Claudia Ismania Samogy Costa

Variações no número de cópias (CNVs) em pacientes brasileiros com transtorno do espectro autista (TEA)

Copy number variations (CNVs) in Brazilian patients with autism spectrum disorder (ASD)

São Paulo

2018

Claudia Ismania Samogy Costa

Variações no número de cópias (CNVs) em pacientes brasileiros com transtorno do espectro autista (TEA)

Copy number variations (CNVs) in Brazilian patients with autism spectrum disorder (ASD)

Versão corrigida da dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo, para obtenção do Título de Mestre em Ciências, na Área de Biologia/Genética. A versão original encontra-se na biblioteca deste mesmo instituto.

Orientadora: Maria Rita dos Santos e Passos-Bueno

São Paulo

2018

Samogy Costa, Claudia Ismania

Variações no número de cópias (CNVs) em pacientes brasileiros com transtorno do espectro autista (TEA)

86 páginas

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

1. Transtorno do Espectro Autista 2. Variação no número de cópias 3. *array*-CGH 4. Síndrome de Phelan-McDermid I. Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva.

Comissão Julgadora:

Profa. Dra. Angela Maria Morgante

Profa. Dra. Ana Beatriz Alvarez Perez

Dr. Guilherme Lopes Yamamoto

Profa. Dra. Maria Rita dos Santos e Passos Bueno

Orientadora

Dedicatória

Aos meus pais e irmã

Agradecimentos

Aos meus pais e irmã, que são e sempre serão minha base e razão de viver. É a eles que dedico todas as vitórias conquistadas e todas as que virão.

À professora Dra. Maria Rita dos Santos e Passos-Bueno, a quem tenho grande admiração, pela oportunidade, confiança, motivação e por todos ensinamentos.

Ao “Lab 200” e todos os seus membros pela amizade, cumplicidade, carinho, aventuras, ensinamentos e pelo apoio em diversos momentos da minha vida dentro e fora do laboratório. Em especial, agradeço à Elisa Varella Branco por ser a melhor dupla que eu poderia desejar e à Eduarda Morgana Montenegro, Isabela Mayá Wayhs Silva, Ana Cristina Sanctis Girardi e Gerson Shigeru Kobayashi pelas contribuições e orientações nas diferentes etapas do desenvolvimento do meu projeto.

À todos os funcionários do IB-USP e do Centro de pesquisa sobre o genoma humano e célula tronco. Especialmente às funcionárias Silvia Souza da Costa e Naila Cristina Vilaça Lourenço, que foram mestres, amigas e parceiras em todos os momentos.

À professora Eliana Maria Beluzzo Dessen, que desde o meu primeiro ano de graduação vem me acompanhando e servindo de exemplo.

À Associação de parentes e amigos da síndrome de Phelan-McDermid, à equipe da Universidade Federal de São Paulo, Dra. Ana Beatriz Alvarez Perez, Dr. Ricardo Henrique Almeida Barbosa e Dr. Rodrigo Ambrósio Fock, ao doutorando em bioinformática Frederico Monfardini e às professoras Dra. Carla Rosenberg e Dra. Ana Cristina Victorino Krepischi por todas as contribuições.

Aos meus amigos, em especial Giovanna Pontillo Guimarães, Phelipe Eduardo Silva e Eduardo Tassoni Tsuchida, que foram essenciais na minha trajetória, me dando força, carinho e motivação em todos os momentos.

Este trabalho contou com apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

LISTA DE SIGLAS

aCGH Microarray-based comparative genomic hybridization

ADHD attention-deficit/ hyperactivity disorder

ASD Autism Spectrum Disorder

CARS Autism Assessment Scale in Childhood

CEGH-CEL Centro de Estudos sobre o Genoma Humano e Células-Tronco CMA chromosomal microarray analysis

CNVs Copy number variations

DGV Database of Genomic Variants

DLRS Derivative log ratio spread

DSM Diagnostic and Statistical Manual of Mental Disorders

FISH Fluorescence in situ hybridization

FoSTeS Fork Stalling and Template Switching

GWAS Genome-wide association study

IB-USP Instituto de Biociências da Universidade de São Paulo ID Intellectual disability

InDels Small insertions and deletions

HD high-quality

MLPA Multiplex ligation-dependent probe amplification

NAHR Nonalelic Homologous Recombination

NDDs Neurodevelopmental disorders

NEHJ Non-homologous End Joining

NGS New Generation Sequencing

PMS Phelan-McDermid Syndrome

qPCR Quantitative polymerase chain reaction

SFARI Simons Foundation Autism Research Initiative

SNP-array Single nucleotide polymorphism array-based

SNVs Single nucleotide variations

UNIFESP Universidade federal de São Paulo

UTR Untranslated region

VOUs Variants of uncertain significance

Sumário

Summary

Chapter 1

General introduction: literature review

Autism Spectrum Disorder (ASD): definition, epidemiology and diagnosis

Autism spectrum disorder (ASD) is a heterogeneous group of diseases that affect neurodevelopment. According to the *Diagnostic and Statistical Manual of Mental Disorders* - 5th edition (DSM-5), patients with ASD are characterized by the presence of stereotyped behavior and restricted interest, as well as persistent interaction and social communication issues. These symptoms are present in patients since early childhood and impair or limit their daily functioning. Moreover, ASD presents a wide spectrum of clinical variability, in which the symptoms have a continuous distribution among patients (American Psychiatric Association 2013).

It is estimated that ASD affects about 1% of the world population and is four times more prevalent among men than women (Elsabbagh et al. 2012). However, this frequency may vary among different population studies. For instance, in the Brazilian population, a lower frequency of 0.27% was estimated, which was attributed to the adopted methodology, awareness of both population and health professionals, and sample size (Paula et al., 2011).

ASD diagnosis is usually established within the first three years of age and is mainly based on behavioral observations, as defined by the DSM-5 (American Psychiatric Association, 2013). In addition to that, some tools such as questionnaires answered by the caregivers and structured protocol of patient observation also aid in the diagnostic process (Elder et al. 2017; Fakhoury 2015). Diagnosis may be hampered by different factors, such as occurrence of comorbidities and overlapping of symptoms with other conditions. It is estimated that in 70% of the cases, ASD occurs in association with comorbidities, such as epilepsy, intellectual development delay, motor control problems, hyperactivity, attention deficit, anxiety, depression and gastrointestinal problems (Lai et al. 2014).

Furthermore, in 10% of the ASD cases, the disorder occurs as a secondary condition associated with a known genetic syndrome, such as X-fragile syndrome, Phelan-McDermid syndrome, tuberous sclerosis and Rett syndrome (Sztainberg and Zoghbi 2016). Finally, gender, cognitive ability, age and some other factors may affect the diagnosis (Lai et al. 2014). Thus, the complexity involved in the ASD diagnostic process supports the importance of developing new tools to improve this process.

Genetics of ASD

The ASD etiology exhibits a strong genetic contribution based on epidemiological and genomic data and twin studies (Devlin and Scherer 2012; Bourgeron 2015; Murphy et al. 2016; Shishido et al. 2014). Studies with monozygotic and dizygotic twins showed a higher concordance for monozygotic twins when compared to dizygotic, ranging from 60-90% and 0-30%, respectively, and the heritability has been estimated at between 38% and 90%, according to different studies (Bourgeron 2015).

The genetic architecture of ASD is complex and heterogeneous. In about 10-25% of the cases, ASD is associated with a single mutation (which might be either a copy number variation (CNV), a single nucleotide variant (SNV) or a chromosome rearrangement), while in the remaining cases, it can be explained by the combination of different genetic alterations, following an oligogenic or multifactorial inheritance model (Bourgeron 2015). The contribution of common single nucleotide polymorphisms (SNPs) has been estimated to be involved in 17% to 60% of the cases (Klei et al. 2012; Cross-Disorder Group of the Psychiatric Genomics Consortium et al. 2013; Gaugler et al. 2014). However, determining the causality of these variants to ASD remains a challenge, since they are numerous, and each variant corresponds to low risk and then, research conducted has mainly focused on identifying rare, *de novo* and high penetrant mutations.

Genomic studies have shown that, with the exception of some chromosomal regions and genes that are recurrently altered, such as duplications and deletions in 15q11-q13 and 16p11 or SNV and CNVs in *CDH8,* the

vast majority of these mutations are specific to each patient (Moreira et al. 2014; Bourgeron 2015).

Proteomic, transcriptomic and functional studies using cellular and animal models showed that ASD candidate genes converge to a few pathways, including synaptic function and plasticity, neuronal development and axon guidance, action dynamics, chromatin remodeling, protein transcription and translation, neurodegeneration, immunity and microglia activation (Gamazon and Stranger, 2015). Moreover, different neurodevelopmental disorders, such as schizophrenia, bipolar disorder, intellectual disability, epilepsy and depression, share many of the candidate genes and chromosomal regions with those identified in ASD (Grayton et al., 2012).

Efforts have been joined to create databases with information on genes that are associated with ASD. One example is the SFARI (Simons Foundation Autism Research Initiative, https://www.sfari.org/) database, which includes nearly 800 autism associated genes (Abrahams et al. 2013). These genes are classified into 5 groups, according to their association with ASD: 1) highly confident, 2) strong candidate, 3) suggestive evidence, 4) minimal evidence and 5) hypothesized association. Among all the genes listed on SFARI, only 89 of them fall into categories 1 and 2, while most of the genes (650) fall into categories 3, 4 and 5. Therefore, despite the large number of ASD candidate genes, their association with phenotype remains unclear, at least for many of them, which reinforces the importance for identifying genetic alterations in ASD patients.

Copy number variations (CNVs)

CNVs correspond to deletions or duplications of DNA segments, which can range from a few dozens of nucleotides to changes with some mega-bases in size (Grayton et al. 2012; Li and Olivier 2013). Together with Indels (deletions or doubling of a small number of nucleotide bases) and inversions, CNVs are part of the group of structural variations that are present in the human genome.

CNVs are originate from different mechanisms: a) non-allelic homologous recombination (NAHR), b) non-homologous end joining and microhomology-mediated end joining, c) mechanisms mediated by retrotransposon elements, and d) error-mediated mechanisms in replication of DNA, such as FoSTeS (fork stalling and template switching) and microhomology-mediated break-induced replication (MMBIR) (Li and Olivier 2013). CNVs derived from NAHR represent the group of alterations that is better characterized. This mechanism is recognized as being causative of recurrent rearrangements in hotspot regions in the human genome. In turn, replication mechanisms are recognized as being the major contributors for originating non-recurrent CNVs (Liu et al. 2012).

It is estimated that up to 15% of the human genome have copy number variations (Li and Olivier 2013; Grayton et al. 2012). In addition to that, Li and Olivier (2013) showed an inversely proportional relation between CNV size and number in the human genome: the smaller the CNVs the higher their frequency. Moreover, it has been shown that the structural variations in the human genome is not always distributed at random and higher concentration sites (hotspots) of CNVs exist. Some of the gene categories displaying the highest concentration of structural variations are: a) genes associated with immunity and cell signaling, b) genes encoding proteins associated with retroviruses and transposons, and c) genes encoding proteins that are associated with the interaction of the organism with the environment. In turn, some regions of the genome have a much smaller amount of structural variation, such as regions with genes that are sensitive to copy number variations. In such genes, the distribution of CNVs tends to be higher in introns and untranslated regions than in exons that encode proteins (Li and Olivier 2013).

Regarding the genotype-phenotype relationship, CNVs may be associated with either phenotypic variability in the normal population, evolutionary advantage or development of diseases. For example, abnormal high enrichment in the number of copies of the human gene participating in the immune response, *CCL3L1*, has been associated with decreased susceptibility to HIV / AIDS (Gamazon and Stranger 2015; Aklillu et al. 2013). In turn, CNVs may also be

associated with the development of diseases in different ways: a) differences in the dosage of a single gene or a group of genes, b) rupture of the coding and / or regulatory sequence of a gene, c) deletions or duplications of exons, introns or regulatory elements of a gene, (d) gene fusion, (e) chromatin position effects, (f) unmasking of pathogenic alleles associated with recessive diseases or functional polymorphisms, or epigenetic effects (Henrichsen et al. 2009; Zhang et al. 2009). As a result of that, both alterations in gene expression and altered protein production or function may occur (Case-Smith and Arbesman 2008).

Currently, CNV can be detected by using different techniques, depending on the study's goals. For detecting CNVs in previously established candidate disease regions, techniques, such as FISH and quantitative PCR, may be applied (Li and Olivier 2013). However, such techniques have the disadvantage of not being able to be applied for detecting CNVs in large scale and discovering new CNVs. These limitations were overcome by the development of chromosomal microarray analysis (CMA) techniques and, more recently, the use of the next generation sequencing (NGS) technique. Microarray platforms of single-base polymorphisms (SNP-array) and comparative genomic hybridization by microarray (aCGH) have been the most used CMA platforms, and they are seen as the gold standard in CNV detection methodology for diagnostic purposes.

The CMA technique consists of labeling genomic DNA samples from a patient and a control by using fluorophores and then hybridizing them in a slide with preselected probes. The fluorescence intensity emitted by both hybridized DNA samples is measured and then compared, allowing the detection of duplications and / or deletions in the patient's genetic material. Although it is a cost-effective method of analyzing the data generated, it presents some limitations, such as failure to detect breakage points of changes and balanced rearrangements, and problems with hybridization noise, which can generate high rates of false positives and false negatives (Jaluria et al. 2007).

Regarding the use of DNA sequencing techniques, in addition to presenting a higher resolution, the technique also allows the detection of inversions and translocations. However, this technique still presents computational challenges (such as cost and alignment of reads in repetitive

regions) and problems with representing regions that are poorly amplified in the PCR step (Zhao et al. 2013). Thus, although replacing CMA by the NGS technique looked promising, CMA still is one of the most widely used methods for detecting CNVs, and it is regarded by the American College of Medical Genetics (ACMG) as the gold standard in diagnosing patients with developmental and congenital abnormalities (Miller et al. 2010).

Copy number variations in ASD

The frequency of CNVs in ASD patient cohorts can vary according to the origin (*de novo* or inherited) and CNV classification (pathogenic, of uncertain significance or benign), or according to the frequency of sporadic *vs.* familial cases, and syndromic *vs.* non-syndromic cases. For example, the prevalence of CNVs in syndromic patients is higher than in non-syndromic patients (Wiśniowiecka-Kowalnik et al. 2013), and the detection rate of potentially pathogenic CNVs in syndromic patients may go up to 27.5% (Jacquemont et al. 2006).

Both *de novo* and inherited CNVs may contribute to ASD (Leppa et al. 2016; Reis et al. 2017). It is estimated that roughly 5% to 15% of ASD patients carry *de novo* CNVs, whereas for the general population, this estimate includes 1% to 2% of individuals (Pinto et al. 2014). Furthermore, in a study of 411 families of sporadic patients, Krumm et al. (2013) demonstrated that, for inherited small CNVs (<50Kb), there is maternal transmission deviation.

In most cases, CNVs are unique among patients. However, some chromosomal regions have recurrent events of duplications and deletions in the number of copies, as it occurs in the regions 15q11-q13, 16p11 and 22q11.2, found in about 3% to 5% of ASD patients. These CNVs may be *de novo* or inherited from asymptomatic parents (Moreira et al. 2014). Furthermore, other regions have been associated with ASD, such as 1q21, 5p15.2, 7q11.23 and 17p11.2 (Shishido et al. 2014). Among these CNVs, the ones in the 22q13.3 region are associated with a syndromic phenotype (corresponding to the Phelan- McDermid syndrome), which will be detailed below.

The functional impact of CNVs associated to ASD generally converges to common molecular functions, involving cell adhesion pathways, neuronal orientation and growth, neurotransmission, cell signaling and transcription, protein translation and degradation, as well as chromatin remodeling and cytoskeleton. These functions may ultimately affect the development, plasticity, and functioning of synaptic networks and neuronal connectivity in the brain (Bourgeron 2015; Krumm et al. 2013; Sbacchi et al. 2010).

Syndromic ASD: Structural changes in 22q13.3 and the Phelan-McDermid Syndrome

Phelan-McDermid Syndrome (PMS) is among the cases of syndromic ASD involving structural changes. This rare genetic syndrome, also known as 22q13.3 deletion syndrome (OMIM 606232), is characterized by developmental delay, intellectual deficit, speech delay (or absence), hypotonia, autism and facial dimorphisms, among other symptoms (Soorya et al., 2013). By 2015, about 600 PMS cases had been described in the literature (Harony-Nicolas et al. 2015) and about 1500 cases had been identified by the Phelan-McDermid Syndrome Foundation (http://www.pmsf.org/).

There is a broad clinical variability among PMS patients and the prevalence of each trait can vary widely, depending on the study (Sarasua et al., 2011, Soorya et al., 2013, Tabet et al, 2017). PMS is primarily caused by changes (duplications, deletions, point mutations or translocations) in the long arm of chromosome 22, which involves the *SHANK3* gene (Cochoy et al. 2015). *SHANK3* is a gene expressed in developing neurons in the cerebral cortex and cerebellum and encodes a protein that is associated with transmission and synaptic plasticity. More specifically, *SHANK3* participates in formation, maturation and synaptic maintenance (Uchino and Waga 2013).

The genetic changes can range from a few hundred kilobases to 9Mb and, although mutations in *SHANK3* seem to be mainly responsible for the neurological symptoms seen in this syndrome, they cannot fully explain PMS clinical variability (Tabet et al. 2017). In this context, efforts have been directed to determine the genetic causes of clinical heterogeneity and the possible

contribution of other genes to this variability. Both size and position of the changes found in PMS patients have been correlated with certain characteristics, such as gastric problems, convulsions and absence of speech (Tabet et al., 2017). Mitz et al. (2018) conducted a study to determine the possible contribution of other genes present in the 22q13.33 region, which are commonly affected by the changes in this region. The authors found different groups of genes that present functions that may be relevant for the development of the syndrome, such as genes associated with cerebellar development and organization, genes associated with other neuropsychiatric diseases, and genes associated with the glutamatergic synaptic signaling pathway.

Regarding ASD, the percentage of patients with PMS who have autism is still debatable, with penetrance ranging from 0% to 94% depending on the work considered (Oberman et al., 2015). Such variations can be explained by both the cohort characteristics and the analysis methodology chosen.

In a study conducted by Oberman et al. (2015), forty patients were evaluated, and, among other characteristics, they investigated the correlation between the size of the present alteration in the patient and ASD diagnosis. As reported before by Sarasua (2011), they also observed an inversely proportional correlation between the size of the genetic alterations found and the severity of autism clinical condition. The smaller the deletion the more severe the condition of restricted and repetitive behavior. In turn, larger alterations correlated positively with greater loss of general adaptive behavior.

So far, there is no characterization of Brazilian patients with PMS in the literature. Since this syndrome is one of the most common among ASD patients (ranging from 0.5% to 2%) (Kolevzon et al. 2014), understanding the genotype-phenotype correlations and the molecular mechanisms underlying PMS may help to uncover the ASD etiology.

Objectives

From the first description of autism in 1943 by Dr. Kanner (Kanner 1968) to the present day, both the definition and the worldwide prevalence of ASD have changed. In general, the criteria grew increasingly broad, until we reached the set of conditions we know today as autism spectrum disorder. Despite the recognition of ASD contributing genetic factors and the large number of genes that have been associated with the disorder, about 75% of the cases of autism remain without etiological explanation (Fernandez and Scherer 2017).

Following the advancement in genetic discoveries associated with ASD, the contribution of CNVs to this disorder's development has been increasingly recognized. However, the causal relationship between CNVs and ASD remains elusive and the clinical significance of an expressive number of CNVs is still uncertain. In addition to that, studies with different populations have found different candidate genes (Hnoonual et al. 2017). In this context, the study of CNVs in the Brazilian case series may add to understanding ASD etiology, since few Brazilian patients were characterized.

