Magic numbers

It's not just which genes you have, it's how many. **Bob Holmes** reveals what makes us who we are

WAY back in 1991, James Lupski was searching for the cause of Charcot-Marie-Tooth syndrome, a common hereditary disease that affects the nervous system. Lupski was comparing genes in people with and without the disease in the hope of finding a gene variant unique to the condition. His team had narrowed their hunt to a small region of chromosome 17 when they got a puzzling result. Instead of having two copies of the region in question – one from each parent – people with the disease had three.

"We thought my gosh, how can you get three?" recalls Lupski, a geneticist at Baylor College of Medicine in Houston, Texas. When they looked more closely, they found that a 1.5-million-base-pair sequence was duplicated on one copy of chromosome 17. The result is that individuals with the disease have a total of three copies of all the genes in this region. All the copies are perfectly normal – it is purely the number of copies that matters.

For years afterward, geneticists viewed Charcot-Marie-Tooth – and a few other diseases of similar origin that turned up later – as an aberration. Everyone was used to thinking that we get one copy of each gene from our mother, one from our father. Two parents, two copies. End of story.

In this standard view, the genetic differences between any two individuals are due to slight differences in the sequence of their genes that mean that the genes "spell out" slightly different proteins. More recently, it has become clear that slight variations in the regulatory sequences that determine when a gene is switched on, or how much of a protein is made, also matter.

But this still isn't the full picture. "What's really amazing is there are lots of people walking around who have a million base pairs of some chromosome missing or amplified, and they're apparently normal and healthy," marvels Stephen Scherer of the Hospital for Sick Children in Toronto, Canada. "We didn't know this a couple of years ago."

These big chunks of missing or extra DNA are found across the genome and often encompass whole genes or blocks of genes – which means that we all have fewer copies of some genes and more copies of others. And it is becoming clear that these differences in the number of gene copies do matter. Many genetic diseases, from obscure birth defects to more common maladies such as schizophrenia and Alzheimer's, are at least partly due to these changes in gene copy number. There are also hints that variations in the number of gene copies may be just as important as variations in gene sequence in generating the normal differences that help to make each of us who we are.

Of course, geneticists have known for decades that there are sometimes massive deletions or additions between individual genomes. After all, that most basic of differences, sex, is due to the presence or absence of a Y chromosome, while Down's syndrome is the result of having three copies of chromosome 21 instead of the usual two. Such major changes, including the chromosome rearrangements that often occur in tumours, are large enough to spot under a microscope.

Dozens of diseases

However, with the advent of molecular techniques, geneticists began to link smaller-scale duplications and deletions with various diseases. Since Lupski's original discovery, gene duplications or deletions have been implicated in dozens of other hereditary diseases, including several forms of mental retardation. Several largely harmless traits are also caused by gene deletions, such as rhesus (Rh) negative blood types and some forms of red-green colour blindness.

In all of these cases, teams stumbled across the gene copy number anomaly while studying a particular disease or trait. It is possible that other examples have been overlooked because researchers have been hunting for mutations rather than copy number variations. If so, copy number variations could be far more common than we realise, even today.

Until recently, it was hard to know for sure, because standard molecular techniques do not reveal how many copies of a particular gene a cell contains. Even if researchers had the resources to sequence and compare vast stretches of DNA from many individuals, it would not help much. Sequencing usually involves chopping DNA up into little fragments, sequencing the fragments and assembling them based on overlaps, a process that often fails to reveal large repeats. "You're studying pieces of DNA that are almost identical," says Scherer. "How does one place pieces of a jigsaw puzzle that have the exact same shape?"

To find large-scale repeats or deletions, researchers have had to develop specialised techniques. At Cold Spring Harbor Laboratory near New York City, a team led by Michael Wigler used the published human genome sequence to identify 85,000 unique DNA sequences scattered evenly across all the ▶

chromosomes and mounted these markers on a "gene chip". Next, the researchers compared the DNA of two people at a time. Pieces of one individual's DNA were labelled with a coloured fluorescent tag; the other person's DNA was labelled with another colour. Then the two samples were mixed and washed over the chip. Any piece of DNA containing the same sequence as a marker on the chip will bind to that marker. If one individual has more copies of a given sequence than the other person then the fluorescence will shift towards their colour at the corresponding marker.

