

REVIEW

Implications of gene copy-number variation in health and diseases

Suhani H Almal and Harish Padh

Inter-individual genomic variations have recently become evident with advances in sequencing techniques and genome-wide array comparative genomic hybridization. Among such variations single nucleotide polymorphisms (SNPs) are widely studied and better defined because of availability of large-scale detection platforms. However, insertion–deletions, inversions, copy-number variations (CNVs) also populate our genomes. The large structural variations (> 3 Mb) have been known for past 20 years, however, their link to health and disease remain ill-defined. CNVs are defined as the segment of DNA > 1 kb in size, and compared with reference genome vary in its copy number. All these types of genomic variations are bound to have vital role in disease susceptibility and drug response. In this review, the discussion is confined to CNVs and their link to health, diseases and drug response. There are several CNVs reported till date, which have important roles in an individual's susceptibility to several complex and common disorders. This review compiles some of these CNVs and analyzes their involvement in diseases in different populations, analyses available evidence and rationalizes their involvement in the development of disease phenotype. Combined with SNP, additional genomic variations including CNV, will provide better correlations between individual genomic variations and health.

Journal of Human Genetics (2012) 57, 6–13; doi:10.1038/jhg.2011.108; published online 29 September 2011

Keywords: CNV; diseases; gene copy number; genetic variants; health; neurodisorders; pharmacogenetics; SNP

INTRODUCTION

Variations in one's genomic DNA make one unique in terms of disease susceptibility and response to drugs. Single nucleotide polymorphisms (SNPs) are the most widely studied form of genetic variations and few of the SNPs have been linked to susceptibility to diseases and serves as a marker for certain disorders. Beside SNPs, submicroscopic copy-number variations (CNVs) are now considered important form of genetic variations. Findings in past few years have indicated a strong association of CNVs with several complex and common disorders that have a profound effect on our health.

CNVs are a segment of DNA that is 1 kb or larger and present at a variable copy number in comparison with a reference genome.^{1,2} CNVs in general are stable and can be inherited. At times they can also develop spontaneously during meiosis. The exact mechanism of the development of new CNV is not clearly understood.³ Deletions, duplications, segmental duplications, insertions, inversions and translocations represent some of the processes resulting in CNV (Figure 1). Variations in the gene copy number can be detected using a variety of platforms, which have evolved rapidly in recent past.^{4,5}

Any of the processes mentioned earlier may result into disruption of genes in the region and therefore may result in the development of diseases. The phenotypic effect on the disease process will, however, depend on how CNVs affect dosage-sensitive genes or their regulatory

elements. It is also conceivable that the development of the phenotypic disease may not depend upon a single CNV but a combination of various CNVs and other genetic variations such as SNPs (Figure 2). Therefore, identification of the all the susceptibility factors affecting the diseases development is a better predictive diagnosis rather than correlating the diseases with fewer susceptibility factors.

The challenge was to identify meaningful differences in the genomic information between individuals. When the database for SNP variations in the human genome became available, it was soon realized that the available technologies were inadequate to detect other forms of variations such as CNVs, insertion–deletions (INDELs) and inversions. It was also realized that the other variations such as CNVs, INDELs and inversions are present in the human genome with frequencies much higher than expected. At the same time these variants were too small to be detected by the microscopic techniques and therefore a new set of convenient and cheaper technology platforms evolved enabling us to map the human genome for CNVs.⁵ In 2006, Redon *et al.*⁶ constructed a first-generation CNV map of the human genome with the use of SNP genotyping arrays and Whole-Genome TilePath BAC arrays and among the 270 individuals studied, 12% of the human genome was found to be covered by CNVs whereas the total number of CNVs included in the Database of Genomic Variants (DGV) was reported to account for 29.7% of the human

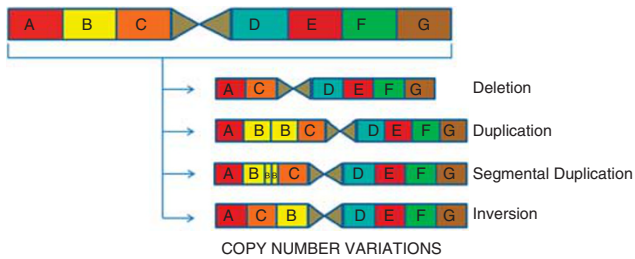


Figure 1 Types of genomic variants. Genomic variants in form of CNVs can be classified primarily as deletion, duplication, segmental duplication and inversion. These variations can encompass the entire gene or a segment of a particular gene represented in the figure.

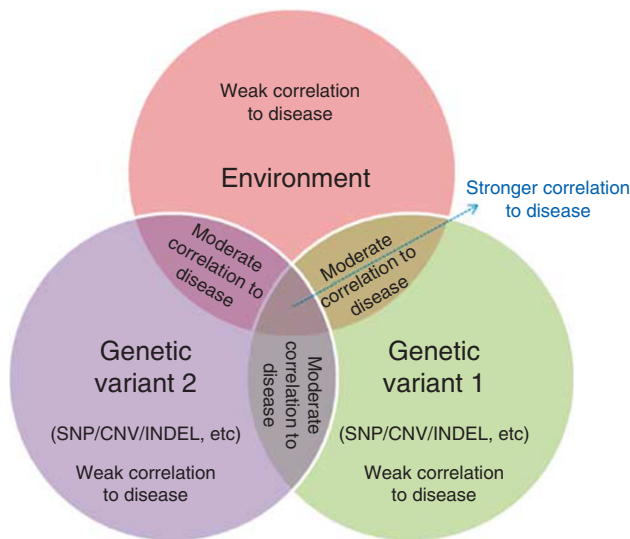


Figure 2 Genetic variations and susceptibility to polygenic diseases. A schematic representation involving several genetic variants (SNP, CNV, INDELS, and so on) and possible involvement of environmental factors in defining susceptibility to multifactorial diseases. A stronger correlation to disease susceptibility can be explained by taking in account the various factors affecting susceptibility to diseases such as environment factor, different type of genomic variants (SNP, CNVs, indels, and so on) as depicted in the figure.