Many of the ASD associated syndromes may present a very large phenotypic variability, both concerning the number and type of comorbidities and the intensity of each clinical sign each patient presents. One of the most prevalent is Phelan-McDermid Syndrome, in which the characterization of the spectrum of clinical variability and correlation between genotype and phenotype are still ongoing. We had the opportunity to study 34 individuals with Phelan-McDermid Syndrome, which represents the second largest study with this syndrome so far. We expect that the analysis of these Brazilian patients will add to better understanding the syndrome and its relationship with ASD in our population.

In this context, this work had the following objectives:

1) Further validate and implement the use of a previously customized array-CGH 180K platform (Silva, 2017) to detect CNVs in ASD patients;
2) To help in better defining the status of previous ASD candidate genes;
3) To identify novel ASD candidate genes;
4) To add to a better clinical and molecular characterization of the PMS;
5) To help disseminate the relevance of PMS across the country.

References

Abrahams, B.S., Arking, D.E., Campbell, D.B., Mefford, H.C., Morrow, E.M., Weiss, L.A., Menashe, I., Wadkins, T., Banerjee-Basu, S. and Packer, A. 2013. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Molecular autism* 4(1), p. 36.

Aklillu, E., Odenthal-Hesse, L., Bowdrey, J., Habtewold, A., Ngaimisi, E., Yimer, G., Amogne, W., Mugusi, S., Minzi, O., Makonnen, E., Janabi, M., Mugusi, F., Aderaye, G., Hardwick, R., Fu, B., Viskaduraki, M., Yang, F. and Hollox, E.J. 2013. CCL3L1 copy number, HIV load, and immune reconstitution in sub-Saharan Africans. *BMC Infectious Diseases* 13, p. 536.

American Psychiatric Association 2013. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Washington, DC: American Psychiatric Association.

Bourgeron, T. 2015. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nature Reviews. Neuroscience* 16(9), pp. 551–563.

Case-Smith, J. and Arbesman, M. 2008. *Evidence-based review of interventions for autism used in or of relevance to occupational therapy.* The American Journal of Occupational Therapy 62(4), pp. 416–429.

Cochoy, D.M., Kolevzon, A., Kajiwara, Y., Schoen, M., Pascual-Lucas, M., Lurie, S., Buxbaum, J.D., Boeckers, T.M. and Schmeisser, M.J. 2015. Phenotypic and functional analysis of SHANK3 stop mutations identified in individuals with ASD and/or ID. *Molecular autism* 6, p. 23.

Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee, S.H., Ripke, S., Neale, B.M., Faraone, S.V., Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., Absher, D., Agartz, I., Akil, H., Amin, F., Andreassen, O.A., Anjorin, A., Anney, R., Anttila, V., Arking, D.E. and et al. 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics* 45(9), pp. 984–994.

Devlin, B. and Scherer, S.W. 2012. Genetic architecture in autism spectrum disorder. *Current Opinion in Genetics & Development* 22(3), pp. 229–237.

Elder, J.H., Kreider, C.M., Brasher, S.N. and Ansell, M. 2017. Clinical impact of early diagnosis of autism on the prognosis and parent-child relationships. *Psychology research and behavior management* 10, pp. 283–292.

Elsabbagh, M., Divan, G., Koh, Y.-J., Kim, Y.S., Kauchali, S., Marcín, C., Montiel-Nava, C., Patel, V., Paula, C.S., Wang, C., Yasamy, M.T. and Fombonne, E. 2012. Global prevalence of autism and other pervasive developmental disorders. *Autism research : official journal of the International Society for Autism Research* 5(3), pp. 160–179.

Fakhoury, M. 2015. Autistic spectrum disorders: A review of clinical features, theories and diagnosis. *International Journal of Developmental Neuroscience* 43, pp. 70–77.

Fernandez, B.A. and Scherer, S.W. 2017. Syndromic autism spectrum disorders: moving from a clinically defined to a molecularly defined approach. *Dialogues in Clinical Neuroscience* 19(4), pp. 353–371.

Gamazon, E.R. and Stranger, B.E. 2015. The impact of human copy number variation on gene expression. *Briefings in functional genomics* 14(5), pp. 352–357.

Gaugler, T., Klei, L., Sanders, S.J., Bodea, C.A., Goldberg, A.P., Lee, A.B., Mahajan, M., Manaa, D., Pawitan, Y., Reichert, J., Ripke, S., Sandin, S., Sklar, P., Svantesson, O., Reichenberg, A., Hultman, C.M., Devlin, B., Roeder, K. and Buxbaum, J.D. 2014. Most genetic risk for autism resides with common variation. *Nature Genetics* 46(8), pp. 881–885.

Grayton, H.M., Fernandes, C., Rujescu, D. and Collier, D.A. 2012. Copy number variations in neurodevelopmental disorders. *Progress in Neurobiology* 99(1), pp. 81–91.

Harony-Nicolas, H., De Rubeis, S., Kolevzon, A. and Buxbaum, J.D. 2015. Phelan mcdermid syndrome: from genetic discoveries to animal models and treatment. *Journal of Child Neurology* 30(14), pp. 1861–1870.

Henrichsen, C.N., Chaignat, E. and Reymond, A. 2009. Copy number variants, diseases and gene expression. *Human Molecular Genetics* 18(R1), pp. R1-8.

Hnoonual, A., Thammachote, W., Tim-Aroon, T., Rojnueangnit, K., Hansakunachai, T., Sombuntham, T., Roongpraiwan, R., Worachotekamjorn, J., Chuthapisith, J., Fucharoen, S., Wattanasirichaigoon, D., Ruangdaraganon, N., Limprasert, P. and Jinawath, N. 2017. Chromosomal microarray analysis in a cohort of underrepresented population identifies SERINC2 as a novel candidate gene for autism spectrum disorder. *Scientific reports* 7(1), p. 12096.

Jacquemont, M.L., Sanlaville, D., Redon, R., Raoul, O., Cormier-Daire, V., Lyonnet, S., Amiel, J., Le Merrer, M., Heron, D., de Blois, M.C., Prieur, M., Vekemans, M., Carter, N.P., Munnich, A., Colleaux, L. and Philippe, A. 2006. Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. *Journal of Medical Genetics* 43(11), pp. 843–849.

Jaluria, P., Konstantopoulos, K., Betenbaugh, M. and Shiloach, J. 2007. A perspective on microarrays: current applications, pitfalls, and potential uses. *Microbial Cell Factories* 6, p. 4.

Kanner, L. 1968. Autistic disturbances of affective contact. *Acta paedopsychiatrica* 35(4), pp. 100–136.

Klei, L., Sanders, S.J., Murtha, M.T., Hus, V., Lowe, J.K., Willsey, A.J., Moreno-De-Luca, D., Yu, T.W., Fombonne, E., Geschwind, D., Grice, D.E., Ledbetter, D.H., Lord, C., Mane, S.M., Martin, C.L., Martin, D.M., Morrow, E.M., Walsh, C.A., Melhem, N.M., Chaste, P. and Devlin, B. 2012. Common genetic variants,

acting additively, are a major source of risk for autism. *Molecular autism* 3(1), p. 9.

Kolevzon, A., Angarita, B., Bush, L., Wang, A.T., Frank, Y., Yang, A., Rapaport, R., Saland, J., Srivastava, S., Farrell, C., Edelmann, L.J. and Buxbaum, J.D. 2014. Phelan-McDermid syndrome: a review of the literature and practice parameters for medical assessment and monitoring. *Journal of neurodevelopmental disorders* 6(1), p. 39.

Krumm, N., O'Roak, B.J., Karakoc, E., Mohajeri, K., Nelson, B., Vives, L., Jacquemont, S., Munson, J., Bernier, R. and Eichler, E.E. 2013. Transmission disequilibrium of small CNVs in simplex autism. *American Journal of Human Genetics* 93(4), pp. 595–606.

Lai, M.-C., Lombardo, M.V. and Baron-Cohen, S. 2014. Autism. *The Lancet* 383(9920), pp. 896–910.

Leppa, V.M., Kravitz, S.N., Martin, C.L., Andrieux, J., Le Caignec, C., Martin-Coignard, D., DyBuncio, C., Sanders, S.J., Lowe, J.K., Cantor, R.M. and Geschwind, D.H. 2016. Rare inherited and *de novo* cnvs reveal complex contributions to ASD risk in multiplex families. *American Journal of Human Genetics* 99(3), pp. 540–554.

Li, W. and Olivier, M. 2013. Current analysis platforms and methods for detecting copy number variation. *Physiological Genomics* 45(1), pp. 1–16.

Liu, P., Carvalho, C.M.B., Hastings, P.J. and Lupski, J.R. 2012. Mechanisms for recurrent and complex human genomic rearrangements. *Current Opinion in Genetics & Development* 22(3), pp. 211–220.

Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J., Faucett, W.A., Feuk, L., Friedman, J.M., Hamosh, A., Jackson, L., Kaminsky, E.B., Kok, K., Krantz, I.D., Kuhn, R.M., Lee, C. and Ledbetter, D.H. 2010. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *American Journal of Human Genetics* 86(5), pp. 749–764.

Mitz, A.R., Philyaw, T.J., Boccuto, L., Shcheglovitov, A., Sarasua, S.M., Kaufmann, W.E. and Thurm, A. 2018. Identification of 22q13 genes most likely to contribute to Phelan McDermid syndrome. *European Journal of Human Genetics* 26(3), pp. 293–302.

Moreira, D.P., Griesi-Oliveira, K., Bossolani-Martins, A.L., Lourenço, N.C.V., Takahashi, V.N.O., da Rocha, K.M., Moreira, E.S., Vadasz, E., Meira, J.G.C., Bertola, D., O'Halloran, E., Magalhães, T.R., Fett-Conte, A.C. and Passos-Bueno, M.R. 2014. Investigation of 15q11-q13, 16p11.2 and 22q13 CNVs in autism spectrum disorder Brazilian individuals with and without epilepsy. *Plos One* 9(9), p. e107705.

Murphy, C.M., Wilson, C.E., Robertson, D.M., Ecker, C., Daly, E.M., Hammond, N., Galanopoulos, A., Dud, I., Murphy, D.G. and McAlonan, G.M. 2016. Autism spectrum disorder in adults: diagnosis, management, and health services development. *Neuropsychiatric Disease and Treatment* 12, pp. 1669–1686.

Oberman, L.M., Boccuto, L., Cascio, L., Sarasua, S. and Kaufmann, W.E. 2015. Autism spectrum disorder in Phelan-McDermid syndrome: initial characterization and genotype-phenotype correlations. *Orphanet Journal of Rare Diseases* 10, p. 105.

Paula, C.S., Ribeiro, S.H., Fombonne, E. and Mercadante, M.T. 2011. Brief report: prevalence of pervasive developmental disorder in Brazil: a pilot study. *Journal of Autism and Developmental Disorders* 41(12), pp. 1738–1742.

Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., Thiruvahindrapuram, B., Xu, X., Ziman, R., Wang, Z., Vorstman, J.A.S., Thompson, A., Regan, R., Pilorge, M., Pellecchia, G., Pagnamenta, A.T., Oliveira, B., Marshall, C.R., Magalhaes, T.R., Lowe, J.K. and Scherer, S.W. 2014. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *American Journal of Human Genetics* 94(5), pp. 677–694.

Reis, V.N. de S., Kitajima, J.P., Tahira, A.C., Feio-Dos-Santos, A.C., Fock, R.A., Lisboa, B.C.G., Simões, S.N., Krepischi, A.C.V., Rosenberg, C., Lourenço, N.C., Passos-Bueno, M.R. and Brentani, H. 2017. Integrative Variation Analysis Reveals that a Complex Genotype May Specify Phenotype in Siblings with Syndromic Autism Spectrum Disorder. *Plos One* 12(1), p. e0170386.

Sarasua, S.M., Dwivedi, A., Boccuto, L., Rollins, J.D., Chen, C.-F., Rogers, R.C., Phelan, K., DuPont, B.R. and Collins, J.S. 2011. Association between deletion size and important phenotypes expands the genomic region of interest in Phelan-McDermid syndrome (22q13 deletion syndrome). *Journal of Medical Genetics* 48(11), pp. 761–766.

Sbacchi, S., Acquadro, F., Calò, I., Calì, F. and Romano, V. 2010. Functional annotation of genes overlapping copy number variants in autistic patients: focus on axon pathfinding. *Current Genomics* 11(2), pp. 136–145.

Shishido, E., Aleksic, B. and Ozaki, N. 2014. Copy-number variation in the pathogenesis of autism spectrum disorder. *Psychiatry and Clinical Neurosciences* 68(2), pp. 85–95.

Silva, I.M.W. 2017. InDels e CNVs pequenas em pacientes com Transtorno do

Espectro Autista.

Soorya, L., Kolevzon, A., Zweifach, J., Lim, T., Dobry, Y., Schwartz, L., Frank, Y., Wang, A.T., Cai, G., Parkhomenko, E., Halpern, D., Grodberg, D., Angarita, B., Willner, J.P., Yang, A., Canitano, R., Chaplin, W., Betancur, C. and Buxbaum, J.D. 2013. Prospective investigation of autism and genotype-phenotype correlations in 22q13 deletion syndrome and SHANK3 deficiency. *Molecular autism* 4(1), p. 18.

Sztainberg, Y. and Zoghbi, H.Y. 2016. Lessons learned from studying syndromic autism spectrum disorders. *Nature Neuroscience* 19(11), pp. 1408–1417.

Tabet, A.-C., Rolland, T., Ducloy, M., Lévy, J., Buratti, J., Mathieu, A., Haye, D., Perrin, L., Dupont, C., Passemard, S., Capri, Y., Verloes, A., Drunat, S., Keren, B., Mignot, C., Marey, I., Jacquette, A., Whalen, S., Pipiras, E., Benzacken, B. and Bourgeron, T. 2017. A framework to identify contributing genes in patients with Phelan-McDermid syndrome. *npj Genomic Medicine* 2, p. 32.

Uchino, S. and Waga, C. 2013. SHANK3 as an autism spectrum disorder-associated gene. *Brain & Development* 35(2), pp. 106–110.

Wiśniowiecka-Kowalnik, B., Kastory-Bronowska, M., Bartnik, M., Derwińska, K., Dymczak-Domini, W., Szumbarska, D., Ziemka, E., Szczałuba, K., Sykulski, M., Gambin, T., Gambin, A., Shaw, C.A., Mazurczak, T., Obersztyn, E., Bocian, E. and Stankiewicz, P. 2013. Application of custom-designed oligonucleotide array CGH in 145 patients with autistic spectrum disorders. *European Journal of Human Genetics* 21(6), pp. 620–625.

Zhang, F., Gu, W., Hurles, M.E. and Lupski, J.R. 2009. Copy number variation in human health, disease, and evolution. *Annual Review of Genomics and Human Genetics* 10, pp. 451–481.

Zhao, M., Wang, Q., Wang, Q., Jia, P. and Zhao, Z. 2013. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC Bioinformatics* 14 Suppl 11, p. S1.

Chapter 2

Detection of rare CNVs in a Brazilian cohort of patients with autism spectrum disorder (ASD) by using a customized array-CGH

Claudia Ismania Samogy Costa[1], Isabela Mayah Wayhs Silva[1], Eduarda Morgana Montenegro[1], Eloísa de Sá Moreira[1], Silvia Souza da Costa[1], Gerson Shigeru Kobayashi[1], Naila Lourenço[1], Carla Rosenberg[1], Ana Krepischi[1] and Maria Rita dos Santos e Passos-Bueno1.

[1]Department of Genetics, Institute of Biosciences, University of São Paulo, Brazil

Key words: Copy number variation (CNV), Autism spectrum disorder (ASD), customized array-CGH

Abstract

Autism spectrum disorder (ASD) is a neurodevelopmental condition with strong genetic component, and copy number variations (CNVs) are one of the most important genetic alterations. Since 2010, chromosomal microarray has been recognized as the first-tier genetic test for detecting copy number variation (CNV) in patients with autism spectrum disorder. This study aimed at identifying and characterizing CNVs in a Brazilian cohort of ASD patients. We first validated a customized array-CGH and screened 105 patients, with a diagnosis yield at 15%. We also described CNVs detected by commercial CMA in 16 patients, adding up to a cohort of 121 Brazilian patients. Overall, we provided further evidence for 27 ASD candidate genes in 26 CNVs (*CSMD1*, *CNTNAP4, CNTN6, DPP6, NR3C2, CACNA2D3, MIB1, LPRRPC, KDM5B, DOCK8, CAMK4, GRPR, POMGNT1, FAT1, PTGER3, AUTS2, PFKP*, *FARP*, *GLRA3*, *CDH13, CDH8, CDH11, CSGALNACT1, PTPNR2, MTNR1A, GRIK1 and FGF2*) and described 15 new CNVs. We also screened a large Brazilian cohort consisting of 1,029 patients with different neurodevelopmental disorders, thus reinforcing the association with neurodevelopmental disruption for three CNVs (including *CDH13, TRIM16* and *PTPNR2* genes), and provided further evidence that at least 70% of our findings are not polymorphisms in our population. Altogether, our data further characterizes CNVs in Brazilian ASD patients and adds to elucidating ASD etiology with a novel customized aCGH platform.

Introduction

Autism spectrum disorder (ASD) is a common genetic condition characterized by problems in social communication/interaction, and restricted, repetitive behavioral patterns (DSM-5, Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition). ASD affects about 1% of the world population and occurs four times more frequently in males than in females (Elisabbagh et al., 2012). In Brazil, the prevalence has been estimated at 0.27% of the population (Paula et al., 2011). However, these differences may be associated with cultural, economic and methodological issues. Additional psychiatric conditions or comorbidities manifest in up to 70% of ASD cases (Liu and Takumi 2014).

The genetic architecture of ASD is complex and heterogeneous. It is estimated that about 10-25% of ASD cases are associated with a single genetic change, and the most common of them is the variation in the number of copies (CNVs) and single nucleotide variants (SNV) (Bourgeron, 2015). However, most of the cases are the result of the combined action of more than one genetic variant (in a multifactorial or oligogenic inheritance model) (Levy, David & Robert, 2009).

The relevance of CNVs in ASD has been confirmed in several reports (Pinto et al. 2014; Krumm et al. 2015; Rosenfeld et al. 2010; Shishido et al. 2014; Eriksson et al. 2015) and, since 2010, chromosomal microarray analysis (CMA) has been recommended as the first-tier clinical diagnostic test for detecting CNVs in patients with ASD (Miller et al. 2010). These CNVs, either inherited or *de novo*, may be both duplications and deletions, and although they are generally unique among patients, some chromosomal regions (e.g. 15q11-q13, 16p11 and 22q11) are most often affected (Bourgeon, 2015). Although both small (<50Kb) and larger CNVs have been reported in ASD cohorts, the small ones remain poorly characterized (Krumm et al. 2013).

Several ASD associated CNVs are shared with other neurodevelopmental disorders (NDDs), such as intellectual disability (ID), schizophrenia and attention-deficit/hyperactivity disorder (ADHD) (Grayton et al. 2012; Rosenfeld et al. 2010). Moreover, CMA in different populations has been useful in detecting novel candidate genes for ASD and other NDDs (Guo et al.

2017; Hnoonual et al. 2017). In Brazil, our population remains poorly characterized regarding CNVs that may be relevant to ASD etiology.

Our main goal was to identify and characterize rare CNVs in a Brazilian cohort of 121 patients with ASD of unknown cause. This was achieved by applying an updated, customized array-CGH (aCGH) that includes 700 additional genes, totalizing 1,527 ASD targets (Moreira et al., 2016). Moreover, our findings were validated in a Brazilian cohort consisting of 1,019 NDD patients.

Methods

Cohort

A total of 121 ASD patients, who were seen at the Centro de Pesquisa sobre o Genoma Humano e Células-tronco (CEGH-CELL, University of São Paulo) for genetic counseling and research participation, were included in this project. This is part of the CEGH-CELL 1000 cohort, for which we systematically interviewed the parents' probands to collect data from pregnancy, proband's development, exams already performed, diagnosis and comorbidities, whenever possible. A consent form allowing the use of information collected was signed by the proband's guardians. ASD individuals were divided into two groups. Group one consisted of 105 ASD individuals (23 females; 82 males) and group two consisted of 16 ASD individuals (15 males and 1 female) with positive CMA result from an outsourced laboratory.

