ARTICLE

Three generation families: Analysis of de novo variants in autism

Claudia I. Samogy Costa[1,4], Gabriele da Silva Campos[1,4], Eduarda Morgana da Silva Montenegro[1,4], Jaqueline Yu Ting Wang[1], Marília Scliar[1], Frederico Monfardini[1], Elaine Cristina Zachi[1], Naila C. V. Lourenço[1], Ada J. S. Chan[2], Sergio L. Pereira[2], Worrawat Engchuan[2], Bhooma Thiruvahindrapuram[2], Mehdi Zarrei[2], Stephen W. Scherer[2,3] and Maria Rita Passos-Bueno[1]✉

De novo variants (DNVs) analysis has proven to be a powerful approach to gene discovery in Autism Spectrum Disorder (ASD), which has not yet been shown in a Brazilian ASD cohort. The relevance of inherited rare variants has also been suggested, particularly in oligogenic models. We hypothesized that three-generation analyses of DNVs could provide new insights into the relevance of de novo and inherited variants across generations. To accomplish this goal, we performed whole-exome sequencing of 33 septet families composed of probands, parents, and grandparents ($n = 231$ individuals) and compared DNV rates (DNVr) between generations and those from two control cohorts. The DNVr in the probands (DNVr = 1.16) was marginally higher than in parents (DNVr = 0.60; $p = 0.054$), and in controls (DNVr = 0.68; $p = 0.035$, congenital heart disorder and DNVr = 0.70; $p = 0.047$, unaffected ASD siblings from Simons Simplex Collection). Moreover, most of the DNVs were found to have paternal origin in both generations (84.6%). Finally, we observed that 40% (6/15) of the DNVs in parents transmitted for probands are in ASD or ASD candidate genes, representing recently emerged risk variants to ASD in their families and suggest *ZNF536*, *MSL2* and *HDAC9* as ASD candidate genes. We did not observe an enrichment of risk variants nor sex bias of transmitted variants in the three generations, that can be due to sample size. These results further reinforce the relevance of de novo variants in ASD.

European Journal of Human Genetics (2023) 31:1017–1022; https://doi.org/10.1038/s41431-023-01398-6

INTRODUCTION

Autism spectrum disorders are mainly characterized by difficulty in social and communication skills, besides restricted/repetitive behaviors and interests [1]. The prevalence of ASD is estimated as at least 1% (reviewed in Lord et al. [2]). Genomic alterations commonly associated with ASD etiology include copy number variants (CNVs), single nucleotide variants (SNVs) and small insertions/deletions (Indels) and the percentage of de novo variants is higher (2 to 3 folds or even higher) in ASD individuals as compared to control population [3–7]. Of note, the rate of the de novo variants in ASD individuals' parents is unknown. The analysis of de novo variants has proven to be a powerful approach in ASD gene discovery [6, 8–10]. However, they do not fully explain phenotype in the most part of the ASD cases given the complex genetic architecture of ASD. Indeed, these patients also present more deleterious inherited variants than their non-affected siblings, favoring oligogenic models as an important mechanism contributing to this phenotype [11–13], where de novo mutation in previous generations can contribute to an accumulation of inherited variants of medium and low effect associated with incomplete penetrance.

Parental increased age has also been associated with higher rates of de novo variants, in both healthy, and disease-associated cohorts, including ASD [14–17]. Considering the increased number of de novo variants in ASD individuals and their advanced parental age, we hypothesized that three-generation analysis of de novo variants could provide new insights into the clinical relevance of variants that have arisen de novo across generations. Therefore, our main goal was to investigate if the proportion of total and damaging de novo variants (DNVs) varies across generations, and verify if the total DNV rates correlate to parental age at conception. We also compared the DNV rates to available controls in the literature: healthy siblings of ASD probands [4] and individuals with congenital heart disorder [18, 19]. We also analyzed the clinical relevance of damaging variants among the DNVs that arose in the parents and probands. Finally, we determined whether there is an accumulation of DNVs plus inherited risk variants through generations or a sex bias in the rare variants' segregation through three generations.