genome, which is often over estimated. Some examples of CNVs include single or multiple genes, which contribute to clinical phenotype whereas smaller CNVs affecting single exons may also account for a proportion of human diseases.⁷ Recently, the 1000 genome project came up with a map of human genome variation based on the population-scale genome sequencing. In this approach, they screened individuals under three categories namely: low-coverage whole-genome sequencing of 179 individuals from four populations; high-coverage sequencing of two mother–father–child trios and exon-targeted sequencing of 697 individuals from seven populations at the pilot scale. They analyzed ~15 million SNPs, 1 million short INDELS and 20 000 structural variants describing the location, allele frequency and local haplotype structure of the variants. They also found that on an average each person carried 250–300 loss of function variants in annotated genes and 50–100 variants previously implicated in inherited disorders.⁸

There are several database where the CNVs are cataloged (DGV; <http://projects.tcag.ca/variation/>), public database such as Toronto

DGV and European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations (ECARUCA, <http://www.ecaruca.net>). Another database, which archives the clinically relevant CNVs is DECIPHER (DatabasE of Chromosomal Imbalance and Phenotype in Human using Ensembl Resources; http://www.sanger.ac.uk/Post_Genomics/decipher/).⁹ Till (May 2009), over 2200 cases of > 50 diseases have been included in DECIPHER.⁷ The structural variation in the human genome and its implication in human health has been recently reviewed by Stankiewicz and Lupski.¹⁰

Furthermore, it is likely that these numbers are underestimated and advances in the technologies will help us in discovering more CNVs in the human genome.⁴ The major limitation in studying CNVs, despite of the recent technical advances, is the size or the breakpoint position, the total number and its gene content. It's conceivable that with the availability of various accessible technologies we might have a high-resolution CNV map of the human genome in the near future. However, the implication of CNV on health will have to wait several large-scale correlation studies not only with one CNV but also with permutations and combinations of various likely variations.

TECHNIQUES TO DETECT CNVS

With the discovery of CNVs as the new form of genetic variations contributing to the disease susceptibility and progression, the task is how to detect these structural variations. The method to detect CNV should be convenient and inexpensive so that it can be applied to a large number of samples from a given population. Only when the data are available from various different populations the real impact of CNVs on health can be assessed. The major hurdle in detection is that a larger fraction of these CNVs do not have defined breakpoint. Demarcated breakpoints usually permit development of simple detection methods around the breakpoints. Beside conventional PCR several other modified PCR-based techniques have evolved, which are considered to be robust assays for screening the targeted region of the genome. Among these quantitative methods are multiplex ligation-dependent probe amplification, multiplex amplification and probe hybridization, quantitative multiplex PCR of short fluorescent fragments, semiquantitative fluorescence *in situ* hybridization, dynamic allele-specific hybridization and paralogue ratio test. The methodologies to detect CNV have been recently reviewed by Dhawan and Padh (2009).⁴

Genome-wide association studies have been widely used techniques to identify gene(s) and its variants associated in disease process. Genome-wide association studies are generally carried out in cases versus control populations.¹¹ Evaluating the technicality of each method, and its advantages and drawbacks, weighed against the objectives of the study can help us select the appropriate techniques.

CNV AND PATHOPHYSIOLOGY

Recently, many CNVs have been reported to affect disease susceptibility. Among them are various complex neurological disorders, cardiovascular diseases, infectious and autoimmune diseases, metabolic diseases, cancer and several other common disorders (Table 1).

CNV AND NEUROLOGICAL DISORDERS

Parkinson's disease

Parkinson's disease (PD) is a progressive nervous disease associated with the symptoms such as muscular rigidity, resting tremors, bradykinesia (slowing of movements) and posture instability. It is found to affect 1% of the population over 50 years of age.¹² A novel triplication in *SNCA* gene, located on 4q21 is found to be linked to autosomal

Table 1 CNV genes associated with diseases studied in different population

<i>Disease</i>	<i>Gene(s) associated</i>	<i>Locus</i>	<i>CNV</i>	<i>Population</i>	<i>References</i>
<i>Neurological disorders</i>					
Parkinson's	<i>SNCA</i>	4q21	Duplication/triplication	Swedish family	13
Alzheimer	<i>APP</i>	21q21	Copy number gain	Dutch population	14–16
Mental retardation	Multiple genes	15q13.3, 17q21.31, Xp11.22	Deletion and duplication	General population	19–26
Autism	Multiple genes	3q24, 15q11-13, 16p11.2	Inherited homozygous deletion/duplication	—	27,29–31
Schizophrenia	<i>MYT1L, CTNND2, ASTN2</i>	1q21.1, 15q11.2, 15q13.3, 22q11.2	Deletion	—	34–37
Bipolar disorder	<i>GSK3β</i>	3q13.3	Duplication	—	38
<i>Susceptibility to other common disorders</i>					
HIV/AIDS susceptibility	<i>CCL3L1</i>	17q21.1	—	Caucasian, New Zealand and UK population	39
Crohn's disease	<i>HBD-2, LCE3B, LCE3C, DEFB4, IRGM</i>	Multiple	Deletion and duplication	European ancestry	41–44
Pancreatitis	<i>PRSS1</i>	7q35	Triplication	French families	45
SLE	<i>FCGR3B</i>	1q23	Copy number gain and loss	UK cohorts	46–47
Asthma	<i>GSTM1</i> and <i>GSTT1</i>	1p13.3 and 22q11.23	Deletion	—	50
<i>Cardiovascular diseases</i>					
Atherosclerosis	<i>LPA</i>	6q26	Copy number gain	—	52,53
Familial hypercholesterolemia	<i>LDLR</i>	19p13.3	Deletion/duplication	—	54
<i>Metabolic diseases</i>					
Type 2 diabetes mellitus	<i>LEPR</i>	1p31	Copy number gain	Korean population	55
Overweight and obesity	<i>SH2B1, PPYR1</i>	16p11.2, 10q11.2-q21.2	Deletion and duplication	Caucasian and Chinese population	57,58
Intestinal disease	<i>AMY1</i>	1p21	Copy number gain	—	59
<i>Drug metabolism</i>					
Nicotine metabolism and tobacco-related diseases	<i>CYP2A6</i>	19q13.2	Deletion/duplication	Chinese, European-American, Korean and Japanese	60,61
Sulfate conjugation	<i>SULT1A1</i>	16p12.1	Deletion and duplication	Caucasian-American and African-American	62
Metabolism of carcinogens	<i>GSTM1, GSTP1</i> and <i>GSTT1</i>	1p13.3, 11q13 and 22q11.23	Deletion	—	63–65
<i>Cancer</i>					
Familial breast cancer	<i>MTUS1</i>	8p22	Exon-specific deletion	German population	68

