The clinical context of copy number variation in the human genome

Charles Lee[1] and Stephen W. Scherer[2,3,*]

During the past five years, copy number variation (CNV) has emerged as a highly prevalent form of genomic variation, bridging the interval between long-recognised microscopic chromosomal alterations and single-nucleotide changes. These genomic segmental differences among humans reflect the dynamic nature of genomes, and account for both normal variations among us and variations that predispose to conditions of medical consequence. Here, we place CNVs into their historical and medical contexts, focusing on how these variations can be recognised, documented, characterised and interpreted in clinical diagnostics. We also discuss how they can cause disease or influence adaptation to an environment. Various clinical exemplars are drawn out to illustrate salient characteristics and residual enigmas of CNVs, particularly the complexity of the data and information associated with CNVs relative to that of single-nucleotide variation. The potential is immense for CNVs to explain and predict disorders and traits that have long resisted understanding. However, creative solutions are needed to manage the sudden and overwhelming burden of expectation for laboratories and clinicians to assay and interpret these complex genomic variations as awareness permeates medical practice. Challenges remain for understanding the relationship between genomic changes and the phenotypes that might be predicted and prevented by such knowledge.

It is now about 50 years since the first recognition of a microscopic human copy number variation (CNV) – trisomy 21 (Ref. 1) – and five years since the first reports of the widespread prevalence of submicroscopic CNVs (Refs 2, 3) (Table 1). Classical genetics was based on the premise that all genes come in pairs, but, in the interval between these two milestones, evidence gradually accumulated to discount this dogma. The earliest examples – trisomy 21, monosomy

[1]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.

[2]The Centre for Applied Genomics and Program in Genetics & Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada.

[3]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada.

*Corresponding author: Stephen W. Scherer, The Hospital for Sick Children, MaRS Centre - East Tower, 101 College Street, Room 14-701, Toronto, Ontario, M5G 1L7, Canada. E-mail: stephen.scherer@sickkids.ca

The clinical context of copy number variation in the human genome

1

Table 1. History and milestones in human copy number variation research

Year	New technology or discovery milestone	Refs
1959	Down syndrome: trisomy 21	1
	Turner syndrome (45,X)	4
	Kleinfelter syndrome (47,XXY)	5
1961	Partial trisomy Down syndrome	159
1963	First inherited deletion syndrome: Cri-du-chat (5p⁻)	160
1969	Chromosome banding: detection of subchromosomal anomalies	161
1978	Duplication of α-globin genes proven by Southern blot	162
1980	First DNA polymorphism due to variable number tandem repeat (VNTR)	163
1980	Fluorescence in situ hybridisation (FISH) technique used for detection of cryptic/ submicroscopic anomalies	164
1982	FISH applied to human chromosomes	165
1985	Minisatellite probes for DNA fingerprinting	166
1985	Dystrophin gene cloning identified structural variants, including intragenic deletions and translocation	167, 168
1986	Contiguous gene syndromes	169
1991	1.4 Mb duplication associated with Charcot–Marie–Tooth disease type 1A identified and suggested to be caused by nonallelic homologous recombination (NAHR)	23
1992	Comparative genomic hybridisation (CGH) technique made possible detection of gains and losses in tumours	170
1996	Complete set of human telomere-specific FISH probes created for clinical detection of variation at human subtelomeric regions	171
1998	Array CGH using bacterial artificial chromosome clones	172
2001	Human Genome Project first reference sequence	42, 44
2004	Human Genome Project 'finished'	43
2004	Recognition of human genomic variation on large scale	2, 3
2006	Copy number variation map for the human genome	16
2007	First diploid sequence of an individual human	40

X (Ref. 4), and XXY (Ref. 5) – had clear clinical consequences (Down, Turner and Kleinfelter syndromes, respectively), but the remarkable revelation associated with the submicroscopic CNVs has been their ubiquity throughout and among all genomes, not just those that come to medical attention. The past five years have yielded rapid developments in technology and analysis, creating a field of investigation that is transforming both our concept of the human genome and the application to clinical practice. CNVs are integral to the full spectrum of human variation and its relationship to health and disease.

Definition and scope of CNV: five years later

After the seminal reports of 2004 (Refs 2, 3), the abbreviation CNV was first formalised by Feuk et al. (Ref. 6), who defined it operationally as 'a segment of DNA that is 1 kb or larger and is present at a variable copy number in comparison with a reference genome' (Box 1). The umbrella classification group of genomic structural variation includes CNVs as well as segments that involve no loss or gain of material but are rearranged relative to a reference (i.e. inversions or balanced translocations). Although all are biologically important and can impact phenotypes, we limit the focus of this review to matters of CNVs. Discussion has persisted as to use of 'variant' in this context. Notwithstanding precedents from cytogenetics and single-nucleotide terminology, our increasing awareness of the inconsistent associations between CNVs and phenotypes reinforces recommendations (Refs 7, 8) concerning nomenclature: to use 'variant' in a generic sense without inherent implications as to pathogenicity, frequency or other characteristics. It seems pragmatic to retain a term without excess denotation, and not attempt a priori to suggest

The clinical context of copy number variation in the human genome

2

Box 1. Terminology

Structural variation/variant
This is the umbrella term to encompass a group of microscopic or submicroscopic genomic alterations involving segments of DNA. We use the term as a neutral descriptor with nothing implied about frequency, association with disease or phenotype, or lack thereof. The structural variation may be quantitative (copy number variants comprising deletions, insertions and duplications) and/or positional (translocations) or orientational (inversions).

Copy number variation/variant (CNV)
CNV refers to DNA segments for which copy number differences have been observed in the comparison of two or more genomes. Without further annotation, CNV carries no implication of relative frequency or phenotypic effect. These quantitative structural variants can be genomic copy number gains (insertions or duplications) or losses (deletions or null genotypes) relative to a designated reference genome sequence.

Insertion/deletion (indel)
Indel is a collective abbreviation to describe relative gain or loss of a segment of one or more nucleotides in a genomic sequence. It allows the designation of a difference between genomes in situations where the direction of sequence change cannot be inferred: for example, when a reference or ancestral sequence has not been defined. It has typically been used to denote relatively small-scale variants (particularly those <1 kb); however, we do not propose any size restriction for its use.

Segmental duplication
This is a segment of DNA >1 kb in size that occurs in two or more copies per haploid genome, with the different copies sharing $>90\%$ sequence identity. These segments can also be CNVs. The duplicated blocks predispose to nonallelic homologous recombination.

Human genome reference assembly
The standard reference DNA sequence (or assembly) of the human genome. The assembly is derived mostly ($>60\%$) of DNA from a single donor, with the rest of the sequence originating from a mosaic of other sources. The current assembly covers most of the euchromatic regions of the human genome.

Single-nucleotide polymorphism (SNP)
A variation in DNA that involves replacement of one nucleotide base for another is called a SNP. Polymorphism implies that the variant (minor) allele has a frequency of at least 1%; however, terminology has come to be applied more loosely by some, to include even rare mutations.

Syndrome
Literally 'running together', syndrome describes a collection of features or symptoms (typically comprising three or more clinical findings), the constellation of which is recognisable as a specified disorder.

Relative risk and odds ratio
Relative risk (RR) and odds ratio (OR) are similar in that they both determine the likelihood that a member of one group (individuals with a CNV) will develop a phenotype, relative to the likelihood that a member of another group (individuals without a CNV) will develop that same phenotype. For RR, this likelihood is measured using probability; for OR, it is measured using odds. With such metrics, researchers have discovered CNVs within clinical cohorts that are risk factors for distinct phenotypes.

---

that it is anything more than an observation of difference. Issues of pathogenicity or polymorphism can be addressed with modifiers (easily adapted as information arises).

However, we suggest that the size component of the CNV definition be reconsidered, and perhaps simply dropped. The initial limitation to segments of at least 1 kb was perhaps more a reflection of the technologies first used to reveal this class of variation [especially array comparative genome hybridisation (aCGH) with bacterial artificial chromosome (BAC) probes] than of a biological or functional threshold. Clearly, most quantitative variation in the genome involves segments smaller than 1 kb (Refs 9, 10, 11, 12, 13). Thus, although 'indels' and di- and tri-nucleotide repeats are not technically CNVs according to the original definition, this term is now often used as a 'catch-all' that encompasses all non-SNP

3

(single-nucleotide polymorphism) unbalanced variation in the genome. Thus, for some of the same reasons discussed above, we suggest that CNV be used in a less restrictive sense and be classified when needed (e.g. CNVs greater than 1 kb).

The microscopically visible CNVs at the larger end of the spectrum (1 Mb or larger) are almost invariably associated with phenotypic consequences that are likely to bring an individual to medical attention. As we move down the size spectrum, some genomic variation is of striking clinical effect, but much contributes to what we understand as normal phenotypic variation: that which simply makes individual humans different from one another (Refs 14, 15). Some CNVs are also likely to be entirely inconsequential (Refs 13, 16). As with variation at the level of the individual nucleotide, many of these CNVs provide the species with a reservoir of adaptive potential for changing environmental circumstances. CNVs, therefore, are involved in every aspect of the phenotype, and whether or not a given CNV is of clinical consequence may be a function of time, place and other factors. While acknowledging this breadth of influence of CNVs, we limit the discussion for this review to examples that are likely to be relevant to medical practice now and in the near future.

CNVs have become the genomic bridge to meld disciplines of molecular genetics and cytogenetics (Table 1). The light microscope revealed the first gains and losses, starting with whole-chromosome aneuploidies, and then partial chromosome changes large enough to be obvious with solid staining. By the mid 1970s, the more indirect tools of molecular genetics, such as Southern blot hybridisation, began to expose quantitative DNA changes from the small end of the spectrum. Later, the hybridisation of molecular probes to human chromosomes, particularly with fluorescence in situ hybridisation (FISH), provided a potent tool for detection of subtle segmental deletions, duplications and rearrangements. With this and, in tandem, the DNA sequencing efforts of the Human Genome Project, segmental changes started to be recognised as a basis for many mendelian disorders as well as contiguous gene syndromes (Ref. 17). The emergence of array technologies, particularly

aCGH, facilitated widespread efficient scanning of the genome for quantitative changes in a size range that had not previously been accessible. In 2003–2004, a few studies started to observe complex CNV and structural variations at multiple loci (Refs 18, 19, 20); however, the Human Genome Project's strong message of 99.9% human sequence identity between two unrelated healthy individuals, with most variation encompassed by SNPs, nonetheless prevailed.

By 2004, it was apparent that CNVs are not just a cause of disease, but are ubiquitous among human genomes and an important aspect of human variation (Refs 2, 3). Despite the profound logistical challenges associated with studying these complex genomic features, progress has been swift. As whole-genome sequences are becoming available for comparison, we foresee greater opportunity for fruitful analyses and applications in personalised medical care.

Forms of CNVs

The gain or loss of genomic material is recognised by comparison of reference and sample genomes through hybridisation or sequence analysis, and is described in relation to the reference (Fig. 1). Simple CNVs take the form of deletions, or tandem or insertional duplications. Sites at which a greater degree of replication has evolved allow a greater variety of copy number alleles among haploid genomes, with the potential for incremental variation in the related individual phenotypes. Many CNVs, however, show highly complex rearrangement of a genomic region, reflecting a history of steps in their generation, sometimes with both gain and loss of material. CNVs may involve whole genes, portions of genes, multiples of contiguous genes, regulatory elements, or none of the above, and the nature and extent of material that is deleted or duplicated is undoubtedly important for the phenotypic consequences.