Customized aCGH 180K:

We redesigned a previously customized aCGH 60K with 269 candidate genes produced by Agilent Technologies (Moreira et al. 2016). The new version now targets 1,527 genes (list of genes available upon request) that are selected following criteria: a) Genes already associated with ASD described on the SFARI database; b) Genes associated with other neuropsychiatric diseases; c) genes involved in biological pathways described as altered in autistic probands; d) genes with similar function to that of genes already associated with ASD and / or e) genes encoding proteins with function in the brain. All the exons and exon junctions, including the untranslated regions (UTRs) of each of the targeted genes, were covered. The high-quality (HD) oligonucleotide probes of the slide were selected by using the SureDesign program (Agilent, https://earray.chem.agilent.com/suredesign/), with a mean coverage of 4 probes per segment (median = 5), with additional probes being designed for regions represented by less than three probes. Each probe has the size of 60 base pairs and, by using the criterion of three consecutive probes altered for detection of CNVs, corresponds to a resolution of 400bp. HD probes were also selected in the 10 Kpb adjacent to the candidate genes, with a median of 3 HD probes in these regions, as well as additional probes in covered areas.

The remaining genome (backbone) is represented by 71,006 probes, with 36Kbp being the median of the spacing between them (corresponding to a resolution of 72Kb, considering 3 consecutive altered probes).

Sample hybridization

All samples were previously tested for concentration and quality by using the NanoDrop ND 1000 device (Peqlab BiotechnologieTM Spectrophotometer). The sample hybridization can be divided into the following steps: enzymatic fragmentation of samples, labeling with fluorophores (cy3 or cy5), purification, hybridization and washing. All these steps followed the Agilent Technologies protocol. A DLRS value equal to or less than 0.2 was used as quality control. The statistical algorithm used in the analysis was ADM-2, with a threshold value of 6.0. The aberration filter used was the DefaultAberrationFilter_v1, which selects variations presenting at least three consecutive probes changed, modifying only the value of the log2 ratio cy3 / cy5, with an absolute minimum value of at least 0.3. The genome reference used was hg19. All samples were tested in duplicate by using a dye-swap strategy with a single sex-matched reference, and we only accepted concordant results.

Customized-Array-CGH 180K Validation

To estimate array sensitivity, we used the DNA from 14 probands with 16 known CNVs (5 CNVs detected by MLPA-343 and 11 detected by different commercial CMA platforms). Since we did not have access to the experiment and raw data from outsourced laboratories, a commercial 180K aCGH was used to determine whether a discordant result between experiments (our aCGH vs. that of the outsourced lab) was a false negative of our customized array or a false positive from the outsourced lab.

Concerning specificity, we tested 14 CNVs (detected in 12 probands), to which we had applied our customized aCGH, by using other methodologies: 6 CNVs (from 5 probands) using a commercial aCGH 180K (from Agilent) and 8 CNVs using qPCR.

Commercial aCGH 180K hybridization followed the manufacturer's protocol. To perform qPCR, primers were designed encompassing the minimum size of each alteration by Primer-BLAST

(https://www.ncbi.nlm.nih.gov/tools/primer-blast/) and UCSC In-Silico PCR (https://genome.ucsc.edu/cgiin/hgPcr?hgsid=663284711_dXRnq0iWPPi4nChvzMipo78JqBFy) software. qPCR was performed in duplicates, with a final volume of 20µL and 20ng of genomic DNA from all probands, by Fast SYBR® Green II Master Mix (Life Technologies), using the ABI Prism 7500 SequenceDetection System (AppliedBiosystems). Ct values were normalized by using the endogenous control genes: *HPTR, GRP15, GAPDH, TRFC* and *ACTB* (depending on the experiment), and the data was normalized by setting a normal control value to 1 (supplementary material S1 shows some examples).

Finally, all CNVs, smaller than 1Kb and detected by our customized aCGH, were also validated by quantitative PCR (qPCR) and CNVs, smaller than 500Kb and detected by outsourced laboratories, were validated by our customized array.

CNV analysis and classification

All CNVs found were characterized by (a) size, (b) type (deletion or duplication), (c) population frequency (based on the data available on the Database of Genomic Variation - DGV), (d) gene content (based on the data of RefSeq genes available on the UCSC genome browser, using the hg19), (e) literature and medical information about disease associated genes, (f) public database of affected individuals (SFARI), (g) inheritance (whenever possible) and (h) gene content expression. Common polymorphisms (CNVs occurring in more than 1% of the population or without genes) were classified as benign and excluded from further analysis. Rare CNVs carrying genes with known neurodevelopmental/brain function or previously classified as ASD risk genes (or neuropsychiatric diseases) were better characterized in silicon to find out what gene regions were compromised (intrinsic, exotic or regulatory element), by using the data available on the ENSEMBL website (https://www.ensembl.org/info/website/tutorials/grch37.html). Ingenuity pathway Analysis (IPA) software (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis) was used to assess cellular functions, while Exac Browser (http://exac.broadinstitute.org) was used to determine genes' pLI score (a

measure of loss-of-function intolerance of each gene based on the ratio between the numbers of variants observed and expected).

This information together with the International Standard Cytogenomic Array (ISCA) and the American College of Medical Genetic (ACMG) standards was taken into account to classify the CNVs as pathogenic, benign, or variant of uncertain clinical significance (VOUS). CNVs classified as VOUs were further subclassified as (a) likely benign, (b) potentially pathogenic or (c) VOUs with no sub-classification. Therefore, we considered *de novo* and inherited CNVs in known ASD genes or previously reported in the literature as potentially pathogenic. CNVs with genes that have known nervous system function and/or association with other NDDs, but no previous association to ASD, were classified as VOUs with no sub-classification.

Finally, the origin of every pathogenic, potentially pathogenic and VOUs with no sub- classification CNV was assessed when parental DNA was available.

Screening of a Brazilian cohort with different neurodevelopmental disorders (NDDs)

We screened a cohort consisting of 1,019 Brazilian patients with intellectual disability, either associated or not to other neurodevelopmental disorders, here referred to as NDD cohort, in order to 1) verify whether our large CNV represents polymorphisms in our population and 2) validate the CNVs found in our cohort as potential players in disturbing the neurodevelopment process. DNA sample of all these probands were submitted to CMA exams by using different platforms, such as the aCGH 8x60K (ID 021924) aCGH 180k (ID 022060), from Agilent Technologies (average probe spacing = 54.5Kb, 17.6Kb, respectively). These exams were performed by the research Laboratory in Human Genetics from the Departamento de Genética e Biologia Evolutiva do Instituto de Biociências da Universidade de São Paulo (leading investigators: Dr. Carla Roenberg and Dr. Ana Krepischi), from 2004 to 2018.

Results

Specificity and sensitivity of the customized aCGH

To determine sensitivity of the new customized aCGH, we analyzed 14 previously genotyped ASD individuals with 16 known CNVs ranging from 2.5Kb to 1.5Mb in size (median size of 136Kb) (Supplementary material, table S2). Fourteen of the sixteen CNVs were also detected in our aCGH. The two remaining CNVs were reported by an outsourced laboratory but not by our platform. One of them, a 76Kb deletion affecting *NRXN1* gene, was also tested by a commercial 180K aCGH platform (Agilent technologies) and by qPCR, and, since it was not validated, it was not included in the sensitivity estimate. The second one was 6Kb duplication in the same gene and was not tested by a second method. Then, according to the data gathered, our customized aCGH presented a sensitivity of 93% (15/16).

Another fourteen CNVs were used to determine *array* specificity (table S2). These CNVs ranged from 91bp to 1.9Mb in size. All nine CNVs that were over 1Kb were validated by using other methods, while CNVs that were smaller than this size presented a high rate of false positives (only 40% (2/5) were validated).

Detected CNVs

Analysis of 105 probands by our customized aCGH revealed 31 rare CNVs in 26 probands, with size ranging from 478bp to 11.9Mb (median size of 198.6Kb) (table 1a). Eight out of 31 (26%) CNVs the customized aCGH detected were smaller than 50Kb, and 42% (13/31) of them are intragenic. Four CNVs (in 3 probands) were classified as pathogenic, whereas 15 CNVs (in 13 probands) were classified as potentially pathogenic and 11 CNVs (in 11 probands) as VOUs with no sub-classification. The remaining CNVs were classified as potentially benign and were excluded from further analyses. Considering the number of patients with pathogenic and/or potentially pathogenic CNVs, the overall detection rate for the customized aCGH was 15% (16 out of 105 patients).

Furthermore, the outsourced aCGH analyses of 16 probands revealed 11 CNVs that were classified as potentially pathogenic, 4 CNVs that were classified as VOUs with no sub-classification and 1 CNV that was classified as likely benign (table 1b), which was excluded from further analysis. These 16 CNVs ranged from 2.5Kb to 3.3Mb in size.

We were able to test the CNV origin in 21 out of the 47 CNVs found in these cohorts. Eight of them were maternally inherited, six were paternally inherited and seven were de novo. For eight out of the twenty remaining CNVs only one parental (maternal or paternal) test was available. Seven of these CNVs are not maternally inherited and one is not paternally inherited (Table 1).

Overall, in 121 probands, we identified four CNVs that were classified as pathogenic and 26 CNVs that were classified as potentially pathogenic.

Pathogenic CNVs

Three of the 4 pathogenic CNVs map to regions previously implicated in known neurodevelopmental syndromes: a 2q32.1-q33.1 deletion of 11.9Mb, compatible with Glass syndrome (proband ID: F11410-1), a 4p16.3-p16.1 deletion of 5.5 Mb, compatible with Wolf-Hirschhorn syndrome (proband ID: F11426-1) and a 1q21 duplication of 11.9Mb, compatible with 1q21 duplication syndrome (proband ID: F5502-1). Probands F11426-1 and F5502-1 carry additional CNVs: proband F11426-1 carries a 4p16.1 pathogenic duplication of 2.6Mb, and F5502-1 carries a potentially pathogenic 560.5Kb duplication at Xp22.2. The phenotype of these patients was congruent with their molecular alteration (Table 1). Parental testing, conducted for proband F5502-1 only, revealed that both CNVs were maternally inherited.

Table 1. ***Summary of all pathogenic, likely pathogenic and VOUs with no sub-classification found in this study***. A) CNVs found in a cohort tested by our customized aCGH 180K. b) CNVs found by outsourced laboratories. Two patients presented negative results and were not included in this table. **De novo*/paternal: only maternal test was available and CNV was not maternally inherited. Abbreviations: ASD: autism spectrum disorder, ID: intellectual disability; DD: developmental delay; delay; NA: not available, MRI: magnetic resonance imaging.

Case ID	Additional clinical features	Previous genetic exams	Sex	Clinical diagnosis	Coordinates (hg19)				Gain/loss	Size (pb)	Gene involved	Origin	Classification
a) Alterations detected by the customized aCGH 180K													
F5502-1	Speech delay, mild hearing loss, irritability.	None	M	ASD and 1q duplication	1	q21.1	145899339	147824207	Dup	1924868	Many	Maternal	Pathogenic
					X	p22.2	15964240	16524779	Dup	560539	*GRPR (5); MAGEB17*	Maternal	potentially pathogenic
F11410-1	Absent speech, campodactyly, facial dysmorphisms and ID	None	M	ASD and 2q deletion	2	q32.1 - q33.1	189161039	201092158	Del	11931119	*Many*	NA	Pathogenic
F11426-1	Cardiac alterations, absent speech, ID, facial dysmorphisms and hypotonia.	None	M	ASD and 4p deletion	4	p16.3 - p16.1	1704380	7212894	Del	5508514	*Many*	NA	Pathogenic
					4	p16.1	7321929	9901111	Dup	2579182	*Many*	NA	Pathogenic
					19	p13.2	12122159	12148401	Del	26242	*ZNF433*	NA	No sub-classification
F10757-1	NA	X-fragile and MLPA (P343)	M	ASD	3	p14.3	54919369	54920237	Dup	868	*CACNA2D3 (2)*	*De novo*	potentially pathogenic
C26025	None	None	M	ASD	7	q11.22	69063762	69064240	Dup	478	*AUTS2 (3)*	Paternal	potentially pathogenic
F10644-1	scruff scarf, macrocrania	MLPA (P343)	M	ASD	9	q33.1	118128604	119249801	Dup	1121197	*ASTN2(3), PAPPA, DEC1 E LOC101928775*	Maternal	potentially pathogenic
F10848-1	cardiac and gastrological alterations and hypotonia	X-fragile and MLPA (P343)	F	ASD	9	p24.3	172364	239485	Del	67121	*CBWD1, DOCK8 (3)*	Maternal	potentially pathogenic
F11367-1	Reflux, convulsion, umbilical hernia	None	F	ASD and Fibromyalgia	18	q11.2	19357843	19395904	Del	38061	*MIB1 (4)*	*De novo*/ paternal	potentially pathogenic

Case ID	Additional clinical features	Previous genetic exams	Sex	Clinical diagnosis	Coordinates (hg19)				Gain/ loss	Size (pb)	Gene involved	Origin	Classification
F11455-1	hypotonia and DD	None	M	ASD	1	p31.1	71415328	71440210	Del	24882	*PTGER3 (not scored)*	Paternal	potentially pathogenic
F11562-1	Development regression and frequent vomiting	None	M	ASD	1	q32.1	202746393	202857089	Dup	110696	*KDM5B (2), LOC148709, RABIF*	*De novo*/ paternal	potentially pathogenic
					5	q22.1	110260870	110730612	Dup	469742	*TSLP, WDR36, CAMK4 (4)*	*De novo*/ paternal	potentially pathogenic
F10291-1	Hypotonia	MLPA (P343)	M	ASD	7	q36.2	153998331	154002788	Del	4457	*DPP6 (4)*	*De novo*	potentially pathogenic
F6281-1	Echolalia	X-fragile and Karyotype	M	ASD	4	q31.2 3	149300414	149555306	Del	254892	*NR3C2 (3)*	*De novo*	potentially pathogenic
F11407-1	Echolalia, delayed speech and food restriction	None	M	ASD	2	p21	44112889	44128874	Dup	15985	*LRPPRC (not scored)*	NA	potentially pathogenic
F11604-1	None	MLPA (P343), X-fragile and Karyotype	F	ASD	10	p15.3 - p15.2	2373294	3174063	Dup	800769	*PFKP*	*De novo*	potentially pathogenic
F11431-1	facial dysmorphisms obesity and ID	X-fragile and MLPA (P343)	M	ASD	17	p13.1	6656136	6794838	Del	138702	*XAF1, FBXO39, TEKT1, ALOX12P2*	NA	No sub-classification
					2	q37.3	242343392	242344421	Del	1029	*FARP2*	NA	potentially pathogenic
F11486-1	None	None	M	ASD	4	q34.1	175649802	175709869	Dup	60067	*GLRA3*	Maternal	potentially pathogenic
F11394-1	frequent vomiting	X-fragile and MLPA (P343)	M	ASD	6	q15	90174326	90359551	Dup	185225	*ANKRD6, LYRM2, MDN1*	NA	No sub-classification
F11089-1	delayed speech and frequent vomiting	X-fragile	M	ASD	19	p12	24007283	24366074	Dup	358791	*RPSAP58, ZNF254, LOC100101266, ZNF276*	*De novo*/ paternal	No sub-classification

Case ID	Additional clinical features	Previous genetic exams	Sex	Clinical diagnosis	Coordinates (hg19)				Gain/ loss	Size (pb)	Gene involved	Origin	Classification
F6642-1	NA	X-fragile	M	ASD	7	p22.2	3142074	3258015	Del	115941	*LOC100129603*	NA	No sub-classification
F10806-1	None	X-fragile and MLPA (P343)	M	ASD	10	p15.1	4049156	4384306	Del	335150	*LOC101927964, LINC00702*	NA	No sub-classification
F11208-1	Hearing alteration	None	M	ASD	18	q23	76782230	76974053	Del	191823	*ATP9B*	NA	No sub-classification
F11564-1	DD, facial dysmorphism, reflux and ID	X-fragile and MLPA (P343)	M	ASD	6	q21	107427262	107682874	Dup	255612	*BEND3, PDSS2*	NA	No sub-classification
F10842-1	Macrocrania, hypotonia, myalgia and facial dysmorphisms	X-fragile and MLPA (P343)	M	ASD	5	p13.2	37272912	37478198	Del	205286	*NUP155, WDR70*	NA	No sub-classification
F11430-1	Hearing alterations	karyotype	M	ASD	17	p12	15561251	15888832	Dup	327581	*TRIM16, ZNF286A, TBC1D26, MEIS3P1, ADORA2B, ZSWIM7*	NA	No sub-classification
F11561-1	None	None	M	ASD	1	q44	246827500	247074460	Dup	246960	*CNST, SCCPDH, LOC149134, AHCTF1*	NA	No sub-classification
F11554-1	NA	X-fragile	M	ASD	Y	q11.2 22	20574793	20936561	Dup	361768	*HSFY2, HSFY1, TTTY9A, TTTY9B*	NA	potentially benign
b) Alterations detected by commercial aCGH platforms													
F11407-1	Food restriction	karyotype and X-fragile	M	ASD	1	p34.1	46595386	46690969	Dup	95583	*PIK3R3, TSPAN1, POMGNT1 (not scored), LURAP1*	*De novo*	potentially pathogenic

Case ID	Additional clinical features	Previous genetic exams	Sex	Clinical diagnosis	Coordinates (hg19)				Gain/loss	Size (pb)	Gene involved	Origin	Classification
F8995-1	NA	X-fragile and MLPA (P343)	M	ASD	3	p26.3	1442910	1445441	Del	2531	*CNTN6 (3)*	Maternal	potentially pathogenic
F11273-1	Delayed speech	karyotype and X-fragile	M	ASD	16	q23.1	76495677	76501480	Del	5803	*CNTNAP4 (4)*	Paternal	potentially pathogenic
F11520-1	Seizures, development regression and speech alterations	X-fragile	M	ASD	7	q36.3	158196079	158631220	Dup	435141	*PTPRN2, NCAPG2, ESYT2*	*De novo*	potentially pathogenic
					16	q21	61949547	65248573	Del	3299026	*CDH8 (4), CDH11 (not scored)*	Maternal	potentially pathogenic
F11164-1	MRI and ophthalmologic alterations	X-fragile	M	ASD	4	q28.1	123804848	124189068	Dup	384220	*FGF2, NUDT6, SPATA5*	*De novo*/ paternal	potentially pathogenic
F11404-1	Development regression, magnetic resonance imaging alteration and ADHD	karyotype and X-fragile	M	ASD	8	p23.2	3502641	5095396	Dup	1592755	*CSMD1 (not scored)*	*De novo*/ paternal	potentially pathogenic
F11473-1	Speech alteration	karyotype and X-fragile	M	ASD	21	q21.3 - q21.11	31160690	31680049	Del	519359	*GRIK1, CLDN17, CLDN8*	*De novo*/ paternal	potentially pathogenic
PCGH1481	NA	NA	F	ASD	4	q35.2	187454411	187584497	Del	130086	*MTNR1A, FAT1 (not scored)*	*De novo*	Potentially pathogenic
F11397-1	Delayed speech and echolalia	None	M	ASD	8	p21.3	19148867	19438015	Dup	289148	*SH2D4A, CSGALNACT1*	NA	potentially pathogenic
F3591-1	NA	None	M	ASD	16	q23.3	83316229	83450041	Del	133812	*CDH13*	Paternal	potentially pathogenic
					Y	q11.221	16159175	16298135	Dup	138961	*VCY, VCY1B*	Paternal	potentially benign
F11517-1	hypotonia, hyperactivity and facial dysmorphisms	karyotype and X-fragile	M	ASD	2	q37.3	238375008	238649938	Dup	274930	*RAB17, PRLH, MLPH, LRRFIP1*	Paternal	No sub-classification

Case ID	Additional clinical features	Previous genetic exams	Sex	Clinical diagnosis	Coordinates (hg19)				Gain/loss	Size (pb)	Gene involved	Origin	Classification
PCH1482	NA	NA	M	ASD	X	q28	152697418	152754991	Dup	57573	*TREX2*	Maternal	No sub-classification
F11403-1	NA	None	M	ASD	16	p11.2 - p11.1	32033397	34765203	Dup	2731806	*Many*	NA	No sub-classification
F11508-1	NA	NA	M	ASD	16	p11.2	34466474	34765203	Dup	298729	*LINC01566, FRG2DP, TP53TG3HP*	*De novo*/ maternal	No sub-classification

Potentially pathogenic CNVs

Eighteen (9 deletions and 9 duplications) of the 26 potentially pathogenic CNVs include at least one gene that was previously described on the SFARI database (table 2). We observed that 61% (11/18) of these CNVs are intragenic (*CSMD1*, *CNTNAP4, CNTN6, CSMD1, DPP6, NR3C2, CACNA2D3, MIB1, LPRRPC, PTGER3, AUTS2*), wherein 5 of them are moderately (pLI = 0.79) or highly (pLI > 0.9) loss-of-function intolerant (*DPP6, NR3C2, CACNA2D3*, *AUTS2* and *PTGER*). The other seven CNVs affect more genes (besides the SFARI one), including a deletion encompassing *CDH11* and *CDH8*, both are highly loss-of-function intolerant (pLI = 1.00).

