Genetics and Molecular Biology, 43, 4, e20190218 (2020)

DOI: https://doi.org/10.1590/1678-4685-GMB-2019-0218

Research Article
Human and Medical Genetics

Copy number variation (CNV) identification, interpretation, and database from Brazilian patients

Victória Cabral Silveira Monteiro de Godoy[1], Fernanda Teixeira Bellucco[1], Mileny Colovati[1], Hélio Rodrigues de Oliveira-Junior[1], Mariana Moysés-Oliveira[1] and Maria Isabel Melaragno[1]

[1]*Universidade Federal de São Paulo, Departamento de Morfologia e Genética, Disciplina de Genética, São Paulo, SP, Brazil.*

Abstract

Copy number variations (CNVs) constitute an important class of variation in the human genome and the interpretation of their pathogenicity considering different frequencies across populations is still a challenge for geneticists. Since the CNV databases are predominantly composed of European and non-admixed individuals, and Brazilian genetic constitution is admixed and ethnically diverse, diagnostic screenings on Brazilian variants are greatly difficulted by the lack of populational references. We analyzed a clinical sample of 268 Brazilian individuals, including patients with neurodevelopment disorders and/or congenital malformations. The pathogenicity of CNVs was classified according to their gene content and overlap with known benign and pathogenic variants. A total of 1,504 autosomal CNVs (1,207 gains and 297 losses) were classified as benign (92.9%), likely benign (1.6%), VUS (2.6%), likely pathogenic (0.2%) and pathogenic (2.7%). Some of the CNVs were recurrent and with frequency increased in our sample, when compared to populational open resources of structural variants: 14q32.33, 22q11.22, 1q21.1, and 1p36.32 gains. Thus, these highly recurrent CNVs classified as likely benign or VUS were considered non-pathogenic in our Brazilian sample. This study shows the relevance of introducing CNV data from diverse cohorts to improve on the interpretation of clinical impact of genomic variations.

Keywords: CNV, copy number variation, database, CNV classification.

Received: July 12, 2019; Accepted: September 25, 2020.

Introduction

CNVs are characterized by losses or gains of DNA sequences that are larger than 50 bp (Alkan *et al.* 2011; MacDonald *et al.* 2014; Zarrei *et al.* 2015). They are a relevant class of variants due to the large number of genome segments that differs in the dosage between individuals, conferring great interindividual diversity (Iafrate *et al.* 2004). Such variants are also present in healthy individuals with no apparent association with disease phenotypes, being considered benign CNVs (Iafrate *et al.* 2004; Sebat *et al.* 2004; Redon *et al.* 2006; Korbel *et al.* 2007; Conrad *et al.* 2010). There are also CNVs that are responsible for the etiology of numerous human diseases, such as multiple syndromes that are associated to congenital anomalies, complex neurodegenerative and neuropsychiatric disorders, intellectual disability, cancer and immunological diseases (Stankiewicz and Lupski 2010; Girirajan *et al.* 2011). The recognition of their influence on the phenotype, however, is not an easy task. Although some CNV maps and databases have been constructed (Zarrei *et al.* 2015) for both, healthy individuals (e.g. DGV- http://dgv.tcag.ca/dgv/app/home) and affected patients (e.g. DECIPHER- https://decipher.sanger.ac.uk/), it is still a great challenge to assess the CNVs' clinical impact (Lee *et al.* 2007; Gijsbers *et al.* 2011; Brnich *et al.* 2019). The American College of Medical Genetics and Genomics (ACMG) presented guidelines for CNVs interpretation and recommended the use of specific standard terminology: "pathogenic", "likely pathogenic", "uncertain significance", "likely benign", and "benign" (Richards *et al.* 2015). This guide was recently updated to assist clinical laboratories in the classification and reporting of CNVs. These professional standards will guide the evaluation of constitutional CNVs and encourage consistency and transparency across clinical laboratories (Riggs *et al.* 2019). However, there are no generally established rules for CNV analysis, interpretation, and classification, and the guidelines can change over time due to the scientific information evolution (de Leeuw *et al.* 2012; Palmer *et al.* 2014; Brnich *et al*, 2019; Riggs *et al*. 2019). Furthermore, it has been shown that the CNV distribution can differ across ethnic populations (Li *et al*. 2009; Collins *et al*. 2019). The Brazilian population is highly admixed and still underrepresented in genomic databanks (Naslavsky *et al*. 2017; Andrade *et al*. 2018). Thus, the main goal of this study was to survey and classify large CNVs to assemble a database from Brazilian patients in order to improve the interpretation of their clinical impact.

