



OPEN

## Chromosomal microarray analyses from 5778 patients with neurodevelopmental disorders and congenital anomalies in Brazil

Ana C. V. Krepischi<sup>1,2,5</sup>, Darine Villela<sup>2,5</sup>, Silvia Souza da Costa<sup>1</sup>, Patricia C. Mazzonetto<sup>2</sup>, Juliana Schauen<sup>2</sup>, Michele P. Migliavacca<sup>2</sup>, Fernanda Milanezi<sup>2</sup>, Juliana G. Santos<sup>2</sup>, Gustavo Guida<sup>2</sup>, Rodrigo Guarischi-Sousa<sup>2</sup>, Gustavo Campana<sup>2</sup>, Fernando Kok<sup>3</sup>, David Schlesinger<sup>3</sup>, Joao Paulo Kitajima<sup>3</sup>, Francine Campagnari<sup>3</sup>, Debora R. Bertola<sup>1,4</sup>, Angela M. Vianna-Morgante<sup>1</sup>, Peter L. Pearson<sup>1</sup> & Carla Rosenberg<sup>1,2</sup>✉

Chromosomal microarray analysis (CMA) has been recommended and practiced routinely since 2010 both in the USA and Europe as the first-tier cytogenetic test for patients with unexplained neurodevelopmental delay/intellectual disability, autism spectrum disorders, and/or multiple congenital anomalies. However, in Brazil, the use of CMA is still limited, due to its high cost and complexity in integrating the results from both the private and public health systems. Although Brazil has one of the world's largest single-payer public healthcare systems, nearly all patients referred for CMA come from the private sector, resulting in only a small number of CMA studies in Brazilian cohorts. To date, this study is by far the largest Brazilian cohort (n = 5788) studied by CMA and is derived from a joint collaboration formed by the University of São Paulo and three private genetic diagnostic centers to investigate the genetic bases of neurodevelopmental disorders and congenital abnormalities. We identified 2,279 clinically relevant CNVs in 1886 patients, not including the 26 cases of UPD found. Among detected CNVs, the corresponding frequency of each category was 55.6% Pathogenic, 4.4% Likely Pathogenic and 40% VUS. The diagnostic yield, by taking into account Pathogenic, Likely Pathogenic and UPDs, was 19.7%. Since the rationale for the classification is mostly based on Mendelian or highly penetrant variants, it was not surprising that a second event was detected in 26% of those cases of predisposition syndromes. Although it is common practice to investigate the inheritance of VUS in most laboratories around the world to determine the inheritance of the variant, our results indicate an extremely low cost–benefit of this approach, and strongly suggest that in cases of a limited budget, investigation of the parents of VUS carriers using CMA should not be prioritized.

Chromosomal microarray analysis (CMA), including array-comparative genomic hybridization (aCGH) and SNP-array, has become the gold standard procedure to detect copy number variations (CNVs) in the clinical setting. Because CMA offers a much higher diagnostic yield (15–20%) than the conventional G-banded karyotype (~3%), the test is recommended as the first-tier cytogenetic test for patients with unexplained neurodevelopmental delay/intellectual disability, autism spectrum disorders, or multiple congenital anomalies<sup>1</sup>. It is noteworthy that G-banded karyotyping should be offered only for patients with obvious chromosomal syndromes (e.g., Down syndrome), a family history of chromosomal rearrangement, or multiple miscarriages. However, in Brazil, detection of chromosomal alterations is still performed mainly by karyotyping, due to the high taxation costs of importing microarray material, and relatively cheap technical labour. As a result, the number of CMA studies in cohorts of patients with neurodevelopmental disorders and congenital anomalies is very scarce, and

<sup>1</sup>The Human Genome and Stem Cell Research Center, Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of São Paulo, Rua do Matão 277, São Paulo ZIP Code 05508-090, Brazil. <sup>2</sup>Diagnósticos da América S.A., DASA, São Paulo, Brazil. <sup>3</sup>Mendelics, São Paulo, SP, Brazil. <sup>4</sup>Instituto da Criança Do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil. <sup>5</sup>These authors contributed equally: Ana C. V. Krepischi and Darine Villela. ✉email: carlarosenberg@ib.usp.br

their sample size is typically small (< 500 individuals)<sup>2–6</sup>. Nonetheless, the few previous investigations reported a diagnostic rate ranging from 15 to 22%, similar to that cited in the literature<sup>7–12</sup>.

The clinical interpretation of CMA results can be challenging. Although it is now possible to screen the human genome for CNVs at high resolution, identifying several clinically recognizable syndromes, there are many variants that are rare or only present in a particular individual or family. While some can be confidently predicted to be either pathogenic or benign, in many cases, definitive evidence is missing, leaving us with many variants of uncertain significance (VUS)<sup>13</sup>. Variant classification is often complicated since the criteria applied in the interpretation of a CNV include inheritance, size, type (duplication or deletion), and gene content, with support of multiple database resources for annotation<sup>14,15</sup>. Because the CMA data from the Brazilian population is underrepresented in the literature and public databases, this study was established to provide a collection of genomic data from 5778 patients with various neurodevelopmental disorders; all were genotyped using a high-resolution SNP-array platform. This study is by far the largest Brazilian cohort investigated in diagnostic CMA; by creating this data resource we aimed to establish an overview of all cytogenetic alterations found in clinical CMA, and document the CNVs that are clinically relevant in the diagnosis of neurodevelopmental disorders.

## Material and methods

**Casuistic.** The cohort presented here results from a joint collaboration between the Human Genome and Stem Cell Research Center of the Institute of Biosciences, University of São Paulo (IB-USP), and three private diagnostic centers located in the state of São Paulo (DASA, Mendelics, and Deoxi Biotechnology) to provide the largest copy number data from patients investigated in a clinical CMA routine in Brazil. Despite the three diagnostic centers being in the same state in Brazil, patients were referred from all regions of the country. A total of 5778 children underwent CMA, between 2010 and 2020, for presenting a general neurodevelopmental disorder and/or congenital abnormalities without evident cause. All relevant variants are publicly available at the DECIPHER database (<https://www.deciphergenomics.org/>). This study was approved by the IB-USP Research Ethical Committee, and an informed consent was obtained from the patients' parents or guardians.

**Chromosomal microarray analysis (CMA): SNP-array.** Genomic DNA samples were extracted from peripheral blood cells or saliva following standard procedures. SNP-array experiments were performed using the Illumina Infinium CytoSNP 850 K BeadChip (Illumina, San Diego, USA), except for 810 cases which were carried out using the Affymetrix CytoScan 750 K Array (Affymetrix, Santa Clara, USA). Data were analyzed using either the BlueFuse™ Multi Analysis (Illumina, San Diego, USA) or the Chromosome Analysis Suite—ChAS Software (Affymetrix, Santa Clara, USA). Log<sub>2</sub> ratio and B Allele Frequency (BAF) values were plotted along chromosomal coordinates, allowing the detection of both copy number changes and copy neutral regions of homozygosity (ROH).