Abbreviations: CNV, copy-number variation; SLE systemic lupus erythematosus.

dominant PD and showed the dosage effect. Further, the gene expression profiling reflected an approximately twofold increase of *SNCA* protein in blood, mRNA in the brain tissue and also the deposition of large aggregates of this protein in the brain tissue.¹³

CNV in terms of gain of the copies of *SNCA* gene showed a profound effect on PD. The *SNCA* triplication was reported in a family of Swedish American descent with autosomal dominant early-onset PD.¹³ *SNCA* duplication was also reported in a Swedish family, suggesting a dosage effect of *SNCA* in selected cases of PD.¹³

Alzheimer's disease

Alzheimer's disease (AD) is a progressive neurological disorder characterized by dementia among elderly people, which is mainly due to the intracellular tangles and extracellular plaques of amyloid getting deposited in the vulnerable regions of the brain.

Copy number gain of the *APP* gene (the amyloid precursor protein) is hypothesized to be one of the causes of AD.¹⁵ In the Dutch population, duplication of the *APP* gene has been reported and was found to be associated with autosomal dominant early-onset AD and cerebral amyloid angiopathy.^{14–16}

It is interesting to note the association of the Alzheimer's with Down's syndrome that is due to trisomy 21. *APP* gene maps on the same chromosome so those with Down's syndrome have three copies of the gene and are more prone to Alzheimer's. Genetic variations in the form of point mutations, in at least 15 genomic loci¹⁷ and genetic variations in the promoter region of *APP* gene are found to be associated with AD.¹⁸

Mental retardation and developmental disorders

Mental retardation (MR) that is a nonprogressive cognitive impairment is also affected by the CNV seen in the genomic loci of several genes. Studies of large cohorts may help in detecting and confirming the roles of rare *de novo* CNVs in MR. MR occurs in 2–3% of newborns in the general population, however, its cause has remained elusive.¹⁹ X-linked MR showed six overlapping duplications at the Xp11.22 in six unrelated males. Further, it was noted that this duplication covered a 320-kb region involving four genes (*SMC1A*, *RIBC1*, *HSD17B10* and *HUWE1*), three candidates of which may convey the phenotype of MR.¹⁹ Apart from duplication, many other forms of genetic variations such as point mutation in *SMC1A* and *HUWE1* genes and a silent mutation in *HSD17B10* gene conveyed the phenotypes of MR along with other distinguished characteristic.²⁰ The conclusion drawn from the above findings showed that it is a dosage-sensitive gene, which confers the MR phenotype in the patients with duplicated genes. *MECP2*, X-linked methyl-CpG-binding protein 2 gene at Xq28 is also found to be associated with the developmental delay, MR and fatal infantile encephalopathy in males, and recently it has been reported that low copy number of *MECP2* gene confers a clinical phenotype, resulting in MR or developmental delay and altered neurological symptoms (particularly seizures) phenotypes in males.^{21–24} Needless to say, this is perhaps one of the most complex phenotype and the clear picture will emerge when all possible loci and their interactions are well defined.

Apart from this several other submicroscopic duplications and deletions in 17q13.3 involving *LIS1* and/or the *14-3-3ε* genes are shown to confer a risk to MR and several other characteristics features.²⁵

The deletion in exonic regions of *NRXN1* gene, located on chromosome 2p16.3, is found to predispose one to a wide spectrum of developmental disorders. These neurexins are a group of highly polymorphic cell surface proteins involved in synapse formation and signaling. Variants comprising a variety of mutations such as missense mutation, translocation, whole-gene deletion and intra-genic copy number changes result into a significant association with a variety of phenotypic changes such as autism, schizophrenia (SZ) and nicotine dependence.²⁶

Autism

Autism is a pervasive neurodegenerative disorder characterized by impaired communication or linguistic skills, social interaction, cognition, some form of repetitive and restricted stereotyped interest, ritual or other behavior. The symptoms vary from person-to-person and no two persons have identical symptoms and hence called Autism spectrum disorder.²⁷ A case–control study using the representational oligonucleotide microarray analysis technology was applied to identify the CNVs involved in Autism. Representational oligonucleotide microarray analysis was first explored for the detection of the genomic aberrations in cancer and healthy humans. In this technique by arraying oligonucleotide probes designed from the human genome sequence, and hybridizing with representations from cancer helped in detecting regions with altered copy number.²⁸ Subsequently with the

help of this technique revealed more spontaneous development of CNVs in Autism spectrum disorder patients than in unaffected controls.²⁹

Both duplication and deletion were observed for Autism spectrum disorder. There are several loci associated with autism susceptibility such as duplication of 15q11-13 (*AUTS4*; MIM 608636) and deletion of 16p11.2 (*AUTS14*; MIM611913). A reciprocal microduplication and recurrent microdeletion at 16p11.2 have been shown to be associated with autism and may account for 1% of the cases.³⁰ Applying the homozygosity mapping analysis in pedigrees several large inherited, homozygous deletions were observed. On analysis, it was found that this deletion spans 886 kb on chromosome 3q24 affecting *DIA1* gene, whose expression level changes in relation to neuronal activity.³¹

Schizophrenia

Schizophrenia (SZ) is a chronic, debilitating illness with extensive neurological and psychiatric features. Its prevalence is ~1% of the population. Several CNVs are found to be associated with SZ.^{32,33}

Various deletions located on positions such as 1q21.1, 15q11.2 and 15q13.3 were found to be associated with SZ and psychosis, when studied in case–control samples analyzed by the International Schizophrenia Consortium and subsequently confirmed by other studies.^{34,35} Furthermore, the previously reported deletion at 22q11.2 associated with SZ phenotype in DGV/velocardiofacial syndrome was also confirmed by these groups.³⁶ Another group of researchers identified 90 CNVs in 54 patients, of which 13 were rare CNVs disrupting genes associated with SZ such as *MYTIL*, *CTNND2* and *ASTN2*.³⁷ To confirm the association of these rare CNVs a large cohort of samples is needed.