The nearby genomic sequence may yield clues as to how the CNV was generated (Ref. 21). Often, a CNV is flanked by nearly identical blocks of sequence, called segmental duplications or low-copy repeats, or by Alu or LINE repetitive elements, which have created the opportunity for misalignment of DNA strands during recombination. This process of nonallelic homologous recombination (NAHR) (Ref. 22)
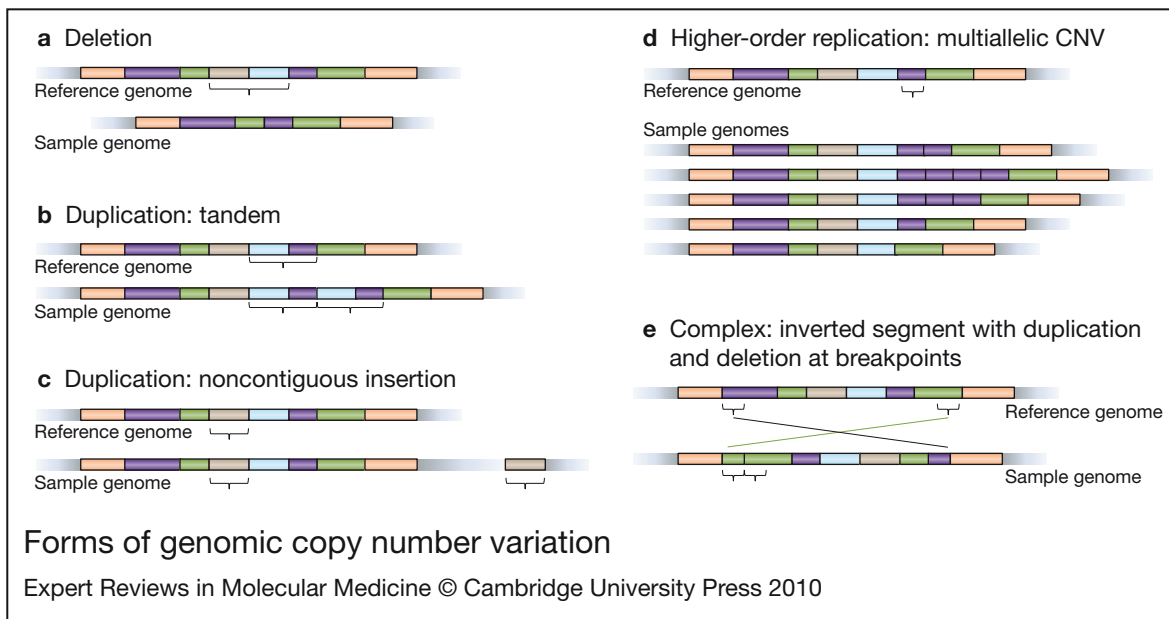
The clinical context of copy number variation in the human genome

4

a Deletion

Reference genome

Sample genome

b Duplication: tandem

Reference genome

Sample genome

c Duplication: noncontiguous insertion

Reference genome

Sample genome

d Higher-order replication: multiallelic CNV

Reference genome

Sample genomes

e Complex: inverted segment with duplication and deletion at breakpoints

Reference genome

Sample genome

Forms of genomic copy number variation

Expert Reviews in Molecular Medicine © Cambridge University Press 2010

Figure 1. Forms of genomic copy number variation. Variations in sample genomes are depicted relative to a reference genome. Colours represent different segments of DNA, such that segments of the same colour contain identical sequences. Schematics show (a) deletion, or loss, of sequence (brown and blue segments) as well as (b, c) duplications of DNA segments. Duplications can be either (b) tandem, where segments (blue and purple) are duplicated into the adjacent sequence, or (c) noncontiguous, where segments (brown) can be duplicated distantly from the original sequence, even on another chromosome. The figure also shows schematics of more complicated variation, including (d) higher-order replication, where a segment (purple) can be duplicated several times and exist in multiple alleles, and (e) a complex rearrangement including an inversion (change in orientation) of sequence associated with duplication (part of the green segment) and deletion (part of the purple segment).

was first suggested as the basis for duplications causing Charcot–Marie–Tooth disease type 1A (CMT1A) (Ref. 23) and subsequently for recurrent changes associated with a wide array of other genomic disorders (Ref. 24).

As more and more genomes are assayed, evidence is accumulating for other mechanisms that generate gains and losses. These include nonhomologous end joining (NHEJ) (Refs 25, 26), a process that is important for generation of B cell and T cell receptor diversity; fork stalling and template switching (FoSTeS) (Ref. 27), first invoked to explain nonrecurrent rearrangements in Pelizaeus–Merzbacher disease (Ref. 28) and more recently for duplications and triplications of the *MECP2* (methyl-CpG-binding protein 2) gene associated with developmental delay and mental retardation in males (Ref. 29); and microhomology-mediated break-induced replication (Refs 30, 31). Although these essentially nonrecurrent mechanisms create a large diversity of CNV

breakpoints with complex architecture, overlapping sets may be associated with some common phenotypic features, reflecting shared dosage-sensitive genes within the deleted or duplicated segments.

Family studies may also be informative with respect to the genesis of a CNV (as well as assessing likelihood of pathogenicity). In particular, an inversion in a parental chromosome may predispose to a de novo unbalanced variant in an offspring. The 17q21.31 microdeletion syndrome is a notable example (Refs 32, 33), as all currently reported cases result from a parental inversion common in Europeans (Ref. 34). Sotos syndrome in Japanese patients is usually due to a paternal microdeletion, which is associated with a paternal inversion (Ref. 35). Williams–Beuren syndrome is often similarly associated with predisposing parental inversions (Refs 36, 37) and other examples continue to emerge, reinforcing the rationale for investigation of parental samples when CNVs are found in

5

clinical investigations, in order to properly counsel about recurrence risk.

Relationship of CNVs and SNPs

SNPs are single base substitutions found throughout the genome, each with a maximum of four possible alleles, although common SNPs usually have only two represented. They can therefore be assayed and documented in binary formats. CNVs are more complex than SNPs, often by orders of magnitude. Either form of variation can involve coding or noncoding sequences, but whereas individual SNPs affect a single site, individual CNVs may encompass multiple contiguous genes. The difference in complexity is even more important collectively, because SNPs are discrete, but CNVs among different chromosomes can be overlapping, with variable DNA portions in common and different endpoints. Furthermore, although resolution for the assays used to determine the extent (i.e. size) of CNVs has improved dramatically, there are usually still limits to the precision with which they are demarcated. As a result, it is an important but challenging task for databases to determine how to document overlapping and nested sets of CNVs in a way that is helpful for clinical research. Aside from the variable size of a CNV segment, there are aspects such as orientation and iterations to accommodate. All in all, CNVs have many more degrees of opportunity for creating variation.

The genetic relationship of CNVs and SNPs to each other (linkage disequilibrium) has been examined by determining the proportion of CNVs that can be 'tagged' well by nearby SNPs (Refs 13, 38). Such 'taggability' was shown to depend on CNV allele frequency and local SNP density, but not CNV size. Overall, the taggability of biallelic CNVs examined was found to be largely similar to that of frequency-matched SNPs, except when rare CNVs were examined, presumably because these events were recent in origin or under negative selection. Interestingly, deletions are found to be better tagged than duplications, which may be a result of the chromosomal dispersion of some duplications and an increased frequency of reversions and multiple new mutations at some duplications.

Prevalence and frequency of CNVs

The remarkable insight of the past five years has been the extent to which CNVs are found as likely explanations, or at least highly suspect candidates for participation, in disease causation, and also their prevalence throughout all genomes, regardless of any association with pathology. Now that researchers are aware of this form of variation, searching for it has become very fruitful. The Database of Genomic Variants (DGV) (Refs 2, 39) documents variation found in population control samples, with more than 29 000 CNVs recorded as of December 2009 (http://projects.tcag.ca/variation/) (Ref. 2). Whether CNVs are more important or more abundant than SNPs as sources of human variation or disease is readily debated. However, it is clear that as a result of their size, CNVs collectively account for more of the variable genome than do SNPs (Refs 16, 40). The first two single human genome sequences (Refs 40, 41) provided an opportunity to look at the number of CNVs in individual genomes [relative to the haploid composites of the Human Genome Project (Refs 42, 43) and Celera Genomics (Ref. 44)]. Recently, we found that ~1.28% of nucleotide variation between the first individual human genome sequence (Ref. 40) and the reference genome assembly was accounted for by CNV, far exceeding the 0.1% encompassed by SNPs (C. Lee and S.W. Scherer, unpublished).

The meaning of the term polymorphism in a genetic context has become muddled, and its use in describing structural variants might well be avoided in the interests of clarity. Descriptors such as 'rare' and 'common' in reference to a CNV apply to the frequency of a given variant rather than to the state of the locus. By convention, a rare variant has a frequency of less than 1% in a population, and this threshold is useful (but should always be specified). Many medical conditions that are relatively common in the population are clearly the result of heritable (and other) risk factors, but the search for common genetic variants to account for the majority of the heritability underlying these phenotypes has generally been unfruitful (Refs 45, 46), first at the level of nucleotides and also with CNV analyses (Ref. 13). What is emerging, however, is evidence that multiple rare CNVs – de novo or inherited – may contribute to the genetic vulnerability for conditions such as schizophrenia or autism (Ref. 47), and likely to many other medically important conditions.

The clinical context of copy number variation in the human genome

6

This creates situations of great complexity to analyse and interpret, and will continue to challenge medical researchers for years to come.

Means of detection: evolution and implications

Array CGH was the technology that disclosed the large but submicroscopic CNVs, first with array probes made from relatively large DNA segments cloned in BACs. Significant refinements have ensued, such as arrays made with smaller oligonucleotide probes (for enhanced resolution), and much greater numbers of probes on each array (for denser coverage of the genome). Arrays designed to genotype SNPs are also exploited for dosage information, by looking for stretches of these markers with increased or decreased signal intensity. Recent strategic modifications to SNP arrays enhance the opportunity to discover CNVs along with concomitant SNP genotypes. The scope of these arrays may be genome-wide (with breadth but with gaps in coverage), targeted (for example to a specific gene or region of interest) or semitargeted (such as only probes for chromosome 21). Particularly for clinical diagnostics, hybrid panels are being developed, with some depth of genome-wide coverage in addition to higher density of regions known to harbour clinically relevant CNVs (Ref. 48).

The alternative to hybridisation methods for detection of CNVs is direct comparison of DNA sequence data between reference and other genomes. As methods for whole-genome sequencing become more efficient and effective, individual genome data will soon accumulate in databases and this method of analysis will undoubtedly predominate. The direct approach to sequence comparison (Refs 10, 40, 41, 49, 50, 51, 52, 53, 54) will eventually allow a much more complete and precise documentation of genomic variation (Ref. 55), but there can be technical obstacles that keep some genomic regions obscured (Ref. 7). Nevertheless, compared with array-based approaches, analysis by, for example, massively parallel sequencing can provide precise determination of breakpoints and copy number, will detect smaller alterations and copy-number-neutral rearrangements and (of particular importance for tumour analysis) can accommodate cellular admixture (Ref. 56).

How CNVs can cause disease

Both SNPs and CNVs provide the basis for phenotypic variability, which is essential for adaptive evolution. They may also be maladaptive in a particular environment, or more globally. In humans, this creates one end of a phenotypic spectrum recognised as disease (Ref. 14), bringing individuals and families to medical attention or seeking clinical intervention. These dysadaptive changes may directly involve genes, but not necessarily, and their pathogenicity can result from quantitative (dosage) or disruptive effects (Table 2; Fig. 2). CNVs that are intragenic or involve a single gene may have functional consequences that are similar to point mutations, behaving much as classical mendelian dominant or recessive traits. Alternatively, CNVs overlapping genes can result in fusion genes that may have phenotypic consequences. More extensive CNVs comprise multiple genes and underlie the 'contiguous gene syndromes' or genomic disorders (Ref. 57). Many other conditions seem to be related to complex combinations of events at noncontiguous loci.

Deletion of a genomic segment causes hemizygosity for the deleted interval, which may also result in haploinsufficiency for dosage-sensitive gene(s). For example, a CNV deletion of *LCE3B* and *LCE3C* (two late cornified envelope genes) has been shown to be a risk factor for psoriasis (Refs 58, 59). Copy number gains, such as duplications, may create imbalances due to excess product of the duplicated genes, or, when intragenic, may alter the structure of a product and thereby its function. The dosage effects of CNVs can be incremental, particularly when associated with higher-order replication (e.g. due to unequal crossover events), in which case, the relationship between copy number and disease states may be more subtle and related to thresholds (Fig. 2c and e). Other studies are revealing certain phenotypes to be associated with a more generalised increase in CNVs. For example, the number of CNVs per genome is strikingly increased in cancer-prone individuals in families with Li–Fraumeni syndrome (Ref. 60), and this observation has prompted similar investigations for neuroblastoma (Ref. 61) and many other phenotypes. CNVs are manifesting the extent to which genomes are unstable, and family studies will allow determination of not just how

The clinical context of copy number variation in the human genome

7

Table 2. Spectrum of copy number variation genotypes and illustrative phenotypes

CNV genotype feature	Illustrative phenotypes (gene symbol or locus)	Refs
Single genes causing phenotype		
Recessive disruption or loss	Duchenne/Becker muscular dystrophy (*DMD*)	173, 174
Dominant disruption or loss	Neurofibromatosis 1 (*NF1*)	175, 176, 177
	Tuberous sclerosis (*TSC1* or *TSC2*)	87, 133
	Sotos syndrome (*NSD1*)	35, 178, 179
	CHARGE syndrome[a] (*CHD7*)	143
Dosage effect	Pelizaeus–Merzbacher disease (*PLP1*)	28, 180
	Early-onset Alzheimer disease (*APP*)	181, 182, 183
	22q autism (*SHANK3*)	146, 184, 185, 186
	Psoriasis (*LCE3C/LCE3B*)	58, 59
Position effect	Aniridia (*PAX6*)	88, 187
	Triphalangeal thumb–polysyndactyly syndrome (ZRS)	188, 189
	Crohn disease (*IRGM*)	126
Multiallelic effects	Crohn disease predisposition (*DEFB4*)	190, 191
	Systemic autoimmunity predisposition (*FCGR3B*)	124, 192, 193
	Parkinson disease (*SNCA*)	194
	HIV/AIDS susceptibility/Kawasaki disease susceptibility/rheumatoid arthritis predisposition (*CCL3L1*)	127, 128, 129, 130, 195
Multiple genes potentially involved in phenotype		
Recurrent	Velo–cardio–facial syndrome/DiGeorge syndrome (22q11.2 deletion)	74, 75, 196, 197
	Williams–Beuren syndrome/7q11.23 duplication syndrome	36, 37, 67, 96, 98, 100, 198, 199
	17q21.3 microdeletion/microduplication syndromes	32, 33, 34, 200, 201
	1q21.1 neuroblastoma	61
	15q13.3 microdeletion syndrome	105, 106, 107, 108, 109
	1q41q42 microdeletion syndrome	202
	16p11.2-p12.2 microdeletion syndrome	203
	16p11.2 microdeletion/microduplication syndrome(s)	145, 147, 148, 149, 204
	3q29 microdeletion/microduplication syndromes	205, 206
Nonrecurrent	Potocki–Lupski syndrome [dup(17)(p11.2p11.2)]	207
	WAGR syndrome[b] (11p13)	88
Heterogeneous	Autism	47, 118, 144, 145, 146, 147, 208, 209, 210, 211, 212
	Bipolar disorder	213, 214, 215
	Schizophrenia	47, 89, 106, 113, 149, 214, 215, 216, 217, 218
	Age-related macular degeneration	219, 220
	Tetralogy of Fallot	90, 221

