



Aatira Vijay¹, Iti Garg¹ and Mohammad Zahid Ashraf²

¹Genomics Division, Defence Institute of Physiology & Allied Sciences, Delhi, India.

²Genomics Division, Defence Institute of Physiology and Allied Sciences, DRDO, Delhi, India.

ABSTRACT: Human genome contains many variations, often called mutations, which are difficult to detect and have remained a challenge for years. A substantial part of the genome encompasses repeats and when such repeats are in the coding region they may lead to change in the gene expression profile followed by pathological conditions. Structural variants are alterations which change one or more sequence feature in the chromosome such as change in the copy number, rearrangements, and translocations of a sequence and can be balanced or unbalanced. Copy number variants (CNVs) may increase or decrease the copies of a given region and have a pivotal role in the onset of many diseases including cardiovascular disorders. Cardiovascular disorders have a magnitude of well-established risk factors and etiology, but their correlation with CNVs is still being studied. In this article, we have discussed history of CNVs and a summary on the diseases associated with CNVs. To detect such variations, we shed light on the number of techniques introduced so far and their limitations. The lack of studies on cardiovascular diseases to determine the frequency of such variants needs clinical studies with larger cohorts. This review is a compilation of articles suggesting the importance of CNVs in multitude of cardiovascular anomalies. Finally, future perspectives for better understanding of CNVs and cardiovascular disorders have also been discussed.

KEYWORDS: copy number variations, cardiovascular diseases, structural variations

RECEIVED: October 31, 2018. **ACCEPTED:** November 8, 2018.

TYPE: Review

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Mohammad Zahid Ashraf, Genomics Division, Defence Institute of Physiology & Allied Sciences, DRDO, Lucknow Road, Timarpur, Delhi 110054, India. Email: mohammadzashraf@gmail.com

Introduction

Copy number variants (CNVs) have gained a lot of interest over the past few years due to the complexities they offer. Apart from altering the expression and phenotype, CNVs are responsible for genetic diversity in population as well. Copy number variants (shown in Figure 1) are sequence of DNA of 50 base pairs (bp) or larger compared with its reference genome and collectively represent, duplications, complex multisite variants, and smaller elements called insertion and deletions (indels).¹ Terminal deletions occur at the terminal region of the chromosome and are combined with neuropsychiatric complications, retarded growth, and dysmorphic facial features as shown in Cri-du-Chat syndrome²; micro-deletions are deletions up to 5 Mb which occur at 16p11.2 in Autism³; duplications include micro-duplications, for example, in 1q21.1 micro-duplication syndrome.⁴ One of the classic examples is duplication of an entire chromosome: chromosome 21 in Down syndrome.⁵ Another type of variants is multisite variants that contribute to the genetic diversity between individuals. Other than mentioned, inversion that is flipped or reversed orientation of chromosomal segments has been established in mutation in Factor VIII gene, which plays a critical role in coagulation and leads to hemophilia A; this mutation is found to be responsible in 43% of the individuals suffering from hemophilia.⁶ Translocation is shifting of large segments of DNA, for example, t(1;11) translocation is considered to have an association with increased risk of psychopathology due to decreased glutamate levels and altered cortical structure and function in the left temporal lobe of the brain.⁷ Microdeletions, a type of interstitial deletions, are

deletions of a very small portion around 1 to 2 Mb of the chromosome, whereas microduplications are comparatively rare, but are prominent reason behind neurodevelopmental genetic diseases.⁸ Copy number variants, when not tolerated, have been identified as a potent genetic variation, regulating both genotype and phenotype in an individual's susceptibility for a disease condition. The risk of changes in phenotype and occurrence of the disease is more of a de novo event, as CNVs are the rearrangements which may not present in the parent karyotype but offspring may carry the changes. Such variants are common among healthy individuals, however, but not necessarily affect the health and phenotype, for example, both deletions and duplications are found in HLA class III genes encoding complement proteins, C2 and C4, which are relatively common.⁹ Apart from the genome-wide CNVs, there are mitochondrial DNA copy numbers (mtDNA-CN), which decide the number of mitochondria and mitochondrial genomes per mitochondrion per cell and variation which may lead to mitochondrial dysfunction.¹⁰ The mtDNA-CN have been studied for predicting type 2 diabetes in a cohort that included 8 years of follow-up.¹¹ Also, leukocyte telomere length and mtDNA-CN and deletions have been proposed as risk markers for breast cancer in a European Prospective Investigation into Cancer and Nutrition (EPIC) study.¹² Interestingly, other species have also been under the influence of CNVs such as rat, mouse strains, and rhesus macaque, but human diseases such as genomic and monogenic disorders, infectious diseases, autoimmune diseases, and carcinomas have well-known links with CNVs.^{13–16} They affect the expression level by altering the transcription via



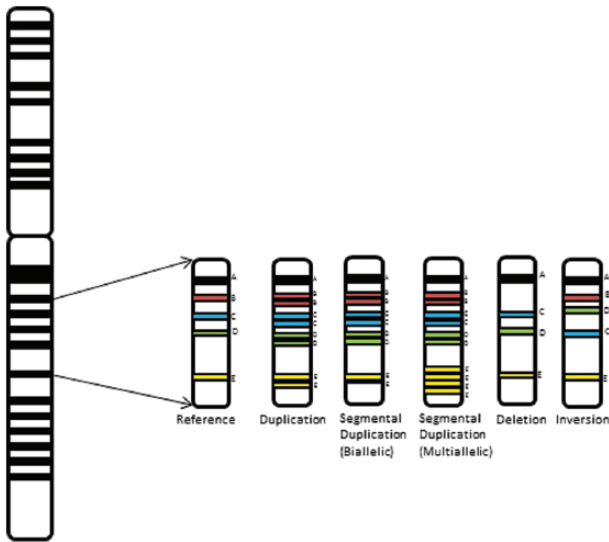


Figure 1. Copy number variation in the chromosome showing normal (reference) and CNV carrying chromosome regions.

disturbing the cis or trans-regulatory regions and other functional regions.¹⁷ The number of copies of a gene present in the coding region is directly proportional to the number of transcript, thus facilitating more translation or vice versa.¹⁸ However, studies have also advocated the parallel effect of copy number alterations in genes of single chromosome leading to detrimental phenotypes.¹⁹ Figure 1 represents human chromosome, for example, with a segment as reference region, followed by different variation occurring in the reference region.

