Copy number variants and pharmacogenomics

Karim Ouahchi1,2, Neal Lindeman1,2 & Charles Lee1,2†
†Author for correspondence
1Brigham and Women's Hospital, Department of Pathology, 20 Shattuck Street, THorn 612A, Boston, MA 02115, USA
Tel.: +1 617 732 5790; Fax: +1 617 264 6861; E-mail: clee@rics.bwh.harvard.edu
2Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA

Keywords: copy number variants, drug response, duplication/deletion, genetic variation, metabolizing enzyme, pharmacogenomics

Pharmacogenomics in medicine & clinical pharmacology
The earliest pharmacogenomic studies focused on highly penetrant sequence polymorphisms in drug-metabolizing enzymes. The recent discovery of the widespread occurrence of copy number variants/polymorphisms in the human genome holds promise for new pharmacogenomic discoveries, aside from the commonly used single nucleotide polymorphism approach. Here we review the discovery of copy number variants and speculate on their implications for pathophysiology and pharmacogenomics.

Copy number variants: discovery, definitions & database
Recent studies have demonstrated that apparently healthy, unrelated individuals have several hundred segments of their genomes that exhibit variation in the number of copies present [7–9]. Iafrate and colleagues applied a 1 Mb resolution, comparative genomic hybridization (CGH) array platform to study 39 healthy, unrelated individuals, in whom they found 255 loci containing CNVs, with an average of 12.4 CNVs per individual [7]. Sebat and colleagues used a modified array-CGH technique, representational oligonucleotide microarray analysis (ROMA), and reported similar findings: 76 CNVs with an average of 11 CNVs per individual [8]. Both studies suggested that at least half of these CNVs contained coding sequence. Shortly thereafter, Tuzun and colleagues used a computational method to compare the reference human genome sequence to end sequence data available from clones in a fosmid DNA library, and identified 241 CNVs: 139 gains and 102 deletions [9]. Their method of analysis identified DNA variants ranging from 8–40 kb in size, with approximately 40% of the variants mapping to gene introns or exons. The vast majority (over 80% of the CNVs identified in this study) had not been previously reported, owing partially to a size selection bias for smaller CNVs by their computational strategy.
These and other CNV discovery studies probably only represent the ‘tip of the iceberg’ as they either sample only a fraction of the genome and/or a limited number of individuals. The real number of CNVs per genome is probably much higher; with some estimates now being as high as 140 in a given individual [Carter N, Jones K, Scherer S, Lee C. Unpublished Data]. A public CNV database [10] now serves as a repository for cataloging CNVs from published studies. More than 600 CNV loci are currently entered into this database. For each CNV, the listing includes known genes that have been mapped within each region. The proteins encoded by these genes have diverse functions, including metabolism, signal transduction, neurotransmission, cell adhesion, and immunological response.

Recent completion of Phase I of the HapMap project [6], which cataloged 1.3 million SNPs among the 269 individuals chosen to identify common genetic variation in people, provided data for McCarroll and colleagues to uncover over 500 new putative deletion variants [10]. Remarkably, whole genes were also found to be deleted, and in some cases both copies of a gene were deleted among these apparently normal individuals. Some of the high frequency, homozygously deleted genes (e.g., glutathione S-transferase μ1 [GSTM1], GST ω1 [GSTT1], cytochrome P450 [CYP]2A6, and uridine diphosphate-glucuronosyltransferase [UGT]2B17) are involved in cellular/drug metabolism, reinforcing the notion that certain copy number variants may influence a person’s ability to metabolize/detoxify specific reagents to which they are exposed.

Copy number variants & pathogenesis CNVs, like other structural chromosomal rearrangements, may arise from nonallelic homologous recombination (NAHR) and unequal crossing-over events. If the recombination event occurs between genes, they would be expected to result in deletion and duplication events, without disruption of the genes themselves (Figure 1A). For example, Charcot-Marie-Tooth disease type 1A (CMT1A) and hereditary neuropathy with pressure palsies (HNPP) are caused by a recombination event occurring within a 24 kb low copy repeat that shares 98.7% nucleotide sequence identity [11]. However, if the recombination event occurs within a gene, the structure, and most likely the function, of that gene is disrupted (Figure 1B). For example, a 9.5 kb DNA sequence in intron 22 of the coagulation Factor VIII is repeated twice at a region near the long arm terminus of the X chromosome. An intra-chromosomal recombination event involving this repeated region causes disruption of the Factor VIII gene and half of hemophilia A cases [12]. If the repeated elements are located on different (nonhomologous) chromosomes, such recombination events could lead to chromosomal insertions or translocations.

Each of these structural alterations, in turn, can lead to altered protein expression or function with subsequent phenotypic effect. Both deletions and duplications can lead to dosage imbalance of genetic material, and therefore affect the expression levels and activity of the protein [13]. Inversions and translocations could lead to protein truncation, altered gene expression (by juxtaposing coding sequence from one protein to the promoter region of another protein), or creation of a novel fusion protein (containing functional domains from two different proteins).

Examples of copy number variant genes CNVs of glutathione transferase genes GSTM1 and GSTT1 are detoxification enzymes belonging to the glutathione S-transferase (GST) supergene family. Deletion variant GSTM1 and GSTT1 alleles have a frequency within the Caucasian population of approximately 50 and 20%, respectively [14,15]. Since GST enzymes are involved in the detoxification of reactive metabolites of carcinogens, it has been suggested that functional GST variants associated with less effective detoxification of carcinogens may lead to increased susceptibility to cancer [15]. Inherited variants of GSTT1 and GSTM1 have been correlated with variable risk and different clinical outcomes in various malignancies, including acute leukemia [16] and carcinomas of the lung [17] and stomach [18]. By extension, chemical substances that are metabolized by GST family enzymes (such as halomethane) may exhibit differences in toxicity or efficacy in patients with different copy numbers of these genes.