METHODS

Subjects and DNA samples

We collected DNA samples from 33 different families ascertained at the Centro de Estudos do Genoma Humano e Células-Tronco (CEGH-CEL), Instituto de Biociências, Universidade de São Paulo, composed of ASD probands, parents and grandparents, herein referred as septets (initial

[1]Departamento de Genética e Biologia Evolutiva, Centro de Estudos do Genoma Humano e Células-tronco, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brasil. [2]The Centre for Applied Genomics, Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada. [3]Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, ON, Canada. [4]These authors contributed equally: Claudia I. Samogy Costa, Gabriele da Silva Campos, Eduarda Morgana da Silva Montenegro. ✉email: passos@ib.usp.br

Received: 19 September 2022 Revised: 10 May 2023 Accepted: 17 May 2023
Published online: 6 June 2023

number of individuals = 231). The main criteria of inclusion of families in our study were: ASD diagnosis based on the current DSM at the time of evaluation, non-syndromic ASD [20], and living grandparents. We clinically evaluated all probands, reviewed parents' and grandparents' age at conception based on date of birth and collected familial histories of neuropsychiatric/neurological disorders. All male probands were negative for Fragile X testing. DNA samples were extracted from whole blood samples in most cases, and for cases of grandparents living in long distances, saliva samples were used.

Exome sequencing and de novo variants rates in probands, parents and controls

All the DNA samples were prepared using the SureSelect QXT Target Enrichment for Illumina Multiplexed Sequencing (Illumina, Inc., California, USA), following the manufacturer's recommendations. The libraries were sequenced with Illumina HiSeq 2500 sequencer in paired end reads of approximately 100 bp. Sequence alignments to the human genome reference (UCSC hg19) were performed with Burrows-Wheeler Aligner, data processing and variant calling were performed with Picard and Genome Analysis Toolkit package (GATK), and variants annotation was performed with ANNOVAR. The mean coverage of exome data from the 231 individuals were 63X, with approximately 80% of the exome being covered by 20 reads, and 70% by 30 reads. The annotation of high confidence de novo variants was performed with DeNovoGear software [21], in addition to a visual inspection of all the identified de novo variants in the software Integrative Genomic Viewer (IGV [22]). For Sanger sequencing of 18 de novo variants, we used BigDye Terminator v3.1 Cycle Sequencing Kit, in the ABI 3730 DNA Analyzer (Thermo Fisher Scientific, Massachusetts, USA). Variants' frequency in the general population was filtered for less than or equal to 1% using GnomAD (https://gnomad.broadinstitute.org/), 1000 Genomes (https://www.internationalgenome.org/) and ABraOM (https://abraom.ib.usp.br/) databases.

For the control population analysis, we obtained de novo variants (missense and loss of function - LoF), in two available whole genome sequencing (WGS) cohorts of non-ASD individuals: 1911 unaffected siblings of ASD patients from Simons Simplex Collection (SSC)[4]; and 2072 individuals with Congenital Heart Defects (CHD) [18, 19], from the international database "denovo-db" (http://denovo-db.gs.washington.edu/denovo-db/index.jsp). We compared the rates (total number of variants divided by the total number of individuals) of de novo variants among probands, parents and these two non-affected ASD cohorts through Fisher's exact test. Variants' prioritization criteria are detailed in Supplementary material.

De novo variants origin, variants classification, parental age

We performed fragment-based phasing algorithm to infer the parental origin of DNVs using dng phaser tool from DeNovoGear software [21]. The analysis parameters were adapted for our exome data, considering SNPs found at 300 pb or 1000 pb distance from the DNVs.

De novo missense (CADD ≥ 20) and LoF variants in genes with brain expression (16,465 genes from the Human Protein Atlas database, HPA, www.proteinatlas.org/) were defined as damaging variants. Regression analysis was performed through decision trees in R software (R Core Team 2019) to evaluate parental age at conception differences, regarding gender and generation, for both parents and grandparents of ASD individuals in our cohort, using the libraries partykit and rpart for the models. Finally, we performed Poisson regression analysis to test whether there is a correlation between the number of de novo variants in the offspring and parental age. Detailed information on variants classification and clinical significance is available in Supplementary material.

Inherited variants transmission and analysis

We also characterized the transmitted variants across generations. For this analysis, we selected ultra-rare (≤ 0.1%) LoF and missense (CADD ≥ 30) variants, in autosomal, brain expressed genes, with pLI ≥ 0.9 (or z-score ≥ 3 for genes with missense variants). All variants were visually curated using IGV [22] and the selected variants are referred as inherited risk variants. We performed Fisher's exact test to determine whether there was a sex bias in variants' transmission. To test the hypothesis that the risk variants accumulate through generations, we performed a Wilcoxon signed-ranks test comparing probands' and grandparents' average number of risk variants (Supplementary material).

