*‘But as it is written, "No eye has seen, no ear has heard, and no mind has imagined the things that God has prepared for those who love him."’*

**1 Corinthians 2:9**

**English Standard Version (©2001)**

**University of Alberta**

# CANDIDATE GENES, METABOLITES AND BIOLOGICAL PATHWAYS ASSOCIATED WITH RESIDUAL FEED INTAKE AND CARCASS QUALITY IN BEEF CATTLE

by

Brian Kahindi Karisa

A thesis submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Animal Science

*Department of Agricultural, Food and Nutritional Science*



Fall 2013

Edmonton, Alberta.

*Dedicated to my late mother; Betty (Mishi) Karisa, whose love and ambition lives in me even after she passed on, and my late father; Karisa Chembe, who gave me the courage and boldness to face all of this life's challenges*

*As they rest in peace, may this thesis be one of the calming echoes from these humble heroes*

## ABSTRACT

The main objective of this study was to identify genes, DNA variants and plasma metabolites associated with residual feed intake (RFI) in beef cattle. In the first study, a total of 117 SNPs were selected and genotyped in 531 steers at the University of Alberta. The population was split into a discovery and validation population and multiple marker association analyses were performed in the discovery, validation and combined populations using ASReml. Twenty two SNPs were associated ($P < 0.05$) with RFI in the discovery population and 7 (of the 22) were also significant ($P < 0.05$) in the validation population. Twenty five SNPs were associated with RFI ($P < 0.05$) in the pooled population. A gene network analysis indicated that the biological processes associated with the significant genes included lipid, glucose, protein and steroid metabolism, growth, energy utilization, and regulation of DNA transcription and translation. The second study was an association analysis using the 117 SNPs and indicated that 7 were associated with various carcass quality traits ($p \leq 0.005$) in the same population. A third association analysis was performed using steers at the University of Guelph, as the discovery population, to identify blood metabolites associated with RFI. Blood samples were collected at 3 periods with period 1, 2 and 3 corresponding to week 2, 6 and 9 into the feeding period respectively. Two, ten and three metabolites were significantly associated with RFI ($P < 0.05$) in period 1, 2 and 3 and accounted for 36%, 74% and 52% of the variation respectively. A validation analysis was performed using steers at the University of Alberta as the validation population. The results indicated that 3 metabolites were significantly associated

with RFI in both discovery and validation populations accounting for 32.8% of the variation in the validation population. A metabolic network analysis indicated that the biological pathways associated with the metabolites included AMPK signaling, growth hormone signaling, lipid and energy metabolism and cholesterol metabolism. The genes, metabolites, biological networks and the biological pathways help contribute to a better understanding of the physiological processes influencing RFI and carcass quality in beef cattle.

## ACKNOWLEDGEMENTS

There are people who made the Canadian winters bearable and the midnight data analysis possible. People who helped me to put my first step in this journey and kept showing me the light at the end of the tunnel. My most sincere gratitude goes to my supervisory committee, Dr. Graham Plastow, Dr. Stephen Moore, Dr Zhiquan Wang and Dr. Heather Bruce for the opportunity and financial support they gave me to undertake this PhD program. They created an environment which enabled me to explode in my creativity, develop research ideas and implement them into successful research objectives; an opportunity I will never take for granted.

I also wish to appreciate the great assistance and constant advice I received from Dr. Jennifer Thomson, thank you very much Jen. In addition, I appreciate my colleagues in the Livestock Gentec group at the University of Alberta; Denis Mujibi, Durunna Obioha, Ghader Manafiazer, Shadi Nayeri, Sungael Nalaila and Elilta Ephrem. I also thank my colleagues who offered their generous assistance in the laboratory work; Dr. Urmila Basu, Yan Meng, Jackson Mah, Fernando Soares, Samuel Yu and the staff at the University of Alberta ranch at Kinsella, especially Barry Irving.

I also express my gratitude to my brothers; Gavuna, Ngumbao, Safari, Charo, Kazungu, Sadik and Katana, and sisters; Dama, Jumwa, Kadzo and Furaha, and my friends who encouraged me and felt close to me even when they were several oceans away. I thank my wife, Lucy Karembo for staying true and calming my mind throughout the PhD duration; thank you so much.

Last and most important, I thank God who put me in the right place at every moment and orchestrated everything according to His will, to Him be the glory, forever and ever.

**TABLE OF CONTENTS**

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ADG | Average daily gain |
| AMP | Adenosine monophosphate |
| AMPK | AMP activated protein Kinase |
| ATP | Adenosine triphosphate |
| BF | Back fat |
| BTA | Bos taurus (chromosome) |
| BW | Body weight |
| CCAC | Canadian Council on Animal Care |
| CCK | Cholecystokinin |
| CMAR | Carcass marbling |
| CREA | Carcass rib eye area |
| DMI | Dry matter intake |
| DNA | Deoxyribonucleic acid |
| ERAD | Endoplasmic reticulum associated degradation (of proteins) |
| EST | Expressed sequence tag |
| FCR | Feed conversion ratio |
| GHRL | Ghrelin |
| GLP | Glucagon like peptide |
| GRDFT | Grade fat |
| GTP | Guanosine triphosphate |
| GWAS | Genomewide association study |

| | |
|---|---|
| Kbp | Kilo base pairs |
| LD | Linkage disequilibrium |
| LMY | Lean meat yield |
| MAS | Marker assisted selection |
| MJ | Mega Joules |
| mRNA | Messenger RNA |
| MW | Metabolic weight |
| NCBI | National center for biotechnology information |
| NGS | Next generation sequencing |
| NPY | Neuropeptide Y |
| PCR | Polymerase chain reaction |
| PYY | Peptide YY3 |
| QGRD | Quality grade |
| QTL | Quantitative trait locus/loci |
| RFI | Residual feed intake |
| RN | Rendement Napole |
| RNA | Ribonucleic acid |
| SI | Selection index |
| SIm | Molecular selection index |
| SIp | Phenotypic Selection index |
| SNP | Single nucleotide polymorphism |
| TCA | Tricarboxylic acid cycle (Krebs cycle) |
| TNF | Tumor necrosis factor |

| | |
|---|---|
| tRNA | Transfer RNA |
| UMAR | Ultrasound marbling |
| UREA | Ultrasound rib eye area |
| UTR | Untranscribed region |
| YGRD | Yield grade |

# CHAPTER ONE

# GENERAL INTRODUCTION

## 1.1 INTRODUCTION

Profitability in beef production is the difference between the level (quality and/or quantity) of products (outputs) and the costs of production (inputs) (MacNeil et al., 1997). In beef production, the major input is feed (Herd et al., 2003) which constitutes about 60 - 70% of the total costs of production (Herd et al., 2003; Arthur et al., 2004) and the major output is meat. Profits can be increased by one or a combination of the following: reducing feed utilization and increasing the quality or quantity of outputs. Therefore in relation to profitability, two of the most important traits assessed for selection in beef production are feed efficiency and carcass quality.

Most measures of feed efficiency are correlated with production traits except residual feed intake (RFI). RFI is defined as the difference between actual feed intake and predicted feed intake (Koch et al., 1963). In beef cattle, the predicted feed intake is estimated based on the individual's maintenance requirements (body weight-BW) and weight gain (average daily gain –ADG). Feed efficient cattle are those that consume less feed than the amount predicted based on their growth and maintenance requirements and thus have a negative RFI value, whereas inefficient cattle have a positive RFI value. RFI is phenotypically independent of the production traits used to estimate it. Therefore, selecting cattle for feed efficiency using RFI is expected to reduce feed intake without significantly affecting the growth rate and the mature body weight of the selected individuals. The

independence of RFI from production traits suggests that RFI represents inherent differences in basic metabolic processes that determine the efficiency of feed utilization (Herd and Arthur, 2009). This concept will be addressed in detail in Chapter 2.

On the other hand, traits related to carcass and meat quality, such as marbling, quality grade and carcass weight may have an impact on the monetary value of the meat and as a result will affect meat prices and profits. Improving feed efficiency and carcass quality traits relies on the ability to accurately select cattle, which are genetically superior for RFI, to be used as the breeding stock and thereby pass their superiority to the offspring. In the past, the traditional selection methods practiced by most breeders were based on the quantitative genetics approach described in detail by Dekkers and Hospital (2002). This approach requires data to be collected on the phenotypes of interest to estimate their heritability and genetic correlations (Dekkers and Hospital, 2002). This approach creates a challenge for RFI because estimating RFI requires expensive equipment to measure each animal's daily feed intake (Herd et al., 2003). Estimation of RFI also requires data on feed intake to be collected over a minimum duration of 63 days using the GrowSafe system (Wang et al., 2006) and some researchers have used 112 days (Mader et al., 2009) and up to 140 days (Montanholi et al., 2010). Where phenotypic data is not sufficient for selection, genetic markers can be used with or without phenotypic information. When used with phenotypic data, genetic markers can increase the accuracy of selection (Dekkers, 1999; Goddard and Hayes, 2002; Villanueva et al., 2002; Villanueva et al., 2005), and therefore result

in increased response to selection (Snelling et al., 2012). Genetic markers are also important for traits, such as carcass traits, which cannot be measured until the animal has been sacrificed making it unavailable for breeding. Genetic markers are effective in selecting for traits that have a moderate heritability (Togashi and Lin, 2010), which indicates that a moderate amount of the observed phenotypic variation is controlled by the additive effect of the genes associated with the trait. RFI has a moderate heritability ranging from 0.16 (Herd and Bishop, 2000) to 0.58 (Crews et al., 2003); therefore it is also a good candidate for marker-assisted selection. A more detailed review of marker-assisted selection for RFI is also discussed in Chapter 2.

Despite the published successes in identifying the genetic basis for economically important traits such as feed efficiency (Herd and Bishop, 2000; Moore et al., 2009; Barendse et al., 2007; Nkrumah et al., 2007; Sherman et al., 2008), utilization of marker-assisted selection tools in beef cattle is still limited, partly due to the lack of reproducibility of phenotype-genotype association studies across different populations/breeds (Sherman et al., 2008).

One of the options available to improve the efficiency of genetic markers for selection is to identify markers located in candidate genes associated with phenotypes. For RFI, this involves identifying SNPs located in genes whose functions are associated with the physiological mechanisms underlying the variation in RFI. These functional genes and SNPs are expected to be more robust across diverse populations and breeds of cattle.

In addition to SNP markers, identification of blood metabolites significantly associated with variation in RFI may be used as intermediate phenotypes that can be refined and developed into biomarkers for selection for RFI in beef cattle.

The regulation of RFI also depends on the combined effect at all the genes, metabolites and other factors that affect gene expression (Hartwell et al., 1999). The combined effects can be analyzed by reconstructing biological networks and an assessment of biological interactions and pathways involved.

## 1.2 OBJECTIVES

The major objective of this study was to identify candidate genes, DNA polymorphisms within the candidate genes, metabolites and the biological processes associated with variation in RFI and carcass quality in beef cattle.

The specific objectives included:

1. Identify ***positional*** and ***functional*** candidate genes for RFI and perform association analyses to identify ***Single Nucleotide Polymorphisms (SNPs)*** associated with RFI in beef cattle.
2. Identify levels of various ***blood metabolites*** and perform association analyses to identify the metabolites associated with RFI in beef cattle and propose a selection tool for RFI using metabolite biomarkers.
3. Use candidate genes and metabolites to reconstruct ***biological networks*** and analyze the networks to identify the biological processes

associated with RFI and assess the interactions between genes and biological processes.

4. Identify, from the genes that are significantly associated with RFI, those that have ***pleiotropic*** effects on carcass traits in beef cattle

## 1.3 REFERENCES


Arthur, P. F., J. A. Archer, R. M. Herd. 2004. Feed intake and efficiency in beef cattle: Overview of recent Australian research and challenges for future. Aust. J. Exp. Agr. 44: 361–369.

Barendse, W., A. Reverter, R. J. Bunch, B. E. Harrison, W. Barris, and M. B. Thomas. 2007. A validated whole-genome association study of efficient food utilization in cattle. J. Anim. Sci. 176:1893-1905.

Crews, D. H. Jr., N. H. Shannon, B. M. A. Genswein, R. E. Crews, C. M. Johnson, and B. A. Kendrick. 2003. Genetic parameters for net feed efficiency of beef cattle measured during post weaning growing versus finishing periods. Proc. West. Sec. Am. Soc. Anim. Sci. 54: 125-128.

Dekkers, J. C. M. 1999. Breeding in the 21st century: Application of molecular technology. Proc. Assoc. Adv. Anim. Breed. Genet. 13:1–16.

Dekkers, J. C. M., and F. Hospital. 2002. Utilization of molecular genetics in genetic improvement of plants and animals. Nature Reviews: Genetics 3: 22-32.

Goddard, M. E., and B. J. Hayes. 2002. Optimisation of response using molecular data. CD-ROM Communication No. 22-01 in Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France.

Hartwell, L. H., J. J. Hopfiled, S. Leiber, and A. W. Murray. 1999. From molecular to modular cell biology. Nature. 402 (2): C47 – C52.

Herd, R. M., and S. C. Bishop. 2000. Genetic variation in residual feed intake and its association with other production traits in British Hereford cattle. Livest. Prod. Sci. 63: 111-119.

Herd, R. M., J. A. Archer, and P. F. Arthur. 2003. Reducing the cost of beef production through genetic improvement of feed intake: Opportunity and challenges to application. J. Anim. Sci. 81(E-Suppl.): E9-E17.

Herd, R. M. and P. F. Arthur. 2009. Physiological basis for residual feed intake. J. Anim. Sci. 87: suppl E64-E71.

Koch, R. M., L. A. Swiger, D. Chambers, and K. E. Gregory. 1963. Efficiency of feed use in beef cattle. J. Anim. Sci. 22:486–494.

Mader, C. J., Y. R. Montanholi, Y. J. Wang, S. P. Miller, I. B. Mandell, B. W. McBride and K. C. Swanson. 2009. Relationships among measures of growth performance and efficiency with carcass traits, visceral organ mass, and pancreatic digestive enzymes in feedlot cattle. J. Anim. Sci. 87:1548-1557.

MacNeil, M. D. R. A. Nugent and W. M. Snelling. 1997. Breeding for Profit: an Introduction to Selection Index Concepts. Range Beef Cow Symposium. Paper 142.

Montanholi, Y. R., K. C. Swanson, R. Palme, F. S. Schenkel, B. W. McBride, D. Lu, and S. P. Miller. 2010. Assessing feed efficiency in beef steers through feeding behavior, infrared thermography and glucocorticoids. Animal. 4(5): 692 – 701.

Moore, S. S., F. D. Mujibi, and E. L. Sherman. 2009. Molecular basis for residual feed intake in beef cattle, J. Anim. Sci. 87:E41-E47.

Nkrumah, J. D., E. L. Sherman, C. Li, E. Marques, D. H. Crews Jr., R. Bartusiak, B. Murdoch, Z. Wang, J. A. Basarab, and S. S. Moore. 2007. Primary genome scan to identify putative quantitative trait loci for feedlot growth rate, feed intake, and feed efficiency of beef cattle. J. Anim. Sci. 85:3170–3181.

Sherman, E. L., J. D. Nkrumah, B. M. Murdoch, and S. S. Moore. 2008. Identification of polymorphisms influencing feed intake and efficiency in beef cattle. Anim. Genet. 39:225–231.

Snelling, W. M., R. A. Cushman, J. W. Keele, C. Maltecca, M. G. Thomas, M. R. S. Fortes, A. Reverter. 2012. Networks and pathways to guide genomic selection. J. Anim. Sci. doi:10.2527/jas.2012-5784.

Togashi, K., and C. Y. Lin. 2010. Theoretical efficiency of multiple-trait quantitative trait loci-assisted selection. J. Anim. Breed. Genet.127(1):53-63.

Villanueva, B., R. Pong-Wong, and J. A. Woolliams. 2002. Marker assisted selection with optimised contributions of the candidates to selection. Genet. Sel. Evol. 34:679–703.

Villanueva, B., R. Pong-Wong, J. Fernández and M. A. Toro. 2005. Benefits from marker-assisted selection under an additive polygenic genetic model. J. Anim. Sci. 83 (8): 1747-1752

Wang, Z., J. D. Nkrumah, C. Li, J. A. Basarab, L. A. Goonewardene, E. K. Okine, D. H., Crews, Jr. and S. S. Moore. 2006. Test duration for growth, feed intake, and feed efficiency in beef cattle using the GrowSafe System. J. Anim. Sci. 84:2289-2298.

# CHAPTER TWO

# LITERATURE REVIEW

## 2.1. UNDERSTANDING RESIDUAL FEED INTAKE

Koch et al., (1963) argued that a robust measure of feed efficiency should allow for the adjustment of feed intake for any of the 'energy sinks' present in various livestock production systems. For example, in beef production, growth is the major energy sink in growing cattle, maintenance of body weight is the major energy sink for mature cattle and reproduction is an additional energy sink in the breeding herd. Koch et al., (1963) therefore proposed residual feed intake (RFI) or net feed intake as a measure of feed efficiency defined as the difference between an individual's actual feed intake and its predicted feed intake. The predicted feed intake is estimated based on energy requirements for maintenance (body weight - BW) and weight gain (average daily gain – ADG). It is expected that individual animals will consume an amount of feed only enough to support maintenance of body weight and daily weight gain therefore the expected value for RFI is 0. RFI, therefore, is not correlated with body weight and average daily gain but is correlated with dry matter intake-DMI (Basarab et al., 2003). In beef cattle, animals which consume more feed than the amount predicted will have a positive value for RFI and are considered relatively inefficient while individuals that consume less-than-the-predicted amount of feed have a negative value for RFI and are considered more efficient (Herd and Arthur, 2009).

## 2.2. STATISTICAL COMPONENTS OF RFI

The computation of RFI for an individual animal applies a multiple linear regression summarized by Crews (2005) as:

$$y = \beta_0 + \beta_1(ADG) + \beta_2(BW) + RFI$$

Where; *y* is daily dry matter feed intake *(DMI)*

$\beta_0$ is the regression intercept

$\beta_1$ is the partial regression of daily intake on average daily gain *(ADG)*

$\beta_2$ is the partial regression of daily intake on body weight *(BW)*

From this equation, RFI is expected to be independent of ADG and BW but is expected to be positively correlated with DMI. Therefore, RFI will also change if ADG and/or BW change at a constant DMI, or if DMI changes at a constant BW and ADG.

## 2.3. PHYSIOLOGICAL REGULATION OF RFI

RFI measures whether the amount of feed an animal consumes is more or less than the predicted amount based on its body weight and average daily gain. Therefore variation in RFI between individuals can be explained by differences in the underlying biological processes that result in variation in energy utilization for maintenance and growth (Herd and Arthur, 2009). About 5 broad biological processes have been implicated in the variation in RFI including: feed intake,

digestion of feed, metabolism, physical activity and thermoregulation (Herd and Arthur, 2009). We review these processes and their roles in influencing RFI.

### 2.3.1. Regulation of RFI by feed intake

There are several factors that can affect feed intake and as a result may affect RFI. These include:

*Diet selection:* Although ruminants consume a wide variety of feeds, some ruminant species are highly selective (Preston and Leng, 1987) and the preferred feed is not necessarily the abundant type in the pasture. When pastures are abundant, feed selection increases, but when less forage is available feed selection is reduced (Preston and Leng, 1987). Sheep and goats tend to be more selective than cattle and other larger ruminants (Langlands and Sanson, 1976). In addition, there are also considerable differences in diet selection between individuals within the same herd, which may result in differences in feed intake (Arnold, 1963). Therefore it is clear that differences in diet selection between individuals may affect their feed intake and may result in differences in residual feed intake.

*Acetate clearance:* Ruminants match the rate of absorption of volatile fatty acids from the rumen with their utilization in metabolism. Weston (1966) showed that acetate clearance was highly correlated with the intake of Lucerne in sheep. The clearance rate of acetate is affected by the balance of nutrients available, particularly the ratio of acetate/propionate and acetate/amino acids. Even if acetate clearance is only associative and not causative it still appears to be ‘the

metabolic parameter which is most highly correlated with feed intake' (Weston, 1966).

*Appetite:* The complex brain networks that regulate appetite may either result in short term or long term effects. Short term effects occur in episodes causing bouts of eating while long term effects arise from tissue stores of fat, which exert a slow but consistent pressure on appetite (Halford and Blundell, 2000). The short term effects on appetite result in a unique feeding behavior for each individual (Halford and Blundell, 2000) and the relationship between RFI and these bouts of feeding behavior was reported previously by Golden et al., (2008) who showed that more efficient steers (low RFI) ate less feed and had fewer bouts of feeding than inefficient steers. Richardson (2003) also observed that steers with low RFI had shorter feeding periods than inefficient steers. *Neuropeptide Y (NPY)* is one of the neurotransmitters associated with appetite regulation in humans and animals (Sherman et al., 2008). Three SNPs in the NPY gene were significantly associated with body weight and growth rate in cattle and showed a trend of association with RFI ($P = 0.10$) (Sherman et al., 2008). If the SNPs affect the functions of NPY, then they may be influencing RFI by altering the appetite and feed intake.

*Cholecystokinin (CCK)* is a hormone released in the proximal small intestine and may possibly mediate the early phase of satiety. CCK reduces meal size and also suppresses hunger before a meal (Halford and Blundell, 2000). The association between cholecystokinin and appetite was observed in several species of animals including dairy cattle (Choi et al., 2000), pigs (Clutter et al., 1998) and rats (Reidelberger et al., 2003) and may affect RFI by altering feed intake.

*Glucagon-like peptide (GLP)-1* is a hormone released from the gut into the blood stream in response to intestinal nutrients such as carbohydrates, fats and proteins. GLP-1 causes increased insulin secretion, decreased glucagon secretion, increased sensitivity to insulin, increased insulin gene expression, inhibited gastric secretion of hydrochloric acid, reduced gastric emptying and decreased food intake (Holst, 2007). These effects if replicated in beef cattle may have profound effects on feed intake and therefore affect RFI.

*Peptide YY3-36 (PYY)* is a protein secreted from endocrine cells in intestines and is similar in structure to neuropeptide Y (NPY). Like NPY, PYY was also shown to reduce food intake and reduce the rate of weight gain in humans and mice (Batterham et al., 2002). PYY therefore may affect RFI by influencing feed intake and weight gain.

*Amylin* is a hormone secreted from the pancreas and also has an effect on both food intake and body weight (Reda et al., 2002). In mice and rats, administration of amylin resulted in reduced food intake, body weight and body mass (Rushing et al., 2001). These effects may eventually influence RFI.

*Leptin* is involved in long term regulation of appetite and occurs through a signal sent to the brain about the state of the adipose tissue (Weigle, 1994). Other similar signals believed to circulate in the blood and indicating the state of tissue energy include satietin, adipsin, tumour-necrosing factor (TNF) or cachectin, adiponectin, resistin and some cytokines (Halford and Blundell, 2000). There is some evidence that leptin interacts with NPY and with the melanocortin system to regulate appetite, body weight and body composition (Maffei et al., 1995). Significant

association between RFI and serum levels of leptin in beef steers have already been reported (Nkrumah et al., 2005; Kelly et al., 2010 and Hoque et al., 2009).

*Ghrelin (GHRL)* stimulates feeding behavior, rather than inhibiting it (Inui et al., 2004). GHRL is a hormone produced in the stomach and pancreas and is a potent stimulator of growth hormone (Inui et al., 2004). It stimulates feed intake and therefore may be associated with RFI. Sherman et al., (2008) reported significant association between a SNP in the Ghrelin gene with RFI and suggested that the role of GHRL in feed efficiency was likely in determining whether fat or carbohydrates would be used as the metabolic substrate for maintenance of energy balance as seen in GHRL knockout mice (Wortley et al., 2004). GHRL has also been shown to interact with NPY and play important roles in the stimulation of appetite and feeding activity (Jarkovska et al., 2004).

*Insulin* levels are positively correlated with adipose tissue mass within the body and may implicate insulin in the long term regulation of appetite in a similar cascade to leptin. Insulin stops the use of fat as an energy source by inhibiting the release of glucagon while glucagon causes the liver to convert stored glycogen into glucose (Halford and Blundell, 2000). Therefore glucagon and insulin play an antagonistic role in the feedback system that keeps blood glucose levels stable. Glucagon also enhances the body's physiological response to stress by increasing energy expenditure (Suzuki et al., 2012). Richardson et al., (2004) reported significant correlations between plasma insulin levels and RFI in beef cattle. However, contradicting results were reported by Kolath et al., (2006) who showed that although plasma glucose concentration was greater in high RFI steers

than in the low RFI steers, plasma insulin concentration was not different between the high RFI and low RFI groups.

### 2.3.2. **Regulation of RFI by digestion and absorption of nutrients**

The association between digestibility of dry matter and feed efficiency measures has been reported by Richardson et al., (1996), who indicated that young bulls and heifers that were ranked low or high for RFI differed in their ability to digest dry matter by about 1%. Digestibility was also correlated with RFI in cattle at a magnitude of r = –0.44 indicating that low RFI (more efficient) steers were better able to digest feed than high RFI steers (Richarson and Herd, 2004). Significant association was also reported between RFI and rumen microbial composition (Hernandez-Sanabria et al., 2012), which may indicate the role of rumen microbes in influencing feed digestibility. *(The role of microbes in regulation of RFI may also relate to the availability of amino acids (microbial protein) for absorption).* RFI can also be affected by feed absorption. There is evidence that animals that differ in RFI also differ in the appearance of amino acids at the portal vein (Lush et al., 1991).

Together, these results indicate that RFI may be associated with differences in both digestibility of feed ingested and the absorption rate of the nutrients. They indicate possible biological mechanism that can cause variation in feed efficiency even before the nutrients reach the tissues for metabolism.

### 2.3.3. **Regulation of RFI by nutrient metabolism**

The various metabolic processes occurring in the body account for over 40% of the phenotypic variation in RFI (Herd and Arthur, 2009). Most of the energy-

producing metabolic processes occur in the mitochondria. However, there are contradicting results reported in relation to the effect of mitochondrial function on RFI. For example, Kolath et al., (2006) observed no difference in the overall mitochondrial function between steers having low or high RFI values. However, there was significant association between RFI and the rate of mitochondrial respiration such that individuals with low RFI had increased rate of respiration and the flux of electrons through the electron transport chain was impaired in the high RFI steers (Kolath et al., 2006). Several other studies have attempted to establish association between mitochondrial enzymes and RFI; Ramos et al., (2011) reported significant associations between the mitochondrial complex 1 protein and RFI. The NADH dehydrogenase gene and the cytochrome gene complex have also been found to be significantly associated with variation in RFI (Zulkifli et al., 2009) though the exact mechanism is not established. In a separate study, the expression profile of mitochondrial genes from steers differing in RFI was used to identify genes that would be associated with RFI (Kelly et al., 2010) using real-time PCR to quantify mRNA transcripts of 17 genes associated with cellular energetic efficiency. The expression of UCP3 (mitochondrial uncoupling protein 3) was up-regulated about 2.2 fold in the high RFI group compared to the low RFI group. Other mRNA transcripts that were up-regulated in the low RFI steers were PGC-1 (PPAR-Gamma Coactivator 1) and COX II (cyclooxygenase-2). Several biological processes related to metabolism have been reported to be associated with RFI including amino acid metabolism (Herd and Arthur, 2009),

methane and nitrogen metabolism (Nkrumah et al., 2006) and glucose and nucleic acid metabolism.

The effect of metabolism on RFI can therefore be pursued in two directions; those metabolic processes that result in production of energy such as glycolysis, TCA cycle and oxidative phosphorylation (catabolism), and those metabolic processes that consume energy such as anabolism, production, thermoregulation and physical activity.

### 2.3.4. Regulation of RFI by energy expenditure

#### *2.3.4.1. Production and reproduction*

In animals, most of the energy is required for maintenance and production. In beef cattle, production encompasses tissue growth, and to a lesser extent growth of fetus and milk production in the breeding herd. Tissue growth is regarded as fat and protein synthesis using substrates including acetate, butyric acid, amino acids and glucose (Preston and Leng, 1987). Individual animals vary in their efficiency of depositing fat and protein (Herd and Arthur, 2009) explaining why associations between RFI and body composition have been reported (Basarab et al., 2003). Studies indicate that steers with low RFI have slightly reduced average back fat and inter-muscular fat than steers with high RFI and that steers with low RFI tend to have higher lean meat yield (Basarab et al., 2003). Metabolite analysis indicated that blood urea which is negatively correlated with protein content in bulls was positively correlated with RFI (Herd and Arthur, 2009). In a separate study, urea was negatively correlated with lean growth and positively correlated

with average back fat (Clarke et al., 1996). Creatinine, which is positively associated with muscle mass (Clarke et al., 1996) was negatively associated with RFI (Richardson et al., 2004)

The relationship between RFI and reproductive performance has had inconsistent results. Schaffer et al., (2010) showed that steers with low RFI took longer to reach puberty than steers with high RFI but showed no correlation with other reproductive measures including conception rates and calving interval. In a breeding herd, low RFI bulls had significantly lower individual progressive sperm motility than high RFI bulls (Wang et al., 2012). However, Wang et al., (2012) showed that using a multiple sire natural mating system, the mean number of progeny per sire was significantly higher in low RFI than in high RFI bulls

2.3.4.2. *Thermoregulation*

Evaporative heat loss through the lungs and nasal turbinates is the principle mechanism of heat loss in ruminants (Blaxter, 1962). RFI has been associated with heat loss traits in beef cattle with Montanholi et al., (2010) reporting that more efficient steers had lower temperatures in the snout and cheeks in comparison to less efficient steers but both groups had similar temperatures in the other body locations tested. Feed intake (which is correlated with RFI) was also shown to be correlated with air temperature, humidity, solar radiation and wind speed (Mujibi et al., 2010), which may relate indirectly with thermoregulation.

2.3.4.3. *Activity*

Energy expenditure may be the root cause of the regulation of appetite by physical activity; however, studies reported that acute exercise caused little or no

immediate effect on levels of hunger or daily energy intake (Halford and Blundell, 2000). The relationship between RFI and physical activity has been reported by several authors reviewed by Herd and Arthur (2009). Luiting et al., (1991) concluded that 80% of the genetic difference in RFI between lines of chickens divergent for RFI could be related to a difference in physical activity. The measures of activity that have been studied relate to feeding behavior such as frequency of feeding and duration of feeding. Durunna et al., (2011) reported that steers with high RFI and steers with low RFI also differed in their feeding frequency and feed duration. In addition, difference in activity was also reported by Richardson et al., (1996) showing a phenotypic correlation of 0.32 for RFI with daily pedometer count indicating that steers with high RFI had higher pedometer counts (locomotion) than steers with low RFI.