The results are stunning. Comparing the genomes of just 20 normal individuals, Wigler's team found 76 duplicated or deleted regions, which contained 70 known genes.

Any two people have an average of 11 duplications or deletions compared with each other, they reported two years ago (*Science*, vol 205, p 525). But this first scan probably underestimates the extent of variation, says Wigler, because all but the largest repeats or deletions were likely to fall into the spaces between markers and thus escape detection.

Through the roof

Sure enough, Wigler and his colleague Jonathan Sebat are finding many more as they repeat the analysis with a more closely spaced array of nearly 400,000 markers. "My guess is there are probably on the order of about 50 of these things that are at least 50 kilobases in size between any two people," says Wigler. That could mean more than 50 genes varying in copy number.

Wigler's study was the first of a host of studies to reach similar conclusions. Barely a month later a team led by Scherer and Charles Lee of Harvard University used a similar technique but different markers to find 255 sites where DNA was deleted or duplicated in a sample of 55 people. And last July, researchers led by Evan Eichler of the University of Washington in Seattle again used a gene chip to find 119 duplications or deletions in a sample of 47 people.

Probably even more common are smaller deletions and duplications of a few thousand base pairs in length – too small to detect by gene chips but large enough to contain a short gene or regulatory sequence. Within the past year, several groups have found hundreds of these smaller structural variants using alternative techniques. "The field's just gone through the roof," says Scherer.

Most researchers think they are only scratching the surface. So far, the lists of variants overlap only slightly from study to study, which suggests that there are a lot more to find. "I'm not surprised at all," says Scherer. "They're testing different individuals, and often using different approaches." To pull all this information together, both Scherer and Eichler have begun compiling databases of structural variants in the human genome. Their lists already include over 1000 different duplications and deletions.

The researchers' next aim is to do a more thorough survey by comparing the genomes of people from different ethnic backgrounds. The International HapMap Project is already comparing single-base-pair changes (known as SNPs) in 269 people of African, European, Japanese and Chinese origin. Now two consortia are starting a survey of the same 269 people, this time looking for variations in gene copy number.

One group, including Scherer and an international team of collaborators, is using gene chips to trawl for repeats and deletions in more detail than ever before. The other, led by Eichler and his colleague Maynard Olson, will try a different method. They will chop the genome up into about 100,000 short bits and compare the length of each bit with the length predicted by the published human genome sequence. Any differences they find will be down to the duplication or deletion of a stretch of DNA – and, as a bonus, they can sequence the fragment to determine exactly what has changed.

The two approaches should complement one another well, says Eichler. What is more, the sample of people includes 60 families, so by looking for gains or losses that appear in children but not in their parents, researchers will be able to estimate how often new gene copy number variants arise.

All of this raises a key question: just how important are these deletions and repeats? We know gene copy number variants can alter the amount of protein produced. Cells with three or more copies of a gene will tend to produce more of the protein the gene codes for than cells with the standard two copies. Because women have two copies of the X chromosome, most of the genes on one of the Xs are switched off to avoid double-dosing on these proteins compared with men, for instance.

However, the production of many proteins is controlled in ways that make the number of gene copies irrelevant. And even if cells do produce more of a protein, it does not always matter. Charcot-Marie-Tooth syndrome turns out to be caused by an extra dose of one particular protein, for instance, but there are another 22 genes within the duplicated region. As far as anyone can tell at the moment, the presence of extra copies of these other genes has no effect.

Nevertheless, many geneticists suspect that copy number is important. "It would be truly astounding if these things were not meaningful," says Wigler. "The problem is, proving that is difficult and expensive, and requires luck."