Although the remaining 8 CNVs do not encompass SFARI genes, they affect genes that are relevant to neural development and/or function and have been previously associated with ASD (table S3). Four of them are intragenic (*FKP*, *FARP*, *GLRA3* and *CDH13*), but none of them encompasses genes that are highly intolerant to loss-of-function variants.

Taken together, 57.7% (15/26) of all these 26 potentially pathogenic CNVs are intragenic, where 33% (5/15) of them affect genes that are moderately or highly loss-of-function intolerant. Moreover, the 11 remaining CNVs affect more than one gene, and 54.5% (6/11) of them include other gene that is likely important for nervous system development/function (*NCAPG2, PAPPA, RABIF, TSLP, PIK3R3* and *SPATA5*). Table S4 in supplementary material presents more details for these six genes.

CNVs of uncertain significance

Fifteen CNVs (six deletions and nine duplications) were classified as VOUs with no sub-classification. Ten of them encompass genes with known functions in the nervous system or that participate in pathways that may be disrupted in ASD (table S5 shows some examples). Only two of them are moderately or highly loss-of-function intolerant (*CNST*; pLI = 0.93 and *WDR70*; pLI = 0.80). One maternally inherited duplication affecting the *TREX2* gene was found in one ASD patient, but not in his healthy brother, and two CNVs map to region 16p11.2 but not in the core of the previous ASD 16p11 region (16p11.2[28511321-34534371], OMIM #614671). Finally, we also detected two CNVs that include long non-coding RNA genes of unknown function.

Table 2. Sfari genes in 18 CNVs found in our cohort. **Both genes are present in the same CNV.*

Sfari score	Gene	Alteration type
Score 2 (strong evidence)	*CACNA2D3* *KDM5B*	Duplication Duplication
Score 3 (suggestive evidence)	*ASTN2* *AUTS2* *DOCK8* *CNTN6* *NR3C2*	Duplication Duplication Deletion Deletion Deletion
Score 4 (Minimal evidence)	*DPP6* *CNTNAP4* *MIB1* *CDH8** *CAMK4*	Deletion Deletion Deletion Deletion Duplication
Score 5 (Hypothesized)	*GRPR*	Duplication
Not scored	*PTGER3* *CSMD1* *POMGNT1* *CDH11** *LRPPRC* *FAT1*	Deletion Duplication Duplication Deletion Duplication Deletion

CNVs validation in the NDD cohort

Ten out of the 1,019 patients of the NDD cohort carry CNVs that overlap our genomic findings (table 3). Six out of the nine CNVs with genomic overlap in the two cohorts encompass genes or chromosomal regions that were previously associated with ASD: Five CNVs include known ASD candidate genes (*AUTS2, DPP6, CSMD1, LRPPRC* and *CAMK4*) and one is a 9q33.1 duplication that partially overlapped a CNV found in our cohort. The 4 remaining CNVs, common in both cohorts, encompass chromosomal regions or specific genes that have yet to be described on ASD databases (7q36.3, 9q33.1, 17p12 and *CDH13*).

Table 3 CNVs found in a cohort of 1,019 patients with NDDs with overlapping CNVs found in our ASD cohort.

ID	Sex	Additional clinical features	Chromosome region		Coordinates (hg19)		Gain/loss	Size	Gene involved
PCGH1192	M	NA	7	q36.3	158215549	158649005	Duplication	433456	*PTPNR2, NCAPG2, ESYT2, WDR60, MIR595, MIR5707, LINC01022*
PCGH400	F	Neuropsychomotor developmental and growth delay	9	q33.1	118458944	118875161	Duplication	416217	*LOC101928775, LINC00474*
PD1316	M	Neuropsychomotor developmental delay and ASD	16	q23.3	82106299	83558802	Duplication	1452503	*HSD17B2, MPHOSPH6, CDH13, MIR8058, LOC101928446, LOC101928417, MIR3182*
PD1701	M	Neuropsychomotor developmental delay	19	p12	23620632	24070297	Deletion	449665	*ZNF675, ZNF681, RPSAP58*
PD1969	F	ASD and dysmorphisms	7	q11.2	69523434	69904091	Deletion	380657	*AUTS2*
PD2128	F	Neuropsychomotor developmental delay and paroxysmal episodes of ataxia and dystonia	8	p23.3p23.2	1857293	3035992	Duplication	1178699	*Many (including CSMD1)*
PD217	M	Neuropsychomotor developmental delay	2	p21	44080324	44347384	Duplication	267056	*LRPPRC*
PD382	M	Behavior abnormalities, seizure and diabetes	7	q36.3	153523345	153643429	Duplication	120084	*DPP6*
PD608	M	Neuropsychomotor developmental delay and low weight-stature gain, esophageal atresia and renal agenesis	5	q22.1	110237388	110756932	Duplication	519544	*TSLP, WDR36, CAMK4*
PD914	M	Neuropsychomotor developmental delay and dysmorphisms	17	p12	15143011	15679601	Duplication	536590	*TRIM16, ZNF286A, TBC1D26, MEIS3P1, CDRT1, TVP23C, CDRT3, CDRT4, TEKT3, PMP22, MIR4731*

Discussion

Customized aCGH validation and overall yield

We conducted a CMA in 105 Brazilian patients with ASD of unknown cause by applying a novel customized aCGH 180K, which presented good specificity and sensitivity for CNVs larger than 1Kb. Our aCGH platform showed an overall diagnostic yield of 15% (considering both pathogenic and potentially pathogenic CNVs), which is in line with what is found in the literature (Bremer et al. 2011; Eriksson et al. 2015; Coulter et al. 2011; Shen et al. 2010; Stobbe et al. 2014). This detection rate is higher than that achieved by the previous aCGH version (Moreira et al., 2016) and is likely due to platform redesign in probe number and distribution per exon, and number of targeted genes. In addition to that, the backbone probes allowed the lower detection limit for CNV size (1Kb) to be better estimated. However, it is difficult to draw precise comparisons among diagnostic yields across studies due to different cohort selection criteria (such as inclusion of syndromic patients), diagnostic criteria, and CNV inclusion criteria. For instance, individuals, who are positive for MLPA P343 and karyotype alterations, represent about 5% of the ASD cases, but they were excluded from our analyses (Shen et al. 2010; Moreira et al. 2014). In this context, if applied to our cohort as a first-tier exam, the overall diagnosis of our customized aCGH should be higher than 15%.

Although some previous work have pointed out the clinical significance of small CNVs to ASD etiology, this class of CNV remains largely uncharacterized (Krumm et al. 2013). In our cohort, 26% of detected CNVs are smaller than 50Kb, further supporting the contribution of these small CNVs to ASD. Notably, our aCGH platform was able to detect a 15.9Kb intragenic duplication (patient F11470-1) affecting *LRPRC* (an unscored SFARI ASD gene), where only a commercial platform would be able to detect 95.5Kb duplication. Finally, although whole genome sequencing has been appointed to replace aCGH (Yamamoto et al. 2016; Abel and Duncavage 2013), its cost still is prohibitive, particularly in socially unequal nations, such as Brazil. Accordingly, given the relevance of CNVs to ASD etiology, our customized aCGH with good sensitivity and specificity to target smaller CNVs, can be a good choice as a first-tier ASD test.

CNV characterization in Brazilian ASD patients

CNV characterization in Brazilian ASD patients was performed considering deletions and duplications that were detected by both customized and commercial aCGH platforms, amounting to four pathogenic cases, 26 potentially pathogenic cases and 15 VOUs with no sub-classification CNVs cases in 121 patients.

The pathogenic CNVs found in our cohort affect regions that are involved in genetic syndromes (Glass syndrome, 1q21.1 duplication syndrome and Wolf-Hirschhorn syndrome). ASD is an established clinical feature in Glass syndrome, which is a very severe NDD condition, in accordance with the clinical presentation of our patient. Even though both duplications and deletions at the 1q21 region have been described in ASD and several NDDs (schizophrenia, bipolar disorder, depression) (Deshpande and Weiss 2018), the penetrance of the phenotype associated with 1q21.1 CNV is incomplete. It is possible that the potentially pathogenic CNV in Xp22 that was identified in our 1q21.1 patient contributes to the ASD phenotype. Finally, ASD is not frequent among 4p16 (WHS) patients (Fisch et al., 2010). Patient F11416 probably has a complex and unbalanced rearrangement with another pathogenic CNV within 4p in addition to the deletion at 4p16, which represents 40%-50% of the WHS cases (Liu et al. 2012). These preliminary findings suggest that ASD in 1q21 and 4p16 syndromes may be modulated by a second hit.

Among the 26 potentially pathogenic CNVs, eighteen of them involve SFARI candidate genes that were classified with different scores, according to the existing evidence of association. However, this association is not strong for 89% (16/18) of them and, thus, our data reinforces their contribution to ASD, particularly for SFARI categories 3, 4, 5 and “not scored” genes. Moreover, for the 8 remaining CNVs not encompassing SFARI genes, our data provides further evidence for them as ASD candidates, since they have been referred to in the literature at least once.

One current challenge is to determine the functional significance and causality of potentially pathogenic CNVs. Here we suggest that if we combine CNV size with the literature available (i.e. pLI score, previous association with ASD, and gene function), the 11 intragenic CNVs should be prioritized as the

strongest functional candidates (*CSMD1, CNTNAP4, CNTN6, CSMD1, DPP6, NR3C2, CACNA2D3, MIB1, LPRRPC, PTGER3 and AUTS2*), in addition to the CNV involving *CDH8* and *CDH11* (SFARI genes with pLI= 1). In turn, the combination of relevant genes in nervous system development and/or function spanned by the same CNV (as seen in 6 of the 11 non-intragenic potentially pathogenic CNVs) may exert additive effects on the development of ASD core symptoms and/or secondary features, thus adding to the variable expressivity seen among ASD patients. In these cases, the evaluation of pLI score is unfitting, since phenotypes dependent on more than one hit (i.e.: recessive conditions or dominant disorders with incomplete penetrance) yield low scores (pLI < 0.5).

Finally, the screening of our population revealed fifteen CNVs that have never been described in ASD patients and were classified as VOUs with no sub-classification. As no substantial information is available for these genes regarding ASD, further studies are necessary to elucidate their potential as ASD candidates.

Altogether, our data reinforces the relevance of screening different populations to determine the contribution of ASD candidate genes and identify new ones.

Validation of our findings in a large Brazilian NDD cohort

Ten CNVs found in our cohort were also found in this larger cohort. Six of them include known ASD candidate genes, while three include genes with less or no evidence of association. These findings reinforce the potential role these 6 CNVs play in neurodevelopment disruptions and suggest the relevance of the remaining four encompassing *PTPRN2*, *CDH13* and *TRIM6,* among other genes. These findings are in line with recent work, thus suggesting a genotypic convergence across different NDD categories (Deshpande & Weiss 2018, Grayton et al. 2012, Jensen & Girirajan 2017). Moreover, this screening provided evidence that CNVs larger than 50Kb (that is, 70% of all the CNVs described in our cohort of 121 patients), are indeed rare in our population. Since the CMA platforms used in this NDD cohort were not able to detect smaller CNVs, we cannot discard the possibility that these CNVs can represent population-specific polymorphism, as recently reported for *SCL17A16* (Moreira et al., 2016).

Conclusion

Our customized aCGH proved to be a good tool for identifying ASD associated CNVs. We provided further evidence that both small and larger CNVs play a part in ASD etiology and described 10 CNVs that are smaller than 50Kb. We also strengthened the association of 19 SFARI genes (encompassed in 18 CNVs) with ASD etiology (of which 16 do not present strong evidence of association) and 8 non-SFARI candidate genes (CDH13, PTPRN2, FKP, FARP, GLRA3, CSGALNACT1, GRIK1 and FGF2). We also described fifteen CNVs that have never been described in ASD cohorts, thus reinforcing the contribution of studying different populations to elucidate ASD etiology. Moreover, the screening of a large Brazilian cohort provided further evidence that at least 70% of all CNVs described in this work (the CNVs larger than 50Kb) are not polymorphisms in our population. Finally, we reinforced that the combination of different genes in one CNV may contribute to both core and secondary ASD feature development, thus highlighting the importance for identifying ASD genetic causes in order to more precisely determine patients' clinical outcome. Altogether, our data provides further CNV characterization in Brazilian ASD patients and adds to elucidating ASD etiology with the aid of a novel customized aCGH platform.

Supplementary material

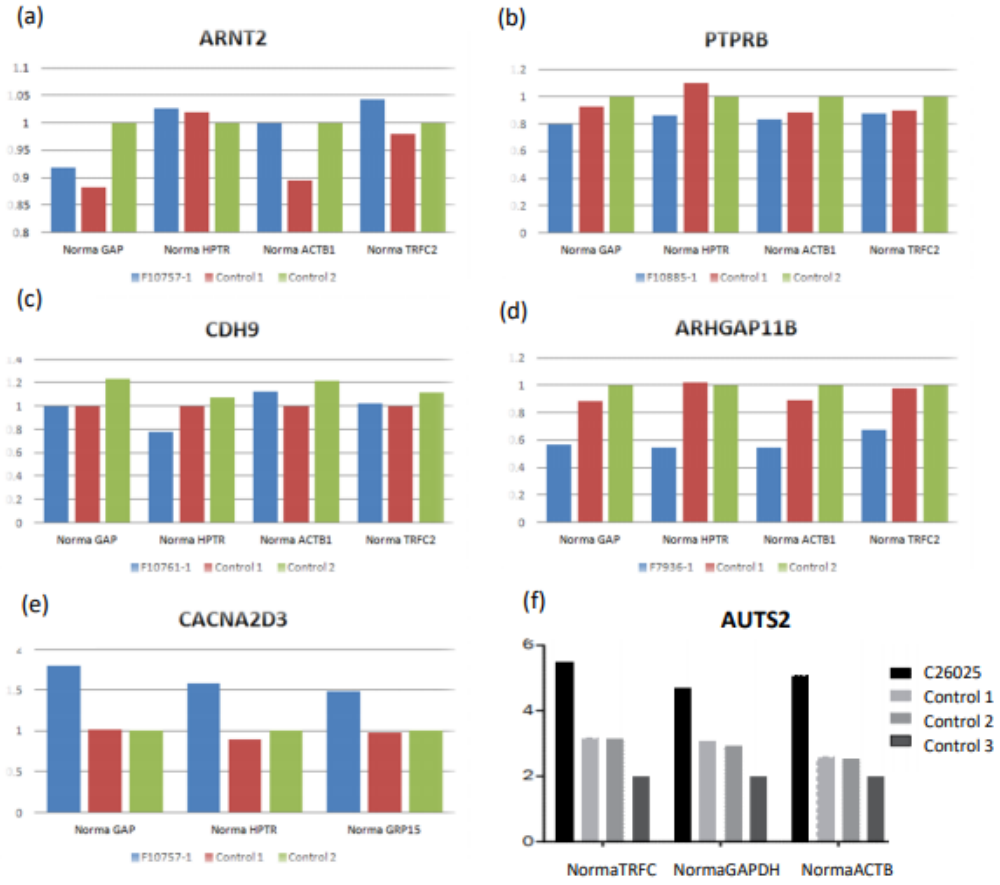

Figure S1. Graphics showing the qPCR experiments for six genes (*ARNT2, PTPRB, CDH9, ARHGAP11B, CACNA2D3 and AUTS2*). The vertical axis represents the copy number of the target region, wherein diploid status was normalized in value "1" for graphs (a) to (e) and "2" for graph (f). Each experiment was conducted by using at least three endogenous genes and two controls. Three of them were validated: *ARHGAP11B, CACNA2D3 and AUTS2*.

Table S2. CNVs used to determine sensitivity and specificity of customized aCGH 180K.

ID	Chromosome region		Coordinates (hg19)		Gain/loss	Size	Gene involved	Genetic exam	Validation
a) CNVs previously detected by other method									
F1626-1	16	p11.2	29517699	30191895	Duplication	674196	Many	MLPA	Validated
F2021-1	22	q13.33	50976445	51178988	Deletion	202543	Many	MLPA	Validated
F4290-1	15	q13.2	30919735	32445252	Deletion	1525517	Many	MLPA	Validated
	16	p12.2	22629047	22709775	Deletion	80728	Many		Validated
F6507-1	16	p11.2	29670772	30198600	Deletion	527828	Many	MLPA	Validated
F7936-1	15	q13.2	30871368	30941572	Deletion	70204	*ULK4P1, ULK4P2, GOLGA8H, ARHGAP11B (not scored), LOC100288637*	aCGH	Validated
F11273-1	16	q23.1	76495677	76501480	Deletion	5803	CNTNAP4	aCGH	Validated
F8995-1	3	p26.3	1442910	1445441	Deletion	2531	CNTN6	aCGH	Validated
F11397-1	8	q21.3	19181933	19434698	Duplication	252765	SH2D4A, CSGALNACT1	aCGH	Validated
F11164-1	4	q28.1	123804848	124189068	Duplication	384220	FGF2, NUDT6, SPATA5	aCGH	Validated
F11407-1	1	p34.1	46595386	46690969	Duplication	95583	PIK3R3, TSPAN1, POMGNT1 (not scored), LURAP1, LOC101929626, LOC110117498	aCGH	Validated
F10806-1	2	p16.3	50539818	50615382	Deletion	75564	NRXN1	aCGH	Not validated
F8402-4	2	p16.3	51253534	51259607	Duplication	6073	NRXN1	aCGH	Not validated
F11508-1	16	p11.2	34466474	34765203	Duplication	298729	LINC01566, FRG2DP, TP53TG3HP	aCGH	Validated
F3591-1	Y	q11.22 1	16159175	16298135	Duplication	138961	VCY, VCY1B	aCGH	Validated
	16	q23.3	83316229	83450041	Deletion	133812	CDH13	aCGH	Validated
b) CNVs detected in 12 patients by the customized aCGH 180K and validated by other method									
F10757-1	15	q25.1	80844922	80845208	Deletion	286	ARNT2	qPCR	Not validated
	3	p14.3	54919369	54920237	Duplication	868	CACNA2D3	qPCR	Validated

ID	Chromosome region		Coordinates (hg19)		Gain/loss	Size	Gene involved	Genetic exam	Validation
F10761-1	5	p14.1	26889849	26889975	Deletion	303	CDH9	qPCR	Not validated
F10885-1	12	q15	70938241	70938332	Deletion	91	*PTPRB (not scored)*	qPCR	Not validated
F10291-1	7	q36.2	153998331	154002788	Deletion	4457	*DPP6 (4)*	qPCR	Validated
C26025	7	q11.22	69063762	69064240	Duplication	478	*AUTS2 (3)*	qPCR	Validated
F6024-1	1	p13.3	110177934	110240126	Duplication	62192	GSTM1, GSTM2, GSTM4	qPCR	Validated
F7936-1	15	q13.2	30871368	30941572	Deletion	70204	*ULK4P1, ULK4P2, GOLGA8H, ARHGAP11B (not scored), LOC100288637*	qPCR	Validated
F10290-1	10	q21.3	68275831	68381747	Deletion	105916	*CTNNA3*	aCGH	Validated
F8293-1	8	p23.1	7088308	8079861	Deletion	991553	Many	aCGH	Validated
F6281-1	4	q31.23	149300414	149555306	Deletion	254892	*NR3C2 (3)*	aCGH	Validated
F10644-1	9	q33.1	118128604	119249801	Duplication	1121197	ASTN2(3), PAPPA, DEC1 E LOC101928775	aCGH	Validated
F5502-1	1	q21.1	145899339	147824207	Duplication	1924868	Many	aCGH	Validated
	X	p22.2	15964240	16524779	Duplication	560539	GRPR; MAGEB17		Validated

Table S3. ASD previously described genes (Non-Sfari) within 8 CNVs found in our cohort.