2.3.4.4. *Immune and stress response*

The resource allocation theory defined by Beilharz et al., (1993) states that in an environmentally limiting situation, animals have limited resources, and when resources are used by a certain biological process they are no longer available for other processes. Ideally, animals have to allocate their energy resources to the biological processes that will improve their fitness. Immune and stress responses are energy consuming processes (Deerenberg et al., 2000) and may be influenced by these energy trade-offs. Theoretically, it may be thought that selecting animals for RFI will result in animals that prioritize their energy to production traits at the cost of immune responses (Van Eerden et al., 2004). Conflicting results have been reported on this topic; Van Eerden et al., (2004) reported no change on measures

of humoral immune response in chicken with different RFI values. However, Galal et al., (2008) showed a positive correlation between cell mediated immune responses such as lymphocyte percentage and RFI. They showed that animals with low RFI had reduced lymphocyte percentage compared to high RFI animals. In beef cattle with divergent RFI values, white blood cell profiles did not differ between the two groups (Gomes et al., 2011). Differences in response to stress in steers selected for divergent RFI were reported (Richardson et al., 2004) showing that plasma cortisol levels were higher in high RFI steers than in low RFI steers and implied that high RFI steers may be more susceptible to stress than the low RFI steers. Stress being an energy consuming process may imply that steers with high RFI spend more energy to initiate these stress responses than their counterparts with low RFI.

## 2.4. UNDERSTANDING CARCASS QUALITY TRAITS

### 2.4.1. Types of Carcass data

Carcass quality data obtained from either actual measurements or estimated by ultrasound can be described in five categories:

a. Quality grade or the corresponding marbling score

b. Yield grade, whose components include carcass weight, fat thickness, percentage kidney, pelvic and heart fat, and ribeye area.

c. Carcass weight

d. Ribeye area

e. Fat (backfat) thickness

*a. Quality grade*

A carcass' quality grade is determined by the amount of marbling on the cut surface of the ribeye between the 12$^{th}$ and 13$^{th}$ ribs, such that increased marbling results in higher quality grade.

*b. Yield grade*

Yield grade estimates the amount of closely trimmed retail cuts of meat that a carcass is likely to yield. Yield grade, as a proportion of lean meat is classified as: 1 = ≥59%; 2 = 54 to 58%; and 3 = <54% (Basarab et al., 2003).

*c. Carcass weight*

Carcass weight is recorded just before the carcass enters the chilling room therefore it is sometimes referred to as hot carcass weight. It reflects the approximate size of the cuts of meat that may be expected after further processing of the carcass for example heavier carcasses result in larger ribeyes.

*d. Ribeye area*

This area is measured as the surface area on the cut surface of the ribeye muscle between the 12$^{th}$ and 13$^{th}$ ribs. The ribeye area is used to calculate the yield grade such that large carcasses have larger ribeyes.

*e. Fat (backfat) thickness*

Average backfat thickness is measured as the amount of fat opposite the ribeye at the cut surface between the $12^{th}$ and $13^{th}$ ribs. External fat is generally considered as a waste of product but some fat is needed to prevent the carcass from cooling too fast, it protects the meat from drying and enhances the tenderization process.

### 2.5. Regulation of traits at the molecular level

Phenotypes are regulated at multiple levels as shown in figure 2.1. Variations arising at any level of regulation may subsequently influence the variation in the respective phenotype (Banks et al., 2000). Distinct levels also interact with other levels so that the phenotype is shaped not only by the individual levels of regulation but also the effects arising from the interaction between the levels. Figure 2.1 summarizes these levels and the interactions existing between them.

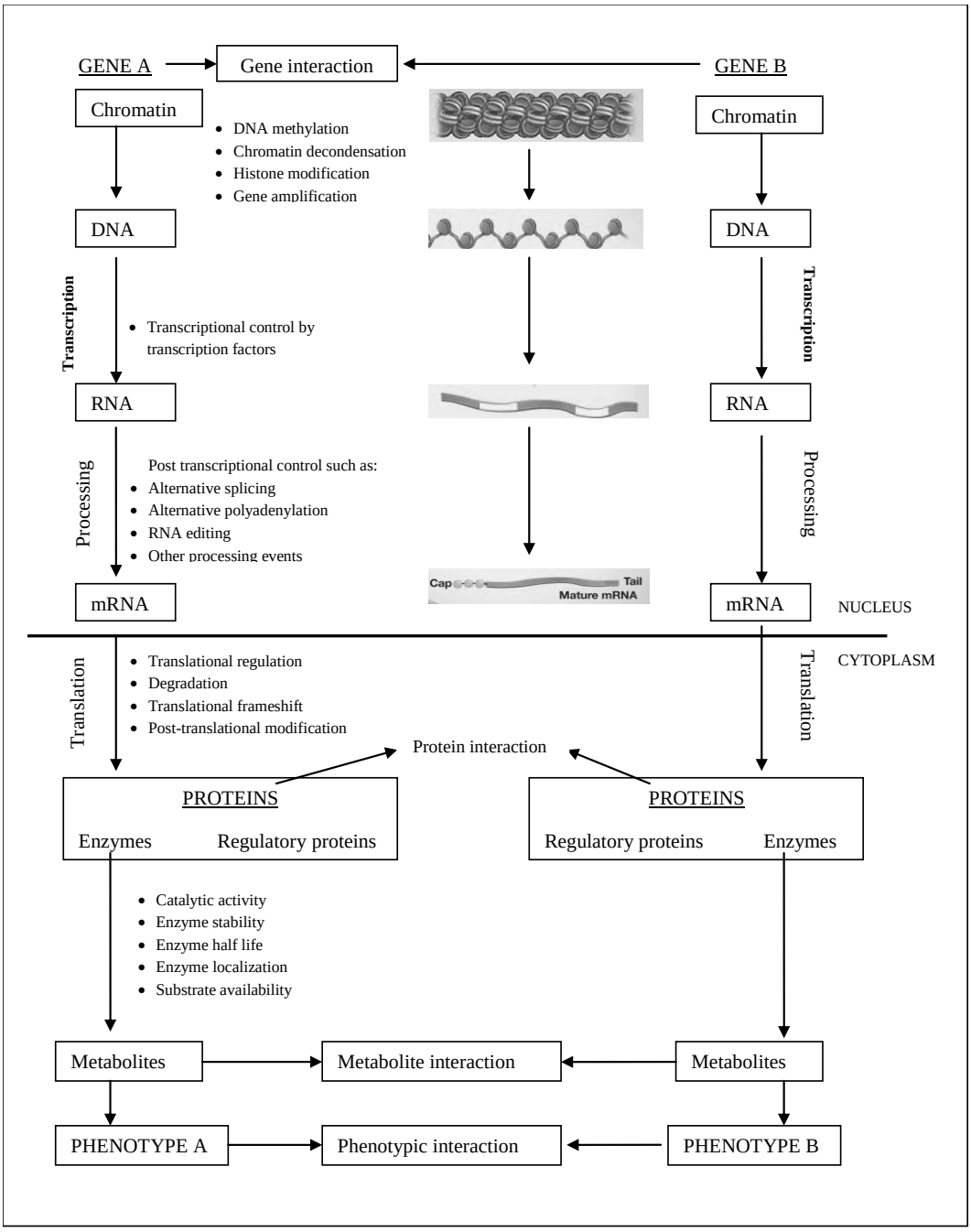



*Figure 2.1: Levels of regulation of gene expression starting from the genomic levels to the products of metabolic reactions at the metabolites levels. Distinct levels of regulation also interact with other levels eventually shaping the variation observed in the traits.*

## 2.6. SELECTION FOR RFI

Traits that are candidates for selection must have specific properties, one of which is that the trait should have considerable phenotypic variation. To select for RFI, whose component traits are feed intake and growth, considerable phenotypic variation must be demonstrated in daily feed intake or dry matter intake and/or growth rate/average daily gain. Several studies have reported considerable phenotypic variation in feed intake (Arthur et al., 2001; Basarab et al., 2003; Crews et al., 2006) therefore providing the opportunity for selection for RFI.

The presence of phenotypic variation alone is not sufficient; the observed differences between animals must have an underlying genetic basis that is partly due to additive genetic effects. RFI has a moderate heritability ranging from 0.16 to 0.58 (Koch et al., 1963; Arthur et al., 2001; Crews et al., 2003; Shenkel et al., 2004). This indicates that benefits of improved RFI can be passed from parents to their offspring.

Lastly, traits that are candidates for selection must be assessed for their genetic correlations with other productivity traits. For example, RFI is strongly correlated with feed conversion ratio (FCR) (0.70, Herd and Bishop, 2000; 0.85, Arthur et al., 2001).

### 2.6.1. Response to selection

The response to selection per year can be estimated using the formula:

$$R_{year} = \frac{i_{male} + i_{female}}{L_{male} + L_{female}} \sigma_p h^2$$

Where;

$i_{male} + i_{female}$ is the selection intensity associated with males and females respectively.

$L_{male} + L_{female}$ is the generation interval for males and females respectively

$\sigma_p$ is the phenotypic standard deviation

$h^2$ is the heritability of the trait.

A detailed discussion of this topic is available at Falconer and Mackay (1996).

In summary, the selection intensity (*i)* represents the number of standard deviation units that selected parents are superior to the mean of the general population. It is obtained from selection intensity tables when the proportion of selected animals is known. When few animals are selected, the proportion of selected animals to the entire population will be low and the selection intensity will be high. The selection intensities between males and females are different because most beef production systems consist of few bulls responsible for mating several females; therefore the selection intensity for males is higher than that for females.

The generation interval (*L*) is the average age of parents when progeny are born or the average time between birth of parents and birth of progeny. L varies widely across species. However, it can be altered within species by changing the age at which animals are selected and bred. The generation interval is also calculated separately for males and females and then averaged.

$\sigma_p$ is the standard deviation of the trait, and from the equation, the response to selection is higher if when $\sigma_p$ is higher.

$h^2$ is the heritability of the trait, and the higher the heritability, the higher the response to selection. Heritability is estimated by the formula $V_A/V_P$ where $V_A$ is the additive genetic variance and $V_P$ is the phenotypic variance; therefore heritability is the ratio of additive genetic variance or variance of breeding value to the overall or phenotypic variance (Hill, 2010).

### 2.6.2. Approaches to selection

The traditional selection practiced by most breeders is based on the quantitative genetic approach (Dekkers and Hospital, 2002). This approach relies on the availability of phenotype data, the heritability of the traits and their genetic correlations to select superior individuals and to predict the performance of the progeny. The quantitative genetics approach assumes that the genetic architecture and number of gene affecting the trait are unknown (Dekkers and Hospital, 2002). It has several limitations because phenotypes are also influenced by the environment making them imperfect predictors of the breeding value of an individual. In some cases, phenotypes are only observable in one gender and not the other such as milk yield observed in lactating cows but not in bulls. In addition, some phenotypes cannot be observed before the time when selection decisions must be made such as carcass traits in beef cattle, which can only be observed when the individual has already been sacrificed and can no longer be used for breeding. Lastly, phenotypic selection is not effective in resolving

negative genetic correlations between traits and/or genes such as the negative correlations caused by pleiotropy or epistasis (Dekkers and Hospital, 2002).

However, despite the assumptions in the quantitative genetics approach, tremendous genetic improvement has been achieved in livestock and crop production exemplified by milk production, broiler body weight and yield of corn (Dekkers and Hospital, 2002)

More recently, to overcome the limitations of phenotypic selection, molecular markers have gained importance in the evaluation and ranking of candidates for selection. Molecular markers can also be used to understand gene functions and exploit the information on the genes themselves and the relationship between gene function and phenotypic variation. The use of molecular markers in selection relies on the ability to determine the genotypes of individuals for the mutations associated with the traits of interest (Dekkers and Hospital, 2002). The mutations fall into two broad types of markers: *indirect markers* and *causal mutations*. Indirect markers are presumed to be non-functional genetic markers that are linked to QTL while causal mutations are located within the genes that directly affect the trait (Dekkers and Hospital, 2002). The causal mutations are more difficult to find and prove and therefore only a few examples are available as reported by Andersson (2001). Several other mutations associated with economically important traits in livestock have been reported and reviewed by Sellner et al., (2007), Goddard and Hayes (2009) and Meuwissen et al., (2013). The indirect markers are more abundant in the genome and their linkage with QTL can be established by evidence of empirical associations of marker

genotypes with trait phenotype (Dekkers and Hospital, 2002). Two approaches are used to identify indirect markers; the gene association analysis approach and the classical QTL mapping (Andersson, 2001). The gene association analysis can be more targeted and may even identify the causal mutations or markers that are very tightly linked to the causal gene. It utilizes unstructured natural populations that have undergone some degree of random mating (Dekkers and Hospital, 2002). On the other hand, classical QTL mapping utilizes specialized populations such as $F_2$ crosses. This approach identifies chromosomal regions associated with the trait. These regions average 10-20cM but the exact position of the QTL or the underlying QTN is unknown and difficult to identify (Dekkers and Hospital, 2002). These approaches are discussed further in Chapter 3.

The use of genetic markers in selection programs relies on the ability to identify the genotypes of individuals for the mutations (direct or indirect markers). This information is then used to develop a molecular score that can be used for selection.

A significant amount of research has been conducted to determine the genetic and molecular basis of RFI and several genetic association studies have resulted in the identification of several SNPs associated with RFI (Moore et al., 2006; Barendse et al., 2007; Nkrumah et al., 2007; Sherman et al., 2008, 2009; Bolormaa et al., 2011; Snelling et al., 2011; Elzo et al., 2012). In addition to these genes, levels of IGF-1 in blood were reported to be associated with increased feed efficiency (Bishop et al., 1989; Stick et al., 1998).

Genome-wide association analyses for RFI have been performed (Nkrumah et al., 2007). Eight QTLs associated with RFI were identified in 8 chromosomes including 1, 5, 7, 8, 12, 16, 17 and 26. Suggestive QTLs were identified in chromosomes 2, 14, 18, 19, 20, 21, 24, 28 and 29.

There are specific examples of mutations associated with variation in RFI:

A C/G mutation at position 2141 in exon 5 of the growth hormone leading to a leucine/valine change in the amino acids sequence at position 127 was reported to be associated with feed efficiency and the GG genotype was associated with lower ADG, BW, meat deposition and low lean yield (Chrenek et al., 1998).

Several polymorphisms located in the leptin gene have also been shown to be associated with feed efficiency: C/T polymorphism at position 207; T/C at position 305; C/T at position 528; and C/G at position 1756 (Buchanan et al., 2002). Some polymorphisms located in the promoter region of the leptin gene were also shown to be associated with feed efficiency (Nkrumah et al., 2005). The SNPs in the promoter region were also significantly ($P < 0.05$) associated with dry matter intake, backfat thickness, marbling score and rib eye area.

An intronic SNP located in the growth hormone receptor gene (GHR) was shown to be associated with RFI (Sherman et al., 2008). Sherman et al., (2008) also identified some SNPs that showed a trend of association with RFI. These included SNPs located in the ghrelin (GHRL) gene and the neuropeptide Y (NPY) gene.

Eight markers located within and between the NCAPG and LOC540095 genes on chromosome 6 were shown to be significantly associated with feed intake and average daily gain in beef cattle (Lindholm-Perry et al., 2010)

A detailed analysis of bovine chromosome 14 was performed to identify genes associated with RFI in beef cattle (Lindholm-Perry et al., 2012). The authors identified 5 markers located downstream of TMEM68, and between TMEM68 and the neighboring XKR4 gene that were predictive for feed intake and gain.

### 2.6.3. Commercialized selection for RFI

There are two commercialized tests available for testing the RFI values in cattle. **Igenity®**-feed efficiency offers analyses to predict RFI in *Bos indicus* (Igenity feed efficiency Indicus) and *Bos taurus* (Igenity feed efficiency Taurus) breeds of cattle. The analysis from Igenity is reported as a series of scores on a scale of 1 – 10 with 1 corresponding to less intake and 10 corresponding to high intake. Igenity believes that these analyses will help producers to identify breeding cattle that will be more efficient and will also produce more efficient offspring while maintaining a good body condition score and without negatively affecting their fertility, (obtained from the Igenity website; http://www.igenity.com/beef/profile/FeedEfficiency.aspx). An independent validation analysis showed that Igenity feed efficiency Indicus was not predictive for RFI in the Brahman breed (NBCEC, 2011). During the validation analysis, the Igenity feed efficiency Taurus test was shown to be inconsistently associated with RFI in *Bos taurus* breeds of cattle. The authors found significant associations

between this test and RFI in two populations of *Bos taurus* cattle but the association was not significant in four other populations of *Bos taurus* cattle (NBCEC, 2011).

**Pfizer GeneSTAR –**feed efficiency offers an analysis for feed efficiency in cattle. A validation test showed that this test successfully predicted feed efficiency in North American *Bos taurus* cattle but not in *Bos indicus* influenced cattle (NBCEC, 2011).

The results obtained by the NBCEC indicated that the available commercialized markers are population specific and will not result to similar genetic progress across populations.

## 2.7. CONCLUSION

Traditionally, selection of beef cattle for RFI was based on phenotypic values without a clear understanding of the underlying molecular mechanisms influencing the trait. This approach is expensive and takes a long time to obtain the RFI values for each individual. The advent of genetic markers has the potential to improve the accuracy of selection and to reduce the generation interval, both of which would result in increased response to selection. Several DNA variants have so far been shown to be associated with RFI and some DNA tests have been developed to test animals for feed efficiency. However, despite the DNA variants shown to be associated with RFI so far and the DNA test developed thereof, only a small portion of the variation in RFI has been accounted for by these variants and there has been frequent failure to reproduce the associations in

other populations of cattle. The candidate gene approach may offer an alternative and superior approach to discover DNA variants that will account for more variation in RFI, markers that may be significantly associated with RFI in different beef cattle breeds and populations.

## 2.8. REFERENCES


Andersson, L. 2001. Genetic dissection of phenotypic diversity in farm animals. Nat. Rev. Genet. 2: 130 - 138.

Arnold, G. W. 1963. Some principles in the investigation of selective grazing. Proc. Australian Soc. Anim. Production. 5:258 - 271

Arthur, P. F., J. A. Archer, D. J. Johnston, R. M. Herd, E. C. Richardson, and P. F. Parnell. 2001. Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. J. Anim. Sci. 79:2805-2811.

Banks, R. E, M. J. Dunn, D. F. Hochstrasser, J. Sanchez, W. Blackstock, D. J. Pappin, P. J. Selby. 2000. Proteomics: new perspectives, new biomedical opportunities. The Lancet. 356 (9243): 1749 - 1756,

Barendse, W., A. Reverter, R. J. Bunch, B. E. Harrison, W. Barris, and M. B. Thomas. 2007. A validated whole-genome association study of efficient food conversion in cattle. Genetics. 176:1893–1905.

Basarab, J. A., M. A. Price, J. A. Aalhus, E. K. Okine, W. M. Snelling, and K. L. Lyle. 2003. Residual feed intake and body composition in young growing cattle. Can. J. Anim. Sci. 83:189–204.

Batterham, R. L., M. A. Cowley, C. J. Small, H. Herzog, M. A. Cohen, C. L. Dakin, A. M. Wren, A. E. Brynes, M. J. Low, M. A. Ghatei, R. D. Cone, S. R. Bloom. 2002. Gut hormone PYY(3-36) physiologically inhibits food intake. Nature. 418(6898):650-4.

Beilharz, R. G., B. G. Luxford, and J. L. Wilkinson. 1993. Quantitative genetics and evolution: Is our understanding of genetics sufficient to explain evolution? J. Anim. Breed. Gen. 110: 161–170.

Bishop, M. D., R. C. M. Simmen, F. A. Simmen, and M. E. Davis. 1989. The relationship of insulin-like growth factor-I with post-weaning performance in Angus beef cattle. J. Anim. Sci. 67:2872–2880.

Blaxter, K. L. 1962. The energy metabolism of ruminants. Hutchinson Scientific and Technical, London, UK.

Bolormaa, S., B. J. Hayes, K. Savin, R. Hawken, W. Barendse, P. F. Arthur, R. M. Herd, and M. E. Goodard. 2011. Genome-wide association studies for feedlot and growth traits in cattle. J. Anim. Sci. 89: 1684-1697.

Buchanan, F. C., C. J. Fitzsimmons, A. G. Van Kessel, T. D. Thue, C. Winkelman-Sim, and S. M. Schmutz. 2002. Association of a missense mutation in the bovine leptin gene with carcass fat content and leptin mRNA levels. Genet. Sel. Evol. 34:105–116.

Choi, B. R., D. L. Palmquist, and M. S. Allen. 2000. Cholecystokinin mediates depression of feed intake in dairy cattle fed high fat diets. Domest Anim Endocrinol. 19(3):159-75.

Chrenek, P., J. Kmet, T. Sakowski, T. Vasicek, J. Huba, J. Chrenek. 1998. Relationships of growth hormone genotypes with meat production traits of Slovak Pied bulls. Czech Journal of Animal Science. 43, 541-544.

Clarke, J. N., D. B. Binnie, K. R. Jones, C. M. Mowat, R. W. Purchas, and A. E. Uljee. 1996. Repeatabilities of blood plasma metabolites and their association with leanness in genotypes showing a wide divergence in carcass composition. Proc. N.Z. Soc. Anim. Prod. 56:180–183.

Clutter, A. C., R. Jiang, J. P. McCann, and D. S. Buchanan. 1998. Plasma cholecystokinin-8 in pigs with divergent genetic potential for feed intake and growth. Domestic Animal Endocrinology. 15(1):9-21.

Crews, D. H. Jr., N. H. Shannon, B. M. A. Genswein, R. E. Crews, C. M. Johnson, and B. A. Kendrick. 2003. Genetic parameters for net feed efficiency of beef cattle measured during postweaning growing versus finishing periods. Proc. West. Sec. Amer. Soc. Anim. Sci. 54:125-128.

Crews, D. H. Jr. 2005. Genetics of feed utilization and national cattle evaluation: A review. Genet. Mol. Res. 4:152–165.

Crews, D. H. Jr., G. E. Carstens, and P. A. Lancaster. 2006. Case Study: A multiple trait selection index including feed efficiency. Prof. Anim. Sci. 22:65-70.

Deerenberg, C., V. Arpanius, S. Daan, and N. Bos. 1997. Reproductive effort decreases antibody responsiveness. Proc. R. Soc. Lond. B. Biol. Sci. 264:1021–1029.

Dekkers, J. C. M., and F. Hospital. 2002. Utilization of molecular genetics in genetic improvement of plants and animals. Nature Reviews: Genetics 3: 22-32.

Durunna, O. N, Z. Wang, J. A. Basarab, E. K. Okine, and S. S. Moore. 2011. Phenotypic and genetic relationships among feeding behavior traits, feed intake, and residual feed intake in steers fed grower and finisher diets. J Anim Sci. 89(11):3401-9.

Elzo, M. A., G. C. Lamb, D. D. Johnson, M. G. Thomas, I. Misztal, D. O. Rae, C. A. Martinez, J. G. Wasdin and J. D. Driver. 2012. Genomic-polygenic evaluation of Angus-Brahman multibreed cattle for feed efficiency and postweaning growth using the Illumina 3K chip. J. Anim. Sci. 90: 2488 – 2497.

Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to quantitative genetics. Essex, England :Prentice Hall.

Galal, A., A. M. H. Ahmed, W. A. H. Ali, M. H. El-Sanhoury, and H. E. Ahmed. 2008. Residual Feed Intake and its Effect on Cell-Mediated Immunity in Laying Hens Given Different Propolis Levels. International Journal of Poultry Science. 7 (11): 1105-1111.

Goddard, M. E., B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programs. Nat. Rev. Genet. 10(6): 381-91.

Golden, J. W., M. S. Kerley, W. H. Kolath. 2008. The relationship of feeding behavior to residual feed intake in crossbred Angus steers fed traditional and no-roughage diets. J Anim Sci. 86(1):180-6.

Gomes, D. R., R. F. De Siqueira, M. A. Ballou, T. R. Stella, and P. R. Leme. 2011. Hematological profile of beef cattle with divergent residual feed intake,

following feed deprivation Pesq. agropec. bras., Brasília. 46 (9):1105-1111.

Halford, J. C. and J. E. Blundell. 2000. Pharmacology of appetite suppression. Prog Drug Res. 54:25-58.

Herd, R. M. and P. F. Arthur. 2009. Physiological basis for residual feed intake. J. Anim. Sci. 87: suppl E64-E71.

Herd, R. M. and S. C. Bishop. 2000. Genetic variation in residual feed intake and its association with other production traits in British Hereford cattle. Livest. Prod. Sci. 63:111-119.

Hernandez-Sanabria, E., L. A. Goonewardene, Z. Wang, O. N. Durunna, S. S. Moore, and L. L. Guan. 2012. Impact of feed efficiency and diet on adaptive variations in the bacterial community in the rumen fluid of cattle. Appl. Environ. Microbiol. 78(4):1203-14.

Hill, W. G. 2010. Understanding and using quantitative genetic variation. Phil. Trans. R. Soc. B. 365 (1537); 73-85.

Holst, J. J. 2007. The Physiology of Glucagon-like Peptide 1. Physiol. Rev. 87 (4): 1409-1439.

Hoque, M. A., K. Katoh, and K. Suzuki. 2009. Genetic associations of residual feed intake with serum insulin-like growth factor-I and leptin concentrations, meat quality, and carcass cross sectional fat area ratios in Duroc pigs. J Anim Sci. 87(10):3069-75.

Inui, A., A. Asakawa, and C. Y. Bowers. 2004. "Ghrelin, appetite, and gastric motility: the emerging role of the stomach as an endocrine organ". FASEB J. 18 (3): 439–56.

Jarkovska, Z., M. Krsek, M. Rosicka, and J. Marek. 2004. Endocrine and metabolic activities of a recently isolated peptide hormone ghrelin, an endogenous ligand of the growth hormone secreta-gogue receptor. Endocr. Regul. 38:80–86.

Kelly, A. K., M. McGee, D. H. Crews, Jr, A. G. Fahey, A. R. Wylie, and D. A. Kenny. 2010. Effect of divergence in residual feed intake on feeding behavior, blood metabolic variables, and body composition traits in growing beef heifers. J Anim Sci. 88(1):109-23.

Kelly, A. K., S. M. Waters, M. McGee, R. G. Fonseca, C. Carberry, and D. A. Kenny. 2011. mRNA expression of genes regulating oxidative phosphorylation in the muscle of beef cattle divergently ranked on residual feed intake. Physiol. Genomics. 43(1):12-23.

Koch, R. M., L. A. Swiger, D. Chambers, and K. E. Gregory. 1963. Efficiency of feed use in beef cattle. J. Anim. Sci. 22:486–494.

Kolath, W. H., M. S. Kerley, J. W. Golden, and D. H. Keisler. 2006. The relationship between mitochondrial function and residual feed intake in Angus steers. J. Anim. Sci. 84:861–865.

Langlands, J. P. and J. Sanson. 1976. Factors affecting the nutritive value of the diet and the composition of rumen fluid of grazing sheep and cattle. Australian Journal of Agricultural Research. 27: 691-707.

Lindholm-Perry, A. K., L. A. Kuehn, W. M. Snelling, T. P. L. Smith, C. L. Ferrell, T. G. Jenkins, H. C. Freetly. 2010. Polymorphisms in POMC are not associated with dry matter intake and average daily gain phenotypes in beef cattle. Animal Genetics. 41(6):669-670.

Lindholm-Perry, A. K., L. A. Kuehn, T. P. Smith, C. L. Ferrell, T. G. Jenkins, H. C. Freetly, W. M. Snelling. 2012. A region on BTA14 that includes the positional candidate genes LYPLA1, XKR4 and TMEM68 is associated with feed intake and growth phenotypes in cattle. Animal Genetics. 43(2):216-219.

Luiting, P., J. W. Schrama, W. Van der Hel, and E. M. Urff. 1991. Metabolic differences between white leghorns selected for high and low residual feed consumption. Br. Poult. Sci. 32:763–782.

Lush, J. M., J. M. Gooden, and E. F. Annison. 1991. The uptake of nitrogenous compounds from the gut of sheep genetically different in wool production. Proc. Nutr. Soc. Aust. 16:144.

Maffei, M., J. Halaas, E. Ravussin, R. E. Pratley, G. H. Lee, Y. Zhang, H. Fei, S. Kim, R. Lallone, and S. Ranganathan. 1995. Leptin levels in human and rodent: measurement of plasma leptin and ob RNA in obese and weight-reduced subjects. Nat Med. 1(11):1155-61.

Meuwissen, T., B. Hayes and M. Goddard. 2013. Accelerating Improvement of Livestock with Genomic Selection. Annu. Rev. Anim. Biosci. 2013. 1:221–237.

Montanholi, Y. R., K. C. Swanson, R. Palme, F. S. Schenkel, B. W. McBride, D. Lu, and S. P. Miller. 2010. Assessing feed efficiency in beef steers through feeding behavior, infrared thermography and glucocorticoids. Animal. 4 (5):692-701.

Moore, S. S., D. H. Crews Jr., and J. D. Nkrumah. 2006. Multiple and candidate gene approaches to genetic evaluation of feed efficiency in beef cattle. Proc. 8th World Congr. Genet. Appl. Livest. Prod. (Brazilian Society of Animal Breeding, Sao Paolo, Brazil). No. 14–01, Belo Horizonte, Brazil.

Mujibi, F. D., S. S. Moore, D. J. Nkrumah, Z. Wang, and J. A. Basarab. 2010. Season of testing and its effect on feed intake and efficiency in growing beef cattle. J Anim Sci. 88(12):3789-99.

Nkrumah, J. D., C. Li, J. Yu, C. Hansen, D. H. Keisler, and S. S. Moore. 2005. Polymorphisms in the bovine leptin promoter associated with serum leptin concentration, growth, feed intake, feeding behavior, and measures of carcass merit. J Anim Sci. 83(1):20-8.

Nkrumah, J. D., E. K. Okine, G. W. Mathison, K. Schmid, C. Li, J. A. Basarab, M. A. Price, Z. Wang, and S. S. Moore. 2006. Relationships of feedlot feed efficiency, performance, and feeding behavior with metabolic rate, methane production, and energy partitioning in beef cattle. J Anim Sci. 84(1):145-53.

Nkrumah, J. D., E. L. Sherman, C. Li, E. Marques, D. H. Crews Jr., R. Bartusiak, B. Murdoch, Z. Wang, J. A. Basarab, and S. S. Moore. 2007. Primary genome scan to identify putative quantitative trait loci for feedlot growth rate, feed intake, and feed efficiency of beef cattle. J. Anim. Sci. 85:3170–3181.

Preston, T. R. and R. A. Leng. 1987. Matching ruminant production systems with available resources in the tropics and sub-tropics. *Penambul Books Armidale.*

Ramos, M. H. 2011. Mitochondrial complex I protein is related to residual feed intake in beef cattle. A Dissertation presented to the Faculty of the Graduate School at the University of Missouri-Columbia.

Reda, T. K., A. Geliebter, and F. X. Pi-Sunyer. 2002. Amylin, food intake, and obesity. Obes. Res. 10 (10):1087-91.

Reidelberger, R. D., D. A. Castellanos, and M. Hulce. 2003. Effects of peripheral CCK receptor blockade on food intake in rats. Am. J. Physiol. Regul. Integr. Comp. Physiol. 285(2):R429-37

Richardson, E. C., R. M. Herd, P. F. Arthur, J. Wright, G. Xu, K. Dibley, and V. H. Oddy. 1996. Possible physiological indicators for net feed conversion efficiency in beef cattle. Proc. Aust. Soc. Anim. Prod. 21:103–106.

Richardson, E. C. 2003. Biological basis for variation in residual feed intake in beef cattle. PhD Diss. Univ. New England, Armidale, Australia.

Richardson, E. C. and R. M. Herd. 2004. Biological basis for variation in residual feed intake in beef cattle. 2. Synthesis of results following divergent selection. Aust. J. Exp. Agric. 44:431–440.