RESULTS

Cohort characterization

Our cohort was composed of 9 females and 24 males diagnosed with non-syndromic ASD (Supplementary Table 1), besides their parents and grandparents ($n = 231$ individuals). Simplex families cases represented 29 out of 33 families (considering one case of monozygotic concordant twins, and three probands with an affected sibling; only probands were included in the sample). Kinship and quality analysis using whole exome sequencing (WES) data revealed ten trios with sample issues (one proband trio and nine parents' trios), which resulted in their exclusion. Only complete trios that fulfilled the established quality criteria were kept in the subsequent analysis, and therefore, we included three sets of trios: 1) proband, mother and father, $n = 32$ trios; b) proband's mother and maternal grandparents, $n = 30$ trios; c) proband's father and paternal grandparents, $n = 27$ trios (Supplementary Table 2). For inherited risk variants, the four families in which the ASD proband carries a pathogenic or likely pathogenic variant in ASD genes (P3, P7, P10, and P15) were also excluded, totalizing 28 probands, 26 mothers, 23 fathers, 49 grandmothers and 49 grandfathers (26 maternal and 23 paternal).

DNVs rates in probands, parents and controls

We identified 37 de novo variants in 20 probands (6 LoF and 31 missense; Table 1, Supplementary Table 3), 20 de novo variants in 15 mothers (1 LoF and 19 missense; Supplementary Table 4), and 14 de novo variants in 10 fathers (3 LoF and 11 missense; Supplementary Table 5). Of note, we validated by Sanger sequencing 100% (18/18) of the set of tested de novo variants in probands (Supplementary Table 3). We observed a significant increase in the combined rate of de novo (DNVr) LoF and missense variants in ASD probands (37/32; DNVr =1.16), compared to congenital heart disease individuals (1,416/2,072; DNVr = 0.68; Fisher's test $p = 0.035$), and non-ASD affected siblings from SSC (1,333/1,911; DNVr = 0.70; Fisher's test $p = 0.047$). We also observed a marginally statistically significant (at 5% level) increase (34/57; DNVr = 0.60; Fisher's test $p = 0.054$) when we compared to DNVr in parents.

De novo variants: Origin and clinical relevance in two generations

Proband's analysis: We were able to infer the parental origin for 16% (6/37) of the DNVs in the probands, in which five had paternal origin and one maternal. Of the 37 variants, 12 were in genes with pLI ≥ 0.9. Five out of the 37 variants were classified as likely pathogenic or pathogenic: four in ASD genes (*WDFY3*, *BRSK2*, *PACS2* and *KAT6A*) and one in the non-associated ASD gene, *EVC* (#MIM193530 and #MIM225500, respectively). *EVC* gene is associated with autosomal dominant (AD) and recessive (AR) conditions, but the patient did not present any clinical features of the AD phenotype when evaluated.

Fourteen out of the 32 remaining DNVs, all classified as VUS, are located in ASD candidate genes (Supplementary Table 3). Five of fourteen were located in genes with pLI ≥ 0.9.Of note, the patient P13-1 harbors two DNVs acting in cis on *MSL2*, recently suggested as an ASD candidate gene [23]. We also highlight, based on their relevance to brain function and being subject to purifying selection (pLI ≥ 0.9), the *ZNF536* and *HDAC9* genes. Therefore, 18 out of 37 (49%) DNVs in probands are pathogenic/likely pathogenic or VUS in genes relevant to ASD.

Parent's analysis: We were able to infer the parental origin for 20% (7/34) of the DNVs in the parents (four in mothers, one had maternal origin and three DNVs had paternal origin; three in the father's group, being all three of paternal origin). Of the 34 DNVs in the parents, 10 were in genes with pLI ≥ 0.9. Two variants (6%, 2/34) in ASD (*CC2D1A*) or ASD candidate (*CHD5*) genes were classified as pathogenic or likely pathogenic (one in a mother, one in a father). Detailed clinical features were not available for

Table 1. Summary of highlighted de novo variants identified in our probands cohort.

