

# Implications of copy number variation in people with chromosomal abnormalities: potential for greater variation in copy number state may contribute to variability of phenotype

Adam J. de Smith · Anne L. Trewick ·  
Alexandra I. F. Blakemore

Received: 14 January 2010/Revised: 21 July 2010/Accepted: 26 July 2010/Published online: 10 August 2010  
© Springer Science+Business Media B.V. 2010

**Abstract** Copy number variation is common in the human genome with many regions, overlapping thousands of genes, now known to be deleted or amplified. Aneuploidies and other forms of chromosomal imbalance have a wide range of adverse phenotypes and are a common cause of birth defects resulting in significant morbidity and mortality. “Normal” copy number variants (CNVs) embedded within the regions of chromosome imbalance may affect the clinical outcomes by altering the local copy number of important genes or regulatory regions: this could alleviate or exacerbate certain phenotypes. In this way CNVs may contribute to the clinical variability seen in many disorders caused by chromosomal abnormalities, such as the congenital heart defects (CHD) seen in ~40% of Down’s syndrome (DS) patients. Investigation of CNVs may therefore help to pinpoint critical genes or regulatory elements, elucidating the molecular mechanisms underlying these conditions, also shedding light on the aetiology of such phenotypes in people without major chromosome imbalances, and ultimately leading to their improved detection and treatment.

**Keywords** Copy number variation · Chromosomal abnormalities · Aneuploidy · Phenotypic variability

## Introduction

One of the fastest-growing research areas in genetics in the past few years has been the investigation of structural

variation in the human genome, with copy number variants (CNVs) found to be much more common than previously imagined. As well as contributing to interindividual phenotypic variation in the general population, these genomic variants may hold the key to understanding the differences in severity seen in people with chromosomal abnormalities. Aneuploidy, for example, can be regarded as a state of large-scale changed copy number resulting from loss or gain of an entire chromosome, arising due to non-disjunction during cell division in either meiosis I or II. Aneuploidies and other chromosomal abnormalities are a common cause of birth defects and are associated with significant morbidity and mortality. The spectrum of phenotypes seen in each affected individual is clearly dependant on the particular chromosomal region concerned, although there is still a great deal of variability in the presentation of phenotypes.

Individuals with Down’s syndrome (DS), for example, like those with other aneuploidies or with chromosomal deletions or duplications, vary greatly in their clinical features, capabilities, disabilities and prognosis. Although many of these people, with the necessary social and clinical support, can and do lead active and fulfilling lives, some DS sub-phenotypes are associated with increased morbidity and mortality in infancy. Congenital heart defects (CHD), for example, occur in ~40% of DS; 50% of these require surgical correction within the first year of life and survival to one year is only ~76% compared with ~91% for non-CHD DS patients (Yang et al. 2002). An appreciation of the molecular mechanisms behind this phenotypic variation and the identification of susceptibility loci may provide diagnostic and prognostic markers to enable us to predict the occurrence of clinically serious phenotypes, such as CHD, in DS. Additionally, analysis of these loci might help elucidate the mechanisms of CHD in non-DS children.

A. J. de Smith (✉) · A. L. Trewick · A. I. F. Blakemore  
Department of Genomics of Common Disease,  
School of Public Health and Epidemiology,  
Imperial College London, Hammersmith Hospital Campus,  
Du Cane Road, London, UK  
e-mail: adam.de-smith@imperial.ac.uk

## Copy number variation and its effects on human phenotype

To assess the implications of genomic copy number variants in the context of chromosomal abnormalities, it is helpful to consider the evidence that they affect phenotype in euploid individuals. Until recently, it was assumed that there are two copies of every gene in euploid individuals, with one on the maternally-inherited chromosome and the other inherited paternally. Any two human genomes were thought to be 99.9% similar, with the main source of genetic variation due to single nucleotide polymorphisms (SNPs). Technological advances often result in paradigm shifts, however, and an unexpected amount of copy number variation was revealed through the advent of array comparative genomic hybridisation (array CGH) and, more recently, of second-generation sequencing technologies. There are currently over 14,000 CNV regions listed on the TCAG Database of Genomic Variants (DGV, updated March 2010), predicted to overlap a significant proportion of the genome. Approximately 11,500 genes, including over 2600 listed in the Online Mendelian Inheritance in Man database, are reportedly overlapped by CNVs (DGV March 2010).

To understand the role of CNVs in human disease, it is important to consider how these genomic variants may affect phenotype. Deletion or duplication of dosage-sensitive genes, or their regulatory regions, could have adverse phenotypic effects, and copy number changes correlate

with expression of affected genes (Henrichsen et al. 2009; Stranger et al. 2007). The majority of common CNVs are much smaller than the subset investigated by Stranger et al. (2007), leading them to propose that a substantial proportion of heritable variation in gene expression may be explained by copy number variation. In addition to direct copy number effects, CNVs may also exert positional effects, for example by shifting them closer to or further away from heterochromatin. Furthermore, adverse phenotypic effects may result from interactions between CNVs and other variants, for example a deletion on one chromosome may unmask a harmful recessive allele on the other. Conversely, a genomic duplication that incorporates a harmful mutation may have a double gain-of-function effect: a duplication of the *PRSS1* and *PRSS2* genes along with a missense mutation was recently shown to cause hereditary pancreatitis (Masson et al. 2008).

