Gene set enrichment analysis for feed efficiency in beef cattle

H.L. Neibergs¹, J.L. Mutch¹, M. Neupane¹, C.M. Seabury², J.F. Taylor³, D.J. Garrick⁴, M.S. Kerley³, D.W. Shike⁵, J.E. Beever⁵, US Feed Efficiency Consortium³, K.A. Johnson¹

¹Washington State University, Pullman
²Texas A&M University, College Station
³University of Missouri, Columbia
⁴Iowa State University, Ames
⁵University of Illinois, Urbana

Abstract

Selection for improved feed efficiency in cattle would decrease the amount of feed consumed for the same or greater levels of production resulting in enhanced profitability and sustainability. Selection for feed efficient cattle has been hampered by a lack of phenotypic data on feed intake and weights from cattle in production due to the cost and difficulty in collecting these data. The aim of this study was to better understand the molecular functions and biological processes associated with residual feed intake (RFI) by identifying gene sets (biological pathways) associated with RFI in Hereford cattle and to use this information to facilitate genomic selection for RFI. Feed intake and body weight were measured on 847 Herefords at Olsen Ranches in Harrisburg, NE. Animals were genotyped with the Illumina BovineSNP50 or BovineHD Bead Chips. Genome-wide association analysis was conducted with GRAMMAR mixed model software as part of the gene set enrichment analysis (GSEA). Single nucleotide polymorphisms were mapped to 19,723 genes based on the UMD 3.1 reference genome assembly based on coordinates within 8.5 kb either side of each gene. Gene sets (4,389) were compiled from Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome, Biocarta and Panther for the GSEA. Gene sets associated (Normalized Enrichment Score > 3.0) with RFI after 10,000 permutations in GenABEL were centrosome (5) (GO:0005813), cytoskeleton organization (4) and peroxisome (KEGG:04146). Most (191) of the identified genes enriched within the gene set for RFI (leading edge genes) were unique to a single gene set although 15 leading edge genes were shared between centrosome (5) and cytoskeleton organization (4) and one gene was shared between cytoskeleton organization and peroxisome gene sets. Further GSEA are being conducted with Angus, Angus x Simmental and crossbred beef cattle to validate these results.

Introduction

Great strides have been made in improving cattle traits that are easily measured and associated with performance or output measures such as weaning and yearling weights. These traits also determine the value of beef products that are sold. Less improvement has been made in traits that are more difficult and costly to measure that are associated with inputs or cost of production. The expense and difficulty in measuring these traits has hampered improvement even though the majority of the costs associated with raising cattle reside in feed (Anderson et al. 2005). Residual feed intake (RFI) is often used as a measure of feed efficiency in beef cattle because it is phenotypically independent of growth and body weight (Koch et al., 1963). Residual feed intake is defined as the difference between the amount of feed actually consumed and the expected feed requirement based on body size and production level of the animal over the period feed consumption was measured. Cattle with low RFI are efficient animals that
consume less feed than expected for their level of performance. An advantage of using RFI as a measure of feed efficiency is that it is not correlated with the traits that are included in its calculation, unlike some other measures such as feed conversion ratio, that may have unintended selection consequences such as large mature cow size (Archer et al., 1999). The failure to measure RFI and use it as an evaluation tool to improve feed efficiency in cattle has largely been due to the cost and difficulty in measuring feed intake.

In the past decade, genomic resources for identifying regions of the genome associated with economically important traits have become available that have accelerated genetic improvement in cattle. As the genetics community has enjoyed these new genomic tools, the impediment to greater genetic improvement has become the collection of cattle phenotypes. The ability to economically collect a full set of phenotypes on cattle lags behind our ability to analyze genomes. This has been particularly true in collecting phenotypes on traits that occur late in life, are expensive or require specialized equipment. Traits such as these are therefore typically not used for routine calculation of estimated progeny differences (EPD), or if used are associated with low EPD accuracies due to the limited number of phenotypes collected on the breeding stock as well as their progeny.

The genotype-phenotype relationship for feed efficiency in cattle is complicated by stage of life, diet, breed, weather and a host of other factors. The use of phenotype-focused approaches to study the genetic basis of feed efficiency may be very helpful in identifying and validating loci that explain a large fraction of genetic variance. The National Program for Genetic Improvement of Feed Efficiency in Beef Cattle has collected RFI data on over 8,000 cattle in several cattle breeds (Angus, Charolais, Gelbvieh, Hereford, Limousin, Red Angus, Wagyu and crossbreds of Charolais x Angus, Piedmontese x Angus x Simmental, Simmental x Angus) to identify quantitative trait loci associated with RFI. To compensate for the lack of RFI phenotypic data on the majority of cattle in the US, the use of genetic markers predictive of RFI could be used to select animals for improved feed efficiency.

