

Copy Number Variation Project Matthew Hurles, Chris Tyler-Smith and Nigel Carter

Introduction

Genetic diseases are caused by several types of variation in DNA sequences. We are investigating gains and losses of DNA sequence consisting of between five hundred and five million letters, known as Copy Number Variation (CNV). This type of variant has often been overlooked in previous surveys of human genomic variation. We do not know what proportion of genetic disease is caused by CNV, but there are a growing number of diseases ranging from Spinal Muscular Atrophy to lupus to malaria in which CNV plays an important role in determining an individual's risk. These previous findings have been essentially serendipitous; a more systematic assessment of the medical importance of CNV requires a comprehensive map of locations in the human genome that are prone to this form of variation. We are using a range of different genomic technologies to map CNVs throughout the human genome, and we are assessing the biological impact of this form of variation in various ways, including disease association studies. Importantly, we are placing all our data on publicly-accessible websites so that the entire scientific community can benefit from our research.

A first-generation map of copy number variation

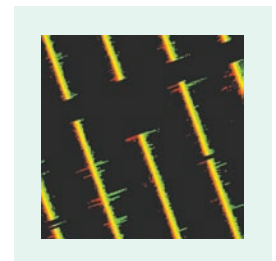
Using array-based methods, we have generated and published a first-generation map of CNV

throughout the human genome based on investigating 270 apparently healthy individuals with ancestry from three continents: Europe, Africa and East Asia. We have shown that CNV is widespread, extensive and complex, and that gains and losses occur with roughly equal frequency. In collaboration with Panos Deloukas and Manolis Dermitzakis we have annotated this map, by identifying which of these CNVs influence the expression levels of nearby genes. We have also identified CNVs whose unusual continental distribution suggests that they may have played a recent role in human evolution and adaptation to environmental challenges. Some of these 'population-genetic outliers' are already known to be involved in genetic disease.

Ongoing work

Most of the CNVs we have identified in the first-generation map are greater than 50,000 bases in size, and yet we know that most CNVs are smaller than 20,000 bases in size. Consequently, we are generating a comprehensive high-resolution map of all common CNV in Europe and Africa. To achieve this level of resolution we are using both array-based methods and new sequencing technologies. We are also investigating DNA copy number and gene expression differences between humans and our closest primate relatives to investigate the role that CNV has played in the origins of humans, and surveying CNV in a larger set of

worldwide populations. Finally, with our genome-wide map of CNV in hand, we are now systematically identifying CNVs that influence traits such as height and weight as well as the risk of suffering from common diseases such as diabetes and cardiovascular disease.



Regions of extensive common copy number variable can be identified from their distinctively enriched pattern of red and green dots, representing losses and gains of DNA copy number, respectively.

Selected Publications

- Redon R *et al.* (2006) Global variation in copy number in the human genome. *Nature* **444**: 444–54
- Stranger BE *et al.* (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**: 848–53
- Conrad DF *et al.* (2006) A high-resolution survey of deletion polymorphism in the human genome. *Nat. Genet.* **38**: 75–81 (2006)
- Fiegler H *et al.* (2006) Accurate and reliable high-throughput detection of copy number variation in the human genome. *Genome Res* **16**: 1566–74.
- Komura D *et al.* (2006) Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. *Genome Res* **16**: 1575–84.
- Khaja R *et al.* (2006) Genome assembly comparison identifies structural variants in the human genome. *Nat. Genet.* **38**: 1413–8.