

Analyze RNASeq Data from G8P2 RFI (Residual Feed Intake) Lines Using QuasiSeq Package

July 11, 2014

Table of Contents

1. Data Summary
2. Model Selection Strategies
 - Number of genes with pvalues less than 0.05
 - Distance between Grenander CDF Estimator to Uniform CDF using
 - Anderson-Darling Statistics
 - Cramers-Von-Miser statistics
 - Kolmogorow-Smirnov statistics
3. Results

RNASeq Data Summary

- RNASeq data set is a 25322×24 table of count data corresponding to 25322 genes of 24 pigs from 2 Lines: high RFI Line and low RFI Line, and 2 Diets: high energy diet (Diet 1) and low energy diet (Diet 2).
- There are 6 pigs for each combination of Diet and Line.

Metadata Summary

- The available metadata consists of information of 9 covariates for 24 samples of 24 pigs.
 - Factors: Diet (2 levels), Line (2 levels), Lane (2 levels), dateRNA (3 levels), and dateGD (5 levels).
 - Quantitative covariates: RFI (RFI values) , RINb (RNA Integrity Number before globin depletion), RINa (RNA Integrity Number after globin depletion), Conc (RNA Concentration after globin depletion).

Number of Genes Used in Analysis

- Models with metadata covariates: The number of genes analyzed is 14781. Those are genes with average counts greater than 1 and for which there are at least four samples with non-zero counts.

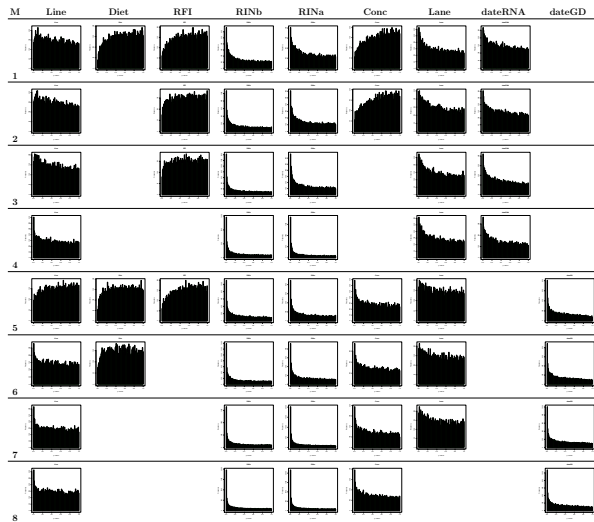
Model Selection Criteria

Starting model includes all interested covariates. For each covariate,

- We conduct a Likelihood Ratio Test using QuasiSeq of the full model vs. the reduced model obtained from the full model by deleting the considered covariate. We collect the set of pvalues of all genes from the tests.
- Obtain the number of genes with pvalues less than or equal 0.05.
- Obtain Grenander CDF estimator of the empirical CDF of the sample from those pvalues.
- Obtain the Anderson-Darling statistics, Cramér-Von-Mises statistics, and Kolmogorov-Smirnov statistics between the Grenander CDF and uniform CDF.

Exclude the covariate corresponding to the smallest value for most of the above criteria.

Backward Model Selection



Results of Model 4 and Model 8

When FDR is controlled at 0.05, 0.10, 0.15, the number of DE Genes between two RFI Lines (DEGs) and the number of DE Genes between two RFI Lines with $\log(\text{fold change})$ at least 1 ($\log(\text{FC}) \geq 1$) are shown in the tables below

FDR	DEGs	$\log(\text{FC}) \geq 1$
0.05	0	0
0.10	10	2
0.15	23	5
0.20	60	8

Model 4: Estimated number of DE Genes between 2 Lines is 1965.

FDR	DEGs	$\log(\text{FC}) \geq 1$
0.05	3	1
0.10	6	1
0.15	28	3
0.20	45	4

Model 8: Estimated number of DE Genes between 2 Lines is 1172.

Model 4 and Model 8 Pvalues of Line Effects