**Variant analysis and clinical interpretation.** Copy number variants were classified for their clinical impact according to the *American College of Medical Genetics (ACMG) guidelines*<sup>15</sup>. The criteria for variant classification were as follow:

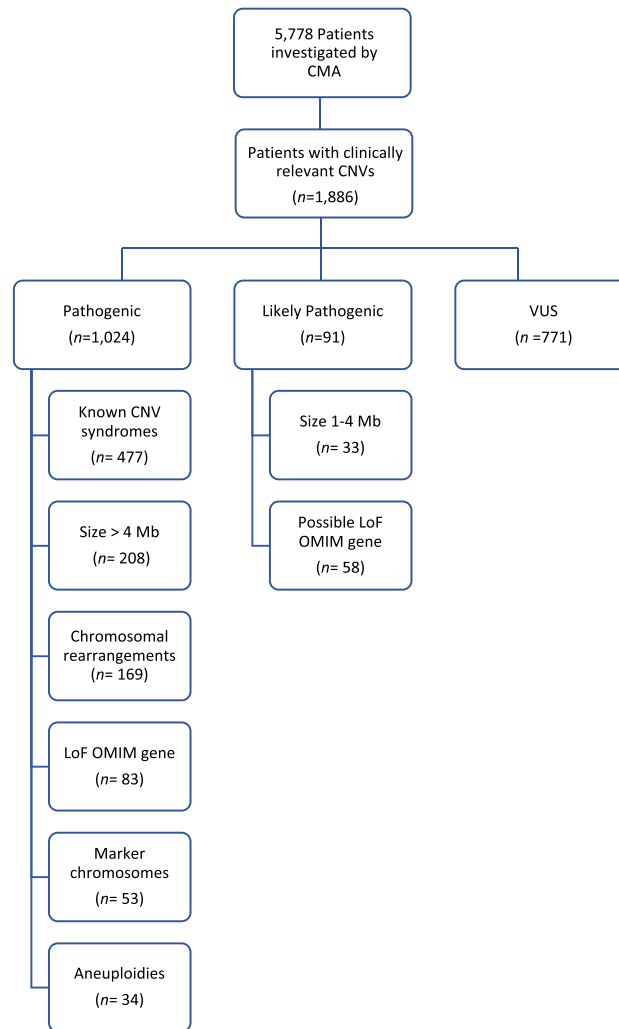
- Pathogenic = (1) when the CNVs were more than 4 Mb in length harboring genes, or; (2) overlapped with regions associated with OMIM morbid genes or DECIPHER/ClinGen microdeletion/microduplication syndromes; (3) deleted haploinsufficiency of OMIM genes;
- Likely Pathogenic = when the CNVs (1) were deletions partially affecting haploinsufficiency of OMIM genes; (2) harboring genes and were 1–4 Mb in length;
- VUS = when the CNVs (1) were duplications containing MIM genes; (2) were deletions, containing recessive MIM genes; or (3) when the segment was larger than 300 kb and harbored genes.

The common variants, i.e., those commonly reported in curated databased (DGV) were disregarded from this study. Chromosomal rearrangements were defined by the presence of more than one large CNV in different or in the same chromosome (e.g.: chromosomes derived from translocations and inversions). In particular, copy neutral ROHs restricted to a single chromosome, known to harbor imprinted regions, were considered pathogenic and likely representing uniparental disomy (UPD). ROH > 10 Mb or at least two ROH > 5 Mb were considered indicative of consanguinity, albeit not Pathogenic per se.

**Ethical approval.** This study is in accordance with ethical standards established in the Declaration of Helsinki (1964), its subsequent revisions, and Resolution 466/2012 of the Brazilian National Health Council. The Research Ethics Committee of the Institute of Biosciences from the University of São Paulo gave ethical approval for this work (CAAE 80921117.5.0000.5464), and an informed consent was obtained from the patients' parents or guardians for genetic testing.

## Results

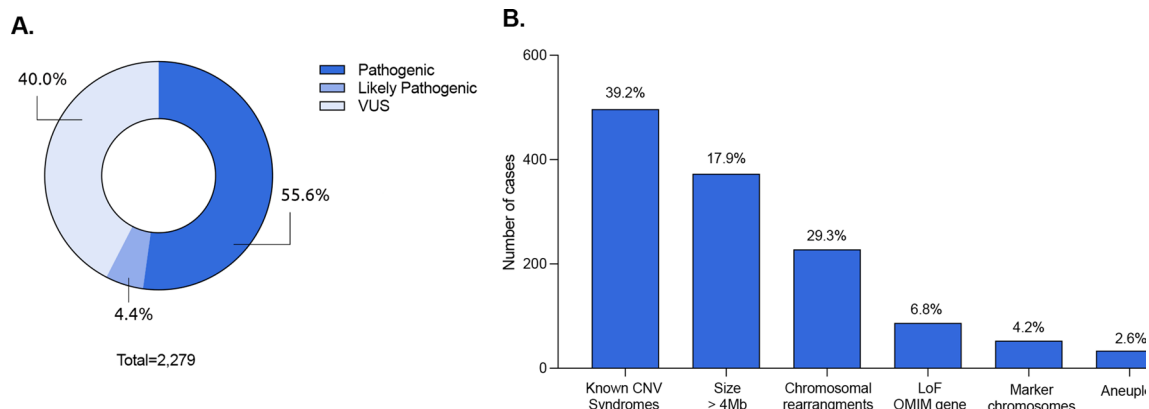
**Diagnostic rate of the cohort.** An overview of the number of individuals with clinically relevant CNVs obtained in the cohort of this study is shown in Fig. 1. Out of the 5778 patients with neurodevelopmental disorders or congenital abnormalities investigated, relevant CNVs were detected in 1886 individuals, and were classified in three main categories: (i) Pathogenic (54%); (ii) Likely Pathogenic (5%) and (iii) Variants of Unknown Significance (VUS; 41%). Taking into account just the Pathogenic, Likely Pathogenic and UPD cases, the overall diagnostic yield in our cohort was 19.7%.



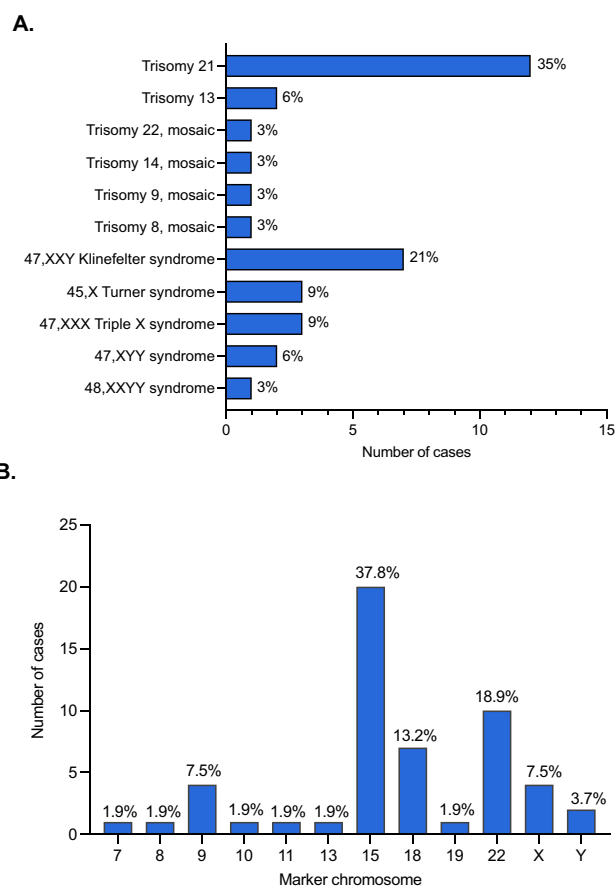
**Figure 1.** An overview of cases with clinically relevant copy number variations (CNVs) identified in the cohort. The figure shows that, from a total of 5778 patients with neurodevelopmental disorders referred for chromosomal microarray analysis (CMA), 1886 carried clinically relevant CNVs, classified into three main categories: (i) Pathogenic CNVs; (ii) Likely Pathogenic CNVs; and (iii) Variants of Unknown Significance (VUS). The total number of cases corresponding to each category is presented in the diagram. Those individuals with more than one alteration were classified within the most clinically relevant category.

