

C. DARKIN



# Patchwork people

For years it was assumed that tiny differences in our genetic make-up gave us our individual traits. Now it seems that those characteristics are caused by rearrangements of large chunks of our DNA — variations that could be the key to understanding disease. **Erika Check** investigates.

**E**xactly one year ago this week, scientists announced that they had finished the 'Book of Life'. The complete sequence of the human genome had been painstakingly reduced to an ordered list of letters representing the four bases of DNA. This text was believed to be virtually identical for every person on Earth — and the major differences between individuals, such as hair colour, were said to be the equivalent of typographical errors, no longer than a single letter. The next major task for scientists was to find out which of these tiny differences can cause disease.

But even as the ink was drying on the complete sequence, some researchers were ques-

tioning whether there was really such a thing as the definitive edition of the Book of Life. By skim-reading individual genomes, these scientists were finding bizarre and unexpected irregularities. In some people, whole paragraphs of the text were duplicated, whereas in others, large passages were missing, or even printed backwards. These major revisions turned up in all kinds of people, including many who seemed healthy and normal. Suddenly, it seemed possible that there was actually no standard version of the Book of Life, and researchers wondered whether we are all much more different from each other than they had thought.

Now, scientists are beginning to link these larger variations in the genome to human health and disease. Some even say that the variations explain normal human diversity and evolution, which have always seemed too intricate to be the result of differences in just single DNA bases. "When we first looked at the human genome, we were so very proud of being able to look at the definitive sequence," says genomics researcher Chris Ponting of the University of Oxford, UK. "Now, in just a few years, we've travelled such a long way. We've gone from looking at the human genome to looking at human genomes, plural."

The widespread existence of all these varia-

tions is a big surprise because such large changes have been associated with devastating genetic diseases<sup>1</sup>. Deleting a chunk of DNA, for example, can eliminate important genes. Extra copies of a gene can cause overproduction of a protein, throwing the cell's finely balanced biochemistry out of kilter. And moving a DNA chunk from one location to another, or even reversing its orientation, can dramatically change the signals that control genes.

So far, most of these variations have not been found to cause overt disease. But researchers now think that they may have a more subtle role to play. They could influence our vulnerability to diseases that are caused by a complex mix of genes and environment. To find out more, two major public projects are cataloguing the extent of these variations in normal humans. One, funded by the US National Institutes of Health, will look for structural variation in the genomes of ten individuals from different ethnic backgrounds. The other, funded by Genome Canada and the UK-based Wellcome Trust, will scan the genome for genetic differences among the 270 people included in the International HapMap Project — a huge study looking at human genetic diversity. And several private donors are funding projects to find out whether the large variations might cause complex diseases such as cardiovascular problems, Parkinson's, psychiatric disorders and autism.

### Shock results

The pioneering work that spurred these large projects began in 2002. Data were pouring out of the efforts to sequence the human genome, as were tools that allowed scientists to compare DNA from different individuals in incredibly fine detail. At first, researchers were surprised and even disturbed by the new findings.

Charles Lee, a cytogeneticist at Brigham and Women's Hospital in Boston, Massachusetts, was investigating whether he could use one of the new technologies as a genetic test. But his experiments kept failing. He frequently found major aberrations in the gene sequences of normal patients he was trying to use in the control group. Some of this group were apparently carrying more copies of certain genes than others, yet they seemed perfectly healthy.

Lee was unsure about what was going on until, in late 2003, he went to Canada to give a talk at a meeting in Toronto. There he discovered that Steve Scherer of the Hospital for Sick Children in Toronto was seeing the same weird phenomenon: normal, healthy patients with different copy numbers for certain genes.

Meanwhile, at Cold Spring Harbor Laboratory in New York, molecular geneticist Michael Wigler was using a different technology to compare the genomes of two men: a Caucasian and an African pygmy. He also saw copy-number changes where he did not expect them — in this case, in a gene that codes for a crucial brain chemical. "We were very excited, and then very afraid," says Wigler,

## IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

**Biology lesson: our ability to learn may be governed by large differences between our respective genomes.**

who was worried that he had discovered a mutation that would predispose one of the men to schizophrenia.

But in the next experiment, his group found copy-number changes in a different gene that is active in human sperm. "That really freaked us out," Wigler says. Over the next year, his group turned up more examples of the same phenomenon. "It didn't take us long to realize we were looking at something that was fairly common," Wigler says.