(continued on next page)

The clinical context of copy number variation in the human genome

Table 2. Spectrum of copy number variation genotypes and illustrative phenotypes (continued)

CNV genotype feature	Illustrative phenotypes (gene symbol or locus)	Refs
Buffering or modifier effects	Thrombocytopaenia absent radius (TAR) syndrome (1q21.1)	77, 78
	Spinal muscular atrophy (*SMN1/SMN2*)	79, 222, 223
Epigenetic effects	Silver–Russell syndrome (11p15 duplication)	80
Somatic mosaicism	Rubinstein–Taybi syndrome (*CREBBP, EP300*)	84, 85, 86
	Tuberous sclerosis (*TSC1, TSC2*)	87
	Aniridia (11p13)	88
Somatic instability	Li–Fraumeni syndrome	60, 133
	Tuberous sclerosis	133
	Neurofibromatosis	224

[a]Coloboma, heart anomaly, choanal atresia, retardation, genital and ear anomalies.
[b]Wilms tumour, aniridia, genitourinary anomalies, mental retardation.
Abbreviations: CNV, copy number variation; HIV/AIDS, human immunodeficiency virus infection and/or acquired immune deficiency syndrome. Full versions of gene names can be found on the HUGO Gene Nomenclature Committee website (http://www.genenames.org/).

commonly CNVs exist, but also how frequently they occur de novo or change during transmission between generations (Ref. 62) (Fig. 3).

Disruptive effects of CNVs result from a variety of mechanisms. A breakpoint within a gene may functionally disable it, but there might also be impact due to disruption or disassociation of promoters or other regulatory elements, or effects on local chromatin structure (Refs 63, 64) (Fig. 2d). These effects may be long-range; for example, microduplication of a conserved noncoding sequence about 110 kb downstream of the *BMP2* (bone morphogenic protein 2) gene, with demonstrated enhancer function, was recently shown to underlie brachydactyly type 2A in two families (Ref. 65). A study of gene expression in HapMap lymphoblasts revealed more than half of the effects of currently known CNV are caused not by altering gene dosage, but by gene disruption or by affecting regulatory or other functional regions, some more than 2 Mb apart (Ref. 66). Analysis of gene expression from within and flanking the region deleted in Williams–Beuren syndrome (Ref. 67) found evidence of significant dysregulation of genes up to 6.5 Mb beyond the deleted region, as well as a lack of direct correlation with copy number for expression of the deleted genes. Clearly the cis-regulatory effects of CNVs can spread well beyond their borders, and genes involved in disease phenotypes may well lie outside of the associated deleted or duplicated segments.

Pathogenicity of a given CNV can be difficult to establish. In investigations initiated by an abnormal phenotype in an individual or cohort (phenotype first), the goal is to find a genotypic explanation to enhance further studies (for clinical research) or to make a diagnosis (in clinical practice) (Fig. 4). The implications of determining pathogenic potential become greater when a CNV is found before a phenotype is known (genotype first), and predictions are expected, upon which interventions may be taken – prenatal diagnosis being, of course, the circumstance of greatest concern in this respect. Various characteristics of pathogenic versus benign variants are outlined in Table 1 of Ref. 8; major considerations involve validation to confirm the chromosomal location and extent of the variation, family studies to determine whether others share the variant genotype or it is de novo, comparison with precedents documented in databases of healthy or affected individuals [such as DGV or DECIPHER (Ref. 68),
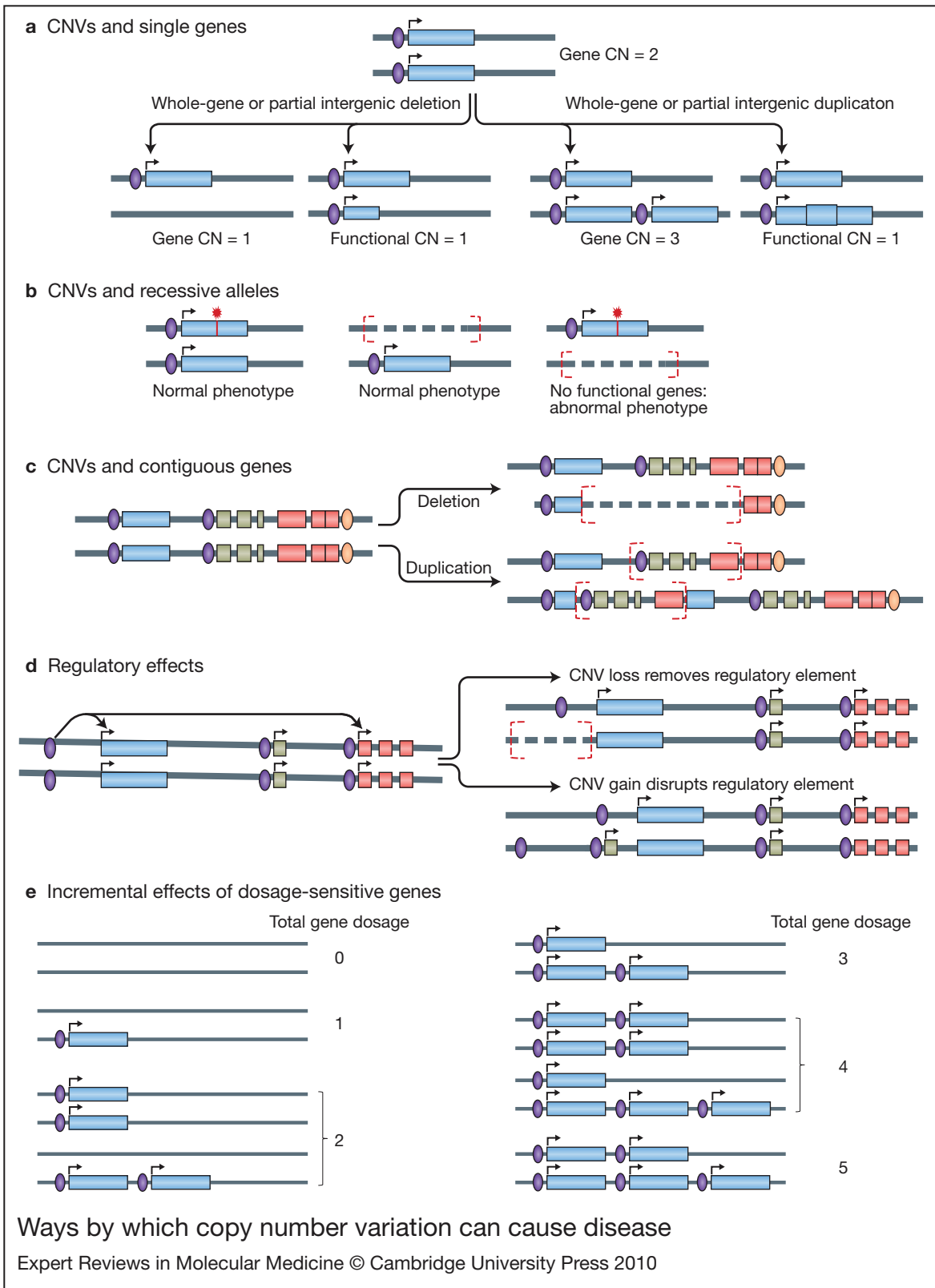
9

a CNVs and single genes

Gene CN = 2

Whole-gene or partial intergenic deletion

Whole-gene or partial intergenic duplicaton

Gene CN = 1   Functional CN = 1   Gene CN = 3   Functional CN = 1

b CNVs and recessive alleles

Normal phenotype   Normal phenotype   No functional genes: abnormal phenotype

c CNVs and contiguous genes

Deletion

Duplication

d Regulatory effects

CNV loss removes regulatory element

CNV gain disrupts regulatory element

e Incremental effects of dosage-sensitive genes

Total gene dosage

0

1

2

Total gene dosage

3

4

5

Ways by which copy number variation can cause disease

Expert Reviews in Molecular Medicine © Cambridge University Press 2010

The clinical context of copy number variation in the human genome

Figure 2. Ways by which copy number variation can cause disease. (*See next page for legend.*)

Figure 2. Ways by which copy number variation can cause disease. (*See previous page for figure.*) This figure illustrates mechanisms underlying quantitative (dosage) or disruptive effects of copy number variation (CNV). Genes are indicated by coloured boxes, while promoters are depicted by coloured ovals. The direction of transcription is indicated by bent arrows above the genes. (a) CNVs can change the number of functional gene copies, through whole or partial deletions or duplications of genes. (b) A recessive mutant allele (indicated by red marker) can be unmasked by a deletion, which causes the loss of both functional copies of the gene. (c) Contiguous gene deletions can also eliminate (green) or disrupt (blue and red) functional genes; additionally, the mechanisms causing contiguous gene deletions can also cause a reciprocal duplication. These duplications can disrupt a dosage-sensitive gene (blue) or increase the copy number of a dosage-sensitive gene (green), which can cause disease. (In this example, another gene, shown in red, has partial duplications of its 3′ end.) (d) CNVs can also cause disease when deletions or duplications interrupt control regions that regulate juxtaposed and distant genes. Lastly, (e) CNVs can have an incremental effect when the copy number of dosage-sensitive genes is modified.

respectively] and knowledge of the genic content of the variant segment. Thus, a CNV that is inherited from a healthy parent or found in healthy family members, or that overlaps variants established in the DGV or does not involve genes of known clinical significance, is more likely to be phenotypically benign. A CNV that is shared by affected family members, or is documented in association with clinical phenotypes [found to be a risk factor using relative risk (RR) and/or odds ratio (OR) (Box 1)], or is gene-rich, particularly if any genes involved are documented in the morbid map of Online Mendelian Inheritance in Man (OMIM) (http://www.ncbi.nlm.nih.gov/omim/), is more likely to be of pathogenic consequence. It is important to note that the characteristics described here rarely allow a definitive determination of whether a given CNV is or is not the explanation for a phenotype (for phenotype-first investigations) or will cause a particular phenotype (for genotype-first studies). Importantly, additional functional studies would need to be conducted to investigate pathogenicity of CNVs and understand the relationship between the genotype and the phenotype. As with most areas of medicine, accurate annotation of the cause and effect of genomic variation requires a combination of analyses, experience and expert judgement for interpretation.

Aside from its content, the overall genomic context of a particular CNV is critical to its phenotypic consequences, and this knowledge is still in its infancy. In a very simple example, a CNV that deletes a dosage-insensitive gene may be completely recessive, but if the remaining allele happens to carry a functional mutation, then a phenotype may ensue (Fig. 2b). Research

may eventually identify specific pathogenic combinations of CNVs that otherwise might be individually benign (Ref. 8), or the converse: certain CNVs that are pathogenic unless a compensatory element is present elsewhere in the genome, epigenome or environment to reduce penetrance. Interpretations of this kind will be enhanced as whole-genome sequence analysis becomes the norm. The apparent lack of phenotype associated with an isolated CNV does not rule out its pathogenic potential in another genomic context. Although a CNV that has arisen de novo is more likely to be pathogenic than one that has escaped selection in a family or population, this is again probabilistic. This assertion is due to the fact that inherited CNVs are present in at least one reproductively viable individual, while de novo CNVs are present in a single individual and may not have been subjected to negative selection. However, this is not a definitive distinction between these classes of CNVs. Finally, CNVs of pathological consequence are more likely to be large (encompassing many genes and/or regulatory sequences), and to involve loss, rather than gain, of genomic material (although early data are somewhat biased because of the relative ease of ascertainment of deletions and larger segments) (Ref. 13).