Send correspondence to Maria Isabel Melaragno. Universidade Federal de São Paulo, Departamento de Morfologia e Genética, Disciplina de Genética, São Paulo, SP, Brazil. E-mail: melaragno.maria@unifesp.br

Subjects and Methods

Individuals studied

A sample composed of 268 microarrays performed in patients with phenotypic alterations was studied (Table S1). The patients were recruited from the Medical Genetics Center of the Universidade Federal de São Paulo, outpatient clinics of the Hospital São Paulo and other genetics centers

in the state of São Paulo, Brazil. This project was approved by the University Ethics Committee and all participants or parents signed informed consents. All procedures performed involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendmentsor comparable ethical standards. We studied microarrays from 143 patients with normal karyotypes (69 presented with diverse phenotypic alterations and 74 were patients with phenotype of the oculoauriculovertebral spectrum - OAVS) and 125 patients with previously identified genomic imbalances/chromosomal alterations, who participated in specific studies in our laboratory (50 presented with 22q11.2 deletion, 23 with apparently balanced translocations, nine with marker chromosomes, five with 18q deletions, four with 18p deletions, four with ring chromosomes, and 30 with other abnormal G-banding karyotypes exclusive to a single patient).

Microarray-based copy number variation assay and quality control

Genomic DNA was obtained from peripheral blood using the Gentra Puregene kit (Qiagen-Sciences, Maryland, USA). DNA samples were then analyzed using the Genome-Wide Human Array 6.0 SNP array (n= 59 individuals), CytoScan 750K (n= 54 individuals), and CytoScan High-Density SNP array (n= 155 individuals), following the manufacturer's instructions (Affymetrix, Santa Clara, CA, US). Array analyses were performed using the Chromosome Analysis Suite software (ChAS), version 3.3 (Affymetrix, Santa Clara, CA, USA). The quality control (QC) parameters were applied according to the manufacturer´s recommendations. For Genome-Wide Human Array 6.0 SNP array platform, samples with Median of the Absolute values of all Pairwise Differences (MAPD) ≤ 0.35 were included in the sample. For CytoScan High-Density (HD) SNP array and CytoScan 750K platforms, samples with MAPD ≤ 0.25, SNP quality control (SNPQC) ≥ 15 (or ≥ 12 when all other parameters met the requirements), and waviness standard deviation (waviness SD) ≤ 0.12 (when all other parameters met the requirements) were also included in the sample.

CNVs analysis and classification

The CNV classification was performed by the same investigator in a blind manner considering array type and patients´ phenotype. Autosomal CNVs that had a minimum coverage of 50 probes and a minimum size of 200 kb for gains and 150 kb for losses were considered for the analysis of pathogenicity, since deletions can be more deleterious for the phenotype. The CNVs previously detected by other cytogenomic tests and undoubtedly causative to the patients' phenotype such as, deletions 18p, 18q and deletion 22q11.2, were excluded from the analysis in order to avoid super representation of these loci. The genomic imbalances were