PI	Gene	SFARI score	Chr	Start Position	Accession	Variant	[b]gnomAD metrics	ACMG classification	[c]DNVs in parents
P1–1[a]	*MTR*	NA	1	237025561	NM_001291939.1	c.2069 A > G:p.(Lys690Arg)	z = 2.04	VUS	No
P1–1[a]	*HDAC9*	NA	7	18833005	NM_001321897.2	c.2120 C > G:p.(Ser707Cys)	z = 2.06	VUS	No
P3–1	*WDFY3*	1	4	85636495	NM_014991.6	c.7917 C > A:p.(Tyr2639Ter)	pLI = 1	Pathogenic	No
P3–1	*TBL1X*	3	X	9656072	NM_001139467.1	c.220 G > A:p.(Asp74Asn)	z = 2.48	VUS	No
P5–1	*TMEM39B*	3	1	32557419	NM_001319677.1	c.353 C > T:p.(Tyr118Met)	z = 1.85	VUS	No
P7–1	*BRSK2*	1	11	1475787	NM_003957.4	c.1617A > C:p.(Lys539Asn)	z = 3.94	Likely pathogenic	Yes, *SLC15A4*
P10–1	*KAT6A*	2 S	8	41798361	NM_006766.5	c.3038 A > G:p.(Lys1013Arg)	z = 2.07	Likely pathogenic	Yes, *IGF2R*
P13–1	*MSL2*	NA	3	135870081	NM_001145417.2	c.1420 G > C:p.(Val474Leu)	z = 2	VUS	Yes, *NBAS*
P13–1	*MSL2*	NA	3	135870098	NM_001145417.2	c.1403 G > T:p.(Ser468Ile)	z = 2	VUS	Yes, *NBAS*
P15–1	*PACS2*	S	14	105834449	NM_001243127.3	c.424 G > A:p.(Glu142Leu)	z = 2.24	Pathogenic	Yes, *C1orf112*
P15–1	*INSYN1*	NA	15	74032592	NM_001039614.3	c.548 G > A:p.(Arg183Gln)	z = 0.83	VUS	Yes, *C1orf112*
P24–1	*SPHKAP*	NA	2	228855786	NM_001142644.2	c.4889 G > A:p.(Trp1630Ter)	pLI = 0.94	VUS	No
P24–1	*ZNF536*	NA	19	30936040	NM_001352260.2	c.1571 G > A:p.(Trp524Ter)	pLI = 1	VUS	No
P28–1	*FBXL5*	NA	4	15627380	NM_001193534.1	c.1342 G > A:p.(Asp448Asn)	z = 1.85	VUS	Yes, *DYTN*
P28–1	*LINS1*	NA	15	101114236	NM_001352508.2	c.797 C > G:p.(Ser266Cys)	z = −0.68	VUS	Yes, *DYTN*
P31–1	*PDK1*	NA	2	173427015	NM_002610.5	c.406 T > C:p.(Tyr136His)	z = −4.32	VUS	No
P32–1[a]	*PLCL1*	NA	2	199011526	NM_006226.4	c.3128 A > G:p.(Asn1043Ser)	z = 1.86	VUS	No
P33–1	*CDH4*	NA	20	60508149	NM_001794.5	c.2346 G > C:p.(Lys782Asn)	z = 2.02	VUS	Yes, *LFNG*

PI Patient identification, *Chr* Chromosome, *NA* Not available, *NS* Not scored, *VUS* Variant of unknown significance, *DNV* de novo variant, *Class*. Classification.
[a]Familial cases.
[b]gnomAD metrics represent gene intolerance to missense (zscore) or loss of function (pLI) variants.
[c]Of note, we reported genes harboring de novo variants in parents that were transmitted to ASD probands.