Richardson, E. C., R. M. Herd, J. A. Archer and P. F. Arthur. 2004. Metabolic differences in Angus steers divergently selected for residual feed intake. Australian Journal of Experimental Agriculture. 44(5): 441 - 452

Rushing, P. A., M. M. Hagan, R. J. Seeley, T. A. Lutz, D. A. D'Alessio, E. L. Air, and S. C. Woods. 2001. Inhibition of central amylin signaling increases food intake and body adiposity in rats. Endocrinology. 142 (11):5035.

Sellner, E. M., J. W. Kim, M. C. McClure, K. H. Taylor, R. D. Schnabel and J. F. Taylor. 2007. Applications of genomic information in livestock. J. Anim. Sci. 85 (12) 3148-3158.

Shaffer, K. S., P. Turk, W. R. Wagner, and E. E. D. Felton. 2011. Residual feed intake, body composition, and fertility in yearling beef heifers. J Anim Sci. 89(4):1028-34.

Schenkel, F. S., S. P. Miller, and J. W. Wilton. 2004. Genetic parameters and breed differences for feed efficiency, growth, and body composition traits of young beef bulls. Can. J. Anim. Sci. 84:177- 184.

Sherman, E. L., J. D. Nkrumah, B. M. Murdoch, C. Li, Z. Wang, A. Fu, and S. S. Moore. 2008. Polymorphisms and haplotypes in the bovine neuropeptide Y, growth hormone receptor, ghrelin, insulin-like growth factor 2, and uncoupling proteins 2 and 3 genes and their associations with measures of

growth, performance, feed efficiency, and carcass merit in beef cattle. J Anim Sci. 86(1):1-16.

Snelling, W. M., M. F. Allan, J. W. Keele, L. A. Kuehn, R. M. Thallman, G. L. Bennett, C. L. Ferrell, T. G. Jenkins, H. C. Freetly, M. K. Nielsen, and K. M. Rolfe. 2011. Partial-genome evaluation of postweaning feed intake and efficiency of crossbred beef cattle. J Anim. Sci. 89:1731-1741.

Stick, D. A., M. E. Davis, S. C. Loerch, and R. C. M. Simmen. 1998. Relationship between blood serum insulin-like growth factor 1 concentration and postweaning feed efficiency of crossbred cattle at three levels of dietary intake. J. Anim. Sci. 76:498–505.

Suzuki, K., C. N. Jayasena, and S. R. Bloom. 2012. Obesity and Appetite Control. Exp. Diabetes Res. 824305. doi: 10.1155/2012/824305.

Van Eerden, E., H. Van Den Brand, H. K. Parmentier, M. C. De Jong, and B. Kemp. 2004. Phenotypic selection for residual feed intake and its effect on humoral immune responses in growing layer hens. Poult Sci. 83(9):1602-9.

Wang, Z., M. G. Colazo, J. A. Basarab, L. A. Goonewardene, D. J. Ambrose, E. Marques, G. Plastow, S. P. Miller, and S. S. Moore. 2012. Impact of selection for residual feed intake on breeding soundness and reproductive performance of bulls on pasture-based multi-sire mating. J Anim Sci. 90(9):2963 -9

Weigle, D. S. 1994. Appetite and the regulation of body composition. FASEB J. 8(3):302-10.

Weston, R. H. 1966. The effect of level of feeding on acetate tolerance in the sheep. Australian Journal of Agricultural Research. 17: 933-937.

Wortley, K. E., K. D. Anderson, K. Garcia, J. D. Murray, L. Malinova, R. Lui, M. Moncrieffe, K. Thabet, H. J. Cox, G. D. Yancopoulos, S. J. Wiegand, and M. W. Sleeman. 2004. Genetic deletion of ghrelin does not deCLMAse food intake but influences metabolic fuel preference. Proc. Natl. Acad. Sci. USA 101:8227–8232.

Zulkifli, N. A., M. Naik, W. S. Pitchford, and C. D. K. Bottema, 2009. Cattle residual feed intake candidate genes. Proc. Assoc. Advmt. Anim. Breed. Genet. 18:668-671

## *PREFACE TO CHAPTER THREE*

*Chapter 3 consists of a review paper published in Springer Science Reviews DOI: 10.1007/s40362-013-0005-8 under the same title. This was a competition by Springer as part of a 'brand new initiative aimed specifically at providing a high-profile vehicle for outstanding early career researchers to showcase their emerging talent.' I was nominated to submit this manuscript by my PhD supervisor Dr. Graham Plastow.*

*This review describes the utilization of the systems biology approach in the analysis of quantitiative traits in animal science. Specifically, it details the processes utilized in the identification of genetic markers associated with economically important traits in animals. Then it is followed by assessing the interactions between genes and metabolites using biological networks.*

# CHAPTER THREE

## Complicated relationships: A review of biological interaction networks and pathways in animal science[1]

### 3.1. INTRODUCTION

Profits in any enterprise are defined as the difference between costs and returns. In livestock production, profits can be increased by reducing the costs of production (inputs) and/or by increasing the returns. Returns can be increased by increasing production and/or by increasing the price of the product. Considering the limited capacity to increase production, farmers may seek to emphasize quality so as to increase the price of the product.

In most livestock production systems, the cost of feed is the largest single expense (input). It accounts for between 60- 70% of the total cost of production in beef cattle (Herd et al. 2003; Arthur et al. 2004), about 65% of total costs in pigs (Hoque et al. 2009) and approximately 70% in broilers (Aggrey et al. 2010). On the other hand, production traits (outputs) differ across different livestock production systems such as eggs and meat in poultry production, milk in dairy production and meat in beef cattle and pig production systems. These production traits influence the economic success in livestock production, which relies on

[1] Published in Springer Science Reviews, 2013.

producing products of high economic value at the lowest cost possible (MacNeil et al. 1997).

Most of the traits that are of economic importance to livestock producers are quantitative in nature; they are influenced by multiple genes, each causing a small effect on the trait. In addition, most economically important traits are correlated. Therefore selecting individuals for one trait could impact upon another trait.

In the past, selection of individuals for economically important traits was solely based on phenotypic measurements and statistical predictions using information collected from relatives (Dekkers and Hospital 2002). Although this strategy has had several successes, the advent of genetic markers for marker assisted selection has the potential to increase the accuracy of selection and reduce generation interval resulting in an increase in the response to selection (Snelling et al. 2012). Genetic markers are especially important when it is expensive to collect the required phenotypic data on the traits of interest, or in traits, such as carcass traits, which cannot be measured until the animal has been sacrificed making it unavailable for breeding.

The growing need to use genetic markers for selecting animals for breeding has resulted in a surge in high throughput genomic techniques and the generation of large amounts of data on DNA polymorphisms and potential candidate genes (Reverter and Fortes 2012). However, despite the increase in DNA markers and candidate genes associated with several economically important traits, only a small proportion of the phenotypic variation in quantitative traits can be explained

by the genetic markers (Reverter and Fortes 2012). The amount of variation explained by DNA markers is even lower when the effects of the same markers are assessed in unrelated populations. This can be partly attributed to the nature of association between the DNA markers and the trait, which may either be statistical or functional. Statistical association is based on linkage disequilibrium (LD) between the marker and the causal DNA variant (Snelling et al. 2012). Statistical association indicates that the marker associated with the variation in the trait of interest may not be causing the phenotypic variation; however, it is inherited together with the unknown functional DNA variant more frequently than by random chance, that is; the marker is in LD with the functional DNA variant. Statistical associations are relatively more difficult to replicate across different populations (Snelling et al. 2012) possibly due to decay of LD resulting in different LD patterns between different populations especially populations separated by geographical isolation (Farnir et al. 2000). Functional association analysis, on the other hand, utilizes DNA markers with functional relevance to the biological mechanisms regulating a trait. As a result, these markers may be associated with a larger effect on the variation of the corresponding trait and because they do not rely necessarily on LD, they are expected to be reproducible across diverse populations offering more accurate and reproducible predictions of the levels of the trait even in untested populations (Snelling et al. 2012). For example, a functional variant observed in the MC4R gene has been shown to have consistent effects on pig growth and fat deposition traits across populations and

environments (Kim et al., 2000, Kim et al. 2004, Kim et al. 2006, Szyndler-Nedza et al. 2010)

This missing heritability has been attributed to several factors reviewed in detail by Maher (2008) and Hill (2010). These factors include:

- Statistical errors arising from the models including markers that are identified to be significantly associated with a trait when they are in fact false positives.

- The contribution of many loci to the trait's variation (Wang et al., 2012b). This is the infinitesimal model first described by Fisher (1918) indicating that quantitative traits are regulated by possibly an infinite number of genes each with a small effect on the trait (Bulmer, 1971). This concept was also illustrated by Nagamine et al. (2012) who developed an analytical approach to detect regions containing multiple alleles that individually contribute too little variance to be detectable by genome-wide association studies.

- Transient epigenetic effects could contribute to heritability estimates from close relatives (Slatkin, 2009).

- Interaction between DNA variants with differing effects on the same trait that may even cause diluting effects to each other. This is illustrated in animals by PRKAG3 (Ciobanu et al. 2004) and MC4R (Fan et al. 2009).

- Interaction between different DNA variants with differing effects on different traits (pleiotropy) though these were suggested to be rare (Wagner et al., 2008).

- Interaction between alleles at the same loci resulting to varying degrees of dominance effects which cannot be detected in the heritability of a trait.

In this review, we discuss some of the approaches used to identify DNA markers associated with economically important traits. We also explore the use of biological networks in analyzing the interactions among candidate genes, DNA polymorphisms, metabolites and understanding the biological pathways involved in regulation of gene expression for productivity traits.

### 3.2. ASSOCIATION ANALYSIS

In biological sciences, association analysis refers to the process of finding variations in different boundaries of cellular processes or molecules and assessing their correlation with the variation in a trait. These boundaries are referred to as levels of constraints or control constraints (Strohman 2002) outlined in Figure 2.1. The molecules may be at the DNA level (marker genotypes) or at the level of metabolites (such as hormones and enzymes). The primary objective is association analysis is to develop markers (genetic, metabolic etc) that could be used to accurately predict the level of a trait without necessarily obtaining the phenotypic data on the trait.

There are two major approaches used to perform association analysis in animal sciences; the quantitative trail locus/loci (QTL) mapping approach and the candidate gene approach.

### 3.2.1. The QTL mapping approach

QTL mapping describes the approach used to identify chromosome regions associated with variation in quantitative traits. This approach assumes the actual genes and functional DNA variants associated with the trait are unknown and instead it identifies DNA markers that are in LD with the causative DNA variants (Hayes, 2007).

QTL mapping can be classified as either family based (classical) or population based (association analysis). Classical QTL mapping relies on experimental populations of animals that are developed by breeding specific individuals or lines to maximize the LD in families. Family-wise LD decays through recombination after a few generations of random mating therefore it is not long term (Hayes 2007). Alternatively, QTL mapping may be performed using naturally occurring populations and thereby exploit population-wise LD as shown in Evans et al. (2003). This approach differs from classical QTL mapping because population-wise LD has undergone several recombination events resulting from random mating. Population-wise LD persists for a longer period across a larger number of generations than family-wise LD (Hayes 2007). A detailed review of QTL mapping was discussed by Hayes (2007).

### 3.2.2. The Candidate gene approach

The candidate gene approach is a functional association approach, which assumes that a gene whose function is related to the physiology of the trait contains a polymorphism that causes variation in that trait (Hayes 2007). Unlike QTL mapping which identifies markers linked to the gene, the candidate gene approach identifies potential causative gene(s) that contain(s) DNA variant(s) associated with phenotypic variation in the trait of interest (Zhu and Zhao 2007).

The candidate gene approach consists of four strategies;

The *comparative genomics strategy* is used to identify candidate genes in particular species by comparing the structure and functions of candidate genes in other species assuming that genes may be functionally conserved or structurally homologous (Zhu and Zhao, 2007). This approach has been utilized in mouse models to identify putative genes that confer susceptibility to human diseases (Moore, 1999). In addition, the approach was also applied after the identification of the role of CAST gene in meat tenderness in cattle (Schenkel et al., 2006) to identify its role in influencing meat tenderness in other species such as sheep (Knight et al., 2012). The predictions are occasionally inefficient because of the biological differences among species due to the genetic heterogeneity and evolutionary differentiation (Zhu and Zhao, 2007).

The *position-dependent strategy* utilizes physical linkage of genes in a QTL region, (Zhu and Zhao 2007). This strategy aims at known QTL regions with the genes located in their vicinity considered as candidates. Successful applications of

the position-dependent strategy have been reported in previous studies including the association between a missense mutation located in the DGAT1 gene and milk yield and composition in cattle (Grisart et al. 2002) and that in PRKAG3 gene with glycogen content in skeletal muscles in pigs (Milan et al. 2000 and Ciobanu et al. 2001).

In addition, DNA polymorphisms located in the GDF8/MSTN gene were shown to be associated with carcass traits in sheep (Johnson et al. 2005), double muscling and racing performance in dogs (Mosher et al. 2007) and double muscling in cattle (McPherron and Lee 1997). Associations between haplotypes of the IGF1 gene with body size in dogs were reported by Sutter et al. (2007). The difficulty in this strategy is to prioritize the positional candidate genes to identify the genes with more functional relevance to the trait.

The *function-dependent strategy* aims at identifying genes whose biological functions are related to the physiological processes causing variation in a trait. It could use gene expression profiles associated with the trait and/or information from gene knock-out and transgenic animals (Zhu and Zhao, 2007). The function dependent strategy was used by (among others) Rothschild et al. (1996) to identify the association between ESR gene and litter size in pigs.

The last strategy is the *combination strategy* that may involve a combination of two or more of the three aforementioned strategies. For example the position-dependent strategy could be combined with mRNA expression profiles to identify candidate genes associated with a trait as reported by Kelly et al. (2012). The

combination strategy was also used by Liu et al. (2001) to identify genes that confer resistance to Marek's disease among resistant and susceptible chickens, and by Schwerin et al. (2004) to identify functional candidate genes associated with mastitis in dairy cattle. Other studies that utilized the combination strategy include Ciobanu et al. 2001; to identify new alleles in the protein kinase adenosine monophosphate-activated gamma (3)-subunit gene and their association with low glycogen content in pig skeletal muscle resulting in improved meat quality and Ciobanu et al. 2004; to identify alleles in the CAST gene and their association with meat quality in pigs. Some additional mutations identified using the candidate gene approaches were reviewed by Andersson and Georges (2004).

### 3.2.3. Metabolites as intermediate phenotypes in association analysis

If the gene is considered as the start point and the phenotype (trait) is considered as the end point (fig 1), all the parameters that are involved in the development of the end point can be considered as intermediate phenotypes. However, certain conditions, as described by Kronenberg (2012), need to be met before the parameter can be considered as a suitable intermediate phenotype. The most important of these conditions is that the parameter should be as close to the end point as possible. The intermediate phenotype should also considerably decrease the heterogeneity of the end point phenotype, which dramatically increases the power to detect a gene influencing the intermediate phenotype. If the association between a gene and an intermediate phenotype is strong and the gene has a huge effect on the intermediate phenotype then we can be optimistic that the gene would also have a strong effect on the endpoint. If the effect is low, then we will

be warned in advance that the association with the endpoint may be even lower and therefore the gene may be of less relevance (Kronenberg 2012).

Significant associations between genetic and metabolic markers and phenotypes have been reported; Weikard et al. (2010) reported significant association between mutations in the *NCAPG* and *GDF8* genes and average daily gain in two independent populations of cattle. These authors also observed a significant association between the mutation in the NCAPG with plasma levels of carnitine, arginine and total dimethylarginine. These results indicate that the significant metabolites could be utilized as intermediate phenotypes linking average daily gain in cattle and the mutation in NCAPG gene. In a separate study, Pliakogiannis et al. (1993) reported significant association between serum carnitine levels and body weight and serum triglyceride levels in humans. If validated, the metabolites reported in these studies may have the potential to be used as intermediate phenotypes in selection of cattle for average daily gain and prediction of body weight in humans respectively.

Therefore metabolites, as intermediate phenotypes, can be used as biomarkers to predict the levels of certain traits, where measuring the levels of the metabolite offers more convenience than measuring the phenotype. However, for these predictions to be highly accurate there should be high correlation between the phenotype and the levels of the metabolite(s).

### 3.3. BIOLOGICAL NETWORKS

When several parameters such as DNA polymorphisms, genes and metabolites are associated with a trait, the level of the endpoint is no longer the sum of the individual effects; it is also determined by the modes of interactions between the parameters generally called biological networks (Hartwell et al. 1999).

In biological networks a *node* defines the component molecule being analyzed; it could be a gene (gene network), a metabolite (metabolic network), a protein (protein network), RNA (RNA network) or regulatory networks (Barabasi et al. 2011). The number of links to a node is defined as *k* or *node's degree or connectivity* and most nodes have multiple links. The highly connected nodes in a network are called *hubs* consisting of molecules having a major biological role and are expected to be older and more conserved across individuals and species (Barabasi et al. 2011). Molecules located at the periphery of the network (*peripheral nodes)* may play a major role on the specific trait of interest but do not have a big impact on other traits due to their low connectivity. It is hypothesized that mutations or deletions of genes at the hub will cause effects on multiple traits compared to mutations located in genes located away from the hub. Trait modules can be developed from the hypothesis that if a gene is involved in regulating a certain trait, the genes that interact with it will also be involved in that trait and the direction of interaction relates to the direction of regulation (Barabasi et al. 2011). An illustration of the features of biological networks is shown in figure 3.1.

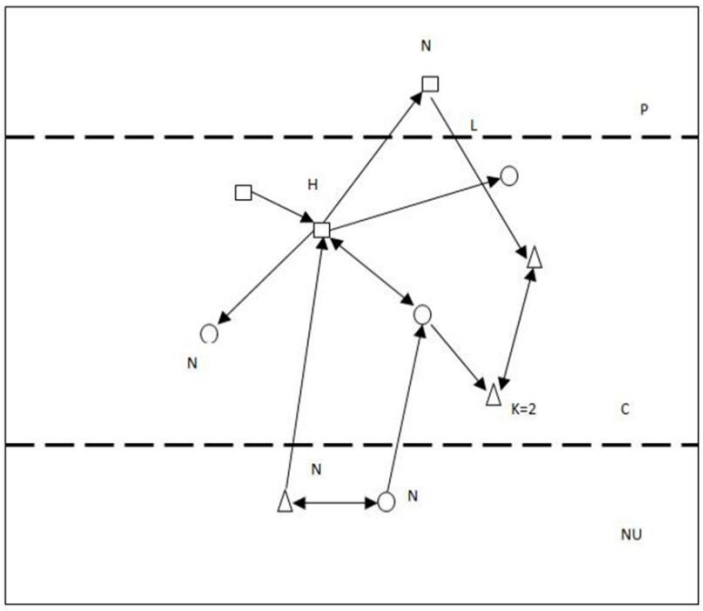



Figure 3.1: General characteristics of a biological network with nodes (N), which may represent a gene or a metabolite in gene and metabolic networks respectively. The locations in the cell where the nodes exert their function are indicated such as plasma membrane (P), cytoplasm (C) or nucleus (NU). Interacting nodes are represented by links (L) whose direction indicates the flow of regulation. The level of connectivity (k) of a node is the number of links to that node and regions with highly connected nodes are called hubs (H). The different shapes of the nodes represent molecules belonging to different functional groups such as enzymes, regulatory proteins or receptors.

There are three major types of biological networks in animal science; gene, metabolic and phenotypic networks

### 3.3.1. Gene and metabolic networks

Gene networks are a reconstruction of the biological processes that result into interactions between multiple genes (Drees et al. 2005). Metabolic networks consist of metabolites as the nodes and biochemical reactions transforming these

metabolites into each other (Pfeiffer et al. 2005). These biochemical processes usually utilize enzymes to convert substrates to products. The product of the reaction becomes the substrate for the subsequent step in the biochemical pathway. Therefore variations in the substrate may cause variation in the levels of products formed, resulting in metabolic interaction. Other metabolites may act as inhibitors or enhancers acting at specific steps along the biochemical pathways and also causing variation in the levels of products formed (Pfeiffer et al. 2005). A metabolic pathway therefore indicates all the possible interactions that can exist between several metabolites and using information on the pathway, one can determine whether the effect is a reduction or an increase in the levels of the product.

### 3.3.2. Phenotypic networks

Phenotypic networks are characterized by phenotypes as nodes and the links represent biological processes and molecular relationships that are common between the interacting traits (Barabasi et al. 2011). Phenotypic networks can therefore be inferred using correlations between individual traits. However, although phenotypic correlation matrices indicate the interactions which can be used to infer the networks, they do not indicate the biological processes that cause the interactions. Phenotypic networks have been used more extensively in human disease studies to develop the 'human diseasome' (Goh et al. 2007). Although no phenotypic network has been constructed for economically important traits in livestock, uncovering links between traits would help to understand how traits that phenotypically appear different may be linked at the molecular level. These

networks may be used to predict the effects of selecting animals for one trait on the other linked traits. This technique therefore offers new approaches to developing complex animal breeding programs and the economic weights used for various traits when developing selection indices.

### 3.3.3. Analysis and manipulation of biological networks

Biological networks can be reconstructed using various data sets including lists of genes or metabolites from association studies, results from gene expression studies or matrices consisting of phenotypic correlations. Programs such as IPA (Ingenuity systems) can reconstruct networks using lists of genes or metabolites while programs such as GenePath (Zupan et al. 2003) infer pathways from gene expression data and Vector PathBlazer (Reshetnikov et al. 2003) reconstructs the networks from a combination of information from different databases.

There are several approaches to the analysis of biological networks including visualization (Junker and Schreiber 2008). Tools such as Pathway Editor (Sorokin 2006) can be used to manually create visualizations while programs such as Pathway Studio (Nikitin et al. 2003), PathwayFinder (Yao et al. 2004) and PubGene (Jenssen et al. 2001) use information in databases to build and create visualization of the pathways. The layout of the biological network is usually automated by the program (Zoubarev 2009). However, the automated layout may lack specific biological information such as sub-cellular localization of the respective gene in the cell (as shown in figure 3.1). In IPA, (and possibly other

programs) the user can customize the information required in the network and its layout.

Using Pathway Studio (Nikitin et al. 2003) the user can interactively expand specific molecules and complexes to show specific regions of the network more clearly. Some programs such as IPA (Ingenuity Systems) assign different shapes for the nodes which represent different molecule classes such that the shape representing an enzyme will be different from the shape representing a receptor.

Comparative analysis of the topology of a biological network can aid in identifying the underlying biological functions associated with the trait. For example if networks were reconstructed for different animal species, the user can use PathBlast (Kelley et al. 2004) to identify network differences and the biological functions associated with them.

Osprey (Breitkreutz et al. 2003) is a program that can superimpose a network on top of another to identify similarities and differences between them (Zoubarev 2009). Biological networks can also be filtered so as to visualize only specific nodes or links (Zoubarev 2009)

In programs such as IPA (Ingenuity Systems) a list of biological pathways will be created with a corresponding significance value (-Log P-value) for each pathway and compared against a threshold to identify the biological pathways significantly associated with the phenotype in study. A detailed analysis of each pathway can be obtained in the canonical pathway analysis section within the program, which

is a database of biochemical pathways, involving genes, proteins or metabolites (Ingenuity Knowledgebase).

By using these analysis tools, the nature of interactions existing between molecules in the nodes of any biological network and the biological processes underlying certain traits can be identified.

### 3.3.4. Examples of biological networks in animal science

The advent of high throughput genotyping and sequencing techniques and the completion of whole genome sequencing in several animal species including cattle (Elsik et al. 2009), pig (Groenen et al. 2012) and chicken (Hillier et al. 2004) have led to a rapid increase in molecular data and the discovery of several genes (Marinus et al. 2011; Reverter and Fortes 2012). This genomic revolution has also provided tools to investigate the interactions among genes and their association with phenotypes of economic importance in livestock production. Once discovered, the genes, networks and biological processes will provide a more complete and accurate understanding of the respective traits and could be used to make predictions of complex traits when phenotypic values are unavailable (Marinus et al. 2011).

#### *a. Cattle*

Productivity in cattle relies on the efficiency of feed utilization (input) and the quality and quantity of production. The economically important traits depend on the production system such as average daily gain and meat quality in beef and pig operations, milk quantity and quality and reproductive traits in dairy operations,

reproductive traits in cow-calf operations and fiber in wool-sheep and alpaca operations. Other traits of importance include resistance or susceptibility to diseases and fertility.

A gene network containing 3,159 genes associated with 22 measures related to puberty, fertility, growth and body composition in beef cattle heifers was reconstructed by Fortes et al. (2010). Among other benefits, this study showed that the gene network approach captured more information than analyses that utilize LD only. This study was followed up with the analysis of gene networks associated with ten growth and fertility traits in Brangus heifers (Fortes et al. 2012). These authors reported the importance of genes involved in biological processes such as axon guidance (a pathway known to influence release of LHRH), regulation of cellular localization, regulation of neurotransmitter secretion and regulation of membrane potential. In addition, the authors identified 5 transcription factors that were located as hubs in the network indicating that their regulatory role may impact the entire network.

Jiang et al. (2009) described a candidate gene association and gene network analysis for 19 traits related to carcass quality and eating quality in Wagyu X Limousin crosses. These traits included phenotypic measurements for carcass weight, carcass rib eye area, subcutaneous fat, pelvic and heart fat and marbling, and taste panel measurements for tenderness, juiciness and flavor. These traits were generally classified into three categories; carcass measurements, eating quality and fatty acid composition. They identified 10 genes associated with carcass measurements, 7 genes associated with eating quality and 5 genes

associated with fatty acid composition. Through gene network analysis, the authors reported that the three classes of phenotypes did not share a lot of gene networks indicating a high degree of genetic independence between them. Therefore, the authors concluded that marker assisted selection to improve one category of these traits would not interfere with the improvement of another category.

Seo and Lewin (2009) set out to create a cattle specific metabolic pathway database using the MetaCyc database (Caspi et al. 2006) and the PathwayTools software (Karp et al. 2002). Using comparative analysis of metabolic pathways, the authors revealed the absence of mammalian genes for 22 metabolic enzymes whose activity was reported in the literature. For example the cattle orthologs of human genes ECGF1, CERK, FAAH2, ALG12 and EARS2 were not identified. This may have resulted from the fact that the generated metabolic network highly depends on the primary genome annotation (Notebaart et al. 2006), which is heavily dependent on sequence homology to human and mouse (Curwen et al. 2004).

Hudson et al (2009) proposed a new algorithm to correctly identify the gene containing causal mutation with microarray data using bovine myostatin mutants. This approach identified the causal mutation by globally contrasting co-expression network dynamics. The authors used the differential wiring method to compare RNA expression levels at several developmental stages and contrasted them between the Wagyu and Piedmontese phenotypic differences. They developed a correlation expression network to identify those nodes in the network

whose links with other nodes changed significantly between the two breeds and termed these as differential wiring. Then using expression levels for the genes and the level of differential wiring, they developed an algorithm that defined a regulatory impact factor such that genes that were highly expressed and had high differential wiring were defined to have a high regulatory impact factor. They further propose that this approach can be applied to other 'omics data because its mathematical approaches mesh well with the known biology of regulatory and non-regulatory molecules. This proposal was tested by Reverter et al. (2010) to identify regulators associated with phenotypic differences in breast cancer and adipocyte differentiation and showed that it appeared universally applicable.

A metabolic network was reconstructed using 1743 metabolites in the mammary gland tissue (Wang et al. 2012a). The authors identified 20 metabolites located in hubs and 11 key enzymes were associated with significant changes in expression during mastitis. Many of the enzymes identified were either involved in amino acid metabolism or had a direct connection to amino acid metabolism.

### *b. Pigs*

Productivity traits in pig production are those related to meat production, litter sizes, feed consumption and disease resistance or susceptibility. There are limited studies that have attempted to use biological networks in pig specific data. Possibly the most relevant study of gene networks in pig productivity traits was an analysis of eQTL performed using whole genome expression microarray using the loin muscle (Steibel et al. 2011). These authors reported 62 unique eQTL and

identified 3 gene networks involved in biological processes related to lipid metabolism, DNA replication and cell cycle regulation. In addition, they identified 2 potential candidate genes; AKR7A2 and TXNDC12 that were part of the gene network associated with lipid metabolism and their location overlapped with QTL for marbling, intramuscular fat and loin muscle area. In a similar study, gene networks were inferred from eQTL using RNA obtained from *Longissimus dorsi* muscle in pigs. The gene networks were inferred from 272 genes having at least one eQTL (Liaubet et al. 2010). Hornshoj et al. (2009) reported a novel study of two porcine tissues based on integrative analysis of data from expression profiling of identical samples using cDNA microarray and iTRAQ based proteomics. They showed that the differences in transcript and protein levels across heart and muscle tissues were positively correlated. These authors did not reconstruct interaction networks but they assessed correlations between protein and transcript levels between microarray and sequencing technologies.

In relation to litter size, differential gene expression analysis was performed using tissues from the ovaries of low and high prolificacy sows during pregnancy (Rodriguez et al. 2011). The sows were categorized in high or low prolificacy depending on their breeding values for prolificacy and 6 sows were selected from each class. The analysis identified 189 differentially expressed genes which were involved in immune system activation, regulation of maternal homeostasis by complement and coagulation cascades and lipid and fatty acid metabolism, which may be involved in steroidogenic pathways. The authors also indicated that 22 of

the differentially expressed genes were located in the same regions with previously reported QTL for litter size traits.

### *c. Poultry*

The completion of the chicken whole genome sequence in 2004 (Wong et al. 2004) enabled investigations into biological interactions (Burt 2005). The chicken genome was also used to create a draft genome assembly for the turkey genome completed in 2010 (Dalloul et al. 2010). To explore biological interactions in chicken, an interactome was created using 8140 genes and established 72000 interactions (Konieczka et al. 2009). This interactome can be used by specific users to extract sub-networks and study specific biological processes.

Several additional studies have aimed at reconstructing biological networks to study specific traits. For example Ciraci et al. (2010) reconstructed a gene network using genes differentially expressed between endotoxin stimulated verses non-stimulated macrophages and showed that endotoxin exposure significantly affected the expression of IL1B, IL6, IL8 and TLR15 and Schokker et al. (2011) reported a gene network reconstructed using gene expression data associated with intestinal salmonellosis in poultry.

## 3.3.5. Application of biological networks in animal breeding

Recent rapid advances in genomic and bioinformatics technologies have allowed researchers in animal sciences to analyze biological networks and, identify and characterize the molecular components of traits, and the variations associated with them (Woelders et al. 2011). However, the quantitative aspects of traits do not

simply arise from the sum of the properties of individual components of the 'trait system' under investigation but depend on dynamic interactions between these components at various biological levels (Woelders et al. 2011). A summary of genetic interactions was discussed by Drees et al., (2005) using a model of two DNA variants/mutations *A* and *B* and included:

*Noninteractive interaction* occurs when *A* has no effect on the trait when *B* is present or *B* has no effect when *A* is present. However, when *A* and *B* have no effect on the trait, but the *AB* combination has an effect, the interaction is termed *Synthetic interaction.*

*Epistatic interaction* occurs when *A* and *B* have different effects (in terms of direction or magnitude) but individuals with both mutations have the same phenotype as having either *A* or *B*. While, when the *A, B,* and the *AB* combination have the same effect the interaction is named *Asynthetic interaction.*

*Conditional interaction* occurs when *A* has an effect only when *B* is present, or the *B* mutant has an effect only when *A* is present. On the contrary, *Suppressive interaction* occurs when *A* has an effect but that effect is abolished by adding the suppressor *B*, which itself shows no single-mutant effect.