Bipolar disorder

Bipolar disorder is a psychiatric disorder characterized with profound and prolong mood swings and depression. It has been found that submicroscopic variation in *GSK3β* gene, which codes for glycogen synthetase kinase, a key component of *Wnt* signaling pathway and a target of lithium salt is involved in the susceptibility to bipolar disorder. The duplication or the increase in the copy number of *GSK3β* gene disrupts the 3'-coding element as well as affects the neighboring genes. Findings from the study suggested that there was a significant increase in the *GSK3β* copy number in the bipolar disorder patients as compared with control.³⁸

CNV AND SUSCEPTIBILITY TO OTHER COMMON DISORDERS

HIV/AIDS susceptibility

Chemokines are secreted proteins involved in the immunoregulatory and inflammatory processes. The *CCL3L1* gene copy number influences the susceptibility to HIV/AIDS. *CCL3L1* gene, which encodes on 17q12, is the major coreceptor for CCR5 and is a dominant suppressive chemokine. Therefore, an increase in the copy number of this gene leads to reduction in susceptibility to HIV, as reported for the Caucasian population. The copy number of the *CCL3L1* gene varies from 0 to 10 copies in the Caucasian population.³⁹

The recruitment of lymphocytes by β -chemokines is a feature of autoimmunity conditions such as rheumatoid arthritis. The finding of the association of CCR5 Δ 32 variant with protection against rheumatoid arthritis led to hypothesis that gene copy number of *CCL3L1* gene influences susceptibility to rheumatoid arthritis and type 1 diabetes. When studied in two independent Caucasian cohorts (New Zealand and UK population) it was found that high copy number (higher than two copies) of *CCL3L1* gene was a risk factor for rheumatoid arthritis.⁴⁰

Crohn disease and psoriasis

Crohn disease (CD) is a chronic inflammatory bowel disease, causing inflammation of the digestive tract. It has been shown that deficient expression of defensins, which are endogenous antimicrobial peptides protecting intestinal mucosa against bacterial invasion, can lead to chronic CD. Therefore, it was hypothesized that the low copy number of the β -defensin gene cluster may also be associated with chronic CD. Various other reported deletions and SNPs in genes have shown a strong correlation with CD. *HBD-2* (human *beta-defensin 2*) gene is found to be associated with CD. When studied, it was found that patients with ulcerative colitis and healthy individuals have a median of 4 copies per diploid genome (range 2–10 copies), whereas patients with CD had lower copy number as compared with controls ($P=0.002$). Further, it was found that individuals with less than three copies of *HBD-2* gene have a significantly higher risk of developing colonic CD as compared with individuals having four or more copies (odds ratio of 3.06).⁴¹

In contrast, the increased copy number of β -defensin genes was shown to be associated with psoriasis, which is a chronic autoimmune skin disease with a prevalence of 2–3% in individuals of the European ancestry.^{42,43} Apart from β -defensin genes, individuals with deletion in *LCE3B* and *LCE3C* genes of late cornified envelope (LCE) gene cluster are found to be susceptible to psoriasis. The absence of the well characterized 32 199 bp region was significantly associated ($P=1.38E^{-08}$) with risk of psoriasis when studied in family-based samples from Spain, The Netherlands, Italy and the United States ($P=5.4E^{-04}$).⁴³

Immunity-related GTPase family, M was found to be associated with CD through earlier SNP studies. Recently, a 20-kb deletion has been found to attribute susceptibility to CD. The deleted portion is located upstream of the gene and is found to be in perfect linkage disequilibrium. It has been hypothesized that as immunity-related GTPase family M expression can affect the autophagy of internalized bacteria, the deletion might alter the expression level of immunity-related GTPase family M, thus, contributing to phenotype associated with CD.⁴⁴

Pancreatitis

Pancreatitis is a multigene-associated disorder, including the cationic trypsinogen gene *PRSS1*. In the earlier studies, the R122H missense mutation was found to increase the activity of trypsin *in vitro*, which led to the suggestion that *PRSS1* might be a dosage-sensitive gene. Upon further analysis, a novel 605 kb triplication was observed in a cohort of 34 French families with hereditary pancreatitis encompassing the *PRSS1* gene resulting into CNV of *PRSS1* gene.⁴⁵

Systemic lupus erythematosus and glomerulonephritis

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease of connective tissues and affects the skin, joints, kidney and serosal membranes, due to failure in regulation of the immune system. A strong correlation was found between the CNV of *FCGR3B* gene and SLE where an increased risk of development of SLE in individuals with fewer than the two copies of *FCGR3B* gene reported in the UK cohorts and the same correlation was confirmed in another Caucasian population.⁴⁶ *FcγR3B* is a glycosylphosphatidylinositol-linked, low-affinity receptor for immunoglobulin G found predominantly on human neutrophils. The low copy number of *FCGR3B* gene is associated with impaired clearance of the immune complex, which is a characteristic feature of SLE.⁴⁷ Complement component 4 (*C4*, including *C4A* and *C4B*) gene mutations were also found to be associated with SLE. On examining the Americans of European

descent, the copy number of *C4* gene varied from 2 to 6 (*C4A*, 0–5; *C4B* 0–4). The risk of SLE was found to be increased in subjects with low *C4* copies but decreased in those with high *C4* copies.⁴⁸

Asthma

Asthma is a chronic inflammatory genetic disorder characterized by a condition marked by recurrent attacks of dyspnea and constriction of the bronchii, also termed as bronchial asthma whereas the bronchial asthma due to allergy is called atopic asthma.

