CLINICAL GENETICS
An International Journal of Genetics and Molecular Medicine

Clin Genet 2010: 77: 201–213
Printed in Singapore. All rights reserved

CLINICAL GENETICS
doi: 10.1111/j.1399-0004.2009.01342.x

Review

Gene copy number variation and common human disease

Fanciulli M, Petretto E, Aitman TJ. Gene copy number variation and common human disease.
Clin Genet 2010: 77: 201–213. © John Wiley & Sons A/S, 2009

Variation in gene copy number is increasingly recognized as a common, heritable source of inter-individual differences in genomic sequence. The role of copy number variation is well established in the pathogenesis of rare genomic disorders. More recently, germline and somatic copy number variation have been shown to be important pathogenic factors in a range of common diseases, including infectious, autoimmune and neuropsychiatric diseases and cancer. In this review, we describe the range of methods available for measuring copy number variants (CNVs) in individuals and populations, including the limitations of presently available assays, and highlight some key examples of common diseases in which CNVs have been shown clearly to have a pathogenic role. Although there has been major progress in this field in the last 5 years, understanding the full contribution of CNVs to the genetic basis of common diseases will require further studies, with more accurate CNV assays and larger cohorts than have presently been completed.

M Fanciulli[a], E Petretto[a,b] and TJ Aitman[a]

[a]Medical Research Council Clinical Sciences Centre, Faculty of Medicine, Imperial College London, Hammersmith Hospital, London, W12 0NN, UK, and [b]Department of Epidemiology and Public Health, Faculty of Medicine, Imperial College, Norfolk Place, London, W2 1 PG, UK

Key words: gene copy number – gene deletion – gene duplication – human common diseases – human genomic disorders

Corresponding author: Timothy J. Aitman, Medical Research Council Clinical Sciences Centre, Faculty of Medicine, Imperial College London, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK.
Tel: +44-(0)208 383 4253/8336;
fax: +44-(0)208 383 8577;
e-mail: t.aitman@csc.mrc.ac.uk

Received 30 July 2009, revised and accepted for publication 28 October 2009

During the past several decades, significant progress has been made in defining the genetic basis of Mendelian diseases. More recently, genome-wide association studies (GWAS) using single nucleotide polymorphism (SNP) markers have localized and, in many cases, identified susceptibility genes for a range of common, complex diseases. Despite the success of GWAS in identifying disease genes, the newly discovered susceptibility genes explain only a modest fraction of trait heritability or risk in common diseases, leading to uncertainty as to the genetic basis of the 'missing heritability' (1–3).

Genomic copy number variants (CNVs) were recognized as a cause of human disease in the 1980s, but their population frequency was presumed to be low and directly related to specific 'genomic disorders' (4, 5). For example, copy number variation at the alpha-globin locus was shown to be a cause of alpha-thalassaemia (6). During the 1990s, deletions and duplications were found as a cause of a number of other single gene disorders, although these are mostly severe syndromes caused by *de novo* mutations that would not be expected to be transmitted from generation to generation. Several examples of these disorders are included in Table 1. These rare genomic disorders and the mechanisms underlying the genomic origins of CNVs have been reviewed in detail elsewhere (5, 7–24) and are beyond the scope of this review.

In the last 5 years, variation in gene copy number has been recognized as a common source of inter-individual variation in the genomes of healthy individuals (25, 26). A CNV may be defined as a DNA segment of 1 kb or more in

Table 1. Human disorders associated with germline CNVs

Disease	Locus	Gene(s)	Structural variation	Reference(s)
Rare genomic disorders				
Cri du chat syndrome	5p15.2	Multiple genes	Deletion	(8)
Spinal muscular atrophy (SMA1-4)	5q12.2-q13.3	*BIRC1, GTF2H2 SERF1A, SERF1B SMN1, SMN2*	Deletion	(10)
Williams-Beuren syndrome	7q11.23	Multiple genes	Deletion/duplication	(11)
CHARGE syndrome	8q12.1	*CHD7*	Deletion	(12)
Charcot-Marie-Tooth disease type 4B2	11p15.4	*ADM SBF2*	Deletion	(13)
Prader-Willi and Angelman syndrome	15q11-q13	*ATP10A, OCA2 OR4M2, OR4N4 UBE3A*	Deletion	(14)
Smith-Magenis syndrome	17p11.2	*ATPF2, COPS3 DRG2, MED9 NT5M, RAI1 SMCR8, SREBF1*	Deletion	(15)
Charcot-Marie-Tooth disease type 1A	17p11.2	*COX10, HS3ST3A1 PMP22, TEKT3 ZNF286*	Deletion	(16)
Neurofibromatosis type 1	17q11.2	*NF1*	Deletion/duplication	(17)
Miller-Dieker lissencephaly syndrome	17p13.3	*LIS1*	Deletion	(18)
DiGeorge/Velocardiofacial syndrome	22q11.2	*GGT2, GNB1L HIC2, TBX1*	Deletion/duplication	(19) (20)
Pelizaeus-Merzbacher disease	Xq22	*PLP1*	Deletion/duplication	(21)
Common disorders				
HIV/AIDS susceptibility	17q11.2	*CCL3L1*	Deletion	(37)
Systemic lupus erythematosus	1q23	*FCGR3B*	Deletion	(38–40)
	6p21.3	*C4*	Deletion	(78)
Rheumatoid arthritis	17q11.2	*CCL3L1*	Duplication	(69)
Kawasaki disease	17q11.2	*CCL3L1*	Duplication	(72)
Crohn's disease	8p23.1	*HBD-2*	Deletion	(74)
	5q33.1	*IRGM*	Deletion	(77)
Psoriasis	8p23.1	*HBD-2*	Duplication	(51)
ANCA-associated vasculitis	1q23	*FCGR3B*	Deletion	(39)
	1q23	*FCGR3B*	Duplication	(40)
Atopic asthma	1q13.3	*GSTM1*	Deletion	(79)
	22q11.2	*GSTT1*	Deletion	(80, 81)
Autism spectrum disorders	15q11-q13	Multiple genes	Deletion/duplication	(85)
	16p11.2	Multiple genes	Deletion/duplication	(86)
	22q13.3	*SHANK3*	Deletion	(87)
	Xp22.33	*NLGN4*	Deletion	(88)
	2p16.3	*NRXN1*	Deletion	(89)
	Multiple	Multiple genes	*De novo* CNVs	
Schizophrenia	2q34	*ERBB4*	Deletion	(91)
	5p13.3	*SLC1A3*	Deletion	(93)
	2q31.2	*RAPGEF4*	Deletion	(90)
	12.24	*CIT*	Deletion	(92)
	Multiple	Multiple genes	*De novo* CNVs	
Epilepsy	15q13.3	*CHRNA7*	Deletion	(92, 93) Helbig et al. (2008)
Parkinson's disease	4q22	*SNCA*	Duplication	(99, 100)
Amyotrophic lateral sclerosis	5q12.2-q13.3	*SMN1 SMN2*	Deletion	(104)
Familial hypercholesterolemia	19p13.2	*LDLR*	Deletion/duplication	(113)

AIDS, acquired immune deficiency syndrome; ANCA, anti-neutrophil cytoplasmic antibody; CNV, copy number variant; HIV, human immunodeficiency virus.

length that is present in a variable number of copies in the genome. These sub-microscopic 'structural' variants include simple deletions, insertions and duplications as well as more complex multi-site variants such as single nucleotide variants with complex characteristics due to CNV or gene conversion (27, 28) (Fig. 1).