Cardiovascular diseases (CVDs) include the diseases of heart and blood vessels including both arteries and veins. Cardiovascular diseases are leading cause of mortality and morbidity worldwide. Cardiovascular diseases include a broad range of disorders including heart and associated vascular system. Most commonly appeared CVDs are coronary artery disease (CAD) and stroke; other than these two, there are myocardial infarction (MI), congenital heart disease, cardiomyopathy, cardiac arrhythmia, heart valve disease, and other vasculature-related diseases such as atherosclerosis, venous thrombosis, pulmonary embolism, and peripheral artery disease. A report from World Health Organization has stated that around 31% of the total deaths in all over the world were caused by CVDs. Different epidemiologic studies on various continents have conducted to assess the early onset in the population, progression of the disease, prevalence, treatment, socioeconomic background of the individuals, and other factors contributing to CVDs.^{20–22} As the risk of CVDs is increasing significantly by the time, the research has been shifted to establish major aspects contributing to the disease, whereas most of the efforts have been put to understand the mechanism of the disease and the common predisposing factors; the global health care services in the field of cardiovascular research are yet to be improvised.²³ However, despite all the research attempts to identify the disease progression during early stages, increased risks have been a major barrier to decrease the global burden of

CVDs as they account for one-half of all the deaths caused by noncommunicable diseases worldwide.²⁴

Having discussed the severity and prevalence of the CVDs and their known risk factors, including hypertension, cholesterol levels in blood, obesity, smoking, diabetes, and others, a major survey on the genetic risk factors is essential to complement the research.²⁵ Major advancements in the field of genetics have proposed genetic variants accountable for predisposing the CVD risk, among which single-nucleotide polymorphisms (SNPs) have majorly been studied due to the improvised technologies for detection, identification, and characterization of annotated as well as novel SNPs. Genome-wide association studies (GWAS) have been considered the most significant studies so far to detect a number of genetic risks for CVDs. These proposed risk factors could be considered as established genetic markers for risk stratification; however, they require more experimental proofs and validation.

Mechanism of Copy Number Variation in Genome

Alterations in the gene copy number are basically changes in the position of a chromosomal region in the chromosome itself; the segments could be of different sizes depending on the type of variation occurred. Mechanism of copy number changes is well studied in model organisms such as *Drosophila*, *Saccharomyces cerevisiae*, and *Escherichia coli*. The reason behind copy number changes in the chromosomes is basically recombination of 2 types: homologous recombination which occurs on more identical sequences or sequences with higher similarity, whereas nonhomologous recombination, on the contrary, requires very limited or microhomology or no homology at all; such recombination occur during replication mechanism as reviewed in a recent study.²⁶ As described in a recent article, when there is extensive similarity in the sequence of the chromosomal segments, nonallelic homologous recombination (NAHR) occurs; the segments with no homology or very less similarity have nonhomologous end-joining and microhomology-mediated end-joining recombination, respectively, for the origin of CNVs. In addition, other common replication errors may also lead to CNVs, such as fork stalling and template switching and microhomology-mediated break-induced replication and transposable element-mediated mechanisms.²⁷ Among all, NAHR are best known to contribute toward recurrent or frequently occurring CNVs; the nonrecurrent-type CNVs are produced by replication-based errors.²⁸ Although the knowledge regarding the origin and development of CNVs have been cited in various studies, mechanisms including damaged DNA followed by its repair mechanism having higher incidences of nonhomologous recombination-induced copy number changes are still unclear.

History of Copy Number Variations and Links with CVDs

A brief study published in 2004 strongly recommended the global role of large-scale copy number variations in human

genome from heritable gain and loss of DNA for the first time; previously, there was no such breakthrough in the field of structural variants.²⁹ The study was conducted with 55 healthy individuals and showed the genetic differences in their genome in the form of structural variants. Going back to the discovery of CNVs in 1987, Nakamura et al³⁰ proposed variable number of tandem repeats (VNTRs), which are heritable regions in the chromosome where short DNA sequence motifs are present in repeats in random variable number, as the name suggests. The complex mutation pattern of one of the multifaceted disorders related to VNTR can be found in dopamine receptor D4 gene exon3 VNTR association in attention deficit hyperactivity disorder in which European and Asian ethnicities display 7- and 2-repeats, respectively.³¹ After VNTRs, in 1989, microsatellite came into light for their distinguished role in complex multifactorial disorders and later they were used as a biomarker for gene map studies in humans; furthermore, these microsatellites were used widely to make the linkage map of all human chromosomes.^{32,33} In a recently published article, authors have provided the evidence that microsatellites may act as enhancers and regulate the expression of genes.³⁴ In later years, these genomic rearrangements, both balanced and imbalanced, were characterized as mutational changes in genome responsible for various ailments.¹⁸ A classic example for such rearrangements is of a muscular atrophy called Charcot-Marie-Tooth disease type IA, in which the CNV responsible for the disease originated in the upstream of peripheral myelin protein 22 (PMP22) gene, which earlier was not documented in the human structural variants database.³⁵ After a long leap, in 2002, after almost 10 years, structural variants came into light as paralogous sequence variants (variants in few nucleotides in closely related gene families).^{36,37} One of the most prominent discoveries in this field was done by Redon et al,³⁸ identifying 1447 copy number variable regions in the genome of 270 healthy individuals from 4 different ethnicities, which suggested that around 12% part of the human genome has CNVs.

Despite the significant improvements in the technologies to identify CNVs, the proposed genetic risk factors for CVDs are not considered clinically significant so far. Notably, a study has challenged the notion for CADs that classical risk factors are responsible for over 50% of the cases, whereas a careful analysis of the articles published in the past has indicated that in fact over 80% of CAD cases have come up with nonclassical risk factors.³⁹ A few diseases with cardiovascular anomalies have been registered previously with smaller CNVs or large cytogenetic changes as the underlying cause. One of the pioneer studies in early 1980s on congenital polyvalvular disease (CPVD) suggested that the presence of an extra copy of chromosomal part of chromosome 18, called trisomy-18 syndrome, leads to congenital anomalies, especially CPVD.⁴⁰ Such diseases resulted after larger changes in the chromosomes, which could be visualized through microscope. Another study from late 1980s karyotyped 41 cases of trisomy-18, among which almost in all autopsy samples different forms of cardiac anomalies

were pronounced including pulmonary, tricuspid, aortic and mitral valve anomalies, polyvalvular disease, and so on.⁴¹ Since 1990s, there were comparatively more studies regarding structural variations in the chromosome linked with cardiac abnormalities.⁴²⁻⁴⁴ However, the problem was to find out clinically significant CNVs and to link them with the cardiac diseases because many variations are found in healthy population as well, which required improvement in the detection and identification technologies with experimental validation.