How common, exactly? Last July, Wigler's group reported that it had looked at 20 normal individuals and found 221 places in the genome where those people had different copy numbers of stretches of DNA<sup>2</sup>. Some of these copy-number changes showed up in more than one person, and so qualify as

'polymorphisms' — shorthand for particular spots in the genome that regularly differ between individuals. In the Book of Life analogy, these polymorphisms represent sections of text where certain paragraphs are repeated different numbers of times in different individuals.

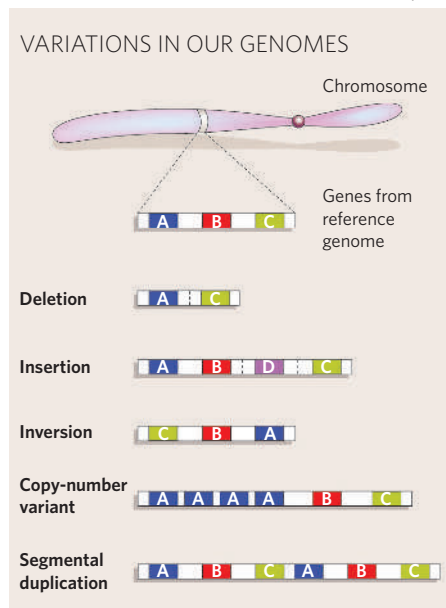
About 76 of the variations Wigler's team found were polymorphisms, and each person had about 11 of them in his or her genome<sup>2</sup>. Soon after, Lee and Scherer reported that in a survey of 55 people they had found 255 copy-number variants, 102 of which were polymorphisms<sup>3</sup>.

### Different strokes

But copy-number polymorphisms tell only part of the story. Earlier this year, a team led by Evan Eichler at the University of Washington in Seattle published evidence of rampant changes of a different sort<sup>4</sup>. Eichler's team compared a portion of an individual woman's genome with the 'reference' human genome sequence produced by the Human Genome Project. The team found 297 potential places where DNA had been rearranged in one of the genomes. Some of these rearrangements were insertions or deletions of stretches of DNA. Others were inversions, where a long stretch of sequence had been reversed in one of the genomes.

To go back to the Book of Life, these places represent missing or repeated pages, or whole sections of text that read backwards (see graphic, left). "We can see this whole new world of variation," Eichler says. "It's going to give us a fresh view of the comprehensive genome, from the single base-pair differences to the really large variants."

Genome researchers now have a catch-all



phrase for the vast array of rearrangements — including copy-number polymorphisms, inversions, deletions and duplications — that occur normally in the human genome. They call it structural variation, and have described at least 800 individual variants that, in total, account for about 3.5% of the human genome. And the sheer number of variants seems likely to catch up with the number of known single nucleotide polymorphisms — the single-letter ‘typos’ in the Book of Life. That makes structural variation a potentially major source of diversity. It is even possible that we’re not all 99.9% similar, as the Human Genome Project predicted.

The biggest question about structural variation is: does it matter? There are already some hints that it does. Eichler’s analysis this year showed that many of the genes found in structural variants negotiate our interactions with the environment<sup>4</sup>. Some make proteins that break down drugs, for example, or help our immune systems respond to disease. So it makes sense that some of these variations explain our unique responses to the stresses or pleasures of life.

### Strength in numbers

In March, a team headed by Sunil Ahuja at the University of Texas Health Science Center, San Antonio, published a new and dramatic proof of that principle<sup>5</sup>. The researchers analysed the average number of copies of an immune-system gene in African, European, Asian and American populations. They found that extra copies of the gene, which makes an immune-system protein called CCL3L1, helped protect people against HIV. If patients with the extra copies became infected by HIV, they progressed more slowly towards full-blown AIDS than those with fewer copies.

There is also good reason to hope that structural variation will shed new light on complex diseases, such as obesity, whose development is triggered by the interaction of many genes, rather than one or two. For one thing, studies such as Eichler’s show that structural variation is involved in our response to the environment — a key factor in complex disease. What’s more, statistical analyses show that regions of structural variation contain genes that are still evolving in humans<sup>6</sup>. If these genes are important enough for evolution to be changing them, they must affect us in some way, for better or worse. And if we can’t detect these effects immediately, it is possible that regions of structural variation are interacting with other parts of our genomes in subtle ways to influence our most crucial traits.

