5 Development, disponível *on-line*).

Visto que o DSM-V entrará em vigor apenas em maio de 2013, no presente trabalho seguimos as definições, classificações e categorias definidas no DSM-IV.

Fombonne, em 2009, relatou que a prevalência estimada dos TEA é de 6,37/1000 (equivalente a uma em cada 157 crianças). Quando o transtorno autista é considerado isoladamente, o estudo aponta uma prevalência de cerca de 2 casos para 1000 indivíduos, com uma razão sexual entre homens e mulheres de 4,2/1. A prevalência de casos de PDD-NOS é a mais alta dentre os TGD, estimada em 3,71/1000, e a de síndrome de Asperger é a mais baixa, calculada em 0,6/1000.

A prevalência dos TEA no Brasil foi recentemente estimada em 2,72/1000 (uma em cada 368 crianças) por um grupo de pesquisadores que avaliaram um bairro da cidade de Atibaia, São Paulo. Os autores discutem que a frequência mais baixa encontrada nesse estudo, em comparação a de outros trabalhos, pode ser explicada pela falta de conscientização da população local sobre esses transtornos, diagnósticos incorretos e falta de sistema para registro dos casos (PAULA *et al.*, 2011).

Existem fortes evidências de que fatores genéticos são determinantes para a etiologia dos TEA: estudos de famílias e de gêmeos demonstram uma taxa de concordância entre gêmeos monozigóticos de 60% a 90% em contraste à taxa de concordância de 1% a 10% vista em gêmeos dizigóticos. O risco de recorrência entre irmãos é de 2% a 8% e a herdabilidade é estimada em 90% (BAILEY *et al.*, 1995; FOLSTEIN; ROSEN-SHEDLEY, 2001; TIDMARSH; VOLKMAR, 2003). Além disso, cerca de 10% a 15% dos casos estão associados a doenças de herança monogênica, tais como: síndrome do X Frágil, esclerose tuberosa, neurofibromatose e em aproximadamente 3% dos casos, os indivíduos possuem alguma alteração cromossômica detectável por cariotípico (VORSTMAN *et al.*, 2006). Porém, em cerca de 70% dos casos de TE, o mecanismo etiológico é ainda desconhecido (SCHAAF *et al.*, 2011).

Até o início da década passada, o modelo de herança mais aceito para TEA era o multifatorial ou oligogênico, no qual o fenótipo resultaria de combinações de variantes alélicas comuns de pequeno efeito, somadas ou não com fatores ambientais. Assim, inicialmente muitos estudos de ligação e de associação foram realizados com o objetivo de identificar regiões e alelos de susceptibilidade para TEA. Yang e Gill (2007) revisaram 14 desses estudos e relataram um total de 94 *loci* descritos como significantes ou sugestivos de ligação. Entretanto, apenas os resultados obtidos para as regiões 7q21.2–q36.2, 16p12.1–p13.3, 1q21–q44, 2q24.1–q33.1, 3q21.3–q29, 5p12–p15.33, 6q14.3–q23.2, 10q23.3–q26.3, 17q11.1–q21.2, 19p13.11–p13.3,

1p12–p21.1, 4q21.3–q35.1, 10p12.1–p14, 13q12.13–q33.1, 15q13.1–q26.1 e 19q12–q13.12 foram replicados por pesquisas independentes. Quanto aos estudos de associação, também observamos baixa replicação de resultados, isto é, enquanto alguns estudos reportaram associação entre a variante de um gene e TEA, outros rejeitaram tal associação.

O surgimento de tecnologias mais avançadas, como por exemplo microarranjos de polimorfismos de base única (Single Nucleotide Polymorphism array; *SNP-array*), possibilitaram a identificação de polimorfismos em todo o genoma humano e não apenas em genes alvos. Isto possibilitou estudos de associação genômica ampla (frequentemente chamados pela sigla *GWAS* do nome em inglês *Genome-wide Association Studies*), onde todo o genoma é analisado em vez de apenas regiões ou genes candidatos. Os dados obtidos a partir desses estudos apontaram novas variações em um único nucleotídeo (SNVs - *single nucleotide variant*), porém, novamente observamos baixa replicação entre diferentes estudos e, além disso, o risco atribuído a cada SNV é baixo e não explica a alta herdabilidade dos TEA (EL-FISHAWAY; STATE, 2010).

Experimentos com microarranjos de DNA [SNP-array e Hibridização Genômica Comparativa por microarranjos (aCGH)] também permitiram a identificação de variações no número de cópias (*Copy Number Variation* - CNVs - duplicações ou deleções submicroscópicas) e, recentemente, tem sido sugerido que essas alterações possuem um importante papel na etiologia dos TEA, podendo estar associados ao mecanismo etiológico em cerca de 20% dos casos (MILES *et al.*, 2010). O grande número de estudos reportando CNVs exclusivas de TEA (não observadas em indivíduos aparentemente saudáveis) que compreendem um único gene, sugere que ao contrário do que se pensava anteriormente, uma parte dos casos de TEA pode ser explicada por modelos monogênicos, seja por herança de um alelo causador ou por mutações *de novo*. Nesse cenário as variantes seriam de alto efeito e, por isso, mais raras na população. O que contribuiria para alta frequência dos TEA seria, neste caso, a existência de diferentes alterações individualmente raras em um mesmo gene ou em genes diferentes (EL-FISHAWAY; STATE, 2010).

Recentemente, crescente esforço tem sido investido em técnicas de sequenciamento de última geração (*NGS – Next Generation Sequencing*) que permitem analisar toda a região codificante (exoma) ou todo o genoma humano em busca de pequenas alterações do DNA, como a troca de um único nucleotídeo e InDels (pequenas inserções ou deleções), as quais, por sua vez,

possam ser causativas do TEA. A taxa de indivíduos afetados nos quais são identificadas, através de NGS, mutações possivelmente causativas têm sido estimada em cerca de 15% (IOSSIFOV *et al.*, 2012; NEALE *et al.*, 2012; O'ROAK *et al.*, 2012; SANDERS *et al.*, 2012).

A seguir faremos uma revisão mais detalhada sobre os estudos de CNVs e sequenciamento de última geração.

1_2 Variações no número de cópias (CNVs) e TEA

A origem das CNVs é atribuída a mecanismos como NAHR (*Nonallelic Homologous Recombination*), NHEJ (*Nonhomologous End-Joining*), FoSTeS (*Fork Stalling and Template Switching*) e retrotransposição (para uma revisão sobre o assunto consultar Stankiewicz e Lupski, 2010). Essas alterações podem englobar parte ou todo um gene, ou podem englobar vários genes ou, ainda, elementos regulatórios. Assim, muitas dessas variações podem estar envolvidas na etiologia de doenças através do mecanismo de dosagem gênica ou da formação de produtos gênicos errôneos devido a ruptura de sequências codificantes (MERIKANGAS *et al.*, 2009; STANKIEWICZ; LUPSKI, 2010). Entretanto, estima-se que 12% a 15% do genoma humano varie significantemente em número de cópias e muitas dessas variações, presentes em indivíduos aparentemente saudáveis, abrangem um ou mais genes (CARTER *et al.*, 2007). Portanto, não é ainda conhecida a razão para algumas alterações apresentarem efeito fenotípicos e outras não, sendo a interpretação do significado clínico de uma CNV um grande desafio.

Lee *et al.* (2007) propôs que alterações *de novo* ou herdadas de um genitor afetado são mais prováveis de serem patogênicas, entretanto, não se pode afirmar que alterações herdadas de genitores saudáveis não estejam contribuindo para a manifestação de um fenótipo, uma vez que essas podem ter penetrância incompleta. De fato, como veremos adiante, uma parcela considerável das CNVs observadas em pacientes com TEA são herdadas de pais não afetados

(DEVLIN; SCHERER, 2012). Esses autores também ressaltaram a importância de observar se a CNV encontrada é similar a outras já catalogadas no banco de dados que compilam variações presentes em indivíduos saudáveis (DGV- *Database of Genomic Variants*; IAFRATE *et al.*, 2004; ZHANG *et al.*, 2006) ou em bancos de dados que copilam CNVs clinicamente relevantes (DECIPHER *Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources*) e também observar se a região deletada ou duplicada contém genes já associados a doenças (LEE *et al.*, 2007).

A frequência comumente reportada de CNVs detectadas em pacientes autistas é de 8 a 11,6%. Considerando-se apenas alterações *de novo* a frequência varia entre famílias simplex (apenas um indivíduo afetado - 5,6% a 10%) e famílias multiplex (mais de um afetado - 2% para 5,5%), sugerindo que nos casos familiais, alelos de susceptibilidade podem ser transmitidos por pais não afetados e atuam em conjunto para modular o risco para TEA (SEBAT *et al.*, 2007; MARSHALL *et al.*, 2008; CHRISTIAN *et al.*, 2008; QIAO *et al.*, 2009; PINTO *et al.*, 2010; ITSARA *et al.*, 2010; GIRIRAJAN *et al.*, 2010; BREMER *et al.*, 2011).

As CNVs são em geral extremamente raras e apenas algumas alterações recorrentes foram descritas, tais como as duplicações 15q11-q13, as microdeleções/microduplicações 16p11.2, as deleções 22q11-13 e alterações nos genes SHANK3, NLGN3, NLGN4 e NRXN1 (JAMAIN *et al.*, 2003; LAUMONNIER *et al.*, 2004; JACQUEMONT *et al.*, 2006; MOESSNER *et al.*, 2007). Embora as CNVs sejam particulares de cada paciente, há evidências de que essas abrangem genes envolvidos em funções biológicas comuns, mais significativamente implicadas na etiologia do TEA, como por exemplo sinapse e orientação axonal. Estes dados mostram que danos em diferentes genes envolvidos na mesma via biológica ou envolvidos no mesmo processo fisiológico podem ter um mesmo efeito final, levando ao fenótipo autista (SBACCHI *et al.*, 2010).

Cerca de 40% das CNVs detectadas em indivíduos autistas são herdadas de pais saudáveis, evidenciando a penetrância incompleta (DEVLIN; SCHERER, 2012). Em alguns casos, o autismo se comporta como um padrão de herança semelhante ao autossômico dominante, onde alterações em apenas um alelo estão relacionadas com a determinação do transtorno. Já em outros, o autismo está relacionado a um padrão de herança autossômico

recessivo, onde se observam mutações em ambos os alelos, como foi descrito em famílias com deleções em homozigose dos genes *PCDH10*, *DIA1* e *NHE92* (MORROW *et al.*, 2008).

Todas essas informações, derivadas dos estudos de CNVs, mostram o quanto os TEA são geneticamente heterogêneos, o que dificulta o entendimento de sua etiologia.

1_3 Sequenciamento de última geração e TEA

Foram publicados recentemente quatro estudos que analisaram o exoma de trios (pais e probando) e/ou quartetos (pais, probando e irmão não afetado). A porcentagem de pacientes com alterações *de novo*, consideradas patogênicas por introduzir um códon de parada prematuro, mudar a matriz de leitura do mRNA ou um sítio de excisão (“*splice*”) do pré-mRNA, variou entre os estudos de 10,3% a 17,2%. Os resultados obtidos foram concordantes quanto a maior ocorrência de mutações *de novo* em crianças cujos pais possuem idade avançada (35,9 a 58 anos) em comparação a crianças com pais mais jovens (30,9 a 35,9 anos). Também foi visto que cerca de 60% a 75% das alterações *de novo* detectadas estão em alelos de origem paterna, sugerindo que essas podem ter ocorrido durante a formação do gameta masculino. Outra descoberta foi maior taxa de alterações *de novo* em mulheres afetadas em comparação aos homens afetados (IOSSIFOV *et al.*, 2012; NEALE *et al.*, 2012; O’ROAK *et al.*, 2012; SANDERS *et al.*, 2012).

Os quatro estudos citados identificaram um total de 127 mutações possivelmente patogênicas porém, novamente nos deparamos com a grande heterogeneidade de *locus*, visto que apenas os genes *CHD8*, *KATNAL2*, *DYRK1A* e *POGZ* foram apontados como candidatos aos TEA por pelo menos dois desses estudos (IOSSIFOV *et al.*, 2012; NEALE *et al.*, 2012; O’ROAK *et al.*, 2012; SANDERS *et al.*, 2012). Avaliando o conjunto de genes alterados (excluindo alterações sinônimas e mudanças de aminoácido), Ben-David e Shifman (2012) relataram que a maior parte desses genes codificam proteínas envolvidas na regulação da transcrição durante o

desenvolvimento do cérebro, com enriquecimento significativo de proteínas reguladoras de cromatina ($p=0,021$ (corrigido); *Enrichment Score = 2,04*).

Assim como elucidar o significado clínico de uma CNV, a interpretação do grande volume de dados gerados nas análises de sequenciamento de última geração também é um grande desafio. As perguntas sobre o quanto cada uma das variações identificadas contribuem para o fenótipo ou, ainda, quantas dessas podem ser causativas independente da existência ou não de outras alterações, permanecem sem respostas. Desse modo, mesmo após o surgimento do sequenciamento de última geração, a maioria dos casos de TEA continuam sem causa conhecida.

1_4 Justificativa

Os TEA causam prejuízos funcionais e impactos graves durante toda a vida, tanto para os indivíduos afetados quanto para os familiares. Estes transtornos possuem alta prevalência sendo atualmente considerados um problema de saúde pública. O diagnóstico nos primeiros anos de vida possibilita que abordagens terapêuticas comportamental, educacional e farmacológica sejam aplicadas precocemente. Isto traz benefícios significativos na aquisição de habilidades funcionais, na redução de comportamentos mal-adaptativos e, consequentemente, aumento no sucesso de integração da criança à sociedade. A investigação dos mecanismos etiológicos de origem genética associados ao TEA pode trazer novas perspectivas para a compreensão desses transtornos bem como para o desenvolvimento de ferramentas de diagnóstico precoce. Visto que 70% dos casos são idiopáticos, hipotetizamos que uma parte desses pode ser causada por pequenas CNVs (>10 kbp), geralmente não detectáveis pelas plataformas comerciais de microarray disponíveis atualmente. Na tentativa de identificar novos genes candidatos e testar essa hipótese, customizamos uma lâmina de aCGH direcionada a detecção de pequenas CNVs em genes potencialmente envolvidos na fisiopatologia dos TEA. Essa lâmina também poderia ser utilizada como uma ferramenta para auxiliar o diagnóstico clínico e o aconselhamento genético, com um

custo mais acessível em comparação a outras comerciais ou ao sequenciamento de última geração.

1_5 Objetivos

O objetivo principal deste projeto é identificar os mecanismos etiológicos genéticos associados ao TEA.

Dentre os objetivos específicos estão: (A) Verificar se pequenas CNVs (>10 Kbp) não detectáveis pelas ferramentas moleculares disponíveis atualmente estão associadas ao TEA idiopáticos; (B) Customizar uma lâmina de aCGH contendo sondas para analisar todos os *exons*, junções *exons-intorns* e UTR de genes criteriosamente selecionados; (C) Verificar a frequência de microdeleções e microduplicações nesses genes em uma amostra de 120 afetados; (D) Verificar a proporção de CNVs herdadas e *de novo*; (e) Verificar se a gravidade do quadro clínico pode estar associada ao número de mutações em genes envolvidos com neurogênese.

2_1 Fluxograma do projeto

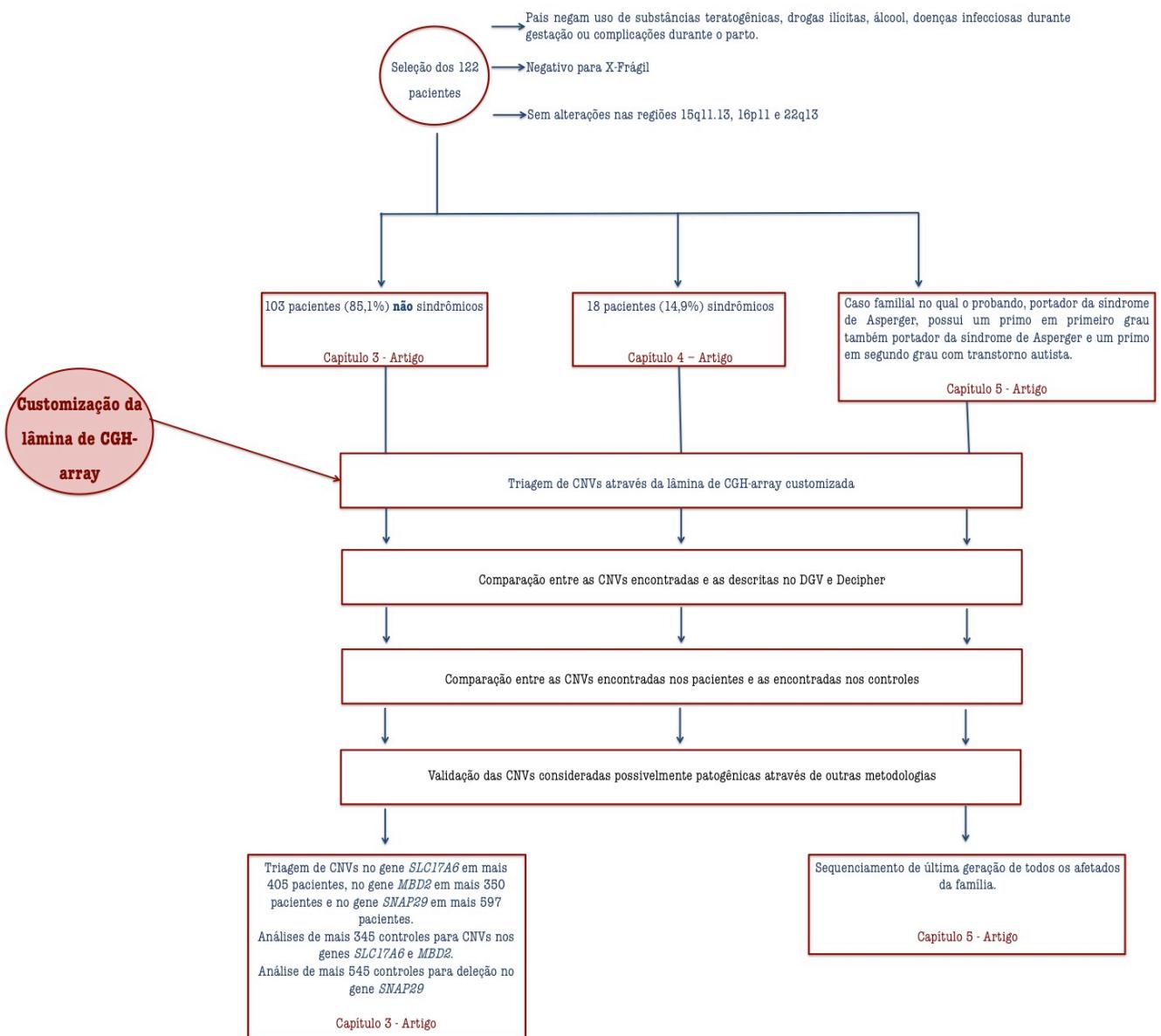


Figura 2.1 - Fluxograma esquematizando os passos executados nesse projeto.

2_2 Pacientes e Controles

Selecionamos inicialmente para o estudo um total de 122 pacientes portadores de TEA, sendo 94 (77%) do sexo masculino e 28 (23%) do sexo feminino. Posteriormente mais 405 pacientes (319 homens e 86 mulheres) com TEA foram analisados para CNVs no gene *SLC17A6*, 350 pacientes (315 homens e 35 mulheres) para alterações no gene *MBD2* e 597 pacientes (441 homens e 156 mulheres) para a deleção 3'UTR do gene *SNAP29*. Os diagnósticos foram realizados por psiquiatras e baseados no DSM-IV, ICD-10, Escala de Avaliação para Autismo Infantil – CARS (SCHOPLER *et al.*, 1980) e Questionário de Triagem para Autismo (ASQ - BERUMENT *et al.*, 1999). Grande parte dos pacientes foram encaminhados ao Centro de Estudos do Genoma Humano pelo Instituto de Psiquiatria, Hospital das Clínicas – FMUSP (equipe liderada pelo psiquiatra Dr. Estevão Vasdaz) e pelo Laboratório de genética do Hospital de Base, Faculdade de Medicina de São José do Rio Preto (equipe liderada pela Dra. Agnes Cristina Fett Conte). A idade dos pacientes analisados variou de 1 a 47 anos ($\mu=12,1$; $\sigma=8$). Na avaliação pela geneticista clínica doutora Débora Romeo Bertola do Instituto de Biociências, Departamento de Genética, foi observado que 18 pacientes possuem características dismórficas e/ou malformações congênitas sendo classificados nesse estudo como sindrômicos (detalhes do fenótipo na Tabela 4.1 – capítulo 4). Do total casos avaliados, 4 são familiais e os parentes afetados também foram posteriormente analisados.

As informações clínicas e o histórico familiar foram obtidas através do preenchimento da ficha de anamnese padrão. Não foram incluídos no estudo pacientes cujos pais reportaram consanguinidade, histórico de uso de medicamentos teratogênicos, drogas ilícitas, álcool ou doenças infecciosas durante a gravidez e complicações durante o parto. Também não incluímos pacientes com resultado positivo para X-Frágil (PCR – HADDAD *et al.*, 1996) ou com alterações nas regiões 15q11.13, 16p11 e 22q13 (SALSA MLPA KIT P343-B1 AUTISM-1; MRC-Holland, Amsterdam, The Netherlands) (testes realizados pelo Centro de Estudos do Genoma Humano - CEGH).

Amostras de DNA de 200 indivíduos brasileiros, aparentemente saudáveis, foram utilizados como controles nos experimentos de aCGH. Posteriormente, amostras de DNA de mais 545 controles brasileiros foram avaliados quanto a presença da deleção 3'UTR do gene *SNAP29* e 345 foram avaliados para CNV nos genes *SLC17A6* e *MBD2*.

A Comissão de Ética do Instituto de Biociências - Universidade de São Paulo aprovou esta pesquisa. Pacientes e controles foram incluídos somente após assinatura do Termo de Consentimento Livre Esclarecido pelas famílias ou outros responsáveis legais.

2_3 Customização da lâmina de aCGH

A técnica de aCGH permite a detecção de perdas (deleção) e ganhos (duplicação) no material genético através da comparação entre a hibridação do DNA teste *versus* DNA controle com as sondas dispostas, ou “imobilizadas”, na superfície de uma lâmina de vidro. Sondas são pequenas sequências de nucleotídeos (nesse caso, aproximadamente 60 pares de bases), complementares a regiões específicas do genoma que se pretende analisar.

O primeiro passo para a customização da lâmina de aCGH foi selecionar os genes. Desse modo, realizamos uma extensa revisão da literatura em busca de genes reportados como associados a etiologia dos TEA por pelo menos dois diferentes estudos. Através dessa criteriosa revisão criamos uma primeira lista de genes selecionados.

O segundo passo foi estudar detalhadamente as vias biológicas das quais participam esses genes com o objetivo de selecionar outros novos genes, ainda não reportados como candidatos aos TEA mas que participam das mesmas vias e desempenham funções tão importantes quanto a de outros já associados a esses transtornos. Por exemplo, o envolvimento do sistema GABAérgico na patogênese dos TEA têm sido sugerido por estudos clínicos e genéticos (BLATT *et al.*, 2001; BLATT, 2005; MA *et al.*, 2005; SCHMITZ *et al.*, 2005; FATEMI *et al.*, 2009),

assim sendo, através do estudo de toda a via GABAérgica, além dos genes já reportados como candidatos, selecionamos para inclusão em nossa lâmina outros genes dessa via que também possam estar envolvidos na etiologia do autismo.

Com o objetivo de melhor ilustrar, citamos um outro exemplo: cerca de 80% das mulheres com síndrome de Rett preenchem os critérios para diagnóstico de autismo. Essa síndrome é causada por mutações no gene *MeCP2*, localizado na região Xq28. O estudo da via de Sinalização da Metilação do DNA e Repressão da Transcrição, da qual a proteína MeCP2 participa, nos possibilitou selecionar outros genes com papel tão importante quanto no mecanismo de silenciamento gênico e que, por este motivo, também podem ser candidatos para TEA. Na figura 2.2 pode-se observar as dez vias canônicas mais significativamente representadas na lâmina de aCGH customizada.