In these 1886 individuals, a total of 2,270 relevant CNVs, were identified and the corresponding frequency of each category was 55.6% Pathogenic, 4.4% Likely Pathogenic and 40% VUS (Fig. 2A). As expected, Pathogenic CNVs accounted for the largest proportion of diagnostic alterations and are divided into seven clinically relevant classes of variants, as presented in Fig. 2B. The description of all individual Pathogenic, Likely pathogenic CNVs, and VUS can be found in Supplementary Tables 1–3.

**Aneuploidies and marker chromosomes.** Sex chromosome (SCA) and autosomal aneuploidies accounted for 34 cases (3.3% of the pathogenic cases) in our cohort. Considering them as a separate groups, the proportion was very similar between SCA and autosomal trisomies: 16 and 18 cases, respectively. SCA comprise 47,XXX, 48,XXYY, 47,XYY, 47,XXY, and 45,X; the most frequent being 47,XXY (Klinefelter syndrome), found in 8/34 cases (23.5%). Among the autosomal trisomies, the most common was trisomy 21, found in 12/34 (35.3%), followed by trisomy 13 in 2/34 (5.9%). Excluding the known viable autosomal trisomies, an extra copy of other autosomes was only observed in mosaics, as it was the case with chromosomes 8, 9, 14 and 22. The frequency of each aneuploidy is shown in Fig. 3A. Marker chromosomes were identified in 53 patients (5.2% of the pathogenic cases). We only considered in this category those markers seen in karyotype, and which did not characterize a well-known OMIM syndrome, such as Pallister-Killian, Cat-eye or Emanuel syndromes. Marker chromosome 15 was the most frequent, corresponding to 20/53 (37.8%). Other markers originated from chromosomes 7, 8, 9, 10, 11, 12, 13, 18, 19, 22, X and Y; except for those derived from the sex chromosomes, all were supernumerary marker chromosomes (Fig. 3B).



**Figure 2.** Distribution of the CNVs identified in the cohort. **(A)** It is shown the frequency of variants in each of the three main CNV categories: (i) Pathogenic; (ii) Likely pathogenic and (iii) Variants of Unknown Significance (VUS). **(B)** Distribution of the pathogenic CNVs, displayed by frequency order: (i) known CNV syndromes, (ii) CNV > 4 Mb, (iii) complex rearrangements, (iv) loss-of-function (LoF) MIM gene, (v) marker chromosomes, (vi) aneuploidies.



**Figure 3.** Frequency of aneuploidies and marker chromosomes. **(A)** Sex chromosome aneuploidies (SCA) and autosomal trisomies accounted for a total of 34 cases, in which 16 correspond to SCA and 18 to autosomal trisomies. The histogram shows the frequency of aneuploidies for each chromosome. **(B)** The frequency of the detected 53 marker chromosomes is displayed according to its chromosome origin.

**Large CNVs and known syndromes.** As shown in Fig. 1, we detected isolated large CNVs (> 4 Mb) in 208/1,024 Pathogenic cases (20.3%), and affecting at least two chromosome segments (chromosome rearrangements) in 169/1,024 cases (16.5%). Loss-of-function (LoF) mutations in haploinsufficiency MIM genes accounted for 83 cases (8.1% of the Pathogenic cases). A total of 477 patients (46.6% of the Pathogenic cases) presented 497 CNVs associated with microdeletion or microduplication syndromes (48 deletions and 23 dupli-

cations). The five most frequent syndromes were 22q11.2 deletion (14.6%; MIM#188400); 15q13.3 reciprocal duplication encompassing only the *CHRNA7* gene<sup>16</sup> (6.6%); 16p11.2 deletion (5%, MIM#611913); 15q11.2 deletion (4.8%; MIM#615656); and Prader-Willi/Angelman syndrome (4.6%; MIM#176270/105830, respectively). A detailed listing and the corresponding number of cases and frequencies of each of the 71 clinical entities are shown in Table 1 and Fig. 4. Importantly, among these syndromes, there were 13 known to confer susceptibility to neurodevelopmental disorders, in other words, presented reduced penetrance; such susceptibility CNVs were detected in 144 patients, 38 of which carried additional variants, observed only in autosomes (Fig. 5).

**Likely Pathogenic variants and VUS.** Lastly, a total of 91 patients carried 99 Likely Pathogenic CNVs, which represent only 4.8% of the individuals carrying CNVs in our cohort (91/1886). The third category of CNVs, the VUS accounted for 908 variants in 771 individuals.

**Uniparental disomy (UPD) and copy neutral regions of homozygosity (ROH).** Copy neutral ROH were observed in 259 patients and were divided in two categories, according to the supposed origin of the ROH. Twenty-six were large or whole-chromosome ROH block(s) restricted to a single chromosome, classified as UPD; of these UPD cases, 14 were classified as Pathogenic by mapping to imprinted chromosomes 6, 7, 11, 14 and 15. The most common UPD was UPD15 (7/26), followed by UPD14 (3/26) and UPD1 (3/26) (Fig. 6). In the remaining 233 individuals carrying ROHs, the presence of blocks of homozygosity in more than one chromosome was indicative of identity by descent.

## Discussion

In this study, we report the largest Brazilian cohort of patients with neurodevelopmental disorder investigated by CMA. An overall diagnostic yield of 19.7% was determined, a result similar to that found in other studies<sup>8–11</sup>. An extensive study with over 15,000 patients<sup>17</sup> found a lower frequency of diagnosis (~14.2%), but considered only those CNVs above 400 kb. The copy number data presented here was deposited in the DECIPHER database, and clearly demonstrated as previous studies the massive importance of copy number changes in postnatal diagnosis.

CMA has been recommended and practiced routinely in the USA and Europe as the first-tier test for patients with neurodevelopmental disorders and congenital abnormalities since 2010<sup>1</sup>. However, the use of CMA tests is still limited in Brazil due to their high costs. It is relevant to mention that the healthcare system in Brazil is a complex mixture of public and private funding, with governance and ownership agreements. The Brazilian public health sector is one of the world's largest single payer healthcare systems. In complementation with this scenario there is a significant and large private sector supported with high investment. It is estimated that only ~26% of Brazilians have a private health insurance, and it is mainly concentrated in the urban areas of the Southeastern part of the country<sup>18</sup>. Although nearly all patients referred for CMA come from the private sector, the health insurances require that G-banded karyotype be used as the first genetic test. The patients with no structural and/or numerical alterations by karyotyping are subsequently referred for investigation by microarray analysis. In contrast, in the public sector, CMA is not even offered to the patients, since the price established by the government for the total genetic investigation of a patient does not pay even the costs of material for a single CMA. In practice, CMA is provided for few patients at Public Universities or Institutions, when it is linked to specific projects and research grants. However, this situation is not just a Brazilian peculiarity, since in countries with better economic conditions, private laboratories can contribute to a class disparity regarding access to medical analyzes.