Means of ascertainment shapes findings

Ascertainment bias is an inevitable component of research, and acceptable as long as it is acknowledged and accounted for. The remarkable aspect of CNVs has been not so much discovery of their association with genomic disorders as their ubiquity throughout genomes of general populations. To study
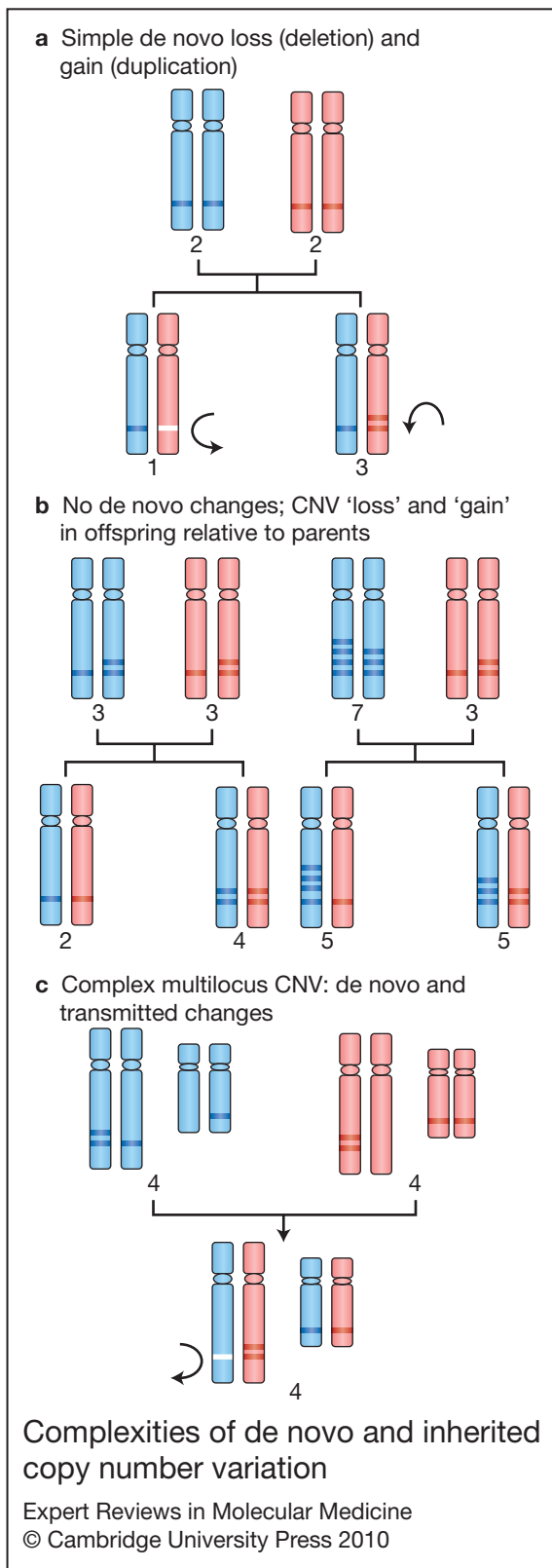
The clinical context of copy number variation in the human genome

a Simple de novo loss (deletion) and gain (duplication)

b No de novo changes; CNV 'loss' and 'gain' in offspring relative to parents

c Complex multilocus CNV: de novo and transmitted changes

Complexities of de novo and inherited copy number variation

Expert Reviews in Molecular Medicine
© Cambridge University Press 2010

Figure 3. Complexities of de novo and inherited copy number variation.

Figure 3. Complexities of de novo and inherited copy number variation. This figure uses a schematic of chromosomes (blue, paternal; pink, maternal) to illustrate transmission of copy number variation (CNV) to offspring. The gene copy number is given below each chromosome pair. Both de novo (indicated by curved arrow) and transmitted changes in CNV copy number are shown. In (a), single de novo deletion and duplication are shown within the maternal chromosome. In (b), no de novo changes are seen, but in each case the offspring has a different copy number than the parents. In the case of the multiallelic variant shown on the right, offspring have the same gene copy number but different gene configurations. Finally, in (c), both de novo and transmitted changes in copy number are combined to show a complex multilocus CNV. In this example, the offspring shows no change in copy number, despite de novo deletion.

cause-and-effect relationships and to put knowledge to use for clinical practice, we need to compare the prevalence of CNVs between defined cohorts and nonclinical samples. The developing means to undertake relatively hypothesis-free and fully comprehensive data collection has begun to provide unprecedented opportunities for such analysis (Fig. 4). For a fully genotype-first approach, data must be gathered from an unselected cohort, such as all newborns, followed by phenotypic comparison of all those who share particular CNV genotypes. A phenotype-first approach involves collecting a cohort with a particular clinical presentation or diagnosis, and looking for CNVs that are more prevalent among them, relative to those without the phenotype. Genotyping can be genome-wide (hypothesis-free), or a targeted search around a candidate locus [compare, for example approaches of Miller et al. (Ref. 69) and Sharp et al. (Ref. 70) with respect to the 15q13.3 deletion syndrome]. Many current studies are somewhat intermediate, in that they involve samples referred for analysis because of some clinical finding, and research outcomes can be strongly influenced by reasons for such referrals – such as developmental delay, behaviour issues, dysmorphic features and so on. A recent commentary (Ref. 71) compares early phenotype-led and more-recent CNV-led studies (Refs 72, 73) that focus on phenotypes
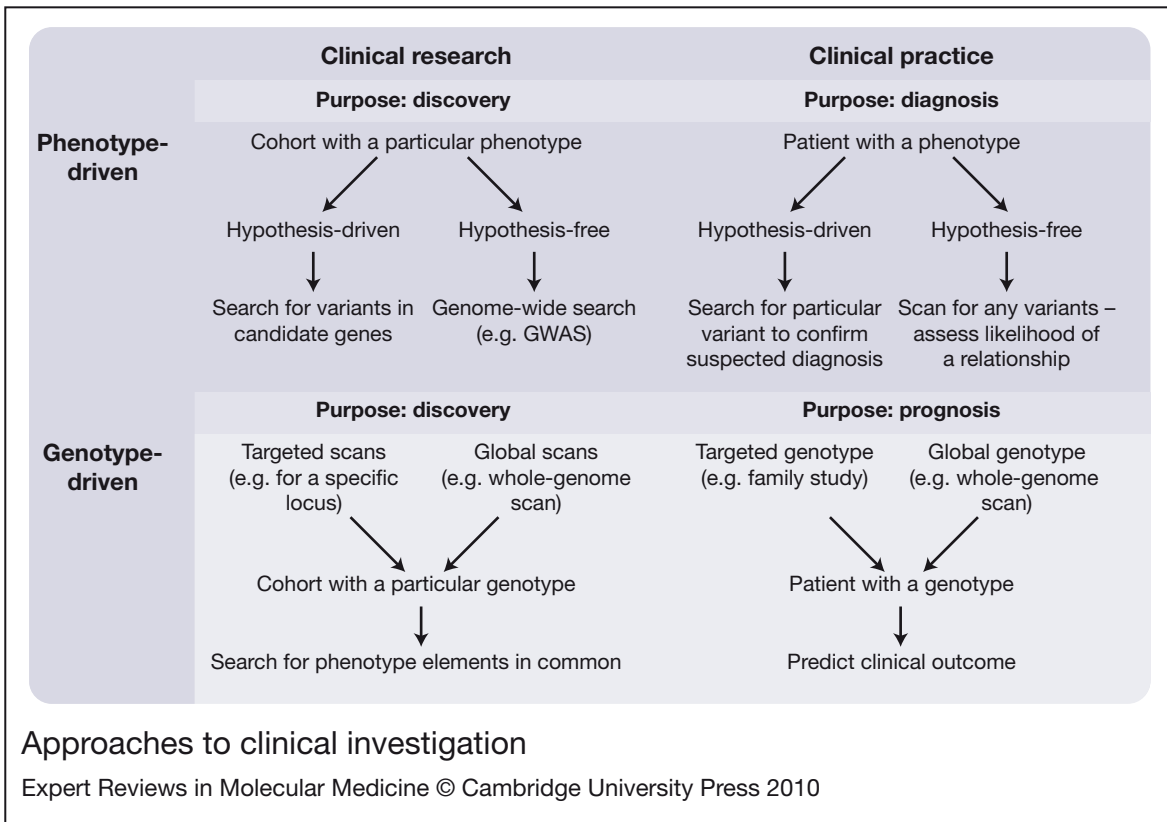
12

Figure 4. Approaches to clinical investigation. This figure breaks the different approaches for clinical investigations into phenotype-driven and genotype-driven approaches. These are further broken into investigations involved in clinical research, aimed at discovery, and investigations involved in clinical practice, aimed at diagnosis or prognosis. Flow charts illustrating different investigations to discover and analyse copy number variation are included in each category. The means of CNV ascertainment, be it phenotype-driven or genotype-driven, can significantly influence the interpretation of disease associations. Abbreviations: CNV, copy number variation; GWAS, genome-wide association study.

associated with deletions and duplications at 1q21.1, with markedly different outcomes from the various approaches – what you find depends largely on what you look for.

Penetrance and expressivity of CNVs

A mutant gene is described as fully penetrant when all individuals with the mutation express the related phenotype, whereas reduced penetrance refers to a situation in which some individuals with a given genotype show phenotypic evidence of it and some do not. To some extent, it is one end of a spectrum of variable expression of the phenotype, and the concepts are inter-related. They apply also to observations about the much more complex genotypes involving CNVs and their associated phenotypes, and examples span a wide range

of relationships. In some situations, there is a consistent relationship between the CNV genotype and at least a core phenotype, such that all individuals with (for example) a given deletion share a definable phenotype, and every individual with that phenotype has a similar or overlapping deletion. Clinical examples of this kind include Williams–Beuren syndrome, Prader–Willi and Angelman syndromes, and the 17q21.31 deletion syndrome. Other CNVs, such as the most common human microdeletion – 22q11.21 – are highly penetrant but with a range of phenotypic expression so broad as to encompass more than one clinically designated syndrome (Refs 74, 75). A nearby multiply ascertained microduplication was associated with such disparate findings as to 'obfuscate the clinical relevance of the

The clinical context of copy number variation in the human genome

13

molecular data' (Ref. 76), and the 1q21.1 microdeletions are said to be so variable as to 'elude syndromic classification' (Ref. 73). Other conditions, notably psychiatric disorders including autism, have a more nuanced connection to the various CNVs emerging as factors that are significantly associated but not independently causative for the phenotype (Ref. 47). Evidence of reduced penetrance abounds in these families with inherited CNVs, although retrospective evaluation of apparently unaffected parents or other relatives sometimes reveals subtle features of the proband's phenotype.

Buffering or modifier effects have been described for the thrombocytopaenia absent radius (TAR) syndrome (Refs 77, 78), for which a deletion at 1q21.1 is necessary but not sufficient to cause the syndrome, and for spinal muscular atrophy (Ref. 79), where the impact of an intragenic deletion of the *SMN1* (survival of motor neuron 1, telomeric) gene may be tempered by normal variation in the number of gene copies of *SMN1* and the closely related *SMN2*. Epigenetic effects may also influence expression of the phenotype, as exemplified with Silver–Russell syndrome (Ref. 80) and possibly developmental verbal dyspraxia (Ref. 81).

Germline and somatic CNVs

Genomic alterations, including CNVs, have one of four origins: they can be (1) inherited from a parent with the same germline variant, (2) inherited from a parent with germline mosaicism for the same variant, (3) arise de novo from a parental germ cell or (4) arise de novo in a somatic cell (Fig. 3). The latter category is especially relevant to the field of cancer genomics because of generalised genomic instability, and clonal expansion and evolution of tumour cells. Somatic mosaicism for CNVs has also been noted in monozygotic twins (Ref. 82) and in different tissues of an individual (Ref. 83), as well as in diseases such as Rubinstein–Taybi syndrome (Refs 84, 85, 86), tuberous sclerosis (Ref. 87) and aniridia (Ref. 88). In clinical genetics applications, it is important to distinguish among these categories, both for clinical research and to predict outcomes for newly ascertained probands and recurrence risks for families. Among de novo events, some are truly random,

but, in contrast to single-base mutations, the structural variants are often associated with vulnerable genomic regions in which similar CNVs tend to recur. These variants can, in turn, beget more genomic instability with disrupted chromatin structure or opportunities for misalignment.

Enigmas in CNV genotype–phenotype relationships

Some approximately similar CNVs have emerged in the context of various different complex phenotypes. Duplications of 17p12 cause CMT1A and the reciprocal deletion is associated with hereditary neuropathy with liability to pressure palsies (HNPP), but the region is also implicated in schizophrenia (Ref. 89). Deletions at 1q21.1 also emerged in schizophrenia phenotype-driven studies (Ref. 89), but CNVs of this region are found enriched in association with phenotypic features such as micro- or macrocephaly, mental retardation, cardiac anomalies or autism (Refs 72, 73). Furthermore, family studies demonstrate that the same CNVs can be without apparent consequence in some individuals.

Some ostensibly similar phenotypes are associated with various different genotypic findings, each of a magnitude to elicit suspicion with respect to pathogenicity. Neurological and psychiatric conditions seem to predominate as examples (Table 2) but the cardiac defect known as tetralogy of Fallot has recently provided similar genotypic characteristics (Ref. 90).

Syndrome, meaning 'running together', describes clinical entities that involve constellations of features from different systems. Certainly there are examples of single-nucleotide mutations that have pleiotropic effects and create multisystem phenotypes, but CNVs are more likely to do so because of their potential to compromise multiple genes, with concomitantly widespread effect. As more information emerges about such genotype–phenotype relationships, we are struck by the enigma that some classical syndromes, Williams–Beuren syndrome for example, have a relatively consistent genotype and phenotype presentation, whereas the highly recognisable phenotypic constellation of Down syndrome can result from CNVs ranging from full trisomy

The clinical context of copy number variation in the human genome

21 to almost any portion thereof. At the same time, some recurrent CNVs have been discovered in clinically unselected cohorts (such as microdeletions of 1q21.1, described above) that, despite considerable genotypic consistency, have no recognisable consistent 'running together' of features.

