

MILESTONE 7

Transcriptomes — a new layer of complexity

Characterizing genes has been a primary objective for researchers in genetics and genomics. However, direct measurement of the gene product, transcribed RNA, provides a functional read-out of the genome that is essential for our understanding of development and disease.

The early 2000s saw both hybridization-based and sequencing-based approaches being used for the transcriptome-wide quantification of RNA. However, microarray technologies were limited in their ability to detect very lowly expressed genes (sensitivity) and differentiate between genes with sequence homology (specificity). An additional limitation is that only known genes and exons are incorporated into the array, and *de novo* discovery is not possible. By contrast, sequence-based approaches such as serial analysis of gene expression (SAGE) or massively parallel signature sequencing (MPSS) relied on expensive Sanger sequencing and

were limited in their ability to detect all transcript isoforms.

The emergence of next-generation DNA sequencing (MILESTONES 3, 5) enabled the development of high-throughput sequencing of whole transcriptomes, known as RNA sequencing (RNA-seq), reported in 2008 in a series of milestone publications across different species. A typical RNA-seq experiment works by isolating messenger RNA (mRNA) with a poly(A) tail, reverse-transcribing the RNA into complementary DNA (cDNA), sequencing this cDNA using a next-generation sequencing instrument and mapping the resulting reads to the reference genome. RNA-seq reveals an overall picture of which parts of the genome are transcribed and enables accurate RNA quantification at higher resolution and greater dynamic range than previous methods because of its digital read-out, allowing the detection of lowly expressed transcripts. It also

“ The ability to discover new genes or transcripts, the high throughput and the scalability of the technology are major advantages of RNA-seq ”

permits new genes, exons and transcript isoforms to be identified.

A study by Nagalakshmi et al. demonstrated the power of RNA-seq by sequencing the yeast transcriptome. The majority of reads mapped to known yeast genes, but many reads mapped to regions of the genome that had previously not been known to be transcribed. This study demonstrated the capacity of RNA-seq to capture gene boundaries at unprecedented resolution, including variability in 3' untranslated regions, which are important for transcript stability and localization. Additionally, it revealed the complexity of the eukaryotic coding genome, uncovering many overlapping genes and alternative transcription start sites.

A contemporaneous study by Lister et al. showed how RNA-seq can be integrated with other genomic data sources to extract functional information about the genome. Here, the researchers combined bisulfite sequencing with RNA-seq in *Arabidopsis thaliana* to study the influence of DNA methylation on the transcriptome.

The ability to discover new genes or transcripts, the high throughput and the scalability of the technology are major advantages of RNA-seq over previous gene expression profiling methodologies. In a study on fission yeast, Wilhelm et al. were able to detect transcriptomes at high resolution over time and under different experimental conditions, capturing the dynamic nature of transcription.

The impact of RNA-seq is far-reaching, and its full potential has yet to be realized. By helping to define the functional genome and quantify RNA over time and under changing conditions, RNA-seq has had a profound impact on research in fields across genetics, biology and medicine, in which it has become a staple research tool.

Margot Brandt, *Nature Communications*

ORIGINAL ARTICLES Nagalakshmi, U. et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349 (2008) | Lister, R. et al. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008) | Wilhelm, B. T. et al. Dynamic repertoire of a eukaryotic transcriptome surveyed at single nucleotide resolution. *Nature* **453**, 1239–1243 (2008)
FURTHER READING Cloonan, N. et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* **5**, 613–619 (2008) | Mortazavi, A. et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008) | Velculescu, V. E. et al. Serial analysis of gene expression. *Science* **270**, 484–487 (1995) | Brenner, S. et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **18**, 630–634 (2000) | Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009)



Credit: Y H Lim / Alamy Stock Photo