Como já citado, cerca de 10 a 15% dos casos de autismo estão associados a uma condição Mendeliana conhecida, tais como: síndrome do X Frágil (*FMR1*), Esclerose Tuberosa (*TSC1*, *TSC2*) Neurofibromatose (*NFI*), entre outras. Portanto, os genes relacionados com essas condições, assim como outros genes cujos produtos proteicos interagem direta ou indiretamente com esses, foram também incluídos.

Uma última estratégia utilizada foi baseada nos relatos de alterações citogenéticas com uma pequena frequência de recorrência em autistas: duplicações da região 15q11-13 (1% a 3% dos casos), microdeleções ou microduplicações na região 16p11.2 (1% dos casos) e deleções nas regiões 2qter e 22qter (<1% casos) (GOIZET *et al.*, 2000; VEENSTRA-VANDERWEELE *et al.*, 2004; KUMAR *et al.*, 2008; WEISS *et al.*, 2008; CHRISTIAN *et al.*, 2008). Genes expressos preferencialmente no cérebro e/ou genes que codificam proteínas com funções cerebrais, localizados nessas regiões, também foram selecionados.

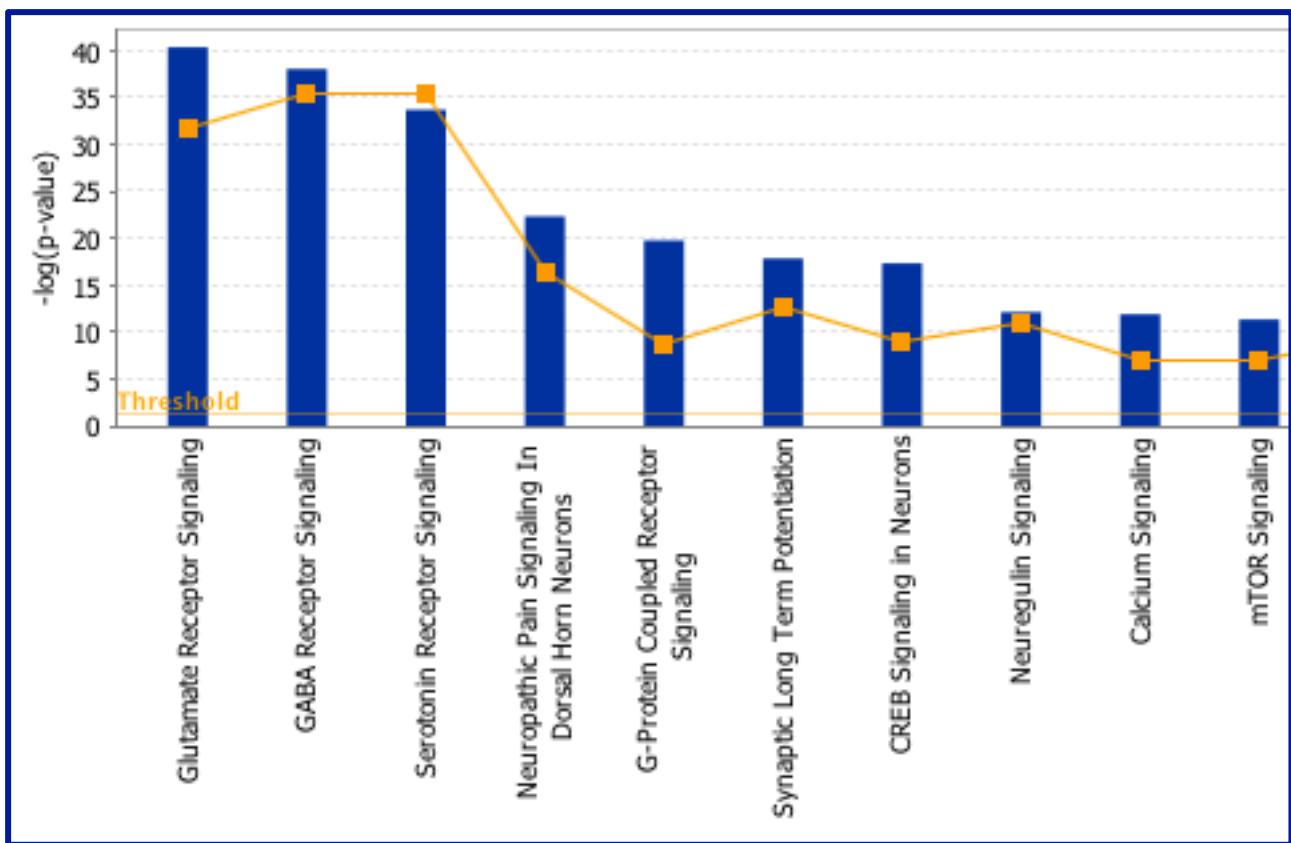


Figura 2.2 Vias canônicas mais significativamente representadas na lâmina de aCGH customizada.

Desse modo, selecionamos para inclusão em nossa lâmina um total 269 genes (Tabela 2.1) O nosso desenho experimental permitiu analisar todos os *exons*, junções *exons/introns* e regiões não traduzidas (UTR, sigla do inglês *Untranslated Region*) de cada um dos 269 genes. Para a seleção das sondas, utilizamos o programa eArray (versão 6.2) disponível no endereço: <https://earray.chem.agilent.com/earray/>. Importante ressaltar que não foram selecionadas sondas em regiões do genoma altamente repetitivas (evitar hibridização inespecífica) e em regiões que correspondem a sítios de restrição para as enzimas *AluI* e *RsaI* (devido aos procedimentos da técnica que será explicada mais adiante). Também com o objetivo de garantir a qualidade da ferramenta de análise, desenvolvemos e aplicamos mais dois filtros:

- O programa eArray fornece uma pontuação (*probe score*) para cada sonda baseada em sua qualidade. Essa pontuação varia de 0 (ruim) a 1 (excelente). Desse modo, excluímos todas as sondas com pontuação menores do que 0,8.

- Excluímos sondas menores ou iguais a 59 pares de base porque quanto maior o comprimento da sonda maior a sua especificidade e, consequentemente, menor a chance de hibridar em uma região não desejada do genoma.

Após a aplicação desses dois filtros, avaliamos a distribuição das sondas remanescentes sob os genes. Nessa etapa as sondas foram selecionadas para que ficássemos com no mínimo três e no máximo seis sondas com sobreposição parcial de suas sequências. Portanto, não há intervalo entre uma sonda e outra. Desse modo, com a sobreposição entre as sondas obtemos maior resolução na região de interesse e, assim, podemos detectar alterações menores cujas metodologias atuais não detectam.

No total foram selecionadas 59.029 sondas, esse número inclui também sondas controles e sondas de normalização necessárias para a boa qualidade do experimento. Na empresa *Agilent Technologies*, com auxílio da robótica, essas sondas foram dispostas de forma ordenada na superfície de uma lâmina semelhante às lâminas de microscópio óptico. Selecionamos a plataforma 8X60K, isto significa que a lâmina tem oito áreas com as 59.029 sondas em cada uma. Desse modo, é possível avaliar oito pacientes em um experimento.

2.4 aCGH

A concentração e a qualidade dos DNAs (obtidos a partir de sangue venoso periférico) foram verificadas no aparelho *NanoDrop ND 1000* (*espectofotômetro da Peqlab Biotechnologie™*) e em gel de agarose 1,2%. Utilizamos 200ng de DNA de cada amostra. Estes quando analisados no NanoDrop exibiam razão A260/280 entre 1,8 e 2,0 (indicando baixa contaminação por proteínas) e razão $A260/280 \geq 2$ (indicando baixa contaminação por compostos orgânicos).

Inicialmente as moléculas de DNA foram fragmentadas pelas enzimas de restrição *AluI* e *RsaI*. Subsequentemente, esses fragmentos foram amplificados e marcados com fluoróforos de cores diferentes (*cyanine3* e *cyanine5*) para serem detectados pelo laser do scanner (*Agilent's*

High-Resolution C Scanner). Para tanto utilizamos o *Genomic DNA Enzymatic Labeling Kit, Agilent Technologies®*. Após a marcação dos DNAs, esses foram purificados utilizando-se colunas *Microcon YM30 (Millipore®)* para a retirada dos nucleotídeos não incorporados. A quantificação do DNA marcado foi realizado no *NanoDrop ND 1000 (espectofotômetro da Peqlab Biotechnologie™)*. Medimos a absorbância em A260nm (DNA), A550nm (cyanine3) e A650nm (cyanine5). Para medir a atividade específica (AE) do DNA genômico marcado, usamos o cálculo:

$$AE = \frac{(\text{pmol}/\mu\text{l} \text{ de cyanine3 + cyanine5})}{(\mu\text{g}/\mu\text{L gDNA})}$$

Em seguida as amostras foram tratadas com *Cot-1 DNA humano*, com o intuito de bloquear ligações inespecíficas, precipitadas com acetato de sódio e etanol e, por fim, ressuspensas em um tampão TE (pH 8,0; *Integrated DNA Technologies, Inc.*, Coralville, IA; United States). Para hibridação utilizamos o *Agilent Oligo aCGH Hybridization Kit*, de acordo com especificações do fabricante.

Os DNAs marcados foram hibridados às lâminas em um forno de hibridação a 65°C. Cada lâmina possui 8 áreas. Em uma área da lâmina o DNA do paciente X está marcado com cy3 sendo comparado com o DNA do paciente Y marcado com cy5 e, em outra área, o DNA do paciente X está marcado com cy5 sendo comparado com o DNA do paciente Z marcado com cy3. Ou seja, cada amostra está em duplicata sendo comparada a duas outras amostras diferentes.

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

As imagens foram obtidas com o uso do scanner Agilent e do programa *Feature Extraction v10.5*. As análises foram feitas com o programa *Agilent Genomic Workbench Lite Edition v6.5*, usando o algoritmo estatístico ADM-2 e limiar 7.0. Apenas as alterações que continham no mínimo três oligonucleotídeos consecutivos com razão $\log_2 \text{cy3/cy5}$ alterada foram

consideradas pelo programa como uma possível alteração no número de cópias de determinado segmento genômico. Nas figuras (2.3 e 2.4) abaixo são mostrados exemplos de que como os resultados são visualizados utilizando o programa citado acima.

Somente as CNVs não observadas nos 200 controles brasileiros e não semelhantes a alterações descritas no DGV foram consideradas possivelmente patogênicas e confirmadas através de outras plataformas de microarray (Agilent SurePrint G3 Human CGH Microarrays 4×180K, SNP-array platform GeneChip Human Mapping 500K Array Set Affymetrix), PCR quantitativo e/ou Amplificação de Múltiplas Sondas Dependentes de Ligação (MLPA, sigla do inglês *Multiplex Ligation-dependent Probe Amplification*, MRC-Holand).

Para os experimentos de aCGH utilizando a plataforma 4X180K da empresa *Agilent Technologies*, os procedimentos foram os mesmos descritos acima. A diferença é que partimos de uma concentração maior de DNA (500ng). As demais técnicas serão descritas abaixo.

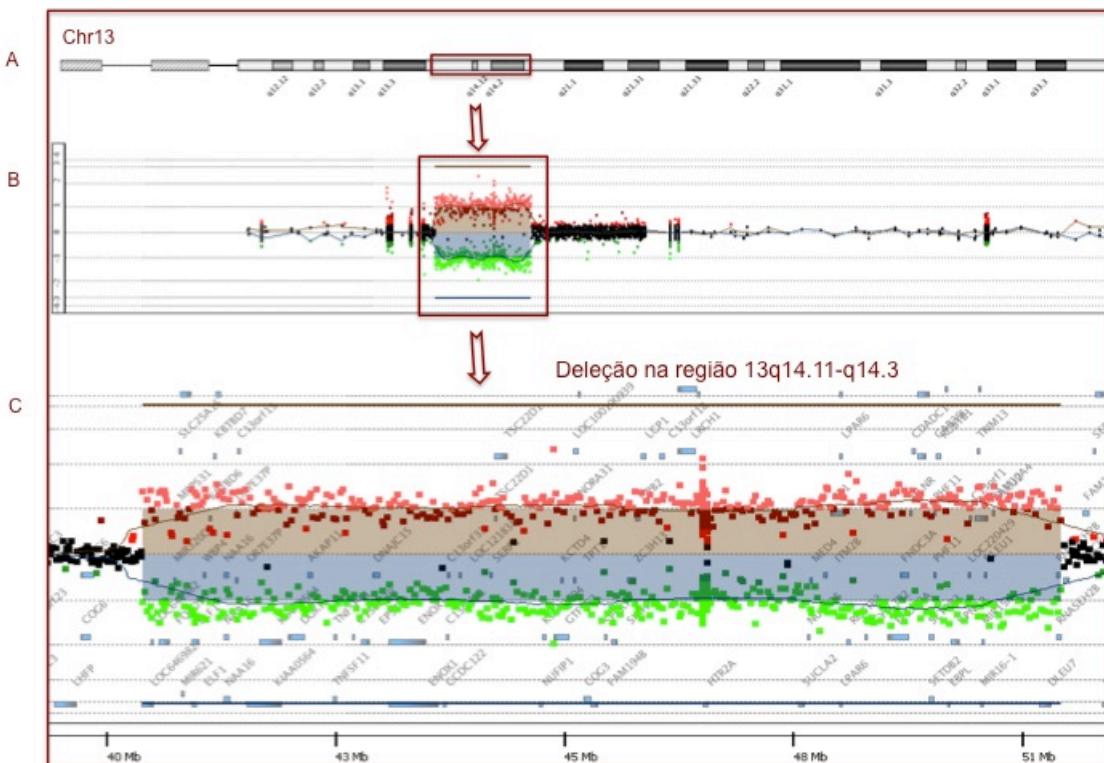


Figura 2.3 **Deleção na região cromossomo 13q14.11-q14.3.** A) e B) Posição da mutação no cromossomo; B) Eixo x= sondas de acordo com a posição genômica. Eixo y= valor do log₂ cy3/cy5. Região hachurada em cinza corresponde a área da lâmina em que o DNA do paciente está marcado com cy3. Região hachurada em marrom corresponde a área da lâmina em que o DNA do paciente está marcado com cy5.

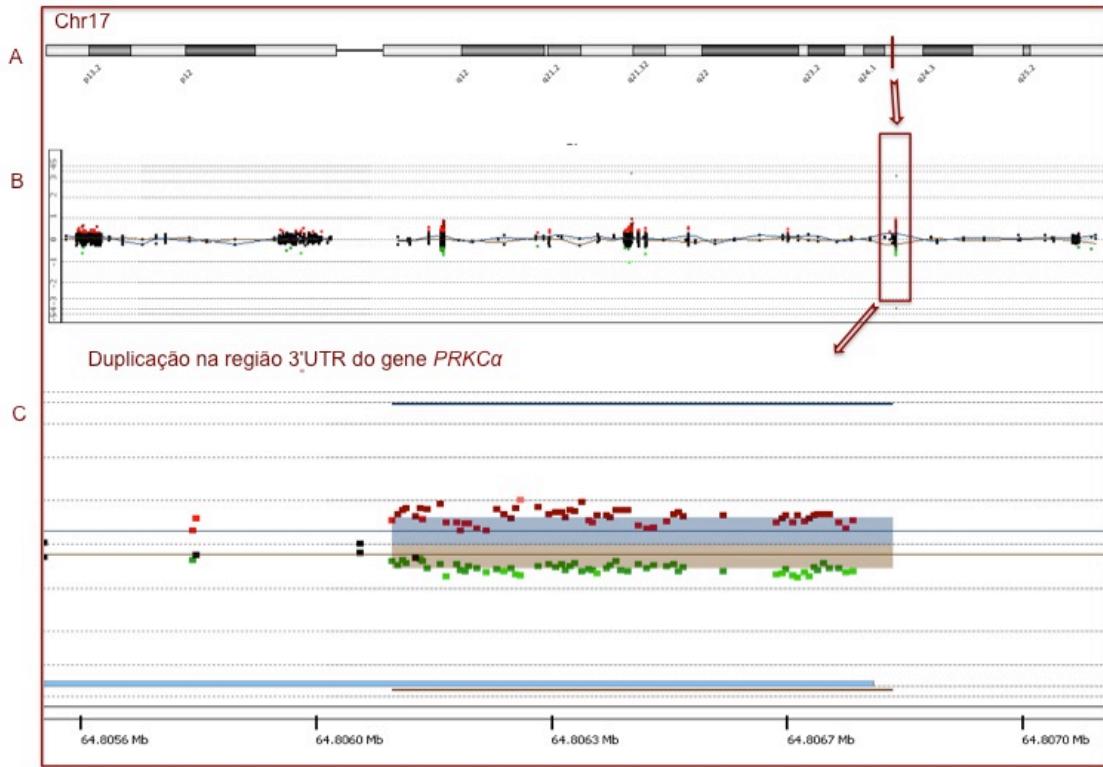


Figura 2.4 **Duplicação 3' UTR do gene PRKCa.** A) e B) Posição da mutação no cromossomo; C) Eixo x= sondas de acordo com a posição genômica. Eixo y= valor do \log_2 cy3/cy5. Região hachurada em cinza corresponde a área da lâmina em que o DNA do paciente está marcado com cy3. Região hachurada em marrom corresponde a área da lâmina em que o DNA do paciente está marcado com cy5;

2.5 Análises in silico

Consultamos os banco de dados de variações genômica para avaliar se as alterações encontradas através do aCGH customizado já haviam sido detectadas em indivíduos aparentemente saudáveis (DGV - <http://dgvbeta.tcag.ca/dgv/app/home>) e/ou em indivíduos portadores de TEA ou outros distúrbios neuropsiquiátricos (<http://decipher.sanger.ac.uk>). Para verificar a possível associação dos genes alterados com outras doenças utilizamos o OMIM

(*Online Mendelian Inheritance in Man* - <http://omim.org>) e o GAD (*Genetic Association Database* - <http://geneticassociationdb.nih.gov>).

As posições genômicas de início e fim dos segmentos UTR deletados ou duplicados foram fornecidas ao *UCSC Genome Browser* (KENT *et al.*, 2002; MEYER *et al.*, 2012) para análises *in silico*. Foram selecionadas as ferramentas *TargetScanHuman* 5.1 (LEWIS *et al.*, 2005; GRIMSON *et al.*, 2007; FRIEDMAN *et al.*, 2009), *phastCons* (MURPHY *et al.*, 2001; CHIAROMONTE *et al.*, 2002; KENT *et al.*, 2003; SCHWARTZ *et al.*, 2003; BLANCHETTE *et al.*, 2004; SIEPEL *et al.*, 2005; POLLARD *et al.*, 2009) e polyA_DB database (ZHANG *et al.*, 2005; CHENG *et al.*, 2006) para verificar a existência de, respectivamente, sítios de regulação por miRNA, elementos altamente conservados entre 46 espécies de vertebrados e sítio de poliadenilação. Para verificar a existência de sítios de ligação para fatores de transcrição utilizamos a ferramenta disponível *online* Consite (<http://asp.ii.uib.no:8090/cgi-bin/CON SITE/consite>; *TF score cutoff 98%*; SANDELIN *et al.*, 2004).

2_6 SNP-array

Para a confirmação das alterações encontradas utilizamos microarray comercial, plataforma *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 *Nsp* e 238 mil para a *Sty*. A distribuição resultante das sondas considerando os 500.000 SNPs não é uniforme nas lâminas, tendo distância física mediana entre eles de 2,5kb e 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; BUCAN; WANG, 2008) e *dChip* (LI E HONG-WONG 2001; LIN *et al.* 2004), com configuração padrão dos parâmetros. Apenas alterações detectadas por pelo menos dois programas diferentes foram consideradas.

2_7 PCR quantitativo

A padronização da técnica de Reação em Cadeia da Polimerase (PCR) em tempo real para quantificar número de cópias do DNA foi realizada conforme descrito por Ponchel *et al.* (2003). Desenhamos oligonucleotídeos localizados dentro das regiões deletadas ou duplicadas e também usamos oligonucleotídeos controles, para os genes referência *G6PDH* e *hGAPDH*, conforme dados publicados anteriormente (WEKSBERG *et al.*, 2005 E PONCHEL *et al.*, 2003) (Tabela 2.2).

As reações de amplificação em tempo real foram realizadas utilizando 100ng de DNA e 0,5X *Syber Green PCR Master Mix* (*Applied Biosystems®*, EUA) em um volume final de reação igual a 25 μ L. Testamos diferentes concentrações de oligonucleotídeos e as mais adequadas para o nosso ensaio foram aquelas em que a combinação de concentrações permitiu uma razão de gene alvo/gene referência igual a 1 em uma amostra de DNA controle. Esse índice é calculado usando o método ΔC_t :

$$\Delta C_t = *C_t \text{ alvo} - *C_t \text{ referência}$$

$*C_t \rightarrow$ número de ciclos em que a fluorescência gerada dentro de uma reação de amplificação do DNA ultrapassa um valor limite. Delta Ct igual a 0 indica uma razão de 1.

Para permitir comparação de número de cópias entre afetado e controle, cada conjunto de oligonucleotídeos (referência e teste) deve ter uma eficiência de amplificação semelhante. Esse **importante** parâmetro foi verificado através da análise do *slope* de uma curva padrão representada como um gráfico de regressão linear semi-log do valor de ΔC_t em comparação ao log da quantidade inicial de ácido nucléico. Na equação $y=ax+b$. O valor que aparece na posição **a** corresponde ao *slope* da reta e, se a eficiência do alvo e o controle estiverem semelhantes, esse valor deverá ser entre **-0,1 e 0,1**.

Os cálculos foram realizados de acordo com a fórmula matemática:

$$KC_t = \left[\frac{AC_{tR} - C_{tR}}{S_R} \right] \times S_T + C_{tT}$$

Onde:

- KC_t controle $\rightarrow AC_{tR}$: média de todos os valores de C_t obtidos para o oligonucleotídeos *referência*, usando DNA controle e teste, em uma mesma corrida de PCR em tempo real; C_{tR} : média dos valores de C_t obtidos para *primer referência* utilizando DNA controle; S_R *slope do primer referência*; S_T : *slope do primer teste*; C_{tT} média dos valores de C_t obtidos para *primer teste* utilizando DNA controle.
- KCt teste $\rightarrow AC_{tR}$: média de todos os valores de C_t obtidos para os oligonucleotídeos *referência*, usando DNA controle e teste, em uma mesma corrida de PCR em tempo real; C_{tR} : média dos valores de C_t obtidos para *primer referência* utilizando DNA teste; S_R *slope do primer referência*; S_T : *slope do primer teste*; C_{tT} média dos valores de C_t obtidos para *primer teste* utilizando DNA teste.

Fold CNV change= KC_t, controle – KC_t, teste

→ Resultado igual a $0 \pm 0,35$ significa que não há alteração; $+1 \pm 0,35$ significa que há ganho de uma cópia e $-1 \pm 0,35$ há perda de uma cópia.

Inicialmente comprovamos a eficiência da técnica na detecção de variações no número de cópias, realizando o *PCR em tempo real* para alterações que já haviam sido confirmadas por outros métodos (duplicação do gene *STS*, confirmada por aCGH plataforma 180K *Agilent Technologies* e deleção do gene *SHANK2* confirmada por array-SNP 500K *Affymetrix*).

2_8 MLPA e PCR

Para validação das CNVs nos genes *SLC17A6* e *MBD2*, detectadas previamente através do aCGH customizado, e também para triagem de novas alterações nesses genes em um número maior de pacientes, realizamos a técnica de MLPA. Foram utilizadas sondas *in house*, desenhadas com o uso do programa *Human MLPA Probe Design* (H-MAPD) (ZHI; HATCHWELL, 2008) em colaboração com a aluna de doutorado da UNESP Ana Luiza Bossolani, orientanda da Profa. Dra. Agnes Cristina Fett Conte. Para garantir a qualidade das sondas todos os parâmetros recomendados pela empresa MRC-Holland foram obedecidos (<http://www.mlpa.com/WebForms/WebFormMain.aspx?Tag=zjCZBtdOUyAt3KF3wRZhAPz9QEm7akikAm7AOEGw1vtZvffaZPOiSig8uqel7Yd>). As sondas foram sintetizadas pela empresa *IDT Technologies*. As sequências dos genes *SLC17A6* e *MBD2* foram obtidas do *Refseq (Reference Sequence Project -* <http://www.ncbi.nlm.nih.gov/books/NBK21091/>; PRUITT, et al., 2012). O gene *SLC17A6* (NM_020346.2) possui 12 *exons* e não foi possível desenhar sonda

apenas para o *exon* 1. O gene *MBD2* (NM_003927.4) possui 7 *exons* e foram desenhadas sondas para os *exons* 1, 2, 5, 6 e 7 (Tabela 2.3).