Although many patients in our cohort had been previously investigated by G-banded karyotyping, we found 34 cases of aneuploidy, in which trisomy 21 was the most frequent chromosomal disorder encountered. The patients with Down syndrome were referred for CMA for presenting autistic features to search for other CNVs associated with autism spectrum disorders; however, in none of the cases, additional CNVs were detected. It is noteworthy that the presence of autism spectrum disorders in individuals with Down syndrome has been well documented for several years<sup>19,20</sup>; therefore, the analysis through CMA in these cases is puzzling. The second most common aneuploidy detected was the Klinefelter syndrome; also in this case, behavioral disorders were the main reason for CMA referral. It is well documented that behavior problems, including autism, are relatively common in Klinefelter syndrome<sup>21</sup>. In fact, autistic features may be more common in individuals with sex chromosome aneuploidies than generally believed<sup>22</sup>.

Marker chromosomes from both autosomes and sex chromosomes represented 4.6% of the diagnostic alterations (Pathogenic, Likely Pathogenic, UPD). Except for the X and Y chromosomes, all were supernumerary. The correlation of specific supernumerary marker chromosomes (SMC) with distinct clinical features have been demonstrated for some syndromes; those were classified as syndromes rather than markers, for example i(12p)-(Pallister-Killian) syndrome (MIM#601803), and Cat eye syndromes (MIM#609029 and MIM#115470, respectively)<sup>23</sup>. The only marker that represents a translocation derivative chromosome in our dataset characterize the Emanuel syndrome (MIM#609029), and results from missegregation of the only known recurrent, non-Robertsonian, constitutional translocation in humans [der(22)t(11;22)(q23;q11.2)]. Although it was not possible to obtain cytogenetic characterization of all markers, it is known that an inverted duplicated chromosome 15<sup>24</sup> is the most common of the heterogeneous group that constitute the supernumerary marker chromosomes.

The microdeletion and microduplication syndromes account for the largest proportion of the diagnosis obtained in our cohort (41.8%). Many of these syndromes harbor genomic hotspots flanked by homologous segmental duplications prone to unequal crossing over, and have high elevated de novo mutation rates, generally with similar CNV sizes<sup>25,26</sup>. In this study, we detected 71 distinct microdeletion/microduplication syndromes in a total of 477 individuals, in which deletions were twice as common than duplications ( $n = 48$  deletions vs  $n = 23$  duplications). Among the five most frequent syndromes, shown in Fig. 4, four have incomplete penetrance and variable expressivity: 22q11.2 deletion (MIM#188400), 15q13.3 duplication<sup>16</sup>, 16p11.2 deletion (MIM#611913)

Known CNV syndromes	No. of cases (%)	Cytoband	MIM (#)
<b>A. Microdeletion syndromes</b>			
22q11.2 deletion syndrome	73 (14.6%)	22q11.21	188,400
16p11.2 deletion syndrome, 593 kb	25 (5.0%)	16p11.2	611,913
15q11.2 deletion syndrome ( <i>NIPA1</i> )	24 (4.8%)	15q11.2	615,656
Prader-Willi/Angelman syndrome	23 (4.6%)	15q11.2	176,270/105,830
1p36 deletion syndrome	19 (3.8%)	1p36.2	607,872
Williams-Beuren syndrome	19 (3.8%)	7q11.23	194,050
Wolf-Hirschhorn syndrome	16 (3.2%)	4p16.3	194,190
22q13.3 deletion syndrome	16 (3.2%)	22q13	606,232
Koolen-De Vries syndrome	13 (2.6%)	17q21.31	610,443
18p deletion syndrome	11 (2.2%)	18p	146,390
Smith-Magenis syndrome	8 (1.6%)	17p11.2	182,290
Miller-Dieker lissencephaly syndrome	8 (1.6%)	17p13.3	247,200
Leri-Weill dyschondrosteosis, <i>SHOX</i> deletion	8 (1.6%)	Xp22.33	127,300
Steroid sulphatase deficiency/Ichthyosis, X-linked	8 (1.6%)	Xp22.31	308,100
8p23.1 deletion syndrome	7 (1.4%)	8p23.1	<a href="https://www.deciphergenomics.org/syndrome/39/overview">https://www.deciphergenomics.org/syndrome/39/overview</a>
Kleefstra syndrome 1	7 (1.4%)	9q34.3	610,253
16p11.2 deletion syndrome, distal, 220 kb – <i>SH2B1</i>	7 (1.4%)	16p11.2	613,444
16p13.11 recurrent microdeletion (neurocognitive disorder susceptibility locus)	6 (1.2%)	16p13.1	<a href="https://www.deciphergenomics.org/syndrome/79/overview">https://www.deciphergenomics.org/syndrome/79/overview</a>
2q37 deletion syndrome	6 (1.2%)	2q37.2	600,430
Sotos syndrome 1	5 (1.0%)	5q35.3	117,550
18q deletion syndrome	5 (1.0%)	18q	601,808
22q11.2 deletion syndrome, distal	5 (1.0%)	22q11.2	611,867
15q13.3 deletion syndrome ( <i>CHRNA7</i> ) 500 kb	5 (1.0%)	15q13.3	612,001
1q21.1 deletion syndrome, proximal	4 (0.8%)	1q21.1	612,474
15q13.3 deletion syndrome	4 (0.8%)	15q13.3	612,001
17q11.2 deletion syndrome, <i>NFI</i>	4 (0.8%)	17q11.2	613,675
1q21.1 deletion syndrome ( <i>GJA5</i> ), distal	3 (0.6%)	1q21.1	612,474
3q29 microdeletion syndrome	3 (0.6%)	3q29	609,425
6pter-p24 deletion syndrome	3 (0.6%)	6p25	612,582
9p deletion syndrome	3 (0.6%)	9p	158,170
10q26 deletion syndrome	3 (0.6%)	10q26	609,625
Jacobsen syndrome	3 (0.6%)	11q23	147,791
Temple syndrome /Kagami-Ogata syndrome	3 (0.6%)	14q32.2	616,222/608,149
17q12 deletion syndrome	3 (0.6%)	17q12	614,527
18q22.3q23 microdeletion	3 (0.6%)	18q22.3q23	607,842
16p12.1 deletion syndrome, 520 kb	3 (0.6%)	16p12.1	136,570
2p16.1-p15 deletion syndrome	2 (0.4%)	2p16.1-p15	612,513
15q24 deletion syndrome	2 (0.4%)	15q24	613,406
Cri-du-Chat syndrome	1 (0.2%)	5p	123,450
7q11.23 deletion syndrome, distal, 1.2 Mb	1 (0.2%)	7q11.23	613,729
8q21.11 deletion syndrome	1 (0.2%)	8q21.11	614,230
Genitopatellar syndrome	1 (0.2%)	10q22.2	606,170
Potocki-Shaffer syndrome	1 (0.2%)	11p11.2	601,224
12q14 microdeletion syndrome	1 (0.2%)	12q14	<a href="https://www.deciphergenomics.org/syndrome/76/overview">https://www.deciphergenomics.org/syndrome/76/overview</a>
Polycystic kidney disease, infantile severe, with tuberous sclerosis	1 (0.2%)	16p13.3	600,273
Alpha-thalassemia/mental retardation syndrome, type 1	1 (0.2%)	16p13.3	141,750
16p13.3 deletion syndrome, proximal	1 (0.2%)	16p13.3	610,543
17q23.1-q23.2 deletion syndrome	1 (0.2%)	17q23.1-q23.2	613,355
<b>B. Microduplication syndromes</b>			
15q13.3 duplication syndrome, reciprocal ( <i>CHRNA7</i> ) 500 kb	33 (6.6%)	15q13.3	<a href="https://dosage.clinicalgenome.org/clingen_gene.cgi?sym=CHRNA7&amp;subject=">https://dosage.clinicalgenome.org/clingen_gene.cgi?sym=CHRNA7&amp;subject=</a>
15q11.2 duplication syndrome, reciprocal ( <i>NIPA1</i> )	15 (3.0%)	15q11.2	<a href="https://dosage.clinicalgenome.org/clingen_gene.cgi?sym=NIPA1&amp;subject=">https://dosage.clinicalgenome.org/clingen_gene.cgi?sym=NIPA1&amp;subject=</a>
16p13.11 recurrent microduplication (neurocognitive disorder susceptibility locus)	11 (2.2%)	16p13.1	<a href="https://www.deciphergenomics.org/syndrome/80/overview">https://www.deciphergenomics.org/syndrome/80/overview</a>
1q21.1 duplication syndrome ( <i>GJA5</i> ), distal	8 (1.6%)	1q21.1	612,475
Continued			

