

Differential Expression with Diagnostic Plots using sam2counts and DESeq

Nikhil Joshi

Bioinformatics Core

UC Davis Genome Center

sam2counts and DESeq

- Sam2counts
 - Python script that creates a counts table from reads aligned to transcripts.
 - Input in Galaxy
 - SAM files of sample read alignments to a transcriptome
 - Sample names
 - Output in Galaxy
 - Counts table of counts for each transcript per sample

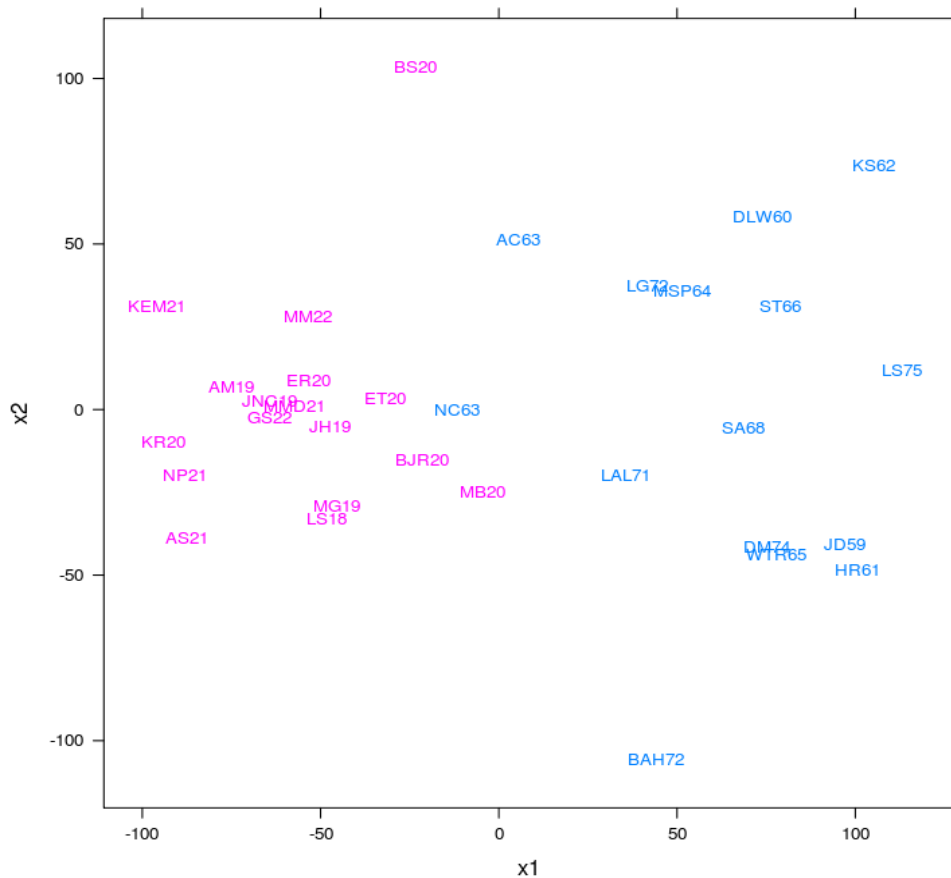
sam2counts and DESeq

- DESeq
 - Bioconductor package in R that is used to analyze differential expression data
 - Input in Galaxy
 - Counts file from sam2counts from reads aligned to a transcriptome
 - Column types and comparison types
 - Output in Galaxy
 - Toplevel sorted by adjusted p-value
 - Page of diagnostic plots

