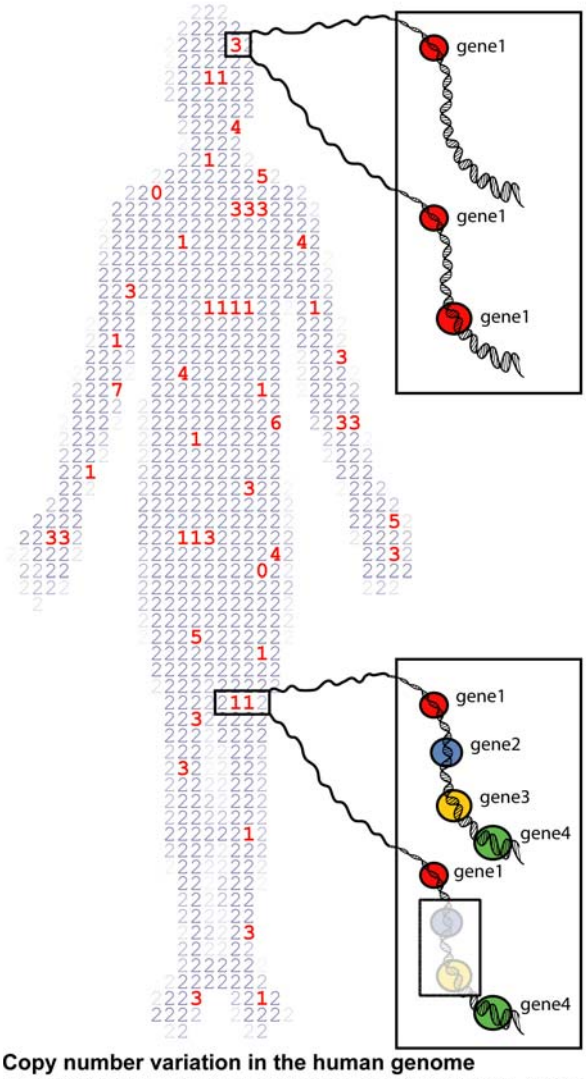
What is copy number variation?
The human genome is comprised of 6 billion chemical bases (or nucleotides) of DNA packaged into two sets of 23 chromosomes, one set inherited from each parent. The DNA encodes 30,000 genes. It was generally thought that genes were almost always present in two copies in a genome. However, recent discoveries have revealed that large segments of DNA, ranging in size from thousands to millions of DNA bases, can vary in copy-number. Such copy number variations (or CNVs) can encompass genes leading to dosage imbalances. For example, genes that were thought to always occur in two copies per genome have now been found to sometimes be present in one, three, or more than three copies. In a few rare instances the genes are missing altogether (see figure below). The new findings indicate that our DNA is less than 99.9% identical, as was previously thought.

Why are CNVs important?
Differences in the DNA sequence of our genomes contribute to our uniqueness. These changes influence most traits including susceptibility to disease. It was thought that single nucleotide changes (called SNPs) in DNA were the most prevalent and important form of genetic variation. The current studies reveal that CNVs comprise at least three times the total nucleotide content of SNPs. Since CNVs often encompass genes, they may have important roles both in human disease and drug response. Understanding the mechanisms of CNV formation may also help us better understand human genome evolution.



Copy number variation in the human genome
The 30,000 genes are usually present in two copies. A new map of the genome has been unveiled that catalogues DNA and genes variable in copy number (those numbers other than 2 are highlighted in red). Duplication of a gene (top) and deletion of two genes (bottom) are depicted.

How does the new CNV map help?
The new global CNV map will transform medical research in four areas. The first and most important area is in hunting for genes underlying common diseases. To date, attempts to identify these genes have not considered the role CNVs may play in human health. Second, the CNV map is being used to study familial genetic conditions. Third, there are thousands of severe developmental defects caused by chromosomal rearrangements. The CNV map is being used to exclude variation found in unaffected individuals, helping researchers to target the region that might be involved. The data generated will also contribute to a more accurate and complete human genome reference sequence used by all biomedical scientists.