In 2006, Redon et al. (29) constructed the first comprehensive CNV map of the human genome, and made several key observations. Amongst the 270 individuals studied, 12% of the human genome (approximately 360 Mb pairs) was found to be covered by CNVs, with a preponderance of smaller size rearrangements (<20 kb). The genomic regions encompassed by these CNVs contain hundreds of genes and other functional elements, and many CNVs reached a population frequency of greater than 1%, in which case they are referred to as copy number polymorphisms (29). These observations, together with the substantial inter-population differences in gene copy number reported by Redon et al. (29) and others (30), suggest that CNVs could play an important role in evolution and adaptation to different environments, and potentially in susceptibility to common diseases. Supporting the importance of structural variants in modulating complex phenotypes, Stranger et al. examined mRNA levels in lymphoblastoid cell lines from 210 unrelated individuals and concluded that CNVs accounted for almost 20% of the genetically determined variation in gene expression independent of other genetic variants, including SNPs (31). Studies in animal models support this view, showing that CNVs can also affect the expression of genes adjacent to the genomic location of CNVs, most likely through alteration of adjacent regulatory sequences (32, 33).

With the growth of information on CNVs in the human genome, the accurate annotation of these structural variations has become increasingly important. At present, several databases are currently available for genome-wide investigation of genomic variants (34–36).

The number of reports of association between CNVs and common diseases has increased markedly over the past 5 years (Fig. 2). CNVs have been associated with a growing number of common complex diseases, including human immunodeficiency virus (HIV) (37), autoimmune diseases (38–40) and a spectrum of neuropsychiatric disorders (7). As well as the associations with heritable CNVs, *de novo* mutations, both germline and somatic, have been reported as causes of several disorders.

Studies on human dystrophin gene suggested that the mutation rate for *de novo* CNVs was equal to one deletion per 8 newborns and one duplication per 50 newborns (41, 42). While heritable *de novo* CNVs have been associated more commonly to neuropsychiatric disorders (7), somatic CNVs that are detected, e.g. in the genomes of cancer cells are believed to play an important role in malignant transformation and cancer pathogenesis (43–45). However, the methodologies for accurate measurement of gene copy number remain challenging, and still limit the scope of present genetic studies associating CNV and disease, both at the individual gene level and at the level of the genome (30, 46).

In this review, we discuss the currently available methods for measurement of CNVs and then focus on the role of transmissible germline CNVs and somatic CNVs in the pathogenesis of common diseases.

Methods for CNV detection

Several methods have been developed for measurement of copy number within genomic CNVs. These assays can be either locus specific or genome wide and are mostly either array based or polymerase chain reaction (PCR) based, giving different degrees of resolution, precision and throughput.

Comparative genomic hybridization arrays

Comparative genomic hybridization (CGH) arrays have been the most widely used method for genome-wide identification of CNVs. For fabrication of CGH arrays, genomic DNA sequences across the genome are spotted onto a solid support (usually glass). Test and reference DNAs (labelled with different fluorescent markers) are then simultaneously hybridized to the array and the respective signal intensities compared to give a measure of copy number. The length of DNA sequences used to construct the arrays ranges from 200 kb (bacterial artificial chromosome or BAC arrays) down to 25 bp (oligonucleotide array) in size. Although BAC arrays provide better coverage of the genome with a higher signal-to-noise ratio, they cannot identify CNVs of <50 kb due to the large size of the BAC DNA probe (47). In the past, this led to an overestimate of the size of discovered CNVs and still represents an important limitation given that the majority of CNVs are typically rearrangements of <20 kb in size (29). In contrast, high-density oligonucleotide arrays provide better resolution with more accurate definition of CNV boundaries, but hybridization intensities have a

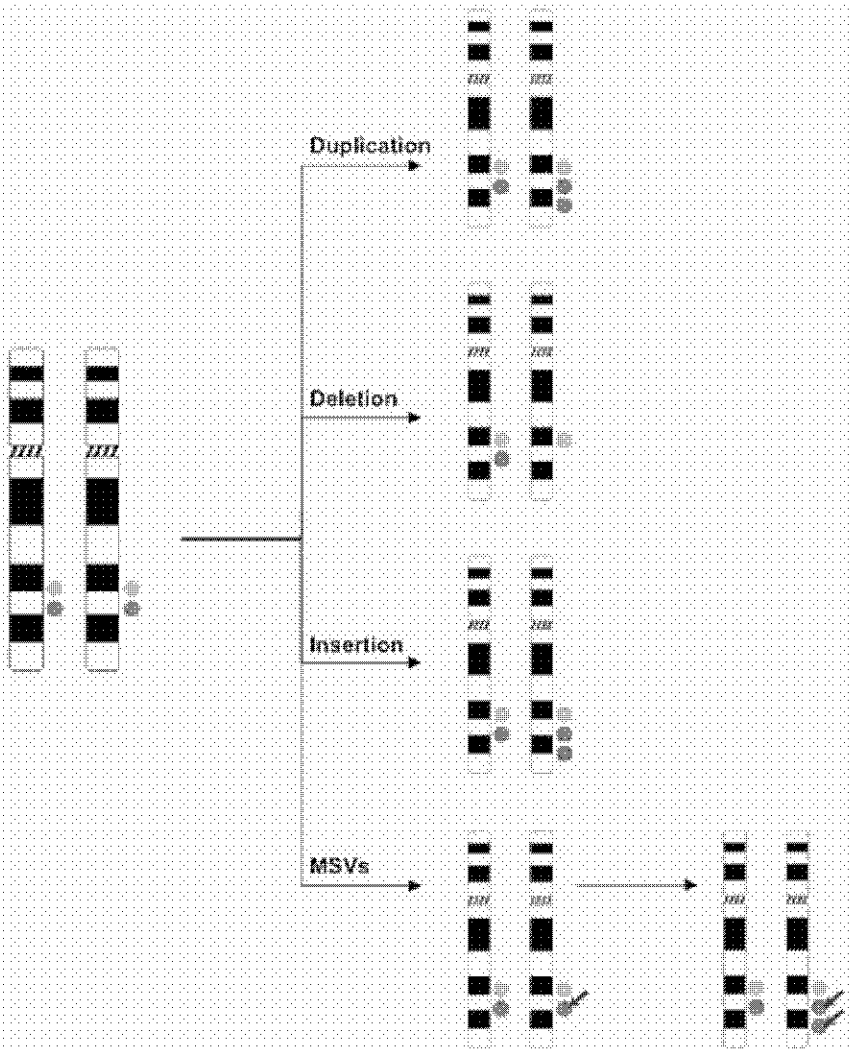

Fig. 1. Different classes of copy number variants (CNVs). Genomic regions or genes (indicated with dots) can undergo duplication, deletion and insertion events. Multi-site variants (MSVs) refer to more complex genomic rearrangements, including concurrent CNV and mutation or gene conversions. The MSVs represented in the figure show a mutation event (blue arrow) followed by gene duplication.

poor signal-to-noise ratio that can lead to considerable experimental variability between studies (48).

