

Multitudes of messages

Tissue-specific expression of mRNA isoforms and interindividual differences in isoform usage can now be quantitated by genome-wide assays, both by custom microarray and by next-generation sequencing.

Multicellular organisms are complex in many different ways, with the cells of different tissues transcribing, processing, transporting and translating different subsets of their DNA code to produce proteins. From sampling of expressed sequence tags and from cDNA collections, we knew that about two-thirds of human genes express at least one spliced variant. More exhaustive investigation with new tools now shows this source of variation to be nearly universal.

These tools now include targeted custom microarrays of increasing sophistication that are economical and reliable to use, but that may still fail to detect unexpected structures. Newer mRNA sequencing methods have the advantage of larger and extensible dynamic range (rare RNAs can be found among an excess of abundant transcripts by increasing sequencing depth) and are not biased against the discovery of new isoforms. However, all forms of sequencing are still relatively expensive, and only a proportion of the short reads from the highest-throughput platforms can be unambiguously mapped to the reference genome build.

On page 1413, Benjamin Blencowe and colleagues generate mRNA-Seq datasets comprising tens of millions of 32-nucleotide reads for a variety of human tissues, finding new splice junctions in one in five multiexon genes and that nearly all multiexon genes undergo alternative splicing, with a mean of seven such events per gene.

Jason Johnson and colleagues built custom microarrays containing probes monitoring exon–exon junctions for nearly 18,000 annotated human genes and use these to characterize the splicing events from 48 human tissues and cell lines (p 1416). By investigating sequence motifs near exon junctions and correlating these to tissue differences in splicing, they are able to deduce that tissue-specific splicing is likely to be influenced by RNA-binding proteins that are expressed in the corresponding tissue. For some such proteins, a consensus motif and a role in alternative splicing is already known.

After applying the mRNA-Seq technique to 15 human

tissues and cell lines, Eric Wang *et al.* (*Nature* advance online publication 2 November 2008; doi:10.1038/nature07509) note not only the differences between tissues, but also the remarkable conservation of alternative splicing patterns across tissues. Their investigation of binding motifs for proteins and miRNAs on or near exons produced another interesting result: FOX-1, FOX-2 and CELF RNA-binding proteins may influence not only tissue-specific alternative splicing, but also alternative polyadenylation events.

Interindividual variation in splicing seems much less prevalent than tissue-specific differences. By examining mRNA of lymphoblastoid HapMap cell lines on exon tiling microarrays, Tony Kwan *et al.* (*Nat. Genet.* **40**, 225–231; 2008) found isoform differences in 178 of 324 genes where the transcription differences correlated with nearby SNPs, 85 of which could be accounted for by alternative splicing alone. Wang *et al.* believe that the exon microarray approach underestimated and that there may be interindividual variation in as many as 21% of alternatively spliced genes. Clearly this question should be reexamined via mRNA-Seq of different individuals.

The proportion of alternative exons that generate functional differences needs to be explored, but at what level ‘function’ should be deduced is far from clear. For example, the implications of isoform redundancy for the evolution of intron-containing genes are just beginning to be investigated. We now know that mutations can frequently arise not singly, but in highly correlated clusters within approximately 30-kb regions (Wang *et al. Proc. Natl. Acad. Sci.* **104**, 8403–8408; 2007). If it is resistance to mutation and not tissue-specific function that drives the evolution of alternative splicing, tissue-specific consequences of mutation may be important.

It would seem that a suitable use of next-generation sequencing would be to accurately determine the rate and mode of mutation of a range of genes in a number of individuals and that this investigation should look at mutation rates both in germline and in a range of somatic tissues. ■