Role of Copy Number Variations as Compared With SNP in CVDs

Single-nucleotide polymorphisms on the contrary to the CNVs occur due to a variation in a single nucleotide in the gene, which may influence the transcription by affecting promoter activity or messenger RNA stability; as a consequence, they are reported to have association with several diseases in humans.⁴⁵ Previously, candidate gene studies to identify presence or absence of common genetic variants were used for disease association analysis which lacked the reproducibility and it was practically not possible to test higher number of polymorphisms for larger cohorts. Genome-wide association studies have emerged as one of the most powerful tools to investigate the whole genome for SNPs in particular.⁴⁶ Genetic studies have led to upsurge the understanding of CVD association with SNPs. For instance, the clinical data suggest that protein levels of Apolipoprotein E (ApoE) in plasma are strikingly associated with the risk for coronary heart disease; hence, common polymorphisms affecting this gene in turn affect the protein level leading to the disease.⁴⁷ Common variants in ATP-binding cassette transporter 1 (ABCA1) gene, low-density lipoprotein receptor gene (LDLR), and others are reported to have association with cardiac diseases.^{48,49} Over the past few years, SNPs have been considered as a potential risk factor for genetic diseases in humans; CNVs, however, were detected lately but their potency to alter the genomic functions and phenotype was characterized in no time.

Copy number variants span more than 50 bases in the genome and are responsible for more variations in the chromosome, leading to disruption of the functional large genomic segments than do SNPs.¹⁰ Copy number variants present in the coding region directly influence the copies of a gene during transcription, hence contributing to the progression of diseases, whereas if CNVs occur in the noncoding region of a gene, they may not influence the copy numbers. Around 4 Mb of CNVs is present in the whole human genome, whereas SNPs account for 2.5 Mb, suggesting CNV accountability for most of the genetic variations in humans.⁵⁰ Many CVDs have been identified having association with SNPs, but the share of CNVs in increasing the complexity of human genome and link with the diseases cannot be overlooked. A review suggesting implications of CNVs in CVDs has underlined majorly monogenic disorders suggesting more investigation in this field.⁵¹ A genome-wide case-control study of CNVs has suggested a

Table 1. Different cardiovascular disorders associated with CNVs and its detection via comprehensive arrays.

S. NO.	CARDIOVASCULAR DISEASE TYPE	TYPE OF CNV	TECHNIQUE USED FOR DETECTION	REFERENCES
1	Dilated cardiomyopathy	Deletion in exon 4 of HSP-co-chaperone BCL2-associated athanogene 3 (BAG3)	Exome sequencing	Norton et al ⁵²
2	Williams-Beuren syndrome	7q11.23 microduplication syndrome	MLPA and array-CGH	Van der Aa et al ⁵⁷
3	d-transposition of the great arteries (d-TGA) and Ebstein's anomaly	Microduplication of 22q11.2 region	FISH	Laitenberger et al ⁵⁸
4	Congenital heart defects (CHD)	Clinical interpretation of known CNVs in CHD	Array-CGH	Breckpot et al ⁵⁴
5	Congenital heart defects	Recurrent and de novo CNVs	SNP arrays and whole exome sequencing	Glessner et al ⁵⁹
6	Myocardial infarction	Rare and common CNVs	GWAS	Kathiresan et al ⁶⁰
7	Coronary artery disease	Increased copy number in LPA gene	Taqman real-time qPCR assay	Wu et al ⁶¹
8	Stroke	Insertions and deletions	Genome-wide SNP array	Matarin et al ⁶²

Abbreviations: Array-CGH, array comparative genomic hybridization; CNVs, copy number variations; FISH, fluorescence in situ hybridization; GWAS, genome-wide association study; HSP, heat shock protein; LPA, lipoprotein(A); MLPA, multiplex ligation-dependent probe amplification; qPCR, quantitative polymerase chain reaction; SNP, single-nucleotide polymorphism.

potent role of rare variants in large cohort responsible for dilated cardiomyopathy, a heart disorder leading to various cardiovascular complications.⁵² Congenital heart disease (CHD) is a primary birth defect leading to early death of newborns worldwide. Reports are available on rare CNVs present in heterotaxy: a type of CHD and also in syndromic and nonsyndromic congenital heart defects on a cohort of 46 nonsyndromic individuals.^{53,54} Copy number variants play a major role in various blood disorders; for example, CNVs in the α -globin genes of hemoglobin lead to α thalassemia, a very common blood disorder, in which deletion in the α -globin gene makes rare disease onset, whereas duplication in the same gene causes severe phenotypic change in β -thalassemic carriers.⁵⁵ Furthermore, CNVs influence important regulatory sites in the gene from distance, leading to dynamic changes in the expression pattern.⁵⁶

CNVs in Cardiovascular Disorders

Cardiovascular diseases have widely been studied on the basis of their origin, type of disorder, organ, and vessels affected and their genetic inheritance, but very few studies are available corroborating the link between CNVs and CVDs. Table 1 represents different CVDs associated with CNVs and different methods to detect those in the genome.

Congenital Heart Disease

Congenital heart disease is one of the major birth defects, leading to cardiac malformations in around 7 in every 1000 live births, resulting in higher incidences of mortality in infants.⁶³

Children with nonsyndromic CHD had a higher risk of potentially pathogenic CNVs when compared with healthy ones and found to be at higher risk of mortality.⁶⁴ Another study to determine pathogenic chromosomal abnormalities was conducted in fetuses diagnosed with CHD by microarray analysis. The results of the study concluded that a significant percentage of cases were found to be associated with different chromosomal anomalies supporting other same studies.⁶⁵ Furthermore, CNVs have also been implicated in congenital left-sided cardiac lesions. A recent study observed large CNVs were in 33 probands (~3%), out of 1139 probands.⁶⁶ Also, global rare copy number variants and de novo mutations in histone-modifying genes have been linked to the risk of sporadic CHD, earlier.^{67,68} High-resolution genome-wide analysis have shown that coarctation of the aorta (CoA), which accounts for 5% to 8% of all CHDs, is due to abundance of large CNVs on the X chromosome of affected men.⁶⁹ Also, in patients with CoA, CNVs alter the expression of genes regulated by FOXC1.⁷⁰

Myocardial Infarction

Myocardial infarction is a prolonged condition of lack of oxygen to heart muscles. A GWAS has identified major SNPs as well as common and rare CNVs for possible association with early-onset MI in larger cohorts. The study tested 554 common CNVs and SNPs at 8 loci in patients, as compared with controls.⁶⁰ Another study has observed 7 CNV segments associated with the patients of hyperlipidemia (a major cause of MI) and MI suggesting the potential role of structural variants in MI.⁷¹

