Published in final edited form as:

Curr Genet Med Rep. 2017 December ; 5(4): 183–190. doi:10.1007/s40142-017-0129-2.

Copy Number Variation Disorders

Tamim H. Shaikh¹

¹Department of Pediatrics, Section of Clinical Genetics and Metabolism, University of Colorado Denver, Aurora, CO 80045

Abstract

Purpose of Review—Copy number variation (CNV) disorders arise from the dosage imbalance of one or more gene(s), resulting from deletions, duplications or other genomic rearrangements that lead to the loss or gain of genetic material. Several disorders, characterized by multiple birth defects and neurodevelopmental abnormalities, have been associated with relatively large (>1 Mb) and often recurrent CNVs. CNVs have also been implicated in the etiology of neuropsychiatric disorders including autism and schizophrenia as well as other common complex diseases. Thus, CNVs have a significant impact on human health and disease.

Recent Findings—The use of increasingly higher resolution, genomewide analysis has greatly enhanced the detection of genetic variation, including CNVs. Furthermore, the availability of comprehensive genetic variation data from large cohorts of healthy controls has the potential to greatly improve the identification of disease associated genetic variants in patient samples.

Summary—This review discusses the current knowledge about CNV disorders, including the mechanisms underlying their formation and phenotypic outcomes, and the advantages and limitations of current methods of detection and disease association.

Keywords

copy number variation; genomic disorder; microarray; microdeletion; microduplication

Correspondence: Tamim H. Shaikh, PhD, Department of Pediatrics, University of Colorado Denver, 12800 E. 19th Ave, Mail Stop 8300, Aurora, CO 80045, tamim.shaikh@ucdenver.edu.

Compliance with Ethics Guidelines

Human and Animal Rights and Informed Consent

This article does not contain any studies with human or animal subjects performed by any of the authors.

Conflict of Interest

The author has no conflicts of interest to report.

Web Resources:

The URLs for websites referred to herein are as follows:

Database of Genomic Variants (<http://projects.tcag.ca/variation/>)

NCBI's Database of genomic structural variation (dbVar) (<http://www.ncbi.nlm.nih.gov/dbvar/>)

DECIPHER (<http://decipher.sanger.ac.uk>),

ECARUCA (<http://umcecaruca01.extern.umcn.nl:8080/ecaruca/ecaruca.jsp>)

ClinGen: (<https://www.clinicalgenome.org/resources-tools/>)

ClinVar: (<https://www.ncbi.nlm.nih.gov/clinvar/>)

ExAC Browser: (<http://exac.broadinstitute.org/>)

Introduction

Copy number variation (CNV) refers to the difference in the dosage of genomic segments, ranging in size from one kilobase (kb) to several megabases (Mb), when compared to a reference human genome. CNVs can result from structural variations within the genome including deletions, duplications, insertions or unbalanced translocations and inversions, which can lead to either a loss or gain of genomic segments. CNVs are widespread within the genomes of healthy individuals with no apparent association to disease phenotypes [1–9]. The systematic analysis of CNVs in large populations of healthy individuals has begun to reveal the frequency and the effect of this variation in the human genome [10,11]. CNVs account for more of the inter-individual variability between genomes in terms of total number of bases involved than all the single nucleotide variations (SNVs) and small insertion-deletions (indels), combined [12–13].

Despite the existence of apparently benign CNVs in the human genome, numerous CNVs have been implicated in the etiology of a wide variety of human diseases including intellectual disability (ID), multiple congenital anomaly syndromes, as well as complex neurodegenerative and neuropsychiatric disorders, cancer and immune deficiency [14,15]. The availability and application of increasingly higher resolution genomic technologies, especially in clinical diagnostics, has greatly improved the detection of CNVs and has led to the discovery of new CNV disorders. Thus, CNVs have become increasingly recognized as a significant source not only of genetic diversity within the general population, but also the mutational burden associated with human disease. This review will focus on genetic disorders arising from CNVs, the mechanisms by which CNVs can lead to disease; the underlying genomic rearrangements associated with both recurrent and non-recurrent CNVs and the advantages and limitations of current and emerging methodologies used in CNV detection and the challenges that remain in the association of CNVs to disease.

Genetic disorders arising from CNVs

CNVs associated with genomic disorders

CNVs such as del and dup have long been recognized as a source of mutation in well-characterized genetic disorders and syndromes. These CNV disorders result from genomic alterations or rearrangements mediated by unstable genomic architecture, and are also referred to as genomic disorders [16]. A few of the well-characterized CNV disorders include Prader-Willi and Angelman syndromes (PWS/AS) [17], which result from del in 15q11-q13; Williams-Beuren syndrome (WBS) [18], which results from del in 7q11.23 and the 22q11.2 deletion syndrome (22q11.2 DS, previously referred to as DiGeorge/Velocardiofacial syndromes) [19]. In a majority of the known CNV disorders, affected individuals usually present with multiple congenital anomalies (MCA) including intellectual disability (ID), developmental delay (DD), dysmorphic features, cardiac defects, limb and digital abnormalities and seizures [20]. Most of these disorders were initially recognized due to a set of overlapping phenotypic features observed in a majority of the affected individuals. Subsequent genetic testing using targeted tests such as Fluorescence *In Situ* Hybridization (FISH) were then used to confirm association of the phenotype with specific del or dup.