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 utilizamos o kit EK1 (SALSA MLPA EK1 reagent kits - MRC-Holland) e para controle dos experimentos utilizamos 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. Na figura (2.5) abaixo é mostrado um exemplo de como os resultados são visualizados utilizando o programa *Genemarker*.

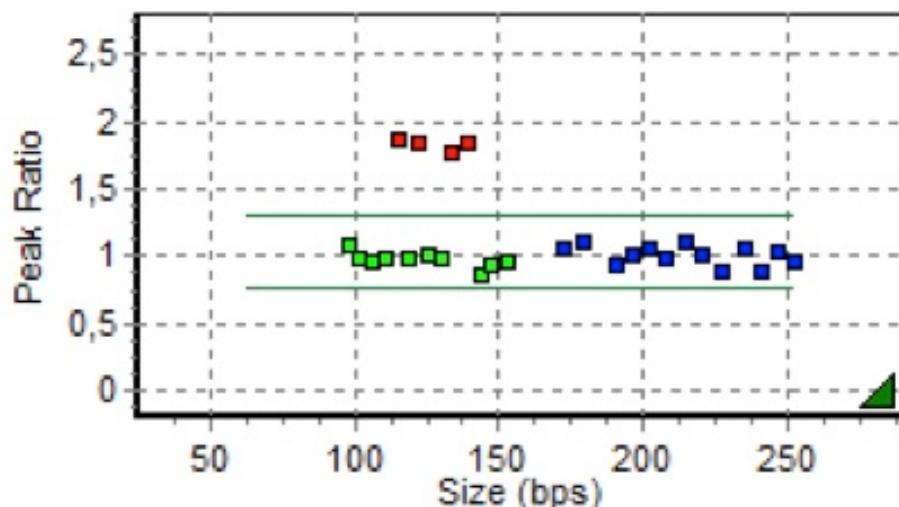


Figura 2.5 MLPA mostrando um paciente com duplicação das sondas localizadas nos *exons* 9, 10, 11 e 12 do gene *SLC17A6* (vermelho), em verde estão as outras sondas, localizadas nos outros *exons* do gene e em azul estão representadas as sondas controle do kit comercial P200 (MRC-Holland), analisados no programa *GeneMarker* v.1.97, com intervalo de classificação de deleção <0,75<normal<1,3< duplicação. As sondas 9, 10, 11 e 12 apresentaram valores numéricos após normalização dos dados de duplicação de 1,847; 1,832; 1,874 e 1,776, respectivamente.

A analise da região 3' UTR do gene *SNAP29* em uma amostra maior de pacientes e de controles foi realizada pela técnica de PCR, e o produto analisado em gel de agarose 1,5%. A presença de deleção pode ser observada pela diferença entre os tamanhos dos fragmentos, isto é, indivíduos portadores da deleção apresentaram dois fragmentos: um de 770pb (normal) e outro de 528pb (deletado), enquanto que indivíduos sem a deleção apresentaram apenas o fragmento de 770pb. Os oligonucleotídeos utilizados na PCR foram desenhados pelo programa Gene Runner, versão 3.0.5 e NCBI/Primer-Blast (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) (Tabela 2.2).

2_9 Sequenciamento de última geração

O sequenciamento de última geração foi realizado no laboratório chefiado pelo professor doutor Steven Scherer, localizado no *Centre for Applied Genomics, Hospital for Sick Children*, Toronto, Canadá. Brevemente, a preparação da biblioteca foi realizada de acordo com o protocolo *Agilent's SureSelect XT Target Enrichment System* (Agilent Technologies; Version 1.5, November 2012), para captura de sequencias exónicas foi utilizado *Agilent SureSelect Human All Exon 50-Mb V4®* (Agilent Technologies), as sequencias capturadas foram amplificadas através do GeneAmp PCR System 9700 (Applied Biosystems) usando *Herculase II Fusion DNA Polymerase* (Agilent Technologies), posteriormente foram purificadas com *AMPure® XP beads* (Agencourt®) e, por fim, para o sequenciamento aos pares (*paired-end sequencing*) foi utilizado o equipamento Illumina HiSeq 2000 (Illumina©) de acordo com as instruções do fabricante. As sequencias foram mapeadas tendo como referência o genoma humano hg19 (obtido no *UCSC Genome Browser*; KENT *et al.*, 2002; MEYER *et al.*, 2012) usando o programa BWA versão 0.6.5a (LI *et al.*, 2009). A ferramenta *Genome Analysis Tool Kit* (GATK – versão 1.1.28) (Broad Institute) foi utilizada para refinar o alinhamento das sequencias. A detecção de SNPs e indels (inserções/deleções) foi realizada também pela ferramenta GATK. SIFT e PolyPhen foram utilizados para análises *in silico* da patogenicidade das variantes encontradas (NG; HENIKOFF *et*

al., 2003; ADZHUBEI *et al.*, 2010). Somente mutações não polimórficas em regiões codificantes e junções *exons-introns* causando códon de parada prematuro, ganho ou perda de sítio de *splice* ou erro na matriz de leitura foram consideradas como potencialmente prejudicais.

Tabela 2.1 Genes selecionados e suas posições genômicas de acordo com o GRCh37/hg19.

Gene	Posição
RFWD2	chr1:175913966-176176370
ASTN1	chr1:176830203-177134024
HTR1D	chr1:23518388-23521222
HTR6	chr1:19991779-20006054
SLC1A7	chr1:53552855-53608289
DPYD	chr1:97543301-98386615
KCNAB2	chr1:6086379-6160523
NOS1AP	chr1:162039580-162339813
MARK1	chr1:220701567-220837798
GABRD	chr1:1950767-1962192
RbBP4	chr1:33116748-33151810
AKT3	chr1:243651534-244006553
HDAC1	chr1:32757707-32799224
DISC1	chr1:231762560-232177016
PRKAA2	chr1:57110989-57181007
SYT11	chr1:155829289-155854989
ARID4B	chr1:235330210-235491532
MTOR	chr1:11166588-11322608
SLC1A4	chr2:242295710-242434255
SLC25A12	chr2:65215578-65250997
GAD1	chr2:172639914-172750813
SAP130	chr2:171673199-171717657
CACNB4	chr2:128698790-128784869
PDK1	chr2:152689287-152955593
ERBB4	chr2:173420778-173463860
INPP1	chr2:212240441-213403352
CNTNAP5	chr2:191208195-191236390
DPP10	chr2:124782863-125672861
SCN2A	chr2:115199898-116601936
SCN1A	chr2:166095911-166248818
RAPGEF4	chr2:166845670-166930149
PIKFYVE	chr2:173600524-173917618
NRXN1	chr2:209130990-209223474
HTR2B	chr2:50145643-51259674
GPC1	chr2:231972951-231989824
HDLBP	chr2:241375114-241407494
FARP2	chr2:242166678-242255254
KIF1A	chr2:241653184-241759624
PASK	chr2:242045514-242088878
GBX2	chr2:237074306-237076652
HTR3E	chr3:121312169-121349139
HTR3C	chr3:183817966-183824782
HTR3D	chr3:183770834-183778459
HTR1F	chr3:183749331-183757156
GRM2	chr3:88031725-88042919
GRM7	chr3:51741080-51752623
SLC6A11	chr3:6902801-7783218
SLC6A1	chr3:10857916-10980144
OXTR	chr3:11034419-11080934
CACNA2D3	chr3:8792096-8811300
CACNA2D2	chr3:54156692-55108582
MBD4	chr3:50400232-50540892
PLD1	chr3:129149792-129158852

<i>CNTN4</i>	chr3:171318619-171528273
<i>NLGN1</i>	chr3:2142246-3099644
<i>CHL1</i>	chr3:173116243-174001116
<i>SYN2</i>	chr3:238649-451095
<i>C3orf58</i>	chr3:12045861-12233530
<i>CNTN3</i>	chr3:143690639-143711209
<i>SLC9A9</i>	chr3:74311721-74570343
<i>GAP43</i>	chr3:142984064-143567346
<i>TRPC1</i>	chr3:115342150-115440334
<i>FBXO40</i>	chr3:142443265-142526727
<i>GRIA2</i>	chr4:158141735-158287224
<i>GABRA2</i>	chr4:46251580-46392056
<i>GABRA4</i>	chr4:46920918-46995580
<i>GABRB1</i>	chr4:47033294-47428445
<i>GABRG1</i>	chr4:46037788-46126082
<i>SAP30</i>	chr4:174292092-174298682
<i>GLRB</i>	chr4:157997276-158093242
<i>GLRA3</i>	chr4:175563197-175750465
<i>PCDH10</i>	chr4:134070469-134112731
<i>NPY5R</i>	chr4:164265090-164273084
<i>NPY1R</i>	chr4:164245116-164253748
<i>TRPC3</i>	chr4:122800182-122872909
<i>ANXA5</i>	chr4:122589152-122618147
<i>HTR1A</i>	chr5:63256278-63257546
<i>HTR4</i>	chr5:147861119-148033741
<i>GRM6</i>	chr5:178405331-178422124
<i>GRIA1</i>	chr5:152870083-153193428
<i>GABRA1</i>	chr5:161274196-161326963
<i>GABRA6</i>	chr5:161112657-161129598
<i>GABRB2</i>	chr5:160715435-160975130
<i>GABRG2</i>	chr5:161494647-161582544
<i>GABRP</i>	chr5:170210722-170241048
<i>NRG2</i>	chr5:139227259-139422879
<i>Rictor</i>	chr5:38938021-39074501
<i>PRKAA1</i>	chr5:40759481-40798297
<i>AHT1</i>	chr6:101846860-102517957
<i>GRIK2</i>	chr6:146348781-146758731
<i>GRM1</i>	chr6:78171947-78173120
<i>HTR1B</i>	chr6:87647023-87726390
<i>HTR1E</i>	chr6:33989628-34101443
<i>GRM4</i>	chr6:25783126-25832287
<i>SLC17A1</i>	chr6:25912990-25930839
<i>SLC17A2</i>	chr6:89967238-90024967
<i>GABBR2</i>	chr6:89887222-89927496
<i>GABBR1</i>	chr6:114257326-114292354
<i>HDAC2</i>	chr6:24428404-24489850
<i>GPLD1</i>	chr6:121756744-121770872
<i>GJA1</i>	chr6:135605111-135818903
<i>PARK2</i>	chr6:161768590-163148834
<i>NHLRC1</i>	chr6:18120717-18122851
<i>RNF216L</i>	chr7:5013615-5037798
<i>DDC</i>	chr7:50526135-50633154
<i>HTR5A</i>	chr7:154862545-154877459
<i>GRM3</i>	chr7:86273229-86494191
<i>GRM8</i>	chr7:126078651-126892428
<i>CACNA2D1</i>	chr7:81579417-82073031
<i>FOXP2</i>	chr7:114055051-114331092

<i>HOXA1</i>	chr7:27132615-27135625
<i>RHEB</i>	chr7:151163097-151217010
<i>WNT2</i>	chr7:116916687-116963343
<i>NrCAM</i>	chr7:107788083-108096826
<i>RELN</i>	chr7:103112232-103629963
<i>LAMB1</i>	chr7:107564245-107643804
<i>HGF</i>	chr7:81331444-81399452
<i>MET</i>	chr7:116312458-116438439
<i>RAC1</i>	chr7:6414125-6443597
<i>CADPS2</i>	chr7:121958478-122526813
<i>CNTNAP2</i>	chr7:145813452-148118086
<i>UBE2H</i>	chr7:129472994-129592789
<i>EN2</i>	chr7:155250823-155257520
<i>DPP6</i>	chr7:153584418-154685994
<i>STX1A</i>	chr7:73113534-73134017
<i>PIK3CG</i>	chr7:106505923-106547585
<i>SEMA3A</i>	chr7:83587664-83824217
<i>MAFK</i>	chr7:1570367-1582678
<i>ORAI2</i>	chr7:102073995-102097268
<i>NPTX2</i>	chr7:98246596-98259181
<i>VGF</i>	chr7:100805790-100808852
<i>KCND2</i>	chr7:119913721-120390387
<i>SLC18A1</i>	chr8:20002366-20040717
<i>DLGAP2</i>	chr8:1449568-1656640
<i>NRG1</i>	chr8:31497267-32600768
<i>GRIN3A</i>	chr9:119187506-120177317
<i>GRIN1</i>	chr9:104331634-104500862
<i>SLC1A1</i>	chr9:140033608-140063207
<i>TSC1</i>	chr9:4490443-4587469
<i>ANXA1</i>	chr9:135766734-135820020
<i>DNM1</i>	chr9:75766780-75785307
<i>STXBP1</i>	chr9:130965662-131017527
<i>ASTN2</i>	chr9:130374485-130454995
<i>SLC18A2</i>	chr10:119000715-119037095
<i>HTR7</i>	chr10:92500577-92617671
<i>GAD2</i>	chr10:26505235-26593491
<i>NRG3</i>	chr10:83635069-84746935
<i>PTEN</i>	chr10:89623194-89728531
<i>ANXA7</i>	chr10:75135188-75173841
<i>HTR3B</i>	chr11:113775588-113817282
<i>HTR3A</i>	chr11:113845796-113861032
<i>GRM5</i>	chr11:88237744-88796816
<i>GRIA4</i>	chr11:105480799-105852819
<i>SLC1A2</i>	chr11:35272752-35441105
<i>SLC17A6</i>	chr11:22359666-22401044
<i>DGKZ</i>	chr11:46354738-46402104
<i>NRXN2</i>	chr11:64373647-64490660
<i>BDNF</i>	chr11:27676443-27743605
<i>RAB3ILL1</i>	chr11:61664773-61684997
<i>PRRG4</i>	chr11:32851488-32876103
<i>TRPC6</i>	chr11:101322295-101454659
<i>SHANK2</i>	chr11:70313961-70935808
<i>TPH2</i>	chr12:72332625-72426221
<i>GRIN2B</i>	chr12:13714409-14133022
<i>ERBB3</i>	chr12:56473891-56497127
<i>CAPS2</i>	chr12:75669759-75723836
<i>SLC6A13</i>	chr12:329790-372002

SLC6A12	chr12:299249-323371
SLC17A8	chr12:100750856-100815836
SLC38A1	chr12:46576842-46663208
AVPR1A	chr12:63540215-63546590
VAMP1	chr12:6571405-6579843
SYT1	chr12:79257772-79845787
HTR2A	chr13:47407512-47471169
SAP18	chr13:21714652-21723223
NBEA	chr13:35516455-36246872
PCDH9	chr13:66876966-67804468
TRPC4	chr13:38210774-38443939
GCH1	chr14:55308723-55369542
AKT1	chr14:105235688-105262080
GPHN	chr14:66974124-67648523
NRXN3	chr14:78870092-80330758
STXBP6	chr14:25281307-25519095
MDGA2	chr14:47308829-48143988
UBE3A	chr15:25582396-25684128
APBA2	chr15:29213839-29410514
NRG4	chr15:76235844-76304785
GABRG3	chr15:27216428-27778372
GABRA5	chr15:27111865-27194356
GABRB3	chr15:26788694-27018927
SIN3A	chr15:75661721-75743926
SCAMP5	chr15:75287900-75313836
UBR	chr15:43235102-43398286
A2BP1	chr16:29882481-29910580
MAPK3	chr16:6069131-7763340
DOC2A	chr16:30125426-30134630
GRIN2A	chr16:30016836-30022401
CACNA1H	chr16:9847266-10276611
TSC2	chr16:1203240-1271771
MLST8	chr16:2097989-2138712
ZNF778	chr16:2255449-2259417
ANKRD11	chr16:89284110-89295363
SEZ6L2	chr16:89334034-89556969
ORAI3	chr16:30960404-30966258
PRKCB	chr16:23847299-24231930
NF1	chr17:29421944-29704694
CACNA1G	chr17:48638448-48704542
GRIN2C	chr17:72838167-72856007
ITGB3	chr17:45331207-45390076
SLC6A4	chr17:28523379-28562954
PRKCA	chr17:64298925-64806862
PLD2	chr17:4710420-4726727
NLGN2	chr17:7311501-7323179
ERBB2	chr17:37844392-37884914
HOXB1	chr17:46606806-46608272
BZRAP1	chr17:56378595-56406152
AANAT	chr17:74449432-74466198
RPTOR	chr17:78518624-78940173
NSF	chr17:44668037-44834833
VAMP2	chr17:8062464-8066293
AQP4	chr18:24432009-24445716
SYT4	chr18:40847858-40857615
MBD2	chr18:51680574-51751158
MBD1	chr18:47795216-47808144

<i>GRIN2D</i>	chr19:48898131-48948187
<i>GRIN3B</i>	chr19:1000436-1009723
<i>RAB3A</i>	chr19:18307610-18314874
<i>AKT2</i>	chr19:40736224-40791265
<i>SLC1A6</i>	chr19:15060990-15083730
<i>SLC17A7</i>	chr19:49932655-49944808
<i>MBD3</i>	chr19:1576677-1592652
<i>PLAUR</i>	chr19:44150247-44174502
<i>DLGAP4</i>	chr20:34995443-35157038
<i>SLC32A1</i>	chr20:37353104-37358016
<i>SNAP25</i>	chr20:10199476-10288065
<i>GSS</i>	chr20:33516236-33543601
<i>CACNA1I</i>	chr22:39966757-40085738
<i>SHANK3</i>	chr22:51113069-51171641
<i>MAPK1</i>	chr22:22113948-22221970
<i>SYN3</i>	chr22:32908541-33454377
<i>NPTXR</i>	chr22:14700001-25900000
<i>PRODH</i>	chr22:18900287-18924066
<i>ARX</i>	chrX:25021814-25034065
<i>HTR2C</i>	chrX:113818550-114144624
<i>GRIA3</i>	chrX:122318095-122624760
<i>GABRA3</i>	chrX:151335633-151619831
<i>GABRE</i>	chrX:151121596-151143151
<i>GABRQ</i>	chrX:151806636-151821825
<i>MAOA</i>	chrX:43515408-43606067
<i>RbBP7</i>	chrX:16862776-16888534
<i>MeCP2</i>	chrX:153287263-153363188
<i>FMR1</i>	chrX:146993508-147032639
<i>ARHGEF9</i>	chrX:62854847-62975031
<i>NLGN4X</i>	chrX:5808083-6146706
<i>NLGN3</i>	chrX:70364680-70391049
<i>SLC9A6</i>	chrX:135067585-135129426
<i>SYN1</i>	chrX:47431300-47479256
<i>SYP</i>	chrX:49044264-49056661
<i>IL1RAPL1</i>	chrX:28605680-29974017
<i>PAK3</i>	chrX:110187512-110464173
<i>GDI1</i>	chrX:153665258-153671812
<i>KDM5C</i>	chrX:53220503-53254604
<i>CDKL5</i>	chrX:18443724-18671747
<i>TRPC5</i>	chrX:111017541-111326004
<i>PTCHD1</i>	chrX:23352984-23414917
<i>HUWE1</i>	chrX:53559071-53713673
<i>CUL4B</i>	chrX:119658447-119709684
<i>BRWD3</i>	chrX:79931682-80065233
<i>UBE2A</i>	chrX:118708498-118718377

Tabela 2. 2 Sequencia dos oligonucleotídeos desenhados para os experimentos de PCR e PCR quantitativo.

Gene	Sequencias
<i>hGAPDH</i>	5' AACAGCGACACCCACTCCTC3'
<i>hGAPDH</i>	5' CATAACCAGGAAATGAGCTTGACAA3'
<i>G6PDH F</i>	5' TCTTCATCACCAAGAGAACTTGC3'
<i>G6PDH R</i>	5' GACCTGGAAGTCACTGGGCA3'
<i>GRM5 F</i>	5' ACCTGTTATCACCATAGTGG3'
<i>GRM5 R</i>	5' CATTAAAATGCATAAGGCTAAGG3'
<i>SLC17A6 F</i>	5' GAAACAAACCTGGGCAGACC3'
<i>SLC17A6 R</i>	5' TCCAACCCTAGGCCAAC3'
<i>PRKCα F</i>	5' TGTTGAGGAAGAGACGGAGG3'
<i>PRKCα R</i>	5' TGGTGTGACGGTTGAATGGC3'
<i>MBD2 F</i>	5' AAGCAGAGGCGTGATAAGC3'
<i>MBD2 R</i>	5' TGCTGACCTGTTGCCTGC3'
<i>GCH1 F</i>	5' AGGCAACAGCTTCCAGGATTAA3'
<i>GCH1 R</i>	5' CCGGTTCCCTTGTTGGTCTATG3'
<i>SEMA3A F</i>	5' GAATCAAAGACAGCTAGAG3'
<i>SEMA3A R</i>	5' GCACTGCACACTCACCTATTG3'
<i>SNAP29 F</i>	5' CTGGTTTAGGACTGGCACTCT3'
<i>SNAP29 R</i>	5' ATAGGCTGCAGACTGTAAAGGG3'
<i>MARK1 F</i>	5' CATCTGCACACTCCTGAAAA 3'
<i>MARK1 F</i>	5' TCACCCCTAGTCTGGAAGCC 3'
<i>GAD2 F</i>	5' ATGATTGGGGCTTCTCGT 3'
<i>GAD2 R</i>	5' ACAGCTAAGAGGGGGTCAAAT 3'
<i>SHANK2 F</i>	5' CCTATAGCAGCGGTACAGAAC3'
<i>SHANK2 R</i>	5' TGCCTTGTTCCATTCTTATTG3'
<i>SNAP29 F PCR</i>	5' CCACGTGCTCCATCTGTAACCTC3'
<i>SNAP29 R PCR</i>	5' GGTGATGCATTATGGGATGGGA3'

Tabela 2.3 Sequencia das sondas desenhadas para os experimentos de MLPA

GENE	EXONS	SONDAS
<i>FOWARD</i>		
<i>SLC17A6</i>		
1	<i>Exon2</i>	GGGTTCCCTAACGGGTTGGATTATGCCATCATGAGCGGCCTGGCTTC
2	<i>Exon3</i>	GGGTTCCCTAACGGGTTGGAGATCCACGGTCTCTTTGGGCTACATC
3	<i>Exon4</i>	GGGTTCCCTAACGGGTTGGATCAGGTTTCGGAGCTGCCATACTCTTACCTCTACCTAA
4	<i>Exon5</i>	GGGTTCCCTAACGGGTTGGATAGAGAGGAGTAGACTGGCAACCACCTCTTT
5	<i>Exon6</i>	GGGTTCCCTAACGGGTTGGAGGTTCTATGCCGGAGCTGTGATTGCAATGCCCTT
6	<i>Exon7</i>	GGGTTCCCTAACGGGTTGGAGTGCTTGAGGAAGCTTGAATGGCTGGTACATGTT
7	<i>Exon8</i>	GGGTTCCCTAACGGGTTGGATCCATGGATCAATCCATGGAGGAAGTTTACATCCATGCCAGCTAT
8	<i>Exon9</i>	GGGTTCCCTAACGGGTTGGAtcaaCTGTTTCATCAGGTTGGATGCTATGCTGTGC
9	<i>Exon10</i>	GGGTTCCCTAACGGGTTGGACTGCTCTGGCTATTCTCATACTAGAGGGTAGCAATCTCAT
10	<i>Exon11</i>	GGGTTCCCTAACGGGTTGGAGCCAGTATCTAATGGCATTGCAATGGTGGCA
11	<i>Exon12</i>	GGGTTCCCTAACGGGTTGGACCAATGGAGGTTGGCTAGTGGTGGAAAAGAAAGAGGAATTGTA
<i>REVERSE</i>		
<i>SLC17A6</i>		
1	<i>Exon2</i>	TGCATCTCCTCGGTATCCGCTGCAACCTCTAGATTGGATCTGCTGGCAC
2	<i>Exon3</i>	ATCACTCAGATTCCGGGAGGCTACATCGCGTCTAGATTGGATCTGCTGGCAC
3	<i>Exon4</i>	TATGCTAATTCCATCAGCAGCCAGAGTCGATTGGATGCTCTAGATTGGATCTGCTGGCAC
4	<i>Exon5</i>	GTGGTGAAGTACTGTGCGAGATGCAGCTATGGTCTAGATTGGATCTGCTGGCAC
5	<i>Exon6</i>	AGCTGGCATTCTGTGCACTACACTGGCTGGCTTCTAGATTGGATCTGCTGGCAC
6	<i>Exon7</i>	TTGGCTTTGGTGTCTATGAAAGTCCTGCAAAGCATCCTAGATTGGATCTGCTGGCAC
7	<i>Exon8</i>	GCAATAATTGTTGCAAACCTCTGCAAGCTGGACTAAAGTCTAGATTGGATCTGCTGGCAC
8	<i>Exon9</i>	CACACTTAGTAATGACAATTATTGTCCTATTGGGGaaagTCTAGATTGGATCTGCTGGCAC
9	<i>Exon10</i>	TCTTGGTACTTGCAGTGGATTCACTGGATTTGCTATATCTGTAAGATCTAGATTGGATCTGCTGGCAC
10	<i>Exon11</i>	CATTGTCAGGAATGGTTGTCCTATCATTGTTGGTGTCTAGATTGGATCTGCTGGCAC
11	<i>Exon12</i>	CAAGGAGAACTACAAGACTCACATAGCTATAAGGACCGAGTGATTCTAGATTGGATCTGCTGGCAC
<i>FOWARD</i>		
<i>MBD2</i>		
1	<i>Exon1</i>	GGGTTCCCTAACGGGTTGGATGATCCGAAAATCTGGCTAAGTGGCAA
2	<i>Exon2</i>	GGGTTCCCTAACGGGTTGGATCTCTAGTCCAAGTGGTAAGAAGTCAGAAGCAAGCCT
3	<i>Intron 4</i>	GGGTTCCCTAACGGGTTGGATGGATGACGTACCTCTGTGTTAGTAAAGCCAG
4	<i>Exon 5</i>	GGGTTCCCTAACGGGTTGGATACTTGTCTCATGGCAGGAGTTGGTCCAGGT
5	<i>Exon 6</i>	GGGTTCCCTAACGGGTTGGACAGGAAGAGCAGTACAGCAAGTACGCAAGAAATTGGAAGAAGCACTGATGGCAGACATCT
6	<i>Exon 7</i>	GGGTTCCCTAACGGGTTGGACTCGTAGCCTCTGATTACATTCAACTCTGATCCCTGGCCTTAGGTTGACATGGAG
<i>REVERSE</i>		
<i>MBD2</i>		
1	<i>Exon1</i>	GAGCGATGTCCTACTACTTCAGGTACCTCCCTCTAGATTGGATCTGCTGGCAC