*Additive interaction:* single-mutant effects combine to give a double-mutant effect. A s*ingle-nonmonotonic interaction* is when *B* shows opposing effects in the *A* background or, *A* shows opposing effects in the *B* background, but not both and a d*ouble-nonmonotonic interaction* is when both *A* and *B* show opposing effects in the background with the other mutant gene.

Therefore, to understand the effects of biological networks on a trait, it is important to evaluate the mode of interactions and the level of dynamicity of the biological networks.

The rationale of the application of biological networks in animal sciences is the development of predictive models of animal traits that help to understand the biology of traits and that can be applied for the prediction, modulation and improvement of traits (Quackenbush, 2007). We will attempt to clarify this concept as follows.

The knowledge derived from biological networks, if used optimally, can be used to generate accurate and comprehensive predictions of performance characteristics of animals kept under normal or specified conditions (Woelders et al. 2011). This can be exemplified by the already existing application of 'predictive biology' in 'genomic selection', which can be used to predict the breeding value of individual animals in the absence of direct phenotypic measurements. Gene networks in this case would be used to expand the genomic selection models to include the expected interactions between the component markers.

Once the breeding values have been estimated (*more accurately)* the knowledge of genotype–phenotype relationships and biological interactions may be used for the development of precision mating systems that maximize the non-additive variation of traits (Woelders et al. 2011). The mating systems will definitely be superior to the current systems because the current systems rely more on the

additive effects of genes, which does not capture the entire contribution from genetic variations (Hallander and Waldmann 2007).

Another important application of biological networks is the sorting of animals to establish the individuals which are best equipped for optimal performance under defined environmental and management conditions (Woelders et al. 2011).

As a result of the biological networks approach, knowledge on the relationships between molecular composition, biological mechanisms and the behavior of trait systems can developed to create a better understanding of the underlying biological mechanisms associated with the traits and may also aid in the identification of genomic variation causally associated with economically important traits in livestock.

### 3.4. CONCLUSION

Following association analysis, reconstruction of biological networks is a critical part in research in livestock science and several other organisms. Firstly, the reconstruction of biological networks is useful in identification of additional candidate molecules associated with the phenotype under study. The assumption is that once the molecules associated with a trait have been identified, all molecules that interact with the already identified molecule(s) are considered as potentially associated with the trait. Secondly, biological networks can be used to identify the biological processes associated with the trait being studied. The biological processes are used to create a better understanding of the underlying biological mechanisms associated with the trait. Therefore even if a trait is

complex and its underlying biological mechanisms are not well understood, biological networks can be used to identify processes that may potentially control the trait. Lastly, biological networks can be used to assess interactions between molecules such as genes or proteins. Understanding the interactions may be useful in predicting the effects of manipulating one molecule on the molecules that interact with it. The network may be used to predict the impact of manipulating one trait on other traits that share similar interacting genes.

Due to these benefits, and the ever increasing amount of molecular data generated by ‘omics’ studies, we anticipate that biological network reconstruction and analysis will remain one of the superior approaches for analysis.

## 3.5. REFERENCES


Aggrey SE, Karnuah AB, Sebastian B and Anthony NB. (2010). Genetic properties of feed efficiency parameters in meat-type chickens. Genetics Selection Evolution 42:25

Andersson L, Georges M (2004) Domestic-animal genomics: deciphering the genetics of complex traits. Nat Rev Genet 5: 202-212.

Arthur PF, Archer JA, Herd RM (2004) Feed intake and efficiency in beef cattle: Overview of recent Australian research and challenges for future. Aust J Exp Agr 44: 361–369.

Barabási AL, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. Nat Rev Genet 12(1):56-68.

Breitkreutz BJ, Stark C, Tyers M (2003) Osprey: a network visualization system. Genome Biol 4(3): R22. Epub 2003 Feb 27.

Bulmer MG (1971). The effect of selection on genetic variability. Am Nat 105: 201-211.

Burt DW (2005) Chicken genome: Current status and future opportunities. Genome Res. 15: 1692-1698.

Caspi R, Foerster H, Fulcher CA, Hopkinson R, Ingraham J, Kaipa P, Krummenacker M, Paley S, Pick J, Rhee SY, Tissier C, Zhang P, Karp PD (2006) MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucl. Acids Res 34 (Database issue): D511-D516.

Ciobanu D, Bastiaansen J, Malek M, Helm J, Woollard J, Plastow G and Rothschild M. (2001). Evidence for new alleles in the Protein Kinase Adenosine Monophosphate-Activated γ3-Subunit gene associated with low glycogen content in pig skeletal muscle and improved meat quality. Genetics 159(3): 1151-1162.

Ciobanu DC, Bastiaansen JW, Lonergan SM, Thomsen H, Dekkers JC, Plastow GS, Rothschild MF. (2004). New alleles in calpastatin gene are associated with meat quality traits in pigs. J. Anim. Sci. 82(10):2829-39.

Ciraci C, Tuggle CK, Wannemuehler MJ, Nettleton D, Lamont SJ (2010) Unique genome-wide transcriptome profiles of chicken macrophages exposed to Salmonella-derived endotoxin. BMC Genomics 11:545.

Curwen V, Eyras E, Andrews TD, Clarke L, Mongin E, Searle SMJ, Clamp M. (2006). The ensembl automatic gene annotation system. Genome Research 14(5): 942 -950.

Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K. (2010) Multi-Platform Next-Generation Sequencing of the Domestic Turkey (Meleagris gallopavo): Genome Assembly and Analysis. PLoS Biol 8(9): e1000475.

Dekkers JCM, Hospital F (2002) Utilization of molecular genetics in genetic improvement of plants and animals. Nature Reviews Genetics 3: 22-32.

Drees BL, Thorsson VN, Carter GV (2005) Derivation of genetic interaction networks from quantitative phenotype data. Genome Biology 6(4):38–39.

Elsik CG, Tellam RL, Worley KC and The Bovine Genome Sequencing and Analysis Consortium. (2009). The genome sequence of taurine cattle: A window to ruminant biology and evolution. Science 324 (5926): 522-528.

Evans GJ, Giuffra E, Sanchez A, Kerje S, Davalos G, Vidal O, Illán S, Noguera JL, Varona L, Velander I, Southwood OI, de Koning DJ, Haley CS, Plastow GS, Andersson L. (2003). Identification of quantitative trait loci for production traits in commercial pig populations. Genetics. 164(2):621-7.

Fan B, Onteru SK, Plastow GS, Rothschild MF. (2009). Detailed characterization of the porcine MC4R gene in relation to fatness and growth. Animal Genetics 40 (4): 401–409.

Farnir F, Coppieters W, Arranz J, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M, Nezer C, Simon P, Vanmanshoven P, Wagenaar D, Georges M (2000) Extensive Genome-wide Linkage Disequilibrium in Cattle. Genome Res. 10: 220-227.

Fisher RA. (1918). The correlations between relatives on the supposition of Mendelian inheritance. Philosophical Transactions of the Royal Society of Edinburgh. 52:399–433.

Fortes MRS, Reverter A, Zhang Y, Collis E, Nagaraj SH, Jonsson NN, Prayaga KC, Barris W, Hawken RJ (2010) Association weight matrix for the genetic dissection of puberty in beef cattle. Proc Natl Acad Sci USA 107:13642-13647.

Fortes MRS, Snelling WM, Reverter A, Nagaraj SH, Lehnert SA, Hawken RJ, DeAtley KL, Peters SO, Silver GA, Rincon G, Medrano JF, Islas-Trejo A, Thomas MG (2012) Gene network analyses of first service conception in Brangus heifers: Use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors. J Anim Sci 90: 2894-2906.

Goh K, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL (2007) The human disease network. PNAS 104 (21): 8685 – 8690.

Grisart B, Coppieters W. Farnir F (2002) Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res 12: 222-231.

Groenen MAM, Archibald AL and the swine genome sequencing consortium. (2012) Analyses of pig genomes provide insight into porcine demography and evolution. Nature. 491: 393–398.

Hallander J and Waldmann P. (2007). The effect of non-additive genetic interactions on selection in multi-locus genetic models. Heredity 98: 349–359.

Hartwell LH, Hopfiled JJ, Leiber S, Murray AW (1999) From molecular to modular cell biology. Nature 402 (2): C47 – C52.

Hayes B (2007) QTL Mapping, MAS and Genomic Selection. Notes from a short-course organized by Animal Breeding & Genetics Department of Animal Science Iowa State University. June 4-8, 2007.

Herd RM, Archer JA, Arthur PF (2003) Reducing the cost of beef production through genetic improvement of feed intake: Opportunity and challenges to application. J of Anim Sci 81: E9–E17.

Hill WG (2010). Understanding and using quantitative genetic variation. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 365(1537): 73–85.

Hillier LW and International Chicken Genome Sequencing Consortium. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432: 695-716.

Hoque MA, Kadowaki H, Shibata T, Oikawac T and Suzuki K. (2009). Genetic parameters for measures of residual feed intake and growth traits in seven generations of Duroc pigs. Livestock Science. 121: 45–49.

Hornshøj H, Bendixen E, Conley LN, Andersen PK, Hedegaard J, Panitz F, Bendixen C (2009) Transcriptomic and proteomic profiling of two porcine tissues using high-throughput technologies. BMC Genomics 10:30.

Hudson NJ, Reverter A, Dalrymple BP. (2009) A Differential Wiring Analysis of Expression Data Correctly Identifies the Gene Containing the Causal Mutation. PLoS Comput Biol. 5(5): e1000382.

Jenssen TK, Laegreid A, Komorowski J, Hovig E (2001) Literature network of human genes for high-throughput analysis of gene expression. Nature Genetics 28: 21-28.

Jiang Z, Michal JJ, Chen J, Daniels TF, Kunej T, Garcia MD, Gaskins CT, Busboom JR, Alexander LJ, Wright RW Jr, MacNeil MD (2009) Discovery of novel genetic networks associated with 19 economically important traits in beef cattle. Int J Biol Sci. 5(6): 528-542.

Johnson PL, McEwan JC, Dodds KG (2005) A directed search in the region of GDF8 for quantitative trait loci affecting carcass traits in Texel sheep. J Anim Sci 83:1988-2000.

Junker BH, Schreiber F (Eds) (2008) Analysis of biological networks. Wiley series on bioinformatics; computational techniques and engineering. John Wiley and Sons. Hoboken. New Jersey.

Karp PD, Paley S, Romero P (2002) The Pathway Tools software. Bioinformatics 18(suppl_1): S225-S232.

Kelley BP, Yuan B, Lewitter F, Sharan R, Stockwell BR, Ideker T (2004) PathBLAST: a tool for alignment of protein interaction networks. Nucleic Acids Res. 32(Web Server issue):W83-8.

Kelly SA, Nehrenberg DL, Hua K, Garland T Jr, Pomp D. (2012). Functional genomic architecture of predisposition to voluntary exercise in mice: expression QTL in the brain. Genetics. 191(2):643-54.

Kim KS, Larsen N, Short T, Plastow G, Rothschild MF. (2000). A missense variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake traits. Mamm Genome. 11(2):131-5.

Kim KS, Thomsen H, Bastiaansen J, Nguyen NT, Dekkers JC, Plastow GS and Rothschild MF (2004) Investigation of obesity candidate genes on porcine fat deposition quantitative trait loci regions. Obesity Research 12: 1981–94.

Kim KS, Lee JJ, Shin HY, Choi BH, Lee CK, Kim JJ, Cho BW, Kim TH. (2006). Association of melanocortin 4 receptor (MC4R) and high mobility group AT-hook 1 (HMGA1) polymorphisms with pig growth and fat deposition traits. Anim Genet. 37(4):419-21.

Knight MI, Daetwyler HD, Hayes BJ, Hayden MJ, Ball AJ, Pethick DW and McDonagh MB. (2012). Discovery and trait association of single nucleotide polymorphisms from gene regions of influence on meat tenderness and long-chain omega-3 fatty acid content in Australian lamb. Animal Production Science. 52(7): 591-600.

Konieczka JH, Drew K, Pine A, Belasco K, Davey S, Yatskievych TA, Bonneau R, Antin PB (2009) BioNetBuilder 2.0: bringing systems biology to chicken and other model organism. BMC Genomics 10: S6.

Kronenberg F (2012) Genetics Meets Metabolomics: from Experiment to Systems Biology, Chapter 15. Springer Science and Business Media, LLC. 255 - 264

Liaubet L, Villa-Vialaneix N, Gamot A, Rossi F, Cherel P, Sancristobal M (2010) The structure of a gene network reveals 7 biological sub-graphs underlying eQTLs in pig. Proceedings of WCGALP 2010 - World Congress on Genetics Applied to Livestock Production, Leipzig : Allemagne (2010)

Liu HC, Cheng HH, Tirunagaru V (2001) A strategy to identify positional candidate genes conferring Marek's disease resistance by integrating DNA microarrays and genetic mapping. Anim. Genet 32: 351-359.

MacNeil MD, Nugent RA, Snelling WM (1997) Breeding for Profit: an Introduction to Selection Index Concepts. *Range Beef Cow Symposium.* Paper 142.

Maher B (2008). Personal genomes: The case of the missing heritability. Nature 456: 18-21.

Marinus FW, Woelders H, Bannik A (eds) (2011) Systems biology and livestock science. Wiley-Blackwell, UK.

McPherron AC and Lee S. (1997). Double muscling in cattle due to mutations in the myostatin gene. PNAS 94 (23): 12457-12461.

Milan D, Jeon JT, Looft C, Amarger V, Robic A, Thelander M, Rogel-Gaillard C, Paul S, Iannuccelli N, Rask L, Ronne H, Lundström K, Reinsch N, Gellin J, Kalm E, Roy PL, Chardon P, Andersson L. (2000). A mutation in

PRKAG3 associated with excess glycogen content in pig skeletal muscle. Science. 288(5469): 1248-51.

Moore KJ (1999) Utilization of mouse models in the discovery of human disease genes. Drug Discov Today 4: 123-128.

Mosher DS, Quignon P, Bustamante CD, Sutter NB, Mellersh CS, Parker HG and Ostrander EA. (2007). A mutation in the Myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. PLoS Genet 3(5): e79. PMC1877876.

Nagamine Y, Pong-Wong R, Navarro P, Vitart V, Hayward C, Rudan I, Campbell H, Wilson J, Wild S, Hicks AA, Pramstaller PP, Hastie N, Wright AF, Haley CS. (2012). Localising loci underlying complex trait variation using regional genomic relationship mapping. PLoS ONE 7(10): e46501.

Nikitin A, Egorov S, Daraselia N, Mazo I (2003) Pathway studio—the analysis and navigation of molecular networks. Bioinformatics Applications Note 19 (16): 2155–2157.

Notebaart RA, van Enckevort FH, Francke C, Siezen RJ, Teusink B. (2004). Accelerating the reconstruction of genome-scale metabolic networks. BMC Bioinformatics 7:296.

Pfeiffer T, Soyer OS, Bonhoeffer S (2005) The evolution of connectivity in metabolic networks. PLoS Biol 3(7): e228.

Pliakogiannis T, Chatzidimitriou C, Evangeliou A, Böhles HJ Kalaitzidis K (1993) Serum carnitine levels, lipid profile, and metabolic status of patients on continuous ambulatory peritoneal dialysis. Perit Dial Int 13 (2): S440-443.

Quackenbush J (2007). Extracting biology from high-dimensional biological data. Journal of Experimental Biology 210: 1507–1517.

Reshetnikov V, Karpov A, Pot D, Tereshchenko F (2003) Vector PathBlazer: A New Pathway Analysis And Visualization Tool. Proc. ISMB. 2003.

Reverter A, Fortes MRS (2012) Building single nucleotide polymorphism-derived gene regulatory networks: Towards functional genome-wide association studies. J Anim Sci doi:10.2527/jas.2012-5780.

Reverter A, Hudson NJ, Nagaraj SH, Pérez-Enciso M, Dalrymple BP (2010) Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. Bioinformatics. 26(7): 896-904.

Rodriguez AF, Munoz M, Fernandez A, Pena RN, Tomas A, Noguera JL, Ovilo C, Fernandez AI (2011) Differential Gene Expression in Ovaries of Pregnant Pigs with High and Low Prolificacy Levels and Identification of Candidate Genes for Litter Size. Biology Of Reproduction 84: 299–307.

Rothschild M, Jacobson C, Vaske D, Tuggle C, Wang L, Short T, Eckardt G, Sasaki S, Vincent A, McLaren D, Southwood O, Van Der Steen H, Mileham A, Plastow G. (1996). The estrogen receptor locus is associated

with a major gene influencing litter size in pigs. Proc. Natl. Acad. Sci. USA. 93: 201-205.

Schenkel FS, Miller SP, Jiang Z, Mandell IB, Ye X., Li H, Wilton JW. (2006). Association of a single nucleotide polymorphism in the calpastatin gene with carcass and meat quality traits of beef cattle. J Anim Sci. 84(2):291-9.

Schokker D, Koning DJ, Rebel JMJ, Smits MA (2011) Comparative Biochemistry and Physiology Part D: Genomics and Proteomics 6 (4): 339-347

Schwerin M, Kuhn C, Brunner R, Goldammer T, Bennewitz J, Reinsch N, Xu N, Thomsen H, Looft C, Weimann C, Hiendleder S, Erhardt G, Medjugorac I, Forster M, Brenig B, Reinhardt F, Reents R, Russ I, Averdunk G, Blumel J, Kalm E. (2004). Animal Science Papers and Reports 22(1): 95-100.

Seo S, Lewin HA (2009) Reconstruction of metabolic pathways for the cattle genome. BMC Systems Biology 3:33.

Slatkin M (2009). Epigenetic inheritance and the missing heritability problem. Genetics. 182: 845–850.

Snelling WM, Cushman RA, Keele JW, Maltecca C, Thomas MG, Fortes MRS, Reverter A (2012) Networks and pathways to guide genomic selection. J Anim Sci. doi:10.2527/jas.2012-5784.

Sorokin A, Paliy K, Selkov A,Demin OV, Dronov S, Ghazal P, Goryanin I (2006) "The pathway editor: a tool for managing complex biological networks." IBM Journal of Research and Development 50 (6): 561 - 573.

Steibel JP, Bates RO, Rosa GJM, Tempelman RJ, Rilington VD, et al (2011) Genome-Wide Linkage Analysis of Global Gene Expression in Loin Muscle Tissue Identifies Candidate Genes in Pigs. PLoS ONE 6(2): e16766.

Strohman R. (2002). Maneuvering in the complex path from genotype to phenotype. Science 296: 701-703.

Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, Padhukasahasram B, Karlins E, Davis S, Jones PG, Quignon P, Johnson GS, Parker HG, Fretwell N, Mosher DS, Lawler DF, Satyaraj E, Nordborg M, Lark KG, Wayne RK, Ostrander EA (2007) A single IGF1 allele is a major determinant of small size in dogs. Science 316(5821): 112-115.

Szyndler-Nędza M, Tyra M, Blicharski T and Piorkowska K. 2010. Effect of mutation in MC4R gene on carcass quality in Pulawska pig included in conservation breeding programme. Animal Science Papers and Reports 28 (1): 37-45.

Wagner GP, Kenney-Hunt JP, Pavlicev M, Peck JR, Waxman D, Cheverud JM. (2008). Pleiotropic scaling of gene effects and the 'cost of complexity'. Nature. 452 (7186): 470-472.

Wang C, Wang J, Ju Z, Zhai R, Zhou L, Li Q, Li J, Li R, Huang J, Zhong J (2012a) Reconstruction of metabolic network in the bovine mammary gland tissue. Mol Biol Rep 39(7): 7311-8.

Wang Y, Gjuvsland AB, Vik JO, Smith NP, Hunter PJ, et al. (2012b) Parameters in Dynamic Models of Complex Traits are Containers of Missing Heritability. PLoS Comput Biol 8(4): e1002459.

Weikard R, Altmaier E, Suhre K, Weinberger KM, Hammon HM, Albrecht E, Setoguchi K, Takasuga A, Kühn C (2010) Metabolomic profiles indicate distinct physiological pathways affected by two loci with major divergent effect on Bos taurus growth and lipid deposition. Physiol Genomics 42A (2): 79-88

Woelders H, Te Pas MFW, Bannink A, Veerkamp RF, and Smits MA. (2011). Systems biology in animal sciences. Animal 5(7): 1036–1047.

Wong GK, Liu B, Wang J, Zhang Y, Yang X, Zhang Z, Meng Q, Zhou J, Li D, Zhang J, et al. 2004. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. Nature 432: 717-722.

Yao D, Wang J, Lu Y, Noble N, Sun H, Zhu X, lin N, Payan DG, Li M, Qu K (2004) PathwayFinder: Bridging the Way Towards Automatic Pathway Extraction. ACM International Conference Proceeding Series. 2004.

Zhu M, Zhao S (2007) Candidate Gene Identification Approach: Progress and Challenges. Int. J. Biol. Sci, 3(7): 420-427.

Zoubarev A, (2009) Tools for visual analysis of biological networks. www.cs.ubc.ca. (15thFeb2013)

Zupan B, Demsar J, Bratko I, Juvan P, Halter JA, Kuspa A, Shaulsky G (2003) GenePath: a system for automated construction of genetic networks from mutant data. Bioinformatics 19 (3): 383-389.

## *PREFACE TO CHAPTER FOUR*

*I have structured the thesis to reflect the different levels of regulation of quantitative (polygenic) traits illustrated and described in detail in figure 2.1.*

*In Chapter 4, which has been published in Journal of Animal Science, 2013 (doi:10.2527/jas.2012-6170), the candidate gene approach is used to identify candidate genes containing SNPs associated with variation in residual feed intake (RFI) in beef cattle.*

*The null hypothesis: No SNP within the selected candidate genes would be significantly associated with RFI*

*RFI is an economically important trait because feed costs account for over 65% of the total costs in beef production. Selecting cattle for RFI will result in individuals who eat less feed without compromising the level production.*

# CHAPTER FOUR

## Candidate genes and single nucleotide polymorphisms associated with variation in Residual Feed Intake in beef cattle[2]

### 4.1. INTRODUCTION

In beef cattle, residual feed intake (RFI) is defined as the difference between actual and predicted feed intake (Koch et al., 1963). Therefore, variation in RFI may be due to variation in biological processes involved in maintenance of body weight and production (Herd and Arthur, 2009). Although the specific processes are unknown (Crews, 2006), Herd and Arthur (2009) estimated that metabolism would directly account for about 42% of the variation in RFI. Other sources of variation may include body composition (5%), digestion (10%), physical activity (9%), thermoregulation and heat increment of feeding (9%) and the remaining 25% is due to unknown factors (Richardson and Herd, 2004; Herd and Arthur, 2009).

The genetic basis of RFI has been investigated indicating that RFI has a moderate heritability ranging from 0.16 (Herd and Bishop, 2000) to 0.58 (Crews et al., 2003) indicating that a moderate amount of the observed phenotypic variation is regulated by the additive effect of the genes associated with the trait. In addition, several DNA markers associated with RFI in beef cattle have also been reported

[2]A version of this chapter has been published in the Journal of Animal Science, doi:10.2527/jas.2012-6170.

(Barendse et al., 2007; Sherman et al., 2008, 2009; Moore et al., 2009; Bolormaa et al., 2011; Snelling et al., 2011; Elzo et al., 2012) and may be used for marker-assisted selection (MAS) for RFI. However, some of these markers are located in introns including those reported by Sherman et al., (2009) and the basis of the biological effect may not be easily explained. In addition, some marker sets accounted for a small proportion of the variation in RFI (for example 6.9% in Sherman et al., 2008). Therefore, there is need to identify functional markers that will explain a larger proportion of variation in RFI and whose effects will be consistently reproduced across genetically diverse beef cattle populations.

One of the approaches that can be used to identify functional markers is the candidate gene approach. This approach entails the identification of positional candidate genes located within QTL associated with RFI and/or its component traits such as feed intake, average daily gain and body weight. The positional candidate genes are then prioritized according to their functions such that genes whose functions are related to the physiology of RFI are considered more important than genes whose functions are not related to RFI. The physiological processes associated with RFI used for this prioritization were reviewed in detail by Richardson and Herd (2004) and Herd and Arthur (2009). DNA polymorphisms located within the functional candidate genes are then identified, genotyped and analyzed for association with RFI. Because the prioritized genes retain their functional relevance across diverse beef cattle populations, the candidate gene approach is well suited for identifying genes underlying the variation in RFI and is expected to account for more phenotypic variation in RFI

and DNA markers identified are also expected to be consistent and reproducible in genetically diverse beef cattle populations.

In this study, we utilized the candidate gene approach to identify SNPs significantly associated with RFI. We also reconstructed gene networks using IPA (Ingenuity Systems, www.ingenuity.com) to identify biological pathways associated with the significant genes. The significant genes, if validated in other beef cattle populations, may be incorporated into a MAS panel for selecting beef cattle for RFI. The biological processes identified will create a better understanding of the physiological processes underlying the variation in feed efficiency especially RFI.

## 4.2. MATERIALS AND METHODS

Phenotypic and genotypic data were obtained from 531 beef cattle steers at the University of Alberta ranch at Kinsella, Canada. The steers used were sired by Angus, Charolais or University of Alberta hybrid bulls. Dams were produced from crosses among 3 composite cattle lines; Beef Synthetic 1, Beef Synthetic 2, and Dairy × Beef Synthetic. The breed composition of the Beef Synthetic groups was described by Goonewardene et al., (2003). The steers were managed and tested under feedlot conditions using the GrowSafe automated feeding system (GrowSafe Systems Ltd., Airdrie, Alberta, Canada) as described by Nkrumah et al., (2004) and all the animals were managed and cared for according to the guidelines of the Canadian Council on Animal Care (CCAC, 1993).

### 4.2.1. Phenotypic data

Phenotypic data was collected from 531 beef cattle steers born between 2003 and 2006. The collection of feed intake data and calculation of RFI was discussed in detail by Basarab et al., (2003). In summary, the actual feed intake for each steer was measured using the GrowSafe automated feeding system (GrowSafe Systems Ltd., Airdrie, Alberta, Canada). This feed intake was converted to dry matter intake (DMI) by multiplying it by the dry matter content of the diet. DMI was then standardized to 10 MJ of metabolizable energy (ME)/kg of dry matter. The predicted feed intake was estimated based on metabolic mid-weight and average daily gain (ADG). ADG was calculated as the slope from the regression of body weight (BW) on test day. Metabolic mid-weight was obtained as the mid-weight on test to the power of 0.75.

### 4.2.2. Genotypic Data

#### *4.2.2.1. Identification of positional candidate genes*

Two sets of positional candidate genes were utilized in this study. The first set consisted of 1100 candidate genes positioned within a range of 500kbp on either side of 203 QTL reported in the Bovine QTL database (cattleQTLdb) (Hu et al., 2007). These QTL were associated with average daily gain (ADG), feed conversion ratio (FCR), body weight (BW), dry matter intake (DMI), metabolic weight (MW), energy balance and RFI. There was at least one QTL for BW on each chromosome except chromosome 5, 9, 13 and 24. QTL for FCR, DMI and ADG were located in 18, 17 and 16 chromosomes respectively. Only three QTL were associated with RFI and were located on BTA25. The second set of

positional candidate genes consisted of 1018 genes positioned within a range of 500kbp on both sides of 310 SNPs previously reported to be significantly associated with RFI in a genomewide association study (GWAS) using the 50K bovine SNP chip (Mujibi et al., 2011).

A total of 2118 positional candidate genes were used in this study.

***SNP Detection***

The SNPs located in the positional candidate genes were identified from the NCBI SNP database (dbSNP) (Sherry et al., 2001) and by comparing cDNA sequences generated from liver samples from steers at the University of Alberta ranch at Kinsella, Canada with reference sequences from Ensembl version 57 (Hubbard et al., 2009).

To generate the cDNA library, RNA was prepared from pooled liver samples using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) using the protocol described in the TRIzol reagent user manual (Conolly et al., 2006). The quality and quantity of RNA was determined using a NanoDrop 2000 spectrophotometer (Nanodrop technologies, USA) (Gallagher and Desjardins 2007). The liver cDNA library was constructed according to the TruSeq RNA and DNA sample preparation kit (Illumina, USA, v2 published by Illumina Inc. 2011, publication number 970-2009-039) and cDNA sequencing was performed on the Genome Analyzer II using the TruSeq RNA and DNA sample preparation kit (Illumina, USA, v2 published by Illumina inc 2011, publication number 970-2009-039).

To generate a list of putative SNPs, Maq (version 0.7.1) (Li et al., 2008) was used to map DNA sequencing reads and the bovine transcript sequences from Ensembl version 57 (Hubbard et al., 2009) were used as reference sequences. The SNPs returned by Maq's SNPfilter command were annotated using NGS-SNP (Grant et al., 2011) by assigning a functional class to each SNP (e.g. nonsynonymous) and then providing Ensembl identifiers and gene ontology (GO) terms for the affected genes, when applicable. Only the non-synonymous SNPs were considered functional candidate SNPs. Functional candidate SNPs were then filtered by discarding those with a minor allele frequency less than 10%.

The positional candidate genes containing functional SNPs were then prioritized according to their functions in relation to the physiology of feed efficiency as reviewed by Herd and Arthur (2009). Positional candidate genes with functions related to metabolism (catabolism and anabolism), carbohydrate, lipid and protein metabolism, body temperature regulation, muscle activity and immunity (Herd and Arthur, 2009) were considered functional candidate genes and a final set of 116 genes was obtained. The non-synonymous SNPs from each gene were prioritized based on the expected effect of the amino acid change such that among several SNPs, a non-conservative amino acid change was considered of higher priority than a conservative amino acid change in the same gene. A total of 117 SNPs were selected within the 116 genes; one SNP from each gene and 2 SNPs from the CAST gene, and genotyped at GeneSeek using DNA samples obtained from steers at the University of Alberta ranch at Kinsella, Canada.

### 4.2.3. Association analysis

The total number of steers that had both genotypic and corresponding phenotypic data was 531. This population was split into a discovery (n = 2/3*531) and a validation population subset (n =1/3*531) and ensured there were no sires common to both sub-populations. The steers in the discovery population were born in 2003 and 2004 and those in the validation population were born in 2005 and 2006. This criterion of splitting the population ensured there was temporal separation between the discovery and the validation and that the subpopulations had experienced different environmental conditions during their growing phase. Considering these factors together, the discovery population consisted of 382 steers and the validation population consisted of 149 steers.

Of the 117 functional SNPs, 113 were successfully genotyped and 39 SNPs were monomorphic in the population leaving 74 polymorphic SNPs to be used in the analysis.

The multiple marker association analysis was performed using the animal model:

$$Y_{ij} = \mu + X_{1i}\beta + \sum_{j=1}^{74} X_{2j} g_j + Za + e \qquad \text{(Equation 1)}$$

$Y_{ij}$ represents the RFI value for animal *i* with *j* SNPs

$\mu$ is the population mean,

$\beta$ is the vector of fixed effects associated with animal *i*, in this case the breed of the sire and the batch.