Clinical exemplars

The ability to undertake whole-genome scans by arrays or sequencing has provided the opportunity to discover individual disease-associated CNVs in the absence of any prior hypotheses as to their chromosomal location. This holistic approach is also revealing combinations of CNVs, both collectively and within individuals, that may become the key to understanding complex phenotypes such as autism, bipolar disorder, schizophrenia, macular degeneration, or tetralogy of Fallot (Table 2). A concept is emerging of CNV load, as cohorts or individuals are recognised to have a higher than average number of CNVs, rather than specific aberrations in candidate genes. Li–Fraumeni syndrome provides a striking prototype (Refs 60, 91). Below, we describe representative examples of the effects of CNVs in clinical conditions.

Down syndrome

Down syndrome is something of a metaphor for the progress of CNV discovery in humans. When we consider CNVs in the broader definition to include microscopic variants, then trisomy 21 was arguably the first to be discovered (Ref. 1) (Table 1). Whole-chromosome aneuploidies have different underlying mechanisms than the submicroscopic variants, but the phenotypic consequences are not categorically distinct; rather, they are part of a continuous spectrum in this respect. After recognition of nondisjunctional trisomy 21, rearrangements such as Robertsonian translocations were found as the basis for duplicated long arms of chromosome 21, and then microscopic partial trisomies, followed by those detectable by FISH. Eventually, arrays have been used to fine-tune the extent of duplicated material with higher-resolution mapping, and to study correlates of specific features of the phenotype with particular genes or regions (Refs 92, 93). A very early study of the reciprocal deletion syndrome (i.e. partial monosomy 21) made

some prescient observations concerning gene-dosage effects: 'Our findings do add weight to the hypothesis that genetic control of enzymes is not a simple gene-dosage affair, but a complex interaction of structural, regulator, and modifying genes which may be located at various loci on different chromosome segments.' (Ref. 94).

Williams–Beuren syndrome and its reciprocal 7q11.23 duplication syndrome

Williams–Beuren syndrome is one of the classic genomic disorders – a contiguous gene syndrome associated with a recurrent microdeletion of 7q11 that is strikingly consistent. The recurrence is mediated by flanking segmental duplications and by a relatively common inversion of the region (carried by up to a third of parents of affected individuals and 5% of the general population), which predisposes to aberrant meiotic recombination in parental chromosomes, with pathological outcomes in the offspring (Refs 36, 95, 96). The deletion phenotype is a relatively predictable syndrome. As anticipated for the CNVs mediated by NAHR, by which deletion outcomes should be matched by reciprocal duplication products (Refs 22, 97), the complementary duplication syndrome was eventually recognised (Refs 98, 99, 100). Its clinical phenotype is distinct from that associated with the deletion, and, particularly with respect to expressive speech ability, is in striking contrast, suggesting some effects of gene dosage. As discussed, the impact of the Williams–Beuren microdeletion extends to genes well beyond the borders of the aberrant segment (Ref. 67).

15q13.3 microdeletion and duplication phenotypes

This recurrent CNV locus has been recognised only recently, by aCGH (Refs 70, 101), but is repeatedly coming to attention from a variety of study groups. It illustrates the challenges in assessing pathogenicity of these variants, and the impact of ascertainment. The region is adjacent to that deleted in Prader–Willi and Angelman syndromes (PWS/AS), which together feature a series of duplication blocks demarcated by recurrent breakpoints (BP1 to BP6) (Refs 102, 103, 104). Just distal to the PWS/AS region is the 1.5 Mb segment BP4–BP5, which is found to be

The clinical context of copy number variation in the human genome

deleted or duplicated in an increasing number of individuals ascertained through routine and targeted clinical investigations, occasionally as part of a larger CNV (Refs 105, 106, 107, 108, 109). Clearly the region is enticing, drawing interest from several clinical directions, but observations are disparate. Among controls, deletions have been mostly limited to a handful of Icelandic individuals (Ref. 106), but several studies that included family investigations have discovered deletions or duplications in parents or other relatives who do not share the probands' phenotypes (Refs 69, 110, 111, 112), indicating that deletion of this region is not necessarily pathogenic, but also not inconsequential.

Even from relatively untargeted clinical investigations (Refs 69, 70, 110, 112), details of phenotypes associated with the 15q13.3 CNVs are skewed by the nature of the referral base – for example, predominantly developmental delay, dysmorphic features, multiple congenital anomalies and behaviour issues. Deletions of 15q13.3 were found in up to 0.3% of such referrals; duplications were rarer. When parental samples were available, the majority of these probands' CNVs were found to be inherited.

In studies of more clinically defined cohorts (phenotype-first), deletions of the BP4–BP5 segment (or more) were rarely, but significantly, associated with schizophrenia (Refs 106, 113) or idiopathic generalised epilepsy (Ref. 114); deletions or duplications of the same segment appear with tantalising frequency when autism or related features such as expressive language delay are part of the phenotype (Refs 69, 110, 111, 112). Despite relative consistency of the CNV genotypes found across a broad range of studies, the phenotypes of these individuals vary greatly. With current evidence, the finding of a CNV involving 15q13.3 in a clinical investigation would raise concern, but would probably be insufficient to explain or predict any particular phenotype. Such observations of uncertainty will consume a great deal of health professionals' time for the foreseeable future (Refs 115, 116).

Chromosome 1q21.1 CNV phenotypes

Evidence of a potentially contiguous gene syndrome at 1q21.1 was first noted in a targeted candidate gene study of a cohort with congenital heart defects (Refs 16, 117). Later, genome-wide phenotype-first surveys detected significant association of similar duplications with autism (Ref. 118) and deletions with schizophrenia (Refs 106, 113). Using the complementary genotype-first approach, two large studies (Refs 72, 73) started with relatively unselected clinical referral cohorts to ascertain, through data from genome-wide or targeted assays, large numbers of individuals with CNVs involving 1q21.1 and then to document the scope of associated clinical phenotypes. Both found large (~1.35 Mb) recurrent deletions of the region as well as reciprocal duplications, but other than some relationship between CNV dosage and micro- or macrocephaly (Ref. 72), there was such a wide range of clinical presentation among index cases that no common manifestations of a syndrome could be recognised. Furthermore, although deletions and duplications of this region are clearly rare in the general population (Ref. 119), family studies for those ascertained in the clinical cohort showed many CNVs to have been inherited from parents with milder or absent features relative to those of their respective offspring. Similar to the situation for 15q13.3, and undoubtedly for many regions yet to be characterised, the finding of a CNV in this region would raise legitimate suspicion, but, with current information, would be sufficient neither to explain nor to predict a particular clinical outcome.

Attention has been drawn recently to a smaller previously known CNV (Ref. 120) just distal to the 1q21.1 region, for which the deletion allele is common (9.1%) in controls but significantly more prevalent (15.6%) among cases with neuroblastoma (Ref. 61). Cis and trans effects appear to be involved as part of dosage effects on susceptibility, and scrutiny as a result of the initial observation led to discovery of a novel transcript of interest from within the deletion interval.

The 1q21.1 region is flanked proximally by another segment of 200 kb that is deleted in all individuals with TAR syndrome and not in controls studied to date (Ref. 77). In at least two families (Refs 77, 78), the microdeletion was inherited from an unaffected parent, indicating that the CNV is necessary but not sufficient to cause the syndrome. At present, therefore, this CNV is a helpful diagnostic tool in the context of other clinical findings, but is not

16

in itself predictive of the TAR phenotype. We also note that larger deletions of 1q21.1 could also influence risk for neuroblastoma (Ref. 61) and that chromosomal inversion encompassing the 1q21.1 region has also been observed (Ref. 16).

Immunity and autoimmunity

Early observations on the impact of CNVs was that they are particularly prevalent among genes that have a role in our interface with the environment (Refs 121, 122, 123), such as those that are part of the immune system. Various gene families, such as the major histocompatibility locus, immunoglobulins, chemokines, receptors, defensins and interleukins, might reflect CNV events throughout evolution, but are also characteristically polymorphic with much of the variation contributed by CNV for individual loci. By contrast to simple deletion variants, these sites are typically multiallelic, reflecting a wide range of copy number and creating particular challenges for discerning their incremental effects and specifying exact copy numbers, but new approaches have been reported (Refs 124, 125).

Disorders with an autoimmune component, such as psoriasis (Ref. 59), systemic lupus erythematosus (SLE), type 1 diabetes, and rheumatoid arthritis, are all beginning to yield some of the mystery of their respective causes as more refined data emerge from these highly variable genomic sites. Interestingly, a recent discovery found that a polymorphic deletion variation upstream of *IRGM* (immunity-related GTPase family, M) was associated with Crohn disease (Ref. 126).

Autoimmunity is a maladaptive consequence of an adaptive immune system, and variation at some of the relevant loci can have a spectrum of clinical consequences. For example, a lack of the chemokine receptor CCR5 or an excess of its ligand CCL3L1 appears to be protective against human immunodeficiency virus (HIV) infection and acquired immune deficiency syndrome (AIDS) (Refs 127, 128, 129), but at the same time is associated with an enhanced inflammatory and autoimmune response, predisposing to rheumatoid arthritis (Ref. 130). This example reminds us that these variants do not act in isolation but as part of functional networks, and powerful analytical tools will be needed in order to interpret clinical data in the appropriate context.

Given the weak effect of HLA (human leukocyte antigen) matching to predict acute organ rejection in lung transplantation, a recent study (Ref. 131) considered other genetic risk factors – in particular, the chemokine ligand CCL4L1, genes for which are within a CNV region on chromosome17q12 that also contains *CCL3L1*. Copy number for *CCL4L1* was significantly greater in patients who experienced acute rejection, and even greater in those with multiple rejection episodes than among those who did not reject their allograft. Another study (Ref. 132) found a similar result when looking at mismatches for the homozygous deletion of *UGT2B17* between donors and recipients, which increased the likelihood of graft-versus-host disease. Whether these observations are a direct effect of gene dosage, or a proxy for nearby variable elements is not yet clear, but undoubtedly they will spark a flurry of research activity.

Modifier effects in Li–Fraumeni syndrome

This familial cancer syndrome is caused by mutations in the *TP53* gene (encoding p53), but the breadth of variation in severity, onset and types of tumours, even among those who share the same *TP53* mutation, prompted a genome-wide search for evidence of modifier loci. Rather than specific genomic sites, the number and size of CNVs was found to be markedly increased in *TP53* mutation carriers, particularly those with cancer (Ref. 60). The instability associated with these CNVs might in turn be the precursor to somatic changes and tumour formation. Information about the CNV load in *TP53* mutation carriers may provide an adjunct for risk prediction and counselling (Ref. 133).

A specific modifier was found within the *TP53* gene, comprising a common (∼10–30%) 16 bp microduplication in intron 3 (TP53PIN3), the presence of which is associated with an average onset of tumour diagnosis 19 years later than in mutation carriers without the duplicated variant (Ref. 134).

Two recent reports (Refs 135, 136) describe deletions of 17p13.1 encompassing the *TP53* gene found from whole-genome scans in three patients with mental retardation and dysmorphic features. In addition to providing a likely explanation for the referring clinical features, they predicted a Li–Fraumeni

17

phenotype for which appropriate risk management could be recommended. These add to a larger series found from among general clinical referrals of CNVs that involve genes with probable predisposition to various cancer syndromes (Ref. 137).

Discussion: clinical implications and applications

How are CNVs changing clinical practice?
The most conspicuous effect of the discovery of CNVs has been in laboratory medicine. There are many more diagnoses being made, but the distinction between classical cytogenetics and molecular diagnostics has become blurred (Ref. 138) as the gap in resolution of analysis is taken up with knowledge of this prevalent form of variation. New laboratory tools and skills are being invoked, and practitioners must broaden their expertise to encompass the entire spectrum of variation. The very particular skill of reading a traditional karyotype is rapidly being usurped by diagnostic arrays with less subjective interpretation, enhanced resolution and competitive costs. From the other end of the spectrum, awareness of interactions among single-nucleotide alterations and structural variations is increasing demand for follow-up diagnostic assays and enhancing the expectation for more comprehensive analysis. As whole-genome sequencing eventually becomes routine, the needed interpretive skills will change yet again.

New awareness of the widespread nature of this form of genomic variation reminds us of the ongoing need for healthy scepticism in diagnostic and predictive analyses – the simplest example being that apparent homozygosity for a SNP may in fact be hemizygosity, where both a SNP and a deletion are present in combination, but on different haplotypes. Only the haplotype containing the SNP would be detected by traditional genotyping assays, leading to the misclassification of the allele as homozygous. For example, in one case, a patient with cystic fibrosis (autosomal recessive) was apparently homozygous for the F508del mutation. This was unremarkable until the mother's sample tested negative for the same mutation; subsequently, she and the newborn were found to share a large deletion that encompassed the same exon (Ref. 139). In another, more complex example, a newborn with strong clinical evidence of cystic fibrosis was negative for all standard sequence-based mutation screens, and only with quantitative assays did the laboratory find large intragenic deletions on each of the patient's *CFTR* alleles (Ref. 140). Recent evidence from global newborn screening programmes demonstrates that larger intragenic deletions of *CFTR* may account for 1–3% of mutant chromosomes (Ref. 141), or more, as awareness permeates and appropriate screening assays for CNVs are invoked (Ref. 142).

For clinicians, CNVs have opened up analytical potential for clinical cases that had previously eluded diagnosis. This potential is creating a huge demand for laboratory tests that are still expensive, and very time-consuming to interpret. Nonetheless, when informative, such results may allow many patients and families the satisfaction of an explanation for their observed challenges, sometimes after years of fruitless investigations. These additional tools may allow earlier diagnosis for conditions, such as autism, for which early intervention in some individuals may be particularly beneficial.