2	<i>Exon 2</i>	CAGTTGGCAAGGTACCTGGAAATACTGTTGATCTCAGC TCTAGATTGGATCTGCTGGCAC
3	<i>Intron 4</i>	GAAGAAGGTACATGAGAAAATTGTATGTTGCAGGG TCTAGATTGGATCTGCTGGCAC
4	<i>Exon 5</i>	AGCAATGATGAGACCCTTTATCTGCTGTTGCC TCTAGATTGGATCTGCTGGCAC
5	<i>Exon 6</i>	TGTCGCGAGCTGCTGATAACAGAAGAGATGGATATTGAAATGGACAGTGGAGATGAAGCCTA TCTAGATTGGATCTGCTGGCAC
6	<i>Exon 7</i>	GTGGAGGAAGATAGCGCATATATTGCAGTATGAACATTGCCTCTGGACGTTGTGAGAATTGT TCTAGATTGGATCTGCTGGCAC

Capítulo 3

Use of a custom-designed array-CGH to identify small CNVs in nonsyndromic Autism Spectrum Disorders

Tese Doutorado

Estudo de genes candidatos aos transtornos do espectro autista

Abstract

The etiology of Autism Spectrum Disorders (ASD) is very heterogeneous and, despite the enormous efforts to identify their underlying genetic causes, about 70% of the cases are idiopathic. We have hypothesized that a proportion of these cases might be caused by small CNVs (<10 Kpb), usually not detected through conventional microarray platforms or next generation sequencing. To address this hypothesis, we designed an array CGH with high density of probes targeted to exons and 5' and 3' untranslated (UTR) regions of 269 genes involved in biological processes potentially significant to the pathophysiology of ASD. We detected CNVs in 11.6% of the 103 analyzed patients with nonsyndromic idiopathic ASD, whom were previously excluded to harbor any of the recurrent CNVs at 15q11.13, 16p11 and 22q13. Most of the potential pathogenic CNVs identified (70%) are smaller than 10 Kbp, suggesting a role of these alterations in ASD etiology and the importance of a molecular tool able to detect them. This study suggests the involvement of glutamate and axonal guidance related functions and disruption of non-coding DNA in ASD etiology. Finally, our findings corroborate the association of *SHANK2*, *DIAPH3*, *GCH1*, *GRM5* and *MARK1* with nonsyndromic ASD and suggest for the first time *SNAP29*, *SLC17A6*, *PRKCa*, *MBD2* and *GAD2* as new candidates to these disorders.

Resumo

A etiologia dos transtornos do espectro autista (TEA) é muito heterogênea e, apesar dos enormes esforços para identificar os fatores genéticos associados a esses transtornos, cerca de 70% dos casos são idiopáticos. Nós levantamos a hipótese de que uma parte desses casos pode ser causada por pequenas CNVs (<10 Kpb), normalmente não detectadas através de plataformas de microarray convencionais ou sequenciamento de última geração. Para testar esta hipótese, nós customizamos uma lâmina de array-CGH com alta densidade de sondas complementares aos *exons* e regiões UTRs de 269 genes envolvidos em processos biológicos potencialmente significativos para a fisiopatologia dos TEA. Com esta plataforma detectamos CNVs em 11,6% dos 103 portadores de TEA não sindrômicos analisados, os quais foram anteriormente analisados para excluir a presença das alterações recorrentes nas regiões 15q11.13, 16p11 e 22q13. A maioria (70%) das CNVs possivelmente patogênicas identificadas são menores do que 10 kpb, sugerindo que essas podem ser relevantes na etiologia TEA e mostrando a importância de uma ferramenta molecular capaz de detectá-las. Este estudo sugere que funções relacionadas com sinapses glutamatérgicas e orientação axonal, assim como, alterações em regiões UTR podem estar envolvidas na etiologia dos TEA. Finalmente, nossos resultados corroboram com a associação dos genes *SHANK2*, *DIAPH3*, *GCH1*, *GRM5* e *MARK1* com TEA não sindrômico e sugere pela primeira vez os genes *SNAP29*, *SLC17A6*, *PRKCa*, *MBD2* e *GAD2* como candidatos a esses transtornos.

Autism Spectrum Disorders (ASD) are neuropsychiatric conditions characterized by restricted, repetitive and stereotyped patterns of behavior, interests, and activities, qualitative impairment in social interaction and communication deficits (AMERICAN PSYCHIATRIC ASSOCIATION, 2000). The ASD prevalence is about 1%, with an affected male-to-affected female ratio of 4:1 (FOMBONNE, 2009). Several studies have provided evidence for a strong genetic component in ASD etiology (BAILEY *et al.*, 1995; SZATMARI *et al.*, 1996; ZWAIGENBAUM, 2005; BETANCUR, 2011), but the underlying genetic cause is still unknown in at least 70% of the cases (SCHAAF; ZOGHBI, 2011). High-resolution genomic methodologies, as array comparative genomic hybridization (CGH) and next generation sequencing (NGS), have demonstrated significant roles for rare copy number variations (CNVs) and point mutations, which account for about 10-20% of the ASD cases (SEBAT *et al.*, 2007; MARSHALL *et al.*, 2008; MILES *et al.*, 2010; PINTO *et al.*, 2010; IOSSIFOV *et al.*, 2012; NEALE *et al.*, 2012; O'ROAK *et al.*, 2012; SANDERS *et al.*, 2012). Except for alterations involving the regions 1q21.1, 7q11.23, 15q11-q13, 1611.2 and 22q13, and the genes *SHANK2*, *NRXN1*, *PTCHD1*, *DYRK1A*, *POGZ* and *CDH8*, the great majority of CNVs or point mutations are private, without recurrence among autistic patients (JACQUEMONT *et al.*, 2006; MARSHALL *et al.*, 2008; WEISS *et al.*, 2008; KUMAR *et al.*, 2009; SANDERS *et al.*, 2011; IOSSIFOV *et al.*, 2012; O'ROAK *et al.*, 2012; SANDERS *et al.*, 2012; GIRIRAJAN *et al.*, 2013). Most genes disrupted by these alterations share common biological functions or converge on a few molecular pathways, suggesting that damage in different genes may cause similar effects during neural development and neuronal function and ultimately lead to the autistic phenotype (GESCHWIND, 2008; CUSCÓ *et al.*, 2009; GLESSNER *et al.*, 2009; BUCAN *et al.*, 2009; SBACCHI *et al.*, 2010).

Typically large CNVs (>400 kbp) have been associated with the ASD etiology. It is possible that small CNVs, still a challenge to be detected through conventional microarray platforms or

NGS (GIRIRAJAN *et al.*, 2013), can contribute significantly to these disorders. In order to improve the detection of small potentially pathogenic CNVs (<10 kbp), we designed an array CGH with high density of probes targeted to exons and 5' and 3' untranslated (UTR) regions of 269 genes involved in at least 25 canonical pathways, as Glutamate Receptor Signaling, GABA Receptor Signaling, Serotonin Receptor Signaling, G-Protein Coupled Receptor Signaling, Synaptic Long Term Potentiation, Synaptic Long Term Depression, mTOR Signaling, among other potentially significant to the pathophysiology of ASD. Identification of small CNVs, in contrast to the large ones, will allow the identification of the specific target genes that contribute to the ASD phenotype.

Materials and methods

Subjects

A total of 700 Brazilian patients (519 males and 181 females, ranging in age from 3 to 46 years) diagnosed with ASD were included in the present study. Among them, four have an affected sibling. While 510 patients were referred to the *Centro do Estudos do Genoma Humano*, São Paulo (most ascertained at the *Instituto de Psiquiatria of the Hospital das Clínicas*, University of São Paulo), 190 were ascertained at the *Hospital de Base*, Faculty of Medicine of São José do Rio Preto, São Paulo.

All patients were diagnosed by psychiatrists using DSM-IV (Diagnostic and Statistical Manual of Mental Disorder), ICD-10 (International Statistical Classification of Diseases and Related Health Problems) and ASQ (Autism Screening Questionnaire) (BERUMENT *et al.*, 1999; ORABONA *et al.*, 2009; GRIESI-OLIVEIRA *et al.*, 2012). Children's behavior was

characterized by score questions of CARS (Childhood Autism Rating Scale) (SCHOPLER *et al.*, 1980).

Patients with autism-related syndromes, metabolic disorders, facial dysmorphisms, congenital malformations, or those whose mothers were exposed to teratogenic drugs and teratogenic infectious agents during pregnancy were not included in the study. All patients were negative for CNVs at 15q11.13, 16p11 and 22q13 through multiplex ligation probe amplification (MLPA) analysis (SALSA MLPA KIT P343-B1 AUTISM-1; MRC-Holland, Amsterdam, The Netherlands).

DNA samples from 745 apparently healthy individuals were used as controls.

The Ethics Committee of Bioscience Institute – University of São Paulo approved this research. Patients and controls were included in this study only after a signed written informed consent by the families or other legal caregivers.

Customization of CGH array

Custom-designed oligonucleotide array-CGH (8x60K – Agilent Technologies) was performed with eArray web software using the filters: similarity score, probe score above 0.8 and probe \geq 59pb, in order to select the best-performing oligonucleotides from Agilent's library (<https://earray.chem.agilent.com/earray/>). This array contains probes targeted to all exons and 5' and 3' UTRs regions of 269 selected genes based on the following criteria: genes previously reported as candidates for ASD and/or other neuropsychiatric disorders in at least two different studies; genes involved in the same pathways performing functions similar to the ones performed by candidate genes and genes encoding proteins with neurological functions located in candidate genomic regions (Table 2.1 – chapter 2). For each exon we designed three to six overlapping probes. This design should provide increased sensitivity for detecting small deletions and duplications. The genomic coordinates are based on the February 2009 assembly of the human

reference genome (GRCh37/hg19).

Investigation of copy number alterations by custom-designed Comparative Genomic Hybridization on Microarrays

In the first step, the screening was performed by custom-designed oligonucleotide array CGH. The procedures for DNA digestion, labeling with Cy3- and Cy5-dCTPs by random priming followed by purification, hybridization and washing were performed according to the manufacturer's protocol (Agilent Technologies). Scanned images of the arrays were processed using the Feature Extraction software (Agilent Technologies, Santa Clara, CA), and the analysis was carried out using the Agilent genomic Workbench 7.0 (Agilent Technologies, Santa Clara, CA). For array CGH analysis, identification of aberrant copy number segments was based on the ADM-2 algorithm with default settings (threshold log₂ ratio of 0.2 or 1.14 was used for gain or high copy gain, and -0.23 and -1.14 used for loss and homozygous loss, respectively), and the significance threshold was set on 7. We considered at least three consecutive probes for calling a segment. All ASD samples were tested in duplicate. Only the autism specific CNVs [not observed in 200 samples from healthy Brazilian individuals and in the Database of Genomic Variants (DGV) (<http://projects.tcag.ca/variation>)] were considered possibly pathogenic and, in the second step, were confirmed using another array CGH platform (Agilent SurePrint G3 Human CGH Microarrays 4×180K [180,000 oligonucleotides probes]), and/or a SNP-array platform (GeneChip Human Mapping 500K Array Set Affymetrix), and/ or quantitative real- time PCR.

Quantitative real-time PCR

Quantitative real- time PCR (qPCR) was performed using the ABI Prism 7900HT Fast Real-

time Sequence Detection System (Applied Biosystems, Foster City, CA). Reactions were carried out in triplicates using SYBR Green PCR MasterMix (Applied Biosystems) in a total volume of 20 μ L. Dissociation curves were analyzed for each primer pair to verify the specificity of the qPCR reaction. The copy number quantification was calculated as described by Ponchel *et al.* (2003) and ΔK_{Ct} (Fold copy number change) values of 0 ± 0.35 indicate an equal ratio of the target and reference, which corresponds to no genetic abnormality, values of -1 ± 0.35 indicate loss of one copy (microdeletion), whilst values of 1 ± 0.35 indicate gain of one copy (microduplication). *G6PDH* (WEKSBERG *et al.* 2005) and *hGAPDH* (PONCHEL *et al.*, 2003) were used as endogenous control genes. The primers are shown in Table 2.2 (chapter 2).

Single Nucleotide Polymorphism Array

The Affymetrix GeneChip Human Mapping 500K Array Set is comprised of two arrays, each capable of genotyping on average 500,000 SNPs (approximately 262,000 for *Nsp* arrays and 238,000 for *Sty* arrays). The assay was performed according to the standard protocol. The quality of hybridization samples was evaluated by QC Call Rate; the threshold chosen for this was 95.0. For the analysis we used Genotyping Console software, Affymetrix (Santa Clara, CA, USA), PennCNV (WANG *et al.*, 2007; WANG; BUCAN, 2008) and dChip (LI; HONG-WONG 2001; LIN *et al.* 2004) with default settings. We considered only CNVs detected by at least two different softwares.

In silico analysis of altered UTR regions

We analyzed the altered UTR segment of CNVs regarding the existence of potential binding sites for transcription factors by Consite (*TF score cutoff* 98%; SANDELIN *et al.*, 2004), miRNA target sites predicted by TargetScanHuman 5.1 (LEWIS *et al.*, 2005; GRIMSON *et al.*, 2007;

FRIEDMAN *et al.*, 2009), elements highly conserved across 46 vertebrates species by phastCons (MURPHY *et al.*, 2001; CHIAROMONTE *et al.*, 2002; KENT *et al.*, 2003; SCHWARTZ *et al.*, 2003; BLANCHETTE *et al.*, 2004, SIEPEL *et al.*, 2005; POLLARD *et al.*, 2009), and mRNA polyadenylation site in the 3'UTR region by PolyA_DB (exon.umdnj.edu/polya_db/; ZHANG *et al.*, 2005; CHENG *et al.*, 2006).

***MBD2* and *SLC17A6* CNVs screening**

SLC17A6 and *MBD2* were screened for CNVs by the MLPA technique (Multiplex Ligation dependent Probe Amplification) with probes designed using the Human MLPA Probe Design algorithm (H-MAPD; <http://genomics01.arcan.stonybrokk.edu/mlpa/cgi-bin/mlpa.cgi>; ZHI; HATCHWELL, 2008) and synthesized by IDT Technologies. For *MBD2* we designed probes to exons 1, 2, 5, 6, 7 and to intron 4 (Transcript ID - ENST00000256429; <http://www.ensembl.org>). As for *SLC17A6* (Transcript ID - ENST00000263160; <http://www.ensembl.org>), it was possible to design probes for all exons but the first one. MLPA reactions were performed according to the manufacturer's instructions (MRC-Holland) and as previously described (AHN *et al.*, 2007). The reference commercial kit P200 was used in combination with custom probes for data normalization and quantity/quality control fragments. The amplified products were separated in an ABI-3730 DNA Analyzer (Applied Biosystem). Results analyses were performed with the help of GeneMapper (version 4.0) and Genemarker (version 1.97) softwares. Ratios below 0.75 and above 1.30 were considered as deletions and duplications, respectively. The probes are shown in Table 2.3 (chapter 2).

Screening of the deletion at the 3'UTR of *SNAP29*

The screening of the deletion at the 3'UTR of *SNAP29* was performed by PCR followed by

electrophoresis on 1.5% agarose gels. Fragments of 770 and 528 bp represent the normal (wild-type) and the deleted alleles, respectively. The primers are shown in Table 2.2 (chapter 2).

Results

Initially, we analyzed 103 non-syndromic ASD cases (99 single and 4 familial) and 200 matched controls by the custom-designed array-CGH. We detected a total of 121 alterations in 62% (64/103) of the patients and 82 alterations in 33% (66/200) of the controls (Qui-square test, $p < 0.0001$).

We considered for further analyses only the autism specific CNVs, i.e., not observed in 200 Brazilian controls initially analyzed and not similar to the others described in Database Genomic Variants (DGV - IAFRATE *et al.*, 2004; ZHANG *et al.*, 2006), supposing that these possibly have higher chance to be ASD causative. Among the single cases, we identified nine different autism specific CNVs in 11 patients, ranging from 242 bp to 376 Kbp (Table 3.1). As to the four familial cases, we found one autism specific CNV in DNA sample from patient 12. Her affected brother also carries this alteration (Table 3.1). Therefore, in total, we detected 10 CNVs possibly pathogenic in 12 probands (11.6%), being two recurrent – one at *SNAP29* [3/103 (2.9%)] and another at *MARK1* [2/103 (1.9%)]. Except for patient 5, who had two CNVs, all the others had only one. All autism specific CNVs were validated by other methodology; hence, we did not observe any false positive results among these 12 patients.

In 8 (cases 1 to 6, 8 and 9) out of 11 ASD cases in which both parents or other informative family members were available for genetic testing, the CNVs were found to be *de novo* (Table 3.1). The rate of *de novo* alterations in the 269 genes included in our custom-designed array among Brazilian autistic patients was estimated as 7.84% (8/102; for the estimate, we excluded

one case that was not possible to determine the CNV origin), while the rate of inherited alterations was 2.94% (3/102) (Table 3.1).

Five out of the 10 CNVs are located at coding regions of five different genes (*DIAPH3*, *GAD2*, *GRM5*, *MBD2* and *SLC17A6*). It was possible to map the alterations breakpoints in *DIAPH3* and *GRM5*: while the deletion at *DIAPH3* encompasses the 5'UTR up to the 3' end of exon 22 (NM_001042517.1), possibly leading to the phenotype by haploinsufficiency, the deletion of exons 3 and 4 of *GRM5* (NM_001143831.2) is in frame and causes the change of the first codon of exon 5 of AGT (Ser) to GGT (Gli). The *in silico* analysis by RaptorX (PENG; XU, 2011) revealed that this deletion alters substantially the tridimensional structure of the protein, possibly leading to loss of its function. The duplications at *MBD2* (NM_003927.4) and *SLC17A6* (NM_020346.2) encompass most exons of these genes and the duplication at *GAD2* (NM_000818.2) encompasses three exons. Therefore, it is possible that these CNVs also lead to the formation of a protein with altered function.

The other five CNVs identified are located at UTR regions of five genes (*SNAP29*, *MARK1*, *PRKCa*, *SHANK2* and *GCH1*) (Table 3.1). *In silico* analysis of the deleted or duplicated segments showed that all of them contain potential binding sites for transcription factors, miRNA target sites, elements highly conserved across 46 vertebrates species, and within the *PRKCa* duplicated segment we also found a mRNA polyadenylation site.

No CNVs at *SLC17A6*, *MBD2* and *SNAP29* were previously found in autistic patients, thereby we considered important to have a better estimate of the prevalence of the alterations at these genes among Brazilian ASD patients as well as to investigate a large number of controls. Therefore, we performed a CNV screening at *SLC17A6* and *MBD2* by MLPA in further 405 and 350 ASD patients, respectively, and 375 controls. We only found recurrence of CNVs at *SLC17A6* in 4 affected and in one control, which involved a duplication of exons 11 and 12 (patient 145, 386), a deletion of exon 11 (patient 182) and duplication of exons 9, 10, 11 and 12 (patient 294 and control). In all patients, the CNVs were *de novo*. Thus, the prevalence of CNVs at *SLC17A6* among ASD patients (0.98% - 5/103+405) is significantly higher than the prevalence in controls (0.17% - 1/375+200) (Qui-square test, p = 0.0365). The estimated prevalence of

CNVs at *MBD2* among ASD patients is 0.2% (1/103+350).

Concerning the deletion at the 3'UTR of *SNAP29*, it was detected in other three out of 597, patients and in other one out of 545 controls analyzed by PCR. Therefore, the prevalence of this alteration among Brazilian ASD patients (0.8571%; 6/103+597) is significantly higher than the prevalence among controls (0.1342%; 1/545+200) (Qui-square test, p = 0.0240).

Table 3.1 Possibly pathogenic CNVs identified in ASD Patients

Case	Sex	Phenotype	Sporadic or Familial	CNV position	Origin	Type	Genes	Size	Confirmed by
1	Male	Autistic disorder	Sporadic	22q11.21; chr22: 21243170- 21243412;	<i>de novo</i>	Deletion	<i>SNAP29</i>	242 bp	1
				3'UTR					
2	Male	Autistic disorder	Sporadic	22q11.21; chr22: 21243170- 21243412	<i>de novo</i>	Deletion	<i>SNAP29</i>	242 bp	1
				3'UTR					
3	Male	Autistic disorder	Sporadic	22q11.21 chr22: 21243170- 21243412	<i>de novo</i>	Deletion	<i>SNAP29</i>	242 bp	1
				3'UTR					
4	Male	Autistic disorder	Sporadic	1q41 chr1: 220836953- 220837254	<i>de novo</i>	Duplication	<i>MARK1</i>	301 bp	1
				3'UTR					

				17q24.2	<i>de novo</i>	Duplication	<i>PRKCa</i>	1.3 Kbp	1
				chr17: 64806141- 64807441					
				3'UTR					
5	Male	Autistic disorder	Sporadic	1q41	<i>de novo</i>	Duplication	<i>MARK1</i>	301 bp	1
				chr1: 220836953- 220837254					
				3'UTR					
6	Male	Autistic disorder	Sporadic	11q13.3	<i>de novo</i>	Duplication	<i>SHANK2</i>	398 bp	1
				chr11: 70314471- 70314869					
				3'UTR					
7	Male	Autistic disorder	Sporadic	13q21.1	Paternal	Deletion	<i>DIAPH3</i>	376 Kbp	2
				chr13: 60373628- 60750097					
				Coding region					
8	Female	Autistic disorder	Sporadic	10p12.1	<i>de novo</i>	Duplication	<i>GAD2</i>	8.4 Kbp	1
				chr10: 26581425- 26589903					
				Coding region					
9	Male	Autistic disorder	Sporadic	11q14.3	<i>de novo</i>	Deletion	<i>GRM5</i>	315 Kbp	1 and 3
				chr11: 88360595- 88675861					
				Coding region					
10	Male	Autistic disorder	Sporadic	18q21.2	Maternal	Duplication	<i>MBD2</i>	34.8 Kbp	1 and 2
				chr18: 51680552- 51715344					
				Coding region					

11	Male	Autistic disorder	Sporadic	11p14.3 chr11: 22391729-22401068 Coding region and 3'UTR	Not maternal father's DNA not available	Duplication	<i>SLC17A6</i>	9.3 Kbp	1 and 4
12	Female	Autistic disorder; <u>Epilepsy</u>	Familial (Affected sibling)	14q22.2 chr14: 55309413-55309822 3'UTR	Paternal	Duplication	<i>GCH1</i>	409 bp	1
Sibling patient 12	Male	Autistic disorder; <u>Epilepsy</u>	Familial	14q22.2 chr14: 55309413-55309822 3'UTR	Paternal	Duplication	<i>GCH1</i>	409 bp	1
*145	Male	Autistic disorder	Sporadic	11q14.3 Coding region and 3'UTR	<i>de novo</i>	Duplication	<i>SLC17A6</i>	exons 11 and 12	4
*386	Female	Autistic disorder	Sporadic	11q14.3 Coding region	<i>de novo</i>	Duplication	<i>SLC17A6</i>	exons 11 and 12	4
*182	Male	Asperger syndrome	Sporadic	11q14.3 Coding region	<i>de novo</i>	Deletion	<i>SLC17A6</i>	exon 11	4
*294	Male	Autistic disorder	Sporadic	11q14.3 Coding region and 3'UTR	<i>de novo</i>	Duplication	<i>SLC17A6</i>	exons 9, 10, 11 and 12	4
#234	Male	Autistic disorder	Sporadic	22q11.21 chr22: 21243170-21243412 3'UTR	<i>de novo</i>	Deletion	<i>SNAP29</i>	242 bp	5

#335	Male	Autistic disorder	Sporadic	22q11.21 chr22: 21243170- 21243412	<i>de novo</i>	Deletion	<i>SNAP29</i>	242 bp	5
#436	Male	Autistic disorder	Sporadic	22q11.21 chr22: 21243170- 21243412	<i>de novo</i>	Deletion	<i>SNAP29</i>	242 bp	5
				3'UTR					
				3'UTR					

*: Alteration detected by MLPA technique; #: Alteration detected by PCR: 1 - Quantitative real time PCR; 2- SNP-array (Affymetrix GeneChip Human Mapping 500K); 3 – Commercial array CGH (Agilent SurePrint G3 Human CGH Microarrays 4×180K); 4 - Customized MLPA (probes *in-house*); 5 – PCR.

