

What You Need to Know

- Genes are specific segments of DNA that contain the instructions for creating proteins.
- Changes in the sequence of the DNA may lead to differences in the structure or amount of protein.
- Many of these differences contribute to the variation found in our physical features such as hair and eye color, as well as the risk to develop disease.
- The most frequent changes involve single letters in the DNA sequence, called SNPs.
- Copy number variants (CNVs) are an additional category of DNA variation involving large scale changes (1,000 to over a million letters).
- Until recently, CNVs were thought to be rare; new findings suggest they occur across as much as 15% of the genome.
- Recent studies suggest CNVs influence the risk of developing various common diseases.
- CNVs may alter the number of copies of a gene or impact the regulatory mechanisms that control gene activity.
- Scientists are just learning how to identify CNVs and associate them with specific disease.

Want more information:

A National Geographic News article that discusses the initial 2006 scientific finding:

news.nationalgeographic.com/news/2006/11/061122-human-genetics.html

An FAQ about CNVs, written by Dr. Steve Scherer, an author of the 2006 initial CNV population study:

www.tcag.ca/scherer/documents/CNV_FAQ.pdf

Other sites:

www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.0030190

projects.tcag.ca/variation/

BIOTECH 101

Copy Number Variation

Another way the genetic recipe differs

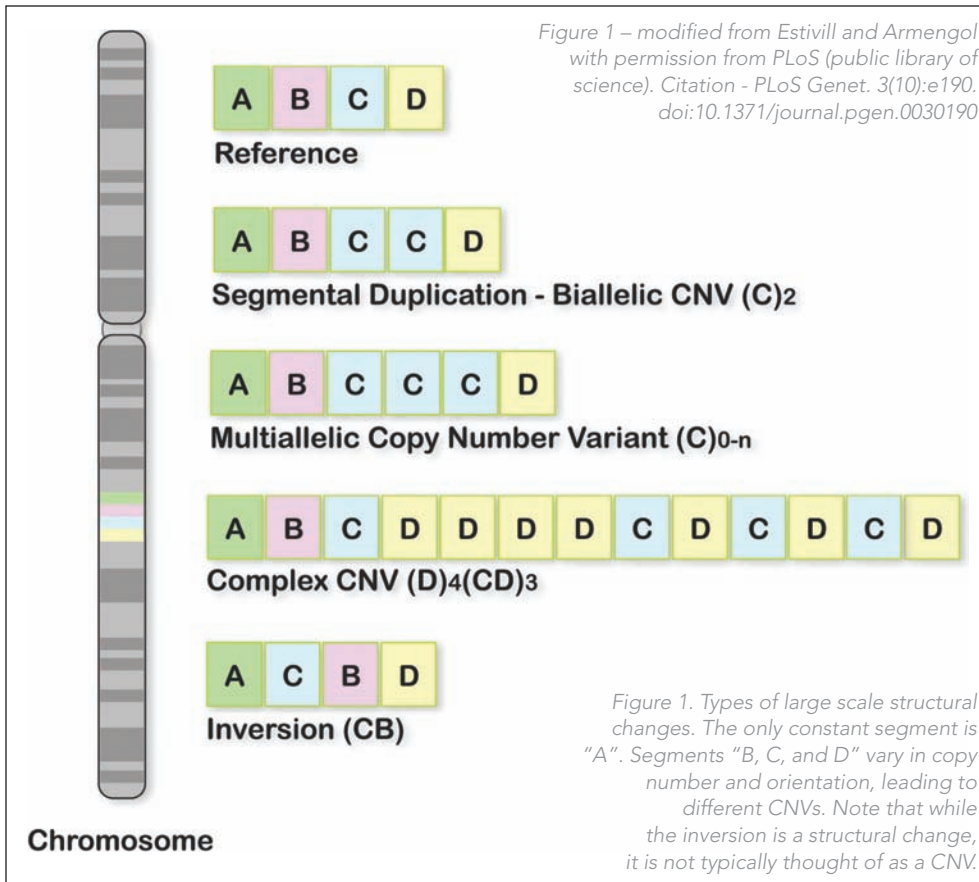
Imagine you are a member of a book club focused on cookbooks. You and your fellow members agree to try the same set of recipes from a particular volume and meet to discuss the cooking (and eating) experience. When you come together at the next meeting, there is a definite opinion split on the resulting dishes. Some members describe the dishes as delicious, others proclaim them too spicy, too bland or just plain inedible. As you begin comparing the various copies of the cookbook, you discover the recipes are similar, but not necessarily exact copies. There are a few copies with subtle changes that replace one measurement with another; some copies substitute certain ingredients or even omit key steps. Intrigued, you examine the recipes in other parts of the book and are surprised to find that in some books, entire pages or chapters

are present in duplicate or even triplicate, but occur only once or not at all in other books. In other copies, sets of pages appear to be incorporated in a different order or inserted in unusual places in the book. As a result, multiple versions of the cookbook exist, leading to very different results.

