TRANSCRIPTOMICS

Every transcript counts

It is becoming clear that much of the genome is transcribed, not just previously annotated genes. But is this just transcriptional noise or is it functionally important? Svante Pääbo and colleagues have found that the expression of intergenic transcripts is subject to the same functional constraints as that of classical exons.

Genome annotation estimates that the human genome encodes 20,000–25,000 protein-coding transcripts. However, evidence from tiling arrays, which measure transcription without regard to existing annotation, indicates that much of the intergenic DNA is also transcribed. It is unclear why this DNA is transcribed — it could be unidentified RNA genes or the unannotated exons of existing genes; alternatively, it might just be noise from the transcriptional machinery, as the generally low levels of intergenic transcription and lack of sequence conservation between mice and humans suggest.

Pääbo and colleagues used tiling arrays to study expression in several tissues from the 1% of the human and chimpanzee genomes that have been characterized by the ENCODE project. They found that a higher proportion of exonic sequence is expressed, but, as the genome contains more intergenic sequence, the majority of the expressed sequence was intergenic. They also found that a lot of intergenic sequence is expressed at low levels, but that a significant proportion shows expression levels that are comparable to that of the exonic sequence. This is true for all but the most highly expressed 5% of sequences, which are predominantly from exonic regions.

The authors' previous work on expression differences in protein-coding genes between the two species showed that expression in the brain has diverged less than in other tissues because of stabilizing selection, and in the testes it has

diverged more because of positive selection. These differences between tissues are only expected to hold for intergenic regions if their expression is functionally constrained — otherwise there is no reason to expect that random variations in expression will be of different magnitudes in different tissues. Although the overall level of divergence was slightly higher for intergenic regions, implying fewer constraints, the differences between tissues followed the same patterns. It is unclear whether this is because all intergenic transcription is slightly less constrained than exonic transcription, or because a proportion of intergenic transcription really does represent unconstrained noise.

Interestingly, in brain and testes, but not other tissues, the authors saw a correlation between the expression level of exonic probes and that of upstream, but not downstream, intergenic probes. This implies that the probes are detecting the same transcript and therefore that alternative promoters are widely used in these tissues.

These results not only demonstrate the extent of intergenic transcription, but they show that it is likely to be functionally relevant and that it might be as important in explaining the differences between species as exonic transcription. In most cases, the functions themselves remain to be discovered, as does the relationship between these expression differences and the underlying DNA sequence variation.

Patrick Goymer

ORIGINAL RESEARCH PAPER Khaitovich, P. et al. Functionality of intergenic transcription: an evolutionary comparison. *PLoS Genet.* 28 August 2006 (doi:10.1371/journal.pgen.0020171.eor)

FURTHER READING Khaitovich, P. et al. Evolution of primate gene expression. *Nature Rev. Genet.* 7, 693–702 (2006)

WEB SITE

ENCODE project: <http://www.genome.gov/10005107>

RNA INTERFERENCE

RNAi misses the target

It is well known that RNAi screens can produce spurious candidates through off-target effects. In *Drosophila*, the use of long dsRNA has been thought to minimize this problem, but Philip Beachy and colleagues now show that off-target effects can dominate the results of a genome-wide RNAi screen in this species.

The authors were looking for novel components of the Wingless (Wg) signalling pathway using more than 20,000 dsRNAs. They found seven dsRNAs that reduced expression of their reporter below a particular threshold. However, further analysis showed that the effects were specific to these dsRNAs and were not replicated by most other dsRNAs that targeted the same ORFs.

By systematically substituting bases in one of the original dsRNAs, the authors honed in on the core sequence that was required for the Wg-repressing effect. A homology search showed that this was a perfect match for part of the mRNA sequence of *armadillo*, a known Wg-pathway component. Equivalent homology was found in the other dsRNAs, and quantitative PCR showed that *armadillo* expression was decreased — in other words, all the effects were off-target.

Further analysis of this screen, an unpublished Hedgehog screen and a previously published Wg screen showed that off-target effects were prevalent. In some cases this was because of a particular trinucleotide repeat, CAN, that is found in many genes. Although these results might seem discouraging, the authors point out that excluding certain sequences from dsRNAs and always using multiple dsRNAs for each gene will significantly reduce the false-positive rate.

Patrick Goymer

ORIGINAL RESEARCH PAPER Ma, Y. et al. Prevalence of off-target effects in *Drosophila* RNA interference screens. *Nature* 10 September 2006 (doi:10.1038/nature05179)