Just as breeds of cattle differ in their abilities to use dietary energy, so do individuals within breeds differ in their ability to convert forages or concentrates into usable energy to grow, maintain their body condition, successfully maintain a pregnancy or raise a calf. Some of the differences in feed intake that are not explained by differences in weight or growth rate are the result of genetic variation (Carstens and Tedeschi, 2006; Herd and Bishop 2000, Basarab et al., 2003). The reported heritability for RFI is moderate (18-49%) suggesting that there is an opportunity to make significant genetic gains in feed efficiency through selection (Bolormaa et al., 2011; Saatchi et al., 2014). There is evidence that measurement of RFI in cattle post-weaning, in mature cows, or across different feeding regimes is repeatable (Herd et al., 2003; Durunna et al., 2011). The consistency of RFI over time and across feedstuffs suggests that it may be a good measure to include in selection indexes for multiple-trait selection of cattle that are feed efficient and productive. It would also be an excellent candidate trait for marker-assisted or genomic selection.

To identify markers for marker-assisted or genomic selection, genome-wide association studies have been conducted to identify genomic regions with major effects on RFI (Barendse et al., 2007; Sherman et al., 2008; Bolormaa et al., 2011; Mujibi et al., 2011; Rolf et al., 2012; Serão et al., 2013; Abo-Ismail et al., 2014; Saatchi et al., 2014). A complementary approach to identify markers for selection is to identify genes that are differentially expressed between cattle
with high and low RFI (Chen et al., 2011; Al-Husseini et al., 2015; Paradis et al., 2015; Tizioto et al., 2015; Xi et al. 2015, Weber et al., 2016). The identification of genes that are differentially expressed would potentially identify genes with large or modest effects on variation in RFI. Genetic markers or single nucleotide polymorphisms (SNPs) within or near these genes could be used for selection.

The use of gene expression data can also be integrated into a third approach; the use of pathway, network or gene set enrichment analysis (GSEA). Pathway, network or GSEA aim to provide insights into genes that individually may have more modest individual effects but may be interacting with one another to cumulatively elicit a large effect on phenotype. The study of pathways or gene networks allows us to better understand the molecular functions and biological processes that are associated with a trait such as RFI. As GSEA has matured, it has used associations with SNPs within or near genes to substitute for gene expression data to identify gene pathways important to a trait. More recent enhancements have combined both SNP and differentially expressed gene data together to identify important gene pathways. Gene set enrichment analysis has previously been conducted with RFI for five breed groups consisting of Angus, ¾ Angus, crossbred Angus and Simmental and purebred Simmental (Serão et al., 2013). Nine clusters of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were associated (p < 0.001, enrichment score > 3.0) with feed efficiency. Those pathways included GO molecular functions of nucleotide binding, protein kinase activity and metallopeptidase activity, and GO biological process of ion transport, phosphorus metabolic process, membrane invagination and proteolysis. The KEGG gene set associated with RFI was the MAPK signaling pathway (Serão et al., 2013). The glycogen synthase kinase 3 beta (GSK3B) gene was the leading edge gene in the KEGG gene set and is involved as a regulator of nutrient storage in adipose tissue and skeletal muscle (Hoehn et al., 2004). Abo-Ismail and colleagues (2014) identified 35 biological process gene sets that were associated with RFI. Of these gene sets, ion transport and cation transport contained the largest number of leading edge genes although proteolysis, protein complex biogenesis and protein amino acid glycosylation were also identified. Ion transport, MAPK signaling, and proteolysis were gene sets that had also been identified by Serão et al. (2013) as associated with RFI in beef cattle. Pathway or network analysis can also be performed using other approaches as has been described by Rolf et al. (2012), Saatchi et al. (2014), and Weber et al. (2016). Several pathways (adherens junction, adipocytokine signaling, apoptosis, long-term depression, calcium signaling, melanogenesis, pancreatic cancer, pathways in cancer and MAPK signaling) were identified in more than one pathway and/or GSEA study as being important in RFI (Rolf et al., 2012; Abo-Ismail et al., 2014; Xi et al., 2015). Although, the specific pathways differed, Weber et al. (2016) identified pathways associated with RFI that are involved in the activation of the immune response and in the down regulation of fat deposition in adipose and muscle tissue in eight steers produced from one high and one low RFI Angus bull.