SNP arrays

Arrays designed to detect SNPs can also be used to identify CNVs by looking for differences in signal intensity that are independent of genotype. SNPs in CNV regions may appear to violate Mendelian inheritance–because they are assaying non-diploid numbers of alleles–or may not to be in Hardy–Weinberg equilibrium (29). Therefore, such SNPs often failed quality control processes (46) and were likely to be discarded during the original design of genotyping arrays, leaving gaps in genome coverage. Furthermore, linkage disequilibrium around CNVs has been reported to be lower than SNPs. This can be explained by the fact that CNVs might undergo recurrent mutation, reversion or transposition events that would modify the non-random associations at different loci. For these reasons, the detection of CNVs, by tagging approaches, using SNP arrays has previously been limited. Newer generation of SNP arrays have now been designed to fill gaps in whole-genome coverage and also to include non-polymorphic probes that are optimized for copy number measurements (46).

Quantitative real-time PCR

Quantitative real-time polymerase chain reaction (qPCR) has been one of the most commonly used methods for screening targeted genomic regions for CNVs. It is an efficient method for detection of deletions or duplications at single loci as well as being a relatively high-throughput and technically straightforward assay. However, qPCR is not suitable for simultaneous amplification of many targets of interest in a single reaction and, in common with the other assays described above, does not permit precise integer measurements

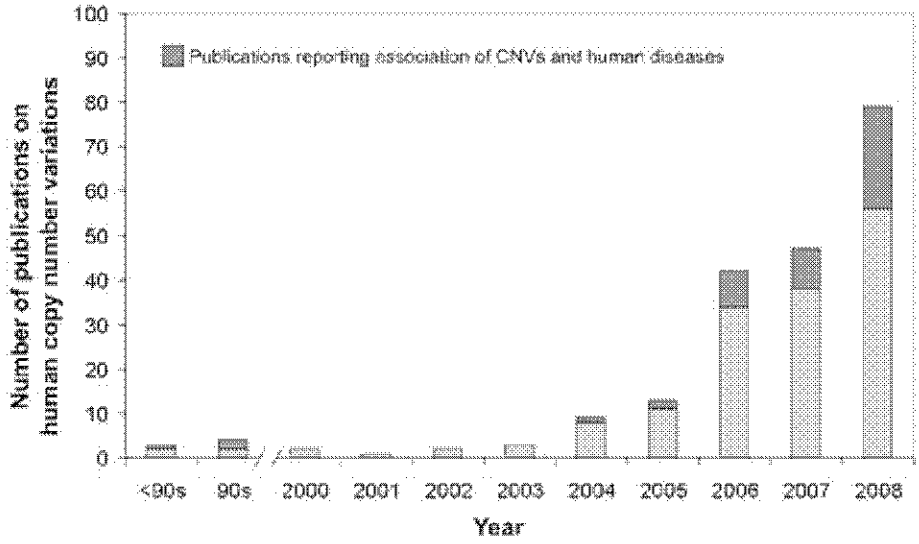

Fig. 2. Number of publications reporting copy number variations (CNVs) in humans. Bars represent cumulative data for human CNVs, including analysis tools and assays for CNV detection. Articles were initially identified in a PubMed search using the query strategy: 'copy number variation' (Title) OR 'copy number polymorphism' (Title) OR 'copy number' (Title) OR 'duplication' (Title) AND (['year' (EDAT): 'year' (EDAT)] AND 'humans' (MeSH Terms) AND 'English' (lang) AND 'pubmed pmc local' (sb)). The resulting list of articles was manually refined by excluding papers reporting somatic rearrangements related to cancer.

of gene copy number because copy number estimates from these assays form a continuous distribution (49–51).

Paralogue ratio test

The paralogue ratio test (PRT) (50, 51) is an accurate and relatively high-throughput method for determining gene copy number at targeted single loci. This method involves the simultaneous amplification of a target element whose copy number is being estimated and another unlinked reference locus that is simultaneously amplified by the same primer pairs in the same reaction tube. The two regions are subsequently distinguished by internal differences (such as amplicon length or restriction digest), quantified and compared to estimate the copy number of the target compared to the reference. In this method, simultaneous amplification of target and control in the same reaction decreases experimental variability and improves precision of the measurements, generating copy number estimates close to integer values.

MLPA and MAPH

Multiplex ligation-dependent probe amplification (MLPA) and multiple amplifiable probe hybridization (MAPH) are alternative targeted PCR-based approaches for simultaneous analysis of multiple genomic regions (up to 40 target sequences) (52). These approaches rely on the use of oligonucleotide probes to generate locus-specific amplicons that can be resolved by capillary electrophoresis: enhanced peak signals indicate duplications while reduced peaks are indicative of deletions. These techniques are simple to apply, sensitive and, although they are more commonly employed for locus-specific studies, they can also be used for screening multiple targets.

Sequencing and genome assembly comparison

Alignment of DNA sequences from different sources has been successfully used to identify structural variants. This method consists of aligning and comparing two completed sequence assemblies and in one report has yielded 13,066 undescribed structural variations including insertions, deletions and CNVs in the human genome (53). The main limitation of this approach was in the assemblies used (International Human Genome Sequencing Consortium and National Centre for Biotechnology Information), that represented mosaics of different donor DNA sources (with neither being fully completed), and therefore were not ideal for sequence comparison. With the development of high-throughput DNA sequencing protocols, so-called next-generation sequences, it has been possible to generate new assemblies of complete genome sequences from single individuals. This enabled more robust and reliable genome comparisons and CNV identification (54, 55). All platforms that are commonly used for massively parallel DNA sequencing (such as Roche/454 FLX (56), Illumina/Solexa Genome Analyzer (57)

and Applied Biosystem Solid™ System) employ single-strand amplification of a fragment library followed by sequencing reaction of the amplified strand, while using different chemistries, detection methods and software (58). For all these platforms, the libraries are obtained directly from the genome of interest, thereby avoiding the bacterial cloning step; moreover, they enable high-throughput generation of sequence reads, ranging from several hundred thousand to tens of millions of reads (58). Next-generation sequencing platforms prompted the development of new computational methods for identification of structural variations. Most of the approaches are based on paired-end read mapping (PEM) (55, 59). This requires the generation of a library of fragments (typically 300–500 bp of genomic DNA), which is subsequently subjected to deep sequencing (massively parallel sequencing generating millions of short sequence reads from a given sample) to determine the 'paired-end spans'. The sizes of paired-end spans are compared to a reference genome to identify structural variations. The distinctive advantage of this method is that it allows the identification of inversions and provides resolution of CNV boundaries at the single nucleotide level. However, PEM is limited to detect only insertions smaller then the average insert size of the library and it cannot reliably identify variants located within complex genomic regions (60). To overcome these limitations, a new approach has been recently proposed (61). This method takes advantage of the depth of single-read coverage and employs a novel CNV calling algorithm (namely Event Wise Testing) designed for *ad hoc* analysis of read depth (RD). The coverage (RD) is estimated in non-overlapping intervals across the genome and it is used as a quantitative measure of copy number. A decrease or increase in RD across multiple consecutive genomic windows is indicative of deletion or duplication events, respectively. However, RD analysis has several limitations. For example, it is not able to ascertain balanced rearrangements or structural variations that involve highly repetitive sequences, and it shows limitations in identifying novel insertions or their precise location. PEM and RD analyses provide different and complementary advantages, and using both approaches in combination will enhance detection of different structural variations employing next-generation sequence data.