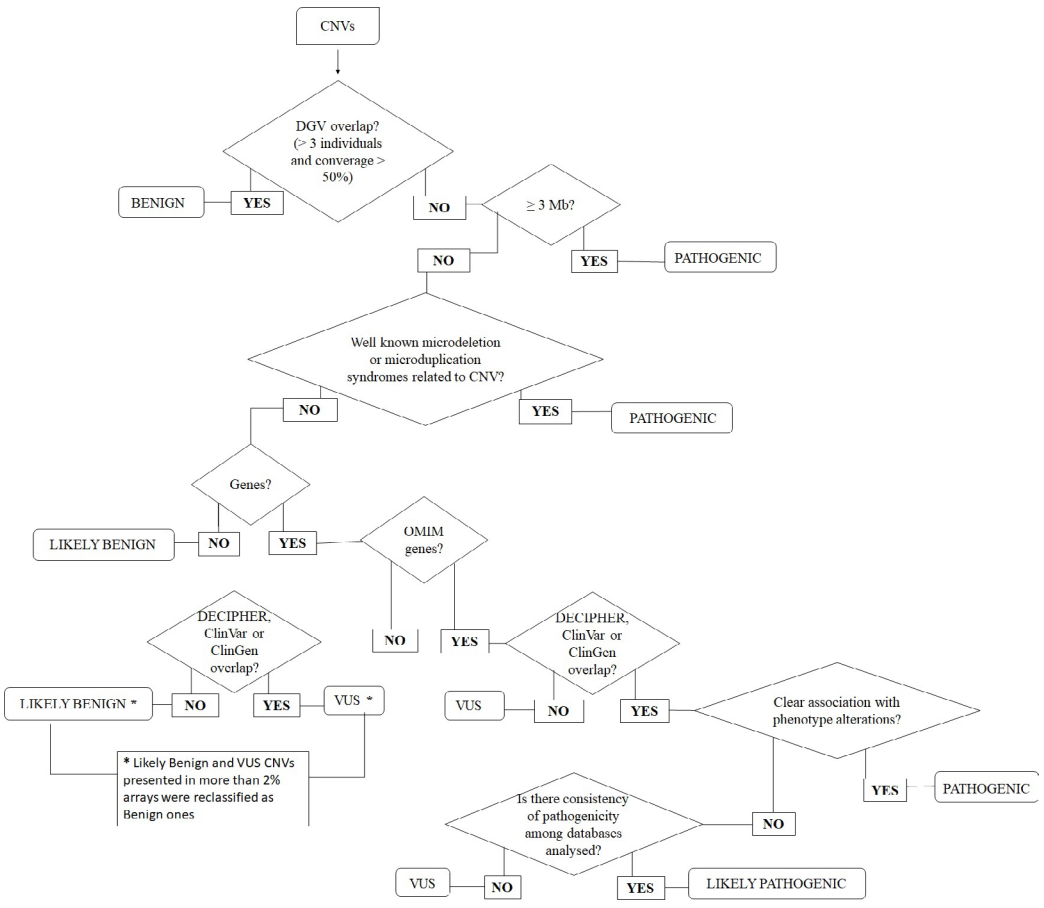

Figure 1 – Flowchart for CNVs analysis and interpretation.

annotated based on the GRCh37/hg19 Genome Build (Feb 2009). CNVs analysis was performed using the UCSC with DGV track (May 15 2016 version), DECIPHER track (Oct 30 2018 version), ClinVar track (Oct 2017 version), ClinGen track (Oct 2017 version), OMIM track for analysis of genes associated with diseases (Oct 10 2018 version) and NCBI RefSeq Genes track (April 19 2017 version). A flowchart for CNVs classification (Figure 1) was built based on the criteria described by Lee *et al*. (2007), Vermeesch *et al*. (2012), Palmer *et al*. (2014), Richards *et al*. (2015), and Nowakowska (2017). CNVs were classified into five categories proposed by the ACMG guidelines (Richards *et al*. 2015): benign, likely benign, variant of uncertain significance (VUS), likely pathogenic and pathogenic. A CNV was considered benign when there was more than 50% overlap, in size and location, with DGV-CNVs of the same nature (i.e., deletion or duplication) from at least three unaffected individuals, and the nonoverlapping segment did not exceed 50 percent of the DGV-CNVs' length. CNVs were considered likely benign when they did not contain genes or if they did, the genes within them were not OMIM genes and the CNVs did not overlap with the ones found in databases of affected individuals (DECIPHER, ClinVar and ClinGen). CNVs were considered VUS when (1) they contained genes that were not OMIM genes, and they overlapped at least one CNV found in databases of affected individuals, with more than 50% overlap, in size and location; (2) when they contained OMIM genes but did not overlap CNVs found in databases of affected individuals; or (3) when they had OMIM genes and overlapped CNVs found in databases of affected individuals, but do not show a clear correlation with phenotypic alterations and consistency of classifications in the analyzed databases. CNVs were considered likely pathogenic when they presented genes described in OMIM, overlapped CNVs present in databases of genomic imbalances in affected individuals, and had no clear association with phenotypic alterations, but with consistency in the databases indicating phenotypic alterations. CNVs were considered pathogenic when (1) they were more than 3 Mb in length; (2) they overlapped with regions associated with DECIPHER microdeletion/microduplication syndromes; or (3) when, even though their size were not more than 3 Mb in length, they harbored OMIM genes, overlapped with CNVs found in databases of affected individuals that showed consistent correlation with phenotypic alterations**.** After this classification, recurrent CNVs found in a high percentage (≥ 2%) of patients in our sample were reclassified as benign ones.

