Analyze RNASeq Data from G9P2 RFI (Residual Feed Intake) Lines Using QuasiSeq Package

July 28, 2014

Table of Contents

RNASeq Data Summary

- RNASeq data set is a 25320 $\times$ 31 table of count data corresponding to 25320 genes of 31 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).

- For Diet 1, the RNA data are from 7 low RFI line pigs and 8 high RFI line pigs. For Diet 2, the RNA data are from 8 low RFI line pigs and 8 high RFI line pigs.

Metadata Summary

- The available metadata consists of infomation of 9 covariates for 31 samples of 31 pigs.
  - Factors: Diet (2 levels), Line (2 levels), Block (4 levels), Blockorder (8 levels).
  - Quantitative covariates: RFI (RFI values) , RINb (RNA Integrity Number before globin depletion), RINa (RNA Integrity Number after globin depletion), Concb (RNA Concentration before globin depletion), Conca(RNA Concentration after globin depletion).

- CBC (Complete Blood Count) data: neutrophils, lymphocytes, monocytes, eosinophils, and basophils. The CBC covariates are in model in the form of log2 transformation.

Number of Genes Used in Analysis

- Models with metadata covariates and with CBC covariates: The number of genes analyzed is 12222. Those are genes with average counts greater than 8 and for which there are at least four samples with non-zero counts.

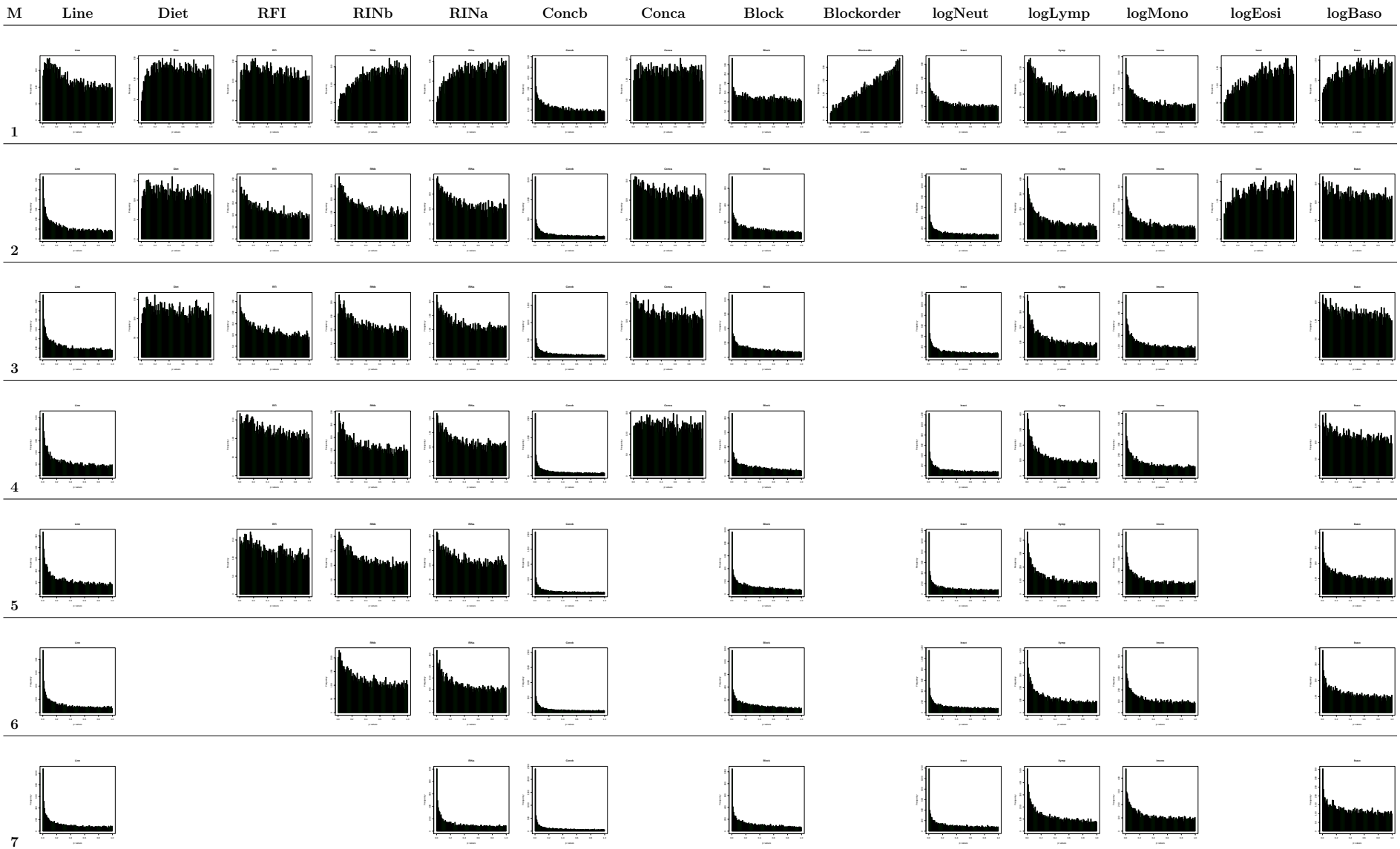
Model Selection Criteria

Starting model includes all covariates of interest. 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 7

- Estimated number of DE Genes between two RFI Lines

```
## [1] 4580
```

- 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 log2(fold change) at least 1 (log2(FC) ≥ 1) are shown in the table below

FDR	DEGs	log2(FC) ≥ 1
0.05	649	60
0.10	1680	89
0.15	2595	102
20.00	3610	107

Comparison to The Paired End Read Count Data

When FDR is controlled at 5%, then there are 444 common DE genes between two RFI lines.

```
## [1] 444
```

Comparison to the G8 Data