

## Editorial

Structural and copy number variants  
in the human genome: implications  
for psychiatry

David St Clair

**Summary**

Copy number variants are small chromosomal deletions and duplications. When they alter the dose of genes critical for normal brain development and adult brain functioning they may cause severe disorders such as autism and schizophrenia. Numerous such loci have recently been

identified. They are offering amazing leads for neuropsychiatric research.

**Declaration of interest**

None.

David St Clair is Professor of Psychiatry at Aberdeen University; he is also Adjunct Professor at Columbia University and Advisory Professor at Shanghai Jiao Tong University.

It is over 50 years since traditional cytogenetics using light microscopy first showed that variations in chromosome structure could be associated with disease in humans. The earliest neuropsychiatric disorder implicated was Down syndrome caused by duplication of the whole of chromosome 21. Chromosome deletions and duplications need to be around two million DNA base pairs in length to be routinely visible by light microscopy and so until recently the vast majority have gone undetected.

**Recent technological advances in sequencing**

Over the past decade high-resolution methods for whole genome interrogation have been developed. The most widely used are single nucleotide polymorphism (SNP) microarrays. These techniques allow DNA to be interrogated for the presence or absence of the SNPs present on the microarray. Almost as a by-product larger DNA variants can also be detected when a string of SNPs on the same chromosome are deleted or duplicated. The terms structural variants or copy number variants (CNVs) are used to refer to DNA variants intermediate in size between those observable by light microscopy and those detectable by traditional sequencing (1–700 base pairs). Some define the lower cut-off size limits as one or more kilobases: others allow any variant greater than 50 base pairs in length.<sup>1</sup> They are remarkably common making up around 13% of the human genome.<sup>2</sup> Most of the published data on CNVs in neuropsychiatric disorders have employed techniques with a resolution of around 30 kilobases, but within the past 12 months data are being reported from arrays with a lower limit of around 3 kilobases. Finally, so-called next-generation sequencing technologies can now sequence and analyse an individual's whole complement of around three billion base pairs of DNA for a few hundred dollars. Progress in DNA sequencing power has been truly staggering and indeed overtaken Moore's famous law that computing power doubles and price halves every year. There are still, however, technical and computational challenges before next-generation sequencing can be routinely applied to investigate genome-wide human structural variants. These challenges are being met and will soon be overcome.<sup>3</sup>

The vast majority of the many hundreds of thousands of structural variants and CNVs are rare and harmless; they only usually have an impact on human health when they bring about a change in dosage/ expression of a critical gene and even then they are rarely more than modestly penetrant. Unless the breakpoints at their boundaries disrupt a gene, they do not routinely affect gene splicing or protein structure. The bigger the CNV, the more likely statistically that it will cause a change in gene dosage, but the event itself is not size dependent. A small deletion or duplication of a key piece of coding region or regulatory machinery can still have a large clinical impact. For this reason it is likely that many of the lessons we have learned over the past few years from studying the hitherto easier to identify larger CNVs will also apply when the clinical and molecular effects of the much more numerous, smaller CNVs are characterised.

**Copy number variant loci implicated  
in neuropsychiatric disorders**

Over the past 5 years a group of seminal papers have reported that compared with normal control groups the rates of large rare (but not of common) CNVs are enriched in the main neurodevelopmental disorders, namely schizophrenia, autism and non-syndromic intellectual disability.<sup>4–7</sup> The picture in bipolar disorder and recurrent major depression, disorders whose neurodevelopmental origins are less clear-cut, are not markedly different from the general population.

The penetrance of these rare pathogenic CNVs is variable. This makes it difficult to be certain whether at the level of the individual patient any particular CNV is pathogenic or is an incidental finding with other disturbances causing the disorder. The number of CNV loci for which we have good statistical evidence of association with neuropsychiatric disorders is small. They include large recurrent CNVs at chromosome 1q21, 15q11.2, 15q13.3, 16p11.2, 16p13.1, 22q12 and Neurexin1. Hundreds more have been reported but they are usually too rare to obtain good statistical evidence of an association. Probably 300 loci have already been implicated in autism with perhaps many hundreds more to come.<sup>8</sup> For this reason it is important to set up international registries to collate and catalogue new findings so that over the years a clearer picture of the evidence implicating or otherwise any individual rare CNV can accumulate.