Results

CNVs type, size and pathogenicity

A total of 1,504 autosome CNVs from 268 microarrays were considered for downstream analysis of pathogenicity. Among them, 1,207 (80.3%) were gains, and 297 (19.7%) were losses. Table 1 present the number and percentage of loss, gain, and total CNVs according to the different sizes and classifications of pathogenicity. According to the classification criteria, 1,397 of them (92.9%) were considered benign, 25 (1.6%) were likely benign, 39 (2.6%) were VUS, 3 (0.2%) were likely pathogenic, and 40 (2.7%) were pathogenic. The mean size of the CNVs was ~763 kb, being ~704 kb for gains and ~1.0 Mb for losses. The mean size of benign, likely benign, VUS, likely pathogenic and pathogenic CNVs was 586 kb, 522 kb, 621 kb, 1.0 Mb and 7.2 Mb, respectively. The sizes of the CNVs ranged from 200 kb to 24.5 Mb for gains and from 150 kb to 20 Mb for losses. The mean number of CNVs per patient was 5.6, being 12.6 for 6.0 SNP array, 2.9 for 750K array, and 3.9 for HD array.

Recurrent CNVs in our sample

Benign, likely benign, and VUS CNVs that were found in, arbitrarily, more than 2% arrays within a certain genomic region, were grouped and were considered as recurrent CNVs in our sample. According to these criteria, our sample presented with recurrent copy gains in four genomic regions: 14q32.33 (97.8% patients), 22q11.22 (32.1%), 1p36.32 (9.7%), and 1q21.1 (6,7%). In our pathogenicity evaluation, these first two copy gains had been classified as benign, the third as VUS and the fourth as likely benign. Given their high frequency, VUS CNVs in 1p36.32 and likely benign CNVs in 1q21.1 were reclassified as benign (i.e. non-pathogenic) CNVs in our Brazilian sample and already included in analyzes (Table S2).

Discussion

In our array sample, a higher number of gain CNVs (80.3%) compared to loss CNVs (19.7%) was found, in agreement with previous studies from the literature (Kang *et al*. 2008; Pietiläinen *et al*. 2011; Palmer *et al*. 2014). Most of the CNVs in the present study were classified as benign (92.9%), and about half (62.6%) of those were smaller than 500 kb. Furthermore, in accordance to the literature (Lee *et al*. 2007), we observed that the CNV mean size increased according to pathogenicity, that is, the greater the imbalance, the greater the pathogenicity. We found that the CNVs distribution differed between platforms. Although the 6.0 SNP array platform was used in only 22% of the performed arrays, 49.6% of the CNVs were identified in this platform, with a mean of 12.6 CNVs/individual, while 750K and HD platforms showed averages of 2.9 and 3.9 CNVs/individual, respectively, that indicate a significant difference ($p<0.001$). The mean number of CNVs per patient was 5.6, being 12.6 for Human Array 6.0 SNP array, 2.9 for 750K array and 3.9 for HD array. These differences could be related to the characteristics of the arrays used. The 6.0 SNP array platform presents more than 1.8 million probes distributed along all the genome including probes in segmental duplication regions and in pseudogenes, while the 750 K and HD arrays, containing about 750,000 and 2.7 million probes, respectively, are more recent technologies than the 6.0 SNP array platform and are composed of probes with greater specificity and sensitivity, focusing in clinically relevant regions that results in a higher accuracy in genomic imbalances detection. It is important to highlight that this difference between platforms is not associated with QC metrics, since only high-quality genotyping reactions were considered in this analysis. Thus, we must consider that the variation of the array type, company origin, and the filtering used for CNV analysis, are factors that may difficult the comparison between published datasets and the CNVs reported in open databases. Some benign, VUS and likely benign CNVs were found to be recurrent in our sample and

Table 1 – Number and frequency (in percentage) of CNVs in our sample according to their type of genomic imbalance, size, and classification.