Over the past several years, the increased application of microarray-based CNV analysis in clinical diagnostics, has led to an explosion in the number of genomic disorders and syndromes that have been directly associated with del and/or dup [20]. Many new microdeletion and microduplication syndromes associated with 1q21.1 [21], 16p11.2 [21] and 15q13.3 [23] have been described (Table 1). The ever-expanding list of these types of novel CNV disorders as well as rare, one-off CNVs associated with disease phenotypes, underscores the significance of CNV based mutations in genetic disorders. Not surprisingly, a majority of these newly described CNV disorders are also associated with MCA and ID/DD (Table 1).

CNVs associated with common complex diseases

CNVs have also been associated with the etiology of complex, neuropsychiatric diseases including autism spectrum disorders (ASD) and schizophrenia (SZ). Although ASD and SZ are amongst the variable phenotypes seen in individuals with genomic disorders like the 22q11.2 DS and WBS, systematic analysis of CNV association with these phenotypes were carried out more recently. The first few case-control based CNV studies of ASD, discovered an enrichment for *de novo* CNVs in cases versus controls [24,25]. While early studies of CNVs in SZ suggested an increased burden of rare CNVs in cases versus controls [26]. Subsequent studies in large cohorts detected enrichment of CNVs including del and/or dup in 1q21, 3q29, 7q11.23, 15q11.2, 15q13.3, 16p11.2 and 22q11.2 in cases with ASD and SZ versus controls [27–30].

CNVs have also been associated with neurodegenerative disorders including Parkinson disease (PD) and Alzheimer disease (AD). In addition to duplications and triplications of the *a-synuclein* gene locus in patients with PD [31,32], recent studies have suggested that CNVs may increase the risk for PD [33,34]. Copy number gains of *APP* due to trisomy 21 in Down syndrome or through dup are associated with early-onset familial AD [35]. CNV studies in patients with AD have detected enrichment of CNVs affecting dosage sensitive, brain-specific genes [36]. Furthermore, multiallelic CNVs that occur at high frequencies (>1%) in the population, have been associated with several complex diseases or traits including susceptibility to HIV infection, autoimmune diseases like systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), Crohn's disease, type 2 diabetes and obesity as well as roles in drug metabolism and toxicity [15].

CNVs involving some genes like the β -defensin cluster on 8p23.1 have been associated with multiple diseases including, Crohn's disease and several skin disorders including Psoriasis, Hidradenitis Supparativa/Acne Inversa [37] and dermatophytosis [38]. Similarly, CNVs of the Fc gamma receptor (FCGR) gene cluster on 1q23.3 have been associated with SLE [39], RA [40] and many other autoimmune disorders. However, many of these CNV associations have failed to replicate, especially in non-European populations [41]. It is believed that the lack of replication may be at least partially related to the limitations in the techniques used to assess copy number of these genes and the complex duplicated architecture of the genomic regions they are located within [15, 42]. Thus, the application of reliable and standard genotyping techniques to large cohorts will be necessary to assess the true association of multiallelic CNVs to complex diseases and traits.

Mechanisms of disease pathogenesis in CNV disorders

CNVs can lead to a disease phenotype by a variety of mechanisms based upon their size, location and gene content. The alteration in the copy number of one or more dosage sensitive genes, encompassed within a large microdeletion or microduplication, is the most common mechanism underlying CNV mediated disease pathogenesis. This is expected to be the mechanism for disease pathogenesis in most CNV disorders including the 22q11.2 DS, WBS, PWS/AS, Charcot-Marie Tooth disease type 1A/Hereditary Neuropathy with Pressure Palsies (CMT1A/HNPP) and Sotos syndrome [16–20]. Alternatively, an intragenic CNV may have the same effect as a disruptive point mutation, effectively inactivating the gene and leading to haploinsufficiency. Indeed, exonic or multiexonic CNVs within dosage sensitive genes represent a significant percentage of mutations in monogenetic disorders [43,44].

CNVs may also alter the expression of genes that are upstream or downstream of the CNV by disrupting regulatory elements like enhancers or promoters [45,46]. A 20 kb deletion which is 2.7 kb upstream of *IRGM* is associated with altered expression of the gene leading to Crohn's disease [47]). Similarly, a 47 kb deletion downstream of *SHOX*, which removes and enhancer element, has been associated with skeletal dysplasia and short stature [48]. CNVs may also reveal recessive disorders, wherein a deletion of one allele of the gene unmasks a point mutation on the other allele, resulting in disease. For example, in three individuals with unexplained intellectual disabilities, the discovery of a CNV that deleted the *COH1* gene led to the subsequent identification of a second pathogenic point mutation in *COH1*, leading to a diagnosis of Cohen syndrome, an autosomal recessive disorder [49]. Deletions of the non-mutated allele of *IL7R* have been uncovered by retrospective analysis of patients with autosomal recessive T-B+NK+ severe combined immunodeficiency (SCID), with a mutation previously detected in only one allele [50].