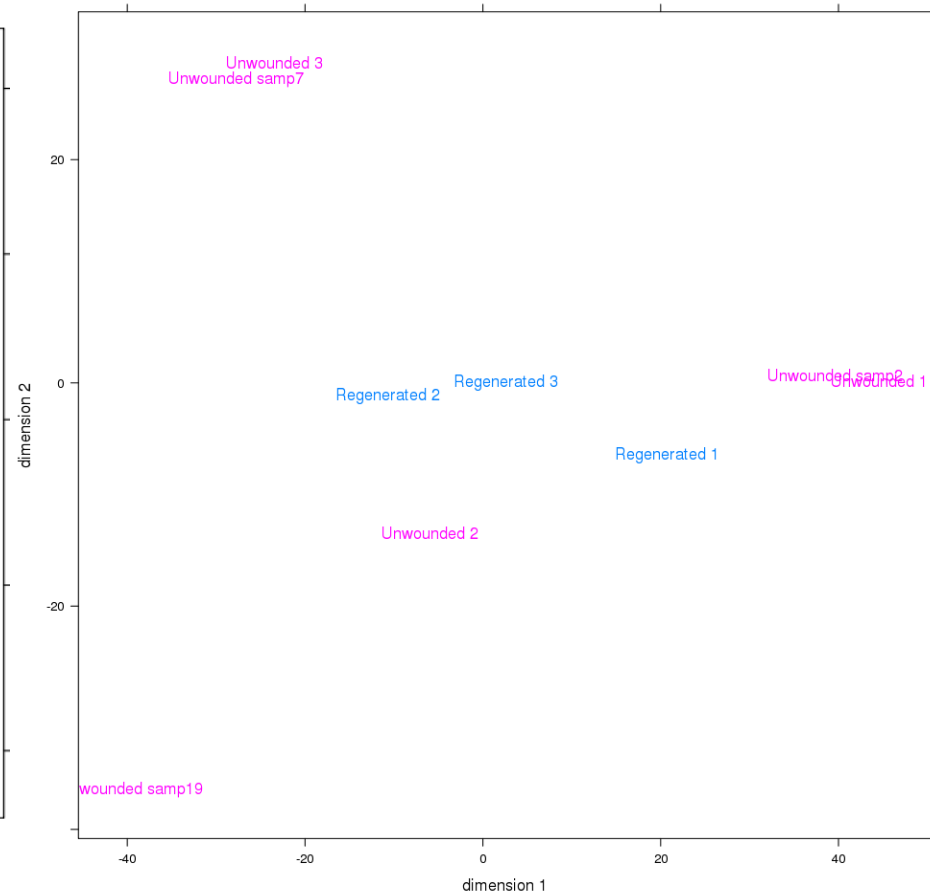
MDS Plot

Represent distances in k-dimensional space in 2 dims