	Losses (%)						Gains (%)						Total (%)					
	<0.5 Mb	0.5 - 1 Mb	1-3 Mb	3-5 Mb	> 5 Mb	Total	<0.5 Mb	0.5 - 1 Mb	1-3 Mb	3-5 Mb	> 5 Mb	Total	<0.5 Mb	0.5 - 1 Mb	1-3 Mb	3-5 Mb	> 5 Mb	Total
B	223 (91.8%)	13 (86.6%)	8 (50%)	0 (0%)	0 (0%)	244 (82.1%)	651 (96.1%)	361 (97.3%)	129 (92.1%)	11 (84.6%)	1 (16.7%)	1153 (95.5%)	874 (95%)	374 (96.9%)	137 (87.8%)	11 (55%)	1 (4.5%)	1397 (92.9%)
LB	12 (4.9%)	0 (0%)	1 (6.3%)	0 (0%)	0 (0%)	13 (4.4%)	8 (1.2%)	1 (0.3%)	3 (2.1%)	0 (0%)	0 (0%)	12 (1%)	20 (2.2%)	1 (0.3%)	4 (2.6%)	0 (0%)	0 (0%)	25 (1.6%)
VUS	6 (2.5%)	0 (0%)	4 (25%)	0 (0%)	0 (0%)	10 (3.4%)	18 (2.7%)	9 (2.4%)	2 (1.5%)	0 (0%)	0 (0%)	29 (2.4%)	24 (2.6%)	9 (2.2%)	6 (3.8%)	0 (0%)	0 (0%)	39 (2.6%)
LP	0 (0%)	1 (6.7%)	1 (6.3%)	0 (0%)	0 (0%)	2 (0.7%)	0 (0%)	0 (0%)	1 (0.7%)	0 (0%)	0 (0%)	1 (0.1%)	0 (0%)	1 (0.3%)	2 (1.3%)	0 (0%)	0 (0%)	3 (0.2%)
P	2 (0.8%)	1 (6.7%)	2 (12.4%)	7 (100%)	16 (100%)	28 (9.4%)	0 (0%)	0 (0%)	5 (3.6%)	2 (15.4%)	5 (83.3%)	12 (1%)	2 (0.2%)	1 (0.3%)	7 (4.5%)	9 (45%)	21 (95.5%)	40 (2.7%)
T	243 (100%)	15 (100%)	16 (100%)	7 (100%)	16 (100%)	297 (100%)	677 (100%)	371 (100%)	140 (100%)	13 (100%)	6 (100%)	1207 (100%)	920 (100%)	386 (100%)	156 (100%)	20 (100%)	22 (100%)	1504 (100%)

P: Pathogenic; LP: Likely pathogenic; VUS: variants of unknown significance; LB: Likely Benign; B: Benign; T: Total.

showed high frequency in four genomic regions (14q32.33, 22q11.22, 1p36.32, and 1q21.1), being exclusively gain CNVs. The 14q32.33 and 22q11.22 CNVs, found in all three different platforms, were classified as benign in all individuals due to overlap with CNVs from the DGV. In contrast, some 1q21.1 and 1p36.32 CNVs were reclassified considering their high frequency in our sample (Table S2). The 1p36.32 CNVs were found exclusively in 750k (35.2% patients) and HD (4.5%) platforms and were reclassified as benign CNVs due to their high frequency. Interestingly, among the seven CNVs found using the HD platform, five were first classified as benign by DGV, one was classified first as VUS and after reclassified as benign due their high frequency, and another was maintained as likely benign because it had a different genomic coordinates (chr1:2412626-2729513) comparing to the others.