CAD and Hypertrophic Cardiomyopathy

Coronary artery disease also called ischemic heart disease is presented by the patients with MI and stroke, making it one of the leading causes of deaths worldwide. A study investigated association between CNV in lipoprotein(a) (LPA) in CAD cases. Higher serum concentration of LPA is associated with increased risk of CAD; however, in this study, a protective role of 1 copy number of the LPA gene has been identified in southern Han Chinese population majorly in women and the elderly.⁶¹ Hypertrophic cardiomyopathy (HCM) is a condition in which the myocardium becomes abnormally thick (hypertrophied), without an obvious reason. People have used high-throughput sequencing technique to screen CNV in genes associated with cardiomyopathy.⁷² Also, people have studied minor genes and CNVs in HCM.⁷³

Essential Hypertension

Essential hypertension, a multifactorial disease, possesses substantial genetic risk factors including CNVs. Deletion of the CNVs *esv27061* and *esv2757747* on the chromosome 1p13.2 was found to be significantly higher in patients with hypertension identified by droplet digital polymerase chain reaction (PCR) technique.⁷⁴ Another study on different strains of rats has confirmed the pivotal role CNVs in hypertension by whole genome array.⁷⁵

Stroke

Stroke is the third leading cause of mortality worldwide after cancer and other heart diseases. A study has investigated the role of CNVs in modulating risks for ischemic stroke. The analysis showed 187 insertions (76%; 135 heterozygous, 25 homozygous duplications or triplications, 2 heterosomic) as well as 60 deletions (24%; 40 heterozygous deletions, 3 homozygous deletions, 14 heterosomic deletions) in the patients when compared with controls; among these, reported CNVs were previously reported.⁶² A very recent detailed description of CNVs influencing stroke probability in individuals has reported and confirmed the positive association of stroke with CNVs.⁷⁶ Evidence are also available for microduplications involved in stroke.⁷⁷

Diseases Other than Cardiovascular Types and their Association With CNVs

Several studies have been done to link the cord between diseases other than cardiovascular types and their phenotypes with CNVs. A number of diseases having a background of structural variations are listed in Table 2. Moreover, an increase or a decrease in the copy numbers of one gene may lead to different diseases as shown in Table 2. A population study has suggested the role of segmental duplication encompassing the gene chemokine (C-C motif) ligand 3-like 1 (CCL3L1) increasing the susceptibility toward HIV-associated immune deficiency or AIDS.⁹⁰ Furthermore, enhanced expression of

the same gene has been underlined as a risk factor for rheumatoid arthritis in a different case-control study.⁷⁹ Moreover, there are major structural rearrangements in the genes which are considered to be rare phenomena but they have their share of clinical significance as shown in Table 1. Among these, some changes or rearrangements were so large that they were cytogenetically visible, for example, more than 4 to 5 copy numbers of gene β -defensin showed significant risk of psoriasis, whereas the low copy numbers of the same gene are considered as a risk for Crohn disease.^{83,84} Besides duplication, deletions in the late cornified envelope 3B and 3C genes have been identified in case of patients with psoriasis.⁸⁵ Despite having an established role of CNVs in the development of autism, the de novo germline mutation has been characterized in recent past as a risk factor for the disease.⁹¹

The effects of these CNVs are at times combined with the environmental factors as well, which may increase or decrease the severity of the disease, for example, in *GSTM1* and *GSTT1* genes in cancer susceptibility.⁹² Within recent years, multiple evidence have led to a conclusion that CNVs play a crucial role in disease susceptibility and development. Also, large-scale studies have been conducted to augment the indications. For example, recently, a study with a cohort of 144 children (patients only) with absence epilepsy subtypes (a type of epileptic disorder found mostly in children) was conducted. The study highlighted duplications and deletions in the gene regions 1p36.33, 1q21.1, 22q11.2, and Xp22.31 increasing the potency to develop the disease.⁹³

Techniques to Identify Copy Number Variations

Before array hybridization technologies, imbalanced chromosomal rearrangements were detected and analyzed by cytogenetic techniques. For large and imbalanced rearrangements, the gold standard Giemsa banding was popular which gives a visible karyotype of all chromosomes,⁹⁴ whereas for molecular cytogenetics, fluorescent in situ hybridization (FISH) and comparative genomic hybridization (CGH) techniques were used.⁹⁵ These techniques come up with a nonnegligible percentage of error which reduces the authenticity of the results. So, an improvised version of FISH was introduced as 3D FISH but it also lacked the speed and the specificity required for high-level comprehensive analysis.⁹⁶ However, because of low resolution and less specificity, these methods were overpowered by the advancements in the detection techniques.

There are 2 classes of chromosomal abnormalities: microscopic, which can be seen by a microscope, and submicroscopic, which are rare and cannot be detected by microscope alone.⁹⁷ For submicroscopic variants, hybridization techniques like CGH and their modified versions like microarray-based CGH (array-CGH) to identify very rare variants were introduced.⁹⁸ Earlier, the arrays were introduced with larger sequence segments of DNA for probing; however, now, increased number of spots and shortened segment size on the

Table 2. Copy number variations in different genes causing diseases, an increased or decreased copy of the same gene can cause different diseases.

S. NO.	GENES	ALTERATION TYPE	DISEASE	REFERENCES
1	CCL3L1	Decreased copies	HIV/AIDS susceptibility	Meddows-Taylor et al ⁷⁸
		Increased copies	Increased risk of rheumatoid arthritis	McKinney et al ⁷⁹
2	FCGR3B/CD 16b	Decreased copies	Increased risk for lupus nephritis	Aitman et al ⁸⁰
3	UGTB17	Deletion	Associated with 2-fold increased risk of osteoporosis	Yang et al ⁸¹
4	SNCA	Triplication	Causes Parkinson disease	Singleton et al ⁸²
5	DEFB4	More than 5 copies	Associated with 1.7-fold increased risk of psoriasis	Hollox et al ⁸³
		Less than 4 copies	Associated with 3-fold increased risk for Crohn disease	Fellermann et al ⁸⁴
6	LCE3B and LCE3C	Multigene deletion	Associated with psoriasis	de Cid et al ⁸⁵
7	AUTS4	Duplication of 15q11-13	Autism spectrum disorders	Sebat et al ⁸⁶
	AUTS14	Deletion of 16p11.2		
8	PMP-22	Duplications at chromosome region 17p11.2	Charcot-Marie-Tooth neuropathy type 1A	Hanemann and Müller ⁸⁷
9	Oncogene Myc	Amplification	Acute myeloid leukemia AML	Jones et al ⁸⁸
10	GSK3b	Duplication	Bipolar disorder	Lachman et al ⁸⁹