Mechanisms of CNV formation

CNVs associated with genetic disorders are classified as either recurrent or unique (non-recurrent) based on the rearrangement mechanisms underlying their formation. Recurrent CNVs are usually flanked by segmental duplications (SDs) or low copy number repeats (LCRs) [51]. Segmental duplications (SDs) are repetitive sequences in the human genome that are defined as being 1 kilobase (kb) in size and whose copies or paralogs share 90% sequence identity with each other and are estimated to account for approximately 5% of the human genome [52]. SDs associated with CNV disorders are often >100 kb, with a complex organization of duplicated modules that may share 96–99% sequence identity with paralogous copies [16–20]. The high level sequence identity shared by paralogous SDs can lead to misalignment between non-allelic SDs on homologous chromosomes. Subsequent recombination between the misaligned SDs can result in the formation of reciprocal deletion and duplication CNVs. This mechanism referred to as non-allelic homologous recombination (NAHR) is known to mediate genomic rearrangements associated with several genetic syndromes [16].

The presence of SDs flanking the genomic regions associated with several known syndromes implicated SD-mediated NAHR in the formation of CNVs associated with Prader-Willi

syndrome (PWS)/Angelman syndrome (AS) [17], Williams-Beuren syndrome (WBS) [18], DiGeorge syndrome (DGS)/velocardiofacial syndrome (VCFS) [and now collectively referred to as the 22q11.2 deletion syndrome (22q11.2DS) 19] and Smith-Magenis (SMS)/Potocki-Lupski syndromes [53]. Based on these findings and the large number of SDs identified in the human genome, suggested that SD-mediated NAHR could be associated with additional, yet unidentified disorders [54,55]. Indeed, the number of SD-mediated genomic disorders has expanded with the discovery of novel disorders [20].

The mechanisms underlying the formation of non-recurrent CNVs are not as well understood. Recent analyses of breakpoints in several non-recurrent CNVs has suggested the involvement of non-homologous end joining (NHEJ), and replication based mechanisms (RBMs) [16,56,57]. NHEJ is a mechanism by which double-stranded breaks (DSBs) are repaired in organisms ranging from bacteria to mammals (58–60). NHEJ does not require sequence homology to mediate recombination. However, the presence of SDs or other repetitive DNA elements appear to stimulate NHEJ mediated rearrangements associated with CNV formation [14,16]. The application of higher resolution techniques in the analysis of CNV breakpoints has suggested that mechanisms other NAHR or NHEJ may be involved in their formation [56,57]. This led to the proposal of RBMs, including break-induced replication (BIR), fork stalling and template switching (FoSTeS) and microhomology-mediated break-induced replication (MMBIR) [16,57]. RBMs are also likely to be stimulated by SDs and other repetitive DNA elements that promote the formation of unstable secondary structures, which can then lead to replication fork stalling or collapse leading to DNA breakage. The subsequent replicative repair of the broken DNA can lead to the formation of deletions, duplications and other rearrangements [16]. RBM-based mechanisms have been proposed in the formation of several non-recurrent CNVs including duplications of Xq28 including MECP2 [16,57].

The evolution of CNV detection methods

The technologies used for CNV detection have rapidly evolved in recent years. In the past, cytogenetic analysis by karyotyping allowed the detection of CNVs >10 Mb, in addition to aneuploidies and gross chromosomal rearrangements like translocations and inversions. However, the detection of CNVs <10 Mb required targeted, low-throughput methods like Fluorescence *In situ* Hybridization (FISH) and real-time PCR [61,62]; or medium throughput techniques such as multiplex amplifiable probe hybridization [63]. A number of technologies can detect copy number gains and losses throughout the human genome and their resolution is continuously increasing. The advent of genome-wide assays using high density microarrays [64] greatly enhanced our ability to detect CNVs at an unprecedented resolution.

High-resolution microarray-based CNV analysis provided a method to detect del and dup as well as novel CNVs that are undetectable by standard karyotype analysis or FISH [65]. CNV analysis using array-based comparative genomic hybridization (aCGH) or single nucleotide polymorphism (SNP)-based microarrays (SNP-arrays) provide efficient tools for researchers and clinicians. CNV microarrays can also detect low-level mosaicism, which would be missed by traditional methods used in cytogenetics [66,67]. Furthermore, SNP-arrays

provide genotype information that can be used to identify genomic regions with copy number neutral, loss of heterozygosity (LOH), which usually result from uniparental disomy (UPD) or parent of common origin effects [68,69]. Exon-focused CNV microarrays have been designed to allow the detection of exonic CNVs in targeted or genomewide assays, since a significant number of mutations in monogenetic disorders involve intragenic CNVs that affect one or a few exons of a gene [43,44, 70].

Thus, CNV microarrays have significantly increased diagnostic yield in genetic disorders, leading to their adoption by clinical laboratories as the first tier of testing for individuals with nonsyndromic intellectual disability (ID), autism spectrum disorders (ASD), and multiple congenital anomalies (MCA) [64,65]. Consequently, the identification of an underlying genetic etiology in patients with ID, ASD and MCA has significantly increased due to the application of CNV microarrays [20,71]. The impact of establishing a genetic diagnosis in congenital disorders is significant as it allows appropriate referrals for therapy, surveillance for potential future involvement of other tissues and organs and assist in better genetic counseling for the family [65]. With the increased application of CNV analysis on large patient cohorts, novel microdeletion and microduplication have been identified and associated with disease phenotypes [72]. CNV detection using microarrays remains the current mainstay of the clinical cytogenetic laboratory [64,65].