The 1q21.1 gain CNVs were found only in the 6.0 SNP array platform (20.3% patients) and HD (3.9%) platforms. Among the 12 CNVs found in the 6.0 SNP array platform, only one was first classified as benign by DGV while the other 11 were further reclassified as benign ones. In the HD platform, the smaller five out the six CNVs found were classified as benign, and the other, larger (952kb), was reclassified as benign. It is important to note that, among the reclassified CNVs, all of them were larger than the CNV classified first as benign or presented little different coordinates. Thus, although these findings reveal differences in CNV distribution between platforms, there was a consistency in recurrent CNVs within each platform, which demonstrates the reliability of our classification criteria.

We looked for these recurrent CNVs in other studies from the literature. Benign 14q32.33 gain CNV was also found in a high frequency (>90%) in some studies from populations with European, African or East Asian ancestry (de Smith *et al*. 2007; Conrad *et al*. 2010) but not in other studies in individuals from Ontario, Thailand and Caucasians and African-Americans (Sebat *et al*. 2004; Shaikh *et al*. 2009; Suktitipat *et al*. 2014; Uddin *et al*. 2015). The 22q11.22 gain CNV, considered as benign, had a higher frequency in our sample comparing with studies from DGV that also referred a CNV in this region (Itsara *et al*. 2009; Shaikh *et al*. 2009; Cooper *et al*. 2011). Deletions in the 22q11.2 region have already been described in patients with OAVS (Xu *et al*. 2008; Digilio *et al*. 2009). Given the previous association of this region to the OAVS spectrum, there could result in a bias in the CNV frequency in this region in our sample. However, among the 268 individuals analyzed in the present study, this 22q11.22 CNV was found in the 30 out of 74 (40.5%) patients with a clinical diagnosis of OAVS and in 56 out of 194 (28.9%) patients with other diseases. Thus, this CNV was found not only in OAVs patients in our sample, but also in patients with different phenotypes, indicating that this CNV can be found in a higher frequency in the Brazilian population. The 1q21.1 gain CNV, classified at first as likely benign, was also considered as recurrent in our sample. In the DGV database, this variant was reported in only one study, in which a 1q21.1 gain was observed in two out of 29,084 individuals from the USA and Canada (Coe *et al*. 2014). Due to the high frequency of this CNV in our sample, we reclassified this CNV as benign. Analogously, the 1p36.32 gain CNV, primely considered as VUS, was reclassified as benign due to its recurrence in our sample. This variant has a higher frequency in our sample when compared to cohorts described in DGV that detected this CNV in other ancestries (Redon *et al*. 2006) (the HapMap collection). These differences in CNV frequencies between the reported data from the literature and our data may indicate distinct composition of the individuals studied since there is a great populational heterogeneity among the publications. However, these frequency divergencies may also reflect differences in the ability in CNV detection, since a diversity of approaches (such as ROMA, BAC-aCGHs, and SNP array) are used for gain and loss detection. This is a relevant factor that does not permit a reliable comparison between data obtained from different papers from the literature, which can show variability in the detection of certain CNVs, as shown in our data from three different platform used, even from the same company. The data obtained in our study indicated that the established analysis flowchart was highly effective in the classification of the CNVs' pathogenicity and allowed the establishment of an CNVs database based on Brazilian individuals. This resource has been remarkably valuable for the diagnostic screenings in our laboratory, considering admixed genetic background of the Brazilian population. The data leveraged in this study may contribute for the pathogenicity interpretation of CNVs in other populations underrepresented in currently available open resources for structural variants.

Acknowledgements

The authors thank the patients and their parents for their participation in this study.

This study was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (grant 2014/11572-8 to M.I.M.) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Conflict of Interest

The authors declare that they have no conflict of interest.

Author Contributions

VCSMG built the database and was a major contributor in writing the manuscript; MMO contributed in the study design. FTB, HRO, and MMO contributed in the database construction, interpretation of variants and in the study design. MC helped in the interpretation of previous reports of some variants. MIM designed the study, supervised the entire work and contributed in the data analysis. All authors read and approved the final manuscript.

References

Alkan C, Coe BP and Eichler EE (2011) Genome structural variation discovery and genotyping. Nat Rev Genet 12:363-76.