Statistical tools for disease association with CNVs

Several methods for CNV calling from aCGH data are now available, ranging from the use of simply defined thresholds of intensity values (62) to more complex statistical modelling (63–66). Examples of widely used algorithms that have been implemented for CNV calling using CGH array data include the Smith–Waterman array (SW-array) (64) and the circular binary segmentation method (63). Both are non-parametric methods, i.e. they do not make any assumption about the nature of the noise distribution and the intensity data. Within a set of ordered array log-ratios values, the SW-array method identifies 'islands' of probes whose intensity differs from a given threshold value. The circular binary segmentation algorithm recognizes 'islands' of probes by using t-test to compare their average intensity. Both methods use permutations to assess the significance of the differences between signal intensities.

When SNP arrays are used to call CNVs, one simple approach searches for unusual patterns of genotype inheritance or allelic segregation within families, and as such is not suitable for analysis in unrelated individuals (67). An alternative tool for deriving integrated CNVs and SNPs from array data has been recently developed (66). This approach assigns CNVs across regions of known common copy number polymorphism, calls SNP genotypes and discovers rare and *de novo* CNVs via a hidden Markov model that takes into account probe-specific means and variances.

For PCR-based assays, CNV calling is carried out through quantification of the genomic DNA, usually followed by binning in copy number classes using pre-defined intensity thresholds. The main challenge in the analysis of association studies of CNVs and human diseases is in minimising false-positive rates that may arise as a consequence of imprecision and inherent noise in copy number measurements (68).

Most of the genotyping platforms described earlier produce copy number estimates, which are distributed on a continuous scale in the population sample. Although forcing the raw copy number measurements into discrete copy number bins has proved adequate to discover CNVs that have a large effect on disease susceptibility (37–40, 69), the approach may result in loss of information and reduced power to detect weaker associations. Moreover, other confounding factors such as different DNA quality or sample handling between cases and controls can affect the accurate quantification of SNPs and CNVs and potentially lead to spurious associations (i.e. high false-positive rate) (68, 70).

Recently, a more appropriate statistical tool for CNV case–control association testing has been developed (68). This methodology takes into account a disease model for the trait of interest

together with a mixture model for the signal which accommodates several determining factors, including different variability and noise levels. These kinds of approach appear to be robust enough to avoid potential confounding factors, and are likely to be particularly helpful when assays with poor signal-to-noise ratio are used for the copy number measurements.

CNVs and common human diseases

Since recognition of the widespread distribution of transmissible germline CNVs in the human genome, associations have been reported to a range of common diseases, including autoimmune and infectious diseases, neuropsychiatric and cardiovascular diseases. Associations have also been reported between both germline and somatic CNVs and cancer. We review here the evidence for and nature of these associations.

CNVs in infectious and autoimmune disease

Gene ontology analysis has shown that regions of the genome affected by copy number variation are significantly enriched for genes involved in inflammation and the immune response (29, 71). To date, a number of studies have reported association between CNVs in these classes of genes and HIV, asthma, Crohn's disease and other autoimmune disorders (Table 1). CNV at *CCL3L1* has been implicated in the pathogenesis of several disease phenotypes (37, 69, 72). *CCL3L1* gene encodes a HIV-1-suppressive chemokine that is the most potent ligand known for the HIV co-receptor CCR5. *CCL3L1* copy number that was lower than the population-specific average was associated, in several ethnic groups, with both HIV susceptibility and with an accelerated rate of HIV progression (37). Subsequent smaller studies in different ethnic groups have been variable in replicating these initial findings, indicating the need for larger replication cohorts to establish firmly the link between *CCL3L1* copy number and HIV susceptibility and progression (73). In the context of autoimmune disease, it is also of interest that high copy number at *CCL3L1* has been associated with disease susceptibility to rheumatoid arthritis and with a protective effect in Kawasaki disease (69, 72).

Similarly, both high and low beta-defensin copy numbers have been associated with susceptibility to different diseases (51, 74). The proteins encoded by these genes act as cytokines linking the innate and adaptive immune responses. Consequently, quantitative variation in gene dosage might be anticipated to contribute to susceptibility to infectious and inflammatory disorders. Low beta-defensin 2 (*HBD-2*) copy number has been reported to predispose to colonic Crohn's disease in three independent cohorts (74), and high *HBD-2* copy number was associated with psoriasis in Dutch and German populations (51).

CNV at *FCGR3B* has been shown to confer increased susceptibility to a range of autoimmune disorders (38–40). The *FCGR3B* gene encodes the low affinity receptor for the Fc fragment of immunoglobulin G. *FCGR3B* is a member of a gene family whose proteins play a key role in the immune response by binding the Fc fragment of immunoglobulins, thereby providing a critical link between the humoral and cellular immune responses. FCGR3B is expressed mainly in neutrophils and is necessary for neutrophil tethering to immune complexes (75).

Variation of copy number at the *Fcgr3* gene locus was initially demonstrated in a rat model of crescentic glomerulonephritis where loss of a newly identified *Fcgr3* paralogue, *Fcgr3*-related sequence (*Fcgr3-rs*) in the Wistar Kyoto rat strain predisposed to the development of disease (38). In humans, low copy number of *FCGR3B* was shown to predispose to glomerulonephritis in systemic lupus erythematosus (SLE). Further studies of unrelated case–control cohorts in different populations found low *FCGR3B* copy number to be a genetic susceptibility factor in the development of several systemic autoimmune disorders, including SLE and anti-neutrophil cytoplasmic antibody (ANCA)-associated vasculitis (39). The association between low *FCGR3B* copy number and SLE susceptibility was replicated in a subsequent case–control study, although this study did not replicate the association between low *FCGR3B* copy number and ANCA vasculitis, instead finding an association with high copy number and ANCA vasculitis (40). More recently, low *FCGR3B* copy number has been found to increase the risk of developing rheumatoid arthritis in a Caucasian cohort from The Netherlands, with smaller cohorts, from UK and New Zealand, showing similar non-significant trends (M. F., T. A., unpublished data, Tony Merriman, personal communication). The association of *FCGR3B* copy number and SLE therefore appears robust, but differences in methodology, sample size and ethnicity may account for some of the varying results in ANCA vasculitis, for which larger cohort sizes studied with the recently developed and more precise PRT assay for *FCGR3B* number (76) would help to provide definitive data.

At least two inflammatory or autoimmune diseases have been shown to be associated with CNVs at different gene loci. For instance, as discussed earlier, Crohn's disease is associated with low copy number at *HBD-2*, but also with a common 20-kb deletion polymorphism immediately upstream of the *IRGM* gene that encodes a member of the p47 immunity-related GTPase family (77). In addition to the association with *FCGR3B* copy number, SLE susceptibility has also been reported to be significantly increased among subjects with a lower number of copies of complement component *C*4 (78).

Associations between genomic deletions at the *GSTM1* and *GSTT1* loci and increased risk of atopic asthma have been reported in a number of independent studies (79–81), although previously Fryer et al. found no association between *GSTM1* and *GSTT1* and asthma or atopy (82). The conflicting results are most likely due to differences in the study population, the phenotype studied or different methodologies, particularly the inability of the assay used by Fryer et al. (82) to distinguish between subjects with one or two gene copies of *GSTM1* or *GSTT1*.

CNV diseases of the nervous system

In addition to the well-established causal relationship between CNV and rare neurological diseases (83, 84) (Table 1), associations have more recently been reported between several common neurological and psychiatric diseases and both *de novo* and inherited CNVs. These diseases include conditions such as autism, schizophrenia and epilepsy as well as some cases of neurodegenerative diseases such as Parkinson's disease, amyotrophic lateral sclerosis (ALS) and autosomal dominant Alzheimer's disease (85–95).