It is likely that many biologically important effects will only become apparent under certain conditions or at certain times in a person's life. "A lot of 'insignificant' structural variations just reflect our total ignorance,"

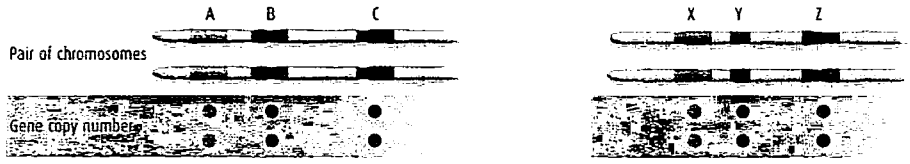
"Lots of apparently healthy people lack big chunks of DNA"

says Lupski. "Maybe a deletion is there all through the life of an individual, but it's not until age 65 that they get prostate cancer." Such subtle effects would be easy to miss unless you already knew what to look for, and when to look for it.

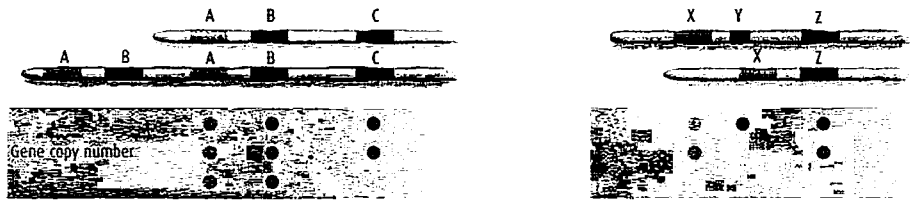
Despite all these difficulties, recent studies are starting to reveal some of the ways in which gene copy number variants can affect individuals. For example, a team led by Sunil Ahuja of the Veterans Administration's AIDS

