

Cue factors



Our genes were once thought to be responsible for shaping who we are. But now scientists are having a rethink. Thanks to a glut of data from new sequencing projects, researchers are beginning to recognise that the regions of the human genome that encode proteins are unlikely to be behind the millions of differences between people. So the question remains: what accounts for these differences?

Searching for an answer, biologists have pored over the few individual genome sequences that have been completed so far. And these researchers have asked: if the rare stretches of DNA that code for proteins are not responsible for many of the differences found between humans, then what about the remaining 98% of the genome that does not encode proteins – the so-called non-coding DNA?

Small regions of non-coding DNA are known to serve as docking sites for regulatory proteins called transcription factors, which are responsible for cueing when and where genes are turned on and off. These tiny transcription factor-binding sites are usually found just upstream of genes – although they can occur anywhere within the genome – and they recruit other proteins that are essential for transforming genes into their respective proteins.

Scientists now suspect that variation in the efficiency of transcription factor binding, and the effect of this binding on how genes are turned on, could be responsible for many of the little quirks that make each of us unique. Jan Korbelt, a genome biologist from EMBL Heidelberg, is particularly interested in variation among people. So he teamed up with Stanford University geneticist Michael Snyder to investigate the effect of differences in transcription factor binding on the variation in gene expression in 10 humans – five Europeans, three Africans, and two East Asians – by looking at their complete genome sequences.

By chemically gluing transcription factors to their DNA-binding sites, the researchers identified the binding patterns

of these regulatory proteins and looked to see whether they bound every site equally strongly in all 10 people. The proteins they studied were NFkB, which is involved in the immune response, and Pol-II, a protein that helps convert DNA to RNA. For NFkB, less than 10% of the 15 000 binding sites varied between people. Yet for Pol-II, about a quarter of its 19 000 binding sites were altered across individuals – considerably more variation than is seen in coding regions, which vary, on average, by only a fraction of a percent among people.

Jan's group was able to trace some of this variation back to either small differences or larger rearrangements in the DNA comprising these binding sites. Still, the majority of binding sites showed no difference in their DNA sequences, despite their variable binding affinity among people. In these cases, Jan explains, the researchers frequently found variations in nearby regions that could explain the differences. "This is a very interesting finding, as it suggests that these regulatory proteins are not working alone, but instead co-operating with nearby transcription factors to regulate how genes are expressed," he says. In fact, Jan's team has just developed a new test for assessing the extent of this transcription factor co-operation.

Jan and his team went on to show that the variation at these binding sites had a profound impact on an individual's gene expression. "Many genes' expression differed by greater than two orders of magnitude between people as a result of variation in transcription factor binding," says Jan.

"What makes this study novel is that it determined how differences in gene expression among people are associated with the genome-wide variation in transcription factor binding, and also with sequence variation in the factor's binding sites," remarks Jan. "Our findings suggest that variation in non-coding DNA may be responsible for many of the differences we see between people."



Lars Steinmetz and Jan Korbel

What's more, these findings also offer an explanation for the differences between humans and their closest cousins. When Jan's team looked in both humans and chimpanzees, they found that nearly a third of the Pol-II-binding sites differed between the two species. Jan explains that there seems to be almost as much variation among humans as between humans and chimpanzees.

The researchers' results were buoyed by a second study in which genome biologist Lars Steinmetz of EMBL Heidelberg, also in collaboration with the Snyder group, used yeast to map genome regions bound by a transcription factor called Ste12 that is important in yeast mating.

Lars and his team compared binding sites between two strains of the common baker's yeast *Saccharomyces cerevisiae*, one of which descended from cells collected from a rotten fig in Merced, California, and the other derived from cells isolated from the lung of a person suffering from AIDS-related immune problems in San Francisco.

The yeast's small genome and short generation time offered Lars and his group the opportunity to fine-map these sites across many more individuals than had previously been possible using human cells. This gave the researchers the power to detect very small differences in transcription factor binding. Owing to the sensitivity of this approach, Lars' group found even higher rates of variation in Ste12 transcription factor binding between the yeasts than was seen in the human study, and because of yeast's frequent genome shuffling, they were able to ascribe this variation to very narrow regions of the genome.

"Our results, and Jan's study involving humans cell lines, suggest that the bulk of the differences among individuals are not found in the genes themselves, but in the non-coding portions of the genome and in particular in the regulatory regions, which we know relatively little about," says Lars.

When mapping the DNA responsible for the variation in transcription factor binding, Lars and his team found that

almost 90% of the binding differences are influenced by sequence variation in nearby regions. This offers tantalising evidence that variation in the regulation of genes by transcription factors is often controlled by sequence differences close to the genes. However, in several cases the data revealed variations far from the genes that also influence gene expression. These regions can encode the transcription factors themselves or signalling molecules that influence the efficiency of a certain transcription factor's binding ability.

Knowing whether genes are regulated by nearby regions or regions on entirely different chromosomes will determine how scientists study variation in gene regulation. If gene expression variation is indeed often controlled locally – by so called cis-regulatory elements – then researchers are safe to continue to take a reductionist approach and look to ascribe the differences between individuals to the regulatory regions around genes. However, if variation in gene regulation turns out to be under the control of regions scattered far and wide throughout the genome – often termed trans-regulatory variation – then researchers need to change tack and study the whole genome simultaneously. Such a broad approach requires the skills of system biologists who specialise in studying entire, intact biological systems.

Either way, once scientists have mapped these regulatory regions, they can look at how individual variations in gene regulation affect how we look, how we think, whether we are susceptible to certain diseases and how we will react to specific treatments. This will open doors to personalised medicine, which seeks to make use of an individual's differences to provide better medical care.

Kasowski M *et al* (2010) Variation in transcription factor binding among humans. *Science* **328**: 232-235

Zheng W, Zhao H, Mancera E, Steinmetz L, Snyder M (2010) Genetic analysis of variation in transcription factor binding in yeast. *Nature* **464**: 1187-1191