High-throughput sequencing (HTS) technologies are becoming more routine tools for the analysis of genetic variation in the human genome. The ability to detect all forms of genetic variation including single nucleotide variations (SNVs), small (<50 bp) insertions/deletions (InDels), CNVs and copy neutral structural variations on a single platform make HTS highly desirable in clinical diagnostics. The use of whole exome sequencing (WES) for the detection of exonic mutations in patients with Mendelian disorders has seen a rapid rise in clinical diagnostics [73]. Several algorithms have been developed to identify CNVs from WES data, most of which use read-depth based approaches for the detection of changes in copy number [74]. However, several limitations both, of WES methodology as well as the algorithms for CNV detection from WES data, prevent the reliable application of WES-based CNV detection in the clinical laboratory [75,76]. Whole genome sequencing (WGS) and analysis pipelines developed for variation detection have become increasingly robust over the past several years enabling the detection of thousands of novel CNVs and other structural variations [10–12]. Furthermore, the recent application of novel, long-read sequencing (77) and genome mapping (78) technologies have greatly enhanced the discovery of structural and copy number variations in whole-genome assays. Thus, the ongoing development and application of novel, high-throughput sequencing and mapping technologies holds great promise for the comprehensive detection of all forms of variation, especially CNVs, which will likely increase the diagnostic yield in patient cohorts.

Challenges in CNV association to disease

CNV analysis with microarrays is offered as routine clinical test, with the results being increasingly used in genetic counseling and clinical management [71]. However, the ability to detect CNVs using high throughput technologies has far outpaced the ability to determine their role in disease pathogenesis. A majority of the microarrays used in clinical laboratories

are designed to have a high density of probes to allow for maximum resolution for the capture of CNVs genomewide. An inadvertent outcome of this improved detection has been the increase in the detection of CNVs that are of unknown clinical significance. This is mainly because the interpretation of novel CNVs is difficult, leading to wide variability in assessment and reporting of the clinical significance of CNVs by different diagnostic laboratories [65]. Some guidelines that allow better interpretation of the clinical significance or disease association of CNVs include criteria such as, overlap with known contiguous gene syndromes; CNV size and gene content; whether the CNV is *de novo* or inherited; comparison to CNV databases and medical literature [71].

CNV data from healthy controls is available in public databases such as the, Database of Genomic Variants and NCBI's Database of genomic structural variation (dbVar) (see Web Resources). Further, clinically relevant CNVs are available from databases such as DECIPHER, ECARUCA, ClinGen and ClinVar (see Web Resources). CNV data generated from whole exome sequencing is now available in the Exome Aggregation Consortium (ExAC) database. These resources have become invaluable in discriminating pathogenic from non-disease associated CNVs. However, once whole genome sequencing (WGS) becomes a more routine test in clinical diagnostics, additional databases of CNVs detected by WGS will be required.

Conclusions

The role of CNVs in human disease as well as normal genetic variation has become increasingly apparent as a result of the higher resolution analysis of the genome. However, CNVs continue to be underappreciated as a source of mutation in disease mainly due to the limitations in the accurate detection of CNVs, especially those that are smaller, affect one or a few exons of a single gene, affect regulatory regions of genes or occur within genomic regions that contain complex genomic architecture, like segmental duplications. Furthermore, in order to associate CNVs with disease will require larger databases with CNV data from control samples analyzed using equivalent high resolution techniques. This appreciation of the limitations of current technology and resources for the detection and disease-association of CNVs has been critical in providing the impetus towards developing newer technologies and better data resources for a truly comprehensive analysis of all forms of variation within the human genome.

Acknowledgments

The author's research projects related to copy number variation were supported by grants from the National Institutes of Health (GM081519, DE025363 and GM120772). The author would like to thank all the personnel and trainees in his laboratory and collaborators for their contributions to projects involving copy number variation analysis.

References

1. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. *Nat. Genet.* 2004; 36:949–951. [PubMed: 15286789]
2. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. *Science.* 2004; 305:525–528. The two studies listed above (1

