


 TRANSCRIPTOMICS

Throwing light on dark matter

Reports over the past few years of extensive transcription throughout eukaryotic genomes have led to considerable excitement. But doubts have been raised about the methods that have detected this pervasive transcription and about how much of it is functional. A recent study confirms these doubts, suggesting that the so-called transcriptomic 'dark matter' — the RNA that is not associated with annotated genes — is far less extensive than previously predicted and might mainly be explained by transcriptional noise.

The studies that have suggested the most extensive intergenic transcription have used tiling microarrays. However, investigations using more recently developed RNA sequencing (RNA-seq) approaches indicate that transcripts outside genes are actually far rarer. A direct comparison of the two methods is now provided by van Bakel and colleagues, who used both approaches to analyse the expression of polyadenylated RNA in a series of mouse

and human tissues. They confirm that tiling microarrays, in contrast to RNA-seq, are susceptible to a high rate of false-positives in identifying transcripts with low expression levels, suggesting that some estimates of the level of intergenic transcription may have been excessively large.

To investigate this possibility, the authors used RNA-seq to determine the proportion of polyadenylated transcripts that can be classified as 'dark matter'. The vast majority of the reads came from within or close to coding genes and most of the remaining 'dark matter' was mapped to introns, suggesting that it is a by-product of mRNA expression. After excluding other common categories, just 2.2–2.5% of the transcripts arose from intergenic regions, and in most cases the levels of these transcripts were so low that the authors propose transcriptional noise as the most likely explanation.

Although their data argue against extensive functions for intergenic

transcripts, van Bakel and colleagues did see expression above background levels for several thousand intergenic regions (still a small proportion of overall transcripts), suggesting potential functions. They provide evidence that most of these transcripts are likely to be non-coding and note that many of them arise from regions that are predicted to contain open chromatin, which suggests possible regulatory roles. As the authors highlight, the only way to ultimately confirm the function of any of this transcriptional 'dark matter' will be to examine the effects of ablating or altering these transcripts.

Louisa Flintoft

ORIGINAL RESEARCH PAPER van Bakel, H., Nislow, C., Blencowe, B. J. & Hughes, T. R.

Most 'dark matter' transcripts are associated with known genes. *PLoS Biol.* **8**, e1000371

FURTHER READING Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Rev. Genet.* **10**, 57–63 (2009) | Jacquier, A. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nature Rev. Genet.* **10**, 833–844 (2009)