The aim of this study was to identify genes and gene sets that were associated with RFI to better understand the molecular functions and biological processes that are associated with RFI and to exploit this information through genomic selection to improve the efficiency of beef cattle production.
Materials and Methods

Eight hundred seventy Hereford cattle were sampled while on feed at Olsen Ranches, Inc. in Harrisburg, Nebraska where they were fed a concentrate ration as previously described (Saatchi et al., 2014). Phenotypes and DNA samples were collected over a three year period (2009-2011). Date of birth (DOB), date of weaning, sex (S), breed composition, days on feed (DOF), feed intake (DMI) and weights were collected. All cattle were fed a minimum of 70 days. Feed intake and body weight gain were measured with a GrowSafe (Airdrie, Alberta Canada) system. There were nine male contemporary groups that consisted of a total of 824 steers and one female contemporary group that consisted of 23 heifers. Four hundred eighty-nine cattle were genotyped with the BovineHD and 358 were genotyped with the BovineSNP50 assays. BovineSNP50 genotypes were imputed with Beagle 4.1 to the density of the Illumina BovineHD BeadChip using the 489 BovineHD genotyped Hereford cattle as a reference. Animals were removed from the analysis if the genotype call rate was less than 90%, if they were predicted to be Klinefelter (XXY) individuals or if phenotypic information (average daily gain [ADG], mid test metabolic weight [MMWT] or DMI) used to calculate RFI was missing. Heterozygosity of > 0.2 for non-pseudoautosomal region markers on the X chromosome was evaluated to confirm anatomical gender. A total of 847 Hereford cattle remained for the GSEA-SNP. Single nucleotide polymorphisms were removed if less than 90% of the genotypes were successfully called, if the minor allele frequency was less than 1% or if they failed the Hardy-Weinberg equilibrium (p < 1.0x10^{-100}) test.

Genome-wide association analysis (GWAA) was performed using the GRAMMAR mixed model software in GenABEL (http://www.genabel.org/; Aulchenko et al., 2007). Residual feed intake was calculated by subtracting expected DMI (dependent variable) from the actual DMI. Expected DMI was calculated by incorporating covariates for ADG, MMWT, contemporary group (CG), S, DOB and DOF to estimate RFI (Animal + e) as shown in equation 1.

Equation 1. Calculation of residual feed intake

\[ DMI = \beta_0 + \beta_1(ADG) + \beta_2(MMWT) + \beta_3(CG) + \beta_4(S) + \beta_5(DOB) + \beta_6(DOF) + \text{Animal} + e \]

The most significantly associated SNPs for each of the 19,723 annotated genes in the UMD3.1 reference assembly (http://bovinegenome.org/?q=node/61) were selected as a proxy for each gene and used for the GSEA-SNP. Gene proxies were only considered for SNPs that were located within 8.5 kb of a gene as this was representative of the average haplotype block size in Herefords as determined by a haplotype block analysis performed in the SNP Variation Suite 8.1 software (Golden Helix, Bozeman, MT) and finding the average distance between the beginning and ending nucleotide positions for each haplotype block. The GSEA-SNP analysis was conducted using a composite of 4,389 gene sets taken from Gene Ontology (GO; http://geneontology.org/) (n= 3,147), Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.jp/kegg/) (n=186), Reactome (http://www.reactome.org/) (n=674), Biocarta (http://www.genecarta.com/) (n=217), and Panther (http://www.pantherdb.org/) (n=165). Significance was calculated using the null distribution estimated from 10,000 permutations for each gene set using GenABEL in R (Holden et al., 2008). An enrichment score was calculated for each pathway using a modified Kolmogorov-Smirnov statistic, and these were normalized.
(NES) based on the size of each gene set. Gene sets with NES > 3.0 were identified as associated with RFI.

Results and Discussion

Two GO gene sets, centrosome (5) (GO:0005813) and cytoskeleton organization (4) (GO:0007010) and one KEGG gene set, peroxisome (KEGG:04146) were associated with RFI (Table 1). The centrosome gene set is a member of the cellular component ontology tree. Centrosomes are critical in mitosis and meiosis as the spindle apparatus of the cell is organized by the centrosomes. Disruption of the centrosome affects the proper segregation of chromosomes in the dividing cell and the stability of the genome (Lerit and Poulton, 2016). Centrosomes are cell organization centers around which hundreds of proteins are found that regulate the cell cycle (Conduit et al., 2015). The centrosome gene set consists of 99 genes, and 37 of those genes are leading edge genes or genes that are enriched among those associated with RFI (Table 1). Fifteen of the leading edge genes in the centrosome gene set were also leading edge genes in the cytoskeleton organization gene set (Table 1). In addition to these genes being involved in centrosome and cytoskeleton organization in the cell, many of these genes play a role in tumor growth in a host of cancers. Of the leading edge genes, CEP120 has been associated with abdominal obesity as measured by waist circumference in humans, although no direct associations with feed efficiency have been reported (Wen et al., 2016).