- and 2) were the first to report high levels of copy number variation/polymorphism in the genomes of apparently healthy individuals. [PubMed: 15273396]
3. Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK. A high-resolution survey of deletion polymorphism in the human genome. *Nat. Genet.* 2006; 38:75–81. [PubMed: 16327808]
 4. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat. Rev. Genet.* 2006; 7:85–97. [PubMed: 16418744]
 5. Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, et al. Mapping and sequencing of structural variation from eight human genomes. *Nature.* 2008; 453:56–64. This study used a novel, clone-based method to detect structural and copy number variations at a higher resolution in eight individuals of diverse geographic ancestry. [PubMed: 18451855]
 6. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, et al. Common deletion polymorphisms in the human genome. *Nat. Genet.* 2006; 38:86–92. [PubMed: 16468122]
 7. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature.* 2006; 444:444–454. [PubMed: 17122850]
 8. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, et al. Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* 2005; 77:78–88. [PubMed: 15918152]
 9. Shaikh TH, Gai X, Perin JC, Glessner JT, Xie H, Murphy K, et al. High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Res.* 2009; 19:1682–1690. [PubMed: 19592680]
 10. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, et al. Mapping copy number variation by population-scale genome sequencing. *Nature.* 2011; 470:59–65. This study reports the first results of copy number variation (CNV) analysis of samples sequenced by the 1000 Genomes Project (1000 GP), demonstrating the utility of population scale sequencing for structural variation (SV) analysis. [PubMed: 21293372]
 11. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. *Nature.* 2010; 464:704–712. [PubMed: 19812545]
 12. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. *Nature.* 2015; 526(7571):75–81. This study expanded on the work reported in #10 by cataloguing SVs and CNVs detected in 2504 individuals sequenced by the 1000 GP. [PubMed: 26432246]
 13. Lupski JR. Structural variation mutagenesis of the human genome: Impact on disease and evolution. *Environ Mol Mutagen.* 2015; 56(5):419–36. [PubMed: 25892534]
 14. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu. Rev. Med.* 2010; 61:437–55. [PubMed: 20059347]
 15. Girirajan S, Campbell CD, Eichler EE. Human copy number variation and complex genetic disease. *Annu Rev Genet.* 2011; 45:203–26. [PubMed: 21854229]
 16. Carvalho CM, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet.* 2016; 17(4):224–38. [PubMed: 26924765]
 17. Vogels A, Fryns JP. The Prader-Willi syndrome and the Angelman syndrome. *Genet Couns.* 2002; 13(4):385–96. [PubMed: 12558108]
 18. Bayés M, Magano LF, Rivera N, Flores R, Pérez Jurado LA. Mutational mechanisms of Williams-Beuren syndrome deletions. *Am J Hum Genet.* 2003; 73(1):131–51. [PubMed: 12796854]
 19. McDonald-McGinn DM, Sullivan KE, Marino B, Philip N, Swillen A, Vorstman JA, et al. 22q11.2 deletion syndrome. *Nat Rev Dis Primers.* 2015; 19(1):15071.
 20. Watson CT, Marques-Bonet T, Sharp AJ, Mefford HC. The genetics of microdeletion and microduplication syndromes: an update. *Annu Rev Genomics Hum Genet.* 2014; 15:215–244. [PubMed: 24773319]
 21. Brunetti-Pierri N, Berg JS, Scaglia F, Belmont J, Bacino CA, Sahoo T, et al. Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nature Genet.* 2008; 40(12):1466–71. [PubMed: 19029900]

22. Rosenfeld JA, Coppinger J, Bejjani BA, Girirajan S, Eichler EE, Shaffer LG, Ballif BC. Speech delays and behavioral problems are the predominant features in individuals with developmental delays and 16p11.2 del and dup. *J. Neurodev. Disorders*. 2010; 2(1):26–38.
23. van Bon BW, Mefford HC, Menten B, Koolen DA, Sharp AJ, Nillesen WM, et al. Further delineation of the 15q13 microdeletion and duplication syndromes: a clinical spectrum varying from non-pathogenic to a severe outcome. *J. Med. Genet*. 2009; 46(8):511–523. [PubMed: 19372089]
24. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong association of de novo copy number mutations with autism. *Science*. 2007; 316(5823):445–9. [PubMed: 17363630]
25. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, et al. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet*. 2008; 82(2):477–88. The above studies (24 and 25) were amongst the first to demonstrate the association of CNVs with autism spectrum disorders. [PubMed: 18252227]
26. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*. 2008; 320(5875):539–43. [PubMed: 18369103]
27. Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, Steinberg S, et al. Large recurrent del associated with schizophrenia. *Nature*. 2008; 455(7210):232–6. [PubMed: 18668039]
28. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*. 2008; 455(7210):237–41. The above studies (26, 27 and 28) were amongst the first to demonstrate the association of CNVs with schizophrenia. [PubMed: 18668038]
29. Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnúsdóttir B, Morgen K, Arnarsdóttir S, et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*. 2014; 505(7483):361–6. [PubMed: 24352232]
30. Kirov G. CNVs in neuropsychiatric disorders. *Hum Mol Genet*. 2015; 24(R1):R45–9. [PubMed: 26130694]
31. Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, et al. alpha-Synuclein locus triplication causes Parkinson's disease. *Science*. 2003; 302(5646):841. [PubMed: 14593171]
32. Ross OA, Braithwaite AT, Skipper LM, Kachergus J, Hulihan MM, Middleton FA, et al. Genomic investigation of alpha-synuclein multiplication and parkinsonism. *Ann Neurol*. 2008; 63(6):743–50. [PubMed: 18571778]
33. Liu X, Cheng R, Ye X, Verbitsky M, Kisselev S, Mejia-Santana H, et al. Increased Rate of Sporadic and Recurrent Rare Genic Copy Number Variants in Parkinson's Disease Among Ashkenazi Jews. *Mol Genet Genomic Med*. 2013; 1(3):142–154. [PubMed: 24073418]
34. Murthy MN, Veerappa AM, Seshachalam KB, Ramachandra NB. High-resolution arrays reveal burden of copy number variations on Parkinson disease genes associated with increased disease risk in random cohorts. *Neurol Res*. 2016; 38(9):775–85. [PubMed: 27399248]
35. McNaughton D, Knight W, Guerreiro R, Ryan N, Lowe J, Poulter M, et al. Duplication of amyloid precursor protein (APP), but not prion protein (PRNP) gene is a significant cause of early onset dementia in a large UK series. *Neurobiol Aging*. 2012; 33(2):426, e13–21.
36. Sekine M, Makino T. Inference of Causative Genes for Alzheimer's Disease Due to Dosage Imbalance. *Mol Biol Evol*. 2017; 34(9):2396–2407. [PubMed: 28666362]
37. Giamarellos-Bourboulis EJ, Platzer M, Karagiannidis I, Kanni T, Nikolakis G, Ulrich J, Bellutti M, et al. High Copy Numbers of β -Defensin Cluster on 8p23.1, Confer Genetic Susceptibility, and Modulate the Physical Course of Hidradenitis Suppurativa/Acne Inversa. *J Invest Dermatol*. 2016; 136(8):1592–8. [PubMed: 27164300]
38. Jaradat SW, Cubillos S, Krieg N, Lehmann K, Issa B, Piehler S, et al. Low DEFB4 copy number and high systemic hBD-2 and IL-22 levels are associated with dermatophytosis. *J Invest Dermatol*. 2015; 135(3):750–758. [PubMed: 25178103]
39. Mueller M, Barros P, Witherden AS, Roberts AL, Zhang Z, Schaschl H, et al. Genomic pathology of SLE-associated copy-number variation at the FCGR2C/FCGR3B/FCGR2B locus. *Am J Hum Genet*. 2013; 92(1):28–40. [PubMed: 23261299]