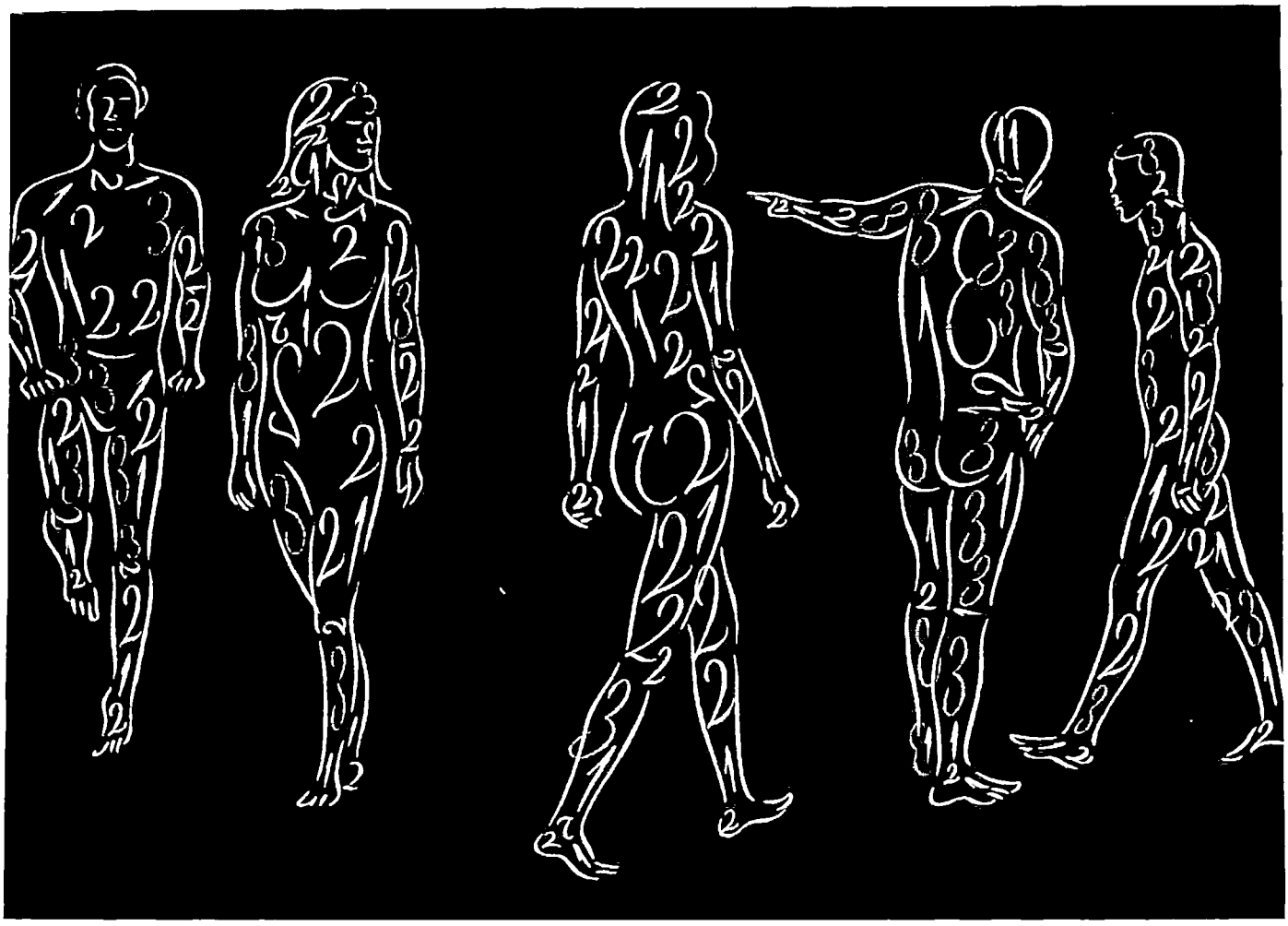
THE NUMBER OF THE GENE

The conventional view is that we have two copies of all genes except those on the sex chromosomes...



...but random duplications and deletions of large segments of DNA mean the number of copies of many genes varies





research centre in San Antonio, Texas, discovered last year that the more copies of an immune-system gene called *CCL3L1* people carry, the less susceptible they are to HIV infection (*Science*, vol 307, p 1434). The gene makes a protein that binds to the receptor HIV uses to gain access to white blood cells; more protein means fewer receptors available for HIV to latch onto.

Meanwhile, Wigler and Sebat have some preliminary evidence linking autism to specific variants in gene copy number. They are keeping mum on the details until they can confirm the find, but their work has already attracted a massive $11 million grant from a charity called the Simons Foundation. Other studies published within the past few months link copy number variation to schizophrenia, bipolar disorder, Parkinson's, a kidney disease and a rare, inherited form of early-onset Alzheimer's disease. One day these findings, if confirmed, might lead to new therapies for these conditions, perhaps based on damping down or increasing the amount of protein produced to compensate for duplicated or missing genes, but no one expects such treatments any time soon.

For now, the focus is still on finding copy-number variations, and working out their effects. Several groups looking at the functions of genes known to vary in copy numbers have found that many appear to play a role in environmental interactions such as immune response or detoxifying environmental poison. "These are the types of genes that are like accordions in our genomes, expanding and contracting all over the place. They're essentially environmental interaction genes," says Eichler.

It is also becoming clear that copy-number variation is not limited to humans. Various teams have begun to find plenty in mice, as well. At Baylor College of Medicine in Houston, Texas, geneticist Wei-Wen Cai has used gene-chip hybridisation to compare 14 inbred strains of lab mice. He found that any two strains typically differ by two or three dozen duplications or deletions that are at least 100,000 base pairs long (*Nature Genetics*, vol 36, p 952). In more recent, unpublished work, he has found copy-number variants associated with traits such as disease resistance and risk of obesity, arthritis, glaucoma and cancer.

To really tease out the links between gene copy number and the bodies and behaviour of normal individuals, researchers are going to need to roll up their sleeves and start doing experiments. Cai's colleague Lupski is altering copy number in mice so that he can directly measure the effect on their behaviour. The initial results look promising, Lupski says. "I think we will find normal differences in behaviour that correlate to differences in copy number." All these findings have led some enthusiasts to suggest that copy-number variation could prove to be more important than single-base-pair changes as a source of genetic variation.

Evolutionary biologists have long regarded gene duplication as important, because once a species acquires a spare copy of a gene, this copy can mutate, take on a new function and turn into a new and different gene. Now it appears that possessing extra copies of a gene might be important in and of itself, and not merely as a route to evolving new genes. After all, Cai notes: "If you want to get a big project done, you don't wait for a Newton or an Einstein to come along and help you. You just put more people into it." ●