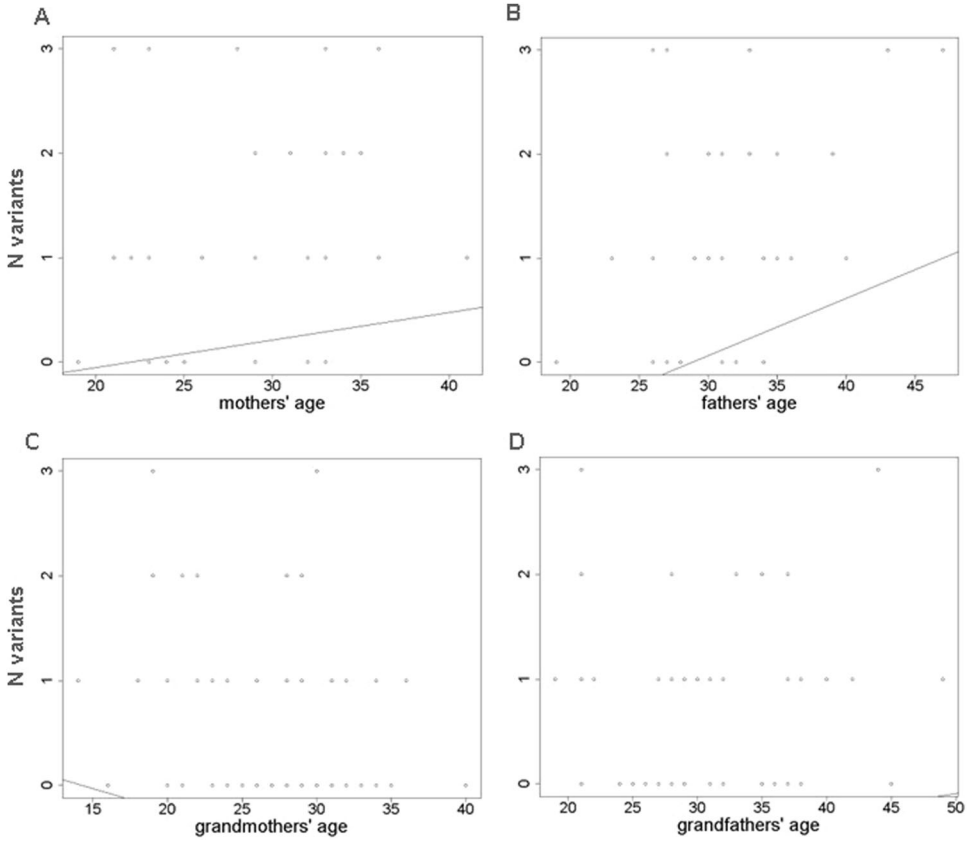

Fig. 1 Poisson regression analysis. Parental age correlation to the number of de novo variants identified in the offspring. **A**, **B** Maternal and paternal conception age correlation to the number of de novo variants in the offspring ($N = 31$ mothers and 27 fathers). **C**, **D** Paternal conception age correlation to the number of de novo variants in the offspring ($N = 58$ grandmothers and 58 grandfathers).

these subjects, but no ASD features were reported. Eleven of the remaining 32 DNVs (32%; 11/34) were classified as VUS for ASD (including one variant in the ASD gene *HNRNPF* and 10 in ASD candidate genes), of which five are located in genes with pLI ≥ 0.9 (Supplementary Tables 4 and 5). Therefore, the probands' parents present 13 (38%; 13/34) pathogenic/likely pathogenic or VUS DNVs in ASD (two variants) or ASD candidate genes (11 variants).

Altogether, 15 out of the 34 DNV variants found in the parents were transmitted (Supplementary Table 6; 9/20 from mothers and 6/14 from fathers), six of them in ASD (*CC2D1A*, potentially pathogenic, maternal origin) or ASD candidate (maternal origin: *RFWD3*, *FGD1*, and paternal origin: *IGF2R, EIF3E*, *LFNG*, all classified as VUS variants) genes. We did not observe a transmission bias of damaging DNVs (missense variants with CADD ⩾ 20 plus LoF variants; 11 transmitted and 14 non-transmitted, Fisher's exact test, $p = 0.777$) or non-damaging variants (four transmitted and five non-transmitted variants). Finally, the search of MSSNG database for other ASD individuals with DNVs in the 15 genes with DNVs identified in the parents that were transmitted to the probands, lead to the identification of additional ASD individuals with DNVs on *NBAS*, *ARHGEF28, IGF2R*, and *LFNG* (Supplementary Table S6), being the last two genes ASD candidate genes.

Parental age at conception and DNVs rates in our cohort

The mean conception age of mothers was 29.18 ± 5 years, compared to 25.85 ± 5 years for the grandmothers. Meanwhile, the mean conception age of fathers was 31.45 ± 6 years, compared to 29.35 ± 7 years in grandfathers (Supplementary Table 2). We observed on the regression decision tree model (Supplementary Fig. 1) that parental age at conception is differentiated by gender and generation (parents or grandparents). Females present reduced conception age compared to males, and probands' grandparents are younger than probands' parents ($p = 0.0008$). A significant correlation was observed only between the number of de novo variants and patient's paternal age (Poisson regression analysis, fathers' age p-value = 0.0358; mothers' age p-value = 0.382; grandfathers' age p-value = 0.452; grandmothers' age p-value = 0.195), Fig. 1. However, the positive association identified for the patient's fathers age was not observed after exclusion of an outlier father, P28-3 ($p = 0.129$).

Transmission patterns of inherited variants

We have performed an analysis of the inherited rare variants through 3 generations for an oligogenic model. In total, 111 different rare variants were selected (107 variants detected in grandparents, two DNVs in probands' parents and two DNVs in probands.;DNVs in probands and probands' parents generations were kept in this analysis, since it is not possible to distinguish the DNVs to inherited risk variants selected in the grandparents generation). The segregation of rare at-risk ASD variants were accordingly to expected to Mendelian law, with no sex bias of the transmitting parent: probands' parents, in addition of two DNVs, inherited 31/55 from grandfather and 24/55 from grandmothers (Fisher's exact test, $p = 0.57$) and probands, in addition of two DNVs, inherited 19/37 from fathers and 18/37 from mothers. In addition, we did not observe a significant difference in the average number of risk variants between grandparents (combined putative de novo and inherited, 99 variants/98 grandparents = 1.01) and probands (de novo and inherited variants, 39 variants/28 probands = 1.39; Wilcoxon test, $p = 0.16$, Supplementary fig. 2).