40. Rahbari R, Zuccherato LW, Tischler G, Chihota B, Ozturk H, Saleem S, et al. Understanding the Genomic Structure of Copy-Number Variation of the Low-Affinity Fc γ Receptor Region Allows Confirmation of the Association of FCGR3B Deletion with Rheumatoid Arthritis. *Hum Mutat.* 2017; 38(4):390–399. [PubMed: 27995740]
41. Cantsilieris S, White SJ. Correlating multiallelic copy number polymorphisms with disease susceptibility. *Hum Mutat.* 2013; 34(1):1–13. [PubMed: 22837109]
42. Cantsilieris S, Western PS, Baird PN, White SJ. Technical considerations for genotyping multi-allelic copy number variation (CNV), in regions of segmental duplication. *BMC Genomics.* 2014; 15:329. [PubMed: 24885186]
43. Boone PM, Bacino CA, Shaw CA, Eng PA, Hixson PM, Pursley AN, et al. Detection of clinically relevant exonic copy-number changes by array CGH. *Hum Mutat.* 2010; 31(12):1326–42. [PubMed: 20848651]
44. Wisniewska J, Bi W, Shaw C, Stankiewicz P, Kang SH, Pursley AN, et al. Combined array CGH plus SNP genome analyses in a single assay for optimized clinical testing. *Eur J Hum Genet.* 2014; 22(1):79–87. [PubMed: 23695279]
45. Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Hum Mol Genet.* 2015; 24(R1):R102–10. [PubMed: 26152199]
46. Chiang C, Scott AJ, Davis JR, Tsang EK, Li X, Kim Y, et al. The impact of structural variation on human gene expression. *Nat Genet.* 2017; 49(5):692–699. [PubMed: 28369037]
47. McCarroll SA, Huett A, Kuballa P, Chlewicki SD, Landry A, Goyette P, et al. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat Genet.* 2008; 40(9):1107–12. [PubMed: 19165925]
48. Verdin H, Fernández-Miñán A, Benito-Sanz S, Janssens S, Callewaert B, De Waele K, et al. Profiling of conserved non-coding elements upstream of SHOX and functional characterisation of the SHOX cis-regulatory landscape. *Sci Rep.* 2015; 5:17667. [PubMed: 26631348]
49. Rivera-Brugués N, Albrecht B, Wiczorek D, Schmidt H, Keller T, Göhring I, et al. Cohen syndrome diagnosis using whole genome arrays. *J Med Genet.* 2011; 48:136–140. [PubMed: 20921020]
50. Engelhardt KR, Xu Y, Grainger A, Germani Batacchi MG, Swan DJ, Willet JD, et al. Identification of Heterozygous Single- and Multi-exon Deletions in IL7R by Whole Exome Sequencing. *J Clin Immunol.* 2017; 37(1):42–50. [PubMed: 27807805]
51. Bailey JA, Eichler EE. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat Rev Genet.* 2006; 7(7):552–64. [PubMed: 16770338]
52. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001; 409(6822):860–921. [PubMed: 11237011]
53. Potocki L, Bi W, Treadwell-Deering D, Carvalho CM, Eifert A, Friedman EM, et al. Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype. *Am J Hum Genet.* 2007; 80(4):633–49. [PubMed: 17357070]
54. Emanuel BS, Shaikh TH. Segmental duplications: an 'expanding' role in genomic instability and disease. *Nature Rev Genet.* 2001; 2(10):791–800. [PubMed: 11584295]
55. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, et al. Recent segmental duplications in the human genome. *Science.* 2002; 297(5583):1003–1007. This study was the first to carry out a comprehensive analysis of the reference genome sequence generated by the Human Genome Project to identify segmental duplications. [PubMed: 12169732]
56. Lee J, Carvalho C, Lupski JR. A DNA Replication Mechanism for Generating Nonrecurrent Rearrangements Associated with Genomic Disorders. *Cell.* 2007; 131(7):1235–1247. [PubMed: 18160035]
57. Zhang F, Khajavi M, Connolly A, Towne CF, Batish SD, Lupski JR. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nature Genetics.* 2009; 41(7):849–853. [PubMed: 19543269]
58. Weller GR, Kysela B, Roy R, Tonkin LM, Scanlan E, Della M, et al. Identification of a DNA nonhomologous end-joining complex in bacteria. *Science.* 2002; 297(5587):1686–1689. [PubMed: 12215643]