It is likely, therefore, that CNVs do significantly affect human phenotypes, and this is reflected by an increasing number of associations reported between CNVs and disease (Tables 1, 2). Several autoimmune disorders have been associated with CNVs, including systemic lupus erythematosus (Fanciulli et al. 2007), Crohn's disease (Fellermann et al. 2006; McCarroll et al. 2008) and psoriasis (Hollox et al. 2008), and a range of other diseases have now been associated with CNVs (reviewed in Zhang et al. 2009). It has been postulated that CNVs could account for a large proportion of the missing heritability of common diseases that has emerged after the recent plethora of genome-wide

**Table 1** Common CNVs associated with disease

Disease	Chromosomal locus	Genes	Risk-associated copy number change	Reference
HIV/AIDS	17q12	<i>CCL3L1</i>	Low copy numbers (<2 in European- and Hispanic-Americans; <4 in African-Americans)	Gonzalez et al. (2005)
Glomerulonephritis	1q23	<i>FCGR3B</i>	Low copy numbers (<2)	Aitman et al. (2006)
Systemic lupus erythematosus (SLE)	1q23	<i>FCGR3B</i>	Low copy numbers (<2)	Fanciulli et al. (2007)
	6p21.3	<i>C4</i>	Low copy numbers (<3)	Yang et al. (2007)
Crohn's disease	8p23.1	Beta-Defensins	Low copy numbers (<4)	Fellermann et al. (2006)
	5q33.1	<i>IRGM</i>	Low copy numbers (<2)	McCarroll et al. (2008)
Psoriasis	8p23.1	Beta-Defensins	High copy numbers (>4)	Hollox et al. (2008)
Osteoporosis	4q13.2	<i>UGT2B17</i>	Presence of gene (>0 copies)	Yang et al. (2008)
Graft-vs.-host disease (GVHD)	4q13.2	<i>UGT2B17</i>	Absence of gene (0 copies)	McCarroll et al. (2009)
Neuroblastoma	1q21.1	Novel neuroblastoma breakpoint family ( <i>NBPF</i> ) gene	Low copy numbers (<2)	Diskin et al. (2009)

**Table 2** Rare CNVs associated with disease

Disease	Chromosomal locus	Genes	Risk-associated copy number change	Reference
Parkinson's disease	4q22.1	<i>SNCA</i>	Triplication (4 copies)	Singleton et al. (2003)
Autosomal dominant early-onset Alzheimer's Disease	21q21	<i>APP</i>	Duplication (3 copies)	Rovelet-Lecrux et al. (2006)
Hereditary pancreatitis	7q34	<i>PRSS1</i> and <i>PRSS2</i>	Triplication (4 copies)	Le Marechal et al. (2006)
Autism	Multiple	Multiple	Deletions and duplications	Sebat et al. (2007) and Glessner et al. (2009)
Bipolar disorder	3q13.3	<i>GSK3β</i>	Duplication	Lachman et al. (2007)
Schizophrenia	Multiple	Multiple	Deletions and duplications	Xu et al. (2008) and Walsh et al. (2008)
Tetralogy of fallot	Multiple	Multiple	Deletions and duplications	Greenway et al. (2009)
Obesity	Multiple	Multiple	Deletions and duplications	Bochukova et al. (2010)
	16p11.2	Multiple	Deletion	Walter et al. (2010)

association studies (GWAS) (reviewed in Manolio et al. 2009), but this is the subject of heated controversy. In a recent large-scale genotyping survey of common CNVs, Conrad et al. (2010) found that the CNVs that they were able to genotype easily were only in linkage disequilibrium with 34 out of 1,554 trait-associated SNPs from GWAS, thus they suggest that common CNVs are unlikely to account for much of the missing heritability. There is, however, a particular challenge in analysis of amplified sequences (which may be remote from the original sequence copy (Conrad et al. 2010)), multi-allelic variants (particularly VNTRs) and recurrent events, so further analysis of those CNV types is required before the full contribution of common CNVs to human disease can be assessed.

As an alternative to the common disease-common variants hypothesis (Chakravarti 1999; Lander 1996), the rare variants hypothesis postulates that a collection of many, individually less frequent copy number changes may collectively significantly contribute to disease susceptibility. Although many CNVs are present in an appreciable proportion of the population, the majority are likely to be less frequent: for example, in our array CGH investigation of 50 apparently healthy French male samples, 809 out of 1469 (~55%) multi-probe CNV regions were identified in only one individual (de Smith et al. 2007). Rare copy number changes have been implicated in neurodevelopmental disorders, including autism (Marshall et al. 2008; Sebat et al. 2007) and schizophrenia (Wilson et al. 2006; Xu et al. 2008). Furthermore, a number of rare genomic structural variants have recently been associated with obesity (Bochukova et al. 2010; Walters et al. 2010), and many more low frequency CNVs with strong effects may contribute to common disease phenotypes: large sample cohorts and different populations will need to be investigated to uncover these rare variants.