Autism and autism spectrum disorders (ASDs) have a population frequency of 15–60 per 10,000 children and have a high concordance rate in monozygotic compared with dizygotic twins (around 80% *vs* 5–10%). Cytogenetic abnormalities have been observed in 5–10% of affected individuals, with duplications at 15q11-q13 observed in 1–3% of cases. Family studies favour an oligogenic model of inheritance, with substantial underlying heterogeneity demonstrated in cytogenetic and linkage studies (96, 97).

In 2007, two landmark studies showed a high prevalence of *de novo* CNVs (i.e. CNVs not present in either parent) in both sporadic and familial cases of ASD compared with controls (85, 86). Sebat et al. examined 264 families with one or more cases of ASD and found *de novo* CNVs in 10% of sporadic cases, compared with 3% of families with multiple cases and 1% in control families (85). The Autism Genome Project Consortium studied over 1400 ASD families, finding a total of 154 CNVs including several recurrent CNVs, either on chromosome 15q11-13 as previously observed in autism cases or at new genomic locations including chromosome 2p16, 1q21 and at 17p12 in a region associated with Smith–Magenis syndrome that has overlaps with ASD (86). These and subsequent confirmatory studies have noted recurrent microdeletion or microduplication events spanning on chromosome 16p11.2, found in approximately 1% of ASD cases and have also highlighted the observation that *de novo* CNVs were often detected at loci that were enriched for genes such as *SHANK3*, neurexin 1 and the neuroglins that are known to regulate synaptic differentiation and regulate glutaminergic neurotransmitter release (85–88). Together, these findings have established copy number variation, in the form of microdeletion and microduplication at multiple loci, as substantial primary risk factors for development of ASD.

As with ASD, schizophrenia has also been associated with multiple, individually rare *de novo* CNVs (91, 93). A recent case–control study identified rare microdeletions and microduplications (>100 Kb) as genetic risk factors for schizophrenia (92). Walsh et al. found novel deletions and duplications in 15% of cases and in 20% of young-onset cases, compared to 5% of controls (90), with similar results reported by others (91, 93). Microdeletions and microduplications associated with schizophrenia contain an overrepresentation of genes belonging to neurodevelopmental and glutaminergic pathways (90), suggesting that multiple, individually rare CNVs affecting genes on these pathways may contribute directly to the pathogenesis of schizophrenia.

Interestingly, an overlap in some of the genes and pathways affected by specific CNVs identified has been observed for different neuropsychiatric disorders including ASD, schizophrenia and mental retardation (7, 98). The reasons for this are as yet unclear, but may be due to pleiotropic effects of the underlying genes and pathways, as well as shared effects of genetic background, or as yet unrecognized environmental factors that are shared in common between these conditions.

Rare CNVs have also been associated with other neurological disorders including epilepsy, Parkinson's disease and ALS. For example, a microdeletion on chromosome 15q13.3, previously associated with several neuropsychiatric features (92, 93), has recently been reported in 1% of cases

of idiopathic generalized epilepsy, making this the most prevalent genetic risk factor for this condition reported to date (94). Rare CNVs in the alpha-synuclein (*SNCA*) locus were shown to underlie autosomal dominant Parkinson's disease in a small number of families (99–101). In ALS, two recent studies have reported an excess of recurrent CNV loci and replicated gene dosage alterations, some of which were specific to ALS (102, 103), while an earlier study showed an overrepresentation in ALS patients of homozygous deletions in the *SMN2* gene that is known to modify the severity of spinal muscular atrophy (104). These data suggest that multiple, often rare CNVs play an important part in the pathogenesis of these neurological disorders.

CNV and cancer

Cancer is caused by dysregulation of the expression and activity of genes that control cell growth and differentiation, leading to abnormal cell proliferation (105). These effects are often mediated by germline or somatic mutations in oncogenes and tumour suppressor genes. With increasing use of whole-genome technologies to detect CNVs, germline and somatic CNVs are now recognized as frequent contributors to the spectrum of mutations leading to cancer development.

Genome-wide analyses with high-resolution SNP arrays have started to define the extent of somatic CNVs in cancer genomes. In acute lymphoblastic leukaemia (ALL), analysis of leukaemia cells for 242 paediatric ALL patients showed structural rearrangements in genes encoding principal regulators of B-lymphocyte development and differentiation in 40% of cases (44). In these patients, 54 recurrent somatic regions of deletion were identified that were not present in matched germline samples. Many of these deletions created fusion proteins in known oncogenes or led to other specific, pathogenic mutations. Copy number changes in *PAX5*, a known gene on the B-cell development pathway, were found in 57 of 192 B-progenitor ALL cases.

Somatic CNVs have also been detected as frequent pathogenic events in solid tumours. In adenocarcinoma of the lung, a total of 57 recurrent copy number changes were detected in a collection of 528 cases (43). Amongst 206 cases of glioblastoma, somatic copy number alterations were also frequent, and concurrent gene expression analysis showed that 76% of genes affected by copy number alteration had expression patterns that correlated with gene copy number (45). Similar reports of somatic CNVs have been reported in prostate cancer (106), breast cancer and colorectal cancer (107). CGH analysis of 64 prostate cancer patients showed numerous recurrent genomic copy number aberrations associated with the disease (106). Specific loss at 8p23.2 was associated with pathological advanced stage disease, and gain at 11q13.1 was found to be predictive of post-operative recurrence. Moreover, the comparison between an independent set of tumours that ultimately metastasized and organ metastases *vs* non-progressors identified 40 genomic aberrations more frequent in primary tumours that metastasized (20–45%) and organ metastases (20–90%), rather then primary tumours that did not progress (0–20%). These structural variants represented strong candidate markers associated with metastatic potential. High-resolution analyses of copy number and nucleotide alterations have been carried out on breast and colorectal cancer (107). Individual colorectal and breast tumours had, on average, 7 and 18 copy number alterations, respectively, with 24 and 9 as the average number of protein coding genes affected by amplification or homozygous deletions. Potential driver genes were considered as those altered in both tumour types and showed at least one amplification or homozygous deletion. The analyses also identified several gene pathways, including the ERBB2, epidermal growth factor receptor (EGFR) and PI3K pathways, that were affected by copy number changes and point alterations in both cancers, thereby pointing to potentially useful targets for diagnosis and therapy.

The correlation between somatic CNV and gene expression in these studies, particularly in certain kinase genes that are targets of anti-cancer drugs, suggests that the analysis of cancer genome CNVs may be useful in informing therapeutic decisions on management of individual patients with particular patterns of mutations (45).

With regard to heritable germline CNVs and cancer, a heritable CNV at chromosome 1q21.1 was found to be associated with neuroblastoma, and a new neuroblastoma breakpoint family gene, *NBPF23*, whose expression correlated with *NBPF23* copy number was implicated in neuroblastoma tumorigenesis (108). Similarly, genome-wide CNV analysis revealed a germline deletion at chromosome 2p24.3 that was more common in prostate cancer patients than in controls, with apparently stronger risk in patients with aggressive compared with non-aggressive prostate cancer (109).

