


 TRANSCRIPTOMICS

Seq-ing regulatory variants

Speculation that next-generation sequencing of RNA (RNA-seq) will connect genome variants with variation in gene expression levels has moved towards reality with the publication of two papers describing RNA-seq on well-characterized human populations.

Gene expression levels are partly controlled by genetic variation, and microarray expression data have been used to look for associations between sequence variants and expression levels (expression quantitative trait loci (eQTLs)). However, RNA-seq provides nucleotide-level resolution and more accurate quantification than microarrays.

Montgomery *et al.* and Pickrell *et al.* each sequenced polyadenylated RNA from lymphoblastoid cell lines derived from ~60 individuals — from Caucasian and Nigerian populations, respectively — that have been extensively genotyped as part of the HapMap project. Using genome-wide collections of SNPs, both studies found a greater number of statistically significant eQTLs than had been identified by

microarray studies. Overlap among the eQTLs found in the two studies and with previous studies suggests that the eQTLs stem from replicable genetic effects.

Both studies agree with previous studies which found that most eQTLs lie close to gene transcriptional start sites or in the 3' UTR. Furthermore, the same data that are used to find eQTLs can also be used to assay allele-specific expression (ASE). Therefore, these studies were able to show directly that eQTLs modulate expression in *cis*. However, the identified eQTLs did not explain all ASE. For example, Montgomery *et al.* found greater haplotype homozygosity when two or three individuals shared an ASE signal, suggesting that recent rare eQTLs — not detectable through standard genotypic association — could be responsible for rare ASE effects.

A further advantage of RNA-seq over microarrays is the improved ability to quantify levels of transcript isoforms. Both studies highlight the effect of genetic variation on exon inclusion and are likely to spur

future research into this aspect of alternative splicing.

Finally, Pickrell *et al.* used their deep RNA-seq data to examine the completeness of current gene annotations. They found >4,000 unannotated exons and a substantial number of new polyadenylation sites.

The combination of RNA-seq with genetic variation data is likely to build on the success of microarray-based studies in identifying disease-associated variants and increasing understanding of gene regulatory architecture and, furthermore, may contribute to our mechanistic understanding of why an individual's transcriptome is unique.

Mary Muers

ORIGINAL RESEARCH PAPERS

Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* 10 Mar 2010 (doi:10.1038/nature08903) | Pickrell, J. K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 10 Mar 2010 (doi:10.1038/nature08872)

FURTHER READING Cheung, V. G. & Spielman, R. S. Genetics of human gene expression: mapping DNA variants that influence gene expression. *Nature Rev. Genet.* **10**, 595–604 (2009)