$\sum_{j=1}^{74} X_{2j} g_j$ is the sum of the SNP effects (g), $X_2 j$ is a design matrix relating an observation ($y$) to one of the genotypes 0, 1 or 2 at the $j$th SNP.

$a$ is a vector of random additive effects (exclusive of SNP g effect(s) in the model) constructed using each animal's pedigree information. It was assumed to be normally distributed with a mean of 0 and variance of $A\sigma^2_{a,}$ where A represented the additive relationship matrix

$e$ was the vector of random residual effects assumed to be normally distributed with a mean of 0 and a variance of $I\sigma^2$ where I was an identity matrix

The heritability was calculated as $h^2 = (\sigma^2_a + \sigma^2_g)/ (\sigma^2_a + \sigma^2_g + \sigma^2)$ where $(\sigma^2_a)$ was the additive variance, $(\sigma^2_g)$ was the variance explained by the SNPs and $(\sigma^2)$ was the residual variance.

In each association analysis, The variation accounted for by the significant SNPs was estimated as a contrast between the residual variance in a reduced model ($Y_{ij} = X_{1i}\beta + Za + e$) and the full model (Equation 1) as described by Yang et al., (2011).

The initial association analysis was performed on the training population using ASReml® 3 software (Gilmour et al., 2009). The three genotypes were coded 0, 1 and 2 such that all heterozygous genotypes were coded 1 and the rare homozygote was coded 0. Missing values were coded 9 and were included in the analysis.

The association analysis was then repeated in the validation population using only the SNPs that were significant in the discovery population.

In addition, to increase the accuracy of the estimate of marker effects, the discovery population and the validation population were combined into a pooled population (n = 531 steers) and association analysis was performed using the same model (equation 1) to estimate the effect of each genotype on RFI.

Genotypic effects were estimated in ASReml 3 and represented as $Y_0$, $Y_1$ and $Y_2$ corresponding to the effect for genotype 0, 1 and 2 respectively. The additive effect (A) for each SNP was calculated using the formula; $A = 0.5(Y_0 - Y_2)$ and dominance effect (D) was calculated using the formula $D = Y_1 - [0.5(Y_0 + Y_2)]$. Contrasts between marker effects for each genotype were estimated using a t-test and the significance was determined as the probability that the effects for the two genotypes were equal; ($Y_0 = Y_1$), ($Y_1 = Y_2$) or the difference between the two genotypic effects is 0; ($Y_0 - Y_1 = 0$), ($Y_1 - Y_2 = 0$).

Linkage disequilibrium (LD) analysis between the significant SNPs was performed using Golden Helix® SNP and Variation suite v7.x (Golden Helix, Bozeman, MT). An $R^2$ between two SNPs would be considered significant LD if it was > 0.5.

### 4.2.4. Assessing the effect of SNPs on protein structure and function

The SIFT (Sorting Intolerant From Tolerant) program (Ng and Henikoff, 2003) was used to predict whether the amino acid substitution from the significant SNPs significantly affected the function of the proteins. Given a protein sequence, SIFT chooses related proteins and obtains an alignment of these protein sequences with

the query sequence. Based on the amino acids appearing at each position in the alignment, SIFT calculates a score, which is the normalized probability that the amino acid change is tolerated and scores less than 0.05 are predicted as deleterious (Ng and Henikoff, 2003).

The effect of the SNPs on the tertiary structure of the respective proteins was assessed using SWISSModel® protein modeling software (Arnold et al., 2006). Models of the tertiary structures of proteins with SNPs were then compared with models from the corresponding reference proteins using the DaliLite® program (Liisa and Jong, 2000).

### 4.2.5. Reconstructing gene networks

Gene interaction networks were reconstructed from the significant genes using IPA software (Ingenuity Systems, www.ingenuity.com). The same program was used to identify biological processes and canonical pathways associated with the significant genes. To reconstruct the gene network, a list of significant genes and their corresponding p-values were imported into the IPA software and the parameters were set to allow the network to include indirect relationships between the imported genes and genes in the IPA knowledgebase. Indirect relationships would assist in the identification of other genes that were not among the genes analyzed but may be associated with RFI. The IPA algorithm generates gene networks by mapping each gene identifier to its corresponding gene in the IPA Knowledge Base (Calvano et al., 2005). The genes are then overlaid onto a global molecular network developed from information contained in the Knowledge Base.

The networks are generated based on their connectivity such that each network has a maximum of 35 imported genes. Each network is assigned a significance score, which represents the likelihood that the imported genes within the network are found therein by random chance (Calvano et al., 2005). A high number of imported genes within a dataset lead to a higher network score. The network score is calculated as the negative of the exponent of the *P* value such that a score of 25 will be equal to a *P*-value of $10^{-25}$ (Calvano et al., 2005). And therefore larger scores correspond to high significance.

## 4.3. RESULTS AND DISCUSSION

### 4.3.1. Analysis of phenotypes

The RFI values ranged from -2.34 to +2.44 $kgday^{-1}$ and as for other biological traits, RFI was normally distributed in the test animals and had a mean of -0.0061 $kgday^{-1}$, a phenotypic variance of 0.64 $(kg/day)^2$ and a standard deviation of 0.80$kgday^{-1}$. The descriptive statistics for the other phenotypes are shown in Table 4.1.

Phenotypic correlations indicated that RFI was independent of ADG but was significantly ($P < 0.0001$) correlated with feed intake and dry matter intake as shown in Table 4.2. FCR was significantly ($P < 0.0001$) correlated with RFI, ADG and DMI. These correlations were consistent with those reported by Sherman et al., (2008).

### 4.3.2. Association analysis

*Discovery population:* Association analysis of the discovery population indicated that 22 SNPs in 21 candidate genes were significantly associated ($P \leq 0.05$) with RFI and accounted for 18.1% of the phenotypic variation. These SNPs were located in chromosome 1, 2, 4, 7, 15, 18, 20 and 29 and the number of SNPs in each chromosome was 4, 6, 1, 2, 2, 1, 5 and 1 respectively. Two SNPs were located in the calpastatin gene.

*Validation population:* The association analysis to validate the 22 significant SNPs was done using the validation population consisting of 149 steers. Seven (of the 22) SNPs were also significantly associated with RFI ($P < 0.05$) in the validation population. Three SNPs are located on chromosome 20 and the other SNPs are located on chromosome 1, 2, 15 and 29. The small number of significant SNPs identified in the validation population may have resulted from the relatively small number of samples when the population was split into a discovery and validation subsets therefore reducing the power and increasing the false negatives.

*Pooled population:* The association analysis in the pooled population indicated that 25 SNPs in 24 candidate genes were significantly associated with variation in RFI ($P \leq 0.05$) accounting for 19.7% of the phenotypic variation in RFI. A summary of the significant SNPs is presented in Table 4.3. These SNPs were located on chromosomes 1, 2, 4, 7, 15, 18, 20 and 29 and the number of SNPs on

each chromosome was 4, 7, 1, 2, 2, 1, 7 and 1, respectively, indicating that the majority of the significant SNPs were located on BTA 1, 2 and 20.

BTA1 contains previously reported QTL for average daily gain and carcass weight (Komatsu et al., 2011), body length, hip height and heart girth (Liu et al., 2010) and post natal growth, carcass rib eye area and calving ease (McClure et al., 2010).

A QTL on BTA2 was previously reported to be associated with average daily feed intake, average daily body weight gain and dressed carcass bone percentage in beef cattle (Martinez et al., 2010). Other QTL on this chromosome were associated with body weight (Kneeland et al., 2004) and body conformation score in beef cattle (Allais et al., 2010).

BTA20 has been reported to contain several QTL related to body weight and average daily gain in cattle. SNPs located in the prolactin receptor (PRLR) were significantly associated with body weight, average daily gain, body height, body length and heart girth in beef cattle (Lu et al., 2011) and rib eye muscle area and calving ease (McClure et al., 2010).

There was significant LD between the two SNPs on chromosome 7 with an $R^2$ value of 0.6 and both were located in the Calpastatin (CAST) gene. The rest of the SNPs had a non-significant $R^2 \leq 0.2$.

The effects associated with each genotype are shown in Table 4.4 as $Y_0$, $Y_1$ and $Y_2$ corresponding to genotype 0, 1 and 2 respectively. The genes with genotypes whose effects were additive included NECAP2, OCLN and OSMR (Figure 4.1a).

Other SNPs showed a dominance effect such as UMPS and CAST (1) (Figure 4.1b). An over-dominant effect was observed in the genotypes for GHR, ACAD11, UGT3A1 and CAST (2) (Figure 4.1c). However, some SNPs had a relatively low number of observations (<20) for at least one of the genotypes (Table 4.4) making the estimates potentially unreliable and may require re-estimation using larger sample sizes.

### 4.3.3. SNPs predicted to cause significant effect on protein structure and/or function

Of the 25 significant SNPs, 3 SNPs were predicted to cause a significant effect on the function of their respective protein using SIFT. These included the Cytochrome P450 subfamily 2B (CYP2B), the Low-density lipoprotein receptor related protein 5 (LRP5) and the Growth hormone receptor precursor (GHR) gene. The SNP in the GHR gene was also predicted to cause a significant effect on the tertiary structure of the protein.

#### *a. Growth hormone receptor precursor (GHR)*

The SNP located in the GHR gene was significantly associated with RFI (P = 0.026). This gene is located in BTA20 within QTL for body weight and energy balance. We did not identify any additional genes with functional relevance to RFI in this region.

The SNP in the GHR was predicted to cause significant effect on both the tertiary structure and the function of the GHR protein.

The genotype effects associated with this SNP were over-dominant with the heterozygous genotype associated with an effect of -135.3 g/day. This effect was 226.4 and 179.5 g/day less than the effects associated with the two homozygous genotypes 0 and 2 respectively; therefore it was more preferred for RFI than the two homozygous genotypes. Combining the observed genotypic effects and the effect of the SNP on the structure and function of the protein, this may be a true overdominant effect. However, the homozygous genotype consisting of the low frequency allele (0) consisted of only 65 individuals so the genotype effects should be confirmed in a larger dataset than used here.

A gene network analysis indicated close interactions between the GHR gene and UBC, LIFR, JAK, OSMR, ERK1/2, CLCF1, STAT5 and STAT 1/3/5 genes as shown in Fig 4.2.

The GHR gene encodes a receptor for growth hormone, which is involved in regulating body growth. Binding of growth hormone to the receptor leads to dimerization of the receptor and the activation of the signal transduction pathway, including the JAK2/STAT5 pathway. The JAK-STAT system consists of three main components: a receptor, Janus kinase (JAK), a signal transducer and activator of transcription (STAT). This system transmits chemical signals from outside the cell, through the cell membrane, into gene promoters on the DNA and cause DNA transcription activity (Aaronson and Horvath, 2002). Some of the genes influenced by the STAT activator of transcription are IGF1, IGF2, IGFBP3 and ALS. The IGF genes are involved in the regulation of growth and cellular

anabolism and have already been reported to be associated with feed efficiency (Bishop et al., 1989; Stick et al., 1998).

A similar association of the GHR gene and RFI was reported in a previous study by Sherman et al., (2008). The SNPs reported by Sherman et al., (2008) were located in the introns and the promoter region. Whilst the SNP in the promoter region may have an effect on gene expression, the SNP in this study was located in an exon and as predicted may have an effect on protein structure and function.

Other studies have reported significant associations between GHR and productivity traits in cattle such as growth and carcass traits in Zebu and cross bred cattle (Curi et al., 2006) and, carcass weight, percentage of valuable cuts, average daily gain and feed conversion ratio (Maj et al., 2004).

***b. The Low density lipoprotein receptor related protein fragment 5 (LRP5)***

The SNP located in the LRP5 gene was significantly associated with RFI (P = 0.018). This gene is located in a body weight QTL on chromosome 29 close to other possible candidate genes for RFI such as Pyruvate carboxylase (PC), NADH dehydrogenase flavoprotein 1 (NDUFV1) and Aldehyde dehydrogenase 3B1 (ALDH3B1). These genes were not included in this study.

This SNP showed a significant (P < 0.05) additive effect and the homozygous genotype 2 was associated with the favorable effect on RFI at -147 g/day. However, in this case there were only 3 and 27 observations for genotypes 0 and 1 respectively; therefore these estimates may be unreliable.

The gene network analysis showed interactions between this gene and the UBC and ERK1/2 genes. An analysis of ERK1/2 indicated that the signal that stimulates the ERK1/2 then stimulates CEBPA. CEBPA (CCAAT/enhancer binding protein) is a transcription factor protein that binds to certain promoters and enhancers and modulates the expression of these genes. The most important genes modulated by this protein are leptin and growth hormone. Several studies have reported the significant association between Leptin and RFI (Hoque et al., 2009; Kelly et al., 2009). The biological processes associated with LRP5 included energy, lipid and cholesterol metabolism (Fujino et al., 2003).

The SNP resulted in an amino acid change from Methionine to Valine. A comparative analysis of the homologs of this gene in ten species was performed in NCBI and indicated that 6 of the 10 species had the amino acid Valine at that position, the mouse (*Mus musculus*) and the rat (*Rattus norvegicus)* had the amino acid Methionine. Other species that differed in their amino acid composition at that position were *Drosophila melanogaster* (Leucine) and the mosquito (*A. gambiae*) with Serine.

The amino acids Met and Val are both non polar amino acids indicating that their chemical properties are similar and may explain why the effect of the amino acid change on the tertiary structure of the protein was not significant.

The LRP5 gene is associated with biological processes that affect bone density in humans (Ashburner et al., 2000). The effect of this gene on bone density may indicate an effect on body weight and as a result an effect on RFI. Other gene

ontology terms associated with this gene are adipose tissue development, cholesterol homeostasis and glucose catabolism (Ashburner et al., 2000). Adipose tissue development influences body composition, a trait that was shown to be associated with RFI (Basarab et al., 2003). Both glucose and cholesterol metabolism fall in the broad category of metabolism, which encompasses all catabolic and anabolic processes whose importance in the regulation of RFI was reviewed in detail by Herd and Arthur (2009).

### *c. Cytochrome P450 subfamily 2B (CYP2B)*

The SNP in the CYP2B gene was significantly associated with RFI (P = 0.014) and was predicted to cause a significant effect on protein function. The homozygous genotype 0 was associated with low RFI with a predicted effect of -75 g/day. However, only 6 steers in the population had this genotype so it may not be accurate and should be regarded with caution. There were 162 heterozygous genotypes with a predicted effect on RFI of 69.4g/day compared to 6.36g/day for the other homozygous genotype (n = 354). The effects associated with genotypes 1 and 2 were not significantly different (P = 0.197). However, both effects were significantly different from the effect of the homozygous genotype 0. The effect of the heterozygous genotype 1 was higher than both homozygous genotypes indicating a possible over-dominant effect where individuals with the heterozygous genotype have the highest RFI.

CYP2B is located on chromosome 18 in a QTL for FCR and close to another cytochrome gene the Cytochrome P450 family 2, subfamily S polypeptide 1

(CYP2S1). The members of subfamily 2B of the cytochrome P450 genes are associated with metabolism of cholesterol, steroid hormones and other lipids. Some of the gene ontology terms associated with CYP2B were electron transport, oxidation-reduction activity and heme binding (Ashburner et al., 2000).

### 4.3.4. Genes with significant additive effect

Some alleles showed additive effects on RFI as shown in figure 4.1a. For example, NECAP2 gene had significant additive effect and the preferred genotype for RFI was the homozygous genotype 2 at -128.1 g/day. Although the dominance effect was not significant (P = 0.12), the three genotypes showed a trend of partial dominance (Table 4.4 and Fig 4.1a). The gene also had an intermediate frequency resulting in relatively good estimates of the allele effects. Other genes in this class include OCLN and OSMR (Figure 4.1a).

### 4.3.5. Genes with significant dominance effects

Two SNPs located in UMPS and CAST(1) genes showed significant dominance effects (Figure 4.1b) and other genes, ACAD11, CAST(2), GHR and UGT3A1 showed an over-dominant effect (Figure 4.1c). Both SNPs located in the CAST gene (CAST1 and CAST2) had significant dominance effects and CAST (1) also showed significant additive effects.

Significant associations between the CAST gene and RFI were previously observed in steers (McDonagh et al., 2001) and pigs (Gandolfi et al., 2011). CAST has also been associated with meat quality (Morgan et al., 1993; McDonagh, 1998; Barendse, 2002; Casas et al., 2006). These associations could

partially explain the correlation between residual feed intake and meat quality traits such as meat tenderness (McDonagh et al., 2001) and the previously reported negative phenotypic correlation (-0.33) between RFI and Warner-Bratzler shear force (Ahola et al., 2007). However, further studies need to be conducted to assess the association between the two specific SNPs analyzed in this study and meat tenderness in beef cattle.

### 4.3.6. Significant genes in the validation population

The 7 genes that were significant in the validation were GHR and LRP5 (see previous section), and UGT3A1, ASNSD1, UBA5, APIP and LIFR. The variation explained by each validated SNP was estimated by performing a single marker association analysis in ASReml 3 with the results as follows; GHR (2.1%), LRP5 (1.9%), UGT3A1 (2.0%), ASNSD1 (2.5%), UBA5 (1.5%), APIP (2.4%) and LIFR (2.3%). These results further reinforce previous observations that quantitative traits are regulated by several genes each with a small effect (Hayes and Goddard, 2001). However, the contributions from each of the significant genes in this study were higher than those reported in previous studies such as Sherman et al., (2009). This may have resulted from one or both of two possible factors reported; firstly, the estimates in this study may have been inflated due to the low numbers of individuals in some of the genotypes. Secondly, the high estimates may have resulted from the candidate gene approach which may have resulted in the identification of markers in genes that cause a larger effect on the trait.

### 4.3.7. Analysis of gene networks and biological processes

The gene interaction network reconstructed included only 13 genes out of the 24 genes found to be significantly associated with RFI in this study (Figure 4.2). The additional genes which had not been analyzed in this study but were present in the network are discussed as potential candidate genes for RFI.

We identified 3 major hubs in the gene network; the first hub was associated with the Ubiquitin C (UBC) gene, the second hub was associated with the insulin induced gene 1 (INSIG1) and a final minor hub was associated with the leukemia inhibitory factor receptor (LIFR) gene.

The hub associated with the UBC gene had multiple interactions with other genes including PLEKHA7, PARP14, SMARCAL1, UBA5, LRP5, CAST and INSIG1 (Fig 4.2). UBC also interacted with ERK1/2 through LRP5, UBA5, GHR, OSMR or LIFR. The UBC hub represented biological processes related to the regulation of phenotypic expression by processes such as endoplasmic reticulum associated degradation of proteins (ERAD), lysosomal degradation, protein degradation via the proteasome, activation of transcription factor NF-kappa-B, cell signaling and DNA repair. This hub therefore, among other processes, indicates the role played by processes that regulate protein function through protein degradation and since most of the proteins associated with the UBC hub were enzymes (figure 4.2 legend) the UBC hub seems to be involved in regulation of enzyme activity through protein degradation.

The INSIG1 gene hub had interactions with other molecules including Lysophosphatidylcholine acyltransferase 3 (LPCAT3), Acyl-CoA synthetase short chain family member 2 (ACSS2), ELOVL fatty acid elongase, Acetoacetyl CoA synthetase (AACS), endoplasmic reticulum lipid raft associated protein (ERLIN2), 3-hydroxyl-3-methylglutaryl-CoA synthase 2 (HMGCS2) and STAR related lipid transfer. This hub represents biological processes related to energy, lipid and steroid metabolism. The importance of lipid and steroid metabolism for RFI was previously reported by Naik et al., (2007) and Richardson et al., (2004).

The LIFR hub (which also interacts with the UBC hub) had links with Janus kinase (JAK) gene and (Glycoprotein 130) GP130. LIFR also interacted with CLCF1, ERK1/2, STAT5, GHR and OSMR. This hub represents biological processes such as growth, cytokine mediated signaling and immune responses. ERK1 is a protein kinase involved in cell growth, adhesion, survival and differentiation by regulating transcription, translation and cytoskeleton rearrangements. It also causes phosphorylation of several transcription factors resulting in regulation of meiosis and mitosis and post mitotic functions of cells. ERK1 is also involved in lysosome processing and endosome cycling (Lancet et al., 2008). ERK1/2 interacted with the UBC hub through UBA5, LRP5, OSMR and GHR and this may represent the relationship in the biological processes associated with both hubs. LIFR combines with GP130, to form a receptor complex while JAK is a non-receptor kinase that transduces the cytokine-mediated signal via the JAK-STAT pathway. Janus kinases are predominantly expressed in immune cells and responsible for signal transduction in response to

activation via tyrosine phosphorylation by interleukin receptors (Lancet et al., 2008). Immune and stress responses were previously reported to be associated with RFI (Richardson et al., 2004) though the results have been inconsistent across different studies.

In summary, the genes and biological processes identified in the candidate gene association analysis and gene network analysis included genes with functions which impact RFI by regulating processes at the genomic level such as regulation of DNA replication, repair and recombination, transcription and translation, and processes related to the metabolism of carbohydrates, lipids, proteins, cholesterol, vitamins, minerals and nucleic acids. In addition, some of the significant processes were involved in immunity such as antigen presentation, immune cell trafficking and inflammatory response.

These processes are similar to those observed by Richardson et al., (2004) and go further to show support for some of the SNPs identified in this study being associated with this variation. However, although the biological processes identified to be associated with variation in RFI have been consistent across several studies (Rolf et al., 2011), there are still several studies reporting different significant genes. For example, from this study, the only genes that were reported in previous studies were; GHR (Sherman et al., 2008; Chen et al., 2012) and CAST (McDonagh et al., 2001).

## 4.4. CONCLUSION

This study reports 22 SNPs associated with variation in RFI in a discovery population and 7 of these SNPs were also significant in a validation population. These validation results indicate that these markers may be useful for marker-assisted selection for RFI in unrelated populations. The effects of the 7 markers were re-estimated using all of the data and they were again significant explaining approximately 15.2% of the variation in RFI in the population. However, the utility of these genes for selection requires some factors to be considered. Firstly, an independent validation should be done before they are used by the industry. Secondly, some of the SNP genotypes showed over-dominance and this needs to be taken into account in their use, for example, they may be useful in crossbreeding animals to combine the different alleles. Thirdly, some of the genes may have pleiotropic effects such as the two SNPs located in the CAST gene, which may influence meat tenderness. If these findings are further validated then these SNPs can be used to select for RFI with increased accuracy.

**Table 4.1: Descriptive statistics for some of the phenotypes related to RFI**

| | ADG | DMI | FCR | RFI |
|---|---|---|---|---|
| Mean | 1.48 | 10.82 | 7.47 | -0.01 |
| Std dev | 0.27 | 1.43 | 1.29 | 0.80 |
| Variance | 0.07 | 2.05 | 1.67 | 0.64 |
| Minimum | 0.63 | 6.20 | 4.59 | -2.34 |
| Maximum | 2.23 | 14.54 | 14.26 | 2.44 |

**Table 4.2: Phenotypic correlation between economically important traits[1] related to RFI in beef cattle**

| | Pearson correlation coefficients | | | | |
|---|---|---|---|---|---|
| | ADG | Feed intake | DMI | FCR | RFI |
| ADG | 1.00 | 0.56 [2] | 0.57 [2] | -0.60 [2] | 0.01 |
| Feed intake | | 1.00 | 0.99 [2] | 0.29 [2] | 0.53 [2] |
| DMI | | | 1.00 | 0.29 [2] | 0.55 [2] |
| FCR | | | | 1.00 | 0.49 [2] |
| RFI | | | | | 1.00 |

[1]*The traits included; ADG - average daily gain, DMI - dry matter intake, FCR- feed conversion ratio and RFI- residual feed intake.*

[2]*Significant at P < 0.0001*

**Table 4.3: Candidate genes significantly associated with residual feed intake in beef cattle steers**

| Gene name | SNP ID[1] | BTA: Position (bp) | P-value | SNP position | SNP allele | Amino acid change | Gene ontology[2] and function related to RFI |
|---|---|---|---|---|---|---|---|
| PARP14 protein fragment (PARP14) | NF | 1:68144657 | 0.028 | 5002 | G/A | Gly -> Arg | Protein synthesis, NAD+ ADP-ribosyltransferase activity |
| Uridine 5 -monophosphate synthase (UMPS) | rs110953962 | 1:70328819 | <0.001 | 572 | C/T | Arg -> Cys | Fat, protein and uridine biosynthesis. |
| Ubiquitin-like modifier activating enzyme 5 (UBA5) | NF | 1:139111130 | 0.048 | 231 | A/T | Glu -> Val | ATP binding and oxidoreductase activity. |
| Acyl-CoA dehydrogenase family member 11 (ACAD11) | rs208270150 | 1:138045480 | 0.004 | 816 | C/T | Ser -> Leu | FAD binding, acyl-CoA dehydrogenase activity and fat metabolism |
| Bridging integrator 1 (BIN1) | rs210348685 | 2:5642793 | 0.026 | 916 | G/A | Ala -> Thr | Regulation of endocytosis and protein synthesis |
| Asparagine synthetase domain-containing protein 1 (ASNSD1) | NF | 2:6949248 | <0.001 | 499 | G/A | Ala -> Thr | Asparagine biosynthesis, glutathione and protein metabolism |
| MKI67 FHA domain-interacting nucleolar phosphoprotein (MKI67IP) | NF | 2:76998684 | 0.036 | 900 | A/T | Arg -> Stop | Regulation of phosphatase activity, RNA metabolic processes and protein synthesis. |
| Aldehyde oxidase (AOX1) | rs110994776 | 2:89545687 | 0.0001 | 1732 | A/G | Ser -> Gly | NAD binding and electron carrier activity |
| SWI/SNF-related matrix-associated actin-dependent | rs109065702 | 2:105138600 | 0.034 | 265 | T/C | Ser -> Pro | Chromatin modification and regulation of transcription. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| regulator (SMARCAL1) | | | | | | | |
| PQ loop repeat containing 2 (PQLC2) | rs209148339[4] | 2:137970404 | 0.023 | 691 | A/G (C/G)[5] | Arg -> Gly | Protein synthesis |
| Adaptin ear-binding coat-associated protein 2 (NECAP2) | NF | 2:140340622 | 0.013 | 568 | A/G | Thr ->Ala | Endocytosis and protein transport |
| Insulin-induced gene 1 (INSIG1) | rs109314460 | 4:121360120 | 0.006 | 308 | A/G | Ser -> Gly | Lipid, cholesterol and steroid metabolism |
| Calpastatin (CAST) | rs109727850 | 7:97480120 | 0.026 | 271 | G/A | Gly -> Asp | Inhibition of calpain which is involved in pre and post mortem degradation of proteins |
| Calpastatin (CAST) | rs210072660 | 7:97526153 | 0.007 | 672 | A/G | Thr ->Ala | - |
| Pleckstrin homology domain containing, family A member 7 (PLEKHA7) | NF | 15:34411065 | 0.04 | 2490 | C/T | Pro -> Ser | Epithelial cell to cell adhesion and protein synthesis |
| APAF1 interacting protein (APIP) | NF | 15:65118633 | <0.001 | 737 | G/A | Val -> Met | Amino acid biosynthesis and apoptosis. |
| Cytochrome P450 subfamily 2B (CYP2B[3]) | NF | 18:49958396 | 0.014 | 16 | G/A | Val -> Met | Electron carrier activity, heme binding and oxidoreductase activity. |
| Occludin (OCLN) | rs109638814[4] | 20:10849769 | 0.008 | 1051 | T/C (A/G)[5] | Cys -> Arg | Cell to cell junction and protein synthesis |
| Growth hormone receptor | rs209676814 | 20:33897128 | 0.026 | 1643 | G/A | Ala -> Thr | Regulates post natal growth and may act as a reservoir of growth |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| precursor (GHR[3,4]) | | | | | | | hormone in plasma. |
| Oncostatin M receptor (OSMR) | rs41580312 | 20:37772898 | 0.006 | 620 | G/T | Arg -> Met | Regulates the growth and maturation of hepatocytes. |
| Leukemia inhibitory factor receptor (LIFR) | NF | 20:38170739 | 0.005 | 1162 | G/A | Ala -> Thr | Stimulation of cell proliferation, body conformation and growth |
| UDP-glucuronosyltransferase 3A1 Precursor (UGT3A1) | rs133951891 | 20:40434540 | 0.023 | 1597 | T/C | Met -> Thr | Glucuronosyltransferase activity and protein metabolism |
| Solute carrier family 45 member 2 (SLC45A2) | NF | 20:42286376 | <0.0001 | 718 | G/A | Ala -> Thr | Melanin biosynthesis, visual perception. |
| Myosin-X (MYO10) | NF | 20:59933885 | 0.0001 | 2375 | G/T | Arg -> Leu | Binds ATP, actin and calmodulin |
| Low-density lipoprotein receptor-related protein 5 fragment (LRP5[3]) | rs42190891 | 29:47717873 | 0.018 | 3166 | A/G | Met -> Val | Bone and mammary gland development, cholesterol metabolism, apoptosis and regulation of insulin secretion. |

*[1]SNPs with an ID indicated as NF did not match any SNP reported in the SNP database*

*[2]Gene ontology and functions were obtained from Ashburner et al., 2000.*

*[3] SNPs in these genes were predicted to cause a significant effect on the function of the protein.*

*[4] SNP in this gene was predicted to cause a significant effect on the tertiary structure of the protein.*

*[5] The SNP alleles reported in the SNP database were different from the alleles obtained in this study but the amino acid and nucleotide positions were the same. The alleles reported in the database are shown in parenthesis in the column 'SNP allele'.*