DISCUSSION

The three-generation families' variant analysis allowed us to determine the rate of de novo variants in non-affected parents of ASD individuals, which had lower DNVs rates (0.60) compared to probands (1.16), and similar rates compared to controls. Despite the differences in sample size and methodology for variant prioritization of other cohorts, DNVs rate in probands was comparable to the literature [4]. Moreover, probands presented 2.1-fold enrichment of variants in genes intolerant to loss of function variants (pLI ≥ 0.9) compared to their parents (probands = 12 variants in 32 individuals; parents = 10 variants in 57 individuals). Although direct comparisons are not possible due to differences in methodology and sample size, notably Satterstrom et al. [8]., using data from ASD affected individuals and their unaffected siblings, observed a 3.5-fold enrichment of DNV (pLI > 0.995) as compared to DNV in non-affected sibs. To our knowledge, this is the first study showing this type of data, and despite our small sample, we were able to observe a higher rate of DNVs between the probands and their parents, even though with a marginal significance level. Higher number of total and ASD relevant variants (here represented by variants in ASD or ASD candidate genes) were observed in probands, further supporting the relevance of de novo variants in the ASD genetic architecture.

We could determine the parental origin for 18% (13/71) of all DNVs identified (probands and their parents), comparably to the previous analysis using this same tool that identified approximately 20% of DNVs parental origin [21]. Also, 84.6% of the DNVs in both generations were of paternal origin, an expected number as seen by Jónsson et al. [15], that identified 80.4% of DNVs with paternal origin. Indeed, this data is in accordance with the observation that the age of the father is an important factor in determining the number of de novo SNVs variants in the offspring [17], although a positive association was not observed in our study possibly due to sample size.

Notably, 43% (6/15) of the transmitted parents' DNVs were in ASD (one gene - *CC2D1A*) or ASD candidate (five genes - *RFWD3*, *FGD1*, *IGF2R*, *EIF3E* and *LFNG*) genes. We observed that de novo variants in two of these ASD candidate genes, *IGF2R* and *LFNG*, were also found in ASD individuals from the MSSNG database. Despite the limited data of this study, the occurrence of de novo and recently emerged variants in ASD or ASD candidate genes in healthy parents can represent a source of genetic factors contributing to ASD following an oligogenic model, which still represents a challenge to be dissected [24, 25].

Probands' analysis of DNVs led to the identification of variants in previous and possible novel ASD candidate genes. We should bear in mind that it is still challenging to infer which gene or variants are indeed highly penetrant or represent medium or minor contributing factors. Nevertheless, our data adds to the literature a novel ASD case with a rare de novo variant in *BRSK2* [26]. Of note, we identified two additional de novo variants in *BRSK2* in the MSSNG database. In addition, we would like to highlight *ZNF536*, *MSL2* and *HDAC9* as new ASD candidate genes. *ZNF536* encodes a zinc finger protein specifically expressed in neuronal cells, that acts as a negative regulator of neural cell differentiation, being also discussed for its role in the maintenance of neuronal cells and development of forebrain neurons implicated in social behavior and stress [27, 28]. We also identified two de novo cis-variants on *MSL2* in one patient (P13-1). *MSL2* was recently reported as an ASD candidate [23], and encodes a protein involved in chromatin modification [29]. Finally, *HDAC9* encodes a histone deacetylase [30]. It has been demonstrated that *HDAC9* is associated with schizophrenia, being widely expressed in post-mitotic neurons, and may play an important role in mature neuron function [31].

The analysis of inherited risk variants in our sample did not reveal any transmission bias across generations or gender, although our sample size may have impacted these results. The relevance and patterns of transmitted variants for ASD has been approached by different authors, generally by comparing the burden of transmitted variants in ASD affected individuals and their siblings [32–34]. However, the origin and features of these variants across generations were not fully explored. More recently, Wilfert et al. (2021) [12] showed that variants transmitted to affected individuals are more recent than variants transmitted to their siblings with similar frequency and type, further suggesting that understanding the patterns of variants' segregation through generations may help understanding the genetic architecture of autism. Genomic larger studies of three-generations of ASD propositus could be a helpful approach to understanding the complex ASD genetic architecture and possibly with a better cost-effective effect than large case-control samples, as exemplified here with the observed difference in the DNV rate between probands and their parents despite our modest sample size.