Discussion

We observed a significantly higher frequency of patients with at least one CNVs at the genes analyzed by the custom-designed array as compared to their frequency in controls (Qui-square test, $p < 0.0001$). These results are different from those obtained through a conventional screening with a genome-wide array platform (MARSHALL *et al.*, 2008), suggesting that the use of our custom-designed array can increase the chance of detecting a type of alteration enriched in ASD patients.

Regarding the possibly pathogenic CNVs, we found 10 different alterations ranging from 242 bp to 376 Kbp, showing the efficiency of the custom-designed platform for the identification of both small and large alterations. Through an *in silico* analysis (UCSC Genome Browser – SNP/CNV) we observed that 80% of these CNVs would not have been detected by a conventional array with the same number of probes (60,000) as our customized platform and 60%

would not have been detected by conventional array with about 1,000,000 of probes.

The detection rate of CNVs of 11.6% among the 103 Brazilian nonsyndromic ASD cases is within the upper limit of the range reported by others (8-10% - BREMER *et al.*, 2010; SORTE *et al.*, 2013). However, it is very difficult to make such a comparison, taking into account the limited number of genes analyzed in our customized array-CGH and the inclusion criteria of the patients, which is also different among the several studies. Despite these differences, it is of note that the rate of *de novo* alterations (7.8%) is within the 5.6% to 10% range previously reported (SEBAT *et al.*, 2007; MARSHALL *et al.*, 2008; PINTO *et al.*, 2010).

We found an equal proportion of CNVs targeting coding and non-coding regions. The *in silico* analysis of the alterations in *DIAPH3* and *GRM5* showed that the CNVs at the coding regions are possibly causing the phenotype by loss or gain of one functional allele. As to the CNVs located at UTR regions, it is possible that they lead to a breakdown in the regulatory machinery, possibly also interfering with the quantity of the encoded proteins, as previously suggested (SHIBAYAMA *et al.*, 2004, BONORA *et al.*, 2005; VOINESKOS *et al.*, 2011). Functional studies will be necessary to validate these hypotheses. The identification of 50% of the CNVs located in 5' or 3'UTR suggest that mutations in non-coding regions might explain a significant proportion of ASD cases negative for most genomic screenings, which still do not explore adequately these regions.

Analyzing the set of altered genes, we observed that four of them encode proteins involved in glutamate pathways (*GRM5*, *SLC17A6*, *GAD2*) or glutamatergic synapses (*SNAP29*). mGluR5 (*GRM5*) and vGlut2 (*SLC17A6*) participate in the Glutamate Receptor Signaling pathway acting, respectively, in regulation of intracellular signal transduction and in glutamate transport into the presynaptic vesicle (Ingenuity® Systems, www.ingenuity.com; XU; ZHU *et al.*, 2009; SIMONYI *et al.*, 2010). GAD2 catalyzes the production of gamma-aminobutyric acid from L-glutamic acid in the Glutamate Degradation III pathway (Ingenuity® Systems, www.ingenuity.com) and, finally, SNAP29 is a ubiquitous synaptosomal-associated SNARE protein that regulates the release of glutamate in synaptic cleft by slowing recycling of the SNARE-based fusion machinery and synaptic vesicle turnover (SU *et al.*, 2001, PAN *et al.*, 2005, SUDHOF *et al.*, 2007). Among them, mGLUR5 and vGlut2 dysfunction have been previously

associated with ASD or with other neurological disorders, such as fragile-X, schizophrenia and ADHD (DEVON *et al.*, 2001; ELIA *et al.*, 2010; VOINEAGU *et al.*, 2011; SKAFIDAS *et al.*, 2012).

We also detected CNVs at *DIAPH3*, *SHANK2* and *PRKCa*, whose encoded proteins are involved in axonal guidance functions. *SHANK2* and *PRKCa* act in Axonal Guidance Signaling while *DIAPH3* act in Actin Cytoskeleton Signaling canonical pathways (*Ingenuity® Systems*, www.ingenuity.com, GUPTON *et al.*, 2007; DENT *et al.*, 2007). Alterations in *PRKCa* are causative of a syndromic ASD form (Timothy syndrome), while those at *DIAPH3* and *SHANK2* have been associated to nonsyndromic forms of ASD (BERKEL *et al.*, 2010; NAVEDO *et al.*, 2010; VORSTMAN *et al.*, 2011). Therefore, 7 out of 10 CNVs here identified might damage the glutamate synapses or axonal guidance related functions, thus supporting that dysfunction in different genes of a common biological pathway or physiological process might lead to the ASD phenotype.

The three remaining CNVs involve *MBD2*, *GCH1* and *MARK1*, which are part of different pathways. *MBD2* belongs to the same pathway of *MECP2*, the gene that when mutated leads to syndromic autism (Rett syndrome) (AMIR *et al.*, 1999). On the other hand, *GCH1* encodes an essential enzyme for *de novo* biosynthesis of the tetrahydrobiopterin, a cofactor required for the biosynthesis of folate, serotonin, dopamine and catecholamines. Overexpression and mutations in *GCH1* have been associated with comorbidities observed in autistic patients, such as mood disorders, pain sensitivity and dystonia (SEROVA *et al.*, 1999; TEGEDEER *et al.*, 2006; TEGEDEER *et al.*, 2008; CAMPBELL *et al.*, 2009; CAO *et al.*, 2010). Finally, *MARK1* encodes a protein that acts as a positive regulator of the *Wnt Signaling* pathway and in the regulation of neuronal migration (DREWES *et al.*, 1997). It is of note that patient 4, in addition to the *MARK1* duplication, has a duplication at *PRKCa*; thus, it is possible that both alterations contribute to the phenotype, in accordance to the multi-hit model previously suggested by others (O'ROAK *et al.*, 2011; LEBLOND *et al.*, 2012).

Despite the possibility that the higher frequency of CNVs at *SLC17A6* and *SNAP29* observed in patients compared to controls ($p = 0.0365$ and $p = 0.0240$, respectively) be due to population stratification, we suggest that these alterations may contribute to the phenotype. It will be

important to screen them in ASD patients from other populations as well as to perform functional studies.

In summary, we have shown that the use of an array specifically designed to detect small CNVs in candidate genes to ASD seems to be an efficient strategy to elucidate the etiologic mechanisms of nonsyndromic ASD. Our study also suggests that disruption of non-coding DNA might contribute to autism etiology in a significant proportion of cases. We found that most of the genes affected by CNVs are involved in two main biological functions – glutamatergic synapses and axonal guidance, suggesting that nonsyndromic ASD can be caused by dysfunction in different genes of a few common physiological processes. Finally, our findings lend support to the involvement of the genes *SHANK2*, *DIAPH3*, *GCH1*, *GRM5* and *MARK1* in ASD etiology and suggest the *SNAP29*, *PRKCa*, *MBD2*, *GAD2* and *SLC17A6* as new candidates to nonsyndromic ASD.

References

- AHN, J. W. et al. Detection of subtelomere imbalance using MLPA: validation, development of an analysis protocol, and application in a diagnostic centre. **BMC Med Genet.** 5;8:9, 2007.
- AMERICAN PSYCHIATRIC ASSOCIATION. **Diagnostic and statistical manual of mental disorders: DSM-IV.** 4th ed. Washington (DC): American Psychiatric Association. 943 p., 2000.
- AMIR, R. E. et al. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. **Nat Genet.** 23(2):185-8, 1999.
- BAILEY, A. et al. Autism as a strongly genetic disorder: evidence from a British twin study. **Psychol Med.**, 25(1):63-77, 1995.

BERKEL, S. et al. Mutations in SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. **Nature Genetics**, v.42, n.6, p.489-491, 2010.

BERUMENT, S. et al. Autism Screening Questionnaire: Diagnostic Validity. **British Journal of Psychiatry**, v. 175, p.444-451, 1999.

BETANCUR, C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. **Brain Res.** 22;1380:42-77, 2011.

BLANCHETTE, M. et al. Aligning multiple genomic sequences with the threaded blockset aligner. **Genome Res.** 14(4):708-15, 2004.

BONORA, E. et al. Mutation screening and association analysis of six candidate genes for autism on chromosome 7q. **Eur J Hum Genet.**, 13(2):198-207, 2005.

BREMER, A. et al. Copy number variation characteristics in subpopulations of patients with autism spectrum disorders. **Am J Med Genet B Neuropsychiatr Genet.**;156(2):115-24, 2011.

BREMER, A. et al. Screening for copy number alterations in loci associated with autism spectrum disorders by two-color multiplex ligation-dependent probe amplification. **Am J Med Genet B Neuropsychiatr Genet.** 5;153B(1):280-5, 2010.

BUCAN, M. et al. Genome-Wide Analyses of Exonic Copy Number Variants in a Family-Based Study Point to Novel Autism Susceptibility Genes. **PloS Genetics**, 5: e1000536, 2009.

CAMPBELL, C. M. et al. Polymorphisms in the GTP cyclohydrolase gene (GCH1) are associated with ratings of capsaicin pain. **Pain** 141(1–2): 114–118, 2009.

CAO L. et al. Four novel mutations in the GCH1 gene of Chinese patients with dopa-responsive dystonia. **Movement Disorders** 25(6): 755–760, 2010.

CHENG, Y.; MIURA, R. M.; TIAN, B. Prediction of mRNA polyadenylation sites by support vector machine. **Bioinformatics**. 1;22(19):2320-5, 2006.

CHIAROMONTE, F.; YAP, V. B.; MILLER, W. Scoring pairwise genomic sequence alignments. **Pac Symp Biocomput.** 115-26, 2002.

- CUSCO, I. et al. Autism-specific copy number variants further implicate the phosphatidylinositol signaling pathway and the glutamatergic synapse in the etiology of the disorder. **Hum. Mol. Genet.** 18, 1795-1804, 2009.
- DENT, E. W. et al. Filopodia are required for cortical neurite initiation. **Nat Cell Biol.** 9:1347–1359, 2007.
- DEVON, R. S. et al. The genomic organisation of the metabotropic glutamate receptor subtype 5 gene, and its association with schizophrenia. **Mol Psychiatry** 6:311-314, 2001.
- DREWES, G. et al. et al. MARK, a novel family of protein kinases that phosphorylate microtubule-associated proteins and trigger microtubule disruption. **Cell**. 18;89(2):297-308, 1997.
- ELIA, J. et al. Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. **Molecular Psychiatry** 15, 637–646, 2010.
- FOMBONNE, E. Epidemiology of pervasive developmental disorders. **Pediatr Res.** 65(6):591-8, 2009.
- FRIEDMAN, R. C. et al. Most Mammalian mRNAs Are Conserved Targets of MicroRNAs. **Genome Research**. 19:92-105, 2009.
- GESCHWIND, D. H. Autism, many genes, common pathways? **Cell**. 391-395, 2008.
- GIRIRAJAN, S. et al. Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. **Am J Hum Genet.** 7;92(2):221-37, 2013.
- GLESSNER, J. T. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. **Nature**. 459, 569-573, 2009.
- GRIESI-OLIVEIRA, K. et al A complex chromosomal rearrangement involving chromosomes 2, 5, and X in autism spectrum disorder. **Am J Med Genet B Neuropsychiatr Genet.** 159B(5):529-36, 2012.

- GRIMSON, A. et al. MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. **Molecular Cell.** 6;27(1):91-105, 2007.
- GUPTON, S. L. et al. mDia2 regulates actin and focal adhesion dynamics and organization in the lamella for efficient epithelial cell migration. **J Cell Sci.** 120:3475–3487, 2007.
- IAFRATE, A. J. et al. Detection of large-scale variation in the human genome. **Nat Genet.** 36(9):949-51, 2004.
- IOSSIFOV, I. et al. De novo gene disruptions in children on the autistic spectrum. **Neuron.** 26;74(2):285-99, 2012.
- 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.
- KENT, W. J. et al. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. **Proc Natl Acad Sci U S A.**; 100(20):11484-9, 2003.
- KUMAR, R. A. et al. SHANK2 redemption: another synaptic protein for mental retardation and autism. **Clin Genet.** 78(6):519-21, 2010.
- 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.
- LEWIS, B. P.; BURGE, C. B.; BARTEL, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. **Cell.** 14;120(1):15-20, 2005.
- LI, C.; HUNG-WONG, W. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. **Genome Biol.** 2(8):RESEARCH0032, 2001.
- LIN, M. et al. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. **Bioinformatics.** 22;20(8):1233-40, 2004.
- MARSHALL, C. R. et al. Structural variation of chromosomes in autism spectrum disorder. **Am.**

J. Hum. Genet. 82:477–488, 2008.

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.

MURPHY, W. J. et al. Resolution of the early placental mammal radiation using Bayesian phylogenetics. **Science**. 2001 Dec 14;294(5550):2348-51.

NAVEDO, M. F. et al. Increased coupled gating of L-type Ca²⁺ channels during hypertension and Timothy syndrome. **Circ Res**. 5;106(4):748-56, 2010.

NEALE, B. M. et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. **Nature**. 4;485(7397):242-5, 2012.

O'ROAK, B. J. et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. **Nat Genet**. 43(6):585-9, 2011.

O'ROAK, B. J. et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. **Science**. 21;338(6114):1619-22, 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.

ORABONA, G. M. et al. HTR1B and HTR2C in autism spectrum disorders in Brazilian families. **Brain Res**. 23;1250:14-9, 2009.

PAN, P. Y. et al. SNAP-29-mediated modulation of synaptic transmission in cultured hippocampal neurons. **J Biol Chem**. 8;280(27):25769-79, 2005.

PENG, J.; XU, J. RaptorX: exploiting structure information for protein alignment by statistical inference. **Proteins**. 79 Suppl 10:161-71, 2011.

PINTO, D. et al. Functional impact of global rare copy number variation in autism spectrum disorders. **Nature**. 15;466(7304):368-72, 2010.

POLLARD, K. S.; HUBISZ, M. J.; SIEPEL, A. Detection of non-neutral substitution rates on

mammalian phylogenies. **Genome Res.** 2009.

PONCHEL, F. et al. Real-time PCR based on SYBR-Green I fluorescence: An alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions. **BMC Biotechnology** 3:18, 2003.

QIAO, Y. et al. Phenomic determinants of genomic variation in autism spectrum disorders. **J Med Genet.** 46(10):680-8, 2009.

SANDELIN, A.; WASSERMAN, W. W.; LENHARD, B. ConSite: web-based prediction of regulatory elements using cross-species comparison. **Nucleic Acids Res.** 1;32(Web Server issue):W249-52, 2004.

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. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. **Neuron**. 9;70(5):863-85. doi: 10.1016/j.neuron.2011.05.002, 2011.

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.; ZOGHBI, H. Y. Solving the autism puzzle a few pieces at a time. **Neuron**. 9;70(5):806-8, 2011.

SCHOPLER, E. et al. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). **Journal of Autism and Developmental Disorders**, v. 10, n. 1, p. 91-103, 1980.

SCHWARTZ, S. et al. Human-mouse alignments with BLASTZ. **Genome Res.** 13(1):103-7, 2003.

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

SEROVA L. I. et al. Heightened transcription for enzymes involved in norepinephrine biosynthesis in the rat locus coeruleus by immobilization stress. **Biol Psychiatry** 45(7): 853–862, 1999.

SHIBAYAMA, A. et al. MECP2 structural and 3'-UTR variants in schizophrenia, autism and other psychiatric diseases: a possible association with autism. **Am J Med Genet B Neuropsychiatr Genet**. 1;128B(1):50-3, 2004.

SIEPEL, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. **Genome Res.** 15(8):1034-50, 2005.

SIMONYI, A.; SCHACHTMAN, T. R.; CHRISTOFFERSEN, G. R. Metabotropic glutamate receptor subtype 5 antagonism in learning and memory. **Eur. J. Pharmacol.** 639, 17–25, 2010.

SKAFIDAS, E. et al. Predicting the diagnosis of autism spectrum disorder using gene pathway analysis. **Mol Psychiatry**. 2012.

SORTE, H. S. et al. Copy number variation findings among 50 children and adolescents with autism spectrum disorder. **Psychiatr Genet**. 23(2):61-9, 2013.

SU, Q. et al. SNAP29: a general SNARE protein that inhibits SNARE disassembly and is implicated in synaptic transmission. **Proc Natl Acad Sci USA**. 98:14038–14043, 2001.

SUDHOF, T. C. The synaptic vesicle cycle. **Annu Rev Neurosci**. 27:509–547, 2004.

SZATMARI, P. et al. High phenotypic correlations among siblings with autism and pervasive developmental disorders. **Am J Med Genet**. 26;67(4):354-60, 1996.

TEGEDER, I et al. Reduced hyperalgesia in homozygous carriers of a GTP cyclohydrolase 1 haplotype. **European Journal of Pain**. 12(8): 1069–1077, 2008.

TEGEDER, I. et al. GTP cyclohydrolase and tetrahydrobiopterin regulate pain sensitivity and persistence. **Nat Med** 12(11): 1269–1277, 2006

VOINEAGU, I. et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. **Nature**. 25;474(7351):380-4, 2011.

- VOINESKOS, A. N. et al. Neurexin-1 and frontal lobe white matter: an overlapping intermediate phenotype for schizophrenia and autism spectrum disorders. **PLoS One**. 6(6):e20982, 2011.
- VORSTMAN, J. A. et al. A double hit implicates DIAPH3 as an autism risk gene. **Mol Psychiatry**. 16(4):442-51, 2011.
- 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 Res**. 17(11):1665-74, 2007.
- WANG, K.; 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, 667–675, 2008.
- WEKSBERG, R. et al. A method for accurate detection of genomic microdeletions using real-time quantitative PCR. **BMC Genomics**. 13;6:180, 2005.
- XU, J. et al. mGluR5 has a critical role in inhibitory learning. **J. Neurosci**. 29, 3676–3684, 2009.
- ZHANG, H.; HU, J.; RECCE, M.; TIAN B. PolyA_DB: a database for mammalian mRNA polyadenylation. **Nucleic Acids Res**. 1;33(Database issue):D116-120, 2005.
- ZHANG, J. et al. Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. **Cytogenet Genome**. 115(3-4):205-14, 2006.
- ZHI, J.; HATCHWELL, E. Human MLPA Probe Design (H-MAPD): a probe design tool for both electrophoresis-based and bead-coupled human Multiplex ligation-dependent probe amplification assays. **BMC Genomics**, v.9, n. 407, 2008.
- ZWAIGENBAUM, L. Behavioral manifestations of autism in the first year of life. **Int J Dev Neurosci**. 23(2-3):143-52, 2005.

Capítulo 4

Characterization of microchromosomal rearrangements in syndromic autistic patients

Tese Doutorado

Estudo de genes candidatos aos transtornos do espectro autista

Abstract

In this study we analyzed 18 patients with autism spectrum disorders (ASD) plus dysmorphic features and/or congenital malformations that do not fit any recognizable syndrome. The type, size and frequency of CNVs detected are compared to alterations identified in another our study of nonsyndromic cases using the same custom-designed microarray. Based in our molecular findings we discuss which genes located in the deleted or duplicated regions may be ASD candidates and the possible inheritance model of these disorders for syndromic cases.

Resumo

Neste estudo foram analisados 18 pacientes com transtornos do espectro autista (TEA) e com características dismórficas e/ou malformações congênitas que não se enquadram em qualquer síndrome conhecida. O tipo, o tamanho e a frequência das CNVs detectadas são comparadas com alterações identificadas em outro estudo de casos não sindrômicos usando a mesma plataforma de microarray customizada. Diante dos resultados obtidos, discutimos o possível modelo de herança para casos de TEA sindrômicos e quais entre os genes localizados nas regiões deletadas ou

duplicados podem ser candidatos aos TEA.

Introduction

Autism spectrum disorders (ASD) (OMIM: 209850) are a group of severe neurodevelopmental conditions, characterized by restricted, repetitive and stereotyped behaviors and impairments in social skills and communication, with onset in early infancy. The ASD prevalence is estimated at 1% and the phenotype is highly heterogeneous exhibiting a wide range of severity (DUVALL *et al.*, 2007; FOMBONNE *et al.*, 2009; SCHAAF; ZOGHBI, 2011). A large proportion of individuals with ASD also have intellectual disabilities, physical/visible malformations, and/or dysmorphic features (BREMER *et al.*, 2010). Despite of many evidences showing that genetic factors play a major role in the genesis of these disorders (BAILEY *et al.*, 1995), the underlying genetic cause is unknown in at least 70% of cases. Advances in whole-genome scanning technologies such as microarrays analysis have resulted in the detection of copy number variations (CNVs) in 10 to 20% of the patients (MILES *et al.*, 2010). Recently, we have lent support to the use of custom-designed arrays, targeted to analyze exons of ASD candidate genes and able to detect small CNVs (<10 kbp), to study cases of isolated autism (RIBEIRO *et al.*, 2013). Here, we evaluated whether the use of our custom-designed array comparative genomic hybridization (CGH) is also an efficient strategy to evaluate syndromic forms of ASD, that is, autistic patients with dysmorphic features and/or congenital malformations that do not fit any recognizable syndrome. This study also aimed to identify new ASD candidate genes.

Patients and Controls

We selected for the study 18 Brazilian patients (ten boys and eight girls, ranging in age from 3 to 29 years) diagnosed with autistic disorder, dysmorphic features and/or congenital malformations (classified in this study as syndromic).

The patients were ascertained at the *Instituto de Psiquiatria do Hospital das Clínicas*, University of São Paulo and referred to the *Centro do Estudos do Genoma Humano*, São Paulo. All of them were diagnosed by psychiatrists using DSM-IV (Diagnostic and Statistical Manual of Mental Disorder), ICD-10 (International Statistical Classification of Diseases and Related Health Problems) and ASQ (Autism Screening Questionnaire) (BERUMENT *et al.*, 1999; ORABONA *et al.*, 2009; GRIESI-OLIVEIRA *et al.*, 2012). Children's behavior was characterized by score questions of CARS (Childhood Autism Rating Scale) (SCHOPLER *et al.*, 1980).

Patients with recognizable autism-related syndromes, metabolic disorders or those whose mothers were exposed to teratogenic drugs and teratogenic infectious agents during pregnancy were not included in this study. All patients were negative for CNVs at 15q11.13, 16p11 and 22q13 through multiplex ligation probe amplification (MLPA) analysis (SALSA MLPA KIT P343-B1 AUTISM-1; MRC-Holland, Amsterdam, The Netherlands). DNA samples from 200 apparently healthy individuals were used as controls in the custom-designed array CGH experiment. The Ethics Committee of Bioscience Institute – University of São Paulo approved this research. Patients and controls were included only after a signed written informed consent by the families or other legal caregivers.