Multi-Dim scaling plot for 32 samples

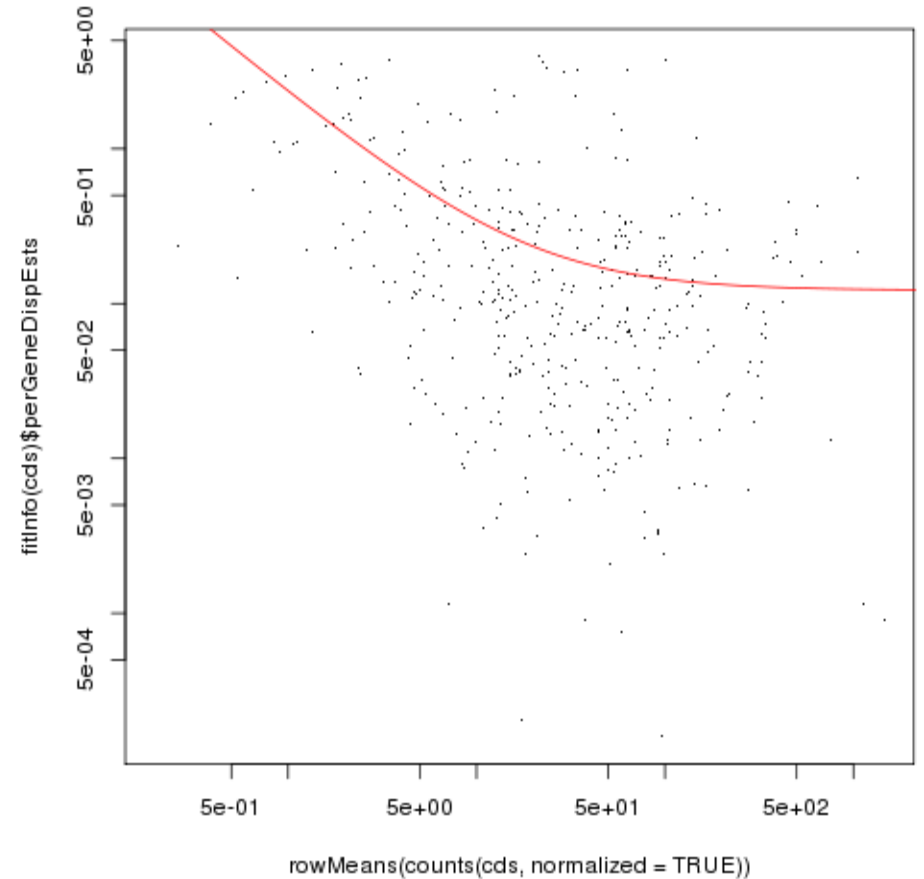
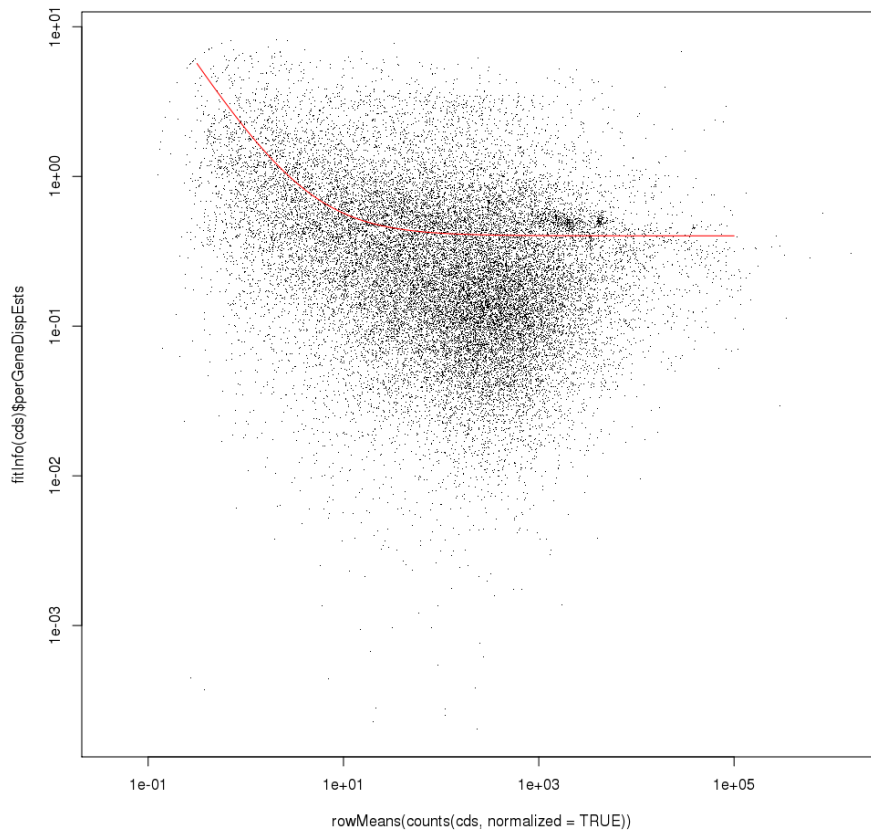


