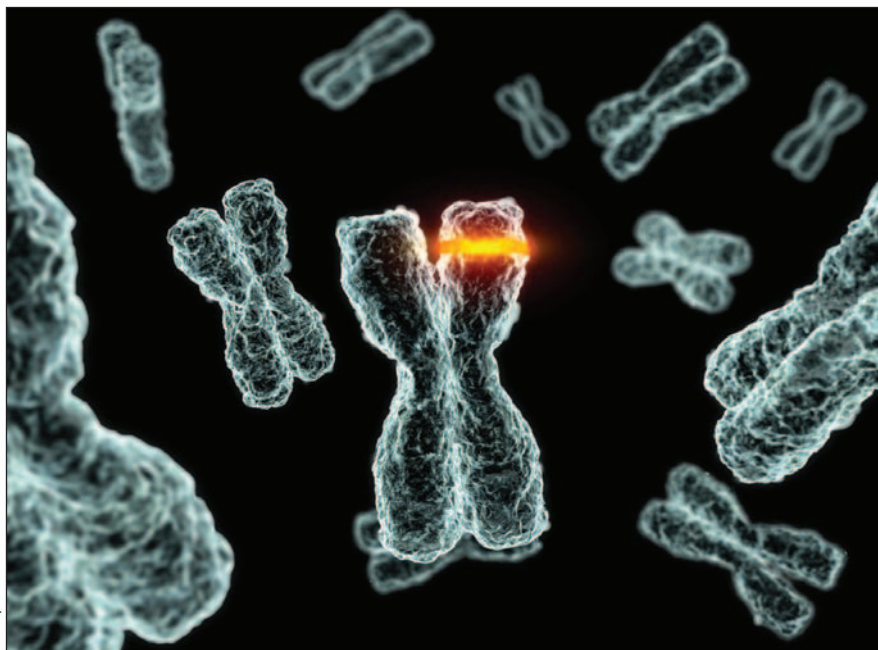


Copy-number variations add a new layer of complexity in the human genome

The first draft of the DNA sequence of the human genome was completed in 2001.^{1,2} This Herculean effort involved scientific teams around the world who together arranged in correct sequence 3 billion base pairs of DNA. Afterward, many thought that the genome had been “conquered,” with only a few remaining i’s to be dotted and t’s to be crossed. The “post-genomic” era had been declared. Researchers’ attention turned away from the composition of the genome toward understanding how genetic information influences the structure and function of proteins, tissues, patients and populations. The challenge was to identify meaningful differences in the genomic information between individuals.

Using automated DNA sequence analysis, scientists identified millions of differences between people at the level of single letters of DNA code. These differences, called “single nucleotide polymorphisms” or SNPs (pronounced *snips*), result in single-letter “misprints” and occur in about 1 in every 300 to 400 bases of DNA sequence. The number of SNP differences may involve as many as 10 million or more nucleotide bases of DNA, or about 0.3% of the total genome. Furthermore, SNPs were proposed to underlie the major differences between people, including features of appearance such as eye and hair colour as well as susceptibility to diseases and differences in responses to medications.

In November 2006, however, a new discovery dramatically expanded the understanding of differences between individuals. The publication of the “copy-number variation” (CNV) map of the genome by Steve Scherer’s group at Toronto’s Sick Children’s Hospital and their international collaborators³ initiated a paradigm shift.



iStockphoto

Variations were previously thought to occur mainly at the level of single base pairs of DNA. However, numerous submicroscopic chromosomal changes involving more than 1000 to half a million or more base pairs have recently been identified. These copy-number variations or CNVs could explain differences in disease susceptibility among individuals and eventually lead to improved prevention and treatment approaches for particular diseases.

The meteoric ascent of CNVs followed from 2 seminal publications in 2004: one from Scherer’s group and another from Michael Wigler’s group, in Cold Spring Harbor, NY.^{4,5} Each team used distinct but complementary methods to detect dosage (copy number) differences of chromosomal regions compared with the standard 2 (maternal and paternal) copies that are expected. Both teams saw numerous submicroscopic chromosomal alterations in the genomes of control subjects. These quantitative genomic variants, eventually called CNVs, were analogous to the chromosomal changes detected by classic cytogenetic methods. Whereas SNPs are analogous to a single-letter misprint in a word in an instruction manual, CNVs are analogous to having a page of the manual torn out completely, or pasted in upside down.

For decades, traditional microscopy-based karyotype analysis and, more

recently, higher resolution fluorescent dye-based visualization would occasionally detect large-scale rearrangements in patients that affected whole chromosomes or sizable chunks of chromosomes. The types of rearrangements detected cytogenetically included deletion (loss of 1 or both copies), duplication (gain of 1 or more copies), inversion (flipped orientation of a chromosomal segment) and translocation (transfer of a piece of 1 chromosome to another). One of these dramatic chromosomal rearrangements typically spanned millions to hundreds of millions of DNA nucleotides. Such rearrangements were considered to be rare events and were almost universally associated with clinical syndromes: a familiar example is the duplication of 1 copy of chromosome 21, entirely or in part, underlying Down’s syndrome. But in contrast to uncommon large pathogenic cytogenetic changes, the much

smaller CNVs are prevalent in the healthy control population.