Abbreviations: AUTS4/14, autism spectrum disorders 4/14; CCL3L1, chemokine C-C motif ligand 3 like 1 gene; DEFB4, defensin B-4 gene; FCGR3B/CD 16b, Fc fragment of IgG, low-affinity IIIb, receptor gene; GSK3b, glycogen synthase kinase 3; LCE3B and LCE3C, late cornified envelope 3B and late cornified envelope 3C; PMP-22, peripheral myelin protein; synuclein, SNCA; UGTB17, uridinediphosphateglycosyl transferase-2 family B-17.

array to improve the resolution are used. Different oligonucleotide arrays, PCR products, clone arrays, genotyping arrays, and others are in use.⁹⁹ A number of array platforms have been recommended in recent past for the detection of complex CNVs with nonexpressive phenotypes. For example, bacterial artificial chromosome arrays and oligo CGH (oligo-CGH) associated with sequencing are improvised versions of CGH, whereas SNP arrays are used to differentiate between 2 SNP alleles. Array techniques have been preferred over conventional karyotyping for diagnosis of fetal and prenatal abnormalities.¹⁰⁰ Customized arrays are also available to choose platforms and target genes and regions according to the need of the research. Furthermore, genome-wide analysis of CNVs can be done by multiplex PCR-based methodologies such as PCR followed by restriction fragment length polymorphism, real-time quantitative PCR for shorter length fragments, multiplex ligation-dependent probe amplification, and multiplex amplicon quantification techniques.¹⁰¹ Such techniques are majorly used for screening and validation of CNVs. Also, whole genome sequencing, exome sequencing, and targeted exome capture followed by sequencing are widely

accepted sequencing techniques to determine copy number variations in human genome.^{102–104}

The advent of next-generation sequencing technologies has enabled accurate and rapid analysis of CNVs with high resolution per base.¹⁰⁵ Moreover, high-resolution array platforms such as NimbleGen (Madison, WI, USA) and Agilent (Santa Clara, CA, USA), pioneer to develop CGH arrays with differently labeled target and reference genome with fluorescent tags, are currently being used for CNVs.⁹¹ Next-generation sequencing platforms such as Illumina, Roche 454 Life Sciences, and SOLiD from Applied Biosystems have provided more accurate CNVs for both genome-wide analysis and targeted gene analysis.^{106–108} Also, a number of computational tools are available which in combination with the sequencing techniques improve the analysis.¹⁰⁹ For better understanding of next-generation techniques for identifying CNVs, comparative studies have been comprehended and more sensitive tools and accurate algorithms have been introduced for CNVs from the big data sets. Flow diagram in Figure 2 shows different CNVs and techniques for detection methods. The CNVs are classified on the basis of detection techniques employed.

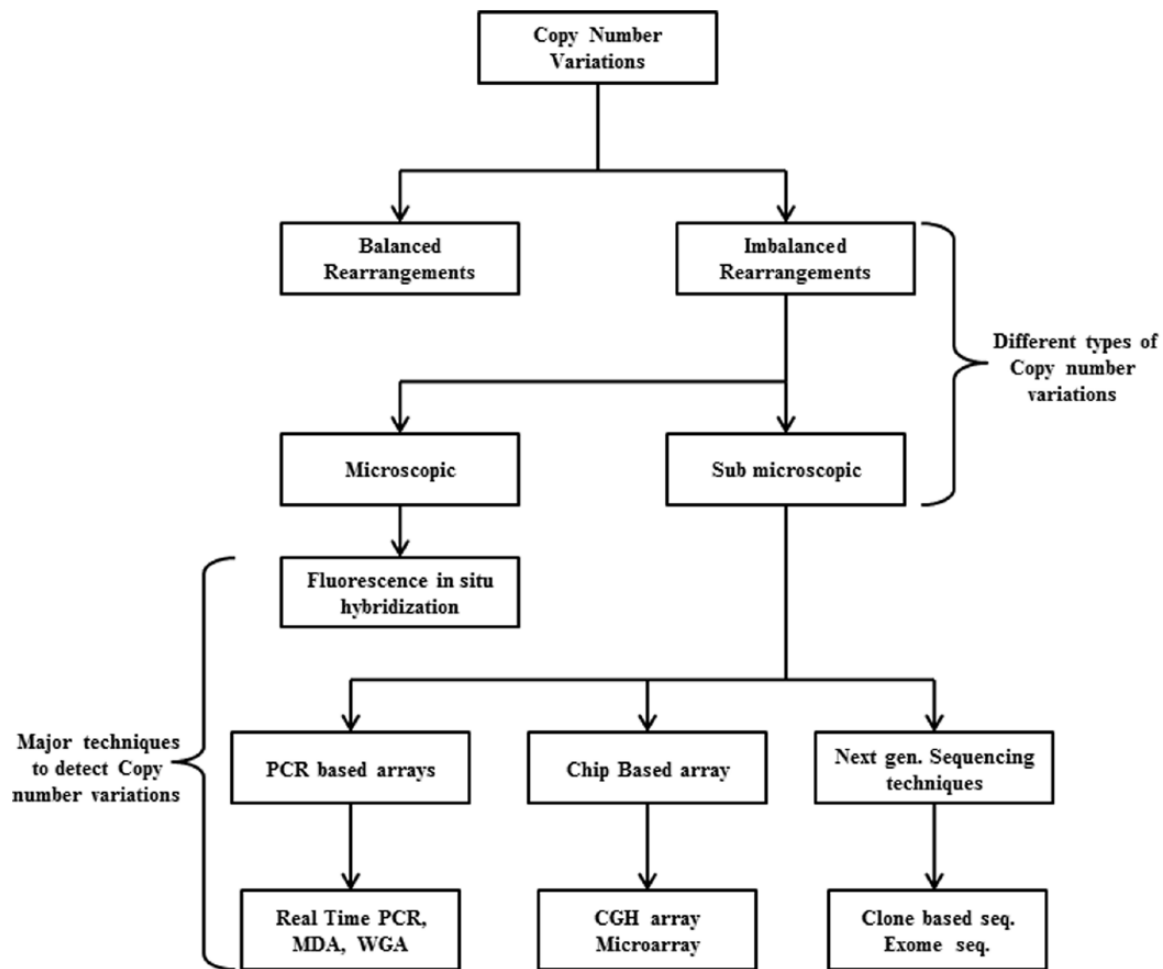


Figure 2. Classification of copy number variants based on the techniques employed for identification.

