An Integrated RNA-seq Atlas of the Murine T-Helper Cell Transcriptome

Andrew Deonarine
MRC Laboratory of Molecular Biology, Hills Road, Cambridge, UK
http://www.mrc-lmb.cam.ac.uk/tcb/

T-helper cells play an important role in mediating the immune response, and with the advent of next generation sequencing, significant insights can be gained into the T-helper cell transcriptome. One of the barriers to analyzing next-generation sequencing data, such as that generated by RNA-seq analyses, is that many of the statistical properties concerning quantification (ie. RPKM [1] vs. FPKM [2]), normalization [3], and differential expression (using methods such as edger [4], DESeq [5], and Cuffdiff [6]) of the data are still not clearly understood. Building on previous investigations into the bimodality of transcript expression [7], a computational pipeline was created to integrate various methods of expression quantification, cell type clustering, differential expression analyses, gene annotation methods, and novel transcript identification into a murine T-helper cell expression atlas. By integrating these various analyses, it was possible to identify key signature genes (transcription factors, cytokines, receptors, and other molecules) that distinguish the various T-helper cell types from each other, in addition to novel transcripts. This expression atlas, which is easily accessible as a user-friendly online resource at http://www.thelpercell.com, will form the basis for future investigations into immune regulation and function using network-based analyses.

This work is relevant to the goals of SEQAHEAD because it represents a major step forward in the integration and comparison of various methods of expression quantification, differential expression analysis, and annotation of RNA-seq data. The computational principles presented here could potentially be applicable to many other fields of molecular biology and medicine.

References

1. Mortazavi, A., Williams, BA., McCue K., Schaeffer, L., Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat Methods* (2008) 5: 621-8.
2. Roberts, A., Trapnell, C., Donaghey, J., Rinn, JL., Pachter, L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* (2011) 12: R22.
3. Robinson, MD., Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* (2010) 11: R25.
4. Robinson, MD., McCarthy, DJ., Smyth, GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* (2010) 26: 139-40.
5. Anders, S., Huber, W. Differential expression analysis for sequence count data. *Genome Biol* (2010) 11: R106.
6. Trapnell, C. Cufflinks Manual. Downloaded from http://cufflinks.cbcb.umd.edu/manual.html on Sept. 12th, 2011.
7. Hebenstreit, D., Fang, M., Gu, M., Charoensawan, V., van Oudenarrden, A., Teichmann, SA. RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Mol. Syst. Biol* (2011) 7: 497.

Relevant Web sites

8. http://www.thelpercell.com