Apart from the Neurexin1 locus all the loci mentioned above are flanked by long stretches of low copy repeats (LCRs) and, by a

process called non-allelic homologous recombination between neighbouring LCRs, are liable to mutate and generate either deletions or duplications. The mutation rate is very high, around ten times that of single bases. This means that new or *de novo* mutations at these loci are entering human populations at a very high rate, but given the severity of the phenotypes and the reduced fecundity associated with the CNVs, they are under negative selection and seldom survive in families for more than a few generations before dying out.<sup>9</sup> The highly penetrant 16p11.2 and 22q12 loci associated with autism, schizophrenia and intellectual disability are most likely to be *de novo*; others are usually familial and passed down one or two generations. Most of the CNVs flanked by LCRs have probably now been identified; the mechanisms causing the much more numerous CNV mutations at the remaining loci not flanked by LCRs are less well understood.

Many CNV loci involve multiple genes and the gene or genes responsible for the associated phenotypes have not yet been determined. On the other hand the main categories of genes affected by CNVs that predispose to neuropsychiatric disorders are now pretty well defined. A high proportion are involved in neurodevelopment in early life and neuronal integrity, connectivity, signalling and synaptic plasticity in the adult. Some such as *Neurexin1* are large and have multiple functions in both fetal and adult brain. They are providing scientists with a cornucopia of leads into the molecular pathology of neurodevelopmental disorders.

### Significance of findings from CNV studies

Perhaps the most intriguing findings to emerge from these CNV studies is the extensive and unexpected overlap of the neuropsychiatric phenotypes associated with both deletions and duplications at individual CNV loci. At the 16p11.2 locus the deletion is more associated with autism, whereas the duplication is associated with schizophrenia. Intriguingly, by contrast deletions of 7q11.2 regions are associated with Williams syndrome, a main feature of which is a hypersocial personality, whereas reciprocal duplications are associated with autism.<sup>10</sup> Epigenetic changes may also determine the clinical features of the phenotype. The classic neurodevelopmental example, originally identified by cytogenetics, is the Prader Willi and Angelman syndrome locus on chromosome 15q13. Both syndromes are caused by the same deletion on chromosome 15, but methylation changes silencing the remaining paternally or maternally inherited chromosome determines the phenotype.

The environmental and genetic factors that predispose to non-recurrent CNVs are poorly understood. There is some evidence

they happen more often in recently and rapidly evolving parts of the genome. Environmental influences usually first disturb the dynamic epigenetic and semi-stable elements of the genome that in turn can have an impact upon and cause traditional genetic mutations.

These recent CNV findings have opened up enormous new opportunities for research into disorders of neurodevelopment. By contrast the implications for clinical psychiatry are likely to be modest. Genetic screening using microarray technologies is now routine, and affected families can be offered more informed genetic counselling. It will be many years, however, before the findings can be translated into better and more effective methods for prevention and treatment of these disorders.

**David St Clair**, BA, MPhil, DM, FRCPsych, FRSE, Division of Applied Medicine, University of Aberdeen, Fosterhill, Aberdeen, UK. Email: d.stclair@abdn.ac.uk

First received 24 Feb 2012, accepted 15 Mar 2012

### References

- 1 Alkan C, Coe B, Eichler E. Genome structural variation and genotyping. *Nat Rev Genet* 2011; **12**: 363–75.
- 2 Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, et al. Mapping and sequencing structural variation in eight human genomes. *Nature* 2008; **453**: 56–64.
- 3 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**: 1061–73.
- 4 Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong association of *de novo* copy number mutations with autism. *Science* 2007; **316**: 445–9.
- 5 Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, et al. Association between microdeletion and microduplication at 16p11.2 and autism. *N Eng J Med* 2008; **358**: 667–75.
- 6 Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, Steinberg S, et al. Large recurrent deletions associated with schizophrenia. *Nature* 2008; **455**: 232–6.
- 7 International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 2008; **445**: 237–41.
- 8 Levy D, Ronemus M, Yamrom B, Lee YH, Leotta A, Kendall J, et al. Rare *de novo* and transmitted copy number variation in autistic spectrum disorders. *Neuron* 2011; **70**: 885–97.
- 9 St Clair D. Copy number variation and schizophrenia. *Schizophr Bull* 2009; **35**: 9–12.
- 10 Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, et al. Multiple recurrent *de novo* CNVs including duplications at the 7q11.23 Williams syndrome region are strongly associated with autism. *Neuron* 2011; **70**: 863–85.