“Given that everywhere we’ve looked hard, we’ve found that copy-number variation influences human disease, it would be strange if complex diseases didn’t also appear on that radar,” says evolutionary geneticist Matt Hurles at the Wellcome Trust Sanger Institute near Cambridge, UK. It’s this hope that is propelling the disease studies already under



## IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

**Structural variation could help explain why some people are prone to obesity (top) and may lead to fresh therapies for this and other diseases.**

way — most notably, Cold Spring Harbor’s \$11-million hunt for copy-number polymorphisms related to autism.

But the link between structural variation and natural selection raises perhaps an even more intriguing possibility. Has structural variation created new mutations that have helped to shape human evolution<sup>7</sup>?

### Brought to book

Eichler has found that copy-number variants often occur in larger blocks of repetitive DNA called segmental duplications<sup>8</sup>. These blocks take up about 5% of the human genome, and occur in the same places in all of our genomes. They also seem to cause trouble. Scientists believe segmental duplications make it difficult for our genome to replicate itself faithfully during the processes that create eggs and sperm. Mistakes result in DNA rearrangements, such as deletions and inversions, meaning that new structural variants are created and passed down the generations.

Icelandic scientists reported earlier this

**“This whole new world of variation is going to give us a fresh view of the comprehensive genome.”**

— Evan Eichler

year, for instance, that 20% of Europeans carry a large genetic inversion that is spreading throughout the population<sup>9</sup>. Women who carry the inversion have more children than those who don’t — a classic sign that it confers some sort of selective advantage.

And studies comparing us with our chimp cousins have already linked structural variation to our divergence from the apes. Last year, scientists from the University of Colorado in Denver and Stanford University found 1,005 genes that differed in copy number among humans and four other primates<sup>10</sup>. This month, Eichler’s group reported 651 likely structural rearrangements between chimps and humans<sup>11</sup>. The group counted 245 genes contained in these variants, including some genes involved in reproduction and drug metabolism. Eichler’s group has also found that segmental duplications have created much more of our genomic differences from chimps than single base-pair differences<sup>7</sup>. There are 177 genes contained within the human-specific duplications. As such duplications are hotspots for evolution, those 177 genes could be partly responsible for creating the traits that make us human.

These genetic differences could also be useful. Scherer’s lab has just released a targeted analysis of inversions between the chimp and human genomes<sup>12</sup>. The group found 1,576 probable inversions, and confirmed 23; three of these differed among human individuals. Not only does this shed some light on primate evolution, but as inversions can often predispose DNA to harmful mutations, these inversions might be involved with human disease. “If you can highlight the structural variations that are inversions, you might be able to highlight where you should look for regions involved in disease,” Scherer says.

In other words, those quirky differences written into the pages of the Book of Life are more than just nonsensical scribbles. Those weirdly rearranged sentences may actually portend life or death. Deciphering the meaning of these cryptic passages could help scientists to diagnose, prevent and treat disease — although that will probably take many years. For now, the realization that we are all reading from individual texts has already altered scientists’ understanding of humanity — and of the library of unique volumes that makes up the human race. ■

**Erika Check is Nature’s Washington biomedical correspondent.**

1. Buckland, P. R. *Ann. Med.* **35**, 308–315 (2003).
2. Sebat, J. et al. *Science* **305**, 525–528 (2004).
3. Iafrate, A. J. et al. *Nature Genet.* **36**, 949–951 (2004).
4. Tuzun, E. et al. *Nature Genet.* **37**, 727–732 (2005).
5. Gonzalez, E. et al. *Science* **307**, 1434–1440 (2005).
6. Check, E. *Nature* **435**, 252–253 (2005).
7. Cheng, Z. et al. *Nature* **437**, 88–93 (2005).
8. Sharp, A. J. et al. *Am. J. Hum. Genet.* **77**, 78–88 (2005).
9. Stefansson, H. et al. *Nature Genet.* **37**, 129–137 (2005).
10. Fortna, A. et al. *PLoS Biol.* **2**, E207 (2004).
11. Newman, T. L. et al. *Genome Res.* **15**, 1344–1356 (2005).
12. Feuk, L. et al. *PLoS Genet.* preprint No. 2005, e56 doi:10.1371/journal.pgen.0010056.eor (2005).