59. Weterings E, van Gent DC. The mechanism of non-homologous end-joining: a synopsis of synopsis. *DNA Repair (Amst)*. 2004; 3(11):1425–35. [PubMed: 15380098]
60. van Gent DC, van der Burg M. Non-homologous end-joining, a sticky affair. *Oncogene*. 2007 Dec 10; 26(56):7731–40. [PubMed: 18066085]
61. D'haene B, Vandesompele J, Hellemans J. Accurate and objective copy number profiling using real-time quantitative PCR. *Methods*. 2010; 50:262–270. [PubMed: 20060046]
62. Weaver S, Dube S, Mir A, Qin J, Sun G, Ramakrishnan R, Jones RC, Livak KJ. Taking qPCR to a higher level: Analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution. *Methods*. 2010; 50:271–276. [PubMed: 20079846]
63. den Dunnen JT, White SJ. MLPA and MAPH: sensitive detection of deletions and duplications. *Curr Protoc Hum Genet*. 2006; (Unit 7.14) Chapter 7.
64. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, 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*. 2010; 86:749–764. [PubMed: 20466091]
65. Coughlin CR 2nd, Scharer GH, Shaikh TH. Clinical impact of copy number variation analysis using high-resolution microarray technologies: advantages, limitations and concerns. *Genome Med*. 2012; 4(10):80. [PubMed: 23114084]
66. Scott SA, Cohen N, Brandt T, Toruner G, Desnick RJ, Edelmann L. Detection of low-level mosaicism and placental mosaicism by oligonucleotide array comparative genomic hybridization. *Genet Med*. 2010; 12:85–92. [PubMed: 20084009]
67. Conlin LK, Thiel BD, Bonnemant CG, Medne L, Ernst LM, Zackai EH, et al. Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. *Hum Mol Genet*. 2010; 19:1263–1275. [PubMed: 20053666]
68. Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. *Hum Mol Genet*. 2006; 15:789–795. [PubMed: 16436455]
69. Li L-H, Ho S-F, Chen C-H, Wei C-Y, Wong W-C, Li L-Y, et al. Long contiguous stretches of homozygosity in the human genome. *Hum Mutat*. 2006; 27:1115–1121. [PubMed: 16955415]
70. Tayeh MK, Chin ELH, Miller VR, Bean LJH, Coffee B, Hegde M. Targeted comparative genomic hybridization array for the detection of single- and multiexon gene deletions and duplications. *Genet Med*. 2009; 11:232–240. [PubMed: 19282776]
71. Martin CL, Kirkpatrick BE, Ledbetter DH. Copy number variants, aneuploidies, and human disease. *Clin Perinatol*. 2015; 42(2):227–42. [PubMed: 26042902]
72. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. *Nat Genet*. 2011; 43(9):838–46. This study demonstrated the power of analyzing large cohorts of patient samples in order to detect very rare, but recurrent CNVs associated disease. [PubMed: 21841781]
73. Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, et al. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am J Hum Genet*. 2015; 97(2):199–215. [PubMed: 26166479]
74. Zhao M, Wang Q, Wang Q, Jia P, Zhao Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC Bioinformatics*. 2013; 14(Suppl 11):S1.
75. Yao R, Zhang C, Yu T, Li N, Hu X, Wang X, Wang J, Shen Y. Evaluation of three read-depth based CNV detection tools using whole-exome sequencing data. *Mol Cytogenet*. 2017; 10:30. [PubMed: 28852425]
76. Zare F, Dow M, Monteleone N, Hosny A, Nabavi S. An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. *BMC Bioinformatics*. 2017 May 31; 18(1):286. [PubMed: 28569140]
77. Huddleston J, Chaisson MJP, Steinberg KM, Warren W, Hoekzema K, Gordon D, et al. Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Res*. 2017; 27(5):677–685. [PubMed: 27895111]

