Visualizing Diseased Transcriptomes with *R*

Tracy Nance[1*]**, Stephen B. Montgomery**[1]

1. Department of Pathology, Stanford University, USA
*Contact author: tracy.nance@gmail.com

Keywords: rna-seq, IPF, genes, visualization, transcriptome

Recent advances and decreasing costs in high-throughput sequencing have made possible the generation of vast quantities of biological data. Hundreds of experiments analyzing gene expression in diseases have been performed using RNA sequencing or older microarray technology, and a correspondingly large amount of statistical machinery has been built within R to analyze this data. However the results of these studies are usually presented to the community as lists of genes with p-values, and are difficult for biological researchers to reproduce, query or visualize.

We performed RNA-Seq analysis on lung tissue from 8 individuals with idiopathic pulmonary fibrosis (IPF) and 7 healthy controls, and found evidence for substantial differential gene expression and differential splicing in IPF as compared to healthy lungs [5]. Using a combination of R packages like **shiny** [6], **ggplot2** [8], **DESeq** [1], **DEXSeq** [2], **limma** [7], and the *JavaScript* library **D3** [3], we created a web application [4] which allows researchers to explore our results interactively, and to compare them with previously published microarray studies of the same disease. This application shows the power and ease of combining multiple R packages for visualization and making them available to the public, and we hope that it will serve as a toy model for making results accessible in large-scale genomic and transcriptomic studies.

References

[1] Anders, S. and W. Huber (2010). Differential expression analysis for sequence count data. *Genome Biology 11*, R106.

[2] Anders, S., A. Reyes, and W. Huber (2012, Oct). Detecting differential usage of exons from RNA-seq data. *Genome Res. 22*(10), 2008–2017.

[3] Bostock, M., V. Ogievetsky, and J. Heer (2011). D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*.

[4] Nance, T. (2014). IPF Explorer. http://stanford.edu/~tnance/ipf.html.

[5] Nance, T., K. S. Smith, V. Anaya, R. Richardson, L. Ho, M. Pala, S. Mostafavi, A. Battle, C. Feghali-Bostwick, G. Rosen, and S. B. Montgomery (2014). Transcriptome Analysis Reveals Differential Splicing Events in IPF Lung Tissue. *PLoS ONE 9*(3), e92111.

[6] RStudio and Inc. (2013). *shiny: Web Application Framework for R*. R package version 0.7.0.

[7] Smyth, G. K. (2005). Limma: linear models for microarray data. In R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber (Eds.), *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, pp. 397–420. New York: Springer.

[8] Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York.