**Table 4.4: Genotype effects on RFI in gday$^{-1}$ in beef cattle at the University of Alberta ranch in Kinsella[1]**

| GENES | Predicted effects and (number of observations) for each Genotype (gday$^{-1}$) | | | $Y_0$-$Y_1$ and ($p [Y_0 = Y_1]$) | $Y_1$-$Y_2$ and ($p [Y_1=Y_2]$) | Additive effect (gday$^{-1}$) | Dominance effect (gday$^{-1}$) |
|---|---|---|---|---|---|---|---|
| | $Y_0$ | $Y_1$ | $Y_2$ | | | - | - |
| PARP14 | 0 (0) | -68.15 (15) | 68.15 (506) | 68.15 (0.178) | -136.3 (0.032) | | |
| UMPS | 172.8 (33) | -80.48 (197) | -92.37 (294) | 253.28 (0.0008) | 11.89 (0.436) | 132.585[3] | -120.695[4] |
| UBA5 | -110.9 (4) | 120.3 (113) | -9.34 (408) | -231.2 (0.0009) | 129.64 (0.04) | -50.78 | 180.42[3] |
| ACAD11 | -47.06 (129) | 114.6 (262) | -67.5 (129) | -161.66 (0.014) | 182.1 (0.006) | 10.22 | 171.88[3] |
| BIN1 | 21.62 (5) | -51.77 (98) | 30.15 (418) | 73.39 (0.161) | -81.92 (0.135) | -4.265 | -77.655 |
| ASNSD1 | 0 (0) | -17.82 (3) | 17.82 (518) | 17.82 (0.405) | -35.64 (0.315) | -- | -- |
| MK1671P | 113.4 (6) | -75.63 (50) | -37.77 (472) | 189.03 (0.005) | -37.86 (0.308) | 75.585 | -113.445[4] |
| AOX1 | 39 (29) | 105.5 (121) | -144.5 (381) | -66.5 (0.19) | 250 (0.0008) | 91.75[4] | 158.25[3] |
| SMARCAL1 | -48.29 (22) | 118.4 (211) | -70.11 (287) | -166.69 (0.012) | 188.51 (0.005) | 10.91 | 177.6[3] |
| PQLC2 | 37.35 (13) | 10.55 (145) | -47.9 (368) | 26.8 (0.359) | 58.45 (0.217) | 42.625 | 15.825 |
| NECAP2 | 89.38 (130) | 38.71 (240) | -128.1 (161) | 50.67 (0.248) | 166.81 (0.012) | 108.74[3] | 58.07 |
| INSIG1 | 18.44 (67) | -39.63 (251) | 21.19 (211) | 58.07 (0.217) | -60.82 (0.206) | -1.375 | -59.445 |
| CAST (1) | 127.5 (30) | 119.1 (180) | -246.6 (311) | 8.4 (0.456) | 365.7 (<0.0001) | 187.05[2] | 178.65[3] |
| CAST (2) | 117.5 (57) | -97.83 (226) | -19.62 (240) | 215.33 (0.001) | -78.21 (0.146) | 68.56 | -146.77[3] |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| PLEKHA7 | -37.64 (2) | 38.66 (72) | -1.02 (450) | -76.3 (0.151) | 39.68 (0.298) | -18.31 | 57.99 |
| APIP | 102.9 (22) | -7.58 (116) | -95.31 (383) | 110.48 (0.068) | 87.73 (0.119) | 99.105[4] | -11.375 |
| CYP2B | -75.76 (6) | 69.41 (162) | 6.36 (354) | -145.17 (0.025) | 63.05 (0.197) | -41.06 | 104.11 |
| OCLN | -94.4 (125) | -10.6 (264) | 105 (137) | -83.8 (0.129) | -115.6 (0.059) | -99.7[4] | -15.9 |
| GHR | 91.09 (65) | -135.3 (204) | 44.17 (255) | 226.39 (0.001) | -179.47 (0.007) | 23.46 | -202.93[2] |
| OSMR | 78.26 (57) | 16.18 (209) | -94.44 (257) | 62.08 (0.2) | 110.62 (0.068) | 86.35[4] | 24.27 |
| LIFR | -29.86 (9) | -53.53 (130) | 83.39 (380) | 23.67 (0.375) | -136.92 (0.032) | -56.625 | -80.295 |
| UGT3A1 | 69.97 (80) | -70.4 (268) | 0.43 (171) | 140.37 (0.028) | -70.83 (0.171) | 34.77 | -105.6[4] |
| SLC45A2 | 6 (7) | 138 (60) | -144 (461) | -132 (0.032) | 282 (<0.0001) | 75 | 207[2] |
| MYO10 | 29 (50) | 80 (233) | -109 (240) | -51 (0.248) | 189 (0.005) | 69 | 120[4] |
| LRP5 | 113.9 (3) | 33.15 (27) | -147 (491) | 80.75 (0.137) | 180.15 (0.007) | 130.45[3] | 49.7 |

[1] *Estimates of additive and dominance effects for PARP14 and ASNSD1 were not performed because only two genotypes occurred. The markers were significant at P < 0.01 [(2)], P < 0.05 [(3)] and P < 0.10 [(4)]*

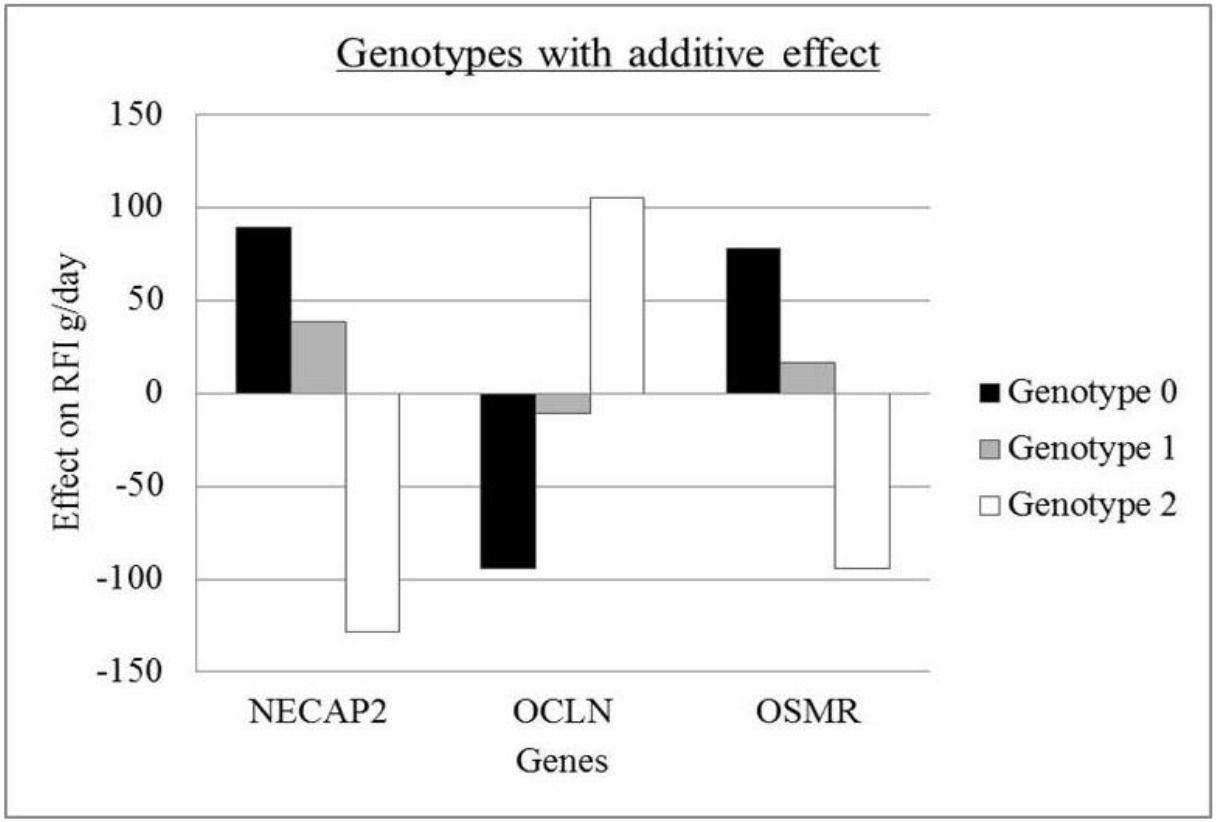



*Figure 4.1a: Predicted genotype effects for genes with additive effects on RFI. The predicted effect of genotype 1, corresponding to the heterozygous genotype, was approximately the average of the two homozygous genotypes 0 and 2, corresponding to the genotype with low and high frequency respectively.*

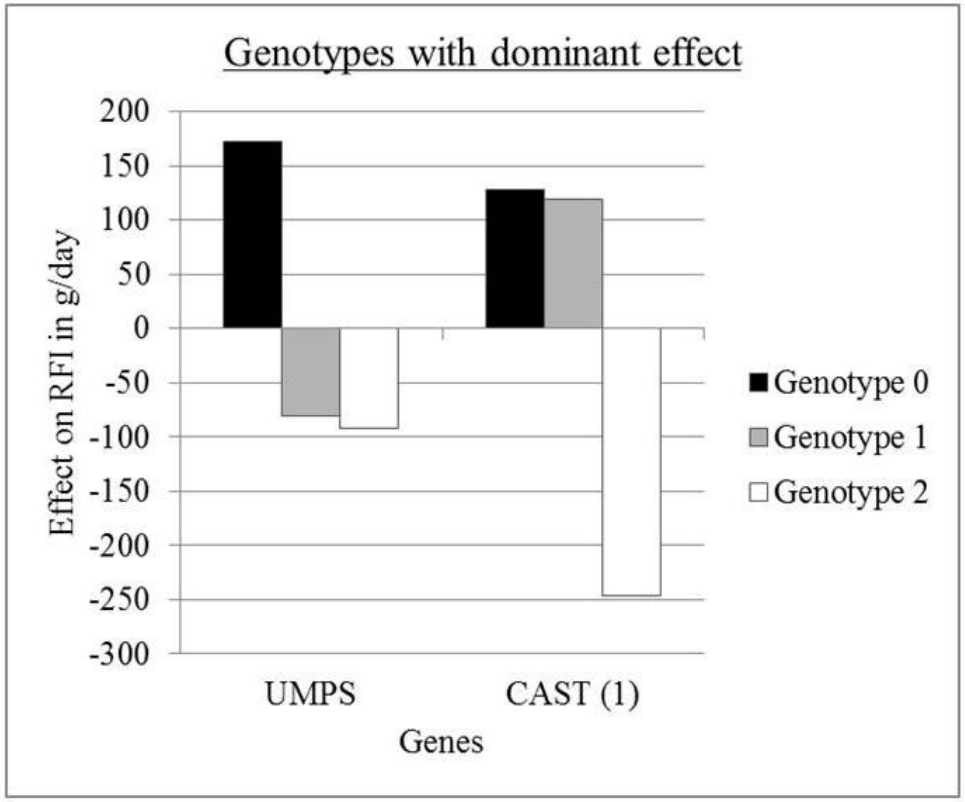



*Figure 4.1b: Predicted genotype effects for genes with dominant effects on RFI. The predicted effect of genotype 1, corresponding to the heterozygous genotype, was closer to one of the homozygous genotype than the other and was not equal to the average of the two homozygous genotypes 0 and 2, corresponding to the homozygous genotype with low and high frequency respectively.*

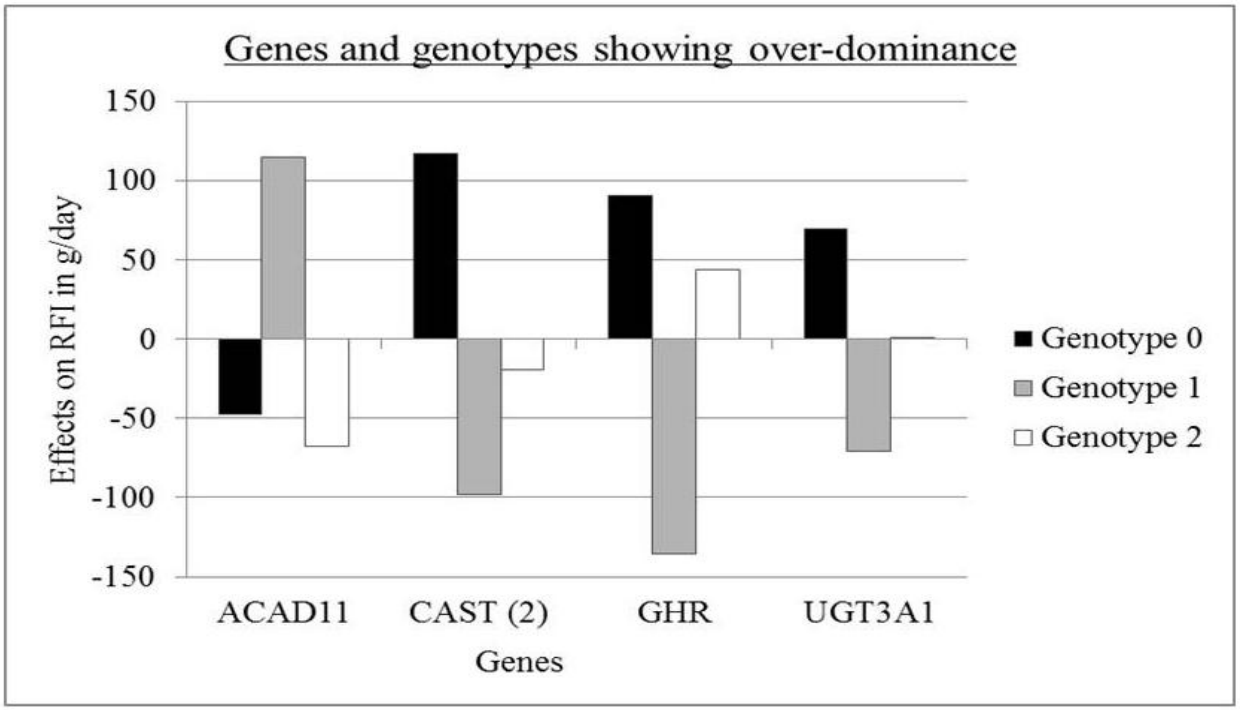



*Figure 4.1c: Predicted genotype effects for genes with over-dominant effects on RFI. The predicted effect of genotype 1, corresponding to the heterozygous genotype, exceeded the value of one of the homozygous genotypes such that the heterozygous genotype conferred more superiority (or inferiority) on the trait than either homozygous genotype 0 or 2, which corresponded to the homozygous genotype with low and high frequency respectively.*

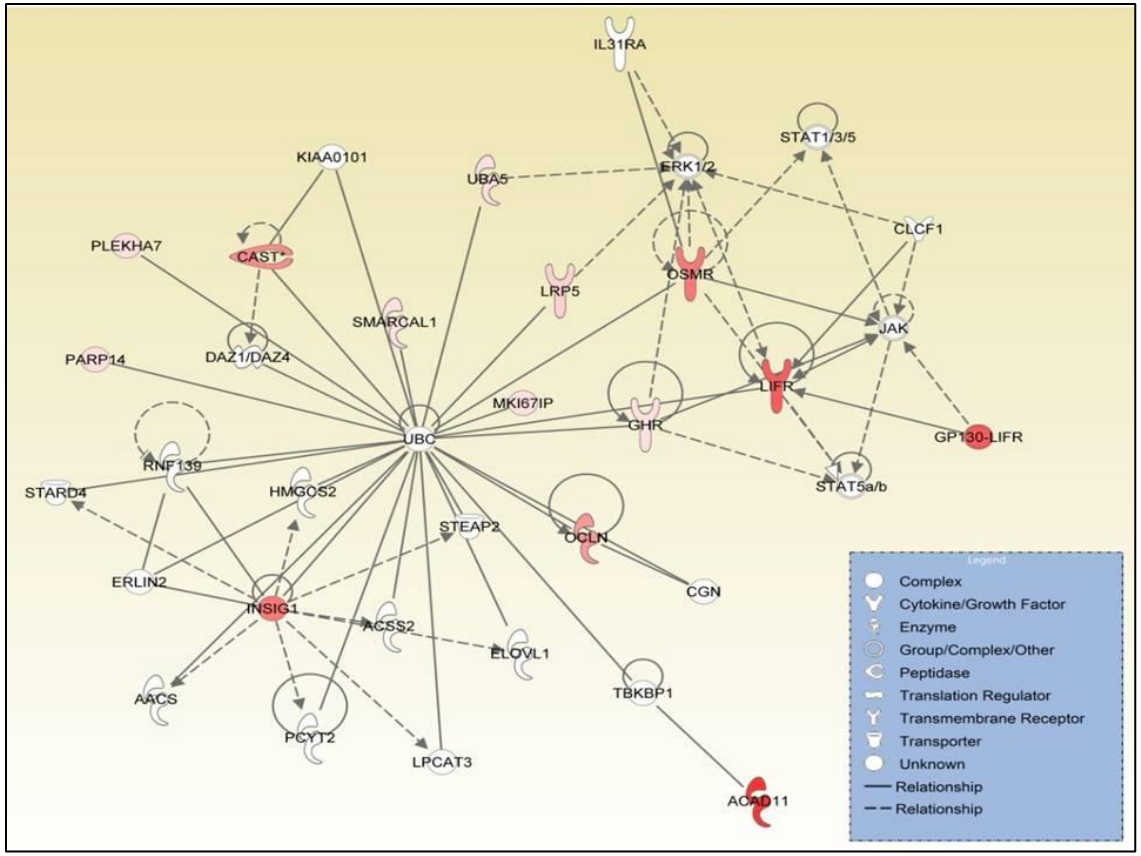



*Figure 4.2: Gene interaction network[1] for candidate genes associated with RFI[2]. [1]Reconstructed using IPA (Ingenuity systems).*

*[2] Only 13 of the 24 significant genes are identified in the network as colored nodes.*

## 4.5.REFERENCES


Aaronson, D. S., and C. M. Horvath. 2002. A road map for those who don't know JAK-STAT. Science. 296:1653-1655.

Ahola, J. K., L. T. Campbell, J. I. Szasz, T. A. Skow, C. W. Hunt, J. B. Glaze. Jr., and R. A. Hill. 2007. Relationship between Residual feed intake and meat quality in steer progeny of divergent intramuscular fat EPD Angus bulls. Proc. of the West. Sec. of Amer. Soc. Anim. Sci. 58:193-199.

Allais, S., H, Levéziel, N. Payet-Duprat, J. F. Hocquette, J. Lepetit, S. Rousset, C. Denoyelle, C. Bernard-Capel, L. Journaux, A. Bonnot, and G. Renand. 2010. The two mutations, Q204X and nt821, of the myostatin gene affect carcass and meat quality in young heterozygous bulls of French beef breeds. J. Anim. Sci. 88:446-454.

Arnold, K., B. Lorenza, K. Jürgen, and S. Torsten. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. Bioinformatics. 22:195-201.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, M. J. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. Gene ontology: tool for the unification of biology. Nat. Genet. 25:25-29.

Barendse W. 2002. DNA markers for meat tenderness. WO02064820. p.1–88.

Barendse, W., A. Reverter, R. J. Bunch, B. E. Harrison, W. Barris, and M. B. Thomas. 2007. A validated whole-genome association study of efficient food utilization in cattle. J. Anim. Sci. 176:1893-1905.

Basarab, J. A., M. A. Price, J. L. Aalhus, E. K. Okine, W. M. Snelling, and K. L. Lyle. 2003. Residual feed intake and body composition in young growing cattle. Can. J. Anim. Sci. 83:189-204.

Bishop, M. D., R. C. M. Simmen, F. A. Simmen, and M. E. Davis. 1989. The relationship of insulin-like growth factor-I with post-weaning performance in Angus beef cattle. J. Anim. Sci. 67:2872-2880.

Bolormaa, S., B. J. Hayes, K. Savin, R. Hawken, W. Barendse, P. F. Arthur, R. M. Herd, and M. E. Goodard. 2011. Genome-wide association studies for feedlot and growth traits in cattle. J. Anim. Sci. 89:1684-1697.

Bulmer, M. G. 1971. The effect of selection on genetic variability. The American Naturalist. 105:201–211.

Calvano, S. E., W. Xiao, D. R. Richards, R. M. Felciano, H. V. Baker, R. J. Cho, R. O. Chen, B. H. Brownstein, J. P. Cobb, S. K. Tschoeke, C. Miller-Graziano, L. L. Moldawer, M. N. Mindrinos, R. W. Davis, R. G. Tompkins, and S. F. Lowry. 2005. A network-based analysis of systemic inflammation in humans. Nature. 437:1032-1037.

Canadian Council on Animal Care. 1993. Guide to the care and use of experimental animals. Vol. 1. E. D. Olfert, B. M. Cross, and A. A McWilliams, eds. 2nd ed. CCAC, Ottawa, ON.

Casas, E., S. N. White, T. L. Wheeler, S. D. Shackelford, M. Koohmaraie, D. G. Riley, C. C. Chase, Jr., D. D. Johnson, and T. P. L. Smith. 2006. Effects of *calpastatin* and μ-*calpain* markers in beef cattle on tenderness traits. J. Anim. Sci. 84:520-525.

Chen, Y., P. F. Arthur, I. M. Barchia, K. Quinn, P. F. Parnell and R. M. Herd. 2012. Using gene expression information obtained by quantitative real-time PCR to evaluate Angus bulls divergently selected for feed efficiency. Anim. Prod. Sci. 52:1058-1067.

Connolly, M. A., P. A. Clausen, and J. G. Lazar. 2006. Purification of RNA from Animal Cells Using Trizol. Cold Spring Harb Protoc, pdb.prot4104.

Crews, D. H. Jr., N. H. Shannon, B. M. A. Genswein, R. E. Crews, C. M. Johnson, and B. A. Kendrick. 2003. Genetic parameters for net feed efficiency of beef cattle measured during postweaning growing versus finishing periods. Proc. West. Sec. Am. Soc. Anim. Sci. 54: 125-128.

Crews, D. H. Jr. 2006. The genetics of feed efficiency in beef cattle. Proc. BIF 38th Annu. Res. Symp. Annu. Meet., Choctaw, MS. p.22–31

Curi, R. A., D. A. Palmieri, L. Suguisawa, A. L. J. Ferraz, H. N. de Oliveira, L. R. Furlan, A. C. Silveira, and C. R. Lopes. Effects of GHR gene polymorphisms on growth and carcass traits in Zebu and crossbred beef cattle. Livest. Sci. 101:94-100.

Elzo, M. A., G. C. Lamb, D. D. Johnson, M. G. Thomas, I. Misztal, D. O. Rae, C. A. Martinez, J. G. Wasdin and J. D. Driver. 2012. Genomic-polygenic

evaluation of Angus-Brahman multibreed cattle for feed efficiency and postweaning growth using the Illumina 3K chip. J. Anim. Sci. 90:2488–2497.

Fan, L. Q., D. R. C. Bailey, and N. H. Shannon. 1995. Genetic parameter estimation of postweaning gain, feed intake, and feed efficiency for Hereford and Angus bulls fed two different diets. J. Anim. Sci. 73:365-372.

Fujino, T., H. Asaba, M. J. Kang, Y. Ikeda, H. Sone, S. Takada, D. H. Kim, R. X. Ioka, M. Ono, H. Tomoyori, M. Okubo, T. Murase, A. Kamataki, J. Yamamoto, K. Magoori, S. Takahashi, Y. Miyamoto, H. Oishi, M. Nose, M. Okazaki, S. Usui, K. Imaizumi, M. Yanagisawa, J. Sakai, and T. T. Yamamoto. 2003. "Low-density lipoprotein receptor-related protein 5 (LRP5) is essential for normal cholesterol metabolism and glucose-induced insulin secretion". Proc. Natl. Acad. Sci. U.S.A. 100:229–234.

Gallagher, S. R. and Desjardins, P. R. 2007. Quantitation of DNA and RNA with absorption and fluorescence spectroscopy. Curr. Protoc. Hum. Genet. Appendix 3:Appendix 3D.

Gandolfi, G., L. Pomponio, P. Ertbjerg, A. H. Karlsson, L. N. Costa, R. Lametsch, V. Russo, and R. Davoli. 2011. Investigation on CAST, CAPN1 and CAPN3 porcine gene polymorphisms and expression in relation to post-mortem calpain activity in muscle and meat quality. Meat Sci. 88:694-700.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2009. ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, UK.

Goonewardene, L. A., Z. Wang, M. A. Price, R. C Yang, R. T. Berg, and M. Makarechian. 2003. The effect of udder type and calving assistance on weaning traits of beef and dairy × beef calves. Livest. Prod. Sci.; 81:47–56.

Grant, J. R., A. S. Arantes, L. Xiaoping, and P. Stothard. 2011. In-depth annotation of SNPs arising from resequencing projects using NGS-SNP. Bioinformatics. 27:2300-2301.

Hayes, B. and Goddard, M. E. 2001. The distribution of the effects of genes affecting quantitative traits in livestock. Genet. Sel. Evol. 33(3):209-29.

Herd, R. M., and S. C. Bishop. 2000. Genetic variation in residual feed intake and its association with other production traits in British Hereford cattle. Livest. Prod. Sci. 63: 111-119.

Herd, R. M. and P. F. Arthur. 2009. Physiological basis for residual feed intake. J. Anim. Sci. 87: (Suppl. 14) E64-71.

Hoque, M. A., K. Katoh, K. Suzuki. 2009. Genetic associations of residual feed intake with serum insulin-like growth factor-I and leptin concentrations, meat quality, and carcass cross sectional fat area ratios in Duroc pigs. J. Anim. Sci. 87:3069-75.

Hu, Z., C. A. Park, E. R. Fritz, and J. M. Reecy. 2010. QTLdb: A comprehensive database tool building bridges between genotypes and phenotypes. 9th World Congress on Genetics Applied to Livestock Production. Leipzig, Germany.

Hubbard, T. J. 2009. Ensembl 2009. Nucleic Acids Res. 37: D690-D697.

Kelly, A. K., M. McGee, D. H. Crews, Jr., A. G. Fahey, A. R. Wylie and D. A. Kenny. 2009. Effect of divergence in residual feed intake on feeding behavior, blood metabolic variables, and body composition traits in growing beef heifers. J. Anim. Sci. 88:109-123.

Kneeland, J., C. Li, J. Basarab, W. M. Snelling, B. Benkel, B. Murdoch, C. Hansen, S. S. Moore. 2004. Identification and fine mapping of quantitative trait loci for growth traits on bovine chromosomes 2, 6, 14, 19, 21, and 23 within one commercial line of Bos taurus. J. Anim. Sci. 82:3405–3414.

Koch, R. M., L. A. Swiger, D. Chambers, and K. E. Gregory. 1963. Efficiency of feed use in beef cattle. J. Anim. Sci. 22:486–494.

Komatsu M, T., Itoh Y. Fujimori, M. Satoh, Y. Miyazaki, H. Takahashi, K. Shimizu, A. E. Malau-Aduli, and M. Morita. 2011. Genetic association between GHSR1a 5'UTR-microsatellite and nt-7(C>A) loci and growth and carcass traits in Japanese Black cattle. Anim. Sci. J. 82:396–405.

Lancet, D., M. Safran, T. Olender, I. Dalah, T. Iny-Stein, A. Inger, A. Harel, and G. Stelzer. 2008. GeneCards tools for combinatorial annotation and dissemination of human genome information. GIACS Conference on Data in Complex Systems, Palermo. Italy.

Li, H., J. Ruan, and R. Durbin. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 18:1851-1858.

Liisa, H. and P. Jong. 2000. DaliLite workbench for protein structure comparison. Bioinformatics. 16: 566-567.

Liu, Y., L. Zan, S. Zhao, Y. Xin, L. Li, W. Cui, Z. Tang, K. Li. 2010. Molecular characterization, polymorphism of bovine ZBTB38 gene and association with body measurement traits in native Chinese cattle breeds. Mol. Biol. Rep. 37: 4041 – 4049.

Lü, A, X. Hu, H. Chen, Y. Dong, and Y. Pang. 2011. Single nucleotide polymorphisms of the prolactin receptor (PRLR) gene and its association with growth traits in Chinese cattle. Mol. Biol. Rep. 38: 261-266.

Maj, A., J. Oprzadek, A. Oprzadek, E. Dymnicki and L. Zwierzchowski. 2004. Polymorphism in the 5'-noncoding region of the bovine growth hormone receptor gene and its association with meat production traits in cattle. Anim. Res. 53:503–514.

Martinez, A., N. Aldai, R. Celaya, K. Osoro. 2010. Effect of breed body size and the muscular hypertrophy gene in the production and carcass traits of concentrate-finished yearling bulls. J. Anim. Sci. 88:1229-1239.

McClure, M. C., N. S. Morsci, R. D. Schnabel, J. W, Kim, P. Yao, M. M. Rolf, S. D. McKay, S. J. Gregg, R. H. Chapple, S. L. Northcutt, and J. F. Taylor. 2010. A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. Anim. Genet. 41:597–607.

McDonagh, M. D. 1998. Calpain activity affects muscle protein turnover and meat tenderisation. PhD Thesis, The University of New England, Armidale, Australia.

McDonagh, M. D., R. M. Herd, E. C. Richardson, V. H. Oddy, J. A. Archerand, and P. F. Arthur. 2001. Meat quality and the calpain system of feedlot steers following a single generation of divergent selection for residual feed intake. Aust. J. Exp. Agric. 41:1013–1021.

Moore, S. S., F. D. Mujibi, and E. L. Sherman. 2009. Molecular basis for residual feed intake in beef cattle, J. Anim. Sci. 87:E41-E47.

Morgan, J. B., T. L. Wheeler, M. Koohmaraie, J. W. Savell, J. D. Crouse. 1993. Meat tenderness and the calpain proteolytic system in longissimus muscle of young bulls and steers. J. Anim. Sci. 71: 1471–1476.

Mujibi, F. D. N., J. D. Nkrumah, O. N. Durunna, J. R. Grant, J. Mah, Z. Wang, J. Basarab, G. Plastow, D. H. Crews, Jr. and S. S. Moore. 2011. Associations of marker panel scores with feed intake and efficiency traits in beef cattle using pre-selected single nucleotide polymorphisms. J. Anim. Sci. 89: 3362-71.

Naik, M. 2008. Identification and characterization of genetic markers and metabolic pathways controlling net feed efficiency in beef cattle. PhD Thesis, University of Adelaide University, Australia

Ng, P. C and S. Henikoff. 2003. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 31: 3812-3814.

Nkrumah, J. D., J. A. Basarab, M. A. Price, E. K. Okine, A. Ammoura, S. Guercio, C. Hansen, C. Li, B. Benkel, B. Murdoch, and S. S. Moore. 2004. Different measures of energetic efficiency and their phenotypic relationships with growth, feed intake, and ultrasound and carcass merit in hybrid cattle. J. Anim. Sci. 82:2451-2459.

Richardson, E. C. and R. M. Herd. 2004. Biological basis for variation in residual feed intake in beef cattle. 2. Synthesis of results following divergent selection. Aust. J. Exp. Agric. 44:431–440.

Rolf, M. M., J. F. Taylor, R. D. Schnabel, S. D. Mckay, M. C. McClure, S. L. Northcutt, M. S. Kerley and R. L. Weaber. 2011. Genome-wide association analysis for feed efficiency in Angus cattle. Anim. Genet. 43: 367 -374.

Sherman, E. L., J. D. Nkrumah, B. M. Murdoch, and S. S. Moore. 2008. Identification of polymorphisms influencing feed intake and efficiency in beef cattle. Anim. Genet. 39:225–231.

Sherman, E. L., J. D. Nkrumah, C. Li, R. Bartusiak, B. Murdoch, and S. S. Moore. 2009. Fine mapping quantitative trait loci for feed intake and feed efficiency in beef cattle. J. Anim. Sci. 87:37-45.

Sherry, S. T., M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, and K. Sirotkin. 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29:308-311.

Snelling, W. M., M. F. Allan, J. W. Keele, L. A. Kuehn, R. M. Thallman, G. L. Bennett, C. L. Ferrell, T. G. Jenkins, H. C. Freetly, M. K. Nielsen, and K.

M. Rolfe. 2011. Partial-genome evaluation of postweaning feed intake and efficiency of crossbred beef cattle. J. Anim. Sci. 89:1731-1741.

Stick, D. A., M. E. Davis, S. C. Loerch, and R. C. M. Simmen. 1998. Relationship between blood serum insulin-like growth factor 1 concentration and postweaning feed efficiency of crossbred cattle at three levels of dietary intake. J. Anim. Sci. 76:498–505.

Yang, J., S. H. Lee, M. E. Goodard and P. M. Visscher. 2011. GCTA: A Tool for Genome-wide Complex Trait Analysis. Am. J. Hum. Genet. 88:76–82.