Investigation of copy number alterations by custom-designed Comparative Genomic Hybridization on Microarrays

The first screening was performed by custom-designed oligonucleotide array CGH with a high probe density targeted to exons of 269 putative candidate genes to autism. We used platforms from Agilent Technologies (Agilent SurePrint G3 Human CGH Microarrays 8×60K [containing 60,000 oligonucleotides probes]). Briefly, samples were labeled with Cy3- and Cy5-dCTPs by random priming, and purification, hybridization, and washing were carried out as recommended by the manufacturer. Scanned images of the arrays were processed using the Feature Extraction software (Agilent Technologies, Santa Clara, CA), and the analysis was carried out using the Agilent genomic Workbench 7.0 (Agilent Technologies, Santa Clara, CA). For array CGH analysis, identification of aberrant copy number segments was based on the ADM-2 algorithm with default settings (threshold log₂ ratio of 0.2 or 1.14 was used for gain or high copy gain, and -0.23 and -1.14 used for loss and homozygous loss, respectively), and the significance threshold was set on 7. We considered at least three consecutive probes for calling a segment. All ASD samples were tested in duplicate. Only the autism exclusive CNVs [not observed in 200 samples from healthy individuals and Database of Genomic Variants (DGV) (<http://projects.tcag.ca/variation>)] were considered possibly pathogenic and were confirmed using another array CGH platform (Agilent SurePrint G3 Human CGH Microarrays 4×180K [180,000 oligonucleotides probes]), SNP-array (GeneChip Human Mapping 500K Array Set Affymetrix) or quantitative real-time PCR.

Quantitative real-time PCR

Quantitative real-time PCR (qPCR) was performed using the ABI Prism 7900HT Fast Real-time Sequence Detection System (Applied Biosystems, Foster City, CA). Reactions were carried out in triplicates using SYBR Green PCR MasterMix (Applied Biosystems) in a total volume of

20 µL. Dissociation curves were analyzed for each primer pair to verify the specificity of the qPCR reaction. The copy number quantification was calculated as described by PONCHEL *et al.* (2003) and ΔK_{Ct} (Fold copy number change) values of 0 ± 0.35 indicate an equal ratio of the target and the reference, which corresponds to no genetic abnormality, values of -1 ± 0.35 indicate loss of one copy (microdeletion), whilst values of 1 ± 0.35 indicate gain of one copy (microduplication). *G6PDH* (WEKSBERG *et al.* 2005) and *hGAPDH* (PONCHEL *et al.*, 2003) were used as endogenous control genes. The primers are shown in table 2.2 (chapter 2).

Single Nucleotide Polymorphism Array

The Affymetrix GeneChip Human Mapping 500K Array Set is comprised of two arrays, each capable of genotyping on average 250,000 SNPs (approximately 262,000 for *Nsp* arrays and 238,000 for *Sty* arrays). The assay was performed according to the standard protocol. The quality of hybridization was evaluated by the QC Call Rate, using 95.0 as threshold. For the analysis we used the softwares Genotyping Console (Affymetrix, Santa Clara, CA, USA), PennCNV (WANG *et al.*, 2007; WANG; BUCAN, 2008) and dChip (LI; HUNG-WONG, 2001; LIN *et al.* 2004), with default settings. We considered only CNVs detected by at least two different softwares.

Results

We analyzed 18 individuals with idiopathic ASD plus dysmorphic features and/or congenital malformations that do not fit any recognizable syndrome and 200 matched controls (RIBEIRO *et al.*, 2013). In seven (four boys and three girls) out of the 18 patients (38.9%), we identified ten

autism specific CNVs (seven deletions and three duplications), which were not observed either in the Database of Genomic Variants (DGV - IAFRATE *et al.*, 2004; ZHANG *et al.*, 2006), a catalog of CNVs found in apparently healthy subjects of diverse populations or in the 200 Brazilian controls. Their sizes ranged from 367 pb to 25 Mbp. The clinical characteristics of these patients and the description of each of these CNVs are summarized in Table 4.1.

The CNVs were subclassified into two groups according to their size. The first group (GI) includes five CNVs spanning more than 500 kbp: three are larger than 10 Mbp (a deletion at 13q14.11-q14.3; a duplication at 1p31.3-p22.3 and a deletion at 3p14.3-p14.1) and two have approximately 2 Mbp (a deletion at 11q13.2-q13 and a deletion at 14q21.2-q21.3). The second group (GII) contains five CNVs spanning less than 500 kbp. In this group, except for one deletion of 66 kbp encompassing three genes, the alterations involve part of one coding region (the deletion of exons 1 to 4 of *GRIA3*) or untranslated regions (UTR) (Table 4.1).

It was possible to determine the origin of the CNVs in five of the seven patients, being the CNVs *de novo* in three of them (1, 3 and 6), and paternally inherited in patient 5, while patient 4 has two CNVs, one maternally inherited and the other *de novo* (Table 4.1). Excluding the two cases in which parental samples were unavailable (patients 2 and 7), the rate of patients with *de novo* alterations was estimated as 25% [(4/16) – patients 1, 3, 4 and 6].

Table 4.1 Possibly pathogenic CNVs identified in syndromic ASD patients

Case	Isolated or Familial	Sex	CNV position (GRCh37/hg19)	Type	Clinical alterations in addition to autism	Origin
			Genes		Size	
1	Isolated	Male	13q14.11-q14.3 131 genes	Deletion	Unilateral retinoblastoma; Brachycephaly; High arched palate; Synophrys; Psychomotor development delay.	<i>de novo</i>
			10 Mbp			
2	Isolated	Female	17q24.2 <i>PRKCa</i> 3'UTR 1.3 Kbp	Duplication	Hypertelorism; Arched eyebrows; Uplanting palpebral fissures and elongate eyelid; Short philtrum; Thick and everted lips; Prominent ears; Hypoplasia of labia minora; Clinodactyly 4 toe.	parental DNA unavailable
			11q13.3 <i>SHANK2</i> 3'UTR 367 bp	Duplication		
3	Isolated	Male	11q13.2-q13.4 39 genes	Deletion	Microcephaly;	<i>de novo</i>
			2 Mbp			

4	Isolated	Female	1p31.3-22.3 223 genes 25 Mbp	Duplication	Cleft palate; Unavailable for more detailed clinical evaluation.	<i>de novo</i>
			14q21.2-q21.3 <i>MDGA2</i> <i>RPL10L</i> 2 Mbp	Deletion		Inherited from father
5	Isolated	Male	16q24.3 <i>SLC22A31</i> <i>CDH15</i> <i>ZNF778</i> 66 Kbp	Deletion	Hypertelorism; Elongated palpebral fissures; Long eyelashes; Short philtrum; Micro-retrognathia; Prominent superior central incisors; Large ears; Elongated neck; Clinodactyly; Epilepsy	Inherited from father
6	Isolated	Female	3p14.3-p14.1 85 genes 12 Mbp	Deletion	Face with flat profile; Straight forehead; Bushy eyebrows; Uplanting palpebral fissures; Elongated nasal tip; Short philtrum; Oral malocclusion; Ears with decreased space between helix and anti-helix; Tapering fingers; Small feet with <i>pes cavus</i> .	<i>de novo</i>
7	Isolated	Male	Xq25 <i>GRIA3</i> 114 Kbp 7q21.11 <i>SEMA3A</i> 3'UTR 380 bp	Deletion	Tongue protrusion; Epicanthic fold; Prominent ears.	Mother does not have the mutation. Father's DNA unavailable

In the present study, we evaluated 18 syndromic autistic patients with a custom-designed array CGH and identified, ten autism specific CNVs in 7 (38.9%) of them, with apparently no preponderance of the sex, as four were males and three females. This detection rate was significantly higher than the 11.6% rate (ten in 103) (Fisher's test; $p = 0.004$) detected among nonsyndromic ASD cases using this same custom-designed array-CGH (RIBEIRO *et al.*, 2013). Indeed, using conventional array-CGH platforms, a higher detection rate (9% to 25%) has also been observed among syndromic ASD as compared to nonsyndromic ASD cases (8% to 10%) (JACQUEMONT *et al.*, 2006; QIAO *et al.*, 2009; BREMER *et al.*, 2010; SORTE *et al.*, 2013). In addition, we observed that large CNVs (>500 Kbp or GI) were found only among syndromic ASD, as all the CNVs among our nonsyndromic cases were below 300 kbp (RIBEIRO *et al.*, 2013). The proportion of deletions (7/10) seems to be higher than duplications among syndromic cases, in contrast to our findings in nonsyndromic cases, where the proportions of duplications and deletions were similar.

The GI CNVs encompass several genes hampering the precise identification of those that contribute to the phenotype. Nevertheless, we will highlight some genes that encode proteins with direct or indirect action on neuronal functions and/or genes previously associated with ASD or other neuropsychiatrists disorders.

The largest alteration of GI is the duplication at 1p31.3-p22.3 (patient 4), comprising 223 genes. Although none of them has been previously reported as an ASD candidate gene, case-control studies related *LRRC7* and *LRRIQ3*, mapped in this interval, as associated with attention deficit-hyperactivity disorder (ADHD), schizophrenia and bipolar disorder (CURTIS *et al.*, 2011; MICK *et al.*, 2011). In addition to the duplication in chromosome 1, this patient also inherited a deletion at 14q21.2-q21.3 that encompasses the whole *MDGA2*, which codifies a protein possibly involved in neural network formation and neuronal plasticity (BUCAN *et al.*, 2009). One exonic inherited deletion in this gene was previously described in an autistic individual (BUCAN *et al.*,

2009). So, two independent studies have detected affected individuals with mutations in this gene, which reinforce the importance of *MDGA2* in autism etiology. It is also possible that the ASD phenotype of patient 4 depends on the haploinsufficiency of *MDGA2* added to the effect of duplication of some of the genes, as *LRRC7* and *LRRKQ3*, located in the altered region of chromosome 1. This would explain the incomplete penetrance of the ASD phenotype in his unaffected mother, who is a carrier of the deletion 14q21.2-q21.3 only.

The smallest alteration of GI encompasses *SHANK2* (patient 3). Heterozygous mutations in *SHANK2*, whose protein acts as a molecular scaffold in postsynaptic membranes (BOECHERS *et al.*, 2002), have been recurrently reported in ASD patients with intellectual disability (BERKEL *et al.*, 2010; PINTO *et al.*, 2010). In addition to *SHANK2*, other 38 genes are located in the deleted region and we cannot disregard that the haploinsufficiency of some of these genes are also contributing to the patient phenotype. Indeed, Leblond *et al.* (2012) suggested that additional genetic, epigenetic or environmental factors seem to be necessary for the emergence of the ASD in individuals with mutations in *SHANK2*.

The deleted region 3p14.3-p14.1 (patient 6) includes, among others, the gene *FHIT*, which was previously found to be deleted and duplicated in individuals with facial dysmorphisms, intellectual disabilities, speech delay and autism (SEBAT *et al.*, 2007; Decipher - <http://decipher.sanger.ac.uk>). Another interesting gene located in this region is *CADPS1*, which encoded protein acts along with the CADPS2 protein in the vesicular exocytosis neuronal process (JOCKUSCH *et al.*, 2007; LIU *et al.*, 2008) and monoamine uptake and storage (BRUNK *et al.*, 2009). Considering that defects of CADPS2 function might contribute to ASD susceptibility (SADAKATA *et al.*, 2007), we suggest that *CADPS1* can also be a candidate gene for ASD.

Among the genes located in the deleted region at 13q14.11-q14.3 (patient 1), we highlight the *HTR2A*, which encodes a serotonin receptor. There is a wealth of evidence that this neurotransmitter plays a role in the pathophysiology of ASD and the therapeutic benefits obtained with the use of serotonergic drugs by autistic patients are widely recognized (ANDERSON *et al.*, 1987; GORDON *et al.*, 1993; COOK; LEVENTHAL, 1996; ORABONA *et al.*, 2009). Another deleted gene in this region with potential relevance to the autistic phenotype is the *NUFIP1*. The protein encoded by this gene acts in association with FMRP (*FMR1* product) regulating the

synthesis of proteins relevant for synaptic plasticity (BARDONI *et al.*, 1999; BARDONI *et al.*, 2003).

Therefore, there are at least two genes previously associated with ASD or with potential to be considered new candidates for ASD within each chromosome interval encompassed by the GI CNVs, suggesting an oligogenic model for the autism phenotype among these syndromic cases.

Most of the GII CNVs are located in UTR regions and deserve further functional investigation. The 3'UTR duplications at the genes *PRKCa* and *SHANK2* found in one of our syndromic patients were also found in two nonsyndromic ASD but not in any of the 200 matched controls neither in the DGV (RIBEIRO *et al.*, 2013), which might indicate their involvement in the emergence of the ASD phenotype. In patient 7, the CNV is located at the 3'UTR of *SEMA3A*, a gene previously associated with behavior impairments usually observed among autistic patients, such as attention problems, anxiety and aggressiveness (MICK *et al.*, 2011). In addition to this alteration, the patient also has a partial deletion of the gene *GRIA3*, which encodes glutamate receptor AMPA3. Mutations at *GRIA3* have been associated with X-linked mental retardation (CHIYONOBU *et al.*, 2007), bipolar disorder (GÉCZ *et al.*, 1999), schizophrenia and autism (GUILMATRE *et al.*, 2009).

Microdeletions in 16q24.3, similar to that observed in patient 5, have been previously reported in autistic males with congenital malformations and facial dysmorphisms (WILLEMSSEN *et al.*, 2010). We observed that the region of overlap among all these CNVs encompasses only the gene *ZNF778*, which encodes a zinc finger protein. Hence, our result reinforces the importance of this gene to the autistic phenotype in patients with the 16q24.3 deletion. Another interesting gene located in the deleted region is *CDH15*, which encodes a calcium-dependent intercellular adhesion glycoprotein. Missense mutations in this gene have been associated with an autosomal dominant form of intellectual disability (BHALLA *et al.*, 2008). We hypothesize that the haploinsufficiency of both, *ZNF778* and *CDH15*, contributes to the emergence of autism in our patient.

Except for the 16q24.3 deletion, all the other GII CNVs comprise just one gene. However, we observed more than one small CNV per patient, suggesting the oligogenic model might account for the autism phenotype in these cases as well. Although our sample is small, it is of

note that three out of the five GII CNVs are located either in 5' or 3'UTR. Interestingly, 50% of the CNVs detected among nonsyndromic cases were also located in this type of region (RIBEIRO *et al.*, 2013). These results reinforce that alterations in regulatory regions of the genome might contribute to ASD, as also previously hypothesized by others (SHIBAYAMA *et al.*, 2004; BONORA *et al.*, 2005; VOINESKOS *et al.*, 2011)

In summary, the percentage of patients with potential pathogenic CNVs detected in our sample (38,9%) was higher than those commonly reported for syndromic ASD (9% to 27.5% - JACQUEMONT *et al.*, 2006; QIAO *et al.*, 2009; BREMER *et al.*, 2010). Therefore, the strategy of using a custom-designed array, enriched with probes targeted to genes with potential to be involved in the ASD etiology and able to detect both large and small CNVs, seems to be relevant in an attempt to elucidate the largest number of cases and to better understand these disorders. Nevertheless, a larger sample should be tested. We observed that most of patients analyzed are either carriers of a CNV comprising multiple ASD candidate genes or of multiple CNVs, which support an oligogenic model for syndromic cases. Finally, besides corroborating the involvement of the genes *MDGA2*, *FHIT*, *HTR2A*, *SHANK2*, *GRIA3*, *ZNF778*, *PRKC α* and *CDH15* in the etiology of ASD, we propose for the first time that *LRRC7*, *LRRIQ3*, *CADPS1*, *NUFIP*, *SEMA3A* are candidates for these disorders.

Acknowledgements

This study makes use of data generated by the DECIPHER Consortium. A full list of centres who contributed to the generation of the data is available from <http://decipher.sanger.ac.uk> and via email from decipher@sanger.ac.uk. Funding for the project was provided by the Wellcome Trust.

References

- ANDERSON, G. M. et al. Whole blood serotonin in autistic and normal subjects. **J. Child Psychol. Psychiatry**, 28:885–900, 1987.
- BAILEY, A. et al. Autism as a strongly genetic disorder: evidence from a British twin study. **Psychol Med.**, 25(1):63-77, 1995.
- BARDONI, B. et al. A novel RNA-binding nuclear protein that interacts with the fragile X mental retardation (FMR1) protein. **Human Molecular Genetics**, 8(13): 2557–2566, 1999.
- BARDONI, B. et al. NUFIP1 (nuclear FMRP interacting protein 1) is a nucleocytoplasmic shuttling protein associated with active synaptoneuroosomes. **Exp Cell Res.** 10;289(1):95-107, 2003.
- BERKEL, S. et al. Mutations in SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. **Nature Genetics**, v.42, n.6, p.489-491, 2010.
- BHALLA, K. et al. Alterations in CDH15 and KIRREL3 in patients with mild to severe intellectual disability. **Am J Hum Genet.** 83(6):703-13, 2008.
- BOECKERS, T. M. et al. ProSAP/Shank proteins - a family of higher order organizing molecules of the postsynaptic density with an emerging role in human neurological disease. **J Neurochem.** 81(5):903-10, 2002.
- BONORA, E. et al. Mutation screening and association analysis of six candidate genes for autism on chromosome 7q. **Eur J Hum Genet.**, 13(2):198-207, 2005.
- BREMER, A. et al. Screening for copy number alterations in loci associated with autism spectrum disorders by two-color multiplex ligation-dependent probe amplification. **Am J Med Genet B Neuropsychiatr Genet.** 5;153B(1):280-5, 2010.

BRUNK, I. et al. Ca²⁺-dependent activator proteins of secretion promote vesicular monoamine uptake. **J Biol Chem.** 9;284(2):1050-6, 2009.

BUCAN, M. et al. Genome-Wide Analyses of Exonic Copy Number Variants in a Family-Based Study Point to Novel Autism Susceptibility Genes. **PloS Genetics**, 5: e1000536, 2009.

CHIYONOBU, T. et al. Partial tandem duplication of GRIA3 in a male with mental retardation. **Am J Med Genet A**. 1;143A(13):1448-55, 2007.

COOK, E.; LEVENTHAL, B. The serotonin system in autism. **Curr. Opin. Pediatr.**, 8:348–354, 1996.

CURTIS, D. Et al. Case-case genome-wide association analysis shows markers differentially associated with schizophrenia and bipolar disorder and implicates calcium channel genes. **Psychiatr Genet.** 21(1):1-4, 2011.

DUVALL, J. A. et al. A quantitative trait locus analysis of social responsiveness in multiplex autism families. **Am J Psychiatry**. 164(4):656-62, 2007.

FOMBONNE, E. Epidemiology of pervasive developmental disorders. **Pediatr Res.** 65(6):591-8, 2009.

GÉCZ, J. et al. Characterization of the human glutamate receptor subunit 3 gene (GRIA3), a candidate for bipolar disorder and nonspecific X-linked mental retardation. **Genomics**, 15;62(3):356-68, 1999.

GORDON, C. et al. A double-blind comparison of clomipramine, desipramine, and placebo in the treatment of autistic disorder. **Arch. Gen. Psychiatry**, 50:441–447, 1993.

GRIESI-OLIVEIRA, K. et al A complex chromosomal rearrangement involving chromosomes 2, 5, and X in autism spectrum disorder. **Am J Med Genet B Neuropsychiatr Genet.** 159B(5):529-36, 2012.

GUILMATRE A, et al. Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. **Arch Gen Psychiatry**. 66(9): 947–956, 2009.

- 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.
- JOCKUSCH, W. J. et al. CAPS-1 and CAPS-2 are essential synaptic vesicle priming proteins. **Cell**. 16;131(4):796-808, 2007.
- LI, C.; HUNG-WONG, W. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. **Genome Biol**. 2(8):RESEARCH0032, 2001.
- LIN, M. et al. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. **Bioinformatics**. 22;20(8):1233-40, 2004.
- LIU, Y. et al. CAPS facilitates filling of the rapidly releasable pool of large dense-core vesicles. **J Neurosci**. 21;28(21):5594-601, 2008.
- MICK, E. Genome-wide association study of the child behavior checklist dysregulation profile. **J Am Acad Child Adolesc Psychiatry**. 50(8):807-17.e8, 2011.
- ORABONA, G. M. et al. HTR1B and HTR2C in autism spectrum disorders in Brazilian families. **Brain Res**. 23;1250:14-9, 2009.
- PINTO, D. et al. Functional impact of global rare copy number variation in autism spectrum disorders. **Nature**. 15;466(7304):368-72, 2010.
- PONCHEL, F. et al. Real-time PCR based on SYBR-Green I fluorescence: An alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions. **BMC Biotechnology** 3:18, 2003.
- QIAO, Y. et al. Phenomic determinants of genomic variation in autism spectrum disorders. **J Med Genet**. 46(10):680-8, 2009.
- SADAKATA, T. et al. Tissue distribution of Ca²⁺-dependent activator protein for secretion family members CAPS1 and CAPS2 in mice. **J Histochem Cytochem**. 55(3):301-11, 2007.

- SEBAT, J. et al. Strong association of de novo copy number mutations with autism. **Science**, 316:445–449, 2007.
- SCHAFF, C.P.; ZOGHBI, H. Y. Solving the autism puzzle a few pieces at a time. **Neuron**. 9;70(5):806-8, 2011.
- SCHOPLER, E. et al. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). **Journal of Autism and Developmental Disorders**, v. 10, n. 1, p. 91-103, 1980.
- SHIBAYAMA, A. et al. MECP2 structural and 3'-UTR variants in schizophrenia, autism and other psychiatric diseases: a possible association with autism. **Am J Med Genet B Neuropsychiatr Genet**. 1;128B(1):50-3, 2004.
- SORTE, H. S. et al. Copy number variation findings among 50 children and adolescents with autism spectrum disorder. **Psychiatr Genet**. 23(2):61-9, 2013.
- VOINESKOS, A. N. et al. Neurexin-1 and frontal lobe white matter: an overlapping intermediate phenotype for schizophrenia and autism spectrum disorders. **PLoS One**. 6(6):e20982, 2011.
- 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 Res**. 17(11):1665-74, 2007.
- WANG, K.; BUCAN, M. Copy Number Variation Detection via High-Density SNP Genotyping. **Cold Spring Harb. Protoc.**; DOI:10.1101/pdb.top46, 2008.
- WEKSBERG, R. et al. A method for accurate detection of genomic microdeletions using real-time quantitative PCR. **BMC Genomics**. 13;6:180, 2005.
- WILLEMSEN, M. H. et al. Identification of ANKRD11 and ZNF778 as candidate genes for autism and variable cognitive impairment in the novel 16q24.3 microdeletion syndrome. **Eur. J. Hum. Genet.**, 18(4):429-35, 2010.

Capítulo 5

Search for genetic factors that modulate the severity of Autism Spectrum Disorders

Tese Doutorado

Estudo de genes candidatos aos transtornos do espectro autista

Abstract

In this study we describe a partial deletion at the *IMMP2L* that segregates in a family with two carriers of Asperger syndrome and one carrier of Autistic disorder. Further investigations showed the presence of additional shared and unshared sequence variants at genes whose proteins perform neuronal functions in the affected individuals. The patient with more severe phenotype also has a duplication of about 2 Mbp at 10q11.22. The possibly damaging genetic effects of all these alterations as well as the hypothesis that the clinical variability of ASD is modulated by the number of hits and possibly by hits in both alleles of a same gene are discussed.

Resumo

Neste estudo descrevemos uma deleção parcial no gene *IMMP2L* que segraga em uma família com dois portadores da síndrome de Asperger e um portador de transtorno autista. Investigações genéticas posteriores mostraram a presença de mutações de ponto ou InDels, compartilhadas e não compartilhadas entre todos os afetados, em genes que codificam proteínas com função neuronal. O paciente com fenótipo mais severo também apresenta uma duplicação de

cerca de 2 Mpb na região 10q11.22. Os possíveis efeitos prejudiciais de todas estas alterações são discutidos, assim como, a hipótese de que a variabilidade clínica dos transtornos do espectro autista pode ser modulada pelo número de mutações e possivelmente por duas diferentes alterações em ambos os alelos do mesmo gene.