Genetic polymorphism of the glutathione S-transferase (*GST*) gene, which is mainly involved in the antioxidant defense is well known as a risk factor for several environmental diseases and was thus hypothesized that CNV in the *GST* genes might be associated with the asthma susceptibility.^{49,50} CNVs of the *GST* genes were examined in patients with atopic asthma and it was found that the null genotypes of *GSTT1* and *GSTM1* together with *GSTP1* Val/Val polymorphism have a significant role in the asthma pathogenesis.⁵¹

CNV AND CARDIOVASCULAR DISEASE

Structural variations are also found to be affecting the susceptibility to many cardiovascular diseases. A common CNV in *LPA* gene on chromosome 6, which encodes for atherogenic apolipoprotein (a), which is the primary determinant of the plasma lipoprotein is a risk factor for atherosclerosis.^{52,53} Apart from this CNVs were also found to be associated with the lipoprotein disorders. It was found that the low-density lipoprotein receptor gene (*LDLR*) is found to be affected in the patients with familial hypercholesterolemia.^{53,54}

CNV AND METABOLIC DISEASES

Type 2 diabetes

Type 2 diabetes is a chronic metabolic disorder characterized by high glucose level or insulin resistance. Various SNPs have been linked to the high risk for diabetes but other genetic variations such as CNV were not explored. Recently, it has been discovered that CNV at leptin receptor, which is mainly involved in satiety and energy expenditure is found to be significantly associated with risk to type 2 diabetes. The role of leptin and leptin receptor in obesity is well established and common genetic variation such as SNPs at the *LEPR* gene locus are found to be associated with obesity, hyperinsulinemia, type 2 diabetes mellitus and variation in the levels of leptin in different populations. The expansion of the map of genetic variation has revealed many new loci associated with the disease. In a recent study, the association of the *LEPR* gene locus encompassing ~200 kb on chromosome 1 was studied in the Korean population using the genome-wide SNP array data and it was found that CNV at the *LEPR* gene locus is significantly associated with metabolic traits and the risk to type 2 diabetes mellitus.⁵⁵

Overweight and obesity

People with body mass index of 25 kg m⁻² are considered overweight and those with body mass index of 30 kg m⁻² are considered obese. As per one estimate 66% of the US population is overweight and among them about 30% are obese.⁵⁶ Obesity represents complex metabolic disorder affecting multiple systems and the cause of obesity is complex and unclear. It still remains a multigenic and multifactorial condition influenced by “environment”. It is conceivable that beside environmental factors, several types of genetic variants such as SNP, CNV and others might be involved in precipitating obesity. In one instance, an early onset severe obesity in the Caucasian population was linked to deletion of 16p11.2 segment resulting in severe hyperphagia and insulin resistance. Examination of the deleted region revealed genes such as *SH2B1*, known to be involved in signaling pathways involving

leptin and insulin. In another case, CNV at 10q11.22 was found associated with body mass index. The region in question has *PPYRI* gene, regulating energy balance.^{57,58}

Amylase gene

CNVs is found not only contributing to the disease in general but also is found to be involved with diet, such as salivary amylase gene (*AMYI* gene), which is associated with the starchy food consumption in a population. It has been reported that higher copy number of *AMYI* gene is correlated positively with salivary amylase protein levels and a population with high-starch food consumption has higher copy number of *AMYI* gene and improves the digestion of starchy foods and may reduce the burden of intestinal diseases.⁵⁹

CNV AND DRUG METABOLISM

Cytochrome P450 (CYP450) is a superfamily of hemoproteins involved in metabolism of xenobiotics such as clinically used drugs, procarcinogens and environmental pollutants.

Among superfamily of CYPs, *CYP2A6* is an important human hepatic P450 enzyme, which is involved in drug metabolism including nicotine. CNVs such as duplication and deletion are found to be associated with the smoking behavior and susceptibility toward lung cancer and tobacco-related diseases. The association of the deletion variant was studied in the Chinese population. Frequency of the deletion variant (*CYP2A6del*) when studied in 96 Chinese subjects, was found to be 15.1%, but only 1% in Finns ($n=100$) and 0.5% in Spaniards ($n=100$).⁶⁰

Apart from deletion, a novel duplication was reported in the African-American population and was found to increase nicotine metabolism and may affect smoking behavior in contrast to the European-American, Korean or Japanese populations.⁶¹

Another gene studied in the Caucasian-American and in the African-American population is *SULT1A1* gene, which catalyzes the sulfate conjugation of a wide variety of drugs and is found to be genetically polymorphic. Apart from established SNPs at the 5'-flanking region, it also shows CNV. The range of copies of *SULT1A1* gene varies from one to approximately five copies in both the populations. When the enzyme activity in the human liver and platelet samples was checked it showed a positive correlation with the number of copies of *SULT1A1* gene.⁶²

Further, CNV mainly in the form of deletion has been reported in GSTs conferring the susceptibility to cancers in various populations hypothesizing that the lack of these enzymes may impair metabolic elimination of various carcinogenic compound thereby increasing risk toward cancers. These phase II GSTs *GSTT1*, *GSTM1* and *GSTP1* catalyse glutathione-mediated reduction of exogenous and endogenous electrophiles. These GSTs have broad and overlapping substrate specificities and it has been hypothesized that allelic variants associated with less effective detoxification of potential carcinogens may confer an increased susceptibility to cancer.⁶³ Lack of *GSTM1* enzyme may impair metabolic elimination of carcinogenic compounds from the body thereby increasing cancer risk.^{64,65}

CNV AND CANCER

Understanding cancer genetics and identifying all possible variant alleles that might predispose one to a variety of cancer is the prime objective. To this end, several biomarkers through SNP studies have revealed an association with cancer and many other complex traits.⁶⁶

In search for the common CNVs, which are associated with malignancy, a map was created that cataloged all the known CNVs whose loci coincide with that of the already known cancer-related

genes. Upon analysis it was found that 49 cancer genes directly encompassed or overlapped by CNV in more than one person in a large reference population. Further, validating the initial observation it was found that many of the genes were reported in the DGV and 40% of the cancer-related genes are disrupted by a CNV as analyzed by DGV. Thus, it can be proposed that structural variations are found to be associated with risk for cancer. Deletions and duplication in the cancer-related genes are found to be polymorphic in different population.⁶⁷

One of the examples of CNV as a risk to cancer with conferred phenotypic effect is *MTUS1* gene, which maps on chromosome 8p spanning a deletion of 1128 bp covering the entire exon 4 of the gene. When studied in the German population, the deletion of exon 4 of *MTUS1* was found to be associated with the slower progression of disease in both familial and high-risk breast cancer patients.⁶⁸

There are many large structural variations, which might predispose to cancer but it has been less appreciated as the deletions and duplication breakpoints are not well characterized in many cases and the PCR method to detect these large structural variation is not reliable. Therefore, to characterize these structural variations involved in cancer syndromes newer methods such as multiplex ligation-dependent probe amplification and others needs to be used, which allow the detection of copy number changes in a single gene or exon.