78. Mak AC, Lai YY, Lam ET, Kwok TP, Leung AK, Poon A, et al. Genome-Wide Structural Variation Detection by Genome Mapping on Nanochannel Arrays. *Genetics*. 2016; 202(1):351–62. [PubMed: 26510793]
79. Rudd MK, Keene J, Bunke B, Kaminsky EB, Adam MP, Mulle JG, et al. Segmental duplications mediate novel, clinically relevant chromosome rearrangements. *Hum Mol Genet*. 2009; 18(16): 2957–62. [PubMed: 19443486]
80. Ciccone R, Mattina T, Giorda R, Bonaglia MC, Rocchi M, Pramparo T, Zuffardi O. Inversion polymorphisms and non-contiguous terminal deletions: the cause and the (unpredicted) effect of our genome architecture. *J Med Genet*. 2006; 43(5):e19-e19. [PubMed: 16648372]
81. Stankiewicz P, Kulkarni S, Dharmadhikari AV, Sampath S, Bhatt SS, Shaikh TH, et al. Recurrent deletions and reciprocal duplications of 10q11.21q11.23 including CHAT and SLC18A3 are likely mediated by complex low-copy repeats. *Hum Mutat*. 2012; 33(1):165–79. [PubMed: 21948486]
82. Van Bon BW, Balciuniene J, Fruhman G, Nagamani SC, Broome DL, Cameron E, et al. The phenotype of recurrent 10q22q23 deletions and duplications. *Eur J Hum Genet*. 2011; 19(4):400–408. [PubMed: 21248748]
83. El-Hattab AW, Zhang F, Maxim R, Christensen KM, Ward JC, Hines-Dowell S, et al. Deletion and duplication of 15q24: molecular mechanisms and potential modification by additional copy number variants. *Genet Med*. 2010; 12(9):573–586. [PubMed: 20860070]
84. Ballif BC, Hornor SA, Jenkins E, Madan-Khetarpal S, Surti U, Jackson KE, et al. Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2. *Nat Genet*. 2007; 39(9):1071–1073. [PubMed: 17704777]
85. Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A, et al. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet*. 2010; 42(3): 203–9. [PubMed: 20154674]
86. Ullmann R, Turner G, Kirchhoff M, Chen W, Tonge B, Rosenberg C, et al. Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum Mutat*. 2007; 28(7):674–82. [PubMed: 17480035]
87. de Kovel CG, Trucks H, Helbig I, Mefford HC, Baker C, Leu C, et al. Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. *Brain*. 2010; 133(1):23–32. [PubMed: 19843651]
88. Mefford HC, Clauin S, Sharp AJ, Moller RS, Ullmann R, Kapur R, et al. Recurrent reciprocal genomic rearrangements of 17q12 are associated with renal disease, diabetes, and epilepsy. *Am J Hum Genet*. 2007; 81(5):1057–1069. [PubMed: 17924346]
89. Koolen DA, Vissers LE, Pfundt R, de Leeuw N, Knight SJ, Regan R, et al. A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat Genet*. 2006; 38(9):999–1001. [PubMed: 16906164]
90. Ben-Shachar S, Ou Z, Shaw CA, Belmont JW, Patel MS, Hummel M, et al. 22q11.2 distal deletion: a recurrent genomic disorder distinct from DiGeorge syndrome and velocardiofacial syndrome. *Am J Hum Genet*. 2008; 82(1):214–221. [PubMed: 18179902]
91. Ou Z, Berg JS, Yonath H, Enciso VB, Miller DT, Picker J, et al. Microduplications of 22q11.2 are frequently inherited and are associated with variable phenotypes. *Genet Med*. 2008; 10(4):267–277. [PubMed: 18414210]
92. Nizon M, Andrieux J, Rooryck C, de Blois MC, Bourel-Ponchel E, Bourgois B, et al. Phenotype-genotype correlations in 17 new patients with an Xp11.23p11.22 microduplication and review of the literature. *Am J Med Genet A*. 2015; 167A(1):111–22. [PubMed: 25425167]

Table 1

CNV disorders identified by microarray analysis (since 2006)

Genomic Region	CNVs	Phenotypes observed	Refs
1q21.1	Del/Dup	MCA, Variable ID, cataracts (del); Heart defects (del/dup)	21
2q11.2	Del/Dup	ID, congenital heart defect(del); DD, hypotonia, macrocephaly (dup)	79
2q13	Del	ID, DD, anxiety, hypotonia	79
7q36.1	Del	DD, speech delay, feeding difficulties	79
8p23.1	Del/Dup	ID, DD, congenital heart disease, behavioral abnormalities	80
10q11.21-q11.23	Del/Dup	ID, DD, MCA, dysmorphic features, seizures	81
10q22-q23	Del/Dup	ID, behavioral abnormalities, hyperactivity	82
15q13.3	Del/Dup	DD, learning disabilities, seizures, behavioral abnormalities	23
15q24	Del/Dup	ID, MCA, dysmorphic features, digital abnormalities	83
16p11.2	Del/Dup	macrocephaly (del)/microcephaly (dup) Learning disability, speech delay, congenital anomaly (both)	22
16p11.2p12.2	Del	ID, DD, dysmorphic facial features	84
16p12.1	Del/Dup	DD, MCA, neuropsychiatric disorders	85
16p13.11	Del/Dup	ID, MCA, neuropsychiatric disorders, epilepsies	86,87
17q12	Del/Dup	ID, renal disease, diabetes, epilepsies	88
17q21.31	Del/Dup	ID, DD, learning disability, dysmorphic features	89
Distal 22q11.23	Del/Dup	ID, MCA (del); Variable ID (dup)	90, 91
Xp11.22-p11.23	Dup	ID, speech delay, EEG abnormalities	92

Abbreviations: Refs: References, ID: Intellectual Disability, DD: Developmental Delay, MCA: Multiple Congenital Anomalies, del: Microdeletion, dup: Microduplication