Introduction

Autism Spectrum Disorders (ASD) are a group of neuropsychiatric disorders, characterized by repetitive and stereotyped behavior, inability to communicate, and lack of reciprocal social interaction. ASD comprise Autistic Disorder (AD), Asperger Syndrome (AS) and Pervasive Developmental Disorders Not Otherwise Specified (PDD-NOS). AD is characterized by the following triad of symptoms: delay or total lack of development of spoken language, not compensated by nonverbal behaviors, impairments in social interaction and repetitive and stereotyped patterns of behavior, interests, and activities. AS differs from AD due to the higher cognitive abilities and to the absence of qualitative impairments in communication, while PDD-NOS is the classification adopted when the triad of symptoms is not fully present or the symptoms arise later in the infancy, after the age of three (Diagnostic and Statistical Manual, IV edition (DSM-IV); VEENSTRA-VANDERWEELE; COOK, 2004).

In the last decade, high-resolution genomic methodologies, as microarray and next generation sequencing (NGS), revealed that rare copy number variations (CNVs) and small mutations represent a causative genetic mechanism in about 10 to 20% of the ASD cases (SEBAT *et al.*, 2007; MARSHALL *et al.*, 2008; MILES *et al.*, 2010; PINTO *et al.*, 2010; IOSSIFOV *et al.*, 2012; NEALE *et al.*, 2012; O'ROAK *et al.*, 2012; SANDERS *et al.*, 2012; RIBEIRO *et al.*, 2013). The data obtained in microarray and NGS studies suggest that in familial cases susceptibility alleles are transmitted by non-affected parents and act together to modulate

the ASD risk (SEBAT *et al.*, 2007; MARSHALL *et al.*, 2008; ITSARA *et al.*, 2010; GIRIRAJAN *et al.*, 2010). AS, AD and PDD-NOS can co-segregate within the same family (ROSENBERG *et al.*, 2009), raising the following hypothesis: would the patients with different severity of the disorder share exactly the same genetic alterations and the variability of the clinical phenotype would be dependent on stochastic effects or epigenetic mechanisms? Or, would patients with different degrees of severity of the disorder share a main genetic alteration that is modulated by other ones, leading to the clinical variation?

Here, we present a genetic study of a Brazilian family with two AS patients and one with AD conducted to test the above mentioned hypothesis.

Methods

Copy number variations (CNVs) analysis

The CNV screening was performed by a custom-designed oligonucleotide array comparative genomic hybridization (CGH) 60K platform with probes targeted to all exons and 5' and 3' untranslated (UTRs) regions of 269 selected genes (RIBEIRO *et al.*, 2013), by a conventional array CGH 180K platform from Agilent Technologies and by a SNP-array platform from Affymetrix.

In the array CGH experiment, the procedures for DNA digestion, labeling with Cy3- and Cy5-dCTPs by random priming followed by purification, hybridization and washing were performed according to the manufacturer's protocol (Agilent Technologies). Scanned images of the arrays were processed using the Feature Extraction software (Agilent Technologies, Santa Clara, CA), and the analysis was carried out using the Agilent Genomic Workbench 7.0 (Agilent

Technologies, Santa Clara, CA). Identification of aberrant copy number segments was based on the ADM-2 algorithm with default settings (threshold log₂ ratio of 0.2 or 1.14 was used for gain or high copy gain, and -0.23 and -1.14 used for loss and homozygous loss, respectively), and the significance threshold was set at 7. We considered at least three consecutive probes for calling a segment. All ASD samples were tested in duplicate.

Genome-wide single-nucleotide polymorphism (SNP) array analysis was performed on DNA isolated from peripheral blood using a 500K SNP array (Affymetrix Inc., Santa Clara, CA, USA) according to the protocol provided by the manufacturer. For the identification of CNVs, the analysis was made using the softwares Genotyping Console (Affymetrix), PennCNV (WANG et al., 2007; WANG; BUCAN, 2008) and dChip (LI; HUNG-WONG, 2001; LIN et al. 2004) with default settings. We considered only CNVs detected by at least two different softwares.

Whole exome sequencing

Exome sequencing was performed by the team of *The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada*, using Agilent SureSelect Human All Exon 50-Mb V4 (Agilent Technologies) and paired end sequenced on a Illumina HiSeq 2000 instrument (Life Technologies) following the manufacturers' protocols. Raw sequence data was mapped to the reference human genome hg19 (downloaded from UCSC Genome Browser) using BWA version 0.6.5a (LI et al., 2009). The Genome Analysis Tool Kit (GATK) version 1.1.28 (Broad Institute) was used to remove PCR duplicates and to refine the alignments, before SNP and indel calling with GATK default parameters. Variant annotation was carried out using SIFT 4.0.3 (NG; HENIKOFF et al., 2003). Only non-polymorphic mutations that had quality by depth (QD) > 5, strand bias (SB) < -0.01, and were in coding regions causing either the creation of a premature stop codon, the gain or loss of a splice site, or frameshifts were considered as potential candidates.

Clinical History

The patient RBR, an 18-year-old boy, was diagnosed with Asperger syndrome (OMIM 608638) by a team of psychiatrists from the *Instituto de Psiquiatria of the Hospital das Clínicas*, University of São Paulo, São Paulo. The diagnostic was performed according to the Diagnosis and Statistical Manual of Mental Diseases (DSM-IV), International Classification of Diseases 10th edition (ICD-10) and Autism Screening Questionnaire (ASQ) (BERUMENT *et al.*, 1999). RBR is the only child of healthy and non-consanguineous parents, born at full-term with normal neonatal measurements and Apgar scores 9/10, after an uneventful pregnancy. His motor development was normal and no facial dysmorphisms or congenital malformations were observed. RBR presents anxiety, irritability, hyperactivity, severe impairment of social interaction, restricted and repetitive patterns of behavior, interests and activities, marked deficiency in the use of nonverbal behaviors, failure to recognize and to use the conventional rules of conversation, failure to recognize the use of irony, slang, sarcasm and metaphors. RBR has one first (SMRF) and one second (MRPA) degree paternal cousins also carriers of ASD.

SMRF, a 24-year-old boy, was born to healthy and non-consanguineous parents, at full-term after an uneventful pregnancy with normal neonatal measurements and Apgar score 9/10. SMRF has two half sisters; both had normal milestones development and are healthy. SMRF does not present dysmorphic features or malformations and his motor development was normal. He was diagnosed as having Asperger syndrome, according to DSM-IV, ICD-10 and ASQ. SMRF presents severe impairment of social interaction, restricted and repetitive patterns of behavior, interests and activities, marked deficiency in the use of nonverbal behaviors, failure to recognize and to use the conventional rules of conversation, failure to recognize the use of irony, slang, sarcasm and metaphors.

MRPA, a 17-years-old boy, was the first child of healthy and non-consanguineous parents, born at full-term after an uneventful pregnancy with normal neonatal measurements and Apgar score 9/10. His younger brother (VRPA) developed normally and is healthy. Differently from RBR and SMRF, MRPA was diagnosed with autistic disorder (DSM-IV, ICD-10 and ASQ). MRPA presents stereotyped and repetitive motor mannerisms, aggressive behavior and lack of verbal communication.

The pedigree is shown in Figure 5.1. We were unable to contact individual II-2 but his family reported that he was diagnosed as having Asperger syndrome.

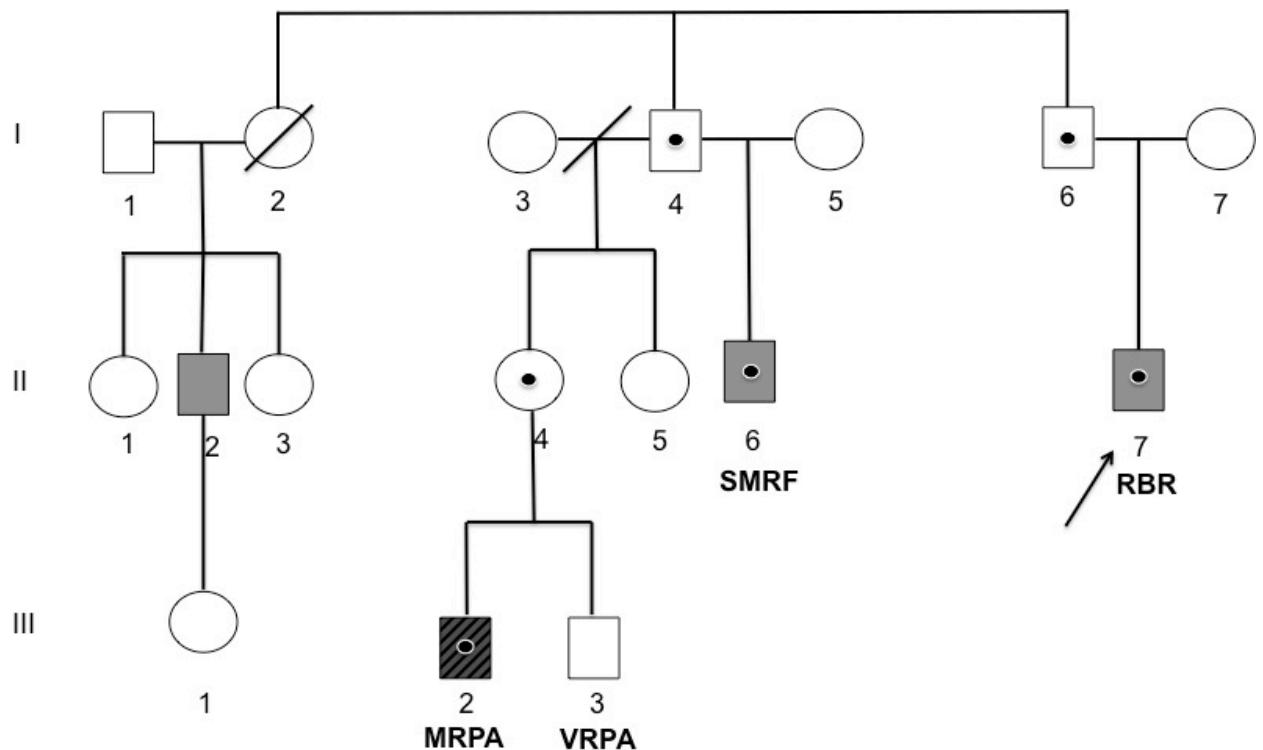


Figure 5.1: Pedigree Family. Black circle represents individuals carrying the *IMMP2L* deletion.

Genetic characterization

A first CNV screening was performed using our custom-designed array CGH in a DNA sample from RBR, which allowed us to detect a deletion of exons 1, 2, 3 and 4 of *IMMP2L* (Transcript ID: ENST0000033176; GRCh37/hg19; www.ensembl.org). This deletion was confirmed through SNP-array and we verified that it has approximately 246 Kbp (SNP_A-4207587 to SNP_A-2254480; GRCh36/hg18) encompassing also the whole 5'UTR. Therefore, it is possible that the deletion also affects the elements necessary for gene transcription. Subsequently, analysis of DNAs from all available individuals of these family by custom-designed array CGH showed that I-4; I-6; II-4, SMRF and MRPA were also carriers of the *IMMP2L* deletion, while VRPA, who is non-affected one, did not harbor this deletion. No other alterations were observed in these analyses (table 5.1).

Next, we have searched for additional CNVs in all affected individuals by another array-CGH platform (180K from Agilent Technologies) and the results revealed that MRPA, in addition to the *IMMP2L* deletion, also has a *de novo* duplication of about 2Mb at 10q11.22 (chr10: 46,169,817-48,334,407; GRCh37/hg19). In patients RBR and SMRF no other CNV was observed by this array platform (table 5.1).

Further, we have performed exome sequencing analysis for RBR, SMRF and MRPA to search for other probably pathogenic alterations. The results showed that these patients share two non-polymorphic mutations located within genes previously reported as candidate to ASD: one nonsense alteration at *SHANK2* and another frameshift alteration at *BZRAP1* (BUCAN *et al.*, 2009; BERKEL *et al.*, 2010; KUMAR *et al.*, 2010; PINTO *et al.*, 2010; BERKEL *et al.*, 2012). *In silico* prediction of the deleterious effects of these alterations did not return information about the one at *BZRAP1* and classified the *SHANK2* mutation as tolerated (SIFT, <http://sift.jcvi.org>; PolyPhen2, <http://genetics.bwh.harvard.edu/pph2/>). Looking at other shared non-polymorphic variants, we found three nonsense mutations, at *DGKH*, *KIF2C* and *PARD3*, classified as probably damaging by SIFT or PolyPhen-2. None of these three genes have been previously associated with ASD. In addition to the shared variants, unshared rare mutations at ASD candidate genes were also found: RBR has a frameshift alteration in *SYNGAP1*; SMRF has one

nonsense (*ANK3*) and one frameshift (*MAP1A*) mutations; and MRPA has two frameshift (*GABRR2* and *LAMC3*) and one nonsense (*LAMC3*) mutations (Table 5.1).

Table 5.1 CNVs and sequence variants identified in ASD patients from family

Patient	Mutations	Gene functions
Shared mutations RBR, SMRF and MRPA.	<i>IMMP2L</i> deletion	<i>IMMP2L</i> encodes a protein involved in processing the signal peptide sequences used to direct mitochondrial proteins to the mitochondria. Immp2l deficiency increases ischemic brain damage by enhancing O ₂ (-) production and damaging mitochondrial functional performance (http://www.ncbi.nlm.nih.gov/gene/83943)
	NM_012309.3(<i>SHANK2</i>): c.435T>A	<i>SHANK2</i> codifies a synaptic scaffold protein (http://www.ncbi.nlm.nih.gov/gene/22941).
	NM_004758.3(<i>BZRAPI</i>):c.4248_4249delAA	<i>BZRAPI</i> codifies an adaptor molecule thought to regulate synaptic transmission (BUCAN <i>et al.</i> , 2009)
	NM_178009.3(<i>DGKH</i>):c.754G>T	<i>DGKH</i> encodes a member of the diacylglycerol kinase (DGK) enzyme family of proteins, specifically the type II DGK subfamily. Members of this family are involved in regulating the intracellular concentrations of diacylglycerol and phosphatidic acid (http://www.ncbi.nlm.nih.gov/gene/160851).
	NM_006845.3(<i>KIF2C</i>):c.1825G>T	The protein encoded by <i>KIF2C</i> is a member of kinesin-like protein family. Proteins of this family are microtubule-dependent molecular motors that transport organelles within cells and move chromosomes during cell division. This protein is important for anaphase chromosome segregation and may be required to coordinate the onset of sister centromere separation (http://www.ncbi.nlm.nih.gov/gene/11004).
	NM_001184794.1(<i>PARD3</i>):c.1136C>A	<i>PARD3</i> encodes a member of the PARD protein family. PARD family members interact with other PARD family members and other proteins; they affect asymmetrical cell division and direct polarized cell growth (http://www.ncbi.nlm.nih.gov/gene/56288).

Unshared mutations	RBR	NM_006772.2(<i>SYNGAPI</i>):c.3955delG	The protein encoded by <i>SYNGAPI</i> is a major component of the postsynaptic density (PSD), a group of proteins found associated with NMDA receptors at synapses (http://www.ncbi.nlm.nih.gov/gene/8831).
	SMRF	NM_020987.3(<i>ANK3</i>):c.4482C>A	Ankyrins are a family of proteins that are believed to link the integral membrane proteins to the underlying spectrin-actin cytoskeleton and play key roles in activities such as cell motility, activation, proliferation, contact, and the maintenance of specialized membrane domains (http://www.ncbi.nlm.nih.gov/gene/288).
		NM_002373.5(<i>MAP1A</i>):c.7152delT	<i>MAP1A</i> encodes a protein that belongs to the microtubule-associated protein family. The proteins of this family are thought to be involved in microtubule assembly, which is an essential step in neurogenesis. Expression of MAP1A is almost exclusively in the brain (http://www.ncbi.nlm.nih.gov/gene?db=gene&cmd=retrieve&dopt=default&list_uids=4133).
MRPA		10q11.22 <i>de novo</i> duplication	Duplicated region encompasses about of 15 genes. Among these, we highlight the <i>GPRIN2</i> , which is highly expressed in the cerebellum and encodes a G protein responsible for the regulation of neurite outgrowth (IIDA; KOZASA, 2004)
		NM_002043.3(<i>GABRR2</i>):c.810_811delCC	The protein encoded by <i>GABRR2</i> is a member of the rho subunit family and is a component of the GABA receptor complex (http://www.ncbi.nlm.nih.gov/gene/2570).
		NM_006059.3(<i>LAMC3</i>):c.534C>G	<i>LAMC3</i> encodes laminin gamma3. Laminins have been implicated in a wide variety of biological processes including cell adhesion, differentiation, migration, signaling, neurite outgrowth and metastasis (http://www.ncbi.nlm.nih.gov/gene/10319).
		NM_006059.3(<i>LAMC3</i>):c.1743delC	

Here we describe an inherited deletion at *IMMP2L* in a patient (RBR) with Asperger syndrome and in his cousins, one also with Asperger Syndrome (SMRF) and another with autistic disorder (MRPA). Genetic variation at *IMMP2L* has previously been associated with Alzheimer's disease, Tourette's syndrome, attention deficit hyperactivity and autism susceptibility (PETEK *et al.*, 2001; ELIA *et al.*, 2010; MAESTRINI *et al.*, 2010; SWAMINATHAN *et al.*, 2010; PATEL *et al.*, 2011). Although we did not detect CNVs at *IMMP2L* in 245 controls from our population, it was reported in Database of Genomic Variants (DGV - IAFRATE *et al.*, 2004; ZHANG *et al.*, 2006), which contains CNVs observed in healthy individuals. Thus, it is possible that the *IMMP2L* deletion segregating in the family contributes to ASD phenotype but it is not enough to cause these disorders. Indeed, the penetrance of the phenotype of the carriers of the *IMMP2L* deletion is about 55.5% (5/9).

MRPA, which is the most severely affected, besides the *IMMP2L* deletion, has a 2Mb *de novo* duplication at 10q11.22. This alteration overlaps multiple CNVs described in DGV but has also been found in patients with intellectual disability and language impairment, milder Zapella variant and autism (ARTUSO *et al.*, 2011; CHILIAN *et al.*, 2013; BARBER *et al.*, 2013). An interesting gene within this duplicated region, is *GPRIN2*. This is highly expressed in cerebellum and encodes a protein responsible for regulation of neurite outgrowth (IIDA; KOZASA, 2004)

RBR, SMRF and MRPA also share five non-polymorphic sequence variants (point mutations and indels): two in genes previously pointed out as candidate for ASD – *SHANK2* and *BZR API*; and three in genes never previously suggested as candidates: *DGKH*, *KIF2C* and *PARD3* (BUCAN *et al.*, 2009; BERKEL *et al.*, 2010; KUMAR *et al.*, 2010; PINTO *et al.*, 2010; BERKEL *et al.*, 2012). Both, *SHANK2* and *BZR API*, codify synaptic proteins (WANG *et al.*, 2000; BERKEL *et al.*, 2010). According *in silico* analysis, *SHANK2* nonsense mutation can be tolerated but we cannot exclude the possibility that it contributes to the phenotype. *DGKH* and *PARD3* were considered as risk genes for bipolar disorder, unipolar depression, adult ADHD and

schizophrenia (WEBER *et al.*, 2011; TAKATA *et al.*, 2011; ZENG *et al.*, 2011; WHALLEY *et al.*, 2012), while *KIF2C* encodes a protein involved in cell division and has not been previously associated with neuropsychiatric disorders (MANEY *et al.*, 1998; MANEY *et al.*, 2001).

We cannot rule out that haploinsufficiency of all these genes contribute to the Asperger phenotype. However, the Asperger patients, still each harbor an unshared mutation predicted as pathogenic in genes whose proteins have neuronal functions: RBR has a nonsense mutation at *SYNGAP1*, previously associated with ASD, which impairs cognitive development by disrupting maturation of dendritic spine synapses (CLEMENT *et al.*, 2012). SMRF has a mutation at *ANK3* and other at *MAP1A*, both previously associated with autism (MAYERS *et al.*, 2011; BI *et al.*, 2012). The first gene encodes a protein that plays key roles in activities such as cell motility, activation, proliferation, contact, and the maintenance of specialized membrane domains; whilst the protein encoded by the latter is involved in microtubule assembly, an essential step in neurogenesis (RefSeq genes, <http://www.ncbi.nlm.nih.gov/gene/>). Analysis of the parents of the Asperger patients is being performed in order to elucidate the origin of each of these mutations. Functional studies in the near future or identification of additional patients with these mutations could help us to elucidate which of them contribute to the phenotype.

Regarding the autistic patient, MRPA, in addition to the six mutations shared with the Asperger cousins and the duplication encompassing about 15 genes, including *GRIP2* as an important ASD candidate, he also has potentially pathogenic mutations in *GABRR2*, that encodes a component of the GABA receptor complex, and *LAMC3*, which encodes a protein involved in several biological processes including cell adhesion, differentiation, migration, signaling, and neurite outgrowth. Alterations in these two genes have been observed in autistic patients (MA *et al.*, 2005; O'ROAK *et al.*, 2011; O'ROAK *et al.*, 2012). It is of note that MRPA harbors two damaging mutations in *LAMC3*. It will be important to verify if these mutations are in *trans*. In this case, there would be loss of function of *LAMC3*, which could be the main cause of the severe phenotype of this patient.

The results obtained up to now lend support to the possibility of several hits leading to the ASD phenotype. In addition, the severity of these disorders seems to be modulated by the number of hits and possibly by hits in both alleles. This is in agreement with previously suggestions that

ASD is caused by several mutations (ITSARA *et al.*, 2010; GIRIRAJAN *et al.*, 2010). As shown in the present study, it is difficult to identify which of the potentially damaging mutations are contributing to the phenotype. In order to better interpret the results, we will test in the near future the presence of these mutations in their non-ASD parents.

Acknowledgements

The authors wish to thank Dr. Sergio Pereira and Dr. Steve Scherer of The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada for assistance with services performed.

References

- AMERICAN PSYCHIATRIC ASSOCIATION. **Diagnostic and statistical manual of mental disorders: DSM-IV**. 4th ed. Washington (DC): American Psychiatric Association. 943 p., 2000.
- ARTUSO, R. et al. Investigation of modifier genes within copy number variations in Rett syndrome. **J Hum Genet**. 56(7):508-15, 2011.
- BARBER, J. C. et al. 16p11.2-p12.2 duplication syndrome; a genomic condition differentiated from euchromatic variation of 16p11.2. **Eur J Hum Genet**. 21(2):182-9, 2013.

BERKEL S, et al. Inherited and de novo SHANK2 variants associated with autism spectrum disorder impair neuronal morphogenesis and physiology. **Hum Mol Genet.** 15;21(2):344-57, 2012.

BERKEL, S. et al. Mutations in SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. **Nature Genetics**, v.42, n.6, p.489-491, 2010.

BERUMENT, S. et al. Autism Screening Questionnaire: Diagnostic Validity. **British Journal of Psychiatry**, v. 175, p.444-451, 1999.

BI, C. et al. Mutations of ANK3 identified by exome sequencing are associated with autism susceptibility. **Hum Mutat.** 33(12):1635-8, 2012.

BUCAN, M. et al. Genome-Wide Analyses of Exonic Copy Number Variants in a Family-Based Study Point to Novel Autism Susceptibility Genes. **PLoS Genetics**, 5: e1000536, 2009.

CHILIAN, B. et al. Dysfunction of SHANK2 and CHRNA7 in a patient with intellectual disability and language impairment supports genetic epistasis of the two loci. **Clin Genet.** 2013.

CLEMENT, J. P. et al. Pathogenic SYNGAP1 mutations impair cognitive development by disrupting maturation of dendritic spine synapses. **Cell.** 9;151(4):709-23, 2012.

GIRIRAJAN, S.; EICHLER, E. E. Phenotypic variability and genetic susceptibility to genomic disorders. **Hum Mol Genet.** 15;19(R2):R176-87, 2010.

IAFRATE, A. J. et al. Detection of large-scale variation in the human genome. **Nat Genet.** 36(9):949-51, 2004.

IIDA, N.; KOZASA, T. Identification and biochemical analysis of GRIN1 and GRIN2. **Methods Enzymol.** 390:475-83, 2004.

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.

