

Fingering the culprit. Extra copies of one gene cause a form of Charcot-Marie-Tooth disease, a neuropathy that affects the feet and hands.

GENOMICS

DNA Duplications and Deletions Help Determine Health

Each human's genome is distinguished by extra, and sometimes missing, DNA that can powerfully impact everything from development to disease

In 1991, both *Science* and *Nature* turned down James Lupski's submission that described an unprecedented link between an inherited human disease and a novel chromosome aberration. "It was rejected without even being sent out for review," recalls Lupski, a geneticist at Baylor College of Medicine in Houston, Texas. Unlike the many studies that fingered tiny mutations in genes as the cause of inherited diseases, Lupski pointed to a relatively large, duplicated region of one chromosome as the culprit. Later research showed that this duplication, inserted within the same chromosome, harbors an additional copy of a gene involved in making a nerve cell's protective sheath. The extra dose of the gene caused the sheaths to disintegrate, interrupting signals between the brain and the hands and feet. At the time, "there was no appreciation that copy number was a mechanism of disease," says Lupski, whose study appeared in *Cell* later that year.

Lupski's link between gene copy number and the peripheral neuropathy known as Charcot-Marie-Tooth disease (CMT) marked the opening of a new chapter in human genetics. Not only has his discovery led to progress in understanding and potentially treating this devastating disease, but it also set the stage for what has recently become a frenzy to find other connections between disease and gene duplications or deletions. As new techniques to spot such genetic differences have become avail-

able, investigations of the human genome have found thousands of variations in the number of copies of a gene, a piece of a gene, or large stretches of DNA—so-called copy number variants, or CNVs. "This whole new world has been opened up in genetic variation," says cytogeneticist Charles Lee of Brigham and Women's Hospital in Boston.

Geneticists are steadily linking more and more of these CNVs to human maladies, including Alzheimer's, Parkinson's, various mental retardations, autism, color blindness, and anatomical deformities. Several studies

Gene Dosage Affects Symptoms

	GENE COPIES	PHENOTYPE
—(□)— —(□)—	2	Normal
—(□)(□)— —(□)—	3	CMT1A
—()— —(□)—	1	HNPP
—(□)(□)— —(□)(□)—	4	Severe CMT1A
—(□)(□)(■)— —(□)—	3	Mild CMT1A HNPP features

Numerous possibilities. The number of extra, missing, and mutated (red) *PMP22* genes leads to neuropathies (CMT1A and HNPP) of varying severity.

have shown that CNVs can also powerfully influence a person's susceptibility to disease and to the side effects of medications. And each connection between a CNV and a disease suggests novel targets for therapies.

CNVs have become such hot commodities that some researchers now contend they're as important as mutations in genes themselves. And several companies that specialize in analyzing DNA have developed new tests to detect CNVs. "We're starting to see CNVs incorporated into most genetic studies now," says Stephen Scherer, a medical geneticist at The Hospital for Sick Children in Toronto, Canada.

But with all the excitement surrounding CNVs, several leading researchers in the field urge colleagues to keep their enthusiasm in check. One caveat is that the main technologies used to pick out duplications and deletions are relatively blunt tools, leading to widespread concern that far more CNVs have been reported than truly exist. "There's a lot of hype in the CNV field right now," says Lee.

Yet there's growing agreement that CNVs can have a profound influence on determining what makes individuals unique, reaching far beyond health status to affecting underlying differences in looks and personalities. "What's cool is that we're a mosaic of pieces of genomes," says Evan Eichler, who studies gene duplications at the University of Washington, Seattle. "None of us is truly normal."

Beyond belief

Geneticists have long known that extra or missing chromosomes or chromosome fragments, visible under a microscope when a cell's DNA is stained, can cause conditions such as Down syndrome. They have also tied scores of diseases to molecular misspellings: mutations in individual bases that make up the DNA in a gene. But, as Lupski discovered, other things can go haywire in the human genome. Pieces of chromosomes too small to see with a microscope can break off, attach in the wrong place, or duplicate, creating "structural variations" that range in size from 2 to 2 million bases. A piece of a gene, a whole gene, or many genes can get caught up in these rearrangements, which occur as DNA is copied during cell division.