### CNVs in aneuploidy

Thus, it is now clear that genomic structural variants may significantly impact phenotype in euploid individuals: these effects may be even more pronounced in people with chromosomal abnormalities. For example, a phenotype associated with a particular monosomy or sub-chromosomal deletion may be ameliorated by the presence of an amplification CNV on the unaffected chromosome, as gene expression within the CNV region could remain at a normal, euploid level. Conversely, a deletion CNV could cause a more severe phenotype.

An example could be Turner syndrome, which is the only whole chromosome monosomy (45, X) that is viable in humans, with approximately 1 in 2,000 female births having one copy of the X chromosome (Nielsen and Wohlert 1990). In addition to the main features of short stature and ovarian failure, present in almost all cases, there are many other phenotypes that may present, including a short webbed neck, kidney malformations, hearing problems and various learning difficulties, as well as increased risk of type 1 diabetes (Gravholt et al. 1998). It is possible that the extensive copy number variation on the X chromosome—37.8% according to the DGV (March 2010)—may contribute to phenotypic variability in this disorder. For example, a critical region for the neurocognitive deficits in Turner syndrome was mapped to Xp22.3, containing 31 genes (Ross et al. 2000), and several CNV loci have been reported in this region.

Similarly, phenotypic variation in genomic disorders could also be a reflection of underlying copy number variation within the relevant genomic region. The most common microdeletion in humans is DiGeorge syndrome (or velo-cardio-facial syndrome) caused by deletion of a 1.5–3 Mb region at chromosome 22q11.2 (Scambler et al.

1992) and occurring in 1 in 5,000 births (Botto et al. 2003). This disorder is also associated with variable phenotypes, such as cleft palate, congenital heart disease (Shprintzen et al. 1978), renal anomalies (Czarnecki et al. 1998) and increased risk of schizophrenia (Murphy et al. 1999). Several CNV loci have been identified within the 22q11.2 deletion region in apparently healthy individuals: of interest, two of the candidate genes for the schizophrenia phenotype, *PRODH* (Li et al. 2004) and *GNBIL* (Williams et al. 2008), are overlapped by a number of CNVs, and these could potentially modify this phenotype in DiGeorge patients.

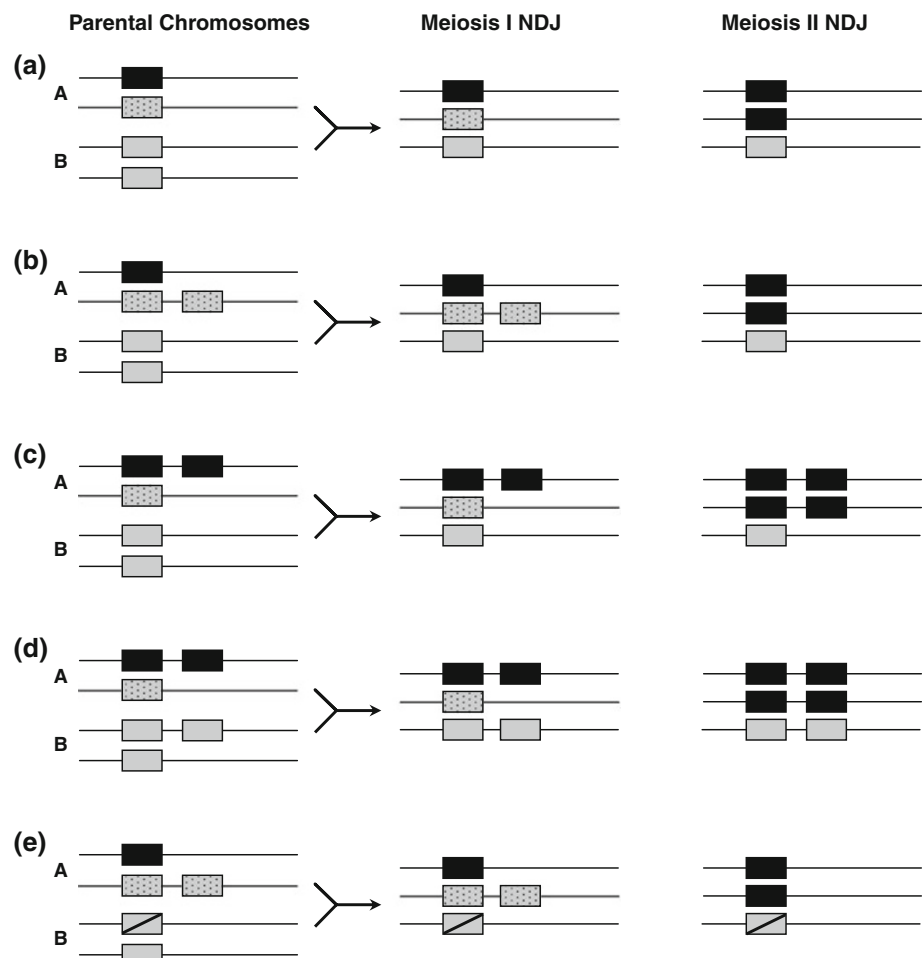
In polyploidy or sub-chromosomal duplications where abnormalities result from extra copies of the chromosome regions, deletion CNVs may “normalise” copy number. Alternatively, there is potential for even greater amplification of copy number where amplification CNVs are present. A simple duplication CNV on the non-disjoining chromosome may lead to the presence of up to six copies of a gene in a trisomic individual (Fig. 1) and the subsequent phenotypic effects associated with increased gene expression levels. Conversely, a gene deletion on the non-

disjoining chromosome may alleviate the expected effect of trisomy by reducing the local copy number state to one or two copies. Thus, the compounded effect of aneuploidy and copy number variation has the potential to generate a much wider range of phenotypes than would be expected on the basis of chromosome number alone. It has been suggested, therefore, that CNVs are likely to act as ‘modifiers of the phenotypic variability of trisomies’ (Beckmann et al. 2007) and the same is probably also true for smaller chromosomal imbalances.