Recently, apart from the constitutional mutations predisposing to cancer there are several acquired (somatic) copy number alterations present in the tumor genome, which need to be analyzed with the help of high-throughput technology.⁶⁷

CONCLUSION

The human genome variations comprise of SNPs, as well as other variants such as CNVs, INDELs, inversions and larger structural alterations. SNPs have been widely studied in various populations primarily because of the ease of detection in large number of samples. However, recent technological advances have opened up investigations into other types of variations mentioned above. In past few years, we have learned a lot about the CNVs and their implication on our health and diseases. Observation from the comprehensive maps generated by Redon *et al.*⁶ and Jakobsson *et al.*⁶⁹ have established CNVs as one of the prominent genetic variation having an important role in inter-individual and inter-ethnic differences in susceptibility to common and complex diseases.⁷⁰

Still more technical advances are awaited for large-scale survey of CNVs, INDELs and other variants in many racial and ethnic populations. CNVs are also thought to have a key role in the evolution by gene duplication and exon reshuffling contributing to major phenotypic consequences. Our attempts to establish such variations and their link to health and disease will remain limited until reliable data of all such genetic variations in several different populations is generated. Only then the reliable genetic basis of disease development and drug response will be understood. It is also conceivable that a combination of several genetic variants will dictate the development of complex diseases. With this limitation in this review an attempt is made to evaluate link between the known CNVs to the development of complex diseases. It is heartening that in several cases an association is established between diseases and CNVs, however, the observations need to be independently replicated in other populations.

FUTURE PERSPECTIVE

Until recently SNP was the basis of studying human genome variability and its contribution to phenotypic variations and disease susceptibility. However, with the discovery of other structural

variations (deletions, duplications and inversions) a better understanding of the genetic variability or diseases susceptibility is emerging. It has been seen that CNVs contribute to phenotypic change or disease by various molecular mechanisms mentioned elsewhere. Based on such mechanism, the structural variations predispose to susceptibility to various complex diseases. The mechanism of disease development may not be very apparent in several cases. For any new CNV locus, it has to be studied in different populations to catalog its extent of diversity. The observation then has to be validated with a large sample size using the high-resolution simple method yielding distribution map of such CNV.

Another important aspect with the discovery of structural variation is to study its evolutionary perspective. For example, it has been reported in a study that segmental duplication events are found to be affecting the genome variability greater than single base-pair change in chimpanzee and human genomes.^{71–73} CNV distribution among various populations should pave the way to understand evolution and mechanism of development of newer CNV. Finally, it is very important to take in account all structural variations to fully understand the mechanism underlying the phenotype, disease development and human diversity. The available variants then can be studied in all possible permutations and combinations to identify a right combination leading to phenotypic changes. In future, when this becomes possible, we may have most phenotypes linked to a combination of few variants. That will be the ultimate outcome of the human genome initiative started in 1980s.

- Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nat. Rev. Genet.* **7**, 85–97 (2006).
- Lupski, J. R. Structural variation in the human genome. *N. Engl. J. Med.* **356**, 1169–1171 (2007).
- Lupski, J. R. & Stankiewicz, P. Molecular mechanisms for rearrangements and their conveyed phenotypes in genomic disorders. *PLoS Genet.* **1**, 627–633 (2005).
- Dhawani, D. & Padh, H. Pharmacogenetics: technologies to detect copy number variations. *Curr. Opin. Mol. Ther.* **11**, 670–680 (2009).
- Boone, P. M., Bacino, C. A., Shaw, C., Eng, P. A., Hixson, P. M., Pursley, A. N. *et al.* Detection of clinically relevant exonic copy-number changes by array CGH. *Hum. Mutat.* **31**, 1326–1342 (2010).
- Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).
- Zhang, F., Gu, W., Hurler, M. E. & Lupski, J. R. Copy number variation in human health, disease, and evolution. *Annu. Rev. Genom. Hum. Genet.* **10**, 45–481 (2009).
- Durbin, R. M., Abecasis, G. R., Altshuler, D. L., Auton, A., Brooks, L. D. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
- Firth, H. V., Richards, S. M., Bevan, A. P., Clayton, S., Corpas, M., Rajan, D. *et al.* DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am. J. Hum. Genet.* **84**, 524–533 (2009).
- Stankiewicz, P. & Lupski, J. R. Structural variation in the human genome and its role in disease. *Annu. Rev. Med.* **61**, 437–455 (2010).
- Estivill, X. & Armengol, L. Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet.* **3**, 1787–1799 (2007).
- Polymereopoulos, M. H., Higgins, J. J., Golbe, L. I., Johnson, W. G., Ide, S. E., Dilorio, G. *et al.* Mapping of a gene for Parkinson's disease to chromosome 4q21–q23. *Science* **274**, 1197–1199 (1996).
- Singleton, A. B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J. *et al.* α -Synuclein locus triplication causes Parkinson's disease. *Science* **302**, 841 (2003).
- Sennvik, K., Fastbom, J., Blomberg, M., Wahlund, L. O., Winblad, B. & Benedikz, E. Levels of alpha and beta-secretase cleaved amyloid precursor protein in the cerebrospinal fluid of Alzheimer's disease patients. *Neurosci. Lett.* **278**, 169–172 (2000).
- Rovelet-Lecrux, A., Hannequin, D., Raux, G., Le Meur, N., Laquerriere, A., Vital, A. *et al.* APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat. Genet.* **38**, 24–26 (2006).
- Cabrejo, L., Guyant-Marechal, L., Laquerriere, A., Vercelletto, M., De La Fourniere, F., Thomas-Anterion, C. *et al.* Phenotype associated with APP duplication in five families. *Brain* **129**, 2966–2976 (2006).
- Goate, A., Chartier-Harlin, M. C., Mullan, M., Brown, J., Crawford, F., Fidani, L. *et al.* Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349**, 704–706 (1991).
- Theuns, J., Brouwers, N., Engelborghs, S., Sleegers, K., Bogaerts, V., Corsmit, E. *et al.* Promoter mutations that increase amyloid precursor-protein expression are associated with Alzheimer disease. *Am. J. Hum. Genet.* **78**, 936–946 (2006).
- Roeleveld, N., Zielhuis, G. A. & Gabreels, F. The prevalence of mental retardation: a critical review of recent literature. *Dev. Med. Child. Neurol.* **39**, 125–132 (1997).
- Froyen, G., Corbett, M., Vandewalle, J., Jarvela, I., Lawrence, O., Meldrum, C. *et al.* Submicroscopic duplications of the hydroxysteroid dehydrogenase *HSD17B10* and the E3 ubiquitin ligase *HUWE1* are associated with mental retardation. *Am. J. Hum. Genet.* **82**, 432–443 (2008).
- Bauters, M., Van Esch, H., Friez, M. J., Boespflug-Tanguy, O., Zenker, M., Vianna-Morgante, A. M. *et al.* Nonrecurrent *MECP2* duplications mediated by genomic architecture driven DNA breaks and break-induced replication repair. *Genome Res.* **18**, 847–858 (2008).
- Carvalho, C. M., Zhang, F., Liu, P., Patel, P., Sahoo, T., Bacino, C. A. *et al.* Some complex rearrangements in patients with duplication of *MECP2* may occur by fork stalling and template switching. *Hum. Mol. Genet.* **18**, 2188–2203 (2008).
- del Gaudio, D., Fang, P., Scaglia, F., Ward, P. A., Craigen, W. J., Glaze, D. G. *et al.* Increased *MECP2* gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. *Genet. Med.* **8**, 784–792 (2006).
- VanEsch, H., Bauters, M., Ignatius, J., Jansen, M., Raynaud, M., Hollanders, K. *et al.* Duplication of the *MECP2* region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *Am. J. Hum. Genet.* **77**, 442–453 (2005).
- Bi, W., Sapir, T., Shchelochkov, O. A., Zhang, F., Withers, M. A., Hunter, J. V. *et al.* Increased *LIS1* expression affects human and mouse brain development. *Nat. Genet.* **41**, 168–177 (2009).
- Ching, M. S., Shen, Y., Tan, W. H., Jeste, S. S., Morrow, E. M., Chen, X. *et al.* Deletion of the *NRXN1* (neurexin-1) predisposes to a wide spectrum of developmental disorders. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **153B**, 937–947 (2010).
- Bailey, A., Phillips, W. & Rutter, M. Autism: towards an integration of clinical, genetic, neuropsychological, and neurobiological perspective. *J. Child. Psychol. Psychiatry* **37**, 89–126 (1996).
- Lucito, R., Healy, J., Alexander, J., Reiner, A., Esposito, D., Chi, M. *et al.* Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res.* **13**, 2291–2305 (2003).
- 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* **305**, 525–528 (2007).
- Weiss, L. A., Shen, Y., Korn, J. M., Arking, D. E., Miller, D. T., Fossdal, R. *et al.* Association between micro deletion and micro duplication at 16p 11.2 and autism. *N. Engl. J. Med.* **358**, 667–675 (2008).
- Morrow, E. M., Yoo, S. Y., Flavell, S. W., Kim, T. K., Lin, Y., Hill, R. S. *et al.* Identifying autism loci and genes by tracing recent shared ancestry. *Science* **321**, 218–223 (2008).
- Walsh, T., McClellan, J. M., McCarthy, S. E., Addington, A. M., Pierce, S. B., Cooper, G. M. *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**, 539–543 (2008).
- Xu, B., Roos, J. L., Levy, S., van Rensburg, E. J., Gogos, J. A. & Karayiorgou, M. Strong association of *de novo* copy number mutations with sporadic schizophrenia. *Nat. Genet.* **40**, 880–885 (2008).
- Stefansson, H., Rujescu, D., Cichon, S., Pietilainen, O. P., Ingason, A., Steinberg, S. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–236 (2008).
- The Int Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237–241 (2008).
- Karayiorgou, M., Morris, M. A., Morrow, B., Shprintzen, R. J., Goldberg, R., Borrow, J. *et al.* Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc. Natl Acad. Sci. USA* **92**, 7612–7616 (1995).
- Vrijenhoek, T., Buizer-Voskamp, J. E., Van Der Stelt, I., Strengman, E., Sabatti, C., Geurts van Kessel, A. *et al.* Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *Am. J. Hum. Genet.* **83**, 504–510 (2008).
- Lachman, H. M., Pedrosa, E., Petruolo, O. A., Cockerham, M., Papolos, A., Novak, T. *et al.* Increase in *GSK3beta* gene copy number variation in bipolar disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **144B**, 259–265 (2007).
- Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G. *et al.* The influence of *CCL3L1* gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**, 1434–1440 (2005).
- McKinney, C., Merriam, M. E., Chapman, P. T., Gow, P. J., Harrison, A. A., Highton, J. *et al.* Evidence for an influence of chemokine ligand 3-like 1 (*CCL3L1*) gene copy number on susceptibility to rheumatoid arthritis. *Ann. Rheum. Dis.* **67**, 409–413 (2008).
- Fellermann, K., Stange, D. E., Schaeffeler, E., Schmalz, H., Wehkamp, J., Bevins, C. L. *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.* **79**, 439–448 (2006).
- Hollox, E. J., Huffmeier, U., Zeeuwen, P. L., Palla, R., Lascorz, J., Rodijk-Olthuis, D. *et al.* Psoriasis is associated with increased beta-defensin genomic copy number. *Nat. Genet.* **40**, 23–25 (2008).
- de Cid, R., Riveira-Munoz, E., Zeeuwen, P. L., Robarge, J., Liao, W., Dannhauser, E. N. *et al.* Deletion of the late cornified envelope *LCE3B* and *LCE3C* genes as a susceptibility factor for psoriasis. *Nat. Genet.* **41**, 211–215 (2009).