Until recently, these submicroscopic changes have escaped routine detection. Indeed, until the human genome was sequenced, researchers had few clues about their existence. Now, with a reference sequence in hand and new techniques to analyze DNA, discovery of structural genetic differences in an individual has become commonplace, and there's an increasing appreciation of how big a role they play in making each

person's genome distinct. "The field is moving very fast," says Lupski.

The identification of these variations has upped the ante on the subtle ways humans differ. Over the past decade, studies that compare DNA differences between individuals have found that single base changes, called single-nucleotide polymorphisms, or SNPs (pronounced "snips"), are abundant, occurring as frequently as once every 100 bases in the 3-billion-base-long human genome. Now CNVs—which by definition are DNA stretches of 500 bases or more that differ from the human genome reference sequence—are proving to be more common than previously thought.

In 2004, two reports that detailed the frequency of CNVs stretched many minds. First came a paper in the 23 July 2004 issue of *Science* (p. 525) in which Jonathan Sebat and Michael Wigler of Cold Spring Harbor Laboratory in New York state and colleagues

described the results of their search for CNVs in 20 "normal" individuals. They found 211 CNVs. That was far more than they and others expected, but welcome confirmation came a week later when *Nature Genetics* published a paper online from the labs of Toronto's Scherer and Brigham and Women's Lee, who analyzed 39 healthy people and found 255 CNVs. In both studies, the average person had about a dozen CNVs. When they first submitted their paper, says Scherer, the total known human CNVs, including the one Lupski had found, numbered just 12. "Two of three reviewers said, 'This just can't be true. We would have seen it before,'" Scherer recalls.

Scherer, Lee, and other colleagues have since created a Database of Genomic Variants to catalog all reported CNVs in normal humans. As of their most recent update in March, they had compiled 6482 CNVs from 40 publications. Just this week, in the 4 Sep-

tember issue of *PloS Biology*, researchers describing J. Craig Venter's genome sequence reported that this sequencing maverick has more than 60 CNVs, half of them losses and half of them gains (see News story by Cohen).

Start making sense

Discovering a variation in gene copy number is now the easy part, says geneticist Matthew Hurles of the Wellcome Trust Sanger Institute in Hinxton, U.K. "The much harder thing is to determine its impact." Toward that goal, Hurles and colleagues have put together DECIPHER, a database that catalogs CNVs linked to disease. And as part of the Genome Structural Variation Consortium, Hurles, Scherer, and Lee are searching for additional functionally relevant CNVs. In the 23 November 2006 issue of *Nature*, the consortium reported analyzing 270 healthy individuals from four distinct ethnic populations. They found that 14.5% of the CNVs in these people included genes already identified as having a role in inherited disorders, suggesting that the variants may be key in a wide variety of afflictions.

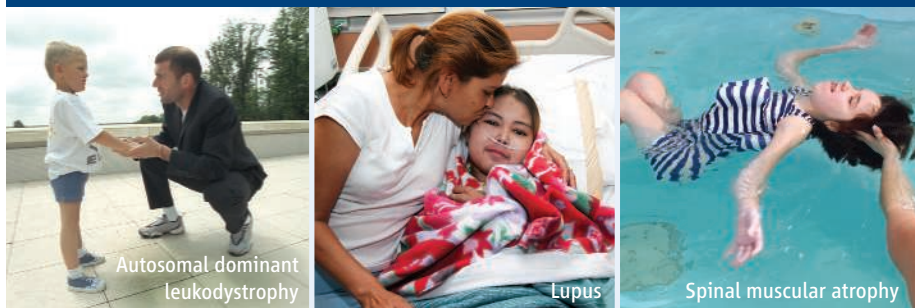
The consortium looked at individuals from the same populations to determine the relative importance of CNVs and SNPs in altering gene expression. In the 9 February issue of *Science* (p. 848), the researchers reported that CNVs accounted for 17.7% of gene-expression differences, whereas SNPs accounted for the rest. These findings, they concluded, underscore that significant genetic variation can involve everything from a single base change to millions of bases.

Yet finding that a CNV alters gene expression or involves a disease-related gene is still a far cry from explaining how it actually causes symptoms. As it turns out, CNVs can cause problems by several different mechanisms. In the simplest scenario, extra or missing copies of a gene alter its overall expression, thus changing production levels of the protein encoded by the gene.