In summary, our study showed for the first time DNVs rates in parents of ASD probands, and suggest enrichment of transmitted de novo risk variants to their offspring. We also reinforce the paternal origin of the vast majority of DNVs with paternal origin in both generations. Notably, in addition to de novo variants other mechanisms such as epigenetics might contribute to increased ASD risk with parental aging [16, 35]. In addition, we cannot rule out the contribution of damaging variants to ASD endophenotypes in the parents, which were not systematically evaluated. Finally, we describe novel variants in genes recently pointed out as ASD genes (*BRSK2*) and reinforce new ASD candidates (*HDAC9*, *MSL2* and *ZNF536*). In conclusion, our study showed that the analysis of three-generations presents a promising strategy to investigate the effect of de novo and inherited risk variants in ASD, with a better cost-effective effect than extreme large population studies.

DATA AVAILABILITY

Data generated as part of this study are available from the corresponding author on reasonable request. All de novo variants described in this work were submitted to the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/), accession numbers SCV003914734 - SCV003914804.

REFERENCES

1. Association AP, Kernberg. Diagnostic and Statistical Manual of Mental Disorders. 5th edition. Washington, D.C.: American Psychiatric Publishing; 2013.
2. Lord C, Brugha TS, Charman T, Cusack J, Dumas G, Frazier T, et al. Autism spectrum disorder. Nat Rev Dis Primers. 2020;6:1–23.
3. Woodbury-Smith M, Scherer SW. Progress in the genetics of autism spectrum disorder. Dev Med Child Neurol. 2018;60:445–51.
4. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014;515:216–21.
5. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, et al. Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders. Am J Hum Genet. 2014;94:677–94.
6. Yuen RK C, Merico D, Bookman ML, Howe J, Thiruvahindrapuram B, Patel RV, et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. Nat Neurosci. 2017;20:602–11.
7. Zarrei M, Burton CL, Engchuan W, Young EJ, Higginbotham EJ, MacDonald JR, et al. A large data resource of genomic copy number variation across neurodevelopmental disorders. NPJ Genom Med. 2019;4:26.
8. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell. 2020;180:568–84.e23.
9. Turner TN, Coe BP, Dickel DE, Hoekzema K, Nelson BJ, Zody MC, et al. Genomic patterns of De Novo mutation in simplex autism. Cell. 2017;171:710–22.e12.
10. da Silva Montenegro EM, Costa CS, Campos G, Scliar M, de Almeida TF, Zachi EC, et al. Meta-analyses support previous and novel autism candidate genes: Outcomes of an unexplored Brazilian cohort. Autism Res. 2020;13:199–206.
11. Ruzzo EK, Pérez-Cano L, Jung JY, Wang LK, Kashef-Haghighi D, Hartl C, et al. Inherited and De Novo genetic risk for autism impacts shared networks. Cell. 2019;178:850–66.e26.