Such findings suggest that CNVs could contribute to increased DNA instability and occurrence of genomic imbalance. This hypothesis was strengthened by Shlien et al. (110) who

investigated the frequency of germline CNVs in families with Li–Fraumeni syndrome, an autosomal dominant disorder characterized by a high frequency of various early-onset cancers in individuals harbouring germline mutations in the *TP53* gene. Shlien et al. found enrichment in germline CNVs in the cancer-prone individuals' family members and, in particular, a significantly higher frequency of CNVs amongst carriers of *TP53* mutations with a family history of cancer.

How CNVs, either somatic or germline, contribute to cancer development is still poorly understood. Possible explanations came from the Knudson's two hit hypothesis (111): tumour suppressor genes can be lost as a consequence of a homozygous deletion leading directly to cancer susceptibility. Alternatively, heterozygous deletions may harbour tumour candidate genes that become unmasked when a functional mutation arises in the other chromosome resulting in tumorigenesis. With respect to duplications, gains of chromosomal regions may result in the increase of gene expression levels of one or more oncogenes. Germline CNVs can provide a genetic basis for subsequent somatic chromosomal changes that arise in tumour DNA. It has been proposed that CNVs are more abundant in cancer-prone individuals because of germline defects in DNA repair processes (24, 110).

CNV and metabolic or cardiovascular diseases

The association between metabolic and cardiovascular traits, such as familial hypercholesterolemia (FH), atherosclerosis and coronary artery disease, and CNVs has been reported in a number of studies (112–114). Moreover, the overlap between CNV regions and loci that have been previously associated with cardiovascular traits suggested a potential role for CNVs in the development of such diseases.

For instance, the MLPA method led to the identification of germline rearrangements, mainly deletions, at the *LDLR* gene in around half of FH patients who carry no other *LDLR* mutations (113). Another example is the *LPA* gene that encodes apolipoprotein(a) (apo(a)) whose plasma concentration is associated with risk of coronary artery disease, myocardial infarction (MI) and stroke. Plasma concentrations of the apo(a) containing lipoprotein Lp(a) vary over 1000-fold between individuals and 90% of this variability is genetically determined at the *LPA* locus, with plasma concentration and Lp(a) isoform size being proportional to a highly variable number of 'kringle 4' repeat sequences (range 5–50) (112). Although these data indicate that CNV in at least two genes can be associated with cardiovascular risk, a recent genome-wide association study for MI in 3000 cases and 3000 controls, that found several SNPs associated with MI, did not find any CNVs, either rare or common, that were associated with MI (115). Further large studies searching specifically for CNV associations with cardiovascular disorders are likely to be required to demonstrate definitively the role of CNVs in the development of these common diseases.

Conclusions and final remarks

The study of CNVs in human disease has evolved rapidly since discovery of the widespread existence of these structural variants in human genomes. Many lines of evidence have now shown that rare, common and *de novo* CNVs play an important pathogenic role in a range of human disorders, from causative high penetrance CNVs in rare genomic disorders to intermediate or low penetrance CNVs in complex multifactorial diseases. The role of germline CNVs in the pathogenesis of autoimmune, inflammatory, infectious and neuropsychiatric disorders appears particularly well established, as does the role of somatic CNVs in cancer. However, the mechanisms through which most common transmissible CNVs confer increased disease susceptibility remain largely unknown.

Many genomic disorders have highly variable penetrance, which may be due to the interaction between structural variants and additional genetic or environmental risk factors, including individual genetic background (116, 117). In addition, somatic mosaicism for CNV in normal individuals has also been reported (118, 119), raising the possibility that certain common human phenotypes other than cancer can arise as a consequence of CNVs that develop in differentiated tissues.

Recent advances in technology have provided powerful tools for detection and analysis of CNVs at the level of the genome as well as for targeted loci. It is clear that copy number variation has a profound effect on inter-individual variation in gene expression, but persistent technical challenges in accurate measurement of copy number mean that the extent of the contribution of common CNVs to the molecular basis of common diseases remains to be established (46). Further studies, using accurate genotyping assays in large population cohorts, will be able to define more precisely the overall role of CNVs in the pathogenesis of common diseases.

Acknowledgements

We thank Angharad Roberts, Richard P. Hull and Ellen Thomas for careful reading of the manuscript.

References

1. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. J Clin Invest 2008: 118: 1590–1605.
2. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science 2008: 322: 881–888.
3. McCarthy MI, Abecasis GR, Cardon LR et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 2008: 9: 356–369.
4. Ghanem N, Uring-Lambert B, Abbal M et al. Polymorphism of MHC class III genes: definition of restriction fragment linkage groups and evidence for frequent deletions and duplications. Hum Genet 1988: 79: 209–218.
5. Freeman JL, Perry GH, Feuk L et al. Copy number variation: new insights in genome diversity. Genome Res 2006: 16: 949–961.
6. Goossens M, Dozy AM, Embury SH et al. Triplicated alpha-globin loci in humans. Proc Natl Acad Sci U S A 1980: 77: 518–521.
7. Cook EH Jr, Scherer SW. Copy-number variations associated with neuropsychiatric conditions. Nature 2008: 455: 919–923.
8. Zhang X, Snijders A, Segraves R et al. High-resolution mapping of genotype-phenotype relationships in cri du chat syndrome using array comparative genomic hybridization. Am J Hum Genet 2005: 76: 312–326.
9. Schadt EE, Lamb J, Yang X et al. An integrative genomics approach to infer causal associations between gene expression and disease. Nat Genet 2005: 37: 710–717.
10. Campbell L, Potter A, Ignatius J et al. Genomic variation and gene conversion in spinal muscular atrophy: implications for disease process and clinical phenotype. Am J Hum Genet 1997: 61: 40–50.
11. Scherer SW, Cheung J, MacDonald JR et al. Human chromosome 7: DNA sequence and biology. Science 2003: 300: 767–772.
12. Vissers LE, Veltman JA, van Kessel AG et al. Identification of disease genes by whole genome CGH arrays. Hum Mol Genet 2005: 14 Spec No. 2: R215–R223.
13. Senderek J, Bergmann C, Weber S et al. Mutation of the SBF2 gene, encoding a novel member of the myotubularin family, in Charcot-Marie-Tooth neuropathy type 4B2/11p15. [erratum appears in Hum Mol Genet. 2004Feb 1;13(3):363]. Hum Mol Genet 2003: 12: 349–356.
14. Horsthemke B, Wagstaff J. Mechanisms of imprinting of the Prader-Willi/Angelman region. Am J Med Genet Part A 2008: 146A: 2041–2052.
15. Lupski JR. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet 1998: 14: 417–422.
16. Lupski JR. Charcot-Marie-Tooth disease: lessons in genetic mechanisms. Mol Med 1998: 4: 3–11.
17. Trovo-Marqui AB, Tajara EH. Neurofibromin: a general outlook. Clin Genet 2006: 70: 1–13.
18. Toyo-oka K, Shionoya A, Gambello MJ et al. 14-3-3epsilon is important for neuronal migration by binding to NUDEL: a molecular explanation for Miller-Dieker syndrome. Nat Genet 2003: 34: 274–285.
19. Carlson C, Sirotkin H, Pandita R et al. Molecular definition of 22q11 deletions in 151 velo-cardio-facial syndrome patients. Am J Hum Genet 1997: 61: 620–629.
20. Edelmann L, Pandita RK, Spiteri E et al. A common molecular basis for rearrangement disorders on chromosome 22q11.[see comment]. Hum Mol Genet 1999: 8: 1157–1167.
21. Woodward KJ, Cundall M, Sperle K et al. Heterogeneous duplications in patients with Pelizaeus-Merzbacher disease suggest a mechanism of coupled homologous and nonhomologous recombination. Am J Hum Genet 2005: 77: 966–987.
22. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders.[see comment]. Cell 2007: 131: 1235–1247.
23. Lupski JR. Genomic disorders ten years on. Genome Med 2009: 1: 4.
24. Hastings PJ. Mechanisms of change in gene copy number. Nat Rev Genet 2009: 10: 14.
25. Iafrate AJ, Feuk L, Rivera MN et al. Detection of large-scale variation in the human genome.[see comment]. Nat Genet 2004: 36: 949–951.
26. Sebat J, Lakshmi B, Troge J et al. Large-scale copy number polymorphism in the human genome. Science 2004: 305: 525–528.
27. Fredman D, White SJ, Potter S et al. Complex SNP-related sequence variation in segmental genome duplications. Nat Genet 2004: 36: 861–866.
28. Feuk L, Marshall CR, Wintle RF et al. Structural variants: changing the landscape of chromosomes and design of disease studies. Hum Mol Genet 2006: 15 Spec No 1: R57–R66.
29. Redon R, Ishikawa S, Fitch KR et al. Global variation in copy number in the human genome. Nature 2006: 444: 444–454.
30. Jakobsson M, Scholz SW, Scheet P et al. Genotype, haplotype and copy-number variation in worldwide human populations. Nature 2008: 451: 998–1003.
31. 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.
32. Henrichsen CN, Vinckenbosch N, Zollner S et al. Segmental copy number variation shapes tissue transcriptomes. Nat Genet 2009: 41: 424–429.
33. Cahan P, Li Y, Izumi M et al. The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells. Nat Genet 2009: 41: 430–437.
34. Zhang J, Feuk L, Duggan GE et al. Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. Cytogenet Genome Res 2006: 115: 205–214.
35. Firth HV, Richards SM, Bevan AP et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet 2009: 84: 524–533.
36. Shaikh TH, Gai X, Perin JC et al. High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. Genome Res 2009: 19: 1682–1690.
37. Gonzalez E, Kulkarni H, Bolivar H et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility.[see comment]. Science 2005: 307: 1434–1440.
38. Aitman TJ, Dong R, Vyse TJ et al. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans.[see comment]. Nature 2006: 439: 851–855.
39. Fanciulli M, Norsworthy PJ, Petretto E et al. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. Nat Genet 2007: 39: 721–723.