Future Perspectives

Although current platforms are able to detect clinically significant CNVs in whole genome with better specificity, there are technical challenges that remain to be directed. The paramount implementation of such techniques is early/prenatal detection of structural variants for diagnostic purposes. The low sample size of human subjects and limitations of high-throughput data analysis and interpretation tools are major bottleneck. Over the past years, CNVs have emerged as one of the possible reasons underlying multifaceted diseases, as the mechanisms of many complex diseases are yet to be elucidated. Although literature evidence are available for different CVDs, their association with CNVs suggests a not-alone role of single-nucleotide variations (SNPs) as genetic variants regulating human diseases and health. More investigation is required to culminate the magnitude of the effect in various CVD subtypes. The current techniques and developments have enabled us to elucidate genomic variations in humans; yet, the complexity itself requires a platform to delineate the neutral or negative effect of any variant on any individual. Current platforms to identify CNVs need improvisation as they need to optimize the threshold for CNV detection and minimize the noise created by other genetic variants. Although genomic maps for CNVs have

suggested their potential effects on the healthy individuals, hence opening up a new obscured area of research, epidemiologic studies with larger cohorts and comprehensive methodologies are prerequisite for a conclusive statement.

Acknowledgements

The authors are extremely thankful to Director, DIPAS, Defence Research and Development Organization (DRDO). A.V. is a DST-INSPIRE fellow.

Author Contributions

MZA and AV conceived and designed the article. AV wrote the manuscript. IG and MZA edited the manuscript.

REFERENCES

- Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nat Rev Genet.* 2015;16:172–183.
- Zhang B, Willing M, Grange DK, et al. Multigenerational autosomal dominant inheritance of 5p chromosomal deletions. *Am J Med Genet A.* 2016;170:583–593.
- Kumar RA, Mohamed SK, Sudi J, et al. Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet.* 2008;17:628–638.
- Dolcetti A, Silversides CK, Marshall CR, et al. 1q21.1 Microduplication expression in adults. *Genet Med.* 2013;15:282–289.
- Tang YC, Amon A. Gene copy-number alterations: a cost-benefit analysis. *Cell.* 2013;152:394–405.
- Antonarakis SE, Rossiter JP. Factor VIII gene inversions in severe hemophilia A: results of an international consortium study. *Blood.* 1995;86:2206–2212.