Genes typically come in pairs, one in each set of chromosomes. A single missing gene can reduce protein production, and if both copies are missing, it can shut down production altogether. Such missing genes cause everything from neurodevelopmental disorders (including the syndromes known as Williams-Beuren, Smith-Magenis, Prader-Willi, Angelman, and Miller-Dieker) to color blindness and predisposition to Crohn's, lupus, and AIDS.

Gene duplications—which can range from one to more than a dozen copies of a gene—in contrast, can increase protein production. As Lupski discovered, such excess leads to a form of CMT. Studies of families with early-onset Alzheimer's disease have found that extra

Copy Number Variants Involved in Human Disease



DISEASE	GENE	PHENOTYPE
Charcot-Marie-Tooth type 1A	<i>PMP22</i>	Demyelination, peripheral neuropathy
X-linked hypopituitarism	<i>SOX3</i>	In males, short stature, mild mental retardation
Autosomal dominant leukodystrophy	<i>LMNB1</i>	Demyelination, white brain matter abnormalities
Parkinson's	<i>SNCA</i>	Neuron degeneration, rigidity, tremor
Alzheimer's	<i>APP</i>	Amyloid β precursor protein buildup
Altered drug metabolism	<i>CYP2D6</i>	Increased side effects, increased or decreased efficacy
HIV/AIDS	<i>CCL3L1</i>	Increased susceptibility to infection and disease
Lupus	<i>FCGR3B</i>	Increased susceptibility to kidney failure
Smith-Magenis syndrome	<i>RAI1</i>	Mental retardation
Pelizaeus-Merzbacher	<i>PLP1</i>	Demyelination, paralysis of legs, involuntary jerking of head
Spinal muscular atrophy	<i>SMN1</i>	Spinal deterioration, milder disease w/ later onset
Rett-like syndrome	<i>MECP2</i>	Mental retardation, spasticity, language/speech problems
Miller-Dieker syndrome	<i>LIS1</i>	Brain malformation, mental retardation, epilepsy
Neurofibromatosis type 1	<i>NF1</i>	Tumors, cognitive deficits

copies of genes can overproduce the amyloid β precursor protein, causing problems for the brain. Similarly, geneticists studying a rare form of early-onset Parkinson's have found that it can result from a CNV that causes the overproduction within the brain of the protein α -synuclein.

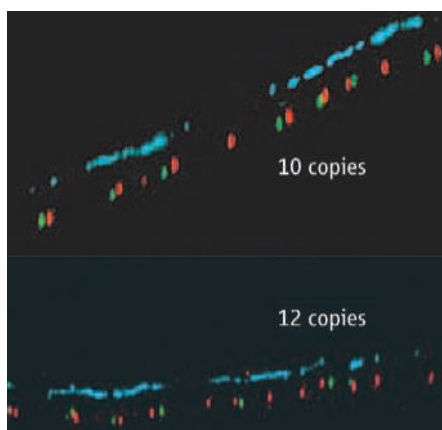
But unraveling the ties between CNVs and disease isn't always straightforward. Having an extra or missing gene often has no appreciable impact. Sometimes an extra copy of a gene includes a SNP, making it different from the normal copies, and it's that single-base mutation that causes problems. The SNP, in turn, might modify the effect of the CNV, lessening or exacerbating the resulting symptoms. Other times, the CNV alters the regulatory mechanism that controls a gene located far away. Or the creation of a CNV can add or delete DNA in the middle of an existing gene, effectively crippling it.

"It's a whole different thing to go from finding a CNV to determining which are the dosage-sensitive genes" relevant to a disorder, says Lupski. He notes that it took his group several years after linking a CNV to CMT to demonstrate that it was extra copies of the gene *PMP22* that led to disease by increasing production of its protein.

Given the difficulty of proving causation, there's a growing concern that some researchers have been too quick to pronounce disease associations with duplications and deletions. "It seems right now that there's this fervor with CNVs, and everyone's testing their disease cohort and saying, 'We found a connection that's never been seen,'" says Eichler. One much-discussed example of a result that hasn't held up: a report in the 10 August 2001 issue of *Cell* that tied susceptibility to phobia and panic disorders to a duplicated region of chromosome 15. Five subsequent studies found no such association.