The cytoskeleton organization gene set is part of the biological process ontology tree. The cytoskeleton is responsible for maintaining the shape of the cell and is involved in cellular movement, cell division, and endocytosis. The cytoskeletal organization gene set has previously been associated with feed efficiency traits in poultry. Cytoskeletal genes have been downregulated in breast muscle in high feed efficiency broiler chickens (Kong et al., 2011). Others have identified the role of regulation of actin cytoskeleton as an important component of feed efficiency and compensatory gain in cattle (Rolf et al., 2012; Keogh et al., 2016). The regulation of actin cytoskeleton organization shares some similarity (3.84%) to that of the cytoskeleton organization gene set in that it is involved in the processes that disassemble cytoskeleton structures or their proteins. The cytoskeleton organization gene set consists of 246 genes of which 97 were associated with RFI.
Table 1. Gene Sets and Leading Edge Genes Associated with Residual Feed Intake.

<table>
<thead>
<tr>
<th>Gene Set</th>
<th>NES¹ (nominal p value)</th>
<th>Gene set size (# leading edge genes)</th>
<th>Leading Edge Genes shared between gene sets²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centrosome (5) (GO:0005813)</td>
<td>3.19 (p=0.0010)</td>
<td>99 (37)</td>
<td>Centrosome &amp; cytoskeleton organization gene sets: SLC9A3R1, CEP120, BIRC5, HEPACAM2, TUBG1, MAP10, USP33, BBS4, TBCCD1, CSNK1D, CEP57, LZTS2, CTNNB1, CYLD, SLAIN2</td>
</tr>
<tr>
<td>Cytoskeleton organization (4) (GO:0007010)</td>
<td>3.07 (p=0.0011)</td>
<td>246 (97)</td>
<td>Cytoskeleton organization &amp; peroxisome gene sets: SOD1</td>
</tr>
<tr>
<td>Peroxisome (KEGG:04146)</td>
<td>3.05 (p=0.0016)</td>
<td>73 (30)</td>
<td></td>
</tr>
</tbody>
</table>

¹Normalized enrichment score; ²Leading edge genes are those that are enriched (significant) within the gene set.

Thirty leading-edge genes were associated with RFI in the peroxisome gene set and one gene (SOD1) was shared between the cytoskeleton organization and peroxisome gene sets. The peroxisome gene set is part of the cellular process ontology tree and is involved in cellular transport and catabolism. Peroxisomes are small organelles whose functions are essential in free radical detoxification, lipid homeostasis and hydrogen peroxide metabolism. The peroxisome transports medium chain fatty acids to the mitochondria where most of the β-oxidation occurs. The efficiency and ability of the peroxisome to function in lipid metabolism and to neutralize free radicals is essential in maintaining cellular membrane integrity and animal health. Mitochondrial biogenesis may play a role in shifts of muscle fiber types which may also impact feed efficiency (He et al., 2016). Although peroxisome proliferator activated protein gamma (PPARG) was not a leading edge gene in this study, it has been identified as an important regulator of food intake and energy homeostasis in rodents (Larsen et al., 2003; Festuccia et al., 2008).

Conclusion

Costs incurred with feeding constitute the major portion of the expense of raising cattle. Cattle that are able to develop, grow and maintain their body weight with less feed are more feed efficient. Identifying cattle that are more resource efficient through collection of daily feed intake and body weights is expensive and requires specialized equipment which has limited the collection of these data. The use of genomic markers to identify efficient cattle will provide a means of selecting breeding stock without the collection of expensive phenotypes on all animals. This study evaluated which biological pathways are involved in those cattle that were more feed efficient using RFI as the phenotype. Three pathways, centrosome (5), cytoskeleton organization (4) and peroxisome were associated with RFI, representing 191 unique genes. These genes are positional and functional candidates for use as genomic markers for RFI in beef cattle. Gene set enrichment analysis of Angus, Angus x Simmental and crossbred cattle are also being conducted to examine and compare in other beef cattle breeds.
Acknowledgement

This project was supported by National Research Initiative competitive Grant No. 2011-68004-30214 from the USDA National Institute of Food and Agriculture.
Literature Cited