Gene	Previous ASD studies
CDH13	A study conducted with 1,124 ASD families found 6 regions of rare recurrent *de novo* CNVs, including one in 16p13.2 embracing CDH13 gene (Guo et al. 2017)
	Genome-scan for IQ discrepancy in autism found two statically significant loci, one of them corresponding to CDH13 gene location (Chapman et al. 2011).
CSGALNACT1	Report of an ASD diagnosed boy carrying a *de novo* 8p22-21.3 duplication of 1Mb, embracing three genes: CSGALNACT1, SH2D4A and PSD3 (Dong et al. 2015)
FARP2	Report of a boy with autism and brachymetaphalangy, with a 2q37.3 deletion of 3.5Mb, including the genes FERM, FARP2, GPC1, HDLBP, KIF1A and PASK. Lymphoblastoid cell expression analyses of these genes were performed and three of these genes, FARP2, HDLBP, and PASK, are considerably downregulated (Felder et al. 2009)
FGF2	Serum FGF-2 levels were found to be low in ASD probands, compared to healthy individuals. In addition, it represented a statistically significant negative correlation identified between serum FGF-2 levels and CARS score (Esnafoglu & Ayyıldız 2017).
GLRA3	A case of a boy diagnosed with autism in which a chromosome 4q interstitial microdeletion was detected. He is hemizygous for AMPA 2, GLRA3, GLRB, NPY1R and NPY5R (Ramanathan et al. 2004).
GRIK1	A *de novo* 8.8-Mb deletion of 21q21.1-q21.3 was found in a boy with developmental delay and autistic behavior, containing 19 genes, including NCAM2 and GRIK1 genes, which the authors argue that may be good candidates for the proband's phenotype (Haldeman-Englert et al. 2010)
PFKP	A paternally inherited deletion was detected in a study conducted with 1,275 ASD cases (and their parents) using the Illumina SNP - microarray platform (Hurles et al. 2008).
PTPNR2	Duplication embracing this gene has been reported in the Sfari CNV database ((Mordaunt et al., 2015))
MTNR1A	Prevalence of deleterious melatonin receptor mutations in a cohort composed by 295 ASD probands and 284 individuals from general population found no difference of prevalence of these deleterious mutations (Miller et al. 2010).
	A 1.2Mb deletion in 4q35.2 was found in a boy with several neurdovelopmental disorders, including ASD. This deletion embraces MTNR1A, FAT1 and F11 genes (Youngs et al. 2012)
	Mutations in the regulatory region of MTNR1A was found in a Swedish group of 109 probands diagnosed with ASD (and were absent in a control group composed by 188 people from general Swedish population) (Krumm et al. 2013).

Table S5. Genes within CNVs found in our cohort that either have a known function in the nervous system that may be disrupted in ASD and/or are candidate for other NDDs.

Gene	Function
ZNF433	ZNF433 was pointed out as a candidate for factor susceptibility for multiple sclerosis, a disease marked by inflammatory lesions in brain and spinal cord during early stages. However, its brain function is unknown (Nischwitz et al. 2010).
XAF1	XAF1 is a XIAP binding protein, which acts on preventing cell apoptosis. Under normal conditions, XAF1 protein is constitutively expressed at low levels in neurons. However, increased expression apparently enhances neuronal susceptibility to degeneration, either by suppressing the ability of XIAP to complex with caspases or by sequestering XIAP in nuclear inclusions (Siegelin et al. 2005).
CNST	Consortin is a binding partner of connexins, which, in turn, are known as players in synaptic formation and function (del Castillo et al. 2010) (Grubišić & Parpura 2015).
PDSS2	PDSS2 is an essential protein for ubiquinone synthesis and cerebellum is one of the most often affected organs in ubiquinone deficiency. In Pdss2 knockout mice, a severe cerebellum hypoplasia was observed by impairing cell migration and eliciting ectopic apoptosis.
ADORA2B	ADORA2B is a receptor expressed in myeloid cells. Its activation transactivates nociceptors of sensory neurons and eventually causes hypersensitive neurons and chronic pain in an IL-6 and sIL-6R trans-signaling-dependent manner (Hu et al. 2016; Liu et al. 1998).
ATP9B	ATP9B is a class 2 P4-ATPase that may play a role in neuronal differentiation. It also interacts with MCP-1, a chemokine expressed in different tissues, including astrocytes and microglial cells, and is associated with neuronal diseases, such as Alzheimer (Barber et al. 2015; Frese et al. 2017).
WDR70	WDR70 is a WD40-repeat protein that takes part in a protein complex and acts in single DNA double-strand break resection (Zeng et al. 2016).
ZNF254	ZNF254 is zinc finger protein not well characterized. In a study conducted with 1,697 patients with bipolar disorder, three of them were found carrying a CNV embracing this gene (Grozeva et al. 2010).
ANKRD6	ANKRD6 is an ankyrin repeat-containing protein expressed in neonatal and adult mouse brain (Tissir et al. 2002). ANKRD6 is an essential component of the Wnt signaling pathway (Allache et al. 2015), which has been pointed out as one of the pathways that may be disrupted in ASD (Kwan et al. 2016).
RAB17	In mouse hippocampal neurons, Rab17 regulates dendritic growth and morphogenesis, in addition to postsynaptic development (Mori et al. 2012).

Table S4. Gene function examples of additional genes embraced by CNVs found in out cohort carrying a previous associated ASD gene and classified as potentially pathogenic.

Genes	Function
NCAPG2	NCAPG2 is a member of the Condensin II system. A deletion followed by alteration in expression has been observed in a boy with microcephaly and mental deficiency (Perche et al. 2013).
PAPPA	Pregnancy associated plasma protein-A (PAPP-A) member of insulin like growth factors (IGFs) axis. Higher levels in maternal serum were correlated to increased risk of macrosomia in newborns (Poon et al. 2011).
RABIF	RABIF is a soluble protein that interacts with specific Rab protein. RABIF can influence vesicular traffic in a secretion assay, which measures neurotransmitter release mediated by synaptic vesicle exocytosis (Gulbranson et al. 2017).
TSLP	TSLP is a pro-allergic cytokine produced by a variety of epithelial cells, including keratinocytes, as well as dendritic cells. It is associated with atopic dermatitis and itch through sensory neurons stimulation (Turner & Zhou 2014)
PIK3R3	PI3K plays an important role during nerve growth factor-stimulated differentiation and in brain development (Nguyen et al. 2010).
SPATA5	SPATA5 protein plays an important role in mitochondrial dynamics and axonal growth. SPATA5 deficiency is associated with intellectual disability, microcephaly, hypotonia, spasticity, seizures, sensorineural hearing loss and cortical visual impairment (Puusepp et al. 2018).

Table S4. Ingenuity Pathway Analysis and pLI information of ASD candidate genes embraced in CNVs classified as potentially pathogenic.

Gene symbol	SFARI Score	Roles in cell (IPA)	pLI
ASTN2	3	NA	0.99
AUTS2	3	Cell reorganization, assembly, extension and migration.	1.00
CACNA2D3	2	**Synaptic transmission**, cell size	**1.00**
CAMK4	4	abnormal morphology, differentiation, long term depression, **synaptic transmission**, growth, phosphorylation in, cell death, degradation in	0.02
CDH11	Not scored	migration, interaction, invasion, invasiveness, cell spreading, expression in, motility, adhesion, formation, plasticity	1.00
CDH13	NA	Cell adhesion, morphology, apico-basal polarity, dissemination, and endhocytosis.	0.22
CDH8	4	Adhesion	1.00
CNTN6	3	abnormal morphology, adhesion	0.00
CNTNAP4	4	accumulation,	0.00
CSGALNACT1	NA	Endochondral ossification and cell organization.	0.00
CSMD1	Not scored	Stabilization	NA
CTNNA3	4	cell-cell adhesion, abnormal morphology, compaction	0.00
DOCK8	3	migration by, formation, migration, function, apoptosis, proliferation, binding, survival, cytotoxicity	0.00
DPP6	4	Transmembrane potential, action potential, afterhyperpolarization, long-term potentiation	0.97
FARP2	NA	Cell differentiation, length, remodeling, reorganization and adhesion.	0.00
FAT1	Not scored	organization, adhesion, morphogenesis, disorganization, transcription in, expression in, invasion, phosphorylation in, migration, cell-cell contact	0.00
FGF2	NA	Cell proliferation, migration, differentiation, survival, mitogenesis and growth.	NA
GLRA3	NA	Cell excitation.	0.04
GRIK1	NA	Transmembrane potential, cell plasticity, long-term potentiation, depolarization, **synaptic transmission** and action potential.	0.00
GRPR	5	Cell proliferation, action potential, mobilization in, expression in	0.14
KDM5B	2	Expression in, cell proliferation, methylation in cell, binding in cell, accumulation, mitosis, migration, demethylation in cell, G2/M phase	0.00
LRPPRC	Not scored	Cellular respiration, translation in, transcription in, oxidative phosphorylation in, density, beta-oxidation in, modification, binding in	0.00
MIB1	4	Cell differentiation, cell death, binding in, ubiquitination in, apoptosis, outgrowth in, morphology, endocytosis by, development, sister chromatid exchange in	0.00
MTNR1A	NA	Melatonin signaling	0.00
NR3C2	3	Expression in, transactivation in, activation in, signaling in, destabilization in, mono-ubiquitination in, association in, reabsorption by, transcription in, ubiquitination in	0.93
PFKP	NA	AMPK signaling and glycolysis.	0.00

POMGNT1	Not scored	Cell proliferation, O-mannosylation in, modification in, infectibility	0.00
PTGER3	Not scored	Fosphorylation in cell, activation in cell, tumorigenicity, cell-cell contact, transactivation in, proliferation, elevation in cell, retraction, transepithelial electrical resistance and aggregation.	0.79
PTPRN2	NA	Cell mitogenesis, morphology and survival.	0.02

References

Abel, H.J. and Duncavage, E.J. 2013. Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. Cancer genetics 206(12), pp. 432–440.

American Psychiatric Association 2013. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Washington, DC: American Psychiatric Association.

Barber, R.C., Phillips, N.R., Tilson, J.L., Huebinger, R.M., Shewale, S.J., Koenig, J.L., Mitchel, J.S., O'Bryant, S.E., Waring, S.C., Diaz-Arrastia, R., Chasse, S., Wilhelmsen, K.C., Alzheimer's Disease Neuroimaging Initiative and Texas Alzheimer's Research and Care Consortium 2015. Can genetic analysis of putative blood alzheimer's disease biomarkers lead to identification of susceptibility loci? *Plos One* 10(12), p. e0142360.

Bourgeron, T. 2015. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nature Reviews. Neuroscience* 16(9), pp. 551–563.

Bremer, A., Giacobini, M., Eriksson, M., Gustavsson, P., Nordin, V., Fernell, E., Gillberg, C., Nordgren, A., Uppströmer, A., Anderlid, B.-M., Nordenskjöld, M. and Schoumans, J. 2011. Copy number variation characteristics in subpopulations of patients with autism spectrum disorders. American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics 156(2), pp. 115–124.

del Castillo, F.J., Cohen-Salmon, M., Charollais, A., Caille, D., Lampe, P.D., Chavrier, P., Meda, P. and Petit, C. 2010. Consortin, a trans-Golgi network cargo receptor for the plasma membrane targeting and recycling of connexins. *Human Molecular Genetics* 19(2), pp. 262–275.

Chapman, N.H., Estes, A., Munson, J., Bernier, R., Webb, S.J., Rothstein, J.H., Minshew, N.J., Dawson, G., Schellenberg, G.D. and Wijsman, E.M. 2011. Genome-scan for IQ discrepancy in autism: evidence for loci on chromosomes 10 and 16. *Human Genetics* 129(1), pp. 59–70.

Coulter, M.E., Miller, D.T., Harris, D.J., Hawley, P., Picker, J., Roberts, A.E., Sobeih, M.M. and Irons, M. 2011. Chromosomal microarray testing influences medical management. Genetics in Medicine 13(9), pp. 770–776.

Deshpande, A. and Weiss, L.A. 2018. Recurrent reciprocal copy number variants: Roles and rules in neurodevelopmental disorders. Developmental Neurobiology 78(5), pp.519-530.

Elsabbagh, M., Divan, G., Koh, Y.-J., Kim, Y.S., Kauchali, S., Marcín, C., Montiel-Nava, C., Patel, V., Paula, C.S., Wang, C., Yasamy, M.T. and Fombonne, E. 2012. Global prevalence of autism and other pervasive developmental disorders. *Autism research : official journal of the International Society for Autism Research* 5(3), pp. 160–179.

Esnafoglu, E. and Ayyıldız, S.N. 2017. Decreased levels of serum fibroblast growth factor-2 in children with autism spectrum disorder. *Psychiatry Research* 257, pp. 79–83.

Eriksson, M.A., Liedén, A., Westerlund, J., Bremer, A., Wincent, J., Sahlin, E., Gillberg, C., Fernell, E. and Anderlid, B.-M. 2015. Rare copy number variants

are common in young children with autism spectrum disorder. Acta Paediatrica 104(6), pp. 610–618.

Felder, B., Radlwimmer, B., Benner, A., Mincheva, A., Tödt, G., Beyer, K.S., Schuster, C., Bölte, S., Schmötzer, G., Klauck, S.M., Poustka, F., Lichter, P. and Poustka, A. 2009. FARP2, HDLBP and PASK are downregulated in a patient with autism and 2q37.3 deletion syndrome. *American Journal of Medical Genetics. Part A* 149A(5), pp. 952–959.

Fisch, G.S., Grossfeld, P., Falk, R., Battaglia, A., Youngblom, J. and Simensen, R. 2010. Cognitive-behavioral features of Wolf-Hirschhorn syndrome and other subtelomeric microdeletions. *American journal of medical genetics. Part C, Seminars in medical genetics* 154C(4), pp. 417–426.

Frese, C.K., Mikhaylova, M., Stucchi, R., Gautier, V., Liu, Q., Mohammed, S., Heck, A.J.R., Altelaar, A.F.M. and Hoogenraad, C.C. 2017. Quantitative Map of Proteome Dynamics during Neuronal Differentiation. *Cell reports* 18(6), pp. 1527–1542.

Grayton, H.M., Fernandes, C., Rujescu, D. and Collier, D.A. 2012. Copy number variations in neurodevelopmental disorders. Progress in Neurobiology 99(1), pp. 81–91.

Grozeva, D., Kirov, G., Ivanov, D., Jones, I.R., Jones, L., Green, E.K., St Clair, D.M., Young, A.H., Ferrier, N., Farmer, A.E., McGuffin, P., Holmans, P.A., Owen, M.J., O'Donovan, M.C., Craddock, N. and Wellcome Trust Case Control Consortium 2010. Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. Archives of General Psychiatry 67(4), pp. 318–327.

Grubišić, V. and Parpura, V. 2015. The second brain in autism spectrum disorder: could connexin 43 expressed in enteric glial cells play a role? *Frontiers in Cellular Neuroscience* 9, p. 242.

Guo, H., Peng, Y., Hu, Z., Li, Y., Xun, G., Ou, J., Sun, L., Xiong, Z., Liu, Y., Wang, T., Chen, J., Xia, L., Bai, T., Shen, Y., Tian, Q., Hu, Y., Shen, L., Zhao, R., Zhang, X., Zhang, F. and Xia, K. 2017. Genome-wide copy number variation analysis in a Chinese autism spectrum disorder cohort. Scientific reports 7, p. 44155.

Haldeman-Englert, C.R., Chapman, K.A., Kruger, H., Geiger, E.A., McDonald-McGinn, D.M., Rappaport, E., Zackai, E.H., Spinner, N.B. and Shaikh, T.H. 2010. A *de novo* 8.8-Mb deletion of 21q21.1-q21.3 in an autistic male with a complex rearrangement involving chromosomes 6, 10, and 21. *American Journal of Medical Genetics. Part A* 152A(1), pp. 196–202.

Hnoonual, A., Thammachote, W., Tim-Aroon, T., Rojnueangnit, K., Hansakunachai, T., Sombuntham, T., Roongpraiwan, R., Worachotekamjorn, J., Chuthapisith, J., Fucharoen, S., Wattanasirichaigoon, D., Ruangdaraganon, N., Limprasert, P. and Jinawath, N. 2017. Chromosomal microarray analysis in a cohort of underrepresented population identifies SERINC2 as a novel candidate gene for autism spectrum disorder. Scientific reports 7(1), p. 12096.

Hu, X., Adebiyi, M.G., Luo, J., Sun, K., Le, T.-T.T., Zhang, Y., Wu, H., Zhao, S., Karmouty-Quintana, H., Liu, H., Huang, A., Wen, Y.E., Zaika, O.L., Mamenko, M., Pochynyuk, O.M., Kellems, R.E., Eltzschig, H.K., Blackburn, M.R., Walters,

E.T., Huang, D. and Xia, Y. 2016. Sustained Elevated Adenosine via ADORA2B Promotes Chronic Pain through Neuro-immune Interaction. *Cell reports* 16(1), pp. 106–119.

Krumm, N., O'Roak, B.J., Karakoc, E., Mohajeri, K., Nelson, B., Vives, L., Jacquemont, S., Munson, J., Bernier, R. and Eichler, E.E. 2013. Transmission disequilibrium of small CNVs in simplex autism. American Journal of Human Genetics 93(4), pp. 595–606.

Krumm, N., Turner, T.N., Baker, C., Vives, L., Mohajeri, K., Witherspoon, K., Raja, A., Coe, B.P., Stessman, H.A., He, Z.-X., Leal, S.M., Bernier, R. and Eichler, E.E. 2015. Excess of rare, inherited truncating mutations in autism. Nature Genetics 47(6), pp. 582–588.

Kwan, V., Unda, B.K. and Singh, K.K. 2016. Wnt signaling networks in autism spectrum disorder and intellectual disability. *Journal of neurodevelopmental disorders* 8, p. 45.

Lai, M.-C., Lombardo, M.V. and Baron-Cohen, S. 2014. Autism. *The Lancet* 383(9920), pp. 896–910.

Liu, H.L., Golder-Novoselsky, E., Seto, M.H., Webster, L., McClary, J. and Zajchowski, D.A. 1998. The novel estrogen-responsive B-box protein (EBBP) gene is tamoxifen-regulated in cells expressing an estrogen receptor DNA-binding domain mutant. *Molecular Endocrinology* 12(11), pp. 1733–1748.

Liu, X. and Takumi, T. 2014. Genomic and genetic aspects of autism spectrum disorder. Biochemical and Biophysical Research Communications 452(2), pp. 244–253.

Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J., Faucett, W.A., Feuk, L., Friedman, J.M., Hamosh, A., Jackson, L., Kaminsky, E.B., Kok, K., Krantz, I.D., Kuhn, R.M., Lee, C. and Ledbetter, D.H. 2010. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. American Journal of Human Genetics 86(5), pp. 749–764.

Mitz, A.R., Philyaw, T.J., Boccuto, L., Shcheglovitov, A., Sarasua, S.M., Kaufmann, W.E. and Thurm, A. 2018. Identification of 22q13 genes most likely to contribute to Phelan McDermid syndrome. European Journal of Human Genetics 26(3), pp. 293–302.