Known CNV syndromes	No. of cases (%)	Cytoband	MIM (#)
22q11.2 microduplication syndrome	7 (1.4%)	22q11.21	608,363
Williams-Beuren region duplication syndrome	6 (1.2%)	7q11.23	609,757
Cat eye syndrome	6 (1.2%)	22q11	115,470
Potocki-Lupski syndrome	5 (1.0%)	17p11.2	610,883
17q12 duplication syndrome	4 (0.8%)	17q12	614,526
16p11.2 duplication syndrome	3 (0.6%)	16p11.2	614,671
Charcot-Marie-Tooth disease	3 (0.6%)	17p12	118,220
Supernumerary der(22)t(11;22) syndrome	3 (0.6%)	22q11.2	609,029
22q13 duplication syndrome	3 (0.6%)	22q13	615,538
Pallister-Killian syndrome	2 (0.4%)	2p	601,803
17p13.3, centromeric, duplication syndrome	2 (0.4%)	17p13.3	613,215
3q29 microduplication syndrome	1 (0.2%)	3q29	611,936
8p23.1 duplication syndrome	1 (0.2%)	8p23.1	<a href="https://www.deciphergenomics.org/syndrome/85/overview">https://www.deciphergenomics.org/syndrome/85/overview</a>
16p13.3 duplication syndrome	1 (0.2%)	16p13.3	613,458
17q21.31 duplication syndrome	1 (0.2%)	17q21.31	613,533
17q23.1-q23.2 duplication syndrome	1 (0.2%)	17q23.1-q23.2	613,618
Xp11.22 microduplication syndrome	1 (0.2%)	Xp11.2	300,705
Xq28 duplication syndrome	1 (0.2%)	Xq28	300,815
Mental retardation, X-linked syndromic, Lubs type, MECP2 duplication	1 (0.2%)	Xq28	300,260
Total	499		

**Table 1.** Frequency of the known copy number variation (CNV) syndromes.

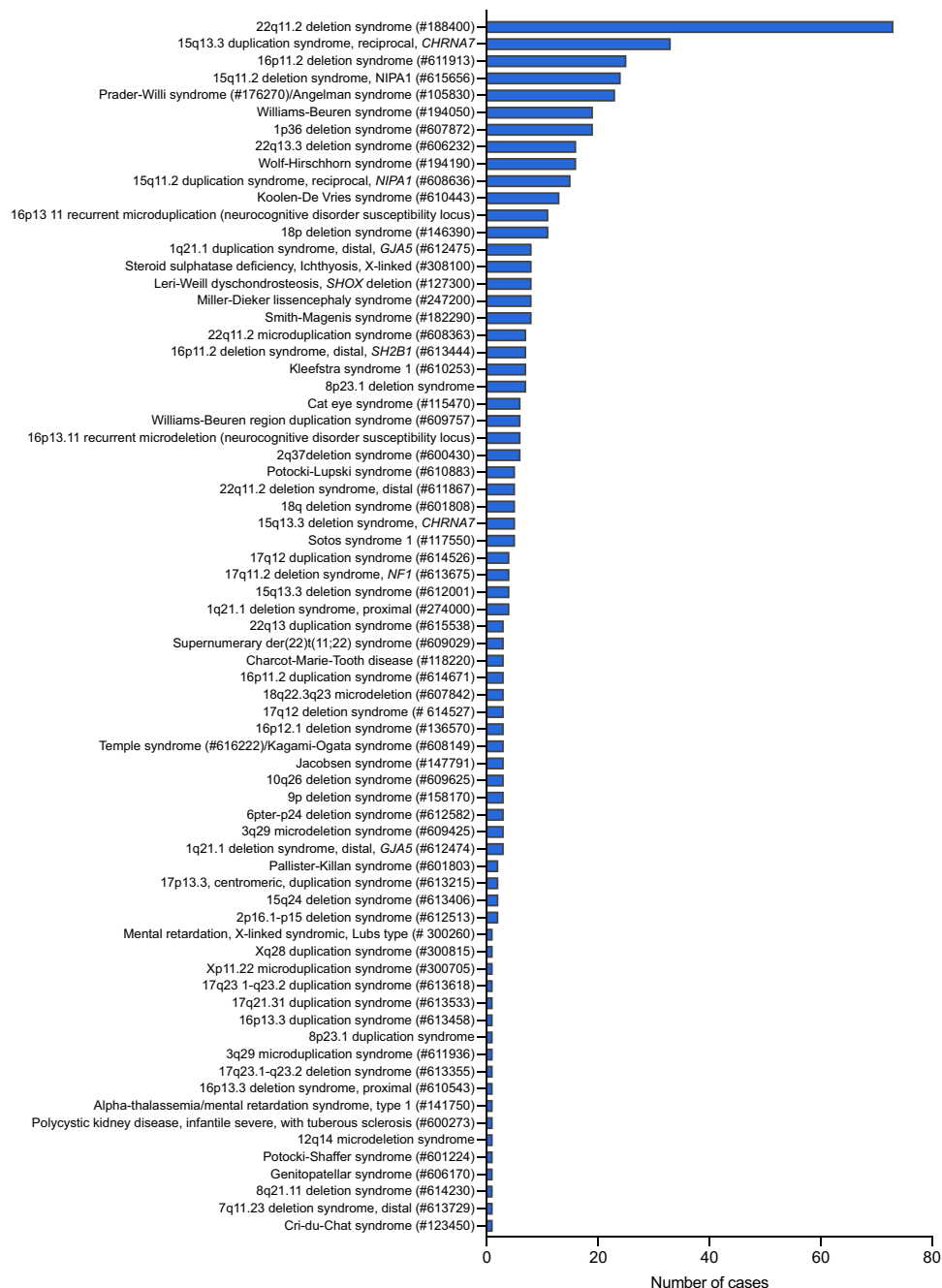
and 15q11.2 deletion (MIM#615656). Such susceptibility CNVs impose a challenge in genetic counseling since they are present in the normal population but enriched in individuals with various neurodevelopmental disorders. Moreover, these CNVs are often inherited from a normal or mildly affected parent, and they lack phenotypic specificity, being associated with a variety of neuropsychiatric disorders, congenital abnormalities, and variable dysmorphisms.