MDS plot



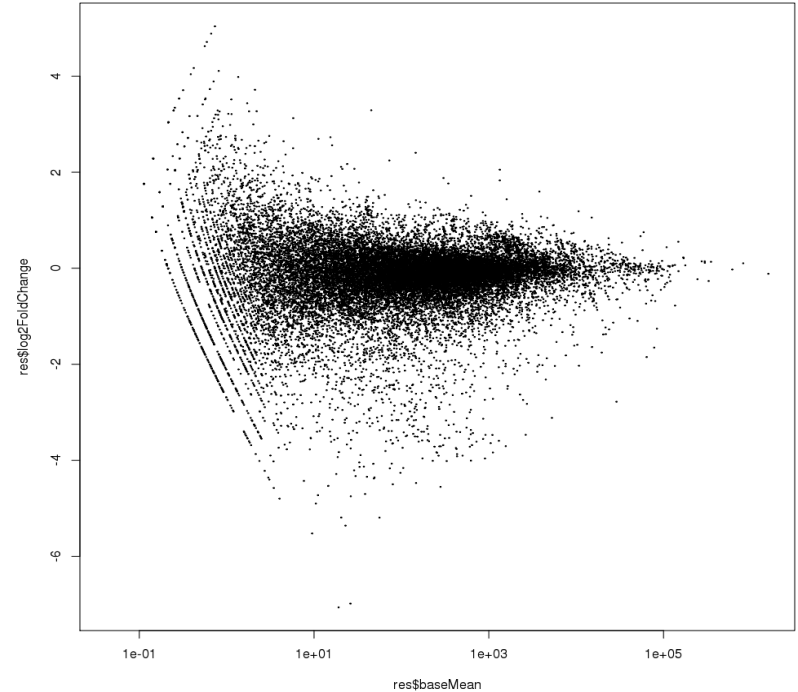
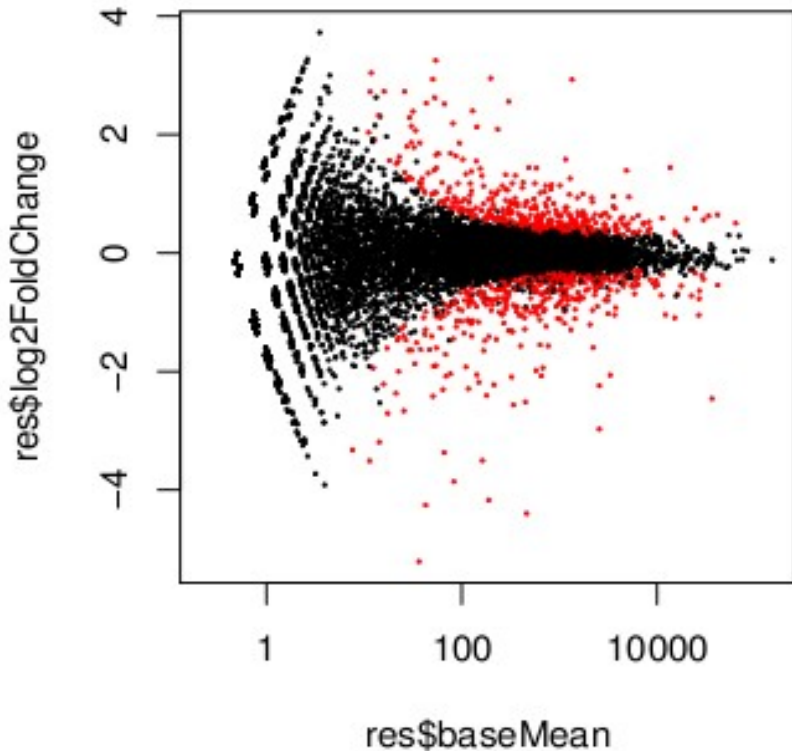
Dispersion Estimates

rowMeans vs. Gene Disp. Est.

