

Differentially Expressed Genes in Blood from Young Pigs between Two Swine Lines Divergently Selected for Feed Efficiency: Potential Biomarkers for Improving Feed Efficiency

A.S. Leaflet R3014

Haibo Liu^{1,2}, Graduate Assistant;
Yet Nguyen³, Graduate Assistant;
Dan Nettleton^{1,2}, Professor;
Jack Dekkers^{1,2}, Professor;
Chris Tuggle^{1,2}, Professor;

^{1,2}Bioinformatics and Computational Biology graduate program, Department of Animal Science; ³Department of Statistics; Iowa State University, Ames, Iowa 50011, USA

Summary and Implications

The goal of this study was to find potential gene expression biomarkers in blood of piglets that can be used to predict pigs' future feed efficiency. Using RNA-seq technology, we found 453 genes were differentially expressed (false discovery rate (FDR) ≤ 0.05) in the blood of two Yorkshire lines of pigs divergently selected for feed efficiency (FE) based on residual feed intake (RFI). Genes involved in several biosynthetic processes were overrepresented among genes more highly expressed in the low RFI line compared to the high RFI line. Weighted gene co-expression network analysis (WGCNA) also revealed genes involved in some of these biosynthesis processes and having similar patterns of expression formed clusters. The average expression in the clusters was highly associated with lines ($p < 3.9E-07$, $R^2 > 0.59$). Current findings implied these biosynthesis pathways might be more active in the high RFI line. After further stringent validation, some of the differentially expressed genes (DEGs) will be selected for validation as biomarkers for feed efficiency.

Introduction

Improving feed efficiency (FE) of domestic animals is of economic, social and environmental significance. However, measuring individual animal's FE is expensive and time-consuming. Thus, predictive biomarkers for FE that can be measured in the early life of the pig are in great need in animal breeding and production. The purpose of this study is to use whole blood from 5-6 week-old pigs to identify differentially expressed genes between two Yorkshire lines of pigs divergently selected for FE based on RFI, with an ultimate goal of developing useful biomarkers for FE.

Material and Methods

Blood samples were collected into Tempus tubes for long-term storage and EDTA tubes for complete blood count (CBC) test from 233 piglets (Generation 9, at 35 ~ 42

days of age) of the two Yorkshire lines of pigs divergently selected for FE based on RFI, the low RFI line and the high RFI line. We subsequently measured the complete blood count (CBC) of all blood samples. These animals were later randomly assigned to two diets of different fiber and energy content, and each pig's RFI was measured on FIRE feeders. Blood samples of 8 animals of extremely low RFI for the low RFI line and extremely high RFI for the high RFI line on each diet were selected for RNA-sequencing, with one sample later dropped due to low RNA quality. RNA samples were prepared and globin transcripts were depleted by an RNase H-mediated method before RNA sequencing. One hundred base pair paired-end sequencing was run on the Illumina HiSeq 2000 platform. Reads were mapped to the pig reference genome (Sscrofa10.2.76) using aligner STAR after quality trimming and featureCounts was used to get read counts for each gene. Differentially expressed gene analysis was performed using QuasiSeq. Weighted gene coexpression network analysis was used to find genes with similar expression patterns, thus potentially with similar functions.

Results and Discussion

Using an analysis package (QuasiSeq) and a model selection strategy based on backward feature selection, we established a model for read counts, including line, CBC covariates (neutrophil, lymphocyte, monocyte and basophil) and technical variables (processing batches, RNA concentration before globin depletion, RNA integrity number (RIN) after globin depletion). With this model, we found 453 DEGs (FDR ≤ 0.05) between the two lines, with 50 DEGs having fold change ≥ 2 . Functional terms including small molecule biosynthetic process, lipid biosynthetic process, steroid biosynthetic process and positive regulation of immune system process, were overrepresented among genes expressed higher in the low RFI line, while signaling, RNA biosynthesis process and cellular response to stimulus were overrepresented among genes higher in the high RFI line (FDR ≤ 0.15). Weighted gene co-expression network analysis showed that the average expression of the genes in four clusters was highly associated with line ($p < 3.9E-07$, $R^2 > 0.59$) and biology process terms (gene ontology) of lipid metabolic process, lipid biosynthetic process, steroid biosynthetic process and response to stimulus were overrepresented among the top two clusters. And the differentially expressed genes were significantly enriched in each of the four top clusters. Both the differentially expressed gene analysis and WGCNA suggested specific biosynthesis pathways might be different

in level of activity between the two lines. We will use qRT-PCR (a lower cost analytical method) to validate the DEGs that showed at least a two-fold change in both the original samples and new representative samples of the two lines. This information will be used to identify a set of DEGs to build a biomarker-based classifier for FE in the future.

Acknowledgements

This study is supported by USDA-NIFA-AFRI- 2011-68004-30336. We thank Nick Boddicker, Ziyanda Mpetile, and the Lauren Christian Memorial Swine Breeding farm staff for their assistance in collecting the blood samples and managing the pig lines for this project.

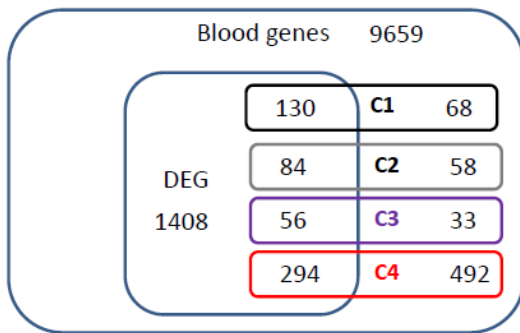


Figure 1. Venn diagram showing differentially expressed genes (false discovery rate ≤ 0.15) were significantly overrepresented among genes of each of the 4 top clusters that were highly associated with line from WGCNA. The clusters were labeled as C1 to C4, in the order of association strength with line. DEG, differentially expressed genes. There were 12282 genes expressed in the blood, 1972 genes differentially expressed between the two lines ($FDR \leq 0.15$). 130 out of 198 genes in C1, 84 out of 142 genes in C2, 56 out of 89 genes in C3, and 294 out of 786 genes in C4 were among the differentially expressed genes. The p values of overrepresentation test for clusters C1 to C4 were respectively 2.55E-56, 9.31E-32, 1.81E-23, and 2.08E-51 (Fisher’s exact test). Regression analysis showed that line is highly associated with the average gene expression of the 4 top clusters with regression coefficients, p-values and R square as follows (regression coefficient, p value, R^2): C1(0.33, 4.73E-13, 0.84); C2(-0.3, 2.93E-09, 0.71); C3(-0.29, 2.72E-08, 0.59); C4(0.28, 3.94E-07, 0.66).