How prevalent? Scherer's group defined a CNV as any submicroscopic chromosomal change affecting more than 1000 — and up to half a million or more — nucleotides of genomic DNA detected using 2 independent methods. They reported a total of 1447 CNVs in the genomes of 270 healthy individuals from 4 different geographic ancestries.³ The extent of the variation was breathtaking: these relatively common CNVs cumulatively affected 360 million nucleotides, or about 12% of the human genome: one of a homologous pair of chromosomes could be a million nucleotides and 20 genes shorter than the other. The comforting idea arising from the SNP map that any 2 humans were more than 99.7% identical at the genomic level was suddenly shattered.

The importance of CNVs to human genetic disorders became evident when a search of the map of single-gene disorders showed that almost 300 proven disease-causing genes overlapped with CNVs.⁵ CNVs can affect phenotypes by altering transcriptional — and presumably translational — levels of genes and their products. For instance, deleting 1 copy of a dosage-sensitive gene results in deficient function that cannot be rescued. CNVs may have a role in polygenic diseases if only for the simplistic reason that certain CNVs span regions containing many genes. Also, genomic deletions in apparently healthy individuals might not directly cause a simple monogenic disease, but in the presence of additional genetic or environmental factors, or both, may contribute to the development later in life of complex polygenic diseases such as diabetes, schizophrenia, cancer and atherosclerosis. Similarly, gene dosage increases are known to cause a few diseases in humans, but

the ubiquity of CNVs implies that this could be a more widespread mechanism underlying both rare and common diseases. So the study of SNPs alone when correlating genomic variation with disease is now inadequate in the context of knowledge of CNVs. A focus on SNPs will literally “miss the forest for the trees.” For instance, my research group recently showed that testing for both SNPs and CNVs expands the molecular diagnosis of familial hypercholesterolemia.⁶

The other side of the coin is that a large proportion of CNVs occur in “gene deserts” outside of regulatory or coding regions. The CNV map indicates a high probability of finding non-pathogenic CNVs in patient samples. How would chance findings of neutral CNVs affect the management and counselling of a patient and his or her family? In general, will ethical issues arising from CNVs simply mirror past issues encountered using cytogenetic methods? Could past medical genetic diagnoses be revised in light of knowledge of CNVs? Should archived specimens be re-evaluated? These questions and others will require attention soon.⁷

Thus, the CNV map adds a new dimension to the study of the human genome. Surprising features of CNVs include their ubiquity in the genome, high population frequency and presence in genomes from healthy people. Future genomic mapping experiments and genome-wide association analyses — and their respective detection technologies — will need to account for the presence of CNVs. Current platforms to study the genome may need to be redesigned either to maximize detection of CNVs or minimize their interference with methods to detect other forms of variation. As the “personal genome” moves closer to becoming a reality, it will be important to interpret the bio-

logical meaning of all forms of genomic variation — including SNPs and CNVs — for any individual. Finally, CNVs are now part of the contemporary discourse on genomic variation studies and their biological, health and societal implications. However, more research is required to fully understand the implications and potential applications of human genomic CNVs.

Robert A. Hegele
Schulich School of Medicine
and Dentistry
University of Western Ontario
Vascular Biology Research Group
Robarts Research Institute
London, Ont.

This article has been peer reviewed.

Acknowledgements: Robert Hegele is supported by the Jacob J. Wolfe Distinguished Medical Research Chair, the Edith Schulich Vinet Canada Research Chair (Tier I) in Human Genetics, a Career Investigator Award from the Heart and Stroke Foundation of Ontario, operating grants from the Canadian Institutes for Health Research, the Heart and Stroke Foundation of Ontario and the Ontario Research Fund, and by Genome Canada through the Ontario Genomics Institute.

REFERENCES

1. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science* 2001;291:1304-51.
2. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome [published erratum in *Nature* 2001;412:565]. *Nature* 2001;409:860-921.
3. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature* 2006;444:444-54.
4. Iafrate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome. *Nat Genet* 2004;36:949-51.
5. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. *Science* 2004;305:525-8.
6. Wang J, Ban MR, Hegele RA. Multiplex ligation-dependent probe amplification of LDLR enhances molecular diagnosis of familial hypercholesterolemia. *J Lipid Res* 2005;46:366-72.
7. Daar AS, Scherer SW, Hegele RA. Implications of copy-number variation in the human genome: a time for questions. *Nat Rev Genet* 2006;7:414.