- 44 McCarroll, S. A., Huett, A., Kuballa, P., Chilewski, S. D., Landry, A., Goyette, P. *et al.* Deletion polymorphism upstream of *IRGM* associated with altered IRGM expression and Crohn's disease. *Nat. Genet.* **40**, 1107–1112 (2008).
- 45 Le Marechal, C., Masson, E., Chen, J. M., Morel, F., Ruzsiewski, P., Levy, P. *et al.* Hereditary pancreatitis caused by triplication of the trypsinogen locus. *Nat. Genet.* **38**, 1372–1374 (2006).
- 46 Aitman, T. J., Dong, R., Vyse, T. J., Norsworthy, P. J., Johnson, M. D., Smith, J. *et al.* Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* **439**, 851–855 (2006).
- 47 Willcocks, L. C., Lyons, P. A., Clatworthy, M. R., Robinson, J. I., Yang, W., Newland, S. A. *et al.* Copy number of *FCGR3B*, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. *J. Exp. Med.* **205**, 1573–1582 (2008).
- 48 Yang, Y., Chung, E. K., Wu, Y. L., Savelli, S. L., Nagaraja, H. N., Zhou, B. *et al.* Gene copy-number variation and associated polymorphisms of complement component *C4* in human systemic lupus erythematosus (SLE): Low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am. J. Hum. Genet.* **80**, 1037–1054 (2007).
- 49 Brasch-Andersen, C., Christiansen, L., Tan, Q., Haagerup, A., Vestbo, J. & Kruse, T. A. Possible gene dosage effect of glutathione-S-transferase on atopic asthma: using real-time PCR for quantitation of *GSTM1* and *GSTT1* gene copy numbers. *Hum. Mutat.* **24**, 208–214 (2004).
- 50 Ivaschenko, T. E., Sideleva, O. G. & Baranov, V. S. Glutathione-S-transferase micro and theta gene polymorphisms as new risk factors of atopic asthma. *J. Mol. Med.* **80**, 39–43 (2002).
- 51 Tamer, L., Alikoglu, M., Ates, N. A., Yildirim, H., Ercan, B., Saritas, E. *et al.* Glutathione-S-transferase gene polymorphisms (*GSTT1*, *GSTM1*, *GSTP1*) as increased risk factors for asthma. *Respirology* **9**, 493–498 (2004).
- 52 Pollex, R. L. & Hegele, R. A. Copy number variation in the human genome and its implications for cardiovascular disease. *Circulation* **115**, 3130–3138 (2007).
- 53 Pollex, R. L. & Hegele, R. A. Genomic copy number variation and its potential role in lipoprotein and metabolic phenotypes. *Curr. Opin. Lipidol.* **18**, 174–180 (2007).
- 54 Lanktree, M. & Hegele, R. A. Copy number variation in metabolic phenotypes. *Cytogenet. Genome. Res.* **123**, 169–175 (2008).
- 55 Jeon, J. P., Shim, S. M., Nam, H. Y., Ryu, G. M., Hong, E. J., Kim, H. L. *et al.* Copy number variation at leptin receptor gene locus associated with metabolic traits and the risk of type 2 diabetes mellitus. *BMC Genomics* **11**, 426 (2010).
- 56 Ogden, C. L., Carroll, M. D., Curtin, L. R., McDowell, M. A., Tabak, C. J. & Flegal, K. M. Prevalence of overweight and obesity in the United States. *JAMA* **295**, 1549–1555 (2006).
- 57 Bochukova, E. G., Huang, N., Keogh, J., Henning, E., Purmann, C., Blaszczyk, K. *et al.* Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* **463**, 666–670 (2010).
- 58 Sha, B. Y., Yang, T. L., Zhao, L. J., Chen, X. D., Guo, Y., Chen, Y. *et al.* Genome-wide association study suggested copy number variation may be associated with body mass index in the Chinese population. *J. Hum. Genet.* **54**, 199–202 (2009).
- 59 Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R. *et al.* Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* **39**, 1256–1260 (2007).
- 60 Oscarson, M., McLellan, R. A., Gullsten, H., Yue, Q. Y., Lang, M. A., Bernal, M. L. *et al.* Characterisation and PCR-based detection of a *CYP2A6* gene deletion found at a high frequency in a Chinese population. *FEBS Lett.* **448**, 105–110 (1999).
- 61 Fukami, T., Nakajima, M., Yamanaka, H., Fukushima, Y., McLeod, H. L. & Yokoi, T. A novel duplication type of *CYP2A6* gene in African-American population. *Drug Metab. Dispos.* **35**, 515–520 (2007).
- 62 Hebring, S. J., Adjei, A. A., Baer, J. L., Jenkins, G. D., Zhang, J., Cunningham, J. M. *et al.* Human *SULT1A1* gene: copy number differences and functional implications. *Hum. Mol. Genet.* **16**, 463–470 (2007).
- 63 Spurdle, A. B., Webb, P. M., Purdie, D. M., Chen, X., Green, A. & Chenevix-Trench, G. Polymorphism at the glutathione S-transferase *GSTM1*, *GSTT1* and *GSTP1* loci: risk of ovarian cancer by histological subtype. *Carcinogenesis* **22**, 67–72 (2001).
- 64 Huang, S. R., Chen, P., Wisel, S., Duan, S., Zhang, W., Cook, E. H. *et al.* Population-specific *GSTM1* copy number variation. *Hum. Mol. Genet.* **18**, 366–372 (2009).
- 65 Rebbeck, T. R. Molecular epidemiology of the human glutathione S-transferase genotypes *GSTM1* and *GSTT1* in cancer susceptibility. *Cancer Epidemiol. Biomarkers Prev.* **6**, 733–743 (1997).
- 66 Barh, D., Agate, V., Dhawan, D., Agate, V., Bajpai, V. & Padh, H. Cancer biomarkers for diagnosis, prognosis and therapy. in *Cellular and Molecular Therapeutics* (eds Whitehouse, D. & Rapley, R.) (John Wiley & Sons Ltd., UK, in press).
- 67 Shlien, A. & Malkin, D. Copy number variations and cancer. *Genome Med.* **1**, 62 (2009).
- 68 Frank, B., Bermejo, J. L., Hemminki, K., Sutter, C., Wappenschmidt, B., Meindl, A. *et al.* Copy number variant in the candidate tumor suppressor gene *MTUS1* and familial breast cancer risk. *Carcinogenesis* **28**, 1442–1445 (2007).
- 69 Jakobsson, M., Scholz, S. W., Scheet, P., Gibbs, J. R., VanLiere, J. M., Fung, H. C. *et al.* Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **451**, 998–1003 (2008).
- 70 Fanciulli, M., Petretto, E. & Aitman, T. J. Gene copy number variation and common human disease. *Clin. Genet.* **77**, 201–213 (2010).
- 71 Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
- 72 Cheng, Z., Ventura, M., She, X., Khaitovich, P., Graves, T., Osoegawa, K. *et al.* A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* **437**, 88–93 (2005).
- 73 Perry, G. H., Tchinda, J., McGrath, S. D., Zhang, J., Picker, S. R., Cáceres, A. M. *et al.* Hotspots for copy number variation in chimpanzees and humans. *Proc. Natl Acad. Sci. USA* **103**, 8006–8011 (2006).