## *PREFACE TO CHAPTER FIVE*

*In addition to improving feed efficiency, profitability in beef production can also be increased by increasing the quantity and/or quality of beef, which is the major output in beef production.*

*In Chapter 5, putative candidate genes for RFI were evaluated for possible association with various carcass traits.*

*Null hypothesis: None of the SNPs selected within the candidate genes would be significantly associated with carcass quality traits.*

*This study has been published in the Canadian Journal of Animal Science (doi:10.4141/CJAS2012-136).*

# CHAPTER FIVE

## Candidate genes and biological pathways associated with carcass quality traits in beef cattle[3]

### 5.1. INTRODUCTION

The economic success in terms of profitability in beef production relies on producing a product of high economic value (output) at the lowest cost possible (input) (MacNeil et al., 1997). One of the options to increase profitability is to increase the meat quantity and/or quality (outputs) through selection for carcass quality traits such as back fat thickness, marbling and rib eye area. Alternatively, beef producers may select cattle for increased feed efficiency using traits such as residual feed intake (Koch et al., 1963). Some of these economically important traits are correlated, for example average back-fat thickness was reported to be highly correlated with Yield Grade (r = 0.86) (Rios-Utrera et al., 2005), both intramuscular and inter-muscular fat were negatively correlated with lean meat yield and feed efficiency (Basarab et al., 2003) and lean meat yield was positively correlated with residual feed intake (Richardson et al., 1998, Herd and Bishop 2000, Basarab et al., 2003). Mirzaei et al., (2011) reported several significant correlations between some growth traits such as body weight and growth, with carcass quality traits such as hot carcass weight, carcass rib eye area and fat depth. These correlations may result in antagonistic effects on the breeding goal resulting

[3] A version of this chapter has been accepted for publication in Canadian Journal of Animal Science, 2013.

in a dilemma that requires producers to be certain of what traits they wish to select and to be aware of what other traits will be affected and in which direction the secondary effects will occur.

The traditional selection practiced by most breeders is based on the quantitative genetics approach (Dekkers and Hospital, 2002). This approach relies on the availability of phenotypes, the heritability of the traits and their genetic correlations. More recently, molecular markers have gained importance in the evaluation and ranking of candidates for selection. The use of molecular markers in selection relies on the ability to determine the genotypes of individuals for the mutations associated with the traits of interest (Dekkers and Hospital, 2002).

Molecular markers such as SNPs have great potential for use in marker assisted selection (MAS) in beef cattle especially for carcass quality traits because these traits have a large impact on profitability. In addition to profitability, carcass quality traits can only be measured after the animal has been slaughtered, which removes its genetics from the breeding population unless semen or oocytes were collected and stored for future use. In addition, beef producers invest significant expense into herd sires which may ultimately have undesirable carcass traits that are discovered late after their purchase when progeny carcass characteristics are realized.

Another advantage of selecting animals using molecular markers relies on the overlap between QTL associated with multiple traits. Markers located in these QTL can be used to select for one trait and also to predict pleiotropic effects of the

markers on other economically important traits. This can be done using bioinformatics tools or where data are available, by calibrating the markers for effects on all economically relevant traits. For example, the calpastatin gene (CAST) was reported to be associated with multiple traits such as RFI in cattle (McDonagh et al., 2001), meat quality in cattle (Morgan et al, 1993, McDonagh, 1998, Barendse, 2002, Casas et al., 2006) and meat quality in pigs (Ciobanu et al., 2004). Due to this pleiotropy, beef producers who wish to embrace marker assisted selection should be able to predict the effects of genetic markers on traits other than those under primary selection.

In this study, we used SNPs in candidate genes for productivity traits to assess their association with carcass quality traits in beef cattle steers at the University of Alberta ranch at Kinsella, Canada. We also reconstructed a gene network using the significant genes to analyze the gene interactions and biological processes associated with the various carcass traits.

## 5.2. MATERIALS AND METHODS

The animals were managed and cared for according to the guidelines of the Canadian Council on Animal Care (CCAC 1993) and the research was approved by the animal care and use committee at the University of Alberta.

Phenotypic data were obtained from 531 beef cattle at the University of Alberta ranch at Kinsella, Canada. The breed composition of this herd was described in detail by Goonewardene et al., (2003), Nkrumah et al., (2007) and Mujibi et al., (2010, 2011a). Cows and heifers were bred on pasture in a multiple-sire breeding

system and the sire of each calf was later determined in a parentage test by using a panel of bovine microsatellite markers (Nkrumah et al., 2007). The steers were managed and tested under feedlot conditions using the GrowSafe automated feeding system (GrowSafe Systems Ltd., Airdrie, Alberta, Canada). A detailed review of the GrowSafe system can be found at Nkrumah et al., (2004).

### 5.2.1. Phenotypic Data

Carcass traits were collected on steers raised at the University of Alberta ranch at Kinsella, Canada. There were two batches of steers tested each year (Mujibi et al., 2010, 2011b) and carcass quality traits were collected as described by Nkrumah et al., (2004) and Nalaila et al., (2011). The carcass traits considered in this analysis included average back fat (BF), grade fat (GRDFT), carcass rib eye area (CREA), ultrasound rib eye area (UREA), carcass marbling (CMAR), ultrasound marbling (UMAR), yield grade (YGRD), quality grade (QGRD) and lean meat yield (LMY). Lean meat yield as a percentage was estimated using the equation % lean meat yield = 57.96 + (0.202 × $cm^2$ *L. thoracis* area) – (0.027 × kg warm carcass weight) – (0.703 × mm average back fat thickness), Yield grade; the proportion of lean meat; was classified as follows: 1 = ≥59%; 2 = 54 to 58%; and 3 = <54% (Basarab et al., 2003).

### 5.2.2. The candidate gene approach

#### *5.2.2.1. Identification of positional candidate genes*

Two sets of positional candidate genes were utilized in this study. The first set consisted of 1100 candidate genes positioned within a range of 500kbp on either

side of 203 QTL reported in the Bovine QTL database (cattleQTLdb) (Hu et al., 2007). These QTL were associated with some production traits including body weight (BW), average daily gain (ADG), feed efficiency, dry matter intake (DMI), metabolic weight (MW) and energy balance. The second set of positional candidate genes consisted of 1018 genes positioned within a range of 500kbp on both sides of 310 SNPs previously reported to be significantly associated with RFI in a GWAS using the 50K bovine SNP chip (Mujibi et al., 2011b).

A total of 2118 positional candidate genes were identified in this study.

#### *5.2.2.2. SNP Detection*

The SNPs located in the positional candidate genes were identified from the NCBI SNP database (dbSNP) (Sherry et al., 2001) and by comparing cDNA sequences generated from liver samples from steers at the University of Alberta ranch at Kinsella, Canada with reference sequences from Ensembl version 57 (Hubbard et al., 2009).

To generate the cDNA library, RNA was prepared from pooled liver samples using TRIzol® reagent (Invitrogen, Carlsbad, CA, USA) using the protocol described in the TRIzol® reagent user manual (Conolly et al., 2006). The quality and quantity of RNA was determined using a NanoDrop 2000 spectrophotometer (Nanodrop technologies, USA) (Sean, 2007). The liver cDNA library was constructed according to the TruSeq® RNA and DNA sample preparation kit (Illumina, USA, v2 published by Illumina Inc. 2011, publication number 970-2009-039) and cDNA sequencing was performed on the Genome Analyzer II

using the TruSeq® RNA and DNA sample preparation kit (Illumina, USA, v2 published by Illumina inc 2011, publication number 970-2009-039).

To generate a list of putative SNPs, Maq (version 0.7.1) (Li et al., 2008) was used to map DNA sequencing reads and the bovine transcript sequences from Ensembl version 57 (Hubbard et al., 2009) were used as reference sequences. The SNPs returned by Maq's SNPfilter command were annotated using NGS-SNP (Grant et al., 2011) by assigning a functional class to each SNP (e.g. nonsynonymous) and then providing NCBI, Ensembl identifiers and gene ontology terms for the affected genes, when applicable. Only the non-synonymous SNPs were considered functional candidate SNPs and were retained if the minor allele frequency was greater than 10%.

The positional candidate genes containing functional SNPs were then prioritized according to their functions and gene ontology terms (Ashburner et al., 2000) in relation to the biological processes associated with feed efficiency (Herd and Arthur, 2009). Genes with functions related to metabolism of carbohydrate, lipid and protein, growth, ATP and body temperature regulation were considered functional candidate genes and a final set of 116 genes was identified spread across 12 chromosomes with some chromosomes containing larger numbers of functional candidate genes than others. The non-synonymous SNPs from each gene were prioritized based on the expected effect of the amino acid change such that among several SNPs, a non-conservative amino acid change was considered of higher priority than a conservative amino acid change in the same gene. A total of 117 SNPs were selected within the 116 genes; one SNP from each gene and 2

SNPs from the CAST gene, and genotyped at GeneSeek® using DNA samples obtained from steers at the University of Alberta ranch at Kinsella, Canada.

### 5.2.3. Association analysis

Association analysis was performed using ASReml 3 software (Gilmour et al., 2009). The homozygous genotype containing the two alleles with lower frequency in the population was coded 0, the homozygous genotype with two alleles of higher frequency was coded 2 and the heterozygote genotypes were coded 1. Of the 117 functional SNPs, 113 were successfully genotyped and 39 SNPs were monomorphic leaving 74 polymorphic SNPs to be used in the analysis.

The multiple marker association analysis was performed using the animal model:

$$Y_{ij} = \mu + X_{1i}\beta + \sum_{j=1}^{74} X_{2j} g_j + Za + e \quad \ldots\ldots \quad \text{(Equation 1)}$$

$Y_{ij}$ represents the specific carcass trait being studied for animal *i* which has *j* SNPs ($j$ = *74 SNPs)*

μ is the population mean,

β is the vector of fixed effects associated with animal *i*, in this case the breed of the sire and the batch (Mujibi et al., 2010)

$\sum_{j=1}^{74} X_{2j} g_j$ is the sum of the SNP effects (g), $X_2 j$ is a design matrix relating an observation (*y*) to one of the genotypes 0, 1 or 2 at the *j*th SNP,

a is a vector of random additive effects (exclusive of SNP g effect(s) in the model) constructed using each animal's pedigree information to construct the relationship

matrix A. It was assumed to be normally distributed with a mean of 0 and variance of $A\sigma^2_a$.

e was the vector of random residual effects assumed to be normally distributed with a mean of 0 and a variance of $I\sigma^2$ where I was an identity matrix

The significance criteria was a P-value corrected for 9 tests; using the same model for 9 different traits in this study, therefore SNPs were considered significant if $P < 0.005$.

A multivariate analysis was performed for significantly correlated ($P < 0.05$) traits (Table 5.1) using ASReml 3 to identify markers that were significantly associated with variation in multiple traits simultaneously. Traits in this analysis were BF, UREA, CREA, CMAR, UMAR, YGRD, QGRD, GRDFT and LMY.

### 5.2.4. Reconstructing gene networks

Gene interaction networks were reconstructed and biological processes were identified using IPA software (Ingenuity® Systems, www.ingenuity.com) for the genes that were significantly associated with single carcass traits. The IPA software was selected because it offered a large knowledge base and can model relationships between genes, proteins, metabolites and can be used to identify biological pathways and interaction complexes (Ingenuity® knowledge base). To reconstruct the gene network, a list of significant genes and their corresponding p-values were imported into the IPA software and the parameters were set to allow the network to include indirect relationships between the imported genes and

genes that were in the database. Indirect relationships would assist in the identification of other genes that were not among the genes analyzed but may be associated with carcass traits. The IPA algorithm generates gene networks by mapping each gene identifier to its corresponding gene in the IPA Knowledge Base. The genes are then overlaid onto a global molecular network developed from information contained in the Knowledge Base. The networks are generated based on their connectivity such that each network has a maximum of 35 imported genes. Each network is assigned a significance score, which represents the likelihood that the imported genes within the network are found therein by random chance. A high number of imported genes within a dataset lead to a higher network score. The network score is calculated as the negative of the exponent of the *P*- value such that a score of 25 is equal to a *P*- value of $10^{-25}$ (Calvano et al., 2005). And therefore large scores correspond to high significance.

## 5.3. RESULTS

### 5.3.1. Correlation analysis

Phenotypic correlation analysis indicated significant correlations between multiple carcass traits. Average back fat was significantly correlated with ultrasound rib eye area, ultrasound marbling, carcass marbling, yield grade, grade fat and lean meat yield. A summary of the correlations are shown in Table 5.1.

### 5.3.2. Association analysis for individual carcass traits

Seven genes were significantly associated with various carcass traits at $P < 0.005$. Three more genes with $P < 0.05$ were considered to show a trend of association because they had been reported to be associated with carcass traits in previous studies. The amount of phenotypic variation explained by the significant SNPs for each trait is discussed in subsequent sections. These genes, their allele substitution effect and their functions are summarized in Table 5.2 and Table 5.3 shows the details of the SNPs and amino acid changes.

The SNPs located in the Diacylglyerol kinase gene (DGKD), Serine/threonine kinase 10 (STK10) and the Ankyrin repeat and BTB domain containing gene (ABTB2) were associated with slaughter weight at $P = 0.005$, 0.0027 and 0.0047 respectively. Two SNPs located in the 17-beta-hydroxysteroid dehydrogenase 12 (HSD17B12) gene and the UBXN4 domain-containing protein 4 were significantly associated with carcass quality grade ($P = 0.0006$ and 0.0048 respectively). The SNP in the NADH dehydrogenase iron sulfur protein 3 (NDUFS3) was associated with carcass rib eye area ($P = 0.0005$) with an effect of $-2.66cm^2$. A SNP located in the low density lipoprotein receptor-related protein 4 (LRP4) was significantly associated with yield grade ($P = 0.0008$).

The three genes showing a trend of significance were the growth hormone receptor precursor (GHR) associated with grade fat and average back fat thickness ($P = 0.03$ and 0.025 respectively) and the solute carrier family 45, member 2 (SLC45A2) associated with grade fat ($P = 0.01$) and ARHGAP1 protein containing fragment (ARHGAP1) associated with carcass weight ($P = 0.008$).

### 5.3.3. Multivariate analysis

A multivariate analysis was performed in ASReml 3 (Gilmour et al., 2009) using the significantly correlated traits to identify genes that were significantly associated with all traits. The highly correlated traits were BF, UREA, CREA, CMAR, UMAR, YGRD, QGRD, GRDFT and LMY.

Four SNPs located in the 17-beta-hydroxysteroid dehydrogenase 12 (HSD17B12), Pyruvate dehydrogenase complex (PDHX), Solute carrier family 30 (zinc transporter), member 5 (SLC30A5) and Ubiquitin like modifier activating enzyme 5 (UBA5) genes were significant ($P < 0.005$). The SNP in the HSD17B12 that was significant in the multivariate analysis was the same as the one that was significantly associated with quality grade, which was significantly correlated with CREA, UREA and UMAR.

### 5.3.4. Gene Network and Biological pathways

Using the 13 significant genes imported into IPA, three networks were reconstructed in IPA. The first network had the highest score of 23 and out of the 13 significant genes, this network consisted of 9 genes including ABTB2, HSD17B12, LRP4, DGKD, STK10, GHR, UBXN4, SLC30A5 and PDHX genes as shown in Figure 5.1a. The second gene network had a score of 6 and out of the 13 significant genes this network consisted of 3 genes including ARHGAP1, NDUFS3 and UBA5 (Fig 5.1b). The third network had a score of 3 and out of the 13 genes this network contained only 1 gene, SLC45A2. Several other genes that formed indirect relationships with these genes were also included in the networks.

However, only the first and the second gene networks, which included more than one of the significant genes, will be discussed further.

The gene network with the highest score had several hubs as shown in figure 5.1a.

The Ins1 hub was composed of several interactions with the insulin 1 gene which encodes the insulin hormone. The insulin hormone plays a role in decreasing blood glucose concentration by increasing cell permeability to glucose. It also increases cell permeability to amino acids and fatty acids. In addition, it accelerates glycolysis, the pentose phosphate cycle and glycogen synthesis in the liver (Rebhan et al, 1997).

An additional hub was centered on the NFkB (complex) transcription factor, which is involved in several processes such as cellular growth, immune and inflammatory responses and developmental processes. A complete review of the NFkB complex was published by Gilmore (2006). There were several interactions including 3 genes involved in estrogen metabolism. These genes are also involved in lipid metabolism and fatty acid biosynthesis.

The second gene network was centered at the Ubiquitin C (UBC) gene (the UBC hub) and also consisted of several minor hubs associated with the NDUF genes. The UBC gene also interacted with NDUFS3 and the several minor hubs associated with the NDUF genes. The UBC gene was not analyzed in this study, but its location on the network and its multiple interactions with the significant genes indicates that it may influence their function and thereby potentially influence carcass traits.

Several genes were included in the gene networks and may be associated with carcass traits including AGRN, MAPK8IP1, CDKN1A, RANBP9, the NDUF genes and several additional genes as shown in Fig 5.1a and Fig 5.1b.

The biological processes represented by the UBC hub relate to regulation of gene expression and include degradation of proteins in the endoplasmic reticulum, lysosomal degradation, and protein degradation via the proteasome, activation of transcription factor NF-kappa-B, cell signaling and DNA repair. Therefore, this hub represents genes and biological processes that influence variation in phenotypes through processes that regulate the levels of gene expression and protein function such as DNA transcription and protein degradation respectively.

The NDUF hubs are associated with energy production and utilization, which is also related to fat and steroid metabolism.

Other important biological processes identified in this pathway analysis included acetyl-CoA biosynthesis, androgen and estrogen biosynthesis, phospholipid degradation, glycerophospholipid metabolism and cytokine signaling.

## 5.4. DISCUSSION

### 5.4.1. Genes associated with individual carcass traits

Protein structure and function have been shown to be influenced by amino acid properties such as the polarity and their interaction with water. In most cases, polar amino acids are found located on the outside of the protein and interact closely with tissue fluids because they are hydrophilic (Branden and Tooze, 1999). Mutations that cause a change from a polar to a non-polar amino acid may

result in a major change in the structure of the protein and as a result affect its function. In this study, all the significant SNPs resulted in amino acid whose polarity was different from the amino acid in the reference sequence (Table 5.3).

### *a. Diacylglycerol Kinase Delta (DGKD)*

In this study, DGKD was significantly associated with slaughter weight and explained 0.4% of the phenotypic variation in slaughter weight. DGKD also had an effect of increasing the average back fat thickness and a reduction in carcass weight though these effects were not significant (P= 0.05 and P=0.02 respectively). The function of DGKD is to catalyze the conversion of Diacylglyerol (DAG) to phosphatidic acid (PA) with ATP as the phosphate group donor. Both DAG and PA are lipid signaling molecules and DGKD acts as the switch by terminating the signaling of one lipid while simultaneously activating signaling by another lipid (Merida et al., 2008). According to the gene ontology database (Ashburner et al., 2000), and the UniProt knowledgebase (UniProt Consortium, 2012), the functions of the DGKD gene relate to fat metabolism and include glycerolipid and glycerophospholipid metabolism and the phosphatidylinositol signaling system, both of which may influence the accumulation of back fat. Although there are no other studies reporting the association between this gene and carcass traits the gene has a role in lipid metabolism explaining the effect it has on average back fat and we hypothesize that this may secondarily affect the carcass weight. Further studies are required to validate this association in other populations.

### *b. 17-beta-hydroxysteroid dehydrogenase 12 (HSD17B12)*

17-beta-hydroxysteroid dehydrogenase 12 (HSD17B12) catalyzes the transformation of estrone (E1) into estradiol (E2), suggesting a central role in estrogen formation. It also has 3-ketoacyl-CoA reductase activity, reducing both long chain 3-ketoacyl-CoAs and long chain fatty acyl-CoAs, suggesting a role in long chain fatty acid elongation (Moon and Horton, 2003). In general the enzyme is involved in lipid metabolism, fatty acid biosynthesis and steroid, especially estrogen, synthesis. In this study, this gene was associated with quality grade (P = 0.0006) and explained 1.36% of the phenotypic variation. HSD17B12 also showed a trend of association with marbling (P = 0.02). Although there has been no previous report on its association with quality grade and marbling, we hypothesize that its role in biological processes related to lipid metabolism may influence the fat levels in meat resulting in variation in marbling and consequently carcass quality grade.

### *c. Low density lipoprotein receptor related protein 4 (LRP4)*

The LRP4 gene was significantly associated with yield grade explaining about 5.5% of the phenotypic variation. LRP4 is a potential cell surface endocytic receptor and has been associated with functions such as calcium ion binding, anatomical structure development, cell differentiation and bone formation (Ashburner et al., 2000, Uniprot Consortium, 2012 and Rebhan et al., 1997). A SNP in the LRP4 gene was significantly associated with bone mineral density and limb development in humans (Unnur et al., 2008). A mutation in this gene caused

syndactyly in Holstein cattle (Duchesne et al., 2006). LRP4 also plays a role in lipid and cholesterol metabolism as a member of the low density lipoprotein receptor gene family which is involved in reducing cholesterol levels in blood (Brown and Goldstein, 1997). The role of this gene in anatomical development and bone mineralization may have an impact on the weight of the carcass. In addition, based on its role in cholesterol metabolism, we hypothesize that it may consequently affect yield grade which is estimated from a combination of carcass weight, fat content and muscle development.

***d. NADH dehydrogenase (NDUFS3)***

The NDUFS3 gene was associated with CREA (P=0.0005) and explained 3.2% of the phenotypic variation in CREA. The gene also showed a trend of association with YGRD (P = 0.03). The enzyme NADH dehydrogenase (NDUFS3) is involved in a complex of reactions in the electron transport chain and oxidative phosphorylation in the mitochondria (Ashburner et al., 2000 and UniProt Consortium, 2012). Other gene ontology terms associated with NDUFS3 are protein binding, apoptosis and negative regulation of cell growth (Ashburner et al., 2000). In a previous study, the NADH dehydrogenase 2 was significantly associated with marbling fat content in the loin muscle (Kim et al., 2009). We hypothesize that its role in energy metabolism in the electron transport chain may influence traits related to growth and fat deposition thereby influencing marbling. The role of the electron transport chain may also be influenced by the muscle fiber type and size where high oxidative fiber types tend to be smaller. A detailed review of the relationship between fiber type and size can be found at Wessel et

al., (2010). Further research is needed to validate the association between this gene and carcass rib eye area in other populations of cattle.

### *e. Serine threonine kinase 10 (STK 10)*

STK10 was significantly associated with slaughter weight (P = 0.02) and explained 0.62% of the phenotypic variation in SLTWT. It also showed a trend of association with carcass weight (P = 0.03), yield grade (P = 0.04) and ultrasound rib eye area (P = 0.04). Serine/Threonine Kinase 10 belong to the family of Serine/Threonine kinases and their functions are to phosphorylate and activate members of the AMPK-related subfamily of protein kinases (Baas et al., 2003). In pigs, the protein kinase AMP-activated γ3 subunit gene, *PRKAG3*, which encodes the γ3 isoform of AMPK was identified by positional cloning as the causative gene for the Rendement Napole (RN) phenotype (Andersson 2003). The RN phenotype is common in Hampshire pigs and is characterized by a 70% increase in skeletal muscle glycogen content, decreased post mortem muscle pH and water content and increased lean meat content (Andersson 2003). There is evidence that the RN phenotype is caused by a missense mutation (Arg to Gln) in *PRKAG3*. Other polymorphisms in porcine PRKAG3 are associated with meat quality traits including meat color (Gunilla et al., 2004) and water holding capacity. In sheep, AMPK was shown to be negatively correlated with muscle adipogenesis (Tong et al., 2008). In beef cattle, a SNP marker in PRKAG3 position T2885C has been significantly associated with meat tenderness (Wu-Feng et al., 2012). We hypothesize that the effects of STK10 on the carcass traits may be through the activation of members of the AMPK-related subfamily of protein kinases and

consequently have effects on multiple carcass traits. Additional studies are required to validate this hypothesis.

### 5.4.2. Genes significant in multivariate analysis

The genes significant in multivariate analysis were HSD17B12 (see above), PDHX, SLC30A5 and UBA5. The multiple traits considered in this analysis were either related to fat content in meat such as carcass and ultrasound marbling, average back fat thickness and grade fat or they had components that were calculated from fat related traits such as yield grade which is estimated from fat content, carcass weight, and muscle development.

The PDHX gene product is a component of the pyruvate dehydrogenase enzyme complex, which catalyses the conversion of pyruvate into acetyl CoA which enters the citric acid cycle (cellular respiration) producing energy for cellular processes (Rebhan et al., 1997). A previous study reported significant association between a SNP in the PDHX gene and body weight and body mass index in humans (Fox et al., 2007). The effect of this gene on energy metabolism may subsequently affect fat composition and growth in steers resulting in variation in carcass and ultrasound measurements of fat, growth and carcass weight. We therefore hypothesize that the gene may play a similar role in cattle as that reported in humans by Fox et al., (2007) resulting in variation in carcass weight. There were no studies that reported association between the SLC30A5 and UBA5 and carcass traits, however, we recommend further studies to validate the significant associations observed.

### 5.4.3. Genes showing a trend of association

The GHR gene codes for the receptor that binds growth hormone and other peptide hormones. In this study, the SNP in GHR gene showed a trend of association with both average back-fat thickness and grade fat explaining 8.8% of the phenotypic variation in grade fat and 9.0% of the phenotypic variation in average back fat. Mogens et al., (1993) showed that growth hormone significantly reduced the fat content and trim on meat. Another allele on the GHR gene has been significantly associated with mean differences in final weight, eye muscle area, marbling score and fat color, but the same allele was not associated with carcass weight, back-fat thickness and final meat quality grade or meat color (Han et al., 2009). A different SNP in the 4$^{th}$ intron of the GHR gene was significantly associated with body weight and feed efficiency in beef cattle (Sherman et al., 2008).

The SLC45A2 gene showed a trend of association with grade fat explaining 5% of the phenotypic variation. The functions of SLC45A2 gene are more related to melanin production and in humans the gene is associated with hair, skin and eye pigmentation. Polymorphisms in the same gene have been associated with silver and white color phenotypes in chickens (Gunnarsson 2007) and using the comparative functional genomics approach, similar effects may be anticipated in beef cattle. The relationship between coat color and growth in cattle was investigated by Finch et al., (1984) where they showed that color had significant effects on growth with white steers gaining 0.13 kg more per day than dark steers. They also showed that coat color affected the feeding behavior of steers where

light colored steers spent more time in the sun grazing than dark ones. The relationship between coat color, obesity and adiposity has also been illustrated with the agouti gene in mice and its interactions with the melanocortin receptors with MC4R knockout mice showing obesity. MC3R has also been implicated in body weight regulation where antagonists of MC4R result into increased fat mass, reduced lean meat and increased feeding (Voisey and Van Daal 2001). In addition, a SNP located in a highly conserved region of the MC4R gene was found to be significantly associated with back-fat thickness, growth rate and appetite in pigs (Kim et al., 2000 and Fan et al, 2009). Therefore, although a direct link between the SLC45A2 and carcass traits has not been reported, we hypothesize that it may interact with the pigment genes and melanocortin receptors and consequently influence fatness, growth rate and appetite.

The ARHGAP1 gene is a member of the cytoskeleton regulator family associated with protein binding and regulation of GTPase activity. Although the ARHGAP1 gene has not been linked to carcass traits before, a SNP in the promoter region of this gene has been associated with osteoporosis in mice (Duncan and Brown, 2010). In another study, ARHGAP1 knockout mice tended to be weaker and most died during their neonatal period. ARHGAP1 knockout mice that survived had a short lifespan, showed premature ageing phenotypes such as reduction in body mass, loss of sub-dermal adipose tissue and osteoporosis (Wang et al., 2007). In this study, the ARHGAP1 gene was associated with carcass weight and accounted for 1.2% of the phenotypic variation. Although there is a lack of previous reports of association between ARHGAP1 and carcass traits, we hypothesize that

variation in this gene may have effects similar to the ones reported by Wang et al (2007) resulting in lower body mass and sub-dermal adipose tissue which would result in reduced carcass weight in beef cattle.

Although two SNPs in the Calpastatin gene were included in this study, there was no significant effect on carcass quality traits. Previous studies have reported significant association between CAST and feed efficiency and carcass quality (see introduction), the main effect is on beef tenderness (Barendse 2002, Casas et al 2006).

When all the genes significantly associated with specific traits were considered together, DGKD, ABTB2 and STK10 accounted for about 5% of the phenotypic variation in slaughter weight and UBXN4 and HSD17B12 accounted for about 14% of the phenotypic variation in quality grade. To be cost effective, a marker set needs to account for at least 10-15% of the genetic variance (Crews, 2008), therefore the two markers associated with quality grade may be validated and incorporated into a marker assisted selection panel for beef quality grade. Although these genes were initially selected as positional candidate genes for traits that may also influence feed efficiency, none of the genes that accounted for more than 10% of the phenotypic variation in a carcass trait was also significantly associated with feed efficiency indicating that their use in marker assisted selection will not have significant effects on feed efficiency traits.

### 5.4.4. Gene networks and biological processes

The gene networks and biological processes were reconstructed using IPA® (Ingenuity® systems) software. There were few interactions between the significant genes, which may indicate a possibility that their effects are relatively independent and only a few genes interact with each other directly. However, these genes may interact indirectly depending on the hub they belong to and the genes that are involved in that hub.

The major hubs identified were associated with the insulin (Ins1) gene in the first network and UBC gene in the second network. There were several hubs associated with other genes including the NDUF genes.

The Ins1 hub corresponded with biological processes associated with the metabolism of glucose, sterol and lipids. Other genes in the first network were involved in sterol metabolism and the regulation of transcription of several genes involved in several biological processes including cellular growth, immunity and cellular development.

The Ubiquitin C (UBC) hub corresponded to biological processes including degradation of proteins in the endoplasmic reticulum, lysosomes and the proteasome, and activation of transcription factor NF-kappa-B, cell signaling and DNA repair.

The NDUF hubs represent biological processes involved in energy production and utilization, and by extension lipid metabolism. These processes are more specific to carcass traits especially the traits related to fat content in meat such as marbling and average back fat thickness.

Other important biological processes identified by the gene network analysis were acetyl-CoA biosynthesis, estrogen biosynthesis and cytokine signaling. Acetyl CoA is important in energy and lipid metabolism and may have an effect on fat traits in carcasses. Estrogen is also involved in lipid and cholesterol metabolism and growth in complex biological pathways that will not be discussed in detail here but can be found in a review by Berthezène (1999). These processes may be implicated in its role in influencing carcass traits in cattle.

## 5.5. CONCLUSIONS

We have reported ten SNPs in ten genes associated with various carcass traits in beef cattle with significant effects on slaughter weight, carcass weight and average back-fat thickness. These genes need to be validated across other diverse breeds and populations in other geographical locations to assess the reproducibility of the results in other populations. If these markers show consistent results across different populations then SNP panels for marker assisted selection could be developed from these markers for selection of carcass traits in diverse beef populations.