Attention is drawn to genes of interest by virtue of their location in a newly recognised CNV [e.g. CHARGE syndrome (Ref. 143)], and this is opening a floodgate of research potential into complex disorders. Eventually, of course, we hope to find therapeutic prospects among such genes, and awareness of their involvement in a given phenotype is the first step. Particularly because of the tendency of larger CNVs to encompass contiguous genes, we are gaining insight into syndromology, with some improvement in explaining the spectrum of variation and the degree of consistency or inconsistency among phenotypic features. As illustrated by CNVs such as at 1q21.1 and 15q13.3 (discussed in this review), and more recently at 16p11.2 [in autism (Refs 144, 145, 146, 147), developmental delay (Ref. 148), schizophrenia (Ref. 149) and obesity (Ref. 150)], this opportunity can also be a Pandora's box. How CNV results are applied to research or medical decision-making needs to be weighted according to the circumstances where it is observed. For example, the relevance of the results will differ if a CNV is uncovered in (1) a known disease gene, (2) in a high-risk setting (such as during prenatal complications), (3) through a targeted list (such as an individual with

a family history of a disease or as confirmation of an existing clinical diagnosis), and (4) in a universal population screen (Refs 14, 47).

Conclusion: research in progress and outstanding research questions

Five years since the first rudimentary scans drew our attention to the widespread presence of genomic CNVs, they have become the focus for a myriad of surveys, both genotype- and phenotype-driven. Compendia such as the DGV are being refined and updated regularly. The apparent size of CNVs is decreasing as tools with enhanced resolution allow more precise definition of breakpoints, and annotation of precise copy number is becoming feasible (Ref. 151). The complexity of these data makes them somewhat recalcitrant, and the means for documentation in an unambiguous and functional way has been significantly challenging. Even more daunting, however, is annotation of the phenotypes of individuals who do and do not carry these variant genotypes, and finding ways to merge the plethora of disparate observations.

In addition, as technologies advance, the ability to detect CNVs in an individual genome increases. This is evidenced by recent diploid genome sequencing projects that find many CNVs that are unique to an individual (Refs 40, 41, 49, 50, 51, 52, 53, 54). Many of these CNVs are large and represent potentially pathogenic variation. As the cost of genome sequencing decreases, the prevalence of such studies will increase. Importantly, sequencing technologies have the potential of combining both SNP and CNV detection strategies into a single analysis, which should increase the power to detect variation that is related to phenotypes and disease.

As the HapMap project followed quickly on the heels of the first consensus genome sequence, so too has considerable effort moved to the study of how these structural variants are differentially distributed among global populations (Refs 13, 16, 152, 153, 154, 155, 156, 157). How does knowledge of such distributions influence the clinical interpretation of data? Can this information tell us anything about different environmental pressures to which these genomic alterations may have provided the means for adaptation? Are the phenotypic consequences different under different environmental circumstances?

Much has already been learned about mechanisms that underlie these genomic rearrangements, and the extent to which they are recurrent or randomly generated, inherited or de novo, and stable or unstable (Ref. 158). This has done little, however, to enlighten us about what determines whether a given CNV will have any pathogenic consequences, or will be associated with a pattern of features that might be recognised as a syndrome.

Major challenges

The connections between genomic observations and clinical implications are not straightforward, and involve complex network relationships. Single-gene disorders will continue to present themselves for medical attention, but the more prevalent and problematic conditions – heart disease, cancer, psychiatric and behavioural disorders, developmental delay, dysmorphic syndromes – require a shift in mindset from genetic-based to genomic-based. Candidate gene searches within a CNV (or group of CNVs) will need to progress to analyses of added dimensions, including gene and protein pathways and networks, for which sophisticated bioinformatics tools will be essential. Moreover, a more complete understanding of CNV and SNPs will be required to better empower genome-wide association studies (GWASs) of disease.

In our recent study, we explored whether CNVs might be plausible candidates for known complex trait associations from SNP-based GWASs. However, we found that CNVs might explain less than 5% of previously reported GWAS hits, suggesting that common CNVs are not likely to account for a large part of the 'missing heritability' (Ref. 45) from complex traits. These results also emphasise the need to consider all classes of variation (CNVs, other structural variants and SNPs, both common and rare) in order to maximise power to detect causal variation in disease association studies.

As formidable as data gathering may be for these CNVs of higher-order complexity, interpretation is far more of a challenge. Examples cited in this review provide some illustration of why particular caution is needed in moving between research findings and applications in a clinical context, in particular

The clinical context of copy number variation in the human genome

19

when studying complex disease. More studies examining CNV mutation rates across chromosomes and the effects of such events on gene dosage and the functional consequences would be beneficial. Undoubtedly, with accumulation of much more information, patterns will emerge to make some sense of what can currently seem, for some CNVs and phenotypes, like an uninterpretable mass of raw data.

We will be challenged to move beyond the obvious benefit of CNVs for explaining (diagnosing) various phenotypes to their utility in prediction and prognosis. A difficulty is that the plethora of CNV data can be provided as information, but without knowledge, and healthcare providers may be burdened for some time with the 'variant of unknown significance'.

Finally, the challenge will be not only to use our knowledge of these variants for explanation and prediction of medically relevant conditions but also to find ways to mitigate their untoward impact, for prevention or treatment of genomic disease. This will require a new level of inspired creativity.

References

1 Lejeune, J., Gautier, M. and Turpin, R. (1959) [Study of somatic chromosomes from 9 mongoloid children.]. Comptes rendus hebdomadaires des séances de l'Académie des sciences 248, 1721-1722 [Article in French]

2 Iafrate, A.J. et al. (2004) Detection of large-scale variation in the human genome. Nature Genetics 36, 949-951

3 Sebat, J. et al. (2004) Large-scale copy number polymorphism in the human genome. Science 305, 525-528

4 Ford, C.E. et al. (1959) A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). Lancet 1, 711-713

5 Jacobs, P.A. and Strong, J.A. (1959) A case of human intersexuality having a possible XXY sex-determining mechanism. Nature 183, 302-303

6 Feuk, L., Carson, A.R. and Scherer, S.W. (2006) Structural variation in the human genome. Nature Reviews Genetics 7, 85-97

7 Scherer, S.W. et al. (2007) Challenges and standards in integrating surveys of structural variation. Nature Genetics 39, S7-15

8 Lee, C., Iafrate, A.J. and Brothman, A.R. (2007) Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. Nature Genetics 39, S48-54

9 Conrad, D.F. et al. (2006) A high-resolution survey of deletion polymorphism in the human genome. Nature Genetics 38, 75-81

10 Khaja, R. et al. (2006) Genome assembly comparison identifies structural variants in the human genome. Nature Genetics 38, 1413-1418

11 Beckmann, J.S., Estivill, X. and Antonarakis, S.E. (2007) Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nature Reviews Genetics 8, 639-646

12 Wain, L.V., Armour, J.A. and Tobin, M.D. (2009) Genomic copy number variation, human health, and disease. Lancet 374, 340-350

13 Conrad, D.F. et al. (2009) Origins and functional impact of copy number variation in the human genome. Nature Oct 7; [Epub ahead of print]

14 Buchanan, J.A. and Scherer, S.W. (2008) Contemplating effects of genomic structural variation. Genetics in Medicine 10, 639-647

15 Varki, A., Geschwind, D.H. and Eichler, E.E. (2008) Explaining human uniqueness: genome interactions with environment, behaviour and culture. Nature Reviews Genetics 9, 749-763

16 Redon, R. et al. (2006) Global variation in copy number in the human genome. Nature 444, 444-454

17 Emanuel, B.S. and Shaikh, T.H. (2001) Segmental duplications: an 'expanding' role in genomic instability and disease. Nature Reviews Genetics 2, 791-800

18 Scherer, S.W. et al. (2003) Human chromosome 7: DNA sequence and biology. Science 300, 767-772

19 Fredman, D. et al. (2004) Complex SNP-related sequence variation in segmental genome duplications. Nature Genetics 36, 861-866

The clinical context of copy number variation in the human genome

20 Cheung, J. et al. (2003) Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. Genome Biology 4, R25

21 Kim, P.M. et al. (2008) Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history. Genome Research 18, 1865-1874

22 Stankiewicz, P. and Lupski, J.R. (2002) Genome architecture, rearrangements and genomic disorders. Trends in Genetics 18, 74-82

23 Lupski, J.R. et al. (1991) DNA duplication associated with Charcot-Marie-Tooth disease type 1A. Cell 66, 219-232

24 Emanuel, B.S. and Saitta, S.C. (2007) From microscopes to microarrays: dissecting recurrent chromosomal rearrangements. Nature Reviews Genetics 8, 869-883

25 Moore, J.K. and Haber, J.E. (1996) Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. Molecular and Cellular Biology 16, 2164-2173

26 Conrad, D.F. and Hurles, M.E. (2007) The population genetics of structural variation. Nature Genetics 39, S30-36

27 Zhang, F. et al. (2009) The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. Nature Genetics 41, 849-853

28 Lee, J.A., Carvalho, C.M. and Lupski, J.R. (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 131, 1235-1247

29 Carvalho, C.M. et al. (2009) Complex rearrangements in patients with duplications of MECP2 can occur by fork stalling and template switching. Human Molecular Genetics 18, 2188-2203

30 Hastings, P.J. et al. (2009) Mechanisms of change in gene copy number. Nature Reviews Genetics 10, 551-564

31 Hastings, P.J., Ira, G. and Lupski, J.R. (2009) A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genetics 5, e1000327

32 Koolen, D.A. et al. (2006) A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. Nature Genetics 38, 999-1001

33 Shaw-Smith, C. et al. (2006) Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. Nature Genetics 38, 1032-1037

34 Stefansson, H. et al. (2005) A common inversion under selection in Europeans. Nature Genetics 37, 129-137

35 Visser, R. et al. (2005) Identification of a 3.0-kb major recombination hotspot in patients with Sotos syndrome who carry a common 1.9-Mb microdeletion. American Journal of Human Genetics 76, 52-67

36 Osborne, L.R. et al. (2001) A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome. Nature Genetics 29, 321-325

37 Scherer, S.W. and Osborne, L.R. (2006) Williams-Beuren syndrome. In Genomic Disorders: The Genomic Basis of Disease (Lupski, J.R. and Stankiewicz, P., eds), pp. 221-236, Humana Press, Totowa, NJ, USA

38 McCarroll, S.A. et al. (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. Nature Genetics 40, 1166-1174

39 Zhang, J. et al. (2006) Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. Cytogenetic and Genome Research 115, 205-214

40 Levy, S. et al. (2007) The diploid genome sequence of an individual human. PLoS Biology 5, e254

41 Wheeler, D.A. et al. (2008) The complete genome of an individual by massively parallel DNA sequencing. Nature 452, 872-876

42 Lander, E.S. et al. (2001) Initial sequencing and analysis of the human genome. Nature 409, 860-921

43 International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. Nature 431, 931-945

44 Venter, J.C. et al. (2001) The sequence of the human genome. Science 291, 1304-1351

45 Maher, B. (2008) Personal genomes: the case of the missing heritability. Nature 456, 18-21

46 Manolio, T.A. et al. (2009) Finding the missing heritability of complex diseases. Nature 461, 747-753

47 Cook, E.H., Jr, and Scherer, S.W. (2008) Copy-number variations associated with neuropsychiatric conditions. Nature 455, 919-923

48 Carter, N.P. (2007) Methods and strategies for analyzing copy number variation using DNA microarrays. Nature Genetics 39, S16-21

The clinical context of copy number variation in the human genome

21

49 Wang, J. et al. (2008) The diploid genome sequence of an Asian individual. Nature 456, 60-65

50 Bentley, D.R. et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456, 53-59

51 Ahn, S.M. et al. (2009) The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. Genome Research 19, 1622-1629

52 Kim, J.I. et al. (2009) A highly annotated whole-genome sequence of a Korean individual. Nature 460, 1011-1015

53 McKernan, K.J. et al. (2009) Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. Genome Research 19, 1527-1541

54 Drmanac, R. et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 327, 78-81

55 Alkan, C. et al. (2009) Personalized copy number and segmental duplication maps using next-generation sequencing. Nature Genetics 41, 1061-1067

56 Chiang, D.Y. et al. (2009) High-resolution mapping of copy-number alterations with massively parallel sequencing. Nature Methods 6, 99-103

57 Lupski, J.R. (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends in Genetics 14, 417-422

58 Hüffmeier, U. et al. (2009) Replication of LCE3C-LCE3B CNV as a risk factor for psoriasis and analysis of interaction with other genetic risk factors. Journal of Investigative Dermatology Dec 17; [Epub ahead of print]

59 de Cid, R. et al. (2009) Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. Nature Genetics 41, 211-215

60 Shlien, A. et al. (2008) Excessive genomic DNA copy number variation in the Li-Fraumeni cancer predisposition syndrome. Proceedings of the National Academy of Sciences of the United States of America 105, 11264-11269

61 Diskin, S.J. et al. (2009) Copy number variation at 1q21.1 associated with neuroblastoma. Nature 459, 987-991

62 Lupski, J.R. (2007) Genomic rearrangements and sporadic disease. Nature Genetics 39, S43-47

63 Cahan, P. et al. (2009) The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells. Nature Genetics 41, 430-437

64 Henrichsen, C.N., Chaignat, E. and Reymond, A. (2009) Copy number variants, diseases and gene expression. Human Molecular Genetics 18, R1-8