### Phenotypic variability in Down’s syndrome may be modified by CNVs

Using Down’s Syndrome as an example, an extra copy of chromosome 21 is not sufficient to cause the full range of phenotypes associated with this disorder, as individuals can present with a range of sub-phenotypes. Some features are almost always present, such as the characteristic facial appearance and mental retardation, although these can vary widely in the severity of their presentation (Kallen et al.

**Fig. 1** Mechanisms of phenotypic variability in trisomy resulting from copy number variation. Amplification of gene copy number during non-disjunction (NDJ) can result in a range of copy number states. Where dosage-sensitive genes are involved, this may result in a wide range of associated phenotypes. This figure illustrates a number of different potential outcomes for a single gene which is susceptible to duplication and/or deletion. **a** illustrates the difference between meiosis I (MI) and meiosis II (MII) NDJ occurring in parent A; **b–d** illustrate just three of the possible scenarios resulting from NDJ in parent A if gene duplications are present in one or both parents, each of which could result in over-expression of a dosage-sensitive gene; **e** illustrates how a deletion on parent B and MII NDJ in parent A, however, that leads to a total of two copies of the gene in the trisomic child which could potentially ameliorate the pathological effects of trisomy



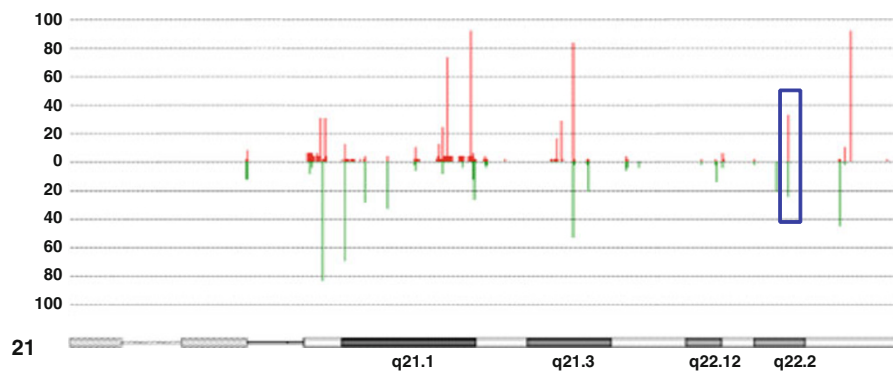
1996). Other features, however, are only present in a fraction of DS cases: for example, congenital heart defects (CHD) are present in ~40% of cases (Park et al. 1977) and gastrointestinal (GI) defects in ~8% of DS (Epstein et al. 1991). Leukaemia is also common in DS, with an ~20% increased risk of acute lymphoblastic leukaemia (ALL) (Hasle et al. 2000), and transient myeloproliferative disorder occurring in ~10% of DS newborns, of which 10–20% develop acute megakaryoblastic leukaemia (AMKL) before the age of 4 (reviewed in Hasle 2001). A detailed knowledge of the genes on the affected chromosome is required to understand the phenotypic effects of aneuploidy. In addition, it is important to determine the overall effects of gene dosage imbalance, which may also be complex.

Individuals with partial trisomy 21 resulting from unbalanced chromosomal translocations will only exhibit those features associated with the extra genomic material present. Molecular analysis of these patients enables ‘phenotypic mapping’ by defining the critical genomic regions that harbour genes associated with various phenotypes (Epstein et al. 1991). As with all regions of the genome, many common CNVs and rarer genomic structural variants are found on chromosome 21. Two recent reports describe structural variations in which the presence of increased copies of genomic regions on chromosome 21 contributes to human phenotype. In one study, duplication of a region at 21q21 that included the amyloid precursor protein (*APP*) gene was found to cause familial Alzheimer’s disease in five separate families (Rovelet-Lecrux et al. 2006). The duplication varied in size from 0.58 to 6.37 Mb and contained from 5 to 12 annotated genes including *APP*. Despite the largest of these duplications being over 6 Mb, none of the families exhibited any clinical evidence of DS (Cabrejo et al. 2006). A second duplication of a 4.3 Mb region at 21q22.13–q22.2,

containing just over 30 genes, did help refine the map of the DS critical region (DCR) as it caused a DS phenotype in three family members (Ronan et al. 2007). These individuals had the facial characteristics of DS and mild cognitive disability indicating that this region includes part of the DCR but not all of it. It is apparent, therefore, that different sub-phenotypes of trisomy 21 are caused by increased copy number at different regions across chromosome 21. It is also evident that some DS individuals who are trisomic for a particular critical region do not always exhibit the associated phenotype, or display a much milder form.