**Table 5.1: Phenotypic and (genetic)[z] correlations between carcass merit traits in cattle from University of Alberta Kinsella ranch[y].**

| TRAITS | BF | GRDFT | CREA | UREA | CMAR | UMAR | YGRD | QGRD | LMY |
|---|---|---|---|---|---|---|---|---|---|
| BF | 1.00 | 0.76[x] (0.69)[x] | 0.01 (0.01) | 0.35[x] (0.08) | 0.47[x] (0.18)[x] | 0.47[x] ((0.14)[x] | 0.64[x] (0.32)[x] | -0.06 (0.04) | -0.66[x] (-0.78)[x] |
| GRDFT | | 1.00 | -0.09(-0.12) | 0.22[x] (-0.16)[x] | 0.53[x] (0.24)[x] | 0.43[x] (0.32)[x] | 0.76[x] (0.36)[x] | -0.03 (0.08) | -0.87[x] (-0.62)[x] |
| CREA | | | 1.00 | 0.67[x] (0.80)[x] | -0.07 (-0.01) | -0.13[x] (-0.08) | -0.30[x] (-0.12)[x] | -0.30[x] (0.45)[x] | 0.45[x] (0.56)[x] |
| UREA | | | | 1.00 | 0.13[x] (0.11) | 0.11[x] (0.13)[x] | 0.06 (0.02) | -0.40[x] (0.52)[x] | 0.03 (0.58)[x] |
| CMAR | | | | | 1.00 | 0.55[x] (0.75)[x] | 0.41[x] (0.65)[x] | 0.03 (0.38)[x] | -0.54[x] (-0.22)[x] |
| UMAR | | | | | | 1.00 | 0.39[x] (0.54)[x] | 0.13[x] (0.15)[x] | -0.44[x] (-0.24)[x] |
| YGRD | | | | | | | 1.00 | 0.06 (0.12)[x] | -0.81[x] (0.19)[x] |
| QGRD | | | | | | | | 1.00 | -0.08 (0.06) |
| LMY | | | | | | | | | 1.00 |

*[y] The phenotypic and genetic correlation between various carcass merit traits, including backfat (BF), grade fat (GRDFT), carcass ribeye area (CREA), ultrasound rib eye area (UREA), carcass marbling (CMAR), ultrasound marbling (UMAR), yield grade (YGRD), quality grade (QGRD) and lean meat yield (LMY).[z] Genetic correlations are in parenthesis*

*[x] Significant correlation (P < 0.05).*

**Table 5.2: SNPs and genes significantly associated with various carcass traits, the position of the SNPs and the functions of the genes[z]**

| Gene symbol | Gene name | Associated trait | P-Value | Number of individuals per genotype 0, 1, 2 | | | Allele effect | Gene function[y] |
|---|---|---|---|---|---|---|---|---|
| UBA5 | Ubiquitin like modifier activating enzyme 5 | Multiple traits | 0.0003 | 4 | 113 | 408 | - | Ubiquitin like post translational modifier protein |
| UBXN4 | UBX domain-containing protein 4 | Quality grade | 0.0048 | 57 | 236 | 230 | -0.1676 | Involved in endoplasmic reticulum-associated protein degradation thus regulation of phenotypic expression |
| DGKD | Diacylglycerol kinase | Slaughter weight | 0.005 | 6 | 160 | 356 | -12.43kg | Involved in Glycerolipid and glycerophospholipd metabolism and phosphatidylinositol signaling system |
| ABTB2 | Ankyrin repeat and BTB (POZ) domain containing 2 | Slaughter weight | 0.0047 | 132 | 238 | 160 | 11.63kg | Suggested to be involved in DNA and protein binding. However the functions of this gene remain largely unknown. |
| PDHX | Pyruvate dehydrogenase complex (complex X) | Multiple traits | 0.0023 | 125 | 264 | 137 | - | The pyruvate dehydrogenase complex catalyzes the conversion of pyruvate to acetyl CoA. |
| HSD17B12 | 17-beta-hydroxysteroid dehydrogenase 12 | Quality grade | 0.0006 | 34 | 160 | 327 | 0.1951 | Catalyzes the transformation of estrone (E1) into estradiol (E2), suggesting a central role in estrogen formation. Also has 3-ketoacyl-CoA reductase |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | activity, reducing both long chain 3-ketoacyl-CoAs and long chain fatty acyl-CoAs, suggesting a role in long fatty acid elongation. |
| ARHGAP1 | ARHGAP1 protein Fragment | Carcass weight | 0.0083 | 60 | 213 | 250 | 12.11kg | One of cytoskeleton regulators and the ontology terms associated with this gene were protein binding and regulation of GTPase activity. Has been associated with reduction in body mass, adipose tissue and osteoporosis |
| LRP4 | Low density lipoprotein receptor-related protein 4 | Yield grade | 0.0008 | 23 | 27 | 471 | -0.4533 | Plays a key role in the formation and the maintenance of the neuromuscular junction.<br><br>It has also been proposed to function as a cell surface endocytic receptor binding and internalizing extracellular ligands for degradation by lysosomes<br><br>In humans and cattle this gene is involved in bone development and growth.<br><br>May also be involved in cholesterol metabolism. |
| NDUFS3 | NADH dehydrogenase | Carcass rib eye area | 0.0005 | 30 | 200 | 294 | -2.663 (cm2) | Involved in a complex of reactions in the electron transport chain and oxidative phosphorylation in the mitochondria. Therefore plays a role in energy |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | metabolism. |
| STK10 | Serine/threonine kinase 10 | Slaughter weight | 0.0027 | 22 | 215 | 284 | 18.14kg | One of Serine Threonine Kinase genes. Their roles are in cell cycle arrest, protein amino acid phosphorylation and regulation of fatty acid oxidation |
| SLC30A5 | Zinc transporter (solute carrier family 30) member 5 | Multiple traits | 0.004 | 80 | 265 | 171 | - | May be involved in zinc transport into cells to form insulin crystals |
| GHR | Growth hormone receptor Precursor (GHR) | Grade fat | 0.03 | 65 | 204 | 255 | 1.59 | This gene encodes the receptor that binds the growth hormone and activates intercellular signals that lead to growth. |
| GHR | Growth hormone receptor Precursor (GHR) | Average back fat | 0.025 | 65 | 204 | 255 | 1.626cm | Encodes the receptor that binds the growth hormone and activates intercellular signals that lead to growth. |
| SLC45A2 | Solute carrier family 45, member 2 | Grade fat | 0.0126 | 7 | 60 | 461 | 0.9835 | Related to melanin production and in humans the gene is associated with hair, skin and eye pigmentation. In animals it is also involved in pigmentation, lipid metabolism and growth. |

[z] *Details of SNP position and type are shown in Table 5.3.*

[y] *Referred from genecards (Rebhan et al., 1997), gene ontology databases (Ashburner et al., 2000)*

**Table 5.3: Details of significant SNPs associated with carcass traits**

| *Gene symbol* | *SNP ID*[x] | *SNP Chr:bp* | *Type of SNP* | *Position of SNP in gene* | *A.A Change* | *Position of A.A in gene* |
|---|---|---|---|---|---|---|
| UBA5 | NF | 1:139081617 | A/T[y] | 231 | Glu > Val | 13 |
| UBXN4 | rs208513069 | 2:61935185 | T/A[y] | 591 | Leu > His | 195 |
| DGKD | NF | 3:120460419 | A/G[z] | 3067 | Thr > Ala | 1023 |
| ABTB2* | rs211653218 | 15:65558664 | A/C[y] | 2339 | Glu > Ala | 780 |
| PDHX | rs211170349 | 15:66290876 | A/T[y] | 782 | Tyr > Phe | 255 |
| HSD17B12 | rs109711563 | 15:74828355 | A/G[z] | 816 | Arg > Gly | 243 |
| ARHGAP1 | NF | 15:76986446 | C/T[z] | 235 | Pro > Ser | 79 |
| LRP4 | NF | 15:77158469 | T/C[z] | 3410 | Met > Thr | 1137 |
| NDUFS3 | NF | 15:77730420 | C/A[y] | 168 | Ala > Asp | 55 |
| STK10* | rs136660541 | 20:3772213 | C/T[z] | 2226 | Thr > Met | 728 |
| SLC30A5 | rs136504424 | 20:10487612 | A/G[z] | 1666 | Ser > Gly | 556 |
| GHR | NF | 20:33915503 | T/A[y] | 873 | Phe > Tyr | 279 |
| SLC45A2 | NF | 20:42286376 | G/A[z] | 718 | Ala > Thr | 240 |

*[z]Transition. [y]Transversion. [x]SNPs with NF in their ID did not match any SNP in the SNP database*
** The SNP reported in the SNP database differed from the SNP alleles in the population we studied but they were located at the same position in the transcript and amino acid (A.A) sequence.*

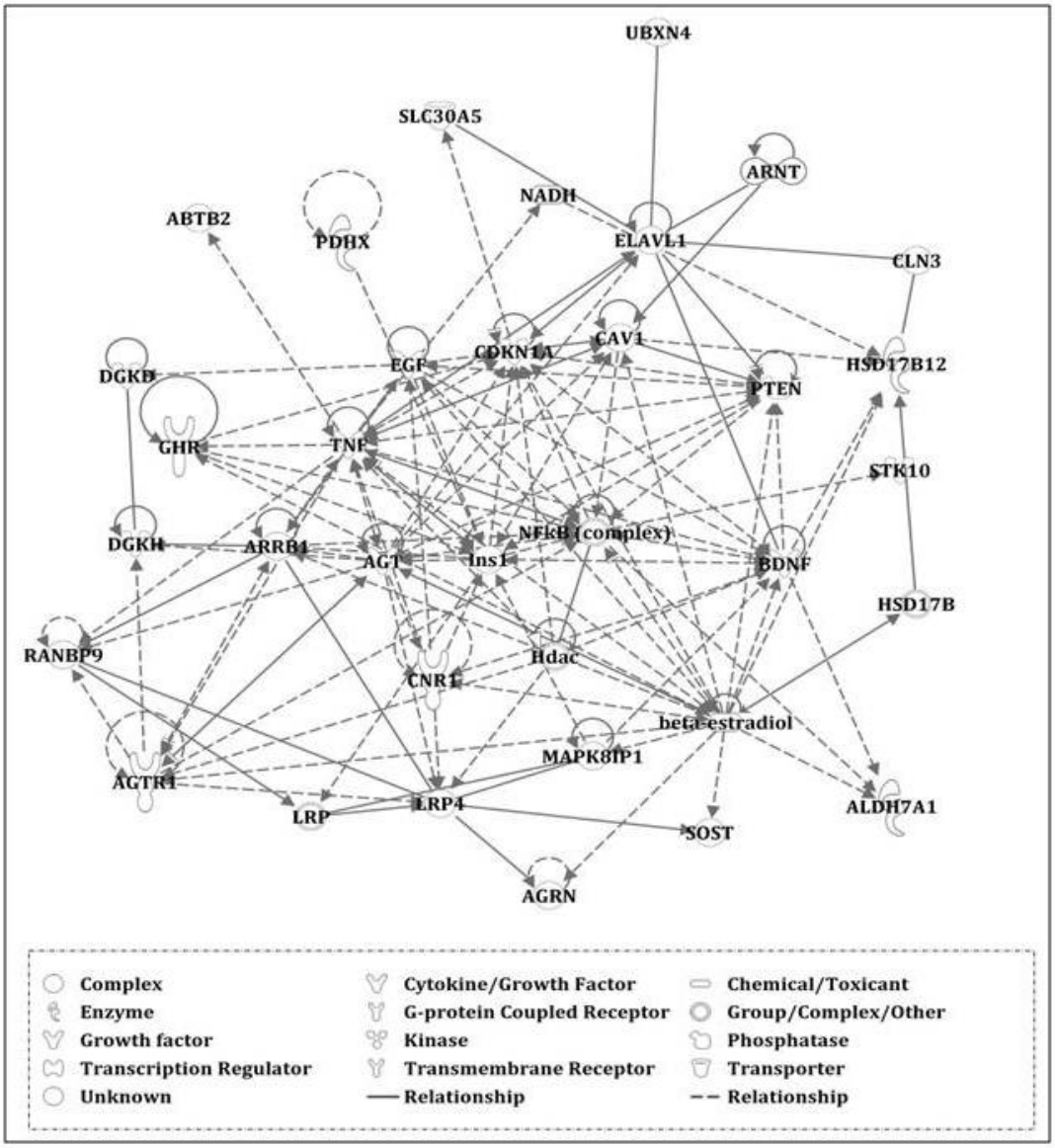



*Figure 5.1a: Gene interaction network 1 associated with carcass quality genes in beef cattle. Genes in the network that were significantly associated with various carcass traits were: PDHX, DGKD, LRP4, ABTB2, SLC30A5, HSD17B12, UBXN4, GHR and STK10.*

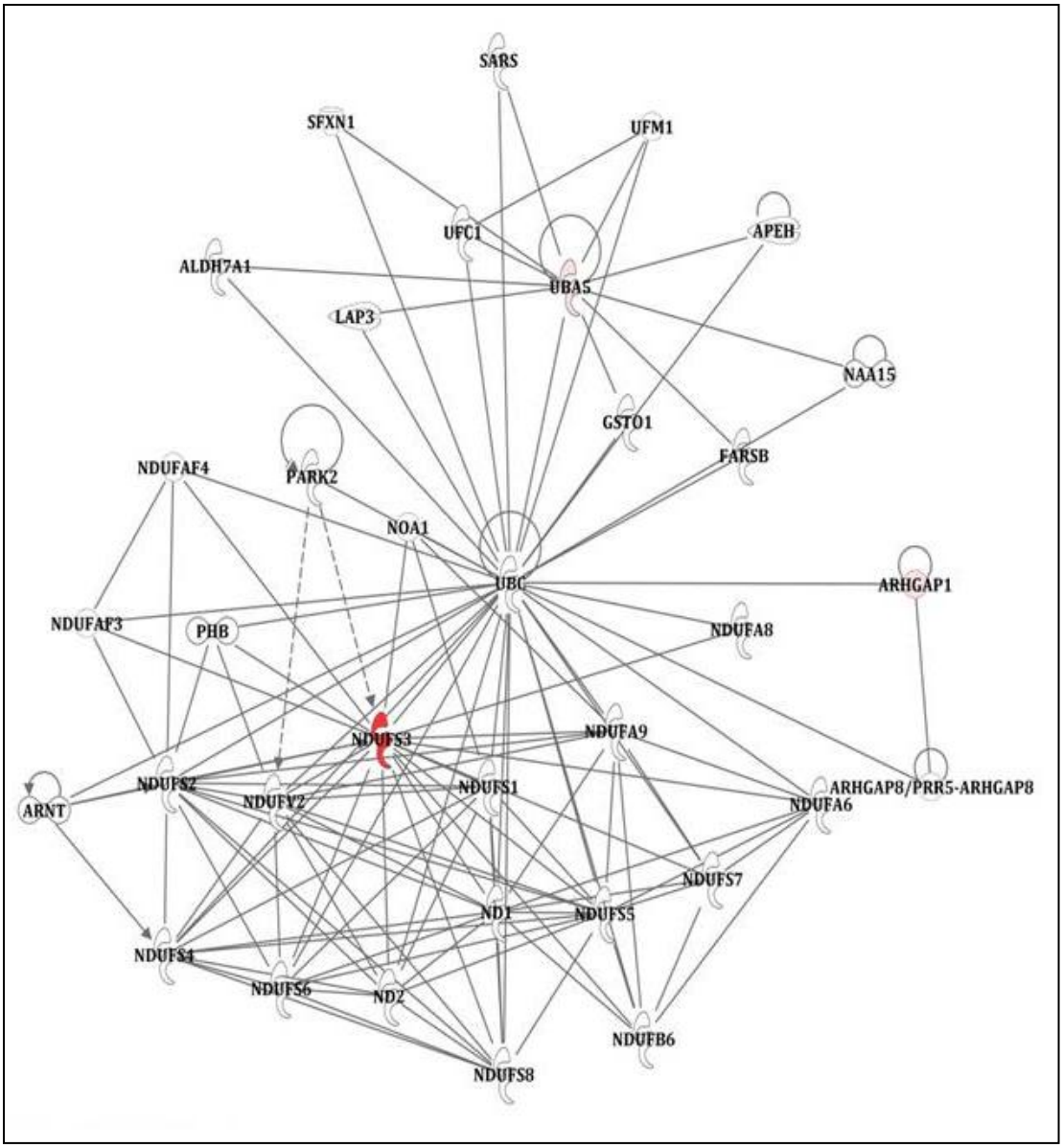



*Figure 5.1b: Gene interaction network 2 associated with carcass quality genes in beef cattle. Genes in the network that were significantly associated with various carcass traits were: NDUFS3, ARHGAP1 and UBA5.*

## 5.6. REFERENCES


Andersson, L. 2003. Identification and characterization of AMPK g3 mutations in the pig Biochemical Society Transactions; 31: 232–235.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, M. J., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. and Sherlock, G. 2000. Gene ontology: tool for the unification of biology. Nat. Genet.; 25(1):25-9.

Baas, A. F., Boudeau, J., Sapkota, G. P., Smit, L., Medema, R., Morrice, N. A., Alessi, D. R. and Clevers, H. C. 2003. Activation of the tumour suppressor kinase LKB1 by the STE20-like pseudokinase STRAD." The EMBO Journal; 22: 3062 – 3072.

Barendse W. 2002. DNA markers for meat tenderness. WO02064820. pp. 1–88.

Basarab, J. A., Price, M. A., Aalhus, J. L., Okine, E. K., Snelling, W. M. and Lyle, K. L. 2003. Residual feed intake and body composition in young growing cattle. Can. J. Anim. Sci.; 83:189-204.

Berthezène, F. 1999. Estrogens and lipid metabolism. Therapie; 54(3):393-5.

Branden, C. I and Tooze, J. 1999. Introduction to protein structure. Edition 2. Published by Garland Publishing, Inc. 19 Union Square west, New York.

Brown, M. S. and Goldstein, J. L. (1997). "The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor". Cell; 89 (3): 331–40.

Calvano, S. E., Xiao, W., Richards, D. R., Felciano, R. M., Baker, H. V., Cho, R. J., Chen, R. O., Brownstein, B. H., Cobb, J. P., Tschoeke, S. K., Miller-Graziano, C., Moldawer, L. L., Mindrinos, M. N., Davis, R. W., Tompkins, R. G., Lowry, S. F. 2005. A network-based analysis of systemic inflammation in humans. Nature; 437(7061):1032-7.

Canadian Council on Animal Care. 1993. Guide to the care and use of experimental animals. Vol. 1. E. D. Olfert, B. M. Cross, and A. A McWilliams, eds. 2nd ed. CCAC, Ottawa, ON.

Casas, E., White, S. N., Wheeler, T. L., Shackelford, S. D., Koohmaraie, M., Riley, D. G., Chase C. C. Jr., Johnson, D. D. and Smith, T. P. L. 2006. Effects of *calpastatin* and μ-*calpain* markers in beef cattle on tenderness traits. J. Anim. Sci.; 84:520-525.

Ciobanu, D. C., Bastiaansen, J. W. M., Lonergan, S. M., Thomsen, H., Dekkers, J. C. M., Plastow, G. S., and Rothschild M. F. 2004. New alleles in calpastatin gene are associated with meat quality traits in pigs. J. Anim. Sci.; 82:2829-2839.

Connolly, M. A., Clausen, P. A. and Lazar, J. G. 2006. Purification of RNA from Animal Cells Using Trizol. Cold Spring Harb Protoc (pdb.prot4104).

Crews, D. H., Jr. 2008. Genetic prediction of feed efficiency and input components. Pages 11–20 in Prediction of Genetic Merit of Animals for Selection, 9th Genetic Prediction Workshop, Beef Improve. Fed., Kansas City, MO. Beef Improvement Federation (BIF), Raleigh, NC.

Dekkers, J. C. M., and Hospital, F. 2002. Utilization of molecular genetics in genetic improvement of plants and animals. Nature Reviews: Genetics 3: 22-32.

Duchesne, A., Gautier, M., Chadi, S., Grohs, C., Floriot, S., Gallard, Y., Caste, G., Ducos, A. and Eggen, A. 2006. Identification of a doublet missense substitution in the bovine LRP4 gene as a candidate causal mutation for syndactyly in Holstein cattle. Genomics; 88:610–21.

Duncan, E. L. and Brown, A. M. 2010. Genetic Determinants of Bone Density and Fracture Risk State of the Art and Future Directions, J. Clin. Endocrinol. Metab. 95:2576-2587.

Fan, B., Onteru, S. K., Plastow, G. S. and Rothschild, M. F. 2009. Detailed characterization of the porcine MC4R gene in relation to fatness and growth. Anim Genet.; 40(4):401-9.

Finch, V. A., Bennett, I. L. and Holmes, C. R. 1984. Coat color in cattle: effect on thermal balance, behavior and growth, and relationship with coat type. The Journal of Agricultural Science; 102: 141- 147.

Fox, C.S., Heard-Costa, N., Cupples, L.A., Dupuis, J., Vasan, R.S. and Atwood, L.D. 2007. Genome-wide association to body mass index and waist circumference: the Framingham Heart Study 100K project. BMC Med Genet 19; 8 Suppl 1:S18.

Gilmore, T. D. (editor) 2006. NF-kB: from basic research to human disease. Oncogene (Reviews); 51: 6679-6899.

Gilmour, A. R., Gogel, B. J., Cullis, B. R., and Thompson, R. 2009. ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Goonewardene, L. A., Wang, Z., Price, M. A., Yang, R.-C., Berg, R. T. and Makarechian, M. 2003. The effect of udder type and calving assistance on weaning traits of beef and dairy × beef calves. Livest. Prod. Sci.; 81:47–56.

Grant, J. R., Arantes, A. S., Xiaoping, L. and Stothard, P. 2011. In-depth annotation of SNPs arising from resequencing projects using NGS-SNP. Bioinformatics; 27:2300-2301.

Gunilla, L., E. Ann-Charlotte, S. Gertrud Von, J. Åsa, H. Ingela, A. J. Henrik, B. Martin, A. Leif and L. Kerstin. 2004. A second mutant allele (V199I) at the PRKAG3 (RN) locus—II. Effect on color characteristics of pork loin. Meat Science. 66 (3): 621-627.

Gunnarsson, U., Hellström, A. R., Tixier-Boichard, M., Minvielle, F., Bed'hom, B., Ito, S., Jensen, P., Rattink, A., Vereijken, A. and Andersson, L. 2007. Mutations in SLC45A2 cause plumage color variation in chicken and Japanese Quail. Genetics; 175:867–877.

Han, S. H., Cho, I. C., Kim, J. H., Ko, M. S., Jeong, H. Y., Oh, H. S. and Lee, S. S. 2009. A*GHR* polymorphism and its associations with carcass traits in Hanwoo cattle, Genes & Genomics; 31 (1): 35-41.

Herd, R. M., Bishop, S. C. 2000. Genetic variation in residual feed intake and its association with other production traits in British Hereford cattle. Livestock production science; 63 (2):111 -119.

Herd, R. M. and Arthur, P. F. 2009. Physiological basis for residual feed intake. Journal of Animal Science, vol. 87 no. 14 suppl E64-E71.

Hu, Z., Fritz, E. R. and Reecy. J. M. 2007. AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond. Nucleic Acids Research, 35 (Database issue):D604-D609.

Hubbard, T. J. 2009. Ensembl 2009. Nucleic Acids Res; 37: 690-697.

Kim, N. K., Cho, Y. M., Jung, Y. S., Kim, G. S., Heo, K. N., Lee, S. H., Lim, D., Cho, S., Park, E. W. and Yoon, D. 2009. Gene Expression Profiling of Metabolism-Related Genes between Top Round and Loin Muscle of Korean Cattle (Hanwoo) J. Agric. Food Chem.; 57: 10898–10903.

Kim, S. K., Larsen, N., Short, T., Plastow, G. and Rothschild, M. F. 2000. A missense variant of the porcine melanocortin-4 receptor *(MC4R)* gene is associated with fatness, growth, and feed intake traits. Mammalian genome; 11: 131-135.

Koch, R. M., Swiger, L. A., Chambers, D. and Gregory, K. E. 1963. Efficiency of feed use in beef cattle. J. Anim. Sci.; 22:486–494.

Li, H., Ruan, J., and Durbin, R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome res.; 18 (11): 1851-1858.

MacNeil, M. D. Nugent, R. A. and Snelling, W.M. 1997. Breeding for Profit: an Introduction to Selection Index Concepts. Range Beef Cow Symposium. Paper 142.

McDonagh, M. D. 1998. Calpain activity affects muscle protein turnover and meat tenderisation. PhD Thesis, The University of New England, Armidale, Australia.

McDonagh, M. D., Herd, R. M., Richardson, E. C., Oddy, V. H., Archer, J. A. and Arthur, P.F. 2001. Meat quality and the calpain system of feedlot steers following a single generation of divergent selection for residual feed intake. Aust. J. Exp. Agric.; 41:1013–1021.

Merida, I., Avila-Flores, A. and Merino, E. 2008. "Diacylglycerol kinases: at the hub of cell signaling.". Biochem. J.; 409 (1): 1–18.

Mirzaei, H. R., Verbyla, A. P., and Pitchford, W. S. 2011. Joint analysis of beef growth and carcass quality traits through calculation of co-variance components and correlations. Genetics and Molecular Research 10 (1): 433-447.

Mogens, V., Kristen, S., John, F., Signe, K. and Bauman, D. E. 1993. The Effect of Bovine Growth Hormone on Growth, Carcass Composition and Meat Quality of Dairy Heifers', Acta Agriculturae Scandinavica, Section A - Animal Science; 43: 3, 165 - 172.

Moon, Y. A. and Horton, J. D. 2003. Identification of two mammalian reductases involved in the two-carbon fatty acyl elongation cascade. J. Biol. Chem.; 278:7335-734.

Morgan, J. B., Wheeler, T. L., Koohmaraie, M., Savell, J. W., Crouse, J. D.1993. Meat tenderness and the calpain proteolytic system in longissimus muscle of young bulls and steers. Journal of Animal Science; 71: 1471–1476.

Mujibi, F. D., Moore, S. S., Nkrumah, J. D., Wang, Z. and Basarab, J. A. 2010. Season of testing and its effect on feed intake and efficiency in growing beef cattle. J Anim Sci.; 88(12):3789-99.

Mujibi, F. D. N., Nkrumah, J. D., Durunna, O. N., Stothard, P., Mah, J., Wang, Z., Basarab, J., Plastow, G., Crews, D. H. Jr., and Moore S. S. 2011a. Accuracy of genomic breeding values for residual feed intake in crossbred beef cattle. J Anim Sci November; 89:3353-3361.

Mujibi, F. D. N., J. D. Nkrumah, O. N. Durunna, J. R. Grant, J. Mah, Z. Wang, J. Basarab, G. Plastow, D. H. Crews, Jr. and S. S. Moore. 2011b. Associations of marker panel scores with feed intake and efficiency traits in beef cattle using pre-selected single nucleotide polymorphisms. J Anim Sci.; 89 (11): 3362-71.

Nalaila, S. M., Stothard, P., Moore, S. S., Wang, Z., and Li, C. 2011. Whole genome fine mapping of QTL for ultrasound and carcass merit traits in beef cattle. Canadian Journal of Animal Science; 91(1): 61-73.

Nkrumah, J. D., Basarab, J. A., Price, M. A., Okine, E. K., Ammoura, A., Guercio, S., Hansen, C., Li, C., Benkel, B., Murdoch, B. and Moore, S. S. 2004. Different measures of energetic efficiency and their phenotypic relationships with growth, feed intake, and ultrasound and carcass merit in hybrid cattle. J. Anim. Sci.; 82:2451.

Nkrumah, J. D., Sherman, E. L., Li, C., Marques, E., Crews, D. H., Jr., Bartusiak, R., Murdoch, B., Wang, Z., Basarab, J. A. and Moore, S. S. 2007. Primary genome scan to identify putative quantitative trait loci for feedlot growth rate, feed intake, and feed efficiency of beef cattle. J. Anim. Sci.; 85:3170.

Rebhan, M., Chalifa-Caspi, V., Prilusky, J. and Lancet, D. 1997. GeneCards: integrating information about genes, proteins and diseases. Trends Genet. 13 (4): 163.

Richardson, E. C., Herd, R. M., Archer, J. A., Woodgate, R. T., Arthur P. F. 1998. Steers bred for improved net feed efficiency eat less for the same feedlot performance. Anim. Prod. Aust.; 22: 213–216.

Rios-Utrera, A., Cundiff, L.V., Gregory, K.E., Koch, R.M., Dikeman, M.E., Koohmaraie, M. and Van Vleck, L.D. 2005. Genetic analysis of carcass traits of steers adjusted to age, weight, or fat thickness slaughter endpoints. J Anim Sci.; 83 (4): 764-776.

Sean R. G. and Desjardins, P. R. 2007. APPENDIX 3D Quantitation of DNA and RNA with Absorption and Fluorescence Spectroscopy (NanoDrop protocols) Published Online: 1 APR 2007.

Sherman, E. L., Nkrumah, J. D., Murdoch, B. M., Li, C., Wang, Z., Fu A. and Moore S. S. 2008. Polymorphisms and haplotypes in the bovine neuropeptide Y, growth hormone receptor, ghrelin, insulin-like growth factor 2, and uncoupling proteins 2 and 3 genes and their associations with measures of growth, performance, feed efficiency, and carcass merit in beef cattle. Journal of animal science, 86:1-16.

Sherman, E. L., Nkrumah, J. D., Li, C., Bartusiak, R., Murdoch, B. and Moore, S. S. 2009. Fine mapping quantitative trait loci for feed intake and feed efficiency in beef cattle. J. Anim Sci.; 87:37-45.

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., and Sirotkin, K. 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29 (1): 308–311.

Tong, J., Zhu, M. J., Underwood, K. R., Hess, B. W., Ford S. P. and Du, M. 2008. AMP-activated protein kinase and adipogenesis in sheep fetal skeletal muscle and 3T3-L1 cells J. Anim. Sci.; 86 (6): 1296-1305.

Uniprot Consortium. 2012. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res.; 40: D71-D75.

Unnur, S., Bjarni, V. H., Solveig, G., Daniel, F. G., Bragi, W., Thorvaldur, I., Thorbjorg, J., Jona, S., Steinunn, S., Jacqueline, R. C., Tuan, V. N., Peter, A., Jeffrey, R. G., John, A. E., Claus, C., Gunnar, S., Augustine, K., Unnur, T. and Kari, S. 2008. New sequence variants associated with bone mineral density. Nature Genetics; 41: 15 – 17.

Voisey, J. and van Daal, A. 2002. Agouti: From Mouse To Man, From Skin To Fat. Pigment Cell Res.; 15: 10-18.

Wang, L., Yang, L., Debidda, M., Witte, D. and Zheng, Y. 2007. Cdc42 GTPase-activating protein deficiency promotes genomic instability and premature aging-like phenotypes. Proc. Natl. Acad. Sci. USA. 104 (4):1248-53.

Wessel, T., Haan, A., Laarse, W. J. and Jaspers, R. T. 2010. The muscle fiber type–fiber size paradox: hypertrophy or oxidative metabolism? Eur J Appl Physiol. 110(4): 665–694.

Wu-Feng, L., Jun-Ya, L., Xue, G., Shang-Zhong, X. and Wen-Bin, Y. 2012. Association analysis of PRKAG3 gene variants with carcass and meat

quality traits in beef cattle. African Journal of Biotechnology; 11(8): 1855-1861.

## *PREFACE OT CHAPTER SIX*

*Post translational regulation of phenotypes can be expressed as variation in metabolite levels in blood. These metabolites, if highly correlated with the trait, may be used as intermediate phenotypes to predict the phenotype of interest.*

*The null hypothesis; that there would be no significant association between metabolite levels and RFI in beef cattle, was tested in Chapter 6 with the intention of proposing the use of metabolites as biomarkers for the prediction of RFI in beef cattle.*

*This Chapter has been