Mordaunt, D.A., McIntyre, L.C., Salvemini, H., Ibrahim, A., Bratkovic, D., Ketteridge, D., Scott, H.S., Kassahn, K.S. and Smith, N. 2015. Presentation of m.3243A>G (MT-TL1; tRNALeu) variant with focal neurology in infancy. American Journal of Medical Genetics. Part A 167A(11), pp. 2697–2701.

Moreira, D.P., Griesi-Oliveira, K., Bossolani-Martins, A.L., Lourenço, N.C.V., Takahashi, V.N.O., da Rocha, K.M., Moreira, E.S., Vadasz, E., Meira, J.G.C., Bertola, D., O'Halloran, E., Magalhães, T.R., Fett-Conte, A.C. and Passos-Bueno, M.R. 2014. Investigation of 15q11-q13, 16p11.2 and 22q13 CNVs in autism spectrum disorder Brazilian individuals with and without epilepsy. Plos One 9(9), p. e107705.

Moreira, E.S., Silva, I.M.W., Lourenço, N., Moreira, D.P., Ribeiro, C.M., Martins, A.L.B., Griesi-Oliveira, K., Lazar, M., Costa, S.S., Naslavsky, M.S., Rocha,

K.M., Aguena, M., Fett-Conte, A.C., Zatz, M., Rosenberg, C., Zachi, E.C., Bertola, D.R., Vadasz, E. and Passos-Bueno, M.R. 2016. Detection of small copy number variations (CNVs) in autism spectrum disorder (ASD) by custom array comparative genomic hybridization (aCGH). Research in autism spectrum disorders 23, pp. 145–151.

Mori, Y., Matsui, T., Furutani, Y., Yoshihara, Y. and Fukuda, M. 2012. Small GTPase Rab17 regulates dendritic morphogenesis and postsynaptic development of hippocampal neurons. *The Journal of Biological Chemistry* 287(12), pp. 8963–8973.

Nguyen, T.L.X., Kim, C.K., Cho, J.-H., Lee, K.-H. and Ahn, J.-Y. 2010. Neuroprotection signaling pathway of nerve growth factor and brain-derived neurotrophic factor against staurosporine induced apoptosis in hippocampal H19-7/IGF-IR [corrected]. *Experimental & Molecular Medicine* 42(8), pp. 583–595.

Nischwitz, S., Cepok, S., Kroner, A., Wolf, C., Knop, M., Müller-Sarnowski, F., Pfister, H., Roeske, D., Rieckmann, P., Hemmer, B., Ising, M., Uhr, M., Bettecken, T., Holsboer, F., Müller-Myhsok, B. and Weber, F. 2010. Evidence for VAV2 and ZNF433 as susceptibility genes for multiple sclerosis. Journal of Neuroimmunology 227(1–2), pp. 162–166.

Paula, C.S., Ribeiro, S.H., Fombonne, E. and Mercadante, M.T. 2011. Brief report: prevalence of pervasive developmental disorder in Brazil: a pilot study. *Journal of Autism and Developmental Disorders* 41(12), pp. 1738–1742.

Perche, O., Menuet, A., Marcos, M., Liu, L., Pâris, A., Utami, K.H., Kervran, D., Cacheux, V., Laudier, B. and Briault, S. 2013. Combined deletion of two Condensin II system genes (NCAPG2 and MCPH1) in a case of severe microcephaly and mental deficiency. *European Journal of Medical Genetics* 56(11), pp. 635–641.

Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., Thiruvahindrapuram, B., Xu, X., Ziman, R., Wang, Z., Vorstman, J.A.S., Thompson, A., Regan, R., Pilorge, M., Pellecchia, G., Pagnamenta, A.T., Oliveira, B., Marshall, C.R., Magalhaes, T.R., Lowe, J.K. and Scherer, S.W. 2014. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. American Journal of Human Genetics 94(5), pp. 677–694.

Poon, L.C.Y., Karagiannis, G., Stratieva, V., Syngelaki, A. and Nicolaides, K.H. 2011. First-trimester prediction of macrosomia. *Fetal Diagnosis and Therapy* 29(2), pp. 139–147.

Puusepp, S., Kovacs-Nagy, R., Alhaddad, B., Braunisch, M., Hoffmann, G.F., Kotzaeridou, U., Lichvarova, L., Liiv, M., Makowski, C., Mandel, M., Meitinger, T., Pajusalu, S., Rodenburg, R.J., Safiulina, D., Strom, T.M., Talvik, I., Vaarmann, A., Wilson, C., Kaasik, A., Haack, T.B. and Õunap, K. 2018. Compound heterozygous SPATA5 variants in four families and functional studies of SPATA5 deficiency. *European Journal of Human Genetics* 26(3), pp. 407–419.

Ramanathan, S., Woodroffe, A., Flodman, P.L., Mays, L.Z., Hanouni, M., Modahl, C.B., Steinberg-Epstein, R., Bocian, M.E., Spence, M.A. and Smith, M. 2004. A case of autism with an interstitial deletion on 4q leading to hemizygosity

for genes encoding for glutamine and glycine neurotransmitter receptor sub-units (AMPA 2, GLRA3, GLRB) and neuropeptide receptors NPY1R, NPY5R. *BMC Medical Genetics* 5, p. 10.

Rosenfeld, J.A., Ballif, B.C., Torchia, B.S., Sahoo, T., Ravnan, J.B., Schultz, R., Lamb, A., Bejjani, B.A. and Shaffer, L.G. 2010. Copy number variations associated with autism spectrum disorders contribute to a spectrum of neurodevelopmental disorders. Genetics in Medicine 12(11), pp. 694–702.

Shen, Y., Dies, K.A., Holm, I.A., Bridgemohan, C., Sobeih, M.M., Caronna, E.B., Miller, K.J., Frazier, J.A., Silverstein, I., Picker, J., Weissman, L., Raffalli, P., Jeste, S., Demmer, L.A., Peters, H.K., Brewster, S.J., Kowalczyk, S.J., Rosen-Sheidley, B., McGowan, C., Duda, A.W. and Autism Consortium Clinical Genetics/DNA Diagnostics Collaboration 2010. Clinical genetic testing for patients with autism spectrum disorders. Pediatrics 125(4), pp. e727-35.

Shishido, E., Aleksic, B. and Ozaki, N. 2014. Copy-number variation in the pathogenesis of autism spectrum disorder. Psychiatry and Clinical Neurosciences 68(2), pp. 85–95.

Siegelin, M., Touzani, O., Toutain, J., Liston, P. and Rami, A. 2005. Induction and redistribution of XAF1, a new antagonist of XIAP in the rat brain after transient focal ischemia. Neurobiology of Disease 20(2), pp. 509–518.

Stobbe, G., Liu, Y., Wu, R., Hudgings, L.H., Thompson, O. and Hisama, F.M. 2014. Diagnostic yield of array comparative genomic hybridization in adults with autism spectrum disorders. Genetics in Medicine 16(1), pp. 70–77.

Turner, M.J. and Zhou, B. 2014. A new itch to scratch for TSLP. *Trends in Immunology* 35(2), pp. 49–50.Zeng, M., Ren, L., Mizuno, K., Nestoras, K., Wang, H., Tang, Z., Guo, L., Kong, D., Hu, Q., He, Q., Du, L., Carr, A.M. and Liu, C. 2016. CRL4(Wdr70) regulates H2B monoubiquitination and facilitates Exo1-dependent resection. *Nature Communications* 7, p. 11364.

Yamamoto, T., Shimojima, K., Ondo, Y., Imai, K., Chong, P.F., Kira, R., Amemiya, M., Saito, A. and Okamoto, N. 2016. Challenges in detecting genomic copy number aberrations using next-generation sequencing data and the eXome Hidden Markov Model: a clinical exome-first diagnostic approach. Human Genome Variation 3, p. 16025.

Chapter 3

Characterization of the Phelan-McDermid Syndrome: study of a Brazilian cohort and genotype-phenotype correlation

Claudia Ismania Samogy Costa[1], Elisa Branco Varella[1*], Claudia Spadoni[2], Helen Ferraz[2], Rodrigo Ambrósio Fock[3], Ricardo Henrique Almeida Barbosa[3], Beatriz Alvarez Perez[3], Naila Lourenço[1], Frederico Monfardini[1], Maria Vibranovski[1], Ana Krepischi[1], Carla Rosenberg[1], Maria Rita dos Santos e Passos-Bueno[1].*

[1]Department of Genetics, Institute of Biosciences, University of São Paulo, Brazil; [2]Amigos e Familiares da Síndrome de Phelan-McDermid (AFSPM); [3]Center of Medical Genetics, Federal University of São Paulo, Brazil.

(*) These authors contributed equally to this work

Key words: Phelan-McDermid syndrome, 22q13 deletion syndrome, Autism spectrum disorder (ASD), *SHANK3*

Abstract

Phelan-McDermid Syndrome (PMS) is a rare genetic disorder characterized by global development delay, intellectual disability, mild dysmorphic features and is associated with several comorbidities. PMS primarily results from *SHANK3* disruption, mainly caused by copy number variations (CNVs), usually 22q terminal deletions of different sizes originated from different molecular mechanisms. Although *SHANK3* disruption has been associated with the major neurological symptoms of the syndrome, it cannot explain the great clinical variability among the PMS cases. Our goal was to clinically and genetically characterize a Brazilian cohort of 34 PMS patients and explore the genotype-phenotype correlation underlying this syndrome. In addition to that, we estimated the frequency of PMS among Brazilian ASD and ID patients. Our clinical and genetic data is mostly consistent with that gathered from cohorts of different populations, including the frequency of ASD among PMS patients. Scarce eyebrow was a prominent clinical feature (78% of the patients), which has not been extensively explored in previous work, and genotype-phenotype correlation provided further evidence of renal abnormalities and speech status with alteration size. Finally, in our ASD and NDD cohort, the frequency of PMS ranged from 0.57% to 0.67. To date, this is the first work describing a cohort of Brazilian PMS patients and it can help understand the etiology of ASD and PMS, particularly in our population.

Introduction

Phelan-McDermid Syndrome (PMS) (OMIM #606232) is a rare genetic disorder characterized by development delay, absent or delayed speech, intellectual disability, seizures, renal, cardiac and gastrological abnormalities and mild dysmorphic feature (Phelan and McDermid 2011; Bonaglia et al. 2011). PMS affects males and females in the same proportion and presents a considerable clinical heterogeneity among patients, with different degrees of functional impairment (Boccuto et al., 2013).

PMS primarily results from *SHANK3* disruption (or haploinsufficiency), mainly caused by copy number variations (CNVs), usually 22q terminal deletions of different sizes that originate from different molecular mechanisms (Cochoy et al. 2015). *SHANK3* codes a protein that is mostly abundant in the brain and is involved in the postsynaptic density of glutamatergic synapses, playing a role in synaptic function by modulating dendritic formation (Monteiro and Feng 2017). In addition to PMS, haploinsufficiency of SHANK3 has been associated with different neurological conditions, such as autism spectrum disorder (ASD), schizophrenia and intellectual disability (ID). At least 0.5% of the ASD cases and up to 2% of ID cases are associated with alterations in this gene (de Sena Cortabitarte et al. 2017).

SHANK3 disruption seems to be associated with the major neurological symptoms of the syndrome, but it cannot explain the great clinical variability among PMS cases, and the underlying molecular basis of this clinical variability is still largely unresolved (Tabet et al. 2017; Sarasua et al. 2014; Sarasua et al. 2011; Soorya et al. 2013; Tabet, Rolland, Ducloy, Levy, et al. 2017; Dhar et al. 2010). Although second genetic hits have been described as a contributing mechanism to the severity of comorbidities seen among PMS patients, it has not been extensively investigated (Tabet et al. 2017). Thus, further studies exploring the molecular basis of PMS are necessary to clarify this syndrome's clinical variability.

Therefore, our main goal was to clinically and genetically characterize a Brazilian cohort of 34 Brazilian patients diagnosed with PMS and explore the genotype-phenotype correlation underlying this syndrome. In addition to that, we estimated the frequency of PMS among Brazilian ASD patients and in a

Brazilian cohort of patients with intellectual deficit (either associated with other neurodevelopmental disorders or not), which will be referred to as neurodevelopmental disorder (NDD) cohort. To date, this is the first characterization of Phelan- McDermid syndrome in Brazilian population.

Methods

Cohort

In total, 34 patients with clinical and genetic diagnosis of PMS participated in this study. This cohort consisted of 16 boys (47%) and 18 girls (53%), ranging from 1.8 to 19.6 years old (mean = 7.54, SD= 3.86). The molecular diagnose was obtained by applying one of these techniques: *array*-CGH, SNP-array, MLPA or FISH. Patients are members of the “Amigos e Familiares da Síndrome de Phelan-Mcdermid - Brasil”, a Brazilian group of parents, relatives and friends of people with Phelan-McDermid Syndrome founded in 2013. All the patients' parents or guardians provided genetic exams and answered a standardized medical history questionnaire. The questionnaire presented queries about the patient's development, comorbidities, autism diagnosis and genetic tests. In addition, 27 families also provided pictures of their affected children.

Two other cohorts were included in this study: one consisting of 525 ASD patients [421 (80.2%) males and 104 (19.8%) females] who were tested for CNV at *SHANK3* by multiplex ligation-dependent probe amplification (MLPA). These patients were attended at the Centro de Pesquisa sobre o Genoma Humano e Células-tronco (CEGH-CEL) between 2012 and 2017; the other is the NDD cohort consisting of 2,297 patients. Chromosomal microarray tests were conducted at the Laboratório de pesquisa em genética humana from the Departamento de Genética e Biologia Evolutiva do Instituto de Biociências da Universidade de São Paulo, from 2004 to 2018. The data collected was included in order to estimate the frequency of 22q13.3 CNVs in Brazilian ASD or ID patients.

Finally, all patients' parents or guardians of both cohorts signed a consent form.

Clinical evaluation

Dysmorphology and comorbidities evaluation: Morphological and comorbidity evaluation was conducted by analyzing the patient's medical report and pictures, whenever available. Parents were asked to answer questions

about comorbidities with three possible answers: “Yes” (if the patient has

comorbidities), “No” (if the patient does not have comorbidities) and “I’m not sure” (if the parents cannot answer). All dubious or “not sure” answers were excluded from our analysis. Moreover, three doctors from the Department of Morphology and Genetic of the Escola Paulista de Medicina – UNIFESP evaluated the pictures of these patients, aiming at analyzing the morphological facial features.

Genetic evaluation

All patients have a deletion at 22q13.3 encompassing *SHANK3*. Different exams and platforms were used: *array*-CGH or SNP array exams were conducted in 29 patients, FISH was applied to four patients and MLPA was applied to one patient. All these exams were conducted by different outsourced labs.

All CNVs found by CMA were characterized by size, gene content and genomic coordinate. All deletions were aligned by using the human genome reference 19 (Hg19) at UCSC genome browser (https://genome.ucsc.edu). For the other CNVs that are not associated with the 22q13.33 region, OMIM genes content and morbid genes were identified by using the Decipher database (https://decipher.sanger.ac.uk). In addition to that, genes that are associated with recessive inheritance diseases in regions with heterozygosity loss were assessed by means of Genomic Oligoarray and SNP array evaluation tool for TRIAL software (http://firefly.ccs.miami.edu).

Multiplex Ligation-dependent Probe Amplification (MLPA)

Multiplex ligation-dependent probe amplification (MLPA) is a multiplex PCR method that is used to detect abnormal copy numbers in candidate chromosomal regions in DNA samples. MLPA reaction can be divided into 5 major steps: 1) Denaturation of the patient's DNA and hybridization of the probes, 2) Binding reaction 3) PCR reaction having the probe sequence as target, 4) Separation of the fragments by capillary electrophoresis, 5) analysis of data. SALSA MLPA kit P343 AUT (Cat. P343-100R, MRC Holland, Amsterdam, Netherlands) was used to detect alterations in autism or autistic-like patients.

This kit detects alterations in 15q11-q13, 16p11.2 and 22q13.33, and assay was conducted following the manufacturer's protocol. The PCR products were detected by the ABI 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, US), using capillary electrophoresis, and raw data was analyzed by GeneMarker (Softgenetics).

Autism Spectrum Disorder assessment

In total, 18 patients were evaluated for ASD diagnosis by a specialist. In addition to medical diagnosis, parents were questioned about the patients' autistic-like behavior. Questionnaire presented four possible answers: (1) No, there is no presence of autism symptoms; (2) No, a detailed assessment was conducted, and autism was ruled out; (3) Yes, symptoms were found in a consultation with a neurologist/psychologist/psychiatry; (4) Yes, there was no diagnosis, but we know by his/her behavior.

Data analysis

Descriptive statistics were determined across all measures into four different domains: comorbidities, morphological features, ASD and language status. Mann–Whitney U tests were conducted to explore potential associations between each medical comorbidity and deletion size, and Kruskal-Wallis test was applied to determine whether there is correlation with speech status and deletion size or not.

Results

Characterization of genetic test results

All 34 patients have deletion of *SHANK3*, and 29 of them present terminal deletion, based on chromosomal microarray analysis (Figure 1). Deletion sizes ranged from 49Kb to 9.1Mb (mean size = 4.1Mb, SD = 3.2). Two of them are mosaics: the first one with 8.3Mb (P4) and the second with 8.8Mb (P25). For nine patients, parental genetic testing was available and all deletions confirmed to be *de novo*.

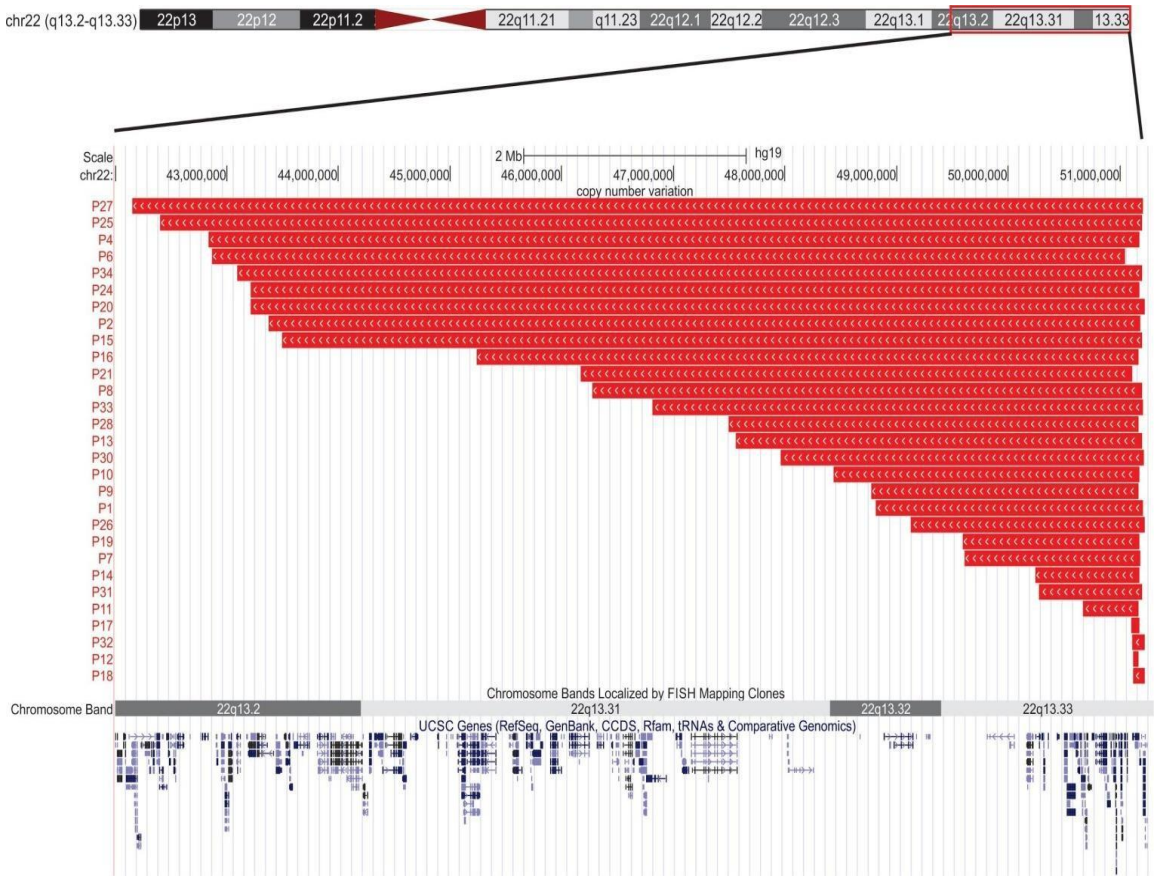

Figure 1. Distribution of deletion sizes at 22q13 based on CMA analysis among 29 PMS patients.