To a rough approximation, we've just described the experience of genomic researchers who've been comparing the DNA sequences (another type of recipe) across various human populations. For years these scientists have focused on small changes present in the DNA. These are equivalent to changes in measurement, ingredient and instruction from the cookbook analogy. These single nucleotide changes (also known as single nucleotide polymorphisms or SNPs) were thought to be responsible for the majority of human variation including differences in physical features and susceptibility to health and disease. Due to their abundance in the human genome (over 10 million identified), SNPs have become the primary tool linking genetic change to many common diseases.

Until recently, larger scale changes - the duplication, insertion or deletion of "chapters" of DNA - were thought to be relatively rare. However, scientists have discovered that these larger differences, known as copy number variants (CNV), occur much more frequently than was suspected. These structural changes alter the number of copies of a specific DNA segment (figure 1). Some individuals may have zero, one, two or more copies of the region per chromosome. Such differences may play a key role in the genetic contribution to human health and disease.





More common than first thought

In 2006, several studies analyzed the genomes of a few hundred healthy individuals for the presence of CNVs. Nearly 1,500 CNV regions were identified, spanning almost 15 percent of the entire human genome. CNVs include genomic regions longer than 1,000 nucleotides and up to several million nucleotides in size. About 100 CNVs were identified for each genome examined in the 2006 studies and the average CNV size was 250,000 nucleotides. It came as a surprise to many scientists just how much of our DNA variation is due to copy number changes. Previous studies based primarily on SNPs suggested that any two randomly selected human genomes would differ by 0.1 percent. CNVs revise that estimate: The two genomes differ by at least 1.0 percent. While this may not seem like a major increase, remember that the human genome is composed of approximately 3 billion nucleotides, so the estimated number of nucleotides that vary between two random individuals has increased from 3 million to 30 million. Humans are still nearly 99 percent

identical at the DNA sequence level, but the CNV research has broadened our understanding of how and where we differ.

CNVs are distributed across the genome and while most occur in areas that contain few or no genes, a substantial number do affect genes. Many of the genes with copy number changes are involved in the immune system, brain development and brain activity. What impact do changes in copy number have on gene activity? It has been suggested that CNV regions influence gene activity in different ways. The specific CNV may directly increase or decrease the number of copies of that gene, leading to a concurrent change in the amount of protein. Alternately, CNVs may alter the performance of nearby regulatory signals that activate or silence genes without directly impacting the copy number of the gene itself.

Clinical impact of CNV

What is the impact of variation in protein levels? Can too much or too little protein be harmful? Based on

our knowledge from other forms of disease, the answer is yes. For example, the presence of an additional chromosome 21 results in Down syndrome. These individuals have a copy number increase for an entire chromosome of genes, leading to the clinical features associated with Down syndrome. Preliminary studies have linked CNVs to lupus, Crohn disease, autism spectrum disorders, Alzheimer disease, HIV-1/AIDS susceptibility, rheumatoid arthritis and Parkinson disease. In some cases the associated CNV is rare, but in other diseases, the identified risk variant is quite common. It is also likely that CNVs may influence individual drug response and susceptibility to infection or cancer.

We are just beginning to identify and catalog human CNVs and our understanding of their impact is in its infancy. Scientists have been hampered in their ability to study CNVs due to a lack of appropriate detection methods. This is changing and newer technologies allow easier identification and classification of CNV regions, especially those involving smaller segments of DNA. What has become clear is that there is not a "single human genome sequence" but instead several different configurations, much like the versions of the cookbook described in the opening paragraph. These different versions contain changes, duplications and deletions on both large and small scales and the interplay between genes, regions that modify the action of genes and the surrounding environment results in the balance between health and disease.

– Dr. Neil Lamb

director of educational outreach
HudsonAlpha Institute for Biotechnology

Note: Dr. Devin Absher, a recent addition to the HudsonAlpha scientific team, is involved in CNV identity for tumors. He was profiled in the Spring 2008 issue of *Through the Microscope*.