- KIM, S. K. et al. Association study between polymorphisms of the PARD3 gene and schizophrenia. **Exp Ther Med.** 3(5):881-885, 2012.
- KORDELI, E.; LAMBERT, S.; BENNETT, V. AnkyrinG. A new ankyrin gene with neural-specific isoforms localized at the axonal initial segment and node of Ranvier. **J Biol Chem.** 270(3):2352-9, 1995.
- KUMAR, R. A, et al. SHANK2 redemption: another synaptic protein for mental retardation and autism. **Clin Genet.** 78(6):519-21, 2010.
- LI, C.; HUNG-WONG, W. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. **Genome Biol.** 2(8):RESEARCH0032, 2001.
- LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics.** 25(14):1754-60, 2009.
- LIN, M. et al. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. **Bioinformatics.** 22;20(8):1233-40, 2004.
- MA, D. Q. et al. Identification of significant association and gene-gene interaction of GABA receptor subunit genes in autism. **Am. J. Hum. Genet.**, 77:377–388, 2005.
- MANEY, T. et al. Association study between polymorphisms of the PARD3 gene and schizophrenia. **J Cell Biol.** 10;142(3):787-801, 1998.
- MANEY, T.; WAGENBACH, M.; WORDEMAN, L. Molecular dissection of the microtubule depolymerizing activity of mitotic centromere-associated kinesin. **J Biol Chem.** 276(14):34753-8, 2001.
- MARSHALL, C. R. et al. Structural variation of chromosomes in autism spectrum disorder. **Am. J. Hum. Genet.** 82:477–488, 2008.
- 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.

- MYERS, R. A. et al. A population genetic approach to mapping neurological disorder genes using deep resequencing. **PLoS Genet.** 7(2):e1001318, 2011.
- NEALE, B. M. et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. **Nature**. 485(7397):242-5, 2012.
- NG, P. C.; HENIKOFF, S. SIFT: Predicting amino acid changes that affect protein function. **Nucleic Acids Res.** 31(13):3812-4, 2003.
- O'ROAK, B. J. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. **Nat Genet**. 43(6):585-9, 2011.
- O'ROAK, B. J. et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. **Science**. 338(6114):1619-22, 2012.
- O'ROAK, B. J. et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. **Nature**. 485(7397):246-50, 2012.
- PINTO, D. et al. Functional impact of global rare copy number variation in autism spectrum disorders. **Nature**. 466(7304):368-72, 2010.
- ROSENBERG, R. E. et al. Trends in autism spectrum disorder diagnoses: 1994-2007. **J Autism Dev Disord**. 39(8):1099-111, 2009.
- SANDERS, S. J. et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. **Nature** 485(7397):237-41, 2012.
- SEBAT, J. et al. Strong association of de novo copy number mutations with autism. **Science**, 316:445–449, 2007.
- TAKATA, A. et al. Nominal association between a polymorphism in DGKH and bipolar disorder detected in a meta-analysis of East Asian case-control samples. **Psychiatry Clin Neurosci**. 65(3):280-5, 2011.
- VEENSTRA-VANDERWEELE J.; COOK, E. H. Jr. Molecular genetics of autism spectrum disorder. **Molecular Psychiatry** 9:819-832, 2004.

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 Res.** 17(11):1665-74, 2007.

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

WANG, Y.; SUGITA, S.; SUDHOFF, T. C. The RIM/NIM family of neuronal C2 domain proteins. Interactions with Rab3 and a new class of Src homology 3 domain proteins. **J Biol Chem** 275: 20033–20044, 2000.

WEBER, H. et al. Cross-disorder analysis of bipolar risk genes: further evidence of DGKH as a risk gene for bipolar disorder, but also unipolar depression and adult ADHD. **Neuropsychopharmacology**. 36(10):2076-85, 2011.

WHALLEY, H. C. Effect of variation in diacylglycerol kinase η (DGKH) gene on brain function in a cohort at familial risk of bipolar disorder. **Neuropsychopharmacology**. 37(4):919-28, 2012.

ZHANG, J. et al. Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. **Cytogenet Genome**. 115(3-4):205-14, 2006.

6.1 Discussão geral

Os TEA possuem alta prevalência e causam prejuízos funcionais graves, sendo atualmente considerados um problema de saúde pública. O estudo de possíveis mecanismos etiológicos pode auxiliar a compreensão desses transtornos e, sob perspectivas terapêuticas, a identificação de genes candidatos pode oferecer evidências de quais vias biológicas estão implicadas de forma significativa, orientando o desenvolvimento de medicamentos que atuem especificamente nessas vias. Adicionalmente, a detecção de alterações gênicas podem assistir o diagnóstico e o aconselhamento genético.

Dessa forma, com o objetivo de identificar genes nos quais alterações podem levar aos TEA e os possíveis mecanismos etiológicos associados, utilizamos as seguintes estratégias: customização de uma lâmina de aCGH que possibilite a detecção não só de grandes CNVs, mas também de alterações menores do que 10 kbp, em *exons* e regiões UTR de genes potencialmente candidatos; a comparação entre os tipos de rearranjos detectados em pacientes sindrômicos e em não sindrômicos e, ainda, a investigação mais detalhada de uma família com indivíduos portadores de transtorno autista e síndrome de Asperger. A tecnologia de microarranjos tem sido amplamente utilizada na detecção de CNVs em pacientes com TEA. Plataformas com maior número de sondas oferecem maior resolução e, em geral, são capazes de detectar alterações maiores ou iguais a 10 Kbp. Entretanto, quanto maior a resolução mais dispendiosa torna-se a técnica. Assim, a abordagem de customizar uma lâmina de aCGH permitiu uma densa cobertura de todos os *exons* e UTRs dos genes alvos com um custo relativamente acessível, o que é importante pensando na utilização dessa lâmina como uma ferramenta que pudesse auxiliar o diagnóstico de TEA na população brasileira.

A sensibilidade da lâmina customizada foi testada em um amostra de 103 portadores de TEA não sindrômicos e em 18 sindrômicos, sendo as taxas de detecção de alterações potencialmente patogênicas, respectivamente, 11,6% e 38,9%. Se avaliarmos conjuntamente pacientes sindrômicos e não sindrômicos, a taxa total de detecção de alterações foi de 15,7%. Comparando com dados da literatura, a taxa de CNVs em autistas, reportada por estudos que avaliaram tanto casos sindrômicos como não sindrômicos, variou de 8-11,6% (BREMER *et al.*, 2011; SORTE *et al.*, 2013) e, em estudos que avaliaram apenas casos sindrômicos, variou de 9-27,5% (JACQUEMONT *et al.*, 2006; QIAO *et al.*, 2009; BREMER *et al.*, 2010). Considerando que diferentemente dos estudos citados, os pacientes que selecionamos foram previamente triados para as CNVs com maior frequência de recorrência entre portadores de TEA (15q11-q13, 16p11.2 e 22q11-13), sendo incluídos em nossa pesquisa apenas aqueles que não apresentam nenhuma destas alterações, a eficiência da nossa lâmina customizada não foi inferior a das lâminas comerciais que possuem um número maior de sondas (85.000 a 2.700.000 sondas – SEBAT *et al.*, 2007; MARSHALL *et al.*, 2008; PINTO *et al.*, 2010; QIAO *et al.*, 2013) e, consequentemente, maior custo.

Os resultados também evidenciaram que o desenho da lâmina, com sobreposição parcial entre as sondas, foi eficaz na visualização de alterações pequenas (<10 kbp). No total, foram detectadas 18 diferentes CNVs possivelmente patogênicas entre os autistas sindrômicos e não sindrômicos, entre essas, 8 (44,4%) são menores do que 10Kb. Assim, sugerimos que a hipótese de que uma parcela dos casos idiopáticos possa ser causada por pequenas CNVs, não detectáveis pelas ferramentas moleculares disponíveis atualmente, parece ser plausível e deve ser testada em uma maior amostra de pacientes e controles.

Cerca de 33,3% das alterações potencialmente patogênicas que identificamos afetam região UTR. Assim, futuramente pretendemos realizar estudos funcionais para melhor investigar o papel de alterações em regiões não codificantes na etiologia dos TEA.

Outra contribuição de nosso estudo é a confirmação de que danos em diferentes genes, cujas proteínas estão envolvidas em uma mesma via ou em um mesmo processo biológico, podem ter o mesmo efeito final e levar a manifestação do fenótipo autista. Das 10 diferentes CNVs detectadas entre os casos não sindrômicos, observamos que 4 afetam genes envolvidos em funções glutamatérgicas e 3 afetam genes envolvidos com orientação axonal. Assim, apesar das alterações

identificadas serem geralmente raras e particulares de cada indivíduo, nossos resultados corroboram que algumas funções biológicas parecem ter maior probabilidade de estarem implicadas na fisiopatologia dos TEA (CUSCÓ *et al.*, 2009; GLESSNER *et al.*, 2009; BUCAN *et al.*, 2009; SBACCHI *et al.*, 2010).

Diferente do que foi observado entre os pacientes não sindrômicos, nos sindrômicos detectamos mais de uma alteração em um mesmo indivíduo ou alterações que englobam mais de um gene, sugerindo o modelo oligogênico para uma parcela dos casos de TEA sindrômico.

CNVs nos genes *SLC17A6*, *SNAP29*, *MARK1* foram recorrentes entre os pacientes não sindrômicos e alterações nos genes *PRKCa* e *SHANK2* foram vistas tanto em sindrômicos como em não sindrômicos. Assim, destacamos a importância de futuramente triar um número maior de portadores de TEA e controles, tanto brasileiros como de outras origens, para deleções e duplicações nesses genes.

Ainda no presente estudo, descrevemos um caso familiar no qual o probando e seu primo em primeiro grau são portadores de síndrome de Asperger e o primo em segundo grau é portador de transtorno autista. Esses compartilham uma deleção no gene *IMMP2L* e mutações de ponto ou InDels não polimórficas em heterozigose e potencialmente patogênicas em pelo menos 5 genes. Esses resultados são consistentes com outros estudos que sugerem que em casos familiais há uma predisposição herdada para TEA porém, esta sozinha não é suficiente para causar esses transtornos visto que os alelos de susceptibilidade são transmitidos pelos pais não afetados (ITSARA *et al.*, 2010; GIRIRAJAN *et al.*, 2010). Portanto, outros fatores são necessários para a manifestação do fenótipo. De fato, mutações potencialmente patogênicas em genes já apontados como candidatos aos TEA, particulares de cada afetado desta família, foram detectadas. O indivíduo portador de transtorno autista, em adição as alterações compartilhadas entre todos os afetados, difere dos seus primos portadores de Asperger pela possibilidade de ser heterozigoto composto para as mutações no gene *LAMC3* e por apresentar uma segunda CNV de aproximadamente 2Mb no cromossomo 10. Esta alteração engloba cerca de 15 genes codificadores de proteínas e, entre esses genes, destacamos o *GPRIN2*, cujo produto proteico regula crescimento de neuritos, função já implicada na etiologia dos TEA (LEPAGNOL-BESTEL *et al.*, 2008; HUSSMAN *et al.*, 2011). Portanto, sugerimos que a gravidade do quadro clínico possa estar relacionada ao número de genes afetados e a relevância funcional desses.

Os nossos resultados, além de corroborar com o envolvimento dos genes *MDGA2*, *FHIT*, *HTR2A*, *SHANK2*, *GRIA3*, *ZNF778*, *PRKCa*, *CDH15*, *DIAPH3*, *GCH1*, *GRM5*, *MARK1*, *SLC17A6*, *IMMP2L*, *BZRAP1*, *SYNGAP1*, *ANK3*, *MAP1A*, *GABRR2* e *LAMC3* nos TEA também sugere novos genes candidatos: *LRRC7*, *LRRIQ3*, *CADPS1*, *NUFIP*, *SEMA3A*, *SNAP29*, *MBD2*, *GAD2*, *DGKH* e *PARD3*.

Por fim, nossos dados são consistentes com os outros estudos que reportam os TEA como distúrbios complexos, com grande heterogeneidade alélica e heterogeneidade de *locus*. Quanto a perspectivas futuras, para uma melhor compreensão dos mecanismos pelos quais as CNVs identificadas podem levar ao fenótipo pretendemos realizar estudos *in vitro* e em *in vivo*.

6.2 Conclusões e Considerações finais

- A triagem de 269 genes candidatos revelou que a taxa de detecção de alterações em pacientes não sindrômicos e sindrômicos foi, respectivamente, 11,6% e 38,9%. A taxa de alterações *de novo* foi maior em sindrômicos (25%) em comparação a taxa observada em não sindrômicos (7,89%);
- A lâmina de aCGH customizada parece ser eficaz na identificação de deleções/duplicações em autistas e poderia ser utilizada como uma ferramenta para auxiliar o diagnóstico;
- Pequenas CNVs (<10 Kbp), não detectáveis pelas plataformas comerciais de *microarray* disponíveis atualmente, em *exons* e regiões UTR, parecem contribuir para o fenótipo de TEA;
- Os genes *LRRC7*, *LRRIQ3*, *CADPS1*, *NUFIP*, *SEMA3A*, *SNAP29*, *MBD2*, *GAD2*, *DGKH* e *PARD3* podem ser novos candidatos aos TEA;
- A gravidade do quadro clínico pode ser modulada pelo número de mutações em genes envolvidos com neurogênese.

- ADZHUBEI, I.; JORDAN, D. M.; SUNYAEV, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. **Curr Protoc Hum Genet.** Chapter 7:Unit7.20, 2013.
- AMERICAN PSYCHIATRIC ASSOCIATION. **Diagnostic and statistical manual of mental disorders: DSM-IV.** 4th ed. Washington (DC): American Psychiatric Association. 943 p., 2000.
- BAILEY, A. et al. Autism as a strongly genetic disorder: evidence from a British twin study. **Psychol Med.**, 25(1):63-77, 1995.
- BEN-DAVID, E.; SHIFMAN. S. Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. **Mol Psychiatry.** 2012.
- BERUMENT, S. et al. Autism Screening Questionnaire: Diagnostic Validity. **British Journal of Psychiatry**, v. 175, p.444-451, 1999.
- BLANCHETTE, M. et al. Aligning multiple genomic sequences with the threaded blockset aligner. **Genome Res.** 14(4):708-15, 2004.
- BLATT, G. J. et al. Density and distribution of hippocampal neurotransmitter receptors in autism: an autoradiographic study. **J. Autism Dev. Disord.**, 31:537–543, 2001.
- BLATT, G. J. et al. GABAergic cerebellar system in autism: a neuropathological and developmental perspective. **Int. Rev. Neurobiol.**, 71:167–178, 2005.
- BREMER, A. et al. Copy number variation characteristics in subpopulations of patients with autism spectrum disorders. **Am J Med Genet B Neuropsychiatr Genet.** 156(2):115-24, 2011
- BREMER, A. et al. Screening for copy number alterations in loci associated with autism spectrum disorders by two-color multiplex ligation-dependent probe amplification. **Am J Med Genet B Neuropsychiatr Genet.** 5;153B(1):280-5, 2010.
- BUCAN, M. et al. Genome-Wide Analyses of Exonic Copy Number Variants in a Family-Based Study Point to Novel Autism Susceptibility Genes. **PloS Genetics**, 5: e1000536, 2009.
- CARTER, N. P. Methods and strategies for analyzing copy number variation using DNA microarrays. **Nat Genet.** 39(7 Suppl):S16-21, 2007.

- CHENG, Y.; MIURA, R. M.; TIAN, B. Prediction of mRNA polyadenylation sites by support vector machine. **Bioinformatics**. 1;22(19):2320-5, 2006.
- CHIAROMONTE, F.; YAP, V. B.; MILLER, W. Scoring pairwise genomic sequence alignments. **Pac Symp Biocomput**. 115-26, 2002.
- CHRISTIAN, S. L. et al. Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. **Biol Psychiatry**, 63:1111–1117, 2008.
- CUSCO, I. et al. Autism-specific copy number variants further implicate the phosphatidylinositol signaling pathway and the glutamatergic synapse in the etiology of the disorder. **Hum. Mol. Genet**. 18, 1795-1804, 2009.
- DEVLIN, B.; SCHERER, S. W. Genetic architecture in autism spectrum disorder. **Curr Opin Genet Dev**. 22(3):229-37, 2012.
- EL-FISHAWY, P.; STATE, M. W. The genetics of autism: key issues, recent findings, and clinical implications. **Psychiatr Clin North Am**. 33(1):83-105, 2010.
- FATEMI, S. H.; FOLSOM, T. D. Dysregulation of fragile × mental retardation protein and metabotropic glutamate receptor 5 in superior frontal cortex of individuals with autism: a postmortem brain study. **Mol Autism**. 6;2:6, 2011.
- FOLSTEIN, S. E.; ROSEN-SHEIDLEY, B. Genetics of autism: complex a etiology for a heterogeneous disorder. **Nat. Rev. Genet.**, 2:943–955, 2001.
- FRIEDMAN, R. C. et al. Most Mammalian mRNAs Are Conserved Targets of MicroRNAs. **Genome Research**. 19:92-105, 2009.
- GIRIRAJAN, S., EICHLER, E. E. Phenotypic variability and genetic susceptibility to genomic disorders. **Hum Mol Genet**. 15;19(R2):R176-87, 2010.
- GLESSNER, J. T. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. **Nature**. 459, 569-573, 2009.
- GOIZET, C. et al. Case with autistic syndrome and chromosome 22q13.3 deletion detected by FISH. **Am. J. Med. Genet.**, 96: 839–844, 2000.
- GRIMSON, A. et al. MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. **Molecular Cell**. 6;27(1):91-105, 2007.
- HADDAD, L. A. et al. A PCR-based test suitable for screening for fragile X syndrome among mentally retarded males. **Hum Genet**. 97(6):808-12, 1996.
- HUSSMAN, J. P. et al. A noise-reduction GWAS analysis implicates altered regulation of neurite outgrowth and guidance in autism. **Mol Autism**. 19;2(1):1, 2011.

- IAFRATE, A. J. et al. Detection of large-scale variation in the human genome. **Nat Genet**. 36(9):949-51, 2004.
- IOSSIFOV, I. et al. De novo gene disruptions in children on the autistic spectrum. **Neuron**. 26;74(2):285-99, 2012.
- ITSARA, A. De novo rates and selection of large copy number variation. **Genome Res**. 20(11):1469-81, 2010.
- 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.
- JAMAIN, S. et al. Linkage and association of the glutamate receptor 6 gene with autism. **Mol. Psychiatry** 7:302–310, 2002.
- KENT, W. J. et al. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. **Proc Natl Acad Sci U S A.**; 100(20):11484-9, 2003.
- KEN, W. J. et al. The human genome browser at UCSC. **Genome Res**. 12(6):996-1006, 2002.
- KUMAR, R. A. et al. SHANK2 redemption: another synaptic protein for mental retardation and autism. **Clin Genet**. 78(6):519-21, 2010.
- LAUMONNIER, F. et al. X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. **Am J Hum Genet**. 74(3):552-7, 2004.
- LEE, C. et al. Copy number variation and clinical cytogenetic diagnosis of constitutional disorders. **Nat. Genet.**, 39:S 48-54, 2007.
- LEPAGNOL-BESTEL, A. M. SLC25A12 expression is associated with neurite outgrowth and is upregulated in the prefrontal cortex of autistic subjects. **Mol Psychiatry**. 13(4):385-97, 2008.
- LEWIS, B. P.; BURGE, C. B.; BARTEL, D. P. Conserved seed pairing, often flanked by adenines, indicates that thousands of human genes are microRNA targets. **Cell**. 14;120(1):15-20, 2005.
- MA, D. Q. et al. Identification of significant association and gene-gene interaction of GABA receptor subunit genes in autism. **Am. J. Hum. Genet.**, 77:377–388, 2005.
- MARSHALL, C. R. et al. Structural variation of chromosomes in autism spectrum disorder. **Am. J. Hum. Genet**. 82:477–488, 2008.
- MERIKANGAS, A. K.; CORVIN, A. P.; GALLAGHER, L. Copy-number variants in neurodevelopmental disorders: promises and challenges. **Trends Genet**.

- Dec;25(12):536-44, 2009.
- MEYER, L. R. et al. The UCSC Genome Browser database: extensions and updates. **Nucleic Acids Res.** 2012 Nov 15, 2013.
- MOESSNER, R. et al. Contribution of SHANK3 mutations to autism spectrum disorder. **Am. J. Hum. Genet.** 81(6):1289-97, 2007.
- MORROW, E. M. et al. Identifying autism loci and genes by tracing recent shared ancestry. **Science**. 11;321(5886):218-23, 2008.
- MURPHY, W. J. et al. Resolution of the early placental mammal radiation using Bayesian phylogenetics. **Science**. 2001 Dec 14;294(5550):2348-51.
- NEALE, B. M. et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. **Nature**. 4;485(7397):242-5, 2012.
- NG, P. C.; HENIKOFF, S. SIFT: Predicting amino acid changes that affect protein function. **Nucleic Acids Res.** 1;31(13):3812-4, 2003.
- O'ROAK, B. J. et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. **Science**. 21;338(6114):1619-22, 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.
- PAULA, C. S. Autism in Brazil: perspectives from science and society. **Rev Assoc Med Bras.** 57(1):2-5, 2011.
- PENA, S. D. J. **Enfoque genético da criança com deficiência mental e sua família.** Deficiencia Mental: Abrodagem Multidisciplinar. Fascículo 4. 2005.
- PINTO, D. et al. Functional impact of global rare copy number variation in autism spectrum disorders. **Nature**. 15;466(7304):368-72, 2010.
- POLLARD, K. S.; HUBISZ, M. J.; SIEPEL, A. Detection of non-neutral substitution rates on mammalian phylogenies. **Genome Res.** 2009.
- RONCHEL, F. et al. Real-time PCR based on SYBR-Green I fluorescence: An alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions. **BMC Biotechnology** 3:18, 2003.
- QIAO, Y. et al. Phenomic determinants of genomic variation in autism spectrum disorders. **J Med Genet.** 46(10):680-8, 2009.
- 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.** 2013
- PRUITT, K. D. et al. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. **Nucleic Acids Res.** 40(Database issue):D130-

5, 2012.

- SANDERS, S. J. et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. **Nature** 485(7397):237-41, 2012.
- 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.
- SCHAFF, C. P. et al. Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorder. **Human Molecular Genetics**, v.20, n.17, p.3366-3375, 2011.
- SCHOPLER, E. et al. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). **Journal of Autism and Developmental Disorders**, v. 10, n. 1, p. 91-103, 1980.
- SCHMITZ, C. et al. Autism: neuropathology, alterations of the GABAergic system, and animal models. **Int Rev Neurobiol.**, 71:1-26. 2005.
- SCHOPLER, E. et al. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). **Journal of Autism and Developmental Disorders**, v. 10, n. 1, p. 91-103, 1980.
- SCHWARTZ, S. et al. Human-mouse alignments with BLASTZ. **Genome Res.** 13(1):103-7, 2003.
- SEBAT, J. et al. Strong association of de novo copy number mutations with autism. **Science**, 316:445–449, 2007.
- SIEPEL, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. **Genome Res.** 15(8):1034-50, 2005.
- SORTE, H. S. et al. Copy number variation findings among 50 children and adolescents with autism spectrum disorder. **Psychiatr Genet.** 23(2):61-9, 2013.
- STANKIEWICZ, P.; LUPSKI, J. R. Structural variation in the human genome and its role in disease. **Annu. Rev. Med.**, 61:437-55, 2010.
- TIDMARSH, M. D.; VOLKMAR, F. R. Diagnosis and epidemiology of autism spectrum disorders. **Can J Psychiatry**, 48:517–525, 2003.
- VEENSTRA-VANDERWEELE J.; COOK, E. H. Jr. Molecular genetics of autism spectrum disorder. **Molecular Psychiatry**. 9:819-832, 2004.
- VORSTMAN, J. A. et al. A double hit implicates DIAPH3 as an autism risk gene. **Mol Psychiatry**. 16(4):442-51, 2011.

- 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 Res.** 17(11):1665-74, 2007.
- WANG, K.; 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, 667–675, 2008.
- WEKSBERG, R. et al. A method for accurate detection of genomic microdeletions using real-time quantitative PCR. **BMC Genomics**. 13;6:180, 2005.
- ZHANG, H.; HU, J.; RECCE, M.; TIAN B. PolyA_DB: a database for mammalian mRNA polyadenylation. **Nucleic Acids Res.** 1;33(Database issue):D116-120, 2005.
- ZHANG, J. et al. Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. **Cytogenet Genome**. 115(3-4):205-14, 2006.
- ZHI, J.; HATCHWELL, E. Human MLPA Probe Design (H-MAPD): a probe design tool for both electrophoresis-based and bead-coupled human Multiplex ligation-dependent probe amplification assays. **BMC Genomics**, v.9, n. 407, 2008.