Introduction to RNASeq

Malay (malay@uab.edu)

March 2, 2015

Contents

1 Introduction

RNASeq is a very vast topic and tons of papers have been and are being written on the topic. The following is just an overview.

Originally the idea was proposed by Mortazavi et al. (2008). Although several modifciation of the original idea have been developed, the basics did not change. In this handout will use the latest in the RNASeq methodology through the use of software called RSEM (Li and Dewey 2011).

2 Normalization of RNASeq data

People have proposed several methods of normalization of RNASeq data. For a comparison see Dillies et al. (2013).

3 RSEM

RSEM is a cutting-edge RNASeq analysis package that is an end-to-end solution for differential expression, and simplifies the whole process. It also intriduces a new more robust unit of RNASeq measurement called TPM.

3.1 Datasets

Every diffential expression measurment should have biological replicates. For demonstration, we will use only 1 replicate for two biological conditions. But in real life, this should never be used. We will use two small datasets from Illumina Body Map project. These are samples prpared from adrenal gland and brain and only from chromosome 19. You can download the datasets here:

http://cmb.path.uab.edu/training/docs/CB2-201-2015/rnaseq_data.tar.gz

Unzip the file.

3.2 Installing RSEM

```
wget http://deweylab.biostat.wisc.edu/rsem/src/rsem-1.2.19.tar.gz
tar -xvzf rsem-1.2.19.tar.gz
cd rsem-1.2.19/
make
export PATH=$PATH:`pwd`

# Install ebseq
module load R/R-3.1.2
make ebseq
cd EBSeq/
export PATH=$PATH:`pwd`
```

3.3 Install Bowtie

Download Bowtie from http://sourceforge.net/projects/bowtie-bio/files/bowtie/1.1.1/

3.4 Prepare reference

```
rsem-prepare-reference --gtf human_chr19.gtf --bowtie --bowtie-path bowtie-1.1.1/ chr19.fa ref/chr19
```

3.5 Calculate expression

```
rsem-calculate-expression --paired-end brain_R1.fq brain_R2.fq ref/chr19 human_chr19
rsem-calculate-expression --paired-end adrenal_R1.fq adrenal_R2.fq ref/chr19 adrenal_chr19
```

3.6 Differential expression

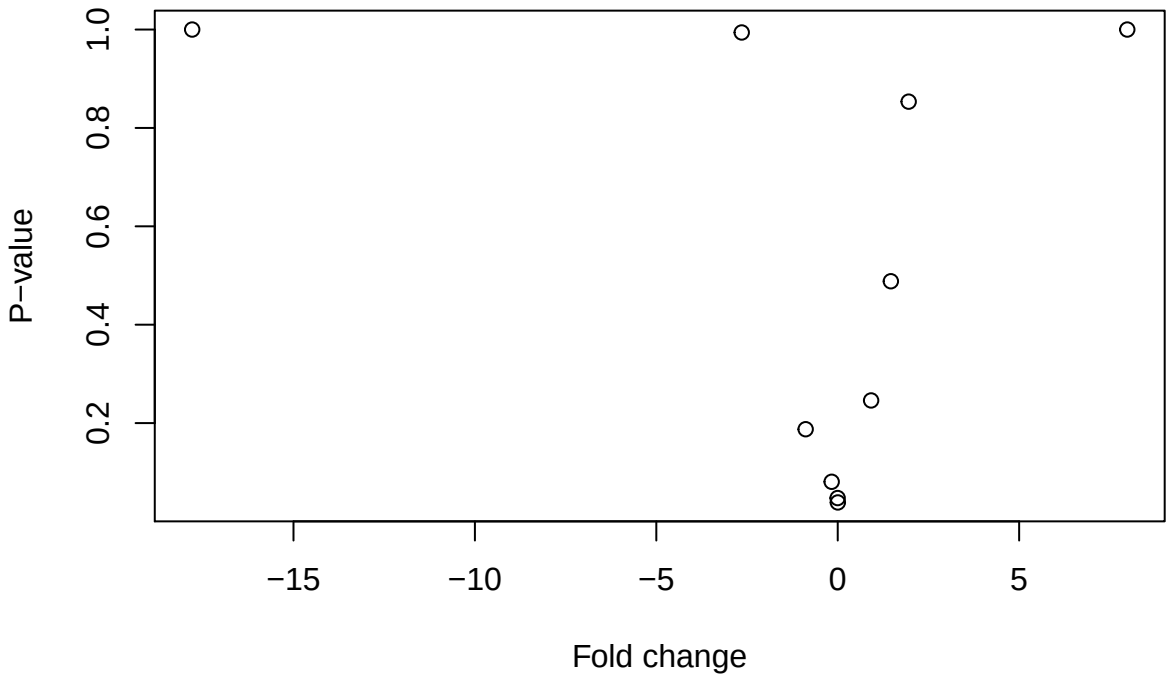
```
rsem-generate-data-matrix adrenal_chr19.genes.results human_chr19.genes.results >diff-brain-adrenal.t
rsem-run-ebseq diff-brain-adrenal.txt 1,1 expression.results.txt
rsem-control-fdr expression.results.txt 0.05 expression_final.txt
```

And we have our differentially expressed genes.

3.7 Volcano plot

Volcano plot is a good way to show the differentially expressed genes. For that we need the p-value for the differentially expressed genes and the the fold change. Given by "PPEE" and "RealFC" values.

```
data<-read.table("expression.results.txt")
plot(log2(data$RealFC),data$PPDE,xlab="Fold change",ylab="P-value")
```



Bibliography

Dillies, Marie-Agnès, Andrea Rau, Julie Aubert, Christelle Hennequet-Antier, Marine Jeanmougin, Nicolas Servant, Céline Keime, et al. 2013. "A Comprehensive Evaluation of Normalization Methods for Illumina High-Throughput RNA Sequencing Data Analysis." *Brief Bioinform* 14 (6): 671–83. doi:10.1093/bib/bbs046.

Li, Bo, and Colin N. Dewey. 2011. "RSEM: Accurate Transcript Quantification from RNA-Seq Data with or Without a Reference Genome." *BMC Bioinformatics* 12 (1): 323. doi:10.1186/1471-2105-12-323.

Mortazavi, Ali, Brian A. Williams, Kenneth McCue, Lorian Schaeffer, and Barbara Wold. 2008. "Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq." *Nat. Methods* 5 (7): 621–28. doi:10.1038/nmeth.1226.