7. Thomson PA, Duff B. Balanced translocation linked to psychiatric disorder, glutamate, and cortical structure/function. *NPJ Schizophr.* 2016;10:16024.
8. Blazejewski SM, Bennison SA, Smith TH, Toyo-Oka K. Neurodevelopmental genetic diseases associated with microdeletions and microduplications of chromosome 17p13.3. *Front Genet.* 9:80.
9. Campbell RD, Carroll MC, Porter RR. The molecular genetics of components of complement. *Adv Immunol.* 1986;38:203–244.
10. Ashar FN, Zhang Y, Longchamps RJ, et al. Association of mitochondrial DNA copy number with cardiovascular disease. *JAMA Cardiol.* 2017;2:1247–1255.
11. Cho SB, Koh I, Nam HY, Jeon JP, Lee HK, Han BG. Mitochondrial DNA copy number augments performance of A1C and oral glucose tolerance testing in the prediction of type 2 diabetes. *Sci Rep.* 2017;7:43203.
12. Campa D, Barrdahl M, Santoro A, et al. Mitochondrial DNA copy number variation, leukocyte telomere length, and breast cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Breast Cancer Res.* 2018;20:29.
13. Guryev V, Saar K, Adamovic T, et al. Distribution and functional impact of DNA copy number variation in the rat. *Nat Genet.* 2008;40:538–545.
14. Adams DJ, Dermitzakis ET, Cox T, et al. Complex haplotypes, copy number polymorphisms and coding variation in two recently divergent mouse strains. *Nat Genet.* 2005;37:532–536.
15. Lee AS, Gutiérrez-Arcelus M, Perry GH, et al. Analysis of copy number variation in the rhesus macaque genome identifies candidate loci for evolutionary and human disease studies. *Hum Mol Genet.* 2008;17:1127–1136.
16. de Smith AJ, Walters RG, Froguel P, et al. Human genes involved in copy number variation: mechanisms of origin, functional effects and implications for disease. *Cytogenet Genome Res.* 2008;123:17–26.
17. Stranger BE, Forrest MS, Dunning M, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science.* 2007;315:848–853.
18. Anthony JF, Griffiths William M, et al. *Modern Genetic Analysis.* New York, NY: W. H. Freeman; 1999.
19. Berletch JB, Yang F, Xu J, et al. Genes that escape from X inactivation. *Hum Genet.* 2011;130:237–245.
20. Prabhakaran D, Jeemon P, Roy A, et al. Cardiovascular diseases in India current epidemiology and future directions. *Circulation.* 2016;133:1605–1620.
21. Bhatnagar P, Wickramasinghe K, Williams J, Rayner M, Townsend N. The epidemiology of cardiovascular disease in the UK 2014. *Heart.* 2015;101:1182–1189.
22. Elizabeth G, Nabel MD. Cardiovascular disease. *N Engl J Med.* 2003;349:60–72.
23. Kreatsoulas C, Anand SS. The impact of social determinants on cardiovascular disease. *Can J Cardiol.* 2010;26:8C–13C.
24. Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: global burden of disease study. *Lancet.* 1997;349:1436–1442.
25. Anderson KM, Odell PM, Wilson PWF, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J.* 1991;121:293–298.
26. Hastings PJ, Ira G, Lupski JR. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.* 2009;5:e1000327.
27. Muñoz-Amatriáin M, Eichten SR, Wicker T, et al. Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome. *Genome Biol.* 2013;14:R58.
28. Liu P, Carvalho CMB, Hastings PJ, Lupski JR. Mechanisms for recurrent and complex human genomic rearrangements. *Curr Opin Genet Dev.* 2012;22:211–220.
29. Iafrate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome. *Nat Genet.* 2004;36:949–951.
30. Nakamura Y, Leppert M, O'Connell P, et al. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science.* 1987;235:1616–1622.
31. Brookes KJ. The VNTR in complex disorders: the forgotten polymorphisms? a functional way forward? *Genomics.* 2013;101:273–281.
32. Litt M, Luty JA. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet.* 1989;44:397–401.
33. Donnelly A, Kozman H, Gedeon AK, et al. A linkage map of microsatellite markers on the human X chromosome. *Genomics.* 1994;20:363–370.
34. Gymrek M, Willems T. Abundant contribution of short tandem repeats to gene expression variation in humans. *Nature Genetics.* 2016;48:22–29.
35. Weterman MA, van Ruisven F, de Wissel M, et al. Copy number variation upstream of PMP22 in Charcot-Marie-Tooth disease. *Eur J Hum Genet.* 2010;18:421–428.
36. Ho MR, Tsai KW, Chen CH, et al. dbDNV: a resource of duplicated gene nucleotide variants in human genome. *Nucleic Acids Res.* 2011;39:D920–D925.
37. Hand ML, Ponting RC, Drayton MC, et al. Identification of homologous, homoeologous and paralogous sequence variants in an outbreeding allopolyploid species based on comparison with progenitor taxa. *Mol Genet Genomics.* 2008;280:293–304.
38. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature.* 2006;444:444–454.
39. Kannel WB, Vasari RS. Adverse consequences of the 50% misconception. *Am J Cardiol.* 2009;103:426–427.
40. Matsuoka R, Misugi K, Goto A, Gilbert EF, Ando M. Congenital heart anomalies in the trisomy 18 syndrome, with reference to congenital polyvalvular disease. *Am J Med Genet.* 1983;14:657–668.
41. Van Praagh S, Truman T, Firpo A, et al. Cardiac malformations in trisomy-18: a study of 41 postmortem cases. *J Am Coll Cardiol.* 1989;13:1586–1597.
42. Hutchinson R, Wilson M, Voullaire L. Distal 8p deletion (8p23.1–8pter): a common deletion? *J Med Genet.* 1992;29:407–411.
43. Lindsay EA, Goldberg R, Jurecic V, et al. Velo-cardio-facial syndrome: frequency and extent of 22q11 deletions. *Am J Med Genet.* 1995;57:514–522.
44. Pauli RM, Scheib-Wixted S, Cripe L, Izumo S, Sekhon GS. Ventricular non-compaction and distal chromosome 5q deletion. *Am J Med Genet.* 1999;85:419–423.
45. Landegren U, Nilsson M, Kwok PY. Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. *Genome Res.* 1998;8:769–776.
46. Bush WS, Moore JH. Chapter 11: genome-wide association studies. *PLoS Comput Biol.* 2012;8:e1002822.
47. Xu M, Zhao J, et al. Apolipoprotein E gene variants and risk of coronary heart disease: a meta-analysis. *Biomed Res Int.* 2016;2016:3912175.
48. Cenarro A, Artieda M, Castillo S, et al. A common variant in the ABCA1 gene is associated with a lower risk for premature coronary heart disease in familial hypercholesterolaemia. *J Med Genet.* 2003;40:163–168.
49. Franceschini N, Muallem H, Rose KM, Boerwinkle E, Maeda N. LDL receptor polymorphisms and the risk of coronary heart disease: the atherosclerosis risk in communities study. *J Thromb Haemost.* 2009;7:496–498.
50. Ionita-Laza I, Rogers AJ, Lange C, et al. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. *Genomics.* 2009;93:22–26.
51. Pollex RL, Hegele RA. Copy number variation in the human genome and its implications for cardiovascular disease. *Circulation.* 2007;115:3130–3138.
52. Norton N, Li D, Rieder MJ, et al. Genome-wide studies of copy number variation and exome sequencing identify rare variants in BAG3 as a cause of dilated cardiomyopathy. *Am J Hum Genet.* 2011;88:273–282.
53. Fakhro KA, Choi M, Ware SM, et al. Rare copy number variations in congenital heart disease patients identify unique genes in left-right patterning. *Proc Natl Acad Sci.* 2011;108:2915–2920.
54. Breckpot J, Thienpont B, Arens Y, et al. Challenges of interpreting copy number variation in syndromic and non-syndromic congenital heart defects. *Cytogenet Genome Res.* 2011;135:251–259.
55. Grimholt RM, Urdal P, Klingenberg O, et al. Rapid and reliable detection of α -globin copy number variations by quantitative real-time PCR. *BMC Hematology.* 2014;14:4.
56. Sebat J. Major changes in our DNA lead to major changes in our thinking. *Nat Genet.* 2007;39:S3–S5.
57. Van der Aa N, Rooms L, Vandeweyer G, et al. Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome. *Eur J Med Genet.* 2009;52:94–100.
58. Laitenberger G, Donner B, Gebauer J, et al. D-transposition of the great arteries in a case of microduplication 22q11.2. *Pediatr Cardiol.* 2008;29:1104.
59. Glessner JT, Bick AG, Ito K, et al. Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. *Circ Res.* 2014;115:884–896.
60. Kathiresan S, Voight BF, Purcell S, et al.; Myocardial Infarction Genetics Consortium. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet.* 2009;41:334–341.
61. Wu Z, Sheng H, Chen Y, et al. Copy number variation of the lipoprotein(a) (LPA) gene is associated with coronary artery disease in a southern Han Chinese population. *Int J Clin Exp Med.* 2014;7:3669–3677.
62. Matarin M, Simon-Sanchez J, Fung HC, et al. Structural genomic variation in ischemic stroke. *Neurogenetics.* 2008;9:101–108.
63. Tennant PW, Pearce MS, Bythell M, Rankin J. 20-year survival of children born with congenital anomalies: a population-based study. *Lancet.* 2010;375:649–656.
64. Kim DS, Kim JH, Burt AA, Crosslin DR, Burnham N. Burden of potentially pathologic copy number variants is higher in children with isolated congenital heart disease and significantly impairs covariate-adjusted transplant-free survival. *J Thorac Cardiovasc Surg.* 2016;51:1147.e4–1151.e4.
65. Zhu X, Li J, Ru T, et al. Identification of copy number variations associated with congenital heart disease by chromosomal microarray analysis and next-generation sequencing. *Prenat Diagn.* 2016;36:321–327.
66. Hanchard NA, Umana LA, D'Alessandro L, et al. Assessment of large copy number variants in patients with apparently isolated congenital left-sided cardiac lesions reveals clinically relevant genomic events. *Am J Med Genet A.* 2017;173:2176–2188.