40. Willcocks LC, Lyons PA, Clatworthy MR et al. Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. J Exp Med 2008: 205: 1573–1582.
41. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet 2006: 7: 85–97.
42. van Ommen GJ. Frequency of new copy number variation in humans. Nat Genet 2005: 37: 333–334.
43. Weir BA, Woo MS, Getz G et al. Characterizing the cancer genome in lung adenocarcinoma. Nature 2007: 450: 893–898.
44. Mullighan CG, Goorha S, Radtke I et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia.[see comment]. Nature 2007: 446: 758–764.
45. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008: 455: 1061–1068.
46. McCarroll SA. Extending genome-wide association studies to copy-number variation. Hum Mol Genet 2008: 17: R135–R142.
47. Sharp AJ. Emerging themes and new challenges in defining the role of structural variation in human disease. Hum Mutat 2009: 30: 135–144.
48. Carter NP. Methods and strategies for analyzing copy number variation using DNA microarrays. Nat Genet 2007: 39: S16–S21.
49. McCarroll SA. Copy-number analysis goes more than skin deep.[comment]. Nat Genet 2008: 40: 5–6.
50. Armour JA, Palla R, Zeeuwen PL et al. Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. Nucleic Acids Res 2007: 35: e19.
51. Hollox EJ, Huffmeier U, Zeeuwen PL et al. Psoriasis is associated with increased beta-defensin genomic copy number.[see comment]. Nat Genet 2008: 40: 23–25.
52. Sellner LN, Taylor GR. MLPA and MAPH: new techniques for detection of gene deletions. Hum Mutat 2004: 23: 413–419.
53. Khaja R, Zhang J, MacDonald JR et al. Genome assembly comparison identifies structural variants in the human genome. Nat Genet 2006: 38: 1413–1418.
54. Levy S, Sutton G, Ng PC et al. The diploid genome sequence of an individual human. PLoS Biol 2007: 5: e254.
55. Wheeler DA, Srinivasan M, Egholm M et al. The complete genome of an individual by massively parallel DNA sequencing.[see comment]. Nature 2008: 452: 872–876.
56. Margulies M, Egholm M, Altman WE et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005: 437: 376–380.
57. Bentley DR. Whole-genome re-sequencing. Curr Opin Genet Dev 2006: 16: 545–552.
58. Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 2008: 9: 387–402.
59. Korbel JO, Urban AE, Affourtit JP et al. Paired-end mapping reveals extensive structural variation in the human genome. Science 2007: 318: 420–426.
60. Tuzun E, Sharp AJ, Bailey JA et al. Fine-scale structural variation of the human genome. Nat Genet 2005: 37: 727–732.
61. Yoon S, Xuan Z, Makarov V et al. Sensitive and accurate detection of copy number variants using read depth of coverage. Genome Res 2009: 19: 1586–1592.
62. Vermeesch JR, Melotte C, Froyen G et al. Molecular karyotyping: array CGH quality criteria for constitutional genetic diagnosis. J Histochem Cytochem 2005: 53: 413–422.
63. Olshen AB, Venkatraman ES, Lucito R et al. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 2004: 5: 557–572.
64. Price TS, Regan R, Mott R et al. SW-ARRAY: a dynamic programming solution for the identification of copy-number changes in genomic DNA using array comparative genome hybridization data. Nucleic Acids Res 2005: 33: 3455–3464.
65. Fiegler H, Redon R, Andrews D et al. Accurate and reliable high-throughput detection of copy number variation in the human genome. Genome Res 2006: 16: 1566–1574.
66. Korn JM, Kuruvilla FG, McCarroll SA et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat Genet 2008: 40: 1253–1260.
67. Conrad DF, Andrews TD, Carter NP et al. A high-resolution survey of deletion polymorphism in the human genome. Nat Genet 2006: 38: 75–81.
68. Barnes C, Plagnol V, Fitzgerald T et al. A robust statistical method for case-control association testing with copy number variation. Nat Genet 2008: 40: 1245–1252.
69. 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.
70. Clayton DG, Walker NM, Smyth DJ et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat Genet 2005: 37: 1243–1246.
71. Cooper GM, Nickerson DA, Eichler EE. Mutational and selective effects on copy-number variants in the human genome. Nat Genet 2007: 39: S22–S29.
72. Burns JC, Shimizu C, Gonzalez E et al. Genetic variations in the receptor-ligand pair CCR5 and CCL3L1 are important determinants of susceptibility to Kawasaki disease. J Infect Dis 2005: 192: 344–349.
73. Nakajima T, Kaur G, Mehra N et al. HIV-1/AIDS susceptibility and copy number variation in CCL3L1, a gene encoding a natural ligand for HIV-1 co-receptor CCR5. Cytogenet Genome Res 2008: 123: 156–160.
74. 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.
75. Fanciulli M, Vyse TJ, Aitman TJ. Copy number variation of Fc gamma receptor genes and disease predisposition. Cytogenet Genome Res 2008: 123: 161–168.
76. Hollox EJ, Detering JC, Dehnugara T. An integrated approach for measuring copy number variation at the FCGR3 (CD16) locus. Hum Mutat 2009: 30: 477–484.
77. McCarroll SA, Huett A, Kuballa P et al. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. Nat Genet 2008: 40: 1107–1112.
78. Yang Y, Chung EK, Wu YL 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 2007: 80: 1037–1054.
79. Piirila P, Wikman H, Luukkonen R et al. Glutathione S-transferase genotypes and allergic responses to diisocyanate exposure. Pharmacogenetics 2001: 11: 437–445.
80. Ivaschenko TE, Sideleva OG, Baranov VS. Glutathione- S-transferase micro and theta gene polymorphisms as new risk factors of atopic bronchial asthma. J Mol Med 2002: 80: 39–43.
81. Brasch-Andersen C, Christiansen L, Tan Q et al. Possible gene dosage effect of glutathione-S-transferases on atopic asthma: using real-time PCR for quantification of GSTM1