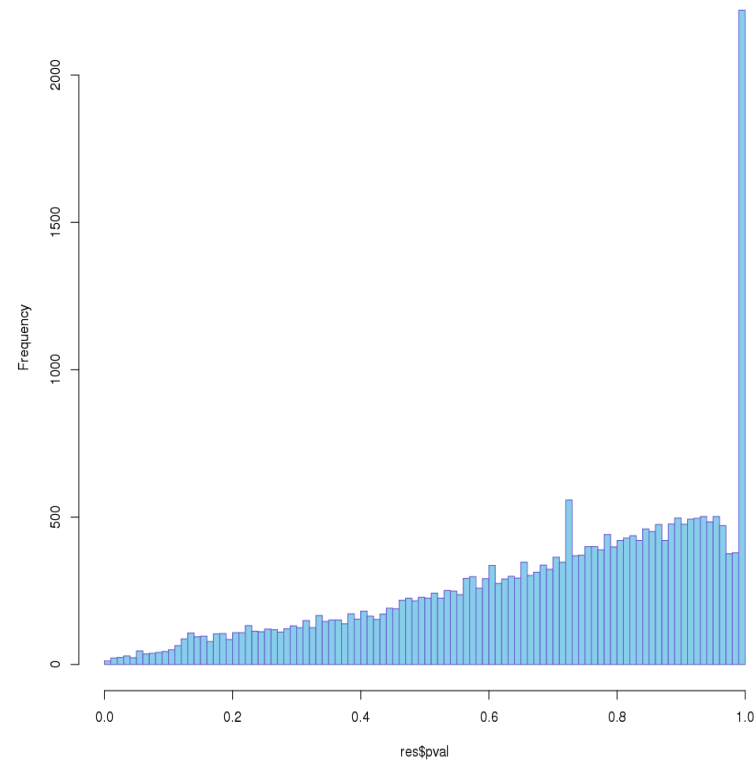
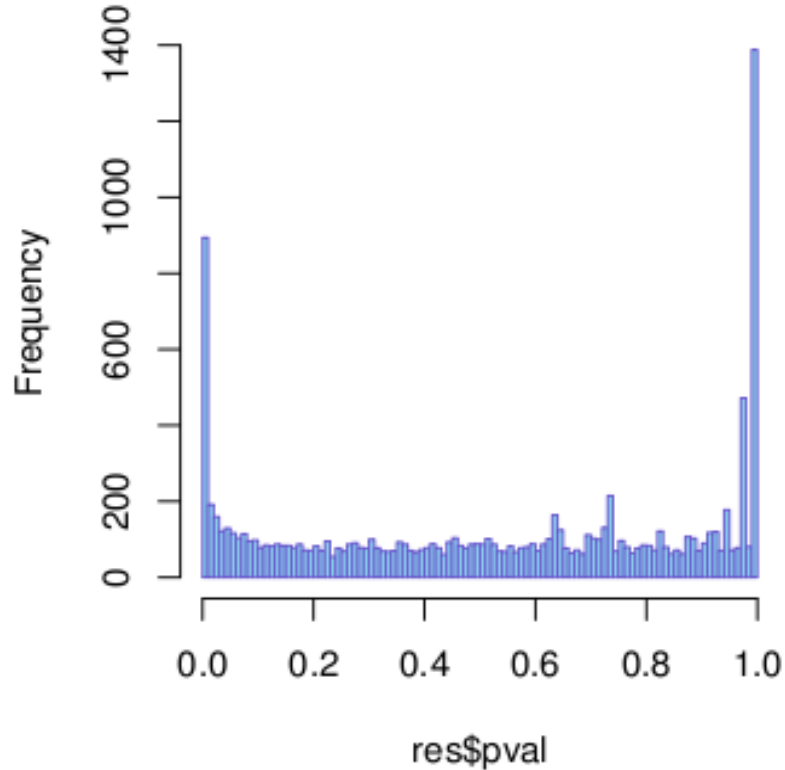


MA-Plot

baseMean vs. log2foldChange



Histogram of p-values



Acknowledgements

The Bioinformatics Core

Dr. Dawei Lin (Director)

Data Analysis

Dr. Joe Fass (Lead)
Dr. Monica Britton
Mr. Nikhil Joshi

Statistical Programming

Mr. Vince Buffalo (Lead)

Application Development (Web/DB)

Mr. Jose Boveda (Lead)

System Admin & HPC

Dr. Zhiwei Lu (Lead)

Visiting members

Ms. Xinran Dong

Campus Scientific Advisory Board Chair

Dr. Craig Benham, Ph.D. (Mathematics)

Members

Dr. Gino Cortopassi, Ph.D. (Molecular Science)
Dr. Vladimir Filkov, Ph.D. (Computer Science)
Dr. Fredric Gorin, Ph.D. (Neuroscience)
Dr. Juan Medrano, Ph.D. (Animal Science)
Dr. Jie Peng, Ph.D. (Statistics)
Dr. David Rocke, Ph.D. (Biostatistics)

Genome Center Director

Dr. Richard Michelmore, Ph.D.

Associate Directors for Bioinformatics

Dr. Ian Korf, Ph.D.
Dr. Patrice Koehl, Ph.D.