67. Soemedi R, Wilson IJ, Bentham J, et al. Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease. *Am J Hum Genet.* 2012;91:489–501.
68. Zaidi S, Choi M, Wakimoto H, et al. De novo mutations in histone-modifying genes in congenital heart disease. *Nature.* 2013;498:220–223.
69. Moosmann J, Uebe S, Dittrich S, et al. Novel loci for non-syndromic coarctation of the aorta in sporadic and familial cases. *PLoS ONE.* 10:e0126873.
70. Sanchez-Castro M, Eldjouzi H, Charpentier E, Busson PF, Hauet Q, et al. Search for rare copy-number variants in congenital heart defects identifies novel candidate genes and a potential role for FOXC1 in patients with coarctation of the aorta. *Circ Cardiovasc Genet.* 2016;9:86–94.
71. Shia WC, Ku TH, Tsao M, et al. Genetic copy number variants in myocardial infarction patients with hyperlipidemia. *BMC Genomics.* 2011;12:S23.
72. Lopes LR, Murphy C, Syrris P, et al. Use of high-throughput targeted exome-sequencing to screen for copy number variation in hypertrophic cardiomyopathy. *Eur J Med Genet.* 2015;58:611–616.
73. Mademont-Soler I, Mates J, Yotti R, et al. Additional value of screening for minor genes and copy number variants in hypertrophic cardiomyopathy. *PLoS ONE.* 2017;12:e0181465.
74. Marques FZ, Prestes PR, Pinheiro LB, et al. Measurement of absolute copy number variation reveals association with essential hypertension. *BMC Med Genomics.* 2014;7:44.
75. Jarinova O, Stewart AFR, et al. Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler Thromb Vasc Biol.* 2009;29:1671–1677.
76. Colaianni V, Mazzei R, Cavallaro S. Copy number variations and stroke. *Neurol Sci.* 2016;37:1895–1904.
77. Nik-Zainal S, Cotter PE, Willatt LR, Abbott K, O'Brien EW. Ring chromosome 12 with inverted microduplication of 12p13.3 involving the Von Willebrand Factor gene associated with cryptogenic stroke in a young adult male. *Eur J Med Genet.* 2011;54:97–101.
78. Meddows-Taylor S, Donninger SL, Paximadis M, et al. Reduced ability of newborns to produce CCL3 is associated with increased susceptibility to perinatal human immunodeficiency virus 1 transmission. *J Gen Virol.* 2006;87:2055–2065.
79. McKinney C, Merriman ME, Chapman PT, et al. Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. *Ann Rheum Dis.* 2008;67:409–413.
80. Aitman TJ, Dong R, Vyse TJ, et al. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature.* 2006;439:851–855.
81. Yang TL, Chen XD, Guo Y, et al. Genome-wide copy-number-variation study identified a susceptibility gene, UGT2B17, for osteoporosis. *Am J Hum Genet.* 2008;83:663–674.
82. Singleton AB, Farrer M, Johnson J, et al. alpha-Synuclein locus triplication causes Parkinson's disease. *Science.* 2003;302:841.
83. Hollox EJ, Huffmeier U, Zeeuwen PL, et al. Psoriasis is associated with increased beta-defensin genomic copy number. *Nat Genet.* 2008;40:23–25.
84. Fellermann K, Stange DE, Schaeffeler E, et al. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am J Hum Genet.* 2006;79:439–448.
85. de Cid R, Riveira-Munoz E, Zeeuwen PLJM, et al. Deletion of the late cornified envelope (LCE) 3B and 3C genes as a susceptibility factor for psoriasis. *Nat Genet.* 2009;41:211–215.
86. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of de novo copy number mutations with autism. *Science.* 2007;316:445–449.
87. Hanemann CO, Müller HW. Pathogenesis of Charcot-Marie-Tooth 1A (CMT1A) neuropathy. *Trends Neurosci.* 1998;21:282–286.
88. Jones L, Wei G, Sevickova S, et al. Gain of MYC underlies recurrent trisomy of the MYC chromosome in acute promyelocytic leukemia. *J Exp Med.* 2010;207:2581–2594.
89. Lachman HM, Pedrosa E, Petruolo OA, et al. Increase in GSK3beta gene copy number variation in bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet.* 2007;144B:259–265.
90. Gonzalez E, Kulkarni H, Bolivar H, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science.* 2005;307:1434–1440.
91. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of de novo copy number mutations with autism. *Science.* 2007;316:445–449.
92. Rebbeck TR. Molecular epidemiology of the human glutathione S-transferase genotypes GSTM1 and GSTT1 in cancer susceptibility. *Cancer Epidemiol Biomarkers Prev.* 1997;6:733–743.
93. Addis L, Rosch RE, Valentin A, et al. Analysis of rare copy number variation in absence epilepsies. *Neurol Genet.* 2016;2:e56.
94. Speicher MR, Carter NP. The new cytogenetics: blurring the boundaries with molecular biology. *Nat Rev Genet.* 2005;6:782–792.
95. Uchida IA, Lin CC. Fluorescent staining of human chromosomes: identification of some common aberrations. *Can Med Assoc J.* 1971;105:479–482.
96. Müller S, Neusser M, Köhler D, Cremer M. Preparation of complex DNA probe sets for 3D FISH with up to six different fluorochromes [published ahead of print May 1, 2007]. *CSH Protoc.* doi:10.1101/pdb.prot4730.
97. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet.* 2006;7:85–97.
98. Thureson AC, Bondeson ML, Edeby C, et al. Whole-genome array-CGH for detection of submicroscopic chromosomal imbalances in children with mental retardation. *Cytogenet Genome Res.* 2007;118:1–7.
99. Carter NP. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet.* 2007;39:S16–S21.
100. Karampetsou E, Morrogh D, Chitty L. Microarray technology for the diagnosis of fetal chromosomal aberrations: which platform should we use? *J Clin Med.* 2014;3:663–678.
101. Li W, Olivier M. Current analysis platforms and methods for detecting copy number variation. *Physiol Genomics.* 2013;45:1–16.
102. Lee JH, Jeon JT. Methods to detect and analyze copy number variations at the genome-wide and locus-specific levels. *Cytogenet Genome Res.* 2008;123:333–342.
103. Pirooznia M, Goes FS, Zandi PP. Whole-genome CNV analysis: advances in computational approaches. *Front Genet.* 2015;6:138.
104. Samarakoon PS, Sorte HS, Kristiansen BE, et al. Identification of copy number variants from exome sequence data. *BMC Genomics.* 2014;15:661.
105. Wang H, Nettleton D, Ying K. Copy number variation detection using next generation sequencing read counts. *BMC Bioinformatics.* 2014;15:109.
106. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature.* 2008;456:53–59.
107. Margulies M, Egholm M, William E, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature.* 2005;437:376–380.
108. McKernan KJ, Peckham HE, Costa GL, et al. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res.* 2009;19:1527–1541.
109. Zhao M, Wang Q, Wang Q, et al. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC Bioinformatics.* 2013;14:S1.