A complex interplay of molecular factors is likely to be responsible for the trisomic phenotypes, possibly involving non-chromosome 21 genes. Furthermore, interaction of trisomy with the presence of certain embedded deleterious alleles or haplotypes may contribute to the variable presentation of the different phenotypes (including degree of mental retardation, congenital heart defects, Hirschprung and other gut diseases, and leukaemia). The increasing amount of copy number variation that has been discovered in the human genome, specifically on chromosome 21, adds another dimension to the molecular consequences of trisomy: it is possible that the phenotypic variability seen in DS may be due to CNVs on this chromosome.

On the DGV, there are 507 reported CNV calls covering 35.0% of chromosome 21, which is similar to the CNV coverage of other chromosomes (33.9%) (DGV March 2010). Many of these are common in the population. For example, our investigation of only 50 healthy French male samples revealed 20 multi-probe CNVs, and an additional 57 single probe CNV signals, on chromosome 21 (de Smith et al. 2007): almost half were detected in multiple samples (47%), and 34% had a frequency of >5% (Fig. 2). In addition, these CNVs overlapped 38 known genes. There were several CNVs within *TIAMI* (T-lymphoma invasion



**Fig. 2** Copy number variation on chromosome 21. This shows the penetrance plot of aberration calls on chromosome 21 from array CGH data from 50 apparently healthy French Caucasian males (de Smith et al. 2007), generated using ADM-2 algorithm with threshold 4 in CGH Analytics 3.4. This displays the percentage of the 50 arrays

that have an aberration at each probe position along the chromosome: deletion aberrations point downwards from 0, and amplifications point upwards. The *box* shows the position of the CNV within *DSCAM*



and metastasis-inducing protein 1), for example, which is expressed in almost all analysed tumour cell lines, including B- and T-lymphomas, melanomas and carcinomas (Habets et al. 1995). A CNV was also found within *DSCAM*, which is a good candidate gene for the CHD phenotype seen in DS, as it lies within the predicted minimum critical region and is expressed in the heart during cardiac development (Korbel et al. 2009). Several other CNVs have since been discovered in this gene, as listed on the DGV, including a common deletion that incorporates 4 exons (Matsuzaki et al. 2009).

Other genes within the CHD critical region also overlap CNVs: a common intronic deletion (10%) was identified in *C2CD2* (Conrad et al. 2010); 6 CNVs have been reported within *PRDM15*, including one deletion overlapping 5 exons that was found in 3/90 Yoruban individuals (Matsuzaki et al. 2009); and one deletion was identified in *BACE2* overlapping the last exon of this gene (Mills et al. 2006).

CNVs also overlap *RUNX1*, with one deletion removing the promoter region and first exon of this gene (Gusev et al. 2009). *RUNX1* codes for a transcription factor involved in haematopoiesis (North et al. 2002) and is associated with

AMKL, which is estimated to be 500-fold more frequent in children with DS than in the general population (reviewed in Zipursky et al. 1992): CNVs in this gene may, therefore, protect against or increase susceptibility to leukaemia in DS. These CNV loci and other regions that could potentially impact variable phenotypes of trisomy 21 are listed in Table 3.

Analysis of copy number variation on chromosome 21 may, therefore, lead to a better understanding of the etiology of DS sub-phenotypes and could possibly shed light on the origin of these conditions in the wider non-trisomic population. Knowledge of which genes are involved in the generation of CHD in DS, for example, and how they interact with other gene products, could have many implications for managing CHD in the non-trisomic population: for example, adult onset CHD may be prevented by therapeutic agents that target critical pathways controlling cardiogenesis, and new approaches may be developed for stem-cell guided cardiac repair (reviewed in Passier et al. 2008).

Since the discovery of trisomy 21, it has been hypothesised that genes present in three copies are over-expressed

**Table 3** Selection of potentially interesting CNV loci on chromosome 21 that may affect phenotypic variability in trisomy 21

Chromosome 21 location (hg18)	Genes	Associated disease/phenotype	Exonic	Copy number loss/gain	Frequency in CNV studies (samples with CNV/total samples in study)	Reference
26,173,965–26,174,696	<i>APP</i>	Alzheimer's disease	No	Gain	1/1	McKernan et al. (2009)
26,196,591–26,196,901			No	Loss	1/1	Wang et al. (2008)
31,337,700–31,424,326	<i>TIAMI</i>	Associated with several different cancers	Exons 27–29	Gain	1/50	de Smith et al. (2007)
35,304,204–35,704,075	<i>RUNX1</i>	Acute megakaryoblastic leukaemia	Exon 1	Loss	6/2906	Gusev et al. (2009)
36,748,886–36,783,426	<i>CLDN14</i>	Hearing loss	Exons 1–3	Loss/Gain	Loss 1/50 Gain 1/50	de Smith et al. (2007)
36,785,973–36,801,336			No	Loss/Gain	Loss 6/90 Gain 6/90	Matsuzaki et al. (2009)
34,648,096–34,829,283	<i>KCNE1</i> <i>KCNE2</i> <i>FAM165</i> <i>BRCAN1</i>	Down syndrome critical region (DCR)	All exons of <i>KCNE1</i> , <i>KCNE2</i> , <i>FAM165B</i> , and exons 2–4 of <i>RCAN1</i>	Gain	5/1190	Zogopoulos et al. (2007)
37,092,300–37,354,491	<i>DSCR6</i> <i>HLCS</i>	DCR	All exons of <i>DSCR6</i> and exons 1–7 of <i>HLCS</i>	Gain	1/776	Pinto et al. (2007)
37,489,474–37,490,172	<i>TTC3</i>	DCR	Exon 41	Gain	1/1	Kim et al. (2009)
40,466,951–40,489,050	<i>DSCAM</i>	Congenital heart defects (CHD)	Exons 12–15	Loss	9/90	Matsuzaki et al. (2009)
40,592,263–40,596,079			No	Loss	1/40	Conrad et al. (2010)
41,568,947–41,572,987	<i>BACE2</i>	CHD	Exon 9	Loss	1/36	Mills et al. (2006)
42,123,984–42,136,309	<i>PRDM15</i>	CHD	Exons 14–18	Loss	3/90	Matsuzaki et al. (2009)
42,223,189–42,226,495	<i>C2CD2</i>	CHD	No	Loss	45/450	Conrad et al. (2010)
46,234,806–46,234,806	<i>COL6A1</i>	CHD	No	Gain	1/1	Levy et al. (2007)
46,367,207–46,399,800	<i>COL6A2</i>	CHD	Exons 21–28	Gain	2/2026	Shaikh et al. (2009)

