➡ **TRANSCRIPTOMICS**

Digging deep with RNA-Seq

The emerging view of eukaryotic transcriptomes is complex, involving overlapping transcripts, transcribed intergenic regions and abundant non-coding RNAs. A recently developed method — known as RNA-Seq — that uses high-throughput sequencing promises a more comprehensive understanding of this complexity.

RNA-Seq involves direct sequencing of cDNAs using high-throughput sequencing technologies, allowing the level of transcription from a particular genomic region to be quantified from the density of corresponding reads. Unlike array-based approaches, RNA-Seq gives a potentially comprehensive view of the transcriptome. Another advantage is its ability to provide information on transcripts that are expressed at very low levels, limited only by the total number of reads that are generated.

Two new studies have surveyed the transcriptomes of model yeast species. An investigation of the *Schizosaccharomyces pombe* transcriptome showed RNA-Seq to be a highly sensitive method that, unlike hybridization-based approaches, is subject to little background noise. More than 90% of the genome was found to be transcribed, and a striking amount of variation in splicing efficiency was uncovered, with differences across introns, genes and conditions. In the case of *Saccharomyces cerevisiae*, almost 75% of the non-repetitive portion of the genome was represented in the transcriptome. The authors report widespread use of alternative initiation codons and upstream ORFs, extensive heterogeneity at the 3′ ends of transcripts, and large numbers of overlapping genes.

As their genomes are more complex — in terms of size, number of introns and gene duplicates — applying RNA-Seq to species such as mice and humans requires a higher degree of sophistication, particularly in terms of data analysis. Two studies using mouse cells from a range of tissues as well as stem cells, and another using human cells, have taken on this challenge. Between them they identify large numbers of previously unknown transcripts, provide new information about the positions of promoters, exons and 3′ ends, and highlight the enormous level of transcript diversity that can be generated by alternative splicing in mammals — the study in human cells identified 4,096 previously unknown splice junctions in 3,106 genes.

RNA-Seq can be combined with other genome-wide methods to provide an integrated view of gene regulation. This approach is illustrated by a study in *Arabidopsis thaliana* that combined the use of RNA-Seq to probe the mRNA and small RNA transcriptomes with a genome-wide survey of DNA cytosine methylation, revealing relationships between sites of transcription and different epigenetic modes of regulation.

In themselves, these papers provide many new findings for further investigation. They also highlight how much more we have to learn about the complexity of eukaryotic transcriptomes, paving the way for many more studies in a range of species and cell types.

Louisa Flintoft

ORIGINAL RESEARCH PAPERS Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008) | Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349 (2008) | Wilhelm, B. T. *et al.* Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**, 1239–1243 (2008) | Mortazavi, A. *et al.* Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628 (2008) | Cloonan, N. *et al.* Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods* **5**, 613–619 (2008) | Sultan, M. *et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 3 Jul 2008 (doi:10.1126/science.1160342)
FURTHER READING Marioni, J. *et al.* RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 11 Jun 2008 (doi:10.1101/gr.079558.108) | Morin, R. D. *et al.* Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *BioTechniques* **45**, 81–94 (2008) | Kapranov, P., Willingham, A. T. & Gingeras, T. R. Genome-wide transcription and the implications for genomic organization. *Nature Rev. Genet.* **8**, 413–423 (2007) | Sandelin, A. *et al.* Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature Rev. Genet.* **8**, 424–436 (2007)