The 22q11.2 deletion was the most frequent syndrome in our cohort, reflecting the same frequency reported by other large population studies<sup>27</sup>. In particular, the 15q13.3 duplication encompassing the *CHRNA7* gene was previously associated to several neurodevelopmental disorders<sup>17,28,29</sup>. However, overlapping duplications in this genomic region were also documented in many individuals of the general population (~0.6%—estimated prevalence of 1:174–186 individuals<sup>30</sup>) and, in almost all cases investigated, patients inherited the duplication from clinically normal parents. The high frequency of duplications of this segment in the general population, together with the lack of enrichment in clinical cohorts<sup>31,32</sup>, indicate that, if this variant has any clinical impact, the penetrance would be very low.

The recurrent 15q11.2 deletion (BP1-BP2), which includes the *CYFIP1*, *NIPAI1*, *NIPA2*, *TUBGCP5* genes, is consistently associated with neurocognitive function. Jonch et al. (2019)<sup>33</sup> performed a comprehensive meta-analysis on individuals with 15q11.2 deletions, comparing data across 20 studies. The case–control study using their clinical cohort compared to controls in the UK Biobank cohort showed enrichment of the deletion in the patient population. Nonetheless, the reciprocal duplication of the 15q11.2 region has refuted clinical significance<sup>31,34</sup>. Duplications of this region are common in the general population and the majority of case–control studies have observed a lack of enrichment in the clinical population. Recent studies of duplication carriers identified through cohort studies in the general population have also shown that carrier individuals perform similarly to non-carrier controls on neurocognitive tests. Therefore, duplication of this segment is considered unlikely dosage sensitivity. Indeed, 15q11.2 and 15q13.3 restricted duplications were historically classified as pathogenic, and the data presented reflects a retrospective analysis; currently, are classified as benign in light of recent evidence, and in accordance with databases such as DECIPHER and ClinGen.

The estimated frequency of the 16p11.2 deletion syndrome is about 1–5/10,000 in the general population. Research based on the ClinGen database suggests that 16p11.2 deletions are the second most common microdeletion, occurring in one of every 235 individuals tested with intellectual and developmental disabilities. Interestingly, this deletion was identified in nearly 1% of individuals with autism<sup>35–37</sup>. Nonetheless, the phenotypic spectrum associated with this deletion is much wider and includes delays in speech or motor development, language impairment, low muscle tone, hypo- or hyperreflexia, a tendency towards obesity, short stature, and several facial dysmorphisms. The syndrome classically involves a heterozygous microdeletion of ~600 kb, containing 29 protein coding genes; although the majority of cases reported are de novo, the deletion is inherited in an autosomal dominant pattern in 20% of the cases; an equal sex ratio has been reported<sup>38</sup>.

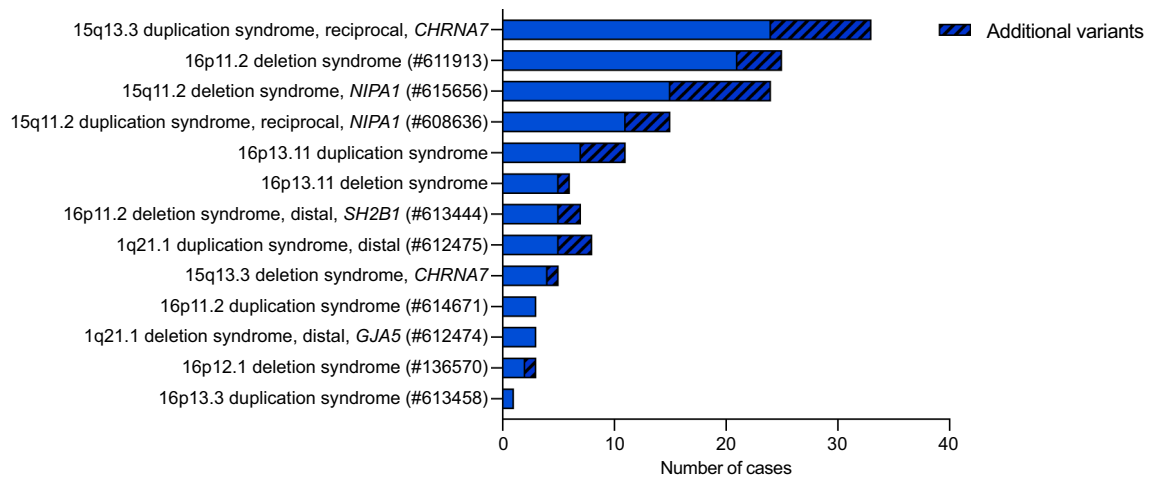
Considering all evidence from association studies about the susceptibility CNVs for neurodevelopmental disorders, the general consensus is that there must be additional modifiers that influence the expression of these variants. A “two-hit”, or second site, model has been suggested for several syndromes<sup>39,40</sup>. Notably, the vast majority of the syndromes with incomplete penetrance, as shown in Fig. 5 from our cohort, present an additional CNV in part of the patients, mostly classified as VUS. Girirajan et al. (2010) demonstrated that 25% of the affected



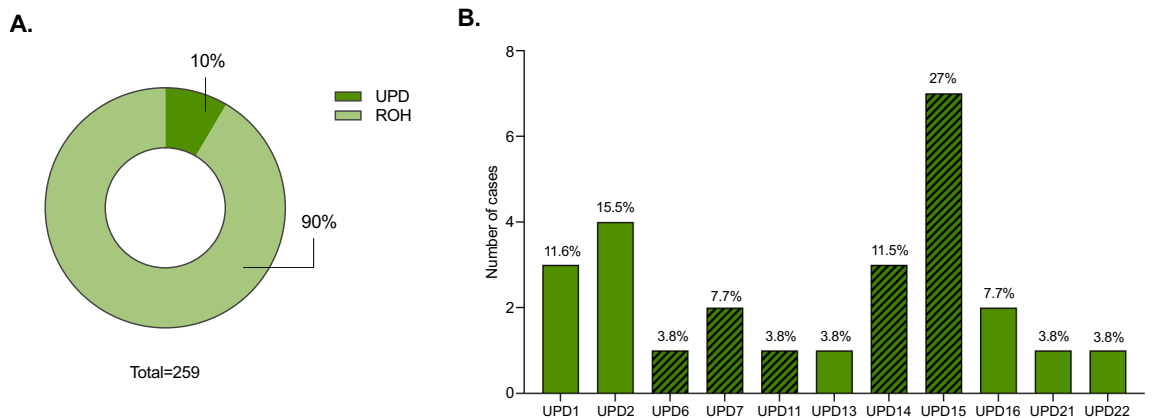
**Figure 4.** Distribution of microduplication and microdeletion syndromes. The histogram shows the frequency of microduplication and microdeletion syndromes identified in a total of 477 patients with neurodevelopmental disorders. Results are displayed in the descending order of frequency.

individuals with a microdeletion on chromosome 16p12.1 carried additional large CNVs<sup>40</sup>. This data supports an oligogenic basis, in which the compound effect of a relatively small number of rare variants of large effect contributes to the heterogeneity of genomic disorders. The authors also identified other known genomic disorders, each defined by a specific CNV, in which the affected children were more likely to carry multiple copy number variants than controls<sup>39</sup>. Overall, a second CNV hit was identified in 10% of their cases. We found that patients with a CNV known to present incomplete penetrance (Fig. 5) carried a CNV second-hit in 26.4% of the cases. It is assumed that part of the second-hits would be only ascertained by sequencing, but SNV/indel second-hits were not investigated in this study. It is relevant to mention that investigation of the parents is mandatory in cases of those susceptibility CNVs for proper genetic counselling<sup>15</sup>. However, in our study we were unable to obtain information about inheritance pattern for all cases because the patients and parents were not necessarily investigated at the same centers or in the same period of time.