by 1.5-fold relative to the euploid state, but this has subsequently been shown to not always be the case. Some genes are over expressed, some are expressed at euploid levels while other genes appear to be down-regulated (Li et al. 2006; Prandini et al. 2007). This may, of course, reflect adaptive regulatory control, but may also reflect the effects of local copy number variation on chromosome 21, which will result in deviations from the predicted three copies of each gene, as shown in Fig. 1.

## Conclusions

Copy number variation may, therefore, make a significant contribution to phenotypic variation in aneuploidy and other chromosomal abnormalities. Targeted analysis of such variation might, therefore, represent an effective strategy for refining critical regions and pinpointing genes central to key pathways (i.e. of cardiogenesis or oncogenesis). This would also help to enable the prediction of which individuals may develop serious phenotypes, such as CHD in DS, which may add an extra dimension to antenatal screening programmes in the future by contributing to the process of ‘informed choice’ by the potential parents of a DS child, as well as well as aiding the long term management of these patients. Additional benefits may also accrue from increased understanding of pathogenic mechanisms relevant to patients with similar phenotypes outside the context of chromosomal imbalance.

In summary, recent technological advances, such as high-resolution array CGH and second-generation sequencing, have vastly and rapidly increased our knowledge of copy number variation in the human genome. As these techniques continue to improve, and greater numbers of samples and populations are investigated, so our understanding of how copy number correlates with phenotype will grow, yielding information not just relevant to individuals with chromosomal imbalances, but to similar conditions in the wider population.

## References

- Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, Mangion J, Robertson-Lowe C, Marshall AJ, Petretto E et al (2006) Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* 439:851–855
- Beckmann JS, Estivill X, Antonarakis SE (2007) Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet* 8:639–646
- Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K, Saeed S, Hamilton-Shield J, Clayton-Smith J, O’Rahilly S et al (2010) Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* 463:666–670
- Botto LD, May K, Fernhoff PM, Correa A, Coleman K, Rasmussen SA, Merritt RK, O’Leary LA, Wong LY, Elixson EM et al (2003) A population-based study of the 22q11.2 deletion: phenotype, incidence, and contribution to major birth defects in the population. *Pediatrics* 112:101–107
- Cabrejo L, Guyant-Marechal L, Laquerriere A, Vercelletto M, De la Fourniere F, Thomas-Anterion C, Verry C, Letournel F, Pasquier F, Vital A et al (2006) Phenotype associated with APP duplication in five families. *Brain* 129:2966–2976
- Chakravarti A (1999) Population genetics—making sense out of sequence. *Nat Genet* 21:56–60
- Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P et al (2010) Origins and functional impact of copy number variation in the human genome. *Nature* 464:704–712
- Czarnecki PM, Van Dyke DL, Vats S, Feldman GL (1998) A mother with VCFS and unilateral dysplastic kidney and her fetus with multicystic dysplastic kidneys: additional evidence to support the association of renal malformations and VCFS. *J Med Genet* 35:348
- Database of Genomic Variants—a curated catalogue of structural variation in the human genome. <http://projects.tcag.ca/variation/>
- de Smith AJ, Tsalenko A, Sampas N, Scheffer A, Yamada NA, Tsang P, Ben-Dor A, Yakhini Z, Ellis RJ, Bruhn L et al (2007) Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males: implications for association studies of complex diseases. *Hum Mol Genet* 16:2783–2794
- Diskin SJ, Hou C, Glessner JT, Attiyeh EF, Laudenslager M, Bosse K, Cole K, Mosse YP, Wood A, Lynch JE et al (2009) Copy number variation at 1q21.1 associated with neuroblastoma. *Nature* 459:987–991
- Epstein CJ, Korenberg JR, Anneren G, Antonarakis SE, Ayme S, Courchesne E, Epstein LB, Fowler A, Groner Y, Huret JL et al (1991) Protocols to establish genotype-phenotype correlations in Down syndrome. *Am J Hum Genet* 49:207–235
- Fanciulli M, Norsworthy PJ, Petretto E, Dong R, Harper L, Kamesh L, Heward JM, Gough SC, de Smith A, Blakemore AI et al (2007) FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat Genet* 39:721–723
- Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, Wehkamp J, Bevins CL, Reinisch W, Teml A, Schwab M, Lichter P 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. *Am J Hum Genet* 79:439–448
- Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP et al (2009) Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 459:569–573
- Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ et al (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307:1434–1440
- Gravholt CH, Juul S, Naeraa RW, Hansen J (1998) Morbidity in Turner syndrome. *J Clin Epidemiol* 51:147–158
- Greenway SC, Pereira AC, Lin JC, DePalma SR, Israel SJ, Mesquita SM, Ergul E, Conta JH, Korn JM, McCarroll SA et al (2009) De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. *Nat Genet* 41:931–935
- Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, Breslow JL, Friedman JM, Pe’er I (2009) Whole population, genome-wide mapping of hidden relatedness. *Genome Res* 19:318–326
- Habets GG, van der Kammen R, Stam JC, Michiels F, Collard JG (1995) Sequence of the human invasion-inducing *TIAM1* gene, its conservation in evolution and its expression in tumor cell lines of different tissue origin. *Oncogene* 10:1371–1376

