Preface

Copy number variants (CNVs) are generally held to be DNA sequences ≥ 1 kilobase in length that are present in a variable number of copies in the genome. The identification of copy number variation as an important source of human genomic diversity was heralded by *Science* as the 'breakthrough of the year' in 2007. It should however be appreciated that CNVs are, qualitatively at least, not an entirely new phenomenon. Indeed, the antecedents of what we have come to know as CNVs are to be found among the whole gene deletions and duplications first reported, either as polymorphisms or pathological mutations, back in the early 1980s. What is new, however, is the realization that such duplications/deletions are not only rather abundant in the human genome but they also often occur at polymorphic frequencies. Thus, two human genomes may differ by more than 20 megabases of DNA and it is likely that the full extent of CNV still remains to be discovered. Nearly 3000 genes are already known to be associated with CNVs, some 10% of the human gene complement. This high degree of variability with regard to gene copy number between two individuals challenges definitions of the wild-type and 'normality'. An indication of how important and far reaching the CNV revolution has been for human molecular genetics comes from the wide range of specialties to which CNVs have already contributed: evolutionary biology, inherited disease (monogenic and complex), cancer genetics, inter-individual variation in diverse human populations, and even somatic mosaicism within normal healthy individuals. Indeed, there is scarcely any discipline in human genetics that remains untouched by the advent of CNVs.

Despite the impressive advances in our understanding of CNVs made over the last few years, many issues remain unresolved. Some questions are procedural (i.e. inter-study definitional differences as to what constitutes a CNV), others are more practical arising from the use of different analytical platforms each with its own set of methodological limitations, or consequent to the generally small sizes of the human population samples which have been evaluated to date. Thus, the extent of coverage of the human genome by CNVs is still largely unknown. In part this is a direct con-

sequence of the lack of standardization, but it is also due to the lack of detailed characterization, by DNA sequencing, of the boundaries of identified CNVs in the vast majority of cases. High resolution breakpoint mapping is clearly a prerequisite for the accurate assessment of CNV size, the identification of the genes and regulatory elements affected, the determination of the consequences for gene expression and their phenotypic *sequelae*, as well as the elucidation of the nature of the molecular mechanisms underlying CNV formation. For most known CNVs, their frequency in diverse human populations is still unknown and may differ quite widely. Appreciation of the full nature and extent of copy number variation, and its frequency in different ethnic populations, awaits the construction of a comprehensive high-resolution CNV map of the human genome that approaches saturation coverage.

The extent to which CNVs are likely to contribute to the diversity of human phenotypes, including 'single gene defects' and genomic disorders, is increasingly being recognized. Indeed, CNVs are now being widely recruited to genome-wide association studies with the aim of assessing their influence on human disease causation/susceptibility. Such programs must however be supported by large-scale population studies involving samples from both controls and patients. It is already clear that CNVs occur as a continuum stretching from 'neutral' polymorphisms on the one hand to directly pathogenetic copy number changes on the other. Between these two extremes may lie those CNVs that are capable of acting as predisposing (or protective) factors in relation to complex disease. This notwithstanding, the identification of CNVs that contribute to the etiology of complex disease remains a very significant challenge. Whereas some CNVs are easy to tag with HapMap SNPs, others (in particular those of a complex or multiallelic nature) are not in linkage disequilibrium with neighbouring SNPs and may have originated by recurrent rearrangement on different haplotypes.

In an evolutionary context, one of our major challenges is to determine how we differ from our closest living relatives, the chimpanzees; to explain in genetic terms the dif-

ferences, not only in anatomy but also in intellect and, in particular, language, that have come about over the last 5–7 million years since the divergence of the human and chimpanzee lineages. It is likely that such major phenotypic differences cannot be explained solely in terms of subtle changes in the structure of specific genes. Instead, species-specific copy number variants could have given rise to key changes in the expression of certain genes, gene expression pathways and ultimately even gene networks. The future thorough characterization of CNVs in the genomes of the great apes promises to potentiate comparative genomic studies with the aim of identifying those CNVs which are human-specific and which may well have experienced positive selection during human evolution.

Our understanding of human CNVs, both in the context of evolution and disease, is contingent upon obtaining answers to the questions outlined above. In this special issue

of *Cytogenetic and Genome Research*, the interested reader will find 40 articles on different aspects of CNVs, arranged into eight different sections viz. CNVs and disease; CNVs and pharmacogenetics; CNVs in normal human populations; characterization of CNVs in the human genome; expression divergence and CNVs; CNVs and evolution; CNVs in the genomes of model organisms; the technology of CNV detection and analysis. It is hoped that this collection of reviews and original articles will not only be of practical assistance to those scientists and clinicians already working in the field but will also serve to encourage others to make their own contribution to unravelling the complexity of this still enigmatic form of polymorphic variation in the human genome.

Hildegard Kehrer-Sawatzki, Ulm
David N. Cooper, Cardiff
September 2008