and GSTT1 gene copy numbers. Hum Mutat 2004: 24: 208–214.
82. Fryer AA, Bianco A, Hepple M et al. Polymorphism at the glutathione S-transferase GSTP1 locus. A new marker for bronchial hyperresponsiveness and asthma. Am J Respir Crit Care Med 2000: 161: 1437–1442.
83. Lee JA, Lupski JR. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. Neuron 2006: 52: 103–121.
84. Gu W, Lupski JR. CNV and nervous system diseases–what's new? Cytogenet Genome Res 2008: 123: 54–64.
85. Sebat J, Lakshmi B, Malhotra D et al. Strong association of de novo copy number mutations with autism. Science 2007: 316: 445–449.
86. Autism Genome Project Consortium, Szatmari P, Paterson AD et al. Mapping autism risk loci using genetic linkage and chromosomal rearrangements.[erratum appears in Nat Genet. 2007 Oct;39(10):1285 Note: Meyer, Kacie J [added]; Koop, Frederike [corrected to Koop, Frederieke]; Langemeijer, Marjolijn [corrected to Langemeijer, Marjolein]; Hijimans, Channa [corrected to Hijmans, Channa]]. Nat Genet 2007: 39: 319–328.
87. Marshall CR, Noor A, Vincent JB et al. Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet 2008: 82: 477–488.
88. Weiss LA, Shen Y, Korn JM et al. Association between microdeletion and microduplication at 16p11.2 and autism. [see comment]. N Engl J Med 2008: 358: 667–675.
89. Kumar RA, KaraMohamed S, Sudi J et al. Recurrent 16p11.2 microdeletions in autism. Hum Mol Genet 2008: 17: 628–638.
90. Walsh T, McClellan JM, McCarthy SE et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia.[see comment]. Science 2008: 320: 539–543.
91. Xu B, Roos JL, Levy S et al. Strong association of de novo copy number mutations with sporadic schizophrenia. Nat Genet 2008: 40: 880–885.
92. International Schizophrenia Consortium Rare chromosomal deletions and duplications increase risk of schizophrenia.[see comment]. Nature 2008: 455: 237–241.
93. Stefansson H, Rujescu D, Cichon S et al. Large recurrent microdeletions associated with schizophrenia.[see comment]. Nature 2008: 455: 232–236.
94. Helbig I, Mefford HC, Sharp AJ et al. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. Nat Genet 2009: 41: 160–162.
95. Rovelet-Lecrux A, Hannequin D, Raux G et al. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy.[see comment]. Nat Genet 2006: 38: 24–26.
96. Veenstra-Vanderweele J, Christian SL, Cook EH Jr. Autism as a paradigmatic complex genetic disorder. Annu Review Genomics Hum Genet 2004: 5: 379–405.
97. Vorstman JA, Staal WG, van Daalen E et al. Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. Mol Psychiatry 2006: 11: 1.
98. Mefford HC, Sharp AJ, Baker C et al. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes.[see comment]. N Eng J Med 2008: 359: 1685–1699.
99. Singleton AB, Farrer M, Johnson J et al. Alpha-synuclein locus triplication causes Parkinson's disease. Science 2003: 302: 841.
100. Ibanez P, Bonnet AM, Debarges B et al. Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. [see comment]. Lancet 2004: 364: 1169–1171.
101. Ross OA, Braithwaite AT, Skipper LM et al. Genomic investigation of alpha-synuclein multiplication and parkinsonism. Ann Neurol 2008: 63: 743–750.
102. Blauw HM, Veldink JH, van Es MA et al. Copy-number variation in sporadic amyotrophic lateral sclerosis: a genome-wide screen.[see comment]. Lancet Neurol 2008: 7: 319–326.
103. Cronin S, Blauw HM, Veldink JH et al. Analysis of genome-wide copy number variation in Irish and Dutch ALS populations. Hum Mol Genet 2008: 17: 3392–3398.
104. Veldink JH, van den Berg LH, Cobben JM et al. Homozygous deletion of the survival motor neuron 2 gene is a prognostic factor in sporadic ALS. Neurology 2001: 56: 749–752.
105. Weir B, Zhao X, Meyerson M. Somatic alterations in the human cancer genome. Cancer Cell 2004: 6: 433–438.
106. Paris PL, Andaya A, Fridlyand J et al. Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors. Hum Mol Genet 2004: 13: 1303–1313.
107. Leary RJ, Lin JC, Cummins J et al. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. Proc Natl Acad Sci U S A 2008: 105: 16224–16229.
108. Diskin SJ. Copy number variation at 1q21.1 associated with neuroblastoma. Nature 2009: 459: 987–991.
109. Liu W, Sun J, Li G et al. Association of a germ-line copy number variation at 2p24.3 and risk for aggressive prostate cancer. Cancer Res 2009: 69: 2176–2179.
110. Shlien A, Tabori U, Marshall CR et al. Excessive genomic DNA copy number variation in the Li-Fraumeni cancer predisposition syndrome. Proc Natl Acad Sci U S A 2008: 105: 11264–11269.
111. Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 1971: 68: 820–823.
112. Lanktree M, Hegele RA. Copy number variation in metabolic phenotypes. Cytogenet Genome Res 2008: 123: 169–175.
113. Wang J, Ban MR, Hegele RA. Multiplex ligation-dependent probe amplification of LDLR enhances molecular diagnosis of familial hypercholesterolemia. J Lipid Res 2005: 46: 366–372.
114. Tosi I, Toledo-Leiva P, Neuwirth C et al. Genetic defects causing familial hypercholesterolaemia: identification of deletions and duplications in the LDL-receptor gene and summary of all mutations found in patients attending the Hammersmith Hospital Lipid Clinic. Atherosclerosis 2007: 194: 102–111.
115. Myocardial Infarction Genetics Consortium, Kathiresan S, Voight BF et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet 2009: 41: 334–341.
116. Taddei I, Morishima M, Huynh T et al. Genetic factors are major determinants of phenotypic variability in a mouse model of the DiGeorge/del22q11 syndromes. Proc Natl Acad Sci U S A 2001: 98: 11428–11431.
117. Klopocki E, Schulze H, Strauss G et al. Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome. Am J Hum Genet 2007: 80: 232–240.
118. Piotrowski A, Bruder CE, Andersson R et al. Somatic mosaicism for copy number variation in differentiated human tissues. Hum Mutat 2008: 29: 1118–1124.
119. Bruder CE, Piotrowski A, Gijsbers AA et al. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. Am J Hum Genet 2008: 82: 763–771.