CNVs of the CYP2D6 gene CYP2D6 was the first example of genetic polymorphism identified in the superfamily of CYP enzymes [19]. Individuals were described as poor, non-, or extensive metabolizers based on the measurement of urinary metabolites of debrisoquine after being administered the parent compound, which is used in the treatment of hypertension. Each phenotype observed was correlated to a specific genotype. More than 46 major polymorphic CYP2D6 alleles have now
been reported, and their frequencies vary among different ethnic groups [20]. CYP2D6 enzymes process more than 30 prescribed and over-the-counter drugs, including antiarrhythmics, antihypertensives, β-blockers, monoamine oxidase inhibitors, antipsychotics, and antidepressants. Although much of the focus has been on SNPs in CYP2D6, CNVs may play a role as well, as suggested by a study reported by Ledesma and colleagues [21] who genotyped 737 unrelated Caucasian individuals and found CYP2D6 CNVs in approximately 12% (5.1% with deletion [CYP2D6*5] and 7.2% with duplication [CYP2D6 2x] variants).

Focused investigations into the role of CNVs in drug metabolism have not yet been fully realized. However, these few examples of critical metabolism genes associated with the earliest reported CNVs provide a tantalizing suggestion of the potential for important future pharmacogenomic discoveries.

Copy number variants of the CCL3L1 gene Duplications in host defense genes, such as chemokines, have been recently investigated in HIV/AIDS susceptibility. Gonzalez and colleagues correlated the number of chemokine (C-C motif) ligand 3-like 1 (CCL3L1) gene copies with susceptibility to HIV in different ethnic populations [22]. CCL3L1, or macrophage inflammatory protein (MIP)-1αp, encodes for an HIV-1 suppressive chemokine and ligand for the HIV coreceptor, CCR5. A low CCL3L1 copy number was found to be associated with enhanced susceptibility to HIV infection and AIDS progression [22]. Other important immunological genes associated with CNVs include certain major histocompatibility complex class II genes (e.g., major histocompatibility complex class II, DM α [HLA-DMA] and HLA-DM β [HLA-DMB]). The role of these CNVs in infectious and autoimmune diseases has not yet been investigated, but the potential exists for CNVs to impact critical molecules involved in response to injury/infection, which may, in turn, provide a prediction of the potential efficacies of different drugs directed at different molecular targets involved in the disease process.

Are CNVs a new genetic marker for pharmacogenomic studies? Studying CNVs is not fundamentally different from studying SNPs. The techniques of detection and genotyping may differ, although the same principles apply. Candidate gene approaches focus on genes that are selected because of an a priori hypothesis regarding their role in the phenotype. Pharmacogenomic studies may focus on specific genes for which a role in drug metabolism has already been established, such as genes belonging to the CYP family, or other important metabolizing enzymes. Since only approximately half of currently listed CNVs overlap with known genes (very few of which are well described in drug metabolism), it will be important to identify more CNVs located in the vicinity of known genes.
### Highlights

- Recent studies have demonstrated the widespread presence of copy number variants (CNVs) in the human genome, with each DNA segment encompassing several to hundreds of thousands of bases of DNA. Ongoing studies suggest that any individual may have over 100 regions in their genome where the copy number of the DNA segments in each region varies from person to person.
- Half of the CNV reported overlap with genes that are encoding proteins with diverse functions. Functional CNVs genes encoding metabolizing enzymes (such as cytochrome P450 [CYP2D6], glutathione S-transferase θ1 [GSTT1] and GST μ1 [GSTM1]), in addition to immunological genes (chemokine [C-C motif] ligand 3-like 1 [CCL3L1] gene), have now been reported.
- CNVs may be useful in association studies to evaluate candidate genes and to further characterize the molecular basis of drug response.

**Association studies** are the marker allele frequency, the extent of linkage disequilibrium (LD) (reviewed in [24]), and the effect of ethnic admixture (25). Moderate to high frequency allele CNVs (i.e., more than 10%) are potentially useful in such association studies. Indeed, Iafrate and colleagues demonstrated that among more than 200 CNVs, 24 CNVs were present in more than 10% of the individuals, and these may therefore be more useful for such association studies [7]. As CNVs continue to be discovered throughout the genome, more of these common CNVs are likely to be identified. Furthermore, CNVs could be used in combination with SNPs in order to provide a larger allelic marker frequency spectrum, and to further confirm an association between phenotype and genotype.

### Conclusions

SNPs may not provide a suitable explanation for all pharmacogenomic variation. Recent discoveries have shown the widespread existence of CNVs in the human genome. These CNVs involve some known metabolizing enzymes, such as CYP2D6, GSTM1 and potential drug targets such as CCL3L1, and can influence phenotype through alteration in gene dosage, structure and expression. Moreover, CNVs may enable discoveries of other genetic mechanisms underlying variation in drug response. Clearly, further study in this area is required to demonstrate the full significance of these CNVs in human pharmacogenomics.

### Acknowledgments

The authors would like to thank David Casavant, Steven M. Carroll and PJ Perry for critical reading of this manuscript and Shona Hidsp for her assistance with preparation of this manuscript. Karim Ouahchi is supported by the Brigham and Women's Hospital Department of Pathology training grant. Some of the work presented in this review is currently supported by grants from the Leukaemia and Lymphoma Society (Charles Lee), Genome Canada, and the Wellcome Trust UK.

### Bibliography


Website