Andrade RB, Amador MAT, Cavalcante GC, Leitão LPC, Fernandes MR, Modesto AAC, Moreira FC, Khayat AS, Assumpção PP, Ribeiro-dos-Santos A *et al*. (2018). Estimating Asian contribution to the Brazilian population: A new application of a validated set of 61 ancestry informative markers. G3 Genes/Genomes/Genetics 8:3577-3582

Brnich S, Abou Tayoun A, Greenblatt MS, Heinen CD, Kanavy D, Luo X, McNulty SM, Starita LM, Tavtigian SV *et al*. (2019). PgmNr 352: Recommendations for determining the clinical validity of functional studies for use in variant interpretation. In: American Society of Human Genetics 69th Annual Meeting, p 445.

Coe BP, Witherspoon K, Rosenfeld JA, van Bon BWM, Vulto-van Silfhout AT, Bosco P, Friend KL, Baker C, Buono S, Vissers LELM *et al*. (2014) Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet 46:1063-1071.

Collins RL, Brand H, Karczewski KJ, Zhao X, Alföldi J, Khera AV, Francioli LC, Gauthier LD, Wang H, Watts NA *et al*. (2019) PgmNr 351: gnomAD-SV: An open resource of structural variation for medical and population genetics. In: American Society of Human Genetics 69th Annual Meeting, p. 443.

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P *et al*. (2010) Origins and functional impact of copy number variation in the human genome. Nature 464:704-712.

Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V *et al*. (2011) A copy number variation morbidity map of developmental delay. Nat Genet 43:838-46.

Digilio MC, McDonald-McGinn DM, Heike C, Catania C, Dallapiccola B, Marino B and Zackai EH (2009) Three patients with Oculo-Auriculo-Vertebral Spectrum and microdeletion 22q11.2. Am J Med Genet A 0:2860-2864.

Gijsbers AC, Schoumans J and Ruivenkamp CA (2011) Interpretation of array comparative genome hybridization data: a major challenge. Cytogenet Genome Res 135:222-7.

Girirajan S, Campbell CD and Eichler EE (2011) Human copy number variation and complex genetic disease. Annu Rev Genet 45:203-226.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW and Lee C (2004) Detection of large-scale variation in the human genome. Nat Genet 36:949-51.

Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, Krauss RM, Myers RM, Ridker PM, Chasman DI *et al*. (2009) Population analysis of large copy number variants and hotspots of human genetic disease. Am J Hum Genet 84:148-161.

Kang Tae-wook, Yeo-jin Jeon, Eunsu Jang, Hee-jin Kim, Hwan Kim, Jong-lyul Park, Siwoo Lee, Yong Sung Kim, Jong Yeol Kim and Seon-young Kim (2008) Copy number variations (CNVs) identified in Korean individuals. BMC Genomics 9:492.

Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L *et al*. (2007) Paired-end mapping reveals extensive structural variation in the human genome. Science 318:420-426.

Lee C, Iafrate AJ and Brothman AR (2007) Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. Nat Genet 39:S48-S54.

de Leeuw N, Dijkhuizen T, Hehir-Kwa JY, Carter NP, Feuk L, Firth HV, Kuhn RM, Ledbetter DH, Martin CL, van Ravenswaaij-Arts CMA *et al*. (2012) Diagnostic interpretation of array data using public databases and internet sources. Hum Mutat 33:930-940.

Li J, Yang T, Wang L, Yan H, Zhang Y, Guo Y, Pan F, Zhang Z, Peng Y, Zhou Q *et al.* (2009) Whole genome distribution and ethnic differentiation of copy number variation in caucasian and asian populations. PLoS One 4:e7958.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL *et al*. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. Science 319:1100-4

MacDonald JR, Ziman R, Yuen RK, Feuk L and Scherer SW (2014) The database of genomic variants: A curated collection of structural variation in the human genome. Nucleic Acids Res 42:986-92.

Naslavsky MS, Yamamoto GL, de Almeida TF, Ezquina SAM, Sunaga DY, Pho N, Bozoklian D, Sandberg TOM, Brito LA, Lazar M *et al.* (2017) Exomic variants of an elderly cohort of brazilians in the ABraOM database. Hum Mutat 38:751-763.

Nowakowska B (2017) Clinical interpretation of copy number variants in the human genome. J Appl Genet 58:449-457.