- Hasle H (2001) Pattern of malignant disorders in individuals with Down's syndrome. *Lancet Oncol* 2:429–436
- Hasle H, Clemmensen IH, Mikkelsen M (2000) Risks of leukaemia and solid tumours in individuals with Down's syndrome. *Lancet* 355:165–169
- Henrichsen CN, Vinckenbosch N, Zollner S, Chaignat E, Pradervand S, Schutz F, Ruedi M, Kaessmann H, Reymond A (2009) Segmental copy number variation shapes tissue transcriptomes. *Nat Genet* 41:424–429
- Hollox EJ, Huffmeier U, Zeeuwen PL, Palla R, Lascorz J, Rodijk-Olthuis D, van de Kerkhof PC, Traupe H, de Jongh G, den Heijer M et al (2008) Psoriasis is associated with increased beta-defensin genomic copy number. *Nat Genet* 40:23–25
- Kallen B, Mastroiacovo P, Robert E (1996) Major congenital malformations in Down syndrome. *Am J Med Genet* 65:160–166
- Kim JI, Ju YS, Park H, Kim S, Lee S, Yi JH, Mudge J, Miller NA, Hong D, Bell CJ et al (2009) A highly annotated whole-genome sequence of a Korean individual. *Nature* 460:1011–1015
- Korbel JO, Tirosch-Wagner T, Urban AE, Chen XN, Kasowski M, Dai L, Grubert F, Erdman C, Gao MC, Lange K et al (2009) The genetic architecture of Down syndrome phenotypes revealed by high-resolution analysis of human segmental trisomies. *Proc Natl Acad Sci USA* 106:12031–12036
- Lachman HM, Pedrosa E, Petruolo OA, Cockerham M, Papolos A, Novak T, Papolos DF, Stopkova P (2007) Increase in GSK3beta gene copy number variation in bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet* 144B:259–265
- Lander ES (1996) The new genomics: global views of biology. *Science* 274:536–539
- Le Marechal C, Masson E, Chen JM, Morel F, Ruszniewski P, Levy P, Ferec C (2006) Hereditary pancreatitis caused by triplication of the trypsinogen locus. *Nat Genet* 38:1372–1374
- Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G et al (2007) The diploid genome sequence of an individual human. *PLoS Biol* 5:e254
- Li T, Ma X, Sham PC, Sun X, Hu X, Wang Q, Meng H, Deng W, Liu X, Murray RM et al (2004) Evidence for association between novel polymorphisms in the PRODH gene and schizophrenia in a Chinese population. *Am J Med Genet B Neuropsychiatr Genet* 129B:13–15
- Li CM, Guo M, Salas M, Schupf N, Silverman W, Zigman WB, Husain S, Warburton D, Thaker H, Tycko B (2006) Cell type-specific over-expression of chromosome 21 genes in fibroblasts and fetal hearts with trisomy 21. *BMC Med Genet* 7:24
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A et al (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753
- Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y et al (2008) Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 82:477–488
- Masson E, Le Marechal C, Delcenserie R, Chen JM, Ferec C (2008) Hereditary pancreatitis caused by a double gain-of-function trypsinogen mutation. *Hum Genet* 123:521–529
- Matsuzaki H, Wang H, Hu J, Rava R, Fu GK (2009) High resolution discovery and confirmation of copy number variants in 90 Yoruba Nigerians. *Genome Biol* 10:R125
- McCarroll SA, Huett A, Kuballa P, Chilewski SD, Landry A, Goyette P, Zody MC, Hall JL, Brant SR, Cho JH et al (2008) Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat Genet* 40:1107–1112
- McCarroll SA, Bradner JE, Turpeinen H, Volin L, Martin PJ, Chilewski SD, Antin JH, Lee SJ, Ruutu T, Storer B et al (2009) Donor-recipient mismatch for common gene deletion polymorphisms in graft-versus-host disease. *Nat Genet* 41:1341–1344
- McKernan KJ, Peckham HE, Costa GL, McLaughlin SF, Fu Y, Tsung EF, Clouser CR, Duncan C, Ichikawa JK, Lee CC et al (2009) Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res* 19:1527–1541
- Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, Devine SE (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res* 16:1182–1190
- Murphy KC, Jones LA, Owen MJ (1999) High rates of schizophrenia in adults with velo-cardio-facial syndrome. *Arch Gen Psychiatry* 56:940–945
- Nielsen J, Wohler M (1990) Sex chromosome abnormalities found among 34,910 newborn children: results from a 13 year incidence study in Arhus, Denmark. *Birth Defects Orig Artic Ser* 26:209–223
- North TE, de Bruijn MF, Stacy T, Talebian L, Lind E, Robin C, Binder M, Dzierzak E, Speck NA (2002) Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. *Immunity* 16:661–672
- Park SC, Mathews RA, Zuberbuhler JR, Rowe RD, Neches WH, Lenox CC (1977) Down syndrome with congenital heart malformation. *Am J Dis Child* 131:29–33
- Passier R, van Laake LW, Mummery CL (2008) Stem-cell-based therapy and lessons from the heart. *Nature* 453:322–329
- Pinto D, Marshall C, Feuk L, Scherer SW (2007) Copy-number variation in control population cohorts. *Hum Mol Genet* 16(Spec No. 2):R168–R173
- Prandini P, Deutsch S, Lyle R, Gagnebin M, Delucinge Vivier C, Delorenzi M, Gehrig C, Descombes P, Sherman S, Dagna Bricarelli F et al (2007) Natural gene-expression variation in Down syndrome modulates the outcome of gene-dosage imbalance. *Am J Hum Genet* 81:252–263
- Ronan A, Fagan K, Christie L, Conroy J, Nowak NJ, Turner G (2007) Familial 4.3 Mb duplication of 21q22 sheds new light on the Down syndrome critical region. *J Med Genet* 44:448–451
- Ross JL, Roeltgen D, Kushner H, Wei F, Zinn AR (2000) The Turner syndrome-associated neurocognitive phenotype maps to distal Xp. *Am J Hum Genet* 67:672–681
- Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerriere A, Vital A, Dumanchin C, Feuillet S, Brice A, Vercelletto M et al (2006) APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet* 38:24–26
- Scambler PJ, Kelly D, Lindsay E, Williamson R, Goldberg R, Shprintzen R, Wilson DI, Goodship JA, Cross IE, Burn J (1992) Velo-cardio-facial syndrome associated with chromosome 22 deletions encompassing the DiGeorge locus. *Lancet* 339:1138–1139
- Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J et al (2007) Strong association of de novo copy number mutations with autism. *Science* 316:445–449
- Shaikh TH, Gai X, Perin JC, Glessner JT, Xie H, Murphy K, O'Hara R, Casalunovo T, Conlin LK, D'Arcy M et al (2009) High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Res* 19:1682–1690
- Shprintzen RJ, Goldberg RB, Lewin ML, Sidoti EJ, Berkman MD, Argamaso RV, Young D (1978) A new syndrome involving cleft palate, cardiac anomalies, typical facies, and learning disabilities: velo-cardio-facial syndrome. *Cleft Palate J* 15:56–62
- Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, Hulihan M, Peuralinna T, Dutra A, Nussbaum R et al (2003) alpha-Synuclein locus triplication causes Parkinson's disease. *Science* 302:841



- Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C et al (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315:848–853
- Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A et al (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320:539–543
- Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J, Falchi M, Chen F, Andrieux J, Lobbens S et al (2010) A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* 463:671–675
- Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Guo Y et al (2008) The diploid genome sequence of an Asian individual. *Nature* 456:60–65
- Williams NM, Glaser B, Norton N, Williams H, Pierce T, Moskvina V, Monks S, Del Favero J, Goossens D, Rujescu D et al (2008) Strong evidence that GNB1L is associated with schizophrenia. *Hum Mol Genet* 17:555–566
- Wilson GM, Flibotte S, Chopra V, Melnyk BL, Honer WG, Holt RA (2006) DNA copy-number analysis in bipolar disorder and schizophrenia reveals aberrations in genes involved in glutamate signaling. *Hum Mol Genet* 15:743–749
- Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M (2008) Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* 40:880–885
- Yang Q, Rasmussen SA, Friedman JM (2002) Mortality associated with Down's syndrome in the USA from 1983 to 1997: a population-based study. *Lancet* 359:1019–1025
- Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, Zhou B, Hebert M, Jones KN, Shu Y, Kitzmiller K et al (2007) Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet* 80:1037–1054
- Yang TL, Chen XD, Guo Y, Lei SF, Wang JT, Zhou Q, Pan F, Chen Y, Zhang ZX, Dong SS et al (2008) Genome-wide copy-number-variation study identified a susceptibility gene, UGT2B17, for osteoporosis. *Am J Hum Genet* 83:663–674
- Zhang F, Gu W, Hurles ME, Lupski JR (2009) Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet* 10:451–481
- Zipursky A, Poon A, Doyle J (1992) Leukemia in Down syndrome: a review. *Pediatr Hematol Oncol* 9:139–149
- Zogopoulos G, Ha KC, Naqib F, Moore S, Kim H, Montpetit A, Robidoux F, Laflamme P, Cotterchio M, Greenwood C et al (2007) Germ-line DNA copy number variation frequencies in a large North American population. *Hum Genet* 122:345–353