12. Wilfert AB, Turner TN, Murali SC, Hsieh P, Sulovari A, Wang T, et al. Recent ultra-rare inherited variants implicate new autism candidate risk genes. Nat Genet. 2021;53:1125–34.
13. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, et al. Excess of rare, inherited truncating mutations in autism. Nat Genet. 2015;47:582–8.
14. Yuen RKC, Merico D, Cao H, Pellecchia G, Alipanahi B, Thiruvahindrapuram B, et al. Genome-wide characteristics of de novo mutations in autism. NPJ Genom Med. 2016;1:160271–10.
15. Jónsson H, Sulem P, Kehr B, Kristmundsdottir S, Zink F, Hjartarson E, et al. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. Nature. 2017;549:519–22.
16. Gao Z, Moorjani P, Sasani TA, Pedersen BS, Quinlan AR, Jorde LB, et al. Overlooked roles of DNA damage and maternal age in generating human germline mutations. Proc Natl Acad Sci USA. 2019;116:9491–500.
17. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. Nature. 2012;488:471–5.
18. Homsy J, Zaidi S, Shen Y, Ware JS, Samocha KE, Karczewski KJ, et al. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. Science. 2015;350:1262–6.
19. Sifrim A, Hitz MP, Wilsdon A, Breckpot J, Turki SHA, Thienpont B, et al. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. Nat Genet. 2016;48:1060–5.
20. Miles JH, Takahashi TN, Bagby S, Sahota PK, Vaslow DF, Wang CH, et al. Essential versus complex autism: Definition of fundamental prognostic subtypes. Am J Med Genet Part A. 2005;135A:171–80.
21. Ramu A, Noordam MJ, Schwartz RS, Wuster A, Hurles ME, Cartwright RA, et al. DeNovoGear: de novo indel and point mutation discovery and phasing. Nat Methods. 2013;10:985–7.
22. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24–6.
23. Du Y, Li Z, Liu Z, Zhang N, Wang R, Li F, et al. Nonrandom occurrence of multiple de novo coding variants in a proband indicates the existence of an oligogenic model in autism. Genet Med. 2020;22:170–80.
24. Wang T, Zhao PA, Eichler EE. Rare variants and the oligogenic architecture of autism. Trends Genet. 2022;38:895–903.
25. Pizzo L, Jensen M, Polyak A, Rosenfeld JA, Mannik K, Krishnan A, et al. Rare variants in the genetic background modulate cognitive and developmental phenotypes in individuals carrying disease-associated variants. Genet Med. 2019;21:816–25.
26. Hiatt SM, Thompson ML, Prokop JW, Lawlor JMJ, Gray DE, Bebin EM, et al. Deleterious Variation in BRSK2 Associates with a Neurodevelopmental Disorder. Am J Hum Genet. 2019;104:701–8.
27. Qin Z, Ren F, Xu X, Ren Y, Li H, Wang Y, et al. ZNF536, a novel zinc finger protein specifically expressed in the brain, negatively regulates neuron differentiation by repressing retinoic acid-induced gene transcription. Mol Cell Biol. 2009;29:3633–43.
28. Thyme SB, Pieper LM, Li EH, Pandey S, Wang Y, Morris NS, et al. Phenotypic landscape of schizophrenia-associated genes defines candidates and their shared functions. Cell. 2019;177:478–91.e20.
29. Wu L, Zee BM, Wang Y, Garcia BA, Dou Y. The RING finger protein MSL2 in the MOF complex is an E3 ubiquitin ligase for H2B K34 and is involved in crosstalk with H3 K4 and K79 methylation. Mol Cell. 2011;43:132–44.
30. Volmar CH, Wahlestedt C. Histone deacetylases (HDACs) and brain function. Neuroepigenetics. 2015;1:20–7.
31. Lang B, Alrahbeni TMA, Clair DS, Blackwood DH. HDAC9 is implicated in schizophrenia and expressed specifically in post-mitotic neurons but not in adult neural stem cells. Am J Stem Cells. 2012;1:31–41.
32. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, et al. Excess of rare, inherited truncating mutations in autism. Nat Genet. 2015;47:582–8.
33. Krumm N, O'Roak BJ, Karakoc E, Mohajeri K, Nelson B, Vives L, et al. Transmission disequilibrium of small CNVs in simplex autism. Am J Hum Genet. 2013;93:595–606.
34. Brandler WM, Antaki D, Gujral M, Kleiber ML, Whitney J, Maile MS, et al. Paternally inherited cis-regulatory structural variants are associated with autism. Science. 2018;360:327–31.
35. Janecka M, Mill J, Basson MA, Goriely A, Spiers H, Reichenberg A, et al. Advanced paternal age effects in neurodevelopmental disorders—review of potential underlying mechanisms. Translational Psychiatry. 2017;7:e1019.

ACKNOWLEDGEMENTS

The authors thank the ASD families that participated and the researchers that contributed indirectly to this study: Ana Cristina de Sanctis Girardi, Susan Walker and Giovanna Pellachia. We also thank Dr. André Fujita for support in statistical analysis. We wish to acknowledge support and resources of MSSNG (https://research.mss.ng), Autism Speaks, and The Centre for Applied Genomics, at The Hospital for Sick Children, Toronto, Canada. We also appreciate the contributions of Wilson WL Sung and Darwin D'Souza.

AUTHOR CONTRIBUTIONS

EMSM and MRPB designed the study. EMSM, GSC, and CISC performed data collection, analysis and helped write the manuscript. SWS conceived and designed MSSNG project and revised the manuscript. FM performed statistical analysis. MS helped with DeNovoGear pipeline and analysis for de novo variants identification. JYTW helped with data annotation and inherited risk variants analysis. AJSC, SLP, WE, BT, and MZ helped perform different components of analysis and data interpretations. MZ revised the manuscript. ECZ contributed to clinical evaluation. NCVL helped with data collection and additional genetic tests of the probands. All authors contributed to data discussion and interpretation of the results.

FUNDING

This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, CEGH-CEL/CEPID process 2013/08028-1, Ph.D. scholarship process 2017/05824-2 and 2018/13743-5), and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, process 466651/2014-7).

COMPETING INTERESTS

SWS is on the Scientific Advisory Committee of Population Bio, and serves as a Highly Cited Academic Advisor for the King Abdulaziz University. The remaining authors declare no conflicts of interest.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The ethics committee of the Instituto de Biociências at Universidade de São Paulo approved the project (accession number 1.133.486), and all the research participants signed a consent term, with written informed consent for publication of individual details obtained.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41431-023-01398-6.

Correspondence and requests for materials should be addressed to Maria Rita Passos-Bueno.