What are the most surprising observations from the recent papers?
It was startling to discover that 12% of the human genome was copy number variable in the 270 DNA samples tested. About 2900 genes, or 10% of those known, are encompassed by these CNVs. Some CNVs found in the general population can be millions of bases in size, affecting numerous genes, yet they have no observable consequence.

How many CNVs are there in the human genome and how big are they?
To date, approximately 2000 CNVs have been described and 1447 of them are from the current study. There could be thousands more CNVs in the human population. About 100 CNVs were detected in each genome examined with the average size being 250,000 bases (an average gene is 60,000 bases). Additional CNVs will be discovered as technologies for detection improve and more DNA samples from worldwide populations are examined.

Can CNVs cause disease?
Most CNVs are benign variants that will not directly cause disease. However, there are several instances where CNVs that affect critical developmental genes do cause disease. For example, recent reviews have listed 17 conditions of the nervous system alone – including Parkinson's Disease and Alzheimer's Disease – that can result from copy number variation. To increase the value of the data, the Hospital for Sick Children has established the 'Database of Genomic Variants' to house CNVs found in the general population. The Wellcome Trust Sanger Institute has developed a database of CNVs (called DECIPHER) associated with clinical conditions.

What types of genes are found to be copy number variable?
Genes that are involved in the immune system and in brain development and activity – two functions that have evolved rapidly in humans – tend to be enriched in CNVs. By contrast, genes that play a role in early development and some genes involved in cell division – both critical to fundamental biology – tend to be spared.

Are there any bioethical considerations that are unique to CNVs?
Since the discovery of CNVs is so new, bioethics studies are just now underway. Compared to other genetic variants, CNVs are larger in size and can often involve complex repetitive DNA sequences. They can also encompass entire genes, many of which have a specific function ascribed to them. For these reasons CNV data could potentially be more amenable to misinterpretation. Some CNVs could be employed to add discrimination power in forensics, but typing them is usually less efficient than other types of genetic markers.

Are there population specific CNVs?
As with all types of genetic variation, CNVs can vary in frequency and occurrence between populations telling us something of our shared history. As a result of our recent common origin, the vast majority of copy-number variation – around 89% – is shared among the diverse human populations studied. Nevertheless, the pattern of CNV that each of us inherits subtly reflects our ancestry and can be used to infer in which continental population our recent background lies. Striking differences in regions of our genome between different populations might define variants that have allowed different groups to adapt to their different environments. One example is the strikingly increased copy number of the HIV-related *CCL3L1* gene in African populations. An understanding of how genetic variation is distributed among populations not only tells us about human prehistory, but also improves our ability to find disease genes. Yet, differences in ancestry are not proof of inherent differences, genetic or otherwise, between individuals.

What's next?
The next-generation of DNA microarray-based technologies will allow equal detection of large and small CNVs. Also on the horizon are new DNA sequencing technologies enabling rapid (and ultimately inexpensive) 'personalized' genome sequencing projects. Coupled together, these technologies will capture almost all the variation in a genome.

Databases:

Database of Genomic Variants: http://projects.tcag.ca/variation/

DECIPHER (Database of Chromosome Imbalances in Phenotypes Using Ensembl Resources): http://www.sanger.ac.uk/PostGenomics/decipher/

Further reading
Check E. Human genome: patchwork people.
Nature. 2005 Oct 20;437(7062):1084-6.

Bob Holmes. Magic Numbers. *New Scientist*, 2006, April 8, 38-41.

Daar AS, Scherer SW, Hegele RA. Implications for copy-number variation in the human genome: a time for questions. *Nature Reviews Genetics*, 2006, 7:414.

Feuk L, Carson AR, Scherer SW. Structural variation in the human genome.
Nature Reviews Genetics, 2006, 7:85-97.