65 Dathe, K. et al. (2009) Duplications involving a conserved regulatory element downstream of BMP2 are associated with brachydactyly type A2. American Journal of Human Genetics 84, 483-492

66 Stranger, B.E. et al. (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes.Science 315, 848-853

67 Merla, G. et al. (2006) Submicroscopic deletion in patients with Williams-Beuren syndrome influences expression levels of the nonhemizygous flanking genes. American Journal of Human Genetics 79, 332-341

68 Firth, H.V. et al. (2009) DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. American Journal of Human Genetics 84, 524-533

69 Miller, D.T. et al. (2009) Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. Journal of Medical Genetics 46, 242-248

70 Sharp, A.J. et al. (2008) A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. Nature Genetics 40, 322-328

71 O'Donovan, M.C., Kirov, G. and Owen, M.J. (2008) Phenotypic variations on the theme of CNVs. Nature Genetics 40, 1392-1393

72 Brunetti-Pierri, N. et al. (2008) Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. Nature Genetics 40, 1466-1471

73 Mefford, H.C. et al. (2008) Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. New England Journal of Medicine 359, 1685-1699

74 Carlson, C. et al. (1997) Molecular definition of 22q11 deletions in 151 velo-cardio-facial syndrome patients. American Journal of Human Genetics 61, 620-629

75 Driscoll, D.A. et al. (1992) Deletions and microdeletions of 22q11.2 in velo-cardio-facial syndrome. American Journal of Medical Genetics 44, 261-268

76 Coppinger, J. et al. (2009) Identification of familial and de novo microduplications of 22q11.21-q11.23 distal to the 22q11.21 microdeletion syndrome region. Human Molecular Genetics 18, 1377-1383

22

77 Klopocki, E. et al. (2007) Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome. American Journal of Human Genetics 80, 232-240

78 Uhrig, S. et al. (2007) Impact of array comparative genomic hybridization-derived information on genetic counseling demonstrated by prenatal diagnosis of the TAR (thrombocytopenia-absent-radius) syndrome-associated microdeletion 1q21.1. American Journal of Human Genetics 81, 866-868

79 Prior, T.W. (2007) Spinal muscular atrophy diagnostics. Journal of Child Neurology 22, 952-956

80 Schonherr, N. et al. (2007) The centromeric 11p15 imprinting centre is also involved in Silver-Russell syndrome. Journal of Medical Genetics 44, 59-63

81 Feuk, L. et al. (2006) Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia. American Journal of Human Genetics 79, 965-972

82 Bruder, C.E. et al. (2008) Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. American Journal of Human Genetics 82, 763-771

83 Piotrowski, A. et al. (2008) Somatic mosaicism for copy number variation in differentiated human tissues. Human Mutation 29, 1118-1124

84 Gervasini, C. et al. (2007) High frequency of mosaic CREBBP deletions in Rubinstein-Taybi syndrome patients and mapping of somatic and germ-line breakpoints. Genomics 90, 567-573

85 Roelfsema, J.H. and Peters, D.J. (2007) Rubinstein-Taybi syndrome: clinical and molecular overview. Expert Reviews in Molecular Medicine 9, 1-16

86 Schorry, E.K. et al. (2008) Genotype-phenotype correlations in Rubinstein-Taybi syndrome. American Journal of Medical Genetics Part A 146A, 2512-2519

87 Kozlowski, P. et al. (2007) Identification of 54 large deletions/duplications in TSC1 and TSC2 using MLPA, and genotype-phenotype correlations. Human Genetics 121, 389-400

88 Robinson, D.O. et al. (2008) Genetic analysis of chromosome 11p13 and the PAX6 gene in a series of 125 cases referred with aniridia. American Journal of Medical Genetics Part A 146A, 558-569

89 Kirov, G. et al. (2009) Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. Human Molecular Genetics 18, 1497-1503

90 Greenway, S.C. et al. (2009) De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. Nature Genetics 41, 931-935

91 Need, A.C. et al. (2009) A genome-wide investigation of SNPs and CNVs in schizophrenia. PLoS Genetics 5, e1000373

92 Lyle, R. et al. (2009) Genotype-phenotype correlations in Down syndrome identified by array CGH in 30 cases of partial trisomy and partial monosomy chromosome 21. European Journal of Human Genetics 17, 454-466

93 Korbel, J.O. et al. (2009) The genetic architecture of Down syndrome phenotypes revealed by high-resolution analysis of human segmental trisomies. Proceedings of the National Academy of Sciences of the United States of America 106, 12031-12036

94 Reisman, L.E. et al. (1966) Anti-mongolism. Studies in an infant with a partial monosomy of the 21 chromosome. Lancet 1, 394-397

95 Ewart, A.K. et al. (1993) Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. Nature Genetics 5, 11-16

96 Osborne, L.R. and Mervis, C.B. (2007) Rearrangements of the Williams-Beuren syndrome locus: molecular basis and implications for speech and language development. Expert Reviews in Molecular Medicine 9, 1-16

97 Lupski, J.R. (2009) Genomic disorders ten years on. Genome Medicine 1, 42

98 Somerville, M.J. et al. (2005) Severe expressive-language delay related to duplication of the Williams-Beuren locus. New England Journal of Medicine 353, 1694-1701

99 Kriek, M. et al. (2006) Copy number variation in regions flanked (or unflanked) by duplicons among patients with developmental delay and/or congenital malformations; detection of reciprocal and partial Williams-Beuren duplications. European Journal of Human Genetics 14, 180-189

100 Torniero, C. et al. (2008) Dysmorphic features, simplified gyral pattern and 7q11.23 duplication reciprocal to the Williams-Beuren deletion. European Journal of Human Genetics 16, 880-887

101 Sharp, A.J. et al. (2006) Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nature Genetics 38, 1038-1042

102 Christian, S.L. et al. (1999) Large genomic duplicons map to sites of instability in the Prader-Willi/Angelman syndrome chromosome region (15q11-q13). Human Molecular Genetics 8, 1025-1037

The clinical context of copy number variation in the human genome

103 Mignon-Ravix, C. et al. (2007) Recurrent rearrangements in the proximal 15q11-q14 region: a new breakpoint cluster specific to unbalanced translocations. European Journal of Human Genetics 15, 432-440

104 Sahoo, T. et al. (2005) Array-based comparative genomic hybridization analysis of recurrent chromosome 15q rearrangements. American Journal of Medical Genetics A 139A, 106-113

105 Sharp, A.J. et al. (2008) A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. Nature Genetics 40, 322-328

106 Stefansson, H. et al. (2008) Large recurrent microdeletions associated with schizophrenia. Nature 455, 232-236

107 Pagnamenta, A.T. et al. (2009) A 15q13.3 microdeletion segregating with autism. European Journal of Human Genetics 17, 687-692

108 Shinawi, M. et al. (2009) A small recurrent deletion within 15q13.3 is associated with a range of neurodevelopmental phenotypes. Nature Genetics 41, 1269-1271

109 Helbig, I. et al. (2009) 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. Nature Genetics 41, 160-162

110 Ben-Shachar, S. et al. (2009) Microdeletion 15q13.3: a locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. Journal of Medical Genetics 46, 382-388

111 Pagnamenta, A.T. et al. (2009) A 15q13.3 microdeletion segregating with autism. European Journal of Human Genetics 17, 687-692

112 van Bon, B.W. et al. (2009) Further delineation of the 15q13 microdeletion and duplication syndromes: a clinical spectrum varying from non-pathogenic to a severe outcome. Journal of Medical Genetics 46, 511-523

113 Consortium, I.S. (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 455, 237-241

114 Helbig, I. et al. (2009) 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. Nature Genetics 41, 160-162

115 Buchanan, J.A. et al. (2009) The cycle of genome-directed medicine. Genome Medicine 1, 16

116 Ali-Khan, S.E. et al. (2009) Whole genome scanning: resolving clinical diagnosis and management amidst complex data. Pediatric Research 66, 357-363

117 Christiansen, J. et al. (2004) Chromosome 1q21.1 contiguous gene deletion is associated with congenital heart disease. Circulation Research 94, 1429-1435

118 Szatmari, P. et al. (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nature Genetics 39, 319-328

119 Mefford, H.C. et al. (2009) A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease. Genome Research 19, 1579-1585

120 Pinto, D. et al. (2007) Copy-number variation in control population cohorts. Human Molecular Genetics 16 (Spec No. 2), R168-173

121 Armengol, L., Rabionet, R. and Estivill, X. (2008) The emerging role of structural variations in common disorders: initial findings and discovery challenges. Cytogenetic and Genome Research 123, 108-117

122 Schaschl, H., Aitman, T.J. and Vyse, T.J. (2009) Copy number variation in the human genome and its implication in autoimmunity. Clinical and Experimental Immunology 156, 12-16

123 Ionita-Laza, I. et al. (2009) Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. Genomics 93, 22-26

124 Hollox, E.J., Detering, J.C. and Dehnugara, T. (2009) An integrated approach for measuring copy number variation at the FCGR3 (CD16) locus. Human Mutation 30, 477-484

125 Nuytten, H. et al. (2009) Accurate determination of copy number variations (CNVs): application to the alpha- and beta-defensin CNVs. Journal of Immunological Methods 344, 35-44

126 McCarroll, S.A. et al. (2008) Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. Nature Genetics 40, 1107-1112

127 Gonzalez, E. et al. (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307, 1434-1440

128 Kulkarni, H. et al. (2008) CCL3L1-CCR5 genotype improves the assessment of AIDS Risk in HIV-1-infected individuals. PLoS One 3, e3165

129 Shostakovich-Koretskaya, L. et al. (2009) Combinatorial content of CCL3L and CCL4L gene copy numbers influence HIV-AIDS susceptibility in Ukrainian children. AIDS 23, 679-688

130 McKinney, C. et al. (2008) Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. Annals of the Rheumatic Diseases 67, 409-413

131 Colobran, R. et al. (2009) Copy number variation in the CCL4L gene is associated with susceptibility

The clinical context of copy number variation in the human genome

to acute rejection in lung transplantation. Genes and Immunity 10, 254-259

132 McCarroll, S.A. et al. (2009) Donor-recipient mismatch for common gene deletion polymorphisms in graft-versus-host disease. Nature Genetics 41, 1341-1344

133 Shlien, A. and Malkin, D. (2009) Copy number variations and cancer. Genome Medicine 1, 62

134 Marcel, V. et al. (2009) TP53 PIN3 and MDM2 SNP309 polymorphisms as genetic modifiers in the Li-Fraumeni syndrome: impact on age at first diagnosis. Journal of Medical Genetics 46, 766-772

135 Schwarzbraun, T. et al. (2009) Predictive diagnosis of the cancer prone Li-Fraumeni syndrome by accident: new challenges through whole genome array testing. Journal of Medical Genetics 46, 341-344

136 Adam, M.P. et al. (2009) Clinical utility of array comparative genomic hybridization: uncovering tumor susceptibility in individuals with developmental delay. Journal of Pediatrics 154, 143-146

137 Adams, S.A. et al. (2009) Impact of genotype-first diagnosis: the detection of microdeletion and microduplication syndromes with cancer predisposition by aCGH. Genetics in Medicine 11, 314-322

138 Speicher, M.R. and Carter, N.P. (2005) The new cytogenetics: blurring the boundaries with molecular biology. Nature Reviews Genetics 6, 782-792

139 Stuhrmann, M. et al. (2009) Testing the parents to confirm genotypes of CF patients is highly recommended: report of two cases. European Journal of Human Genetics 17, 417-419

140 Girardet, A. et al. (2007) Negative genetic neonatal screening for cystic fibrosis caused by compound heterozygosity for two large CFTR rearrangements. Clinical Genetics 72, 374-377

141 Tomaiuolo, R. et al. (2008) Epidemiology and a novel procedure for large scale analysis of CFTR rearrangements in classic and atypical CF patients: a multicentric Italian study. Journal of Cystic Fibrosis 7, 347-351

142 McDevitt, T. and Barton, D. (2009) When good CF tests go bad. European Journal of Human Genetics 17, 403-405

143 Vissers, L.E. et al. (2004) Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. Nature Genetics 36, 955-957

144 Fernandez, B.A. et al. (2009) Phenotypic spectrum associated with de novo and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of autism spectrum disorder. Journal of Medical Genetics Sep 24; [Epub ahead of print]

145 Kumar, R.A. et al. (2008) Recurrent 16p11.2 microdeletions in autism. Human Molecular Genetics 17, 628-638

146 Marshall, C.R. et al. (2008) Structural variation of chromosomes in autism spectrum disorder. American Journal of Human Genetics 82, 477-488

147 Weiss, L.A. et al. (2008) Association between microdeletion and microduplication at 16p11.2 and autism. New England Journal of Medicine 358, 667-675

148 Shinawi, M. et al. (2009) Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioral problems, dysmorphism, epilepsy, and abnormal head size. Journal of Medical Genetics Nov 12; [Epub ahead of print]

149 McCarthy, S.E. et al. (2009) Microduplications of 16p11.2 are associated with schizophrenia. Nature Genetics 41, 1223-1227