**Figure 5.** Frequency of syndromes with incomplete penetrance associated with additional variants. The histogram shows the frequency of a secondary CNV associated with a syndrome with incomplete penetrance.



**Figure 6.** Frequency of copy neutral regions of homozygosity (ROH). (A) A total of 259 patients carried ROHs, 26 (10%) corresponding to uniparental disomy (UPD) cases. The remaining ROH in the other 233 patients (90%) were considered associated to different degrees of identity by descent. (B) The histogram shows the frequency of UPDs per chromosome detected in our cohort. The crossed pattern represents pathogenic UPD known to encompass imprinting regions.

In cases where the patient present one or more VUS, the recommendation is to investigate the parents to determine whether the CNV has been inherited or represents a de novo mutation<sup>15</sup>. While the latter may lead to a reclassification of the VUS as pathogenic or likely pathogenic, the inherited variants remain classified as VUS. In our cohort, we have information on segregation in a minority of the cases: among 126 patients who presented an autosomal VUS and segregation was tested, we found that 125 of the variants were inherited. It is important to remember that, regardless of being inherited, these VUS may still contribute to the patients' phenotype, and must be reported. While it is undeniable that the resulting 0.8% de novo alterations have some impact in diagnosis and genetic counselling, the healthcare context should be considered in the decision to test the parents. In the present cohort, we performed 252 CMA tests in parents, but were able to reclassify the VUS in a single case. For all remaining 125 cases, the parental tests did not add any useful information. When resources are limited, such as in Brazil and in many other countries, we obtain a better cost–benefit testing other 252 patients instead of investigating VUS segregation. Therefore, we would not recommend testing segregation of VUS using CMA in the public healthcare in Brazil or in other developing countries in the current situation. Nonetheless, segregation analysis can be performed with cheaper techniques such as real time PCR.

Importantly, with the incorporation of many robust SNP-array platforms in the clinical routine, many studies have identified large ROH in patients with a wide variety of clinical features<sup>41–44</sup>. Depending on chromosomal distribution and cumulative extent, it may either indicate UPD or parental consanguinity<sup>45</sup>. When ROH > 10 Mb are detected in a single chromosome, the first possibility to be considered is UPD; this event arises as a consequence of a trisomy rescue, which may have further implications, such as the presence of an undetected trisomic cell line. In cases of chromosomes subject to imprinting, the presence of two copies of the same chromosome inherited from only one of the parents is considered as pathogenic per se. UPD in non-imprinted chromosomes still increases the probability of deleterious mutations in homozygosity<sup>46</sup>. Contrarily, the presence of many ROH

throughout the genome is an indication of consanguinity, and the chance of inheritance of recessive monogenic disorders increases with the degree of relatedness. It has been demonstrated that the occurrence of multiple congenital anomalies and other significant clinical problems is higher among children of first cousins (4.4%) and second cousins (3.6%), compared to unrelated parents<sup>47</sup>. The rate of consanguinity in the different regions of Brazil is very heterogeneous and estimates are scarce. A recent paper indicates that some degree of inbreeding may be present in 26.5% of patients with developmental disorders of the South of Brazil<sup>2</sup>. However, when clinically more relevant kinship of 1st–5th degree is considered, they find consanguinity in ~ 8.5% of the cases. In our cohort, only 233 out of the 259 cases of ROH were interpreted as the result of consanguinity, reflecting a lower frequency of consanguinity of 4.0% (233/5778).

In summary, we reported copy number data from patients with neurodevelopmental disorders and congenital anomalies in the largest Brazilian cohort investigated by CMA reported so far. These data, available in the DECIPHER database, can be used as a valuable resource for other genomic studies.

### Data availability

The datasets generated and/or analyzed during the current study are available in the DECIPHER database (<https://www.deciphergenomics.org/>). The reference number that corresponds to the DECIPHER patients ID are presented on the Supplementary Tables. All genomic coordinates from the variant data that support the findings of this study are available on the Supplementary Tables. Raw data supporting the findings of this study are available on request from the corresponding author—C.R. The raw data are not publicly available due to patients privacy/consent restriction.

