

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

August 26, 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
 - Cramér-Von-Mises statistics
 - Kolmogorov-Smirnov statistics

3. Results

RNASeq Data Summary

- RNASeq data set is a 25320×31 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 information 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

- Single End Read Count Data: 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.
- Paired End Read Count Data: The number of genes analyzed is 12280. Those are genes with average counts greater than 8 and for which there are at least four samples with non-zero counts.

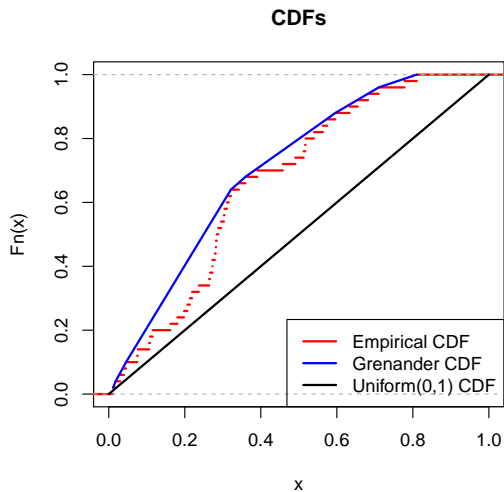
Model Selection Criteria

Starting model includes all covariates of interest. For each covariate,

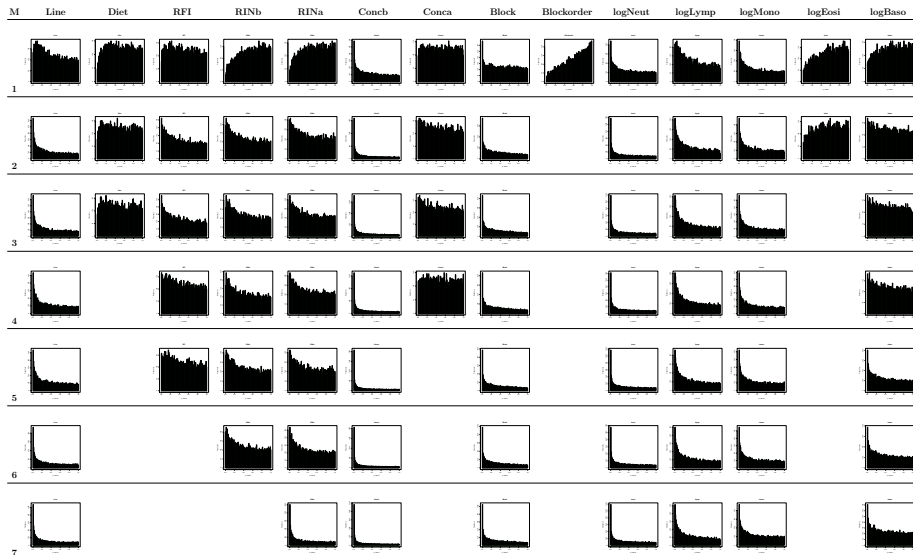
- We conduct a Likelihood Ratio Test using QuasiSeq of the full model vs. the reduced model obtained from the full model by deleting the considered covariate. We collect the set of pvalues of all genes from the tests.
- Obtain the number of genes with pvalues less than or equal 0.05.
- Obtain Grenander 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.