150 Bochukova, E.G. et al. (2010) Large, rare chromosomal deletions associated with severe early-onset obesity. Nature 463, 666-670

151 Perry, G.H. et al. (2008) The fine-scale and complex architecture of human copy-number variation. American Journal of Human Genetics 82, 685-695

152 Zogopoulos, G. et al. (2007) Germ-line DNA copy number variation frequencies in a large North American population. Human Genetics 122, 345-353

153 Jakobsson, M. et al. (2008) Genotype, haplotype and copy-number variation in worldwide human populations. Nature 451, 998-1003

154 Armengol, L. et al. (2009) Identification of copy number variants defining genomic differences among major human groups. PLoS One 4, e7230

155 Matsuzaki, H. et al. (2009) High resolution discovery and confirmation of copy number variants in 90 Yoruba Nigerians. Genome Biology 10, R125

156 Yim, S.H. et al. (2010) Copy number variations in East-Asian population and their evolutionary and functional implications. Human Molecular Genetics Jan 15; [Epub ahead of print]

157 Brookes, A.J. et al. (2009) Genomic variation in a global village: report of the 10th annual Human Genome Variation Meeting 2008. Human Mutation 30, 1134-1138

The clinical context of copy number variation in the human genome

158 Stankiewicz, P. and Lupski, J.R. (2010) Structural variation in the human genome and its role in disease. Annual Review of Medicine 61, 437-455

159 Ilbery, P.L., Lee, C.W. and Winn, S.M. (1961) Incomplete trisomy in a mongoloid child exhibiting minimal stigmata. Medical Journal of Australia 48, 182-184

160 Lejeune, J. et al. (1963) [3 cases of partial deletion of the short arm of a 5 chromosome.] Comptes rendus hebdomadaires des séances de l'Académie des sciences 257, 3098-3102 [Article in French]

161 Caspersson, T. et al. (1969) Chemical differentiation with fluorescent alkylating agents in Vicia faba metaphase chromosomes. Experimental Cell Research 58, 128-140

162 Orkin, S.H. (1978) The duplicated human alpha globin genes lie close together in cellular DNA. Proceedings of the National Academy of Sciences of the United States of America 75, 5950-5954

163 Wyman, A.R. and White, R. (1980) A highly polymorphic locus in human DNA. Proceedings of the National Academy of Sciences of the United States of America 77, 6754-6758

164 Bauman, J.G. et al. (1980) A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA. Experimental Cell Research 128, 485-490

165 Van Prooijen-Knegt, A.C. et al. (1982) In situ hybridization of DNA sequences in human metaphase chromosomes visualized by an indirect fluorescent immunocytochemical procedure. Experimental Cell Research 141, 397-407

166 Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985) Individual-specific 'fingerprints' of human DNA. Nature 316, 76-79

167 Monaco, A.P. et al. (1985) Detection of deletions spanning the Duchenne muscular dystrophy locus using a tightly linked DNA segment. Nature 316, 842-845

168 Ray, P.N. et al. (1985) Cloning of the breakpoint of an X;21 translocation associated with Duchenne muscular dystrophy. Nature 318, 672-675

169 Schmickel, R.D. (1986) Contiguous gene syndromes: a component of recognizable syndromes. Journal of Pediatrics 109, 231-241

170 Kallioniemi, A. et al. (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 258, 818-821

171 [No authors listed] (1996) A complete set of human telomeric probes and their clinical application. National Institutes of Health and Institute of Molecular Medicine collaboration. Nature Genetics 14, 86-89

172 Pinkel, D. et al. (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nature Genetics 20, 207-211

173 Stockley, T.L. et al. (2006) Strategy for comprehensive molecular testing for Duchenne and Becker muscular dystrophies. Genetic Testing 10, 229-243

174 White, S.J. and den Dunnen, J.T. (2006) Copy number variation in the genome; the human DMD gene as an example. Cytogenetic and Genome Research 115, 240-246

175 De Luca, A. et al. (2007) Deletions of NF1 gene and exons detected by multiplex ligation-dependent probe amplification. Journal of Medical Genetics 44, 800-808

176 Raedt, T.D. et al. (2006) Conservation of hotspots for recombination in low-copy repeats associated with the NF1 microdeletion. Nature Genetics 38, 1419-1423

177 Wimmer, K. et al. (2006) Spectrum of single- and multiexon NF1 copy number changes in a cohort of 1,100 unselected NF1 patients. Genes, Chromosomes & Cancer 45, 265-276

178 Saugier-Veber, P. et al. (2007) Heterogeneity of NSD1 alterations in 116 patients with Sotos syndrome. Human Mutation 28, 1098-1107

179 Fagali, C. et al. (2009) MLPA analysis in 30 Sotos syndrome patients revealed one total NSD1 deletion and two partial deletions not previously reported. European Journal of Medical Genetics 52, 333-336

180 Woodward, K.J. (2008) The molecular and cellular defects underlying Pelizaeus-Merzbacher disease. Expert Reviews in Molecular Medicine 10, e14

181 Brouwers, N. et al. (2006) Genetic risk and transcriptional variability of amyloid precursor protein in Alzheimer's disease. Brain 129, 2984-2991

182 Rovelet-Lecrux, A. et al. (2006) APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nature Genetics 38, 24-26

183 Sleegers, K. et al. (2006) APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy. Brain 129, 2977-2983

184 Durand, C.M. et al. (2007) Mutations in the gene encoding the synaptic scaffolding protein

SHANK3 are associated with autism spectrum disorders. Nature Genetics 39, 25-27

185 Moessner, R. et al. (2007) Contribution of SHANK3 mutations to autism spectrum disorder. American Journal of Human Genetics 81, 1289-1297

186 Gauthier, J. et al. (2009) Novel de novo SHANK3 mutation in autistic patients. American Journal of Medical Genetics Part B, Neuropsychiatric Genetics 150B, 421-424

187 Fantes, J. et al. (1995) Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. Human Molecular Genetics 4, 415-422

188 Klopocki, E. et al. (2008) A microduplication of the long range SHH limb regulator (ZRS) is associated with triphalangeal thumb-polysyndactyly syndrome. Journal of Medical Genetics 45, 370-375

189 Sun, M. et al. (2008) Triphalangeal thumb-polysyndactyly syndrome and syndactyly type IV are caused by genomic duplications involving the long range, limb-specific SHH enhancer. Journal of Medical Genetics 45, 589-595

190 Fellermann, K. et al. (2006) A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. American Journal of Human Genetics 79, 439-448

191 Hollox, E.J. et al. (2008) Defensins and the dynamic genome: what we can learn from structural variation at human chromosome band 8p23.1. Genome Research 18, 1686-1697

192 Aitman, T.J. et al. (2006) Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature 439, 851-855

193 Fanciulli, M. et al. (2007) FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. Nature Genetics 39, 721-723

194 Ibanez, P. et al. (2009) Alpha-synuclein gene rearrangements in dominantly inherited parkinsonism: frequency, phenotype, and mechanisms. Archives of Neurology 66, 102-108

195 Burns, J.C. et al. (2005) Genetic variations in the receptor-ligand pair CCR5 and CCL3L1 are important determinants of susceptibility to Kawasaki disease. Journal of Infectious Diseases 192, 344-349

196 Saitta, S.C. et al. (2004) Aberrant interchromosomal exchanges are the predominant cause of the 22q11.2 deletion. Human Molecular Genetics 13, 417-428

197 Shaikh, T.H. et al. (2007) Low copy repeats mediate distal chromosome 22q11.2 deletions: sequence analysis predicts breakpoint mechanisms. Genome Research 17, 482-491

198 Berg, J.S. et al. (2007) Speech delay and autism spectrum behaviors are frequently associated with duplication of the 7q11.23 Williams-Beuren syndrome region. Genetics in Medicine 9, 427-441

199 Cusco, I. et al. (2008) Copy number variation at the 7q11.23 segmental duplications is a susceptibility factor for the Williams-Beuren syndrome deletion. Genome Research 18, 683-694

200 Grisart, B. et al. (2009) 17q21.31 microduplication patients are characterised by behavioural problems and poor social interaction. Journal of Medical Genetics 46, 524-530

201 Kirchhoff, M. et al. (2007) A 17q21.31 microduplication, reciprocal to the newly described 17q21.31 microdeletion, in a girl with severe psychomotor developmental delay and dysmorphic craniofacial features. European Journal of Medical Genetics 50, 256-263

202 Shaffer, L.G. et al. (2007) The discovery of microdeletion syndromes in the post-genomic era: review of the methodology and characterization of a new 1q41q42 microdeletion syndrome. Genetics in Medicine 9, 607-616

203 Ballif, B.C. et al. (2007) Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2. Nature Genetics 39, 1071-1073

204 Ghebranious, N. et al. (2007) A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation. American Journal of Medical Genetics Part A 143A, 1462-1471

205 Ballif, B.C. et al. (2008) Expanding the clinical phenotype of the 3q29 microdeletion syndrome and characterization of the reciprocal microduplication. Molecular Cytogenetics 1, 8

206 Goobie, S. et al. (2008) Molecular and clinical characterization of de novo and familial cases with microduplication 3q29: guidelines for copy number variation case reporting. Cytogenetic and Genome Research 123, 65-78

207 Potocki, L. et al. (2007) Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism

The clinical context of copy number variation in the human genome

27

phenotype. American Journal of Human Genetics 80, 633-649

208 Tabor, H.K. and Cho, M.K. (2007) Ethical implications of array comparative genomic hybridization in complex phenotypes: points to consider in research. Genetics in Medicine 9, 626-631

209 Ullmann, R. et al. (2007) Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. Human Mutation 28, 674-682

210 Schaefer, G.B. and Mendelsohn, N.J. (2008) Genetics evaluation for the etiologic diagnosis of autism spectrum disorders. Genetics in Medicine 10, 4-12

211 Glessner, J.T. et al. (2009) Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature 459, 569-573

212 Abrahams, B.S. and Geschwind, D.H. (2008) Advances in autism genetics: on the threshold of a new neurobiology. Nature Reviews Genetics 9, 341-355

213 Lachman, H.M. et al. (2007) Increase in GSK3beta gene copy number variation in bipolar disorder. American Journal of Medical Genetics Part B, Neuropsychiatric Genetics 144B, 259-265

214 Burmeister, M., McInnis, M.G. and Zollner, S. (2008) Psychiatric genetics: progress amid controversy. Nature Reviews Genetics 9, 527-540

215 Alaerts, M. and Del-Favero, J. (2009) Searching genetic risk factors for schizophrenia and bipolar disorder: learn from the past and back to the future. Human Mutation 30, 1139-1152

216 Walsh, T. et al. (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 320, 539-543

217 Xu, B. et al. (2008) Strong association of de novo copy number mutations with sporadic schizophrenia. Nature Genetics 40, 880-885

218 Rujescu, D. et al. (2009) Disruption of the neurexin 1 gene is associated with schizophrenia. Human Molecular Genetics 18, 988-996

219 Hughes, A.E. et al. (2006) A common CFH haplotype, with deletion of CFHR1 and CFHR3, is associated with lower risk of age-related macular degeneration. Nature Genetics 38, 1173-1177

220 Maller, J. et al. (2006) Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. Nature Genetics 38, 1055-1059

221 Barber, J.C. et al. (2008) 8p23.1 duplication syndrome; a novel genomic condition with unexpected complexity revealed by array CGH. European Journal of Human Genetics 16, 18-27

222 Hendrickson, B.C. et al. (2009) Differences in SMN1 allele frequencies among ethnic groups within North America. Journal of Medical Genetics 46, 641-644

223 Alias, L. et al. (2009) Mutation update of spinal muscular atrophy in Spain: molecular characterization of 745 unrelated patients and identification of four novel mutations in the SMN1 gene. Human Genetics 125, 29-39

224 Mantripragada, K.K. et al. (2009) Genome-wide high-resolution analysis of DNA copy number alterations in NF1-associated malignant peripheral nerve sheath tumors using 32K BAC array. Genes, Chromosomes & Cancer 48, 897-907

Further reading, resources and contacts

The Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources (DECIPHER) provides tools that allow researchers to share information about copy number changes in patients:

> https://decipher.sanger.ac.uk/

The Database of Genomic Variants (DGV) provides a resource for researchers looking for known or previously identified genomic variation. It documents variation found in population control samples:

> http://projects.tcag.ca/variation/

Online Mendelian Inheritance in Man is a comprehensive resource that describes associations between human genes and genetic phenotypes. It contains referenced information focusing on the link between genotype and phenotype, and contains entries on all known mendelian disorders and over 12 000 genes:

> http://www.ncbi.nlm.nih.gov/omim/

The clinical context of copy number variation in the human genome

Features associated with this article

Figures
Figure 1. Forms of genomic copy number variation.
Figure 2. Ways by which copy number variation can cause disease.
Figure 3. Complexities of de novo and inherited copy number variation.
Figure 4. Approaches to clinical investigation.

Tables
Table 1. History and milestones in human copy number variation research.
Table 2. Spectrum of copy number variation genotypes and illustrative phenotypes.

Box
Box 1. Terminology.

Citation details for this article

Charles Lee and Stephen W. Scherer (2010) The clinical context of copy number variation in the human genome. Expert Rev. Mol. Med. Vol. 12, e8, March 2010, doi:10.1017/S1462399410001390

The clinical context of copy number variation in the human genome