Palmer E, Speirs H, Taylor PJ, Mullan G, Turner G, Einfeld S, Tonge B and Mowat D (2014) Changing interpretation of chromosomal microarray over time in a community cohort with intellectual disability. Am J Med Genet A 164A:377-85.

Pietiläinen OPH, Rehnström K, Jakkula E, Service SK, Congdon E, Tilgmann C, Hartikainen AL, Taanila A, Heikura U, Paunio T *et al.* (2011) Phenotype mining in CNV carriers from a population cohort. Hum Mol Genet 20:2686-2695.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W *et al.* (2006) Global variation in copy number in the human genome. Nature 444:444-454.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E *et al.* (2015) Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17:405-24.

Riggs E, Andersen E, Cherry A, Kantarci S, Kearney H, Patel A, Raca G, Ritter D, South S, Thorland E *et al.* (2019) PgmNr 353: Validation of scoring metrics to guide the classification of constitutional copy number variants. In: American Society of Human Genetics 69th Annual Meeting, p. 447.

Riggs E, Andersen E, Cherry A, Kantarci S, Kearney H, Patel A, Raca G, Ritter DI, South S, Thorland E *et al*. (2019) Technical standards for the interpretation and reporting of constitutional copy-number variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet Med 22:245-257.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Månér S, Massa H, Walker M, Chi M *et al.* (2004) Large-scale copy number polymorphism in the human genome. Science 305:525-28.

Shaikh TH, Gai X, Perin JC, Glessner JT, Xie H, Murphy K, O'Hara R, Casalunovo T, Conlin LK, D'Arcy M *et al.* (2009) High-resolution mapping and analysis of copy number variations in the human genome: A data resource for clinical and research applications. Genome Res 19:1682-1690.

de Smith AJ, Tsalenko A, Sampas N, Scheffer A, Yamada NA, Tsang P, Ben-Dor A, Yakhini Z, Ellis RJ, Bruhn L *et al.* (2007) Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males: Implications for association studies of complex diseases. Hum Mol Genet 16:2783-94.

Stankiewicz P and Lupski JR (2010) Structural variation in the human genome and its role in disease. Annu Rev Med 61:437-55.

Suktitipat B, Naktang C, Mhuantong W, Tularak T, Artiwet P, Pasomsap E, Jongjaroenprasert W, Fuchareon S, Mahasirimongkol S, Chantratita W *et al.* (2014) Copy number variation in Thai population. PLoS One 9:e104355.

Uddin M, Thiruvahindrapuram B, Walker S, Wang Z, Hu P, Lamoureux S, Wei J, MacDonald JR, Pellecchia G, Lu C *et al.* (2015) A high-resolution copy number variation resource for clinical and population genetics. Genet Med 17:747-752.

Vermeesch JR, Brady PD, Sanlaville D, Kok K and Hastings RJ (2012) Genome-wide arrays: Quality criteria and platforms to be used in routine diagnostics. Hum Mutat 33:906-15.
Xu J, Fan YS and Siu VM (2008) A child with features of Goldenhar Syndrome and a novel 1.12 Mb deletion in 22q11.2 by cytogenetics and oligonucleotide array CGH : Is this a candidate region for the syndrome?. Am J Med Genet A 146A:1886-1889.
Zarrei M, MacDonald JR, Merico D and Scherer SW (2015) A copy number variation map of the human genome. Nat Rev Genet 16:172-83.

Internet Resources

University of California-Santa Cruz Human Genome Browser (UCSC) (http://genome.ucsc.edu/).
Database of Genomic Variants (DGV) (http://projects.tcag.ca/variation).
Database of genomic variation and phenotype in humans using ensembl Resources (Decipher) (http://decipher.sanger.ac.uk).
ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/).
ClinGen (https://www.clinicalgenome.org/).
PubMed (http://www.ncbi.nlm.nih.gov/pubmed/).

Supplementary Material

Table S1 – Detailed information about all CNVs analyzed in our sample.
Table S2 – Recurrent gain CNVs in 1p36.32 and 1q21.1 showing their interpretation after re-classification accounting for CNV populational frequency.

Associate Editor: Catarina Satie Takahashi