Eichler and other researchers worry, too, that the literature—and thus the Database of Genomic Variants—is littered with too many large CNVs, which typically stretch for several thousand bases. Eichler notes that all of the bases encompassed by the 6500 CNVs in the database added together account for more than 20% of the genome, a notion he brands as "absolutely bogus." Says Eichler, "They're overinflating the real estate." Scherer, on the other hand, suspects that the literature currently underestimates the true number of CNVs.

Some of this debate stems from limitations of the technique most commonly used to detect CNVs, comparative genome hybridization. Adapted from cancer research, the technique takes advantage of the fact that a single strand of DNA will stick, or hybridize, to its comple-



Red alert. Two chromosomes contain different copy numbers (red dots) of a gene that codes an amylase, an enzyme that breaks down starch.

mentary strand. So researchers use a known strand of DNA as a target, and then compare how much of an unidentified strand hybridizes in comparison to the amount of hybridization seen with the known complementary strand. Powerful as the technique is, it sometimes detects DNA differences that do not exist or reports the same CNVs as unique. Ideally, researchers would compare a person's actual DNA sequence to the reference genome. Although this is a much more expensive process, Eichler and colleagues plan to employ this sequence-level strategy to gather details about the deletions and duplications in more than 50 individuals, thereby getting a handle on the true numbers of CNVs.

Into practice

Once a CNV is conclusively linked to a disease, how does that help people? "That's a burning question for the whole field and for SNP associations as well: How do you translate that information into a pharmacological strategy?" says Steve McCarroll, a geneticist who works on both CNVs and SNPs at the Broad Institute in Cambridge, Massachusetts.

Researchers have made tangible headway with CMT type 1A disease, about 70% of which is caused by an extra copy of *PMP22*. In the April 2004 issue of *Nature Medicine*, a team led by Michel Fontés of the French biomedical research agency INSERM in Marseille, France, described engineering mice to overexpress the human *PMP22* gene, mimicking the CNV Lupski had found in people. The researchers demonstrated that high doses of vitamin C reduced the expression of *PMP22* in the mice, which led to remyelination of their damaged axons and improvements in their locomotion. In April, researchers at Wayne State University in Detroit, Michigan, began studies with high doses of vitamin C in patients

with the CMT type 1A duplication. Results are expected in 2010. "It's all very, very early," says Lupski.

Pharmacogenomics researchers, who examine how genetic variations affect individual responses to drugs, are finding it useful to consider CNVs as they evaluate new treatments and unravel why some people don't respond to existing medicines. They now know, for example, that in the subset of cancer patients who have extra copies of the *PMP22* gene, the chemotherapeutic vincristine accelerates the neuropathies seen in CMT type 1A and should not be used. Another potential application of CNV identification to personalized medicine involves *CYP2D6*, a gene that codes for a liver enzyme and partially controls how people metabolize drugs.

And researchers last year discovered that variations in the number of copies of a gene involved with steroid metabolism, *UGT2B17*, might explain variations in testosterone levels in athletes. In a study, *UGT2B17* was deleted more frequently in Korean than in Swedish men, leading to higher testosterone levels in the Swedes. Such work could improve drug-doping testing of competitors by distinguishing people who take artificial testosterone from those with normally high levels in their body.

Figuring out the true role CNVs ultimately play will depend heavily on the accuracy of the technologies used to detect them and the populations that researchers tap to assess links to disease. Many genomics companies recently have put serious muscle into developing better assays and have begun to collaborate with groups that have collected DNA from well-characterized populations. Iceland's deCODE Genetics, which has made great strides identifying associations between SNPs and disease in its large database of more than 100,000 Icelanders, has teamed up with assaymaker Illumina in San Diego, California. "We have a very significant effort looking for CNVs," says deCODE CEO Kári Stefánsson. McCarroll and David Altshuler at the Broad Institute have joined forces with Affymetrix in Santa Clara, California, to develop assays that simultaneously detect both SNPs and CNVs. And NimbleGen Systems in Madison, Wisconsin, announced in July that the Genome Structural Variation Consortium will use the company's new comparative genome hybridization assay, which can map CNVs that are as few as 500 bases long—100 times finer resolution than the group used in the map it published last year.

Lupski, who has watched the field grow from its inception, says to expect many more surprises. "We're charting new territory, and we don't know where we're going to sail to," he says. "And that's fun." **—JON COHEN**