Grenander's Estimator



Single End Read Count Data-Backward Model Selection



Single End Read Count Data-Results of Model 7

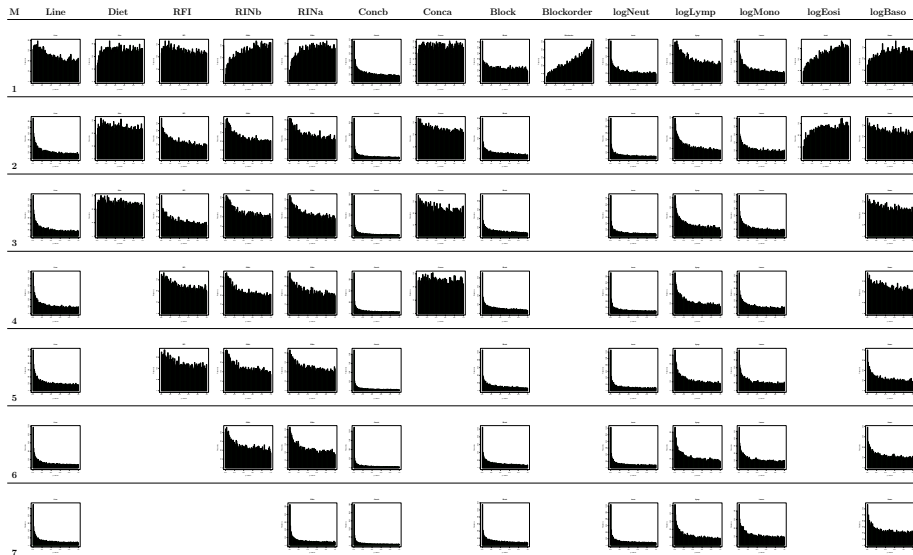
- Estimated number of DE Genes between two RFI Lines

```
## [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 $\log_2(\text{fold change})$ at least 1 ($\log_2(\text{FC}) \geq 1$) are shown in the table below

FDR	DEGs	$\log_2(\text{FC}) \geq 1$
0.05	649	60
0.10	1680	89
0.15	2595	102
20.00	3610	107

Paired End Read Count Data-Backward Model Selection



Paired End Read Count Data-Result of Model 7

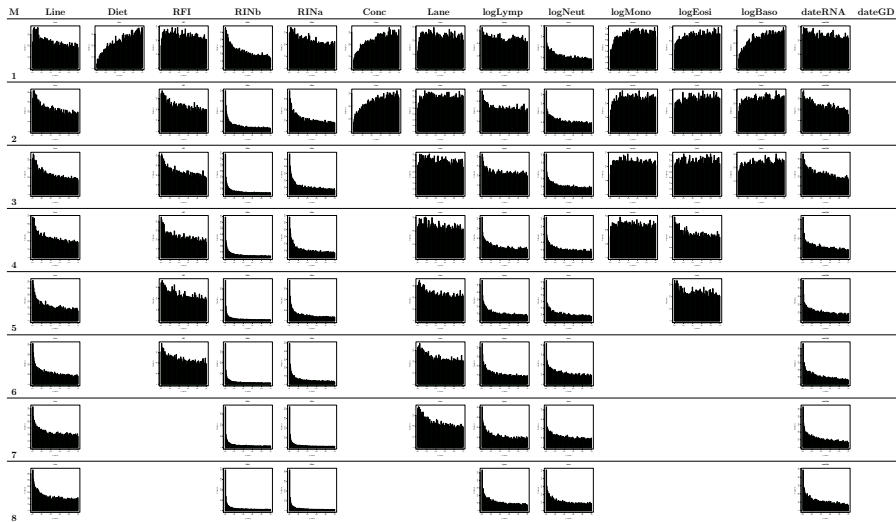
FDR	DEGs	$\log_2(\text{FC}) \geq 1$
0.05	771	61
0.10	1837	92
0.15	2936	101
20.00	3945	109

When FDR is controlled at 5%, then there are 444 common DE genes between two RFI lines in which 1 common DE genes with $\log_2\text{fc}$ at least 1.

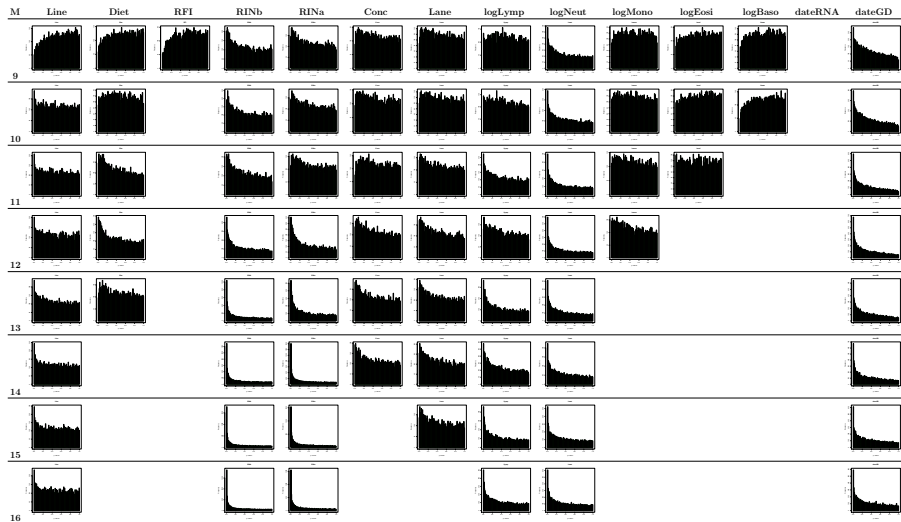
```
## [1] 444
```

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

G8 - Backward Model Selection I



G8 - Backward Model Selection II



Comparison G8 Data and Single End Count Data G9

When FDR is controlled at 20% for DE Analysis of G8 data, and FDR is controlled at 5% for DE Analysis of single end count G9 data, there are 23 common DE genes between RFI lines

```
## [1] "ENSSSCG00000000810" "ENSSSCG000000009051"
## [3] "ENSSSCG000000027314" "ENSSSCG000000006457"
## [5] "ENSSSCG000000002792" "ENSSSCG000000007517"
## [7] "ENSSSCG000000024791" "ENSSSCG0000000021802"
## [9] "ENSSSCG000000000625" "ENSSSCG000000008614"
## [11] "ENSSSCG000000012141" "ENSSSCG000000016426"
## [13] "ENSSSCG000000013293" "ENSSSCG000000007656"
## [15] "ENSSSCG000000021646" "ENSSSCG000000008193"
## [17] "ENSSSCG000000001511" "ENSSSCG000000015766"
## [19] "ENSSSCG000000009829" "ENSSSCG000000006309"
## [21] "ENSSSCG000000006378" "ENSSSCG000000006912"
## [23] "ENSSSCG000000003716"
```

and among those there are 3 genes with $\log_2\text{fc} \geq 1$.

```
## [1] "ENSSSCG000000000625" "ENSSSCG000000015766"
## [3] "ENSSSCG000000006912"
```