Received: 14 March 2022; Accepted: 26 August 2022

Published online: 07 September 2022

### References

1. Miller, D. T. *et al.* Consensus statement: Chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am. J. Hum. Genet.* <https://doi.org/10.1016/j.ajhg.2010.04.006> (2010).
2. Chaves, T. F. *et al.* Long contiguous stretches of homozygosity detected by chromosomal microarrays (CMA) in patients with neurodevelopmental disorders in the South of Brazil. *BMC Med. Genom.* **12**, 1–13 (2019).
3. Pratte-Santos, R., Ribeiro, K. H., Santos, T. A. & Cintra, T. S. Analysis of chromosomal abnormalities by CGH-array in patients with dysmorphic and intellectual disability with normal karyotype. *Einstein (Sao Paulo)*. **14**, 30–34 (2016).
4. Vianna, G. S., Medeiros, P. F. V., Alves, A. E., Silva, T. O. & Jehee, F. S. Array-CGH analysis in patients with intellectual disability and/or congenital malformations in Brazil. *Genet. Mol. Res.* **15**, 1 (2016).
5. Chaves, T. F. *et al.* Copy number variations in a cohort of 420 individuals with neurodevelopmental disorders from the South of Brazil. *Sci. Rep.* **9**, 1–20 (2019).
6. Pereira, R. R. *et al.* Screening for intellectual disability using high-resolution CMA technology in a retrospective cohort from Central Brazil. *PLoS ONE* **9**, 1 (2014).
7. Baldwin, E. L. *et al.* Enhanced detection of clinically relevant genomic imbalances using a targeted plus whole genome oligonucleotide microarray. *Genet. Med.* **10**, 415–429 (2008).
8. Wong, A. *et al.* Detection and calibration of microdeletions and microduplications by array-based comparative genomic hybridization and its applicability to clinical genetic testing. *Genet. Med.* **7**, 264–271 (2005).
9. Rosenberg, C. *et al.* Array-CGH detection of micro rearrangements in mentally retarded individuals: Clinical significance of imbalances present both in affected children and normal parents. *J. Med. Genet.* **43**, 180–186 (2006).
10. Miné, M., & Chen, J. *et al.* Rapid communication a large genomic deletion in the PDHX gene caused by the retrotranspositional insertion of a full-length LINE-1 element. *Hum. Mutat.* 1–6 (2006).
11. Xiang, B. *et al.* Analytical and clinical validity of whole-genome oligonucleotide array comparative genomic hybridization for pediatric patients with mental retardation and developmental delay. *Am. J. Med. Genet. Part A* **146**, 1942–1954 (2008).
12. Miné, M., & Chen, J. *et al.* Rapid communication a large genomic deletion in the PDHX gene caused by the retrotranspositional insertion of a full-length LINE-1 element. *Hum. Mutat.* 1–6 (2006).
13. Hoffman-Andrews, L. The known unknown: The challenges of genetic variants of uncertain significance in clinical practice. *J. Law Biosci.* **4**, 648–657 (2017).
14. Silva, M. *et al.* European guidelines for constitutional cytogenomic analysis. *Eur. J. Hum. Genet.* **27**, 1–16 (2019).
15. Kearney, H. M., Thorland, E. C., Brown, K. K., Quintero-Rivera, F. & South, S. T. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet. Med.* **13**, 680–685 (2011).
16. Coe, B. P. *et al.* Genes associated with developmental delay. *Nat. Genet.* **46**, 1063–1071 (2014).
17. Cooper, G. M. *et al.* A copy number variation morbidity map of developmental delay. *Nat. Genet.* **43**, 838–846 (2011).
18. Lewis, M., Penteado, E. & Malik, A. M. Brazil's mixed public and private hospital system. *World Hosp. Health Serv.* **51**, 22–26 (2015).
19. Kent, L., Evans, J., Paul, M. & Sharp, M. Comorbidity of autistic spectrum disorders in children with Down syndrome. *Dev. Med. Child Neurol.* **41**, 153–158 (1999).
20. Howlin, P., Wing, L. & Gould, J. the recognition of autism in children with down syndrome-implications for intervention and some speculations about pathology. *Dev. Med. Child Neurol.* **37**, 406–414 (1995).
21. Jha, P., Sheth, D. & Ghaziuddin, M. Autism spectrum disorder and Klinefelter syndrome. *Eur. Child Adolesc. Psychiatry* **16**, 305–308 (2007).
22. Marco, E. J. & Skuse, D. H. Autism-lessons from the X chromosome. *Soc. Cogn. Affect. Neurosci.* **1**, 183–193 (2006).
23. Liehr, T., Claussen, U. & Starke, H. Small supernumerary marker chromosomes (sSMC) in humans. *Cytogenet. Genome Res.* **107**, 55–67 (2004).
24. Battaglia, A. The inv dup (15) or idic (15) syndrome (Tetrasomy 15q). *Orphanet J. Rare Dis.* **3**, 1–7 (2008).
25. Carvalho, C. M. B. & Lupski, J. R. Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* <https://doi.org/10.1038/nrg.2015.25> (2016).
26. Lupski, J. R. Genomic disorders: Structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* **14**, 417–422 (1998).
27. Zarrei, M. *et al.* A large data resource of genomic copy number variation across neurodevelopmental disorders. *NPJ Genomic Med.* **4**, 1 (2019).
28. Gillentine, M. A. & Schaaf, C. P. The human clinical phenotypes of altered CHRNA7 copy number. *Biochem. Pharmacol.* **97**, 352–362 (2015).

29. Zhou, D., Gochman, P., Broadnax, D. D., Rapoport, J. L. & Ahn, K. 15Q13.3 Duplication in two patients with childhood-onset Schizophrenia. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **171**, 777–783 (2016).
30. Szafranski, P. *et al.* Structures and molecular mechanisms for common 15q13.3 microduplications involving CHRNA7: Benign or pathological?. *Hum. Mutat.* **31**, 840–850 (2010).
31. Coe, B. P. *et al.* Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat. Genet.* **46**, 1063–1071 (2014).
32. Williams, N. M. *et al.* Genome-wide analysis of copy number variants in attention deficit hyperactivity disorder: The role of rare variants and duplications at 15q13.3. *Am. J. Psychiatry* **169**, 195–204 (2012).
33. Jonch, A. E. *et al.* Estimating the effect size of the 15Q11.2 BP1-BP2 deletion and its contribution to neurodevelopmental symptoms: Recommendations for practice. *J. Med. Genet.* **56**, 701–710 (2019).
34. Kendall, K. M. *et al.* Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: Analysis of the UK Biobank. *Br. J. Psychiatry* **214**, 297–304 (2019).
35. Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* **358**, 667–675 (2008).
36. Stefansson, H. *et al.* CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* **505**, 361–366 (2014).
37. Kumar, R. A. *et al.* Recurrent 16p11.2 microdeletions in autism. *Hum. Mol. Genet.* **17**, 628–638 (2008).
38. Zufferey, F. *et al.* A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. *J. Med. Genet.* **49**, 660–668 (2012).
39. Girirajan, S. *et al.* Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N. Engl. J. Med.* **367**, 1321–1331 (2012).
40. Girirajan, S. *et al.* A recurrent 16p121 microdeletion supports a two-hit model for severe developmental delay. *Nat. Genet.* **42**, 203–209 (2010).
41. Kearney, H. M., Kearney, J. B. & Conlin, L. K. Diagnostic implications of excessive homozygosity detected by SNP-Based microarrays: Consanguinity, uniparental disomy, and recessive single-gene mutations. *Clin. Lab. Med.* **31**, 595–613 (2011).
42. Gibson, J., Morton, N. E. & Collins, A. Extended tracts of homozygosity in outbred human populations. *Hum. Mol. Genet.* **15**, 789–795 (2006).
43. Papenhausen, P. *et al.* UPD detection using homozygosity profiling with a SNP genotyping microarray. *Am. J. Med. Genet. Part A* **155**, 757–768 (2011).
44. Sund, K. L. *et al.* Regions of homozygosity identified by SNP microarray analysis aid in the diagnosis of autosomal recessive disease and incidentally detect parental blood relationships. *Genet. Med.* **15**, 70–78 (2013).
45. Wang, J. C. *et al.* Regions of homozygosity identified by oligonucleotide SNP arrays: Evaluating the incidence and clinical utility. *Eur. J. Hum. Genet.* **23**, 663–671 (2015).
46. Musante, L. & Ropers, H. H. Genetics of recessive cognitive disorders. *Trends Genet.* **30**, 32–39 (2014).
47. Vissers, L. E. L. M., Gilissen, C. & Veltman, J. A. Genetic studies in intellectual disability and related disorders. *Nat. Rev. Genet.* **17**, 9–18 (2016).

## Acknowledgements

This work was supported by FAPESP grant (2013/08028-1), and FAPESP Innovative Research in Small Business (12/50981-5).

## Author contributions

A.C.V.K. contributed to manuscript conceptualization and conducted data curation. D.V. wrote the manuscript. G.G. and D.R.B. were responsible for referring and clinically characterizing the largest number of patients to CMA analysis in this study. M.P.M., F.K. and F.C. also contributed to the clinical characterization of the patients and referral to genetic testing. S.S.C., P.C.M. and J.G.S. performed the experiments and contributed to data analysis. F.M. coordinated the clinical diagnostic routine. R.G.S. and J.P.K. conducted the bioinformatic analysis. G.C. and D.S. led the responsibility of DASA and Mendelics diagnostic centers, respectively. A.M.V.M., and P.L.P. discussed the results and contributed to the final version of the manuscript. C.R. was the principal investigator and supervised the study. All authors read and approved the final version of the manuscript.

## Competing interests

A.C.V.K. and C.R. report that they served as a consultant for DASA, Mendelics, and Deoxi Biotechnology. All other authors report no conflicts of interest relevant to this article.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-19274-6>.

**Correspondence** and requests for materials should be addressed to C.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022