Table 1. CNVs of 34 patients with Phelan-McDermid syndrome.

ID	Sex	Ascertainment method	Rearrangement	Array coordinates arr[GRCh37]	Deletion size (Mb)	Genes	Inheritance
P1	F	SNP array	Deletion	48810119-51211393	2401274	Many	
P2	F	aCGH	Deletion	43371148-51186249	7815101	Many	NA
P3	F	FISH	Deletion**	NA	NA	*SHANK3*	NA
P4	M	aCGH	Deletion (mosaic)	42839080-51174293	8335213	Many	NA
P5	M	FISH	Deletion**	NA	NA	*SHANK3*	NA
P6	M	aCGH 60K	Deletion	42864966-51050847	8185881	Many	*De novo*
P7	F	aCGH 180K	Deletion	49608334-51186249	1577915	Many	NA
P8	F	SNP array 750K	Deletion	46276401-51197766	4921365	Many	NA
P9	F	SNP array 400K	Deletion	48771374-51171678	2400304	Many	*De novo*
P10	F	aCGH	Deletion	48434307-51178213	2743906	Many	*De novo*
P11	F	SNP array 400K	Deletion	50667787-51171678	503891	Many	*De novo*
P12	M	aCGH	Deletion	51122360-51171678	49318	Many	*De novo*
P13	M	SNP array 750K	Deletion	47557877-51197766	3639889	Many	NA
P14	M	aCGH	Deletion	50241089-51178213	937124	ARSA e *SHANK3*	NA
P15	M	SNP array 400K	Deletion	43492638-51197766	7705128	Many	NA
P16	M	SNP array 400K	Deletion	45235285-51171678	5936393	Many	*De novo*

ID	Sex	Ascertainment method	Rearrangement	Array coordinates arr[GRCh37]	Deletion size (Mb)	Genes	Inheritance
P17	F	aCGH	Deletion	51104247-51178574	74327	*SHANK3* e ACR	NA
P18	F	aCGH 180K	Deletion	51123491-51224252	100761	Many	NA
P19	F	aCGH 44K	Deletion	49595567-51178405	1582838	Many	*De novo*
P20	F	aCGH 180K	Deletion	43213659-51224252	8010593	Many	NA
P21	F	SNP array 750K	Deletion	46168628 –51115526	4946898	Many	NA
P22	M	MLPA	Deletion**	NA	NA	ARSA e *SHANK3*	NA
P23	M	MLPA/ FISH	Deletion**	NA	NA	*SHANK3*	NA
P24	F	aCGH 60K	Deletion	43213481-51178354	7964873	Many	NA
P25	F	SNP array 750K	Deletion (mosaic)	42399686-51197766	8798080	Many	*De novo*
P26	M	aCGH 180K	Deletion	49123097-51224252	2101155	Many	NA
P27	M	SNP array 850K	Deletion	42152988-51211392	9058404	Many	NA
P28	F	SNP array 400K	Deletion	47497833-51171678	3673845	Many	*De novo*
P29	F	FISH	Deletion**	NA	NA	*SHANK3*	NA
P30	M	aCGH 180K	Deletion	47963467-51219009	3255542	Many	NA
P31	M	SNP array 750K	Deletion	50274217-51197766	923549	Many	NA
P32	M	aCGh 180K	Deletion	51112491-51224252	111761	Many	NA
P33	F	SNP array 850K	Deletion	46814671-51211392	4396721	Many	Na
P34	M	SNP array	Deletion	43094876- 51197838	8102962	Many	NA

**Coordinates originally in Hg18; ** extension of the deletion not precisely estimated because of the method used (FISH or MLPA).*

Other genetic findings

We also investigated the presence of additional genetic hits, here considered CNV or LOH, as they could add to the severity of the clinical features. Nine patients presented one (n= 3) or more (n= 6) genetic alterations in addition to 22q deletion (table 2). The size of deletions and duplications ranged from 31Kb to 14.8Mb in different genomic regions. Three patients carry a terminal duplication and four patients have interstitial CNVs. Three out of these CNVs encompass genomic regions or genes associated with autosomal dominant or recessive X-linked disorders. Patient P31 carries a duplication that is compatible with the 16p13.3 duplication syndrome and patients P20 and P28 carry duplications affecting IDS (associated with Mucopolysaccharidosis II) and LIPC genes (associated with Diabetes mellitus, insulin non-dependent), respectively. Finally, two out of the seven patients, who underwent SNP-array, presented large regions, larger than 5Mb, with heterozygosity loss. None of the regions was associated with imprinting diseases. However, all LOH regions presented at least one recessive disease gene.

Development

In our cohort, we were able to evaluate speech and motor development based on the information parents reported. Concerning speech, patients were sorted into four different classes: no words, babbling, few words, sentences. Four children were excluded from the analysis because they were under three years old. Seventy seven percent (21/27) of children babbled or used no words, and only 2 children (2/27; 7%) used sentences (P14 and P17). Regarding motor development, 75% (25/33) of them were able to ambulate, being the onset at 2.27 years old (SD=0,84). The remaining 8 patients did not walk at the time they were evaluated (aged from 1.8 to 10 years).

Table 2. Non-22q CNV genomic findings in nine patients with PMS.

ID	Additional genomic findings (Hg19)	Size	Position	Genes	OMIM disease associated
P4	Yq11.21q11.23(13872502-28644194)x0	14.8Mb	Interstitial	32 protein coding genes (USP9Y)	Spermatogenic failure, Y-linked, 2 SPGFY2
	12q24.23q24.33(118841028-133773528)x3 (mosaic)	4.9Mb	Terminal	127 protein coding genes (18 morbid)	And Chromosome Yq11 interstitial deletion syndrome
P11	10q21.3q23.1(68502854-85951641)x2 hmz 13q22.2q31.3(75997199-90483180)x2 hmz 14q21.1q22.1(40349362-51838641)x2 hmz	17.5Mb 14.5Mb 11.5Mb	Interstitial Interstitial Interstitial	117 protein coding genes (96 morbid) 21 protein coding genes (19 morbid) 65 protein coding genes (28 morbid)	Many
P12	6q26(162514945-162652533)x1	1367.6Kb	Interstitial	PRKN	Adenocarcinoma of lung, somatic, Adenocarcinoma, ovarian, somatic, Parkinson disease, juvenile, type 2(AR), {Leprosy, susceptibility to}
	Xq28(148685454-148716519)x3	31Kb	Interstitial	FRAXF	No OMIM diseases associated.
P15	2q11.1(95550957-100920233)x2 hmz	5.4Mb	Interstitial	58 protein coding genes (40 morbid)	Many
P20	18p11.32(64847-464868)x3 Xq28(148094899-148607966)x3	400Kb	Terminal	USP14, THOC1, COLEC12	No OMIM diseases associated.
		513Kb	Interstitial	IDS	Mucopolysaccharidosis II (XLR)
P21	22q13.31(44257185-45143609)x3 (mosaic)	797Kb 620.9Kb	Interstitial	10 protein coding genes (none morbid)	No OMIM diseases associated.
	Xp22.33(7514750-8135644)x3		Interstitial	VCX, PNPLA4, MIR651	No OMIM diseases associated.
P27	21q22.3 (46,807,499-47,126,660)x3	319.2Kb	Interstitial	COL18A1, PCBP3, SLC19A1	Knobloch syndrome, type 1 (AR) – COL18A1
P28	15q21.3(58801559-58861468)x1	59.9Kb	Interstitial	LIPC	Hepatic lipase deficiency (AR), [High density lipoprotein cholesterol level QTL 12], {Diabetes mellitus, non-insulin-dependent} (AD)
P31	10q24.2(100686179-100903763)x1	217.6Kb	Interstitial	HPSE2	Urofacial syndrome 1 (AR)
	11p14.3(23032300-24843680)x3	1.8Mb	Interstitial	LUZP2, MIR8054	No OMIM diseases associated.
	16p13.3(85880-3998442)x3	3.9Mb	Terminal	158 protein coding genes (28 morbid)	Chromosome 16p13.3 duplication syndrome

Dysmorphic features

Among the several dysmorphisms observed (supplementary material table S1), the most common morphological features were ear abnormalities (20/26), large/wide nose (11/26), long / deep or wide philtrum (13/26), scarce eyebrows (20/26), distichiasis (16/26), and large/fleshy hands (21/26), sharp fingers (22/26) and clinodactyly (15/26), elongated phalanges (13/26).

Comorbidities

The most commonly reported comorbidities were increased pain tolerance (24/30) and hypotonia (28/33), occurring in 80% and 85% of the patients, respectively. In addition to that, 20 patients (20/33) had a history of recurring infection of the upper respiratory tract (Table 3). Other common conditions related were gastroesophageal reflux disease (17/30), sleep disturbance (13/33), constipation and/or diarrhea (14/33), renal abnormalities (10/33), chewing difficulty (19/33), swallowing difficult (9/29) and hypohydrosis (9/30).

Autism Spectrum Disorder in PMS

18 of the 34 patients were clinically assessed for ASD. Thirteen of these 18 individuals have ASD (13/18; 72%). Among the 16 remaining patients, twelve of them (12/16: 75%) presented “autism-like behaviors”, such as deficient social interaction, empathy, communication skills, and flexible behavior, as reported by the patients’ caregivers on a questionnaire. Overall, 25 out of the 34 patients (73.5 %) present ASD or autism-like behavior.

SHANK3 alterations frequency in ASD and in neurodevelopmental disorders

We determined the frequency of *SHANK3* CNVs in a Brazilian cohort of 525 patients with ASD diagnosis and observed that three of them (all males) presented a 22q13.3 alteration (0.57%). We also determined *SHANK3* CNVs frequency in the 2,297 patients from the NDD cohort, wherein fifteen patients presented deletions

or duplications in *SHANK3* (0.67%).

Table 3. Comorbidities reported in patients with PMS in our cohort as compared to the literature

Medical comorbidity	N	Total	%	Literature frequency
Increased pain tolerance	24	30	80%	0% - 88%
Hypotonia	28	33	85%	29% - 100%
Recurring upper respiratory tract infections	20	33	61%	8% - 53%
Gastroesophageal reflux	17	30	57%	>25% - 44%
Sleep disturbance	13	33	39%	41% - 46%
Seizures (febrile and/or non-febrile)	9	34	26%	14% - 41%
Constipation and/or diarrhea	14	33	42%	38% - 41%
Renal abnormalities	10	33	30%	17% - 38%
Lymphedema	5	30	17%	22% - 29%
Strabismus	8	33	24%	6% - 26%
Cardiac abnormalities	2	34	6%	3% - >25%
Chewing difficult	19	33	58%	-
Swallowing difficult	9	29	31%	-
Hypohydrosis	9	30	30%	-
Precocious or delayed puberty	4	28	14%	0% - 12%
Hearing loss	1	34	3%	-

Genotype phenotype correlation

Twenty-nine patients presented both CMA result and comorbidity feature information and 20 of them did not present a second hit. Both groups were included for genotype-phenotype correlation analysis. Every comorbidity feature was individually tested to determine whether there was correlation with deletion size. Renal abnormalities were the only comorbidity with significant correlation in both groups (Mann Whitney test: p value = 0.011 for the analysis with 29 patients and 0.03664 for the analysis with 20 patients). Speech status presented a significant correlation with deletion size only in the group with 29 patients (Kruskal-Wallis test: p value = 0.03951 and p value = 0.1511, for the 29 and 20 patient groups, respectively).

Discussion

We analyzed the clinical and genetic data collected from 34 patients diagnosed with PMS. Our work is consistent with the current data described in the literature, as all our patients presented deletions in *SHANK3* (and the size of these deletions was quite variable) and all the patients, who were tested for parental origin, were *de novo (Bonaglia et al. 2011)*.

Overall, our clinical data (comorbidity, dysmorphic features, development and ASD) is consistent with that in the literature. The most common comorbidities in our cohort were increased pain tolerance and hypotonia, as it was previously reported (Soorya et al. 2013). We also evaluated chewing and swallowing difficulties, which were present in 58% of the patients and have not been previously described in the literature. These observations were interpreted as the result of hypotonia, which is a commonly occurring comorbidity in our cohort, as well as in others (Soorya et al. 2013; Costales and Kolevzon 2015; Bonaglia et al. 2011). Regarding dysmorphic features, we observed a high frequency of scarce eyebrows in our cohort (78%), which is a feature that is rarely mentioned in the literature (Dhar et al., 2010). If confirmed in other cohorts, this feature could be included in the morphological evaluation of patients to help clinically identify PMS patients. Finally, only two children were able to formulate full sentences in our cohort and about 75% of our patients were able to walk, but with a delayed onset, which is in line with a global developmental delay observed in different PMS cohorts (Soorya et al. 2013; Costales and Kolevzon 2015; Bonaglia et al. 2011); Tabet et al. 2017; Sarasua et al. 2014; Sarasua et al. 2011; Tabet, Rolland, Ducloy, Levy, et al. 2017; Dhar et al. 2010).

We observed the occurrence of a high frequency of ASD (~73.5%) in our cohort, which is consistent with previous work that report the same in PMS patients (Betancur and Buxbaum, 2013; Costales and Kolevzon, 2015; Leblond et al., 2014; Oberman et al., 2015; Qiu et al., 2018; Soorya et al., 2013; Uchino and Waga, 2013). PMS is a relatively common syndrome among patients with ASD and NDD. We observed a similar frequency of PMS in our ASD patients (0.57%) and in the NDD cohort (0.67%). The proportion of PMS in ASD patients is comparable to other estimates in the literature, which varies from 0.5 to 0.69% (Leblond et al. 2014; Moreira et al. 2014). However, this frequency among NDD is

more variable, ranging from 0.5% to 2.2% (Leblond et al. 2014), possibly because NDD cohorts are clinically more heterogeneous than those that include only ASD patients.

In our genotype-phenotype analysis, we only observed significant correlation of renal abnormalities and speech impairment with deletion size. No patient with deletions smaller than 3.6Kb presented renal abnormalities, but the penetrance of this comorbidity is not complete, since only 53% (8/15) of patients with larger deletions presented it. *ZBED4, CELSR1* and *FBLN1* have been pointed out as potential candidates for renal abnormalities in PMS patients, since these genes are intolerant to loss of function and play a role in kidney development (Mitz et al. 2018). However, they do not seem to correlate with the presence of renal abnormality in our cohort, since *ZBED4* has been deleted in several of our patients that do not present this comorbidity, and *CELSR1* and *FBLN1* were found to be present in several patients with this comorbidity.

Up to now, this work is one of the first ones to report additional genetic findings in PMS patients (Tabet et al., 2017). Nine out of 29 patients (31%) carry additional genetic alterations, wherein three of them carry a terminal duplication, thus suggesting the presence of a balanced translocation at parents' genome. Although it is difficult to determine the combined effect of different genetic alterations, these additional alterations may contribute to phenotype by either worsening the clinical features that are common in PMS or by being responsible for features that are not commonly seen in this syndrome. Finally, two patients out of seven, for whom we had SNP-array report, presented large blocks of LOH regions. However, as no DNA sequencing data is available for any patient, it is difficult to determine whether alterations contribute to phenotype or not.

Conclusion

To date, this is the first work that describes a cohort of Brazilian patients with PMS. Our clinical and genetic data is mostly consistent with that of cohorts of different populations, including the frequency of ASD among PMS patients. Scarce eyebrow was a prominent clinical feature observed in our cohort (78% of the patients) and we suggest that it should be reevaluated in other cohorts, as this clinical sign may represent a potential clinical marker for PMS diagnosis. Genotype-phenotype correlation provided further evidence of renal abnormalities and speech status with size alteration association. In our ASD and NDD cohort, the frequency of PMS was 0.57%-0.67%, thus confirming its relevance among neurodevelopmental disorders. Taken together, our data can add to better understanding PMS.

Supplementary material

Table S1. Frequency of several dysmorphic features seen in PMS patients.

Feature	%		Feature	%
Unusual pattern of the whorl	7.69%		Short philtrum	23.08%
Frontal tuft	7.69%		Absent/flat philtrum	23.08%
Asymmetrical ears	3.85%		Large philtrum	7.69%
Dysplastic Ears	3.85%		Lips facing down	15.38%
Big ears	46.15%		Upper lip Cupid's bow-shaped	7.69%
Low-set ears	30.77%		Macrostomia	3.85%
Ears rolled posteriorly	30.77%		“Open mouth” appearance	11.54%
Auricular fistula	3.85%		Thick lower lip	15.38%
Prominent ears	34.62%		Prominent upper lip	3.85%
Simplified ears	42.31%		Thin upper lip	19.23%
Prominent ears	26.92%		Elongated / extended hands	80.77%
Prominet ear helix	15.38%		Camptodactyly	3.85%
Earlobe attached	19.23%		Clinodactyly	57.69%
Epicanthal folds	7.69%		Long philtrum	15.38%
Upward slanted eyes	11.54%		Tapered fingers	84.62%
Downward slanted eyes	7.69%		Falanges alargadas	11.54%
Distichiasis	61.54%		Elongated phalanges	50.00%
Strabismus	19.23%		Dystrophic fingernails	19.23%
Rarefy eyebrows	76.92%		hyperconvex nails	3.85%
Large / wide nose	42.31%		Elongated feet	38.46%
Short / small nose	7.69%		Small feet	3.85%
Flat face	7.69%		Large feet	26.92%
Hypoplasia of the middle third of the face	19.23%		Hálux valgo	3.85%
Face longa/afilada	15.38%		Enlarged toes	19.23%
Triangular face	7.69%		Clinodactyly (feet)	53.85%
Mento pontudo	3.85%		Short toes	3.85%

Feature	%		Feature	%
Flat zygomatic region	7.69%		Micrognathia	11.54%
Prognathism	7.69%		Prominent philtrum	26.92%

Table S2. Correlation between deletion size and the presence of comorbidities (Mann-Whintey Test)

Feature	P value (29 patients)	P value (20 patients)	Feature	P value (29 patients)	P value (20 patients)
Increased pain tolerance	0.8501	0.9529	*Lymphedema*	0.2572	0.4912
Hypotonia	0.8501	0.9529	*Strabismus*	0.3739	0.7676
Recurring upper respiratory tract infections	0.3669	0.1806	*Cardiac abnormalities*	0.4926	0.5158
Gastroesophageal reflux	0.106	0.778	*Chewing difficult*	0.1257	0.1051
Sleep disturbance	0.942	0.395	*Swallowing difficult*	0.8419	0.8784
Seizures (febrile and/or non-febrile)			*Hypohidrose*	0.2154	0.1949
Constipation and/or diarrhea	0.5765	0.8916	*Precocious or delayed puberty*	0.911	0.9412
Renal abnormalities	0.01137	0.03664	*Hearing loss*	0.1379	0.1

References

Bonaglia, M.C., Giorda, R., Beri, S., De Agostini, C., Novara, F., Fichera, M., Grillo, L., Galesi, O., Vetro, A., Ciccone, R., Bonati, M.T., Giglio, S., Guerrini, R., Osimani, S., Marelli, S., Zucca, C., Grasso, R., Borgatti, R., Mani, E., Motta, C. and Zuffardi, O. 2011. Molecular mechanisms generating and stabilizing terminal 22q13 deletions in 44 subjects with Phelan/McDermid syndrome. *PLoS Genetics* 7(7), p. e1002173.

Boccuto, L., Lauri, M., Sarasua, S.M., Skinner, C.D., Buccella, D., Dwivedi, A., Orteschi, D., Collins, J.S., Zollino, M., Visconti, P., Dupont, B., Tiziano, D., Schroer, R.J., Neri, G., Stevenson, R.E., Gurrieri, F. and Schwartz, C.E. 2013. Prevalence of SHANK3 variants in patients with different subtypes of autism spectrum disorders. *European Journal of Human Genetics* 21(3), pp. 310–316.

Cochoy, D.M., Kolevzon, A., Kajiwara, Y., Schoen, M., Pascual-Lucas, M., Lurie, S., Buxbaum, J.D., Boeckers, T.M. and Schmeisser, M.J. 2015. Phenotypic and functional analysis of SHANK3 stop mutations idrentified in individuals with ASD and/or ID. *Molecular autism* 6, p. 23.

Costales, J.L. and Kolevzon, A. 2015. Phelan-McDermid Syndome and SHANK3: Implications for Treatment. *Neurotherapeutics* 12(3), pp. 620–630.

Dhar, S.U., del Gaudio, D., German, J.R., Peters, S.U., Ou, Z., Bader, P.I., Berg, J.S., Blazo, M., Brown, C.W., Graham, B.H., Grebe, T.A., Lalani, S., Irons, M., Sparagana, S., Williams, M., Phillips, J.A., Beaudet, A.L., Stankiewicz, P., Patel, A., Cheung, S.W. and Sahoo, T. 2010. 22q13.3 deletion syndrome: clinical and molecular analysis using array CGH. *American Journal of Medical Genetics. Part A* 152A(3), pp. 573–581.

Disciglio, V., Lo Rizzo, C., Mencarelli, M.A., Mucciolo, M., Marozza, A., Di Marco, C., Massarelli, A., Canocchi, V., Baldassarri, M., Ndoni, E., Frullanti, E., Amabile, S., Anderlid, B.M., Metcalfe, K., Le Caignec, C., David, A., Fryer, A., Boute, O., Joris, A., Greco, D. and Renieri, A. 2014. Interstitial 22q13 deletions not involving SHANK3 gene: a new contiguous gene syndrome. *American Journal of Medical Genetics. Part A* 164A(7), pp. 1666–1676.

Harony-Nicolas, H., De Rubeis, S., Kolevzon, A. and Buxbaum, J.D. 2015. Phelan mcdermid syndrome: from genetic discoveries to animal models and treatment. *Journal of Child Neurology* 30(14), pp. 1861–1870.

Kolevzon, A., Angarita, B., Bush, L., Wang, A.T., Frank, Y., Yang, A., Rapaport, R., Saland, J., Srivastava, S., Farrell, C., Edelmann, L.J. and Buxbaum, J.D. 2014. Phelan-McDermid syndrome: a review of the literature and practice parameters for medical assessment and monitoring. *Journal of neurodevelopmental disorders* 6(1), p. 39.

Leblond, C.S., Nava, C., Polge, A., Gauthier, J., Huguet, G., Lumbroso, S., Giuliano, F., Stordeur, C., Depienne, C., Mouzat, K., Pinto, D., Howe, J., Lemière, N., Durand, C.M., Guibert, J., Ey, E., Toro, R., Peyre, H., Mathieu, A., Amsellem,

F. and Bourgeron, T. 2014. Meta-analysis of SHANK Mutations in Autism Spectrum Disorders: a gradient of severity in cognitive impairments. PLoS Genetics 10(9), p. e1004580.

Mitz, A.R., Philyaw, T.J., Boccuto, L., Shcheglovitov, A., Sarasua, S.M., Kaufmann, W.E. and Thurm, A. 2018. Identification of 22q13 genes most likely to contribute to Phelan McDermid syndrome. European Journal of Human Genetics 26(3), pp. 293–302.

Monteiro, P. and Feng, G. 2017. SHANK proteins: roles at the synapse and in autism spectrum disorder. *Nature Reviews. Neuroscience* 18(3), pp. 147–157.

Moreira, D.P., Griesi-Oliveira, K., Bossolani-Martins, A.L., Lourenço, N.C.V., Takahashi, V.N.O., da Rocha, K.M., Moreira, E.S., Vadasz, E., Meira, J.G.C., Bertola, D., O'Halloran, E., Magalhães, T.R., Fett-Conte, A.C. and Passos-Bueno, M.R. 2014. Investigation of 15q11-q13, 16p11.2 and 22q13 CNVs in autism spectrum disorder Brazilian individuals with and without epilepsy. Plos One 9(9), p. e107705.

Oberman, L.M., Boccuto, L., Cascio, L., Sarasua, S. and Kaufmann, W.E. 2015. Autism spectrum disorder in Phelan-McDermid syndrome: initial characterization and genotype-phenotype correlations. *Orphanet Journal of Rare Diseases* 10, p. 105.

Phelan, K. and McDermid, H.E. 2011. The 22q13.3 Deletion Syndrome (Phelan-McDermid Syndrome). *Molecular syndromology* 2, pp 186-201.

Sarasua, S.M., Boccuto, L., Sharp, J.L., Dwivedi, A., Chen, C.-F., Rollins, J.D., Rogers, R.C., Phelan, K. and DuPont, B.R. 2014. Clinical and genomic evaluation of 201 patients with Phelan-McDermid syndrome. *Human Genetics* 133(7), pp. 847–859.

Sarasua, S.M., Dwivedi, A., Boccuto, L., Rollins, J.D., Chen, C.-F., Rogers, R.C., Phelan, K., DuPont, B.R. and Collins, J.S. 2011. Association between deletion size and important phenotypes expands the genomic region of interest in Phelan-McDermid syndrome (22q13 deletion syndrome). *Journal of Medical Genetics* 48(11), pp. 761–766.

de Sena Cortabitarte, A., Degenhardt, F., Strohmaier, J., Lang, M., Weiss, B., Roeth, R., Giegling, I., Heilmann-Heimbach, S., Hofmann, A., Rujescu, D., Fischer, C., Rietschel, M., Nöthen, M.M., Rappold, G.A. and Berkel, S. 2017. Investigation of SHANK3 in schizophrenia. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics* 174(4), pp. 390–398.

Soorya, L., Kolevzon, A., Zweifach, J., Lim, T., Dobry, Y., Schwartz, L., Frank, Y., Wang, A.T., Cai, G., Parkhomenko, E., Halpern, D., Grodberg, D., Angarita, B., Willner, J.P., Yang, A., Canitano, R., Chaplin, W., Betancur, C. and Buxbaum, J.D. 2013. Prospective investigation of autism and genotype-phenotype correlations in 22q13 deletion syndrome and SHANK3 deficiency. *Molecular autism* 4(1), p. 18.

Tabet, A.-C., Rolland, T., Ducloy, M., Lévy, J., Buratti, J., Mathieu, A., Haye, D., Perrin, L., Dupont, C., Passemard, S., Capri, Y., Verloes, A., Drunat, S., Keren, B., Mignot, C., Marey, I., Jacquette, A., Whalen, S., Pipiras, E., Benzacken, B. and Bourgeron, T. 2017. A framework to identify contributing genes in patients with Phelan-McDermid syndrome. *npj Genomic Medicine* 2, p. 32.

Chapter 4

General discussion and main conclusions

Autism spectrum disorders (ASD) correspond to a group of neurodevelopmental disorders that are characterized by stereotyped behavior, social communication and interaction issues, and restricted interest (DSM-5, Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition). ASD has a strong genetic component and copy number variations (CNVS) are pointed out as one of the most important genetic factors (Pinto et al. 2014; Shishido et al. 2014; Yingjun et al. 2017). It is estimated that it affects roughly 1% of the world population and the majority of these cases (about 75%) remain unsolved (Fernandez and Scherer 2017). Thus, in order to characterize both small and large size CNVs in a Brazilian cohort, in this work, we further validated a previously customized array-CGH 180k (Silva 2017) and characterized rare CNVs found in a cohort consisting of 121 ASD patients.

First, we improved the customized aCGH specificity by reanalyzing the patient's data and keeping only experiments that were previously conducted by Silva (2017) with DLRS lower than 0.2 for validation (before, values lower than 0.3 were accepted) and CNVs larger than 400pb (based on the smallest CNV validated by qPCR technique) that were unique to our cohort. With these new parameters, the aCGH specificity significantly improved for CNVs larger than 1Kb and 65 more patients were tested, adding up to 105 Brazilian patients with ASD. The overall detection rate (15%) was higher than the previous version of this customized aCGH, possibly associated with the higher number of probes and targeted genes (1527) included, and, if used as first-tier test, our detection rate is consistent with literature reported higher detection rate (Hnoonual et al. 2017).

Although other techniques, such as whole genome sequencing (WES), will probably replace the CMA technique, the cost remains prohibitive, particularly for patients in developing countries. Thus, our customized platform represents a strategy to detect genetic alterations in ASD patients that is affordable to our population.

In total, we described 4 pathogenic, 26 potentially pathogenic and 15 VOUs with no sub-classification. Altogether, this data provided further evidence on the

contribution of 27 candidate genes to ASD etiology (nineteen of them are SFARI candidate and eight are non-Sfari candidate genes). We also identified fifteen rare CNVs that have never been described in ASD patients, thus reinforcing the relevance of screening different populations so as to uncover ASD etiology (Hnoonual et al. 2017).

We also provided evidence that the larger CNVs found in our cohort are not polymorphisms of our population. This type of validation is challenging for smaller CNVs, because of the scarce information about them (in both public databases and in the Brazilian cohort of 1,019 NDD patients) and, so, more studies are necessary to elucidate the contribution of this type of CNVs to ASD. Moreover, we found overlap of ten CNVs among the ASD patients and the NDD cohort, further corroborating the relevance of six ASD candidate genes and providing further evidence of nervous system disruption for genes with less (*PTPRN2, CDH13*) or no evidence (*TRIM16*) of ASD association. Thus, our data supports the idea of a shared molecular etiology among different NDDs (Jensen and Girirajan 2017; Chen et al. 2014).

Finally, considering syndromic ASD cases, Phelan-McDermid Syndrome (PMS) is among the most frequently occurring syndromes among ASD patients. Despite the efforts of different researchers, the frequency and molecular basis of several comorbidities present in this syndrome, including ASD, remain unclear (Harony-Nicolas et al. 2015; Kolevzon et al. 2014; Mitz et al. 2018; Oberman et al. 2015; Tabet et al. 2017; Uchino and Waga 2013). Thus, as more patients are described in the literature, both clinical characterization and genotype-phenotype correlation ruling the clinical variability seen among PMS patients will become more accurate. Accordingly, in our cohort consisting of 34 patients, we described our main clinical and genetic findings (such as higher frequency of scarce eyebrows in PMS patients than it has been previously reported) and explored genotype-phenotype correlations, corroborating the association of renal abnormalities and speech status with CNV size.

All in all, this data helps to better characterize the contribution of CNVs in Brazilian cohort of ASD patients and to characterize the Phelan McDermid syndrome in Brazil. In short, in this work, we:

1. Further validated a customized array-CGH 180K;

2. Described both small and larger rare CNVs that can contribute to ASD (and provide evidence that the larger ones are not polymorphisms of our population by screening a larger Brazilian cohort);
3. Reinforced the contribution of 26 candidate genes for ASD;
4. Identified 15 CNVs that have never been reported in ASD cases;
5. Pointed out 9 CNVs that are shared among patients with different NDDs, including ASD and ID;
6. Clinically and genetically characterized a cohort of Brazilian patients with PMS;
7. Contributed for further understanding the genotype-phenotype correlation in PMS.
8. Contributed to estimating ASD occurrence in patients with PSM
9. Estimated the frequency of CNV occurrence in *SHANK3* in a Brazilian ASD cohort and in another Brazilian cohort consisting of patients with different NDDs.

References

American Psychiatric Association 2013. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Washington, DC: American Psychiatric Association.

Chen, E.S., Gigek, C.O., Rosenfeld, J.A., Diallo, A.B., Maussion, G., Chen, G.G., Vaillancourt, K., Lopez, J.P., Crapper, L., Poujol, R., Shaffer, L.G., Bourque, G. and Ernst, C. 2014. Molecular convergence of neurodevelopmental disorders. American Journal of Human Genetics 95(5), pp. 490–508.

Fernandez, B.A. and Scherer, S.W. 2017. Syndromic autism spectrum disorders: moving from a clinically defined to a molecularly defined approach. Dialogues in Clinical Neuroscience 19(4), pp. 353–371.

Harony-Nicolas, H., De Rubeis, S., Kolevzon, A. and Buxbaum, J.D. 2015. Phelan mcdermid syndrome: from genetic discoveries to animal models and treatment. *Journal of Child Neurology* 30(14), pp. 1861–1870.

Hnoonual, A., Thammachote, W., Tim-Aroon, T., Rojnueangnit, K., Hansakunachai, T., Sombuntham, T., Roongpraiwan, R., Worachotekamjorn, J., Chuthapisith, J., Fucharoen, S., Wattanasirichaigoon, D., Ruangdaraganon, N., Limprasert, P. and Jinawath, N. 2017. Chromosomal microarray analysis in a cohort of underrepresented population identifies SERINC2 as a novel candidate gene for autism spectrum disorder. *Scientific reports* 7(1), p. 12096.

Jensen, M. and Girirajan, S. 2017. Mapping a shared genetic basis for neurodevelopmental disorders. *Genome Medicine* 9(1), p. 109.

Kolevzon, A., Angarita, B., Bush, L., Wang, A.T., Frank, Y., Yang, A., Rapaport, R., Saland, J., Srivastava, S., Farrell, C., Edelmann, L.J. and Buxbaum, J.D. 2014. Phelan-McDermid syndrome: a review of the literature and practice parameters for medical assessment and monitoring. *Journal of neurodevelopmental disorders* 6(1), p. 39.

Mitz, A.R., Philyaw, T.J., Boccuto, L., Shcheglovitov, A., Sarasua, S.M., Kaufmann, W.E. and Thurm, A. 2018. Identification of 22q13 genes most likely to contribute to Phelan McDermid syndrome. *European Journal of Human Genetics* 26(3), pp. 293–302.

Oberman, L.M., Boccuto, L., Cascio, L., Sarasua, S. and Kaufmann, W.E. 2015. Autism spectrum disorder in Phelan-McDermid syndrome: initial characterization and genotype-phenotype correlations. *Orphanet Journal of Rare Diseases* 10, p. 105.

Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., Thiruvahindrapuram, B., Xu, X., Ziman, R., Wang, Z., Vorstman, J.A.S., Thompson, A., Regan, R., Pilorge, M., Pellecchia, G., Pagnamenta, A.T., Oliveira, B., Marshall, C.R., Magalhaes, T.R., Lowe, J.K. and Scherer, S.W. 2014. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *American Journal of Human Genetics* 94(5), pp. 677–694.

Shishido, E., Aleksic, B. and Ozaki, N. 2014. Copy-number variation in the pathogenesis of autism spectrum disorder. *Psychiatry and Clinical Neurosciences* 68(2), pp. 85–95.

Silva, I. M. W. InDels e CNVs pequenas em pacientes com Transtorno do Espectro Autista[doi:10.11606/D.41.2017.tde-29052017-145246]. São Paulo : Instituto de Biociências, Universidade de São Paulo, 2017. Dissertação de Mestrado em Biologia (Genética).

Tabet, A.-C., Rolland, T., Ducloy, M., Lévy, J., Buratti, J., Mathieu, A., Haye, D., Perrin, L., Dupont, C., Passemard, S., Capri, Y., Verloes, A., Drunat, S., Keren, B., Mignot, C., Marey, I., Jacquette, A., Whalen, S., Pipiras, E., Benzacken, B. and Bourgeron, T. 2017. A framework to identify contributing genes in patients with Phelan-McDermid syndrome. *npj Genomic Medicine* 2, p. 32.

Uchino, S. and Waga, C. 2013. SHANK3 as an autism spectrum disorder-associated gene. *Brain & Development* 35(2), pp. 106–110.

Yingjun, X., Haiming, Y., Mingbang, W., Liangying, Z., Jiaxiu, Z., Bing, S., Qibin, Y. and Xiaofang, S. 2017. Copy number variations independently induce autism spectrum disorder. *Bioscience reports* 37(4).

Chapter 5

Abstract

Autism Spectrum Disorder (ASD) is a heterogeneous group of neurodevelopmental disorders that affect about 1% of the world population and has a strong genetic component. Stereotyped behavior and restricted interest, as well as social interaction and communication issues characterize ASD. Moreover, in 10% of the cases, ASD occurs as a secondary condition in addition to a syndrome, such as Phelan-McDermid syndrome (PMS), which is associated with great clinical variability. Among genetic factors, copy number variations (CNVs) are one of the most important. However, the clinical significance of many CNVs remains unclear and there is an underrepresentation of small ASD associated CNVs in the literature. In this context, this project aimed at 1) characterizing large and small CNVs in Brazilian patients with ASD, by using a previously customized array-CGH in our laboratory; 2) clinically and genetically describing a cohort of Brazilian patients with PMS, as well as determining the frequency of this syndrome among Brazilian patients with ASD and other neurodevelopmental disorders. As a result of that, we 1) further validated the customized array-CGH, 2) provided additional evidence of association with ASD for 27 candidate genes, 3) described 15 CNVs that have never before been reported in the literature in association with this disorder, 4) presented evidence that roughly 70% of CNVs found in our cohort are not polymorphism of our population, and 5) reinforced the idea of shared molecular pathways among different neurodevelopmental disorders. In addition to that, for the first time, we described a Brazilian cohort of PMS patients and contributed to the molecular and clinical characterization of this syndrome. We also provided additional evidence of genotype-phenotype association regarding the presence of renal problems and speech status in PMS patients and estimated the frequency of occurrence of this syndrome among Brazilian patients with ASD and intellectual disability (syndromic or not). With these results, we hope to lead to a better understanding of the etiology of ASD and PMS, particularly among our population.

Resumo

O Transtorno do Espectro Autista (TEA) corresponde ao um grupo heterogêneo de alterações no neurodesenvolvimento que afeta cerca de 1% da população mundial e apresenta um forte componente genético. O TEA é caracterizado pela presença de comportamento estereotipado e interesses restritos, além de problemas de interação social e comunicação. Além disso, em 10% dos casos, o TEA ocorre como uma condição secundária somada a uma síndrome. Um exemplo é a síndrome de Phelan-McDermid (PMS), associada a uma grande variabilidade clínica. Dentre os fatores genéticos, as variações no número de cópias (CNVs) são um dos mais importantes. No entanto, o significado clínico de muitas CNVs permanece incerto, além de haver juma sub-representação de CNVs pequenas associadas ao TEA na literatura. Dentro deste contexto, este projeto teve como objetivos 1) caracterizar CNVs grandes e pequenas em pacientes brasileiros com TEA utilizando uma lâmina de array-CGH previamente customizada no Laboratório de Genética do Desenvolvimento – USP. 2) descrever clínica e geneticamente uma casuística de pacientes brasileiros com PMS, bem como determinar a frequência desta síndrome em pacientes com TEA e com outras alterações de neurodesenvolvimento. Como resultados, nós 1) validamos a lâmina customizada,

2) fornecemos evidencia adicional de associação com o TEA para 27 genes, 3) descrevemos 15 CNVs nunca reportadas em associação com o transtorno 4) apresentamos evidências de que cerca de 70% das CNVs encontradas em nossa coorte não são polimorfismo de nossa população e 5) reforçamos a ideia de vias moleculares compartilhadas entre diferentes alterações do neurodesenvolvimento. Além disso, descrevemos pela primeira vez uma casuística brasileira de pacientes com PMS e contribuímos para a caracterização molecular e clínica da síndrome. Fornecemos evidência adicional de associação genótipo-fenótipo no que diz respeito à presença de problemas renais e status de fala em pacientes com PMS e estimamos a frequência da síndrome entre pacientes brasileiros com TEA e com deficiência intelectual (sindrômica ou não). Com estes resultados, esperamos ter contribuído para o entendimento da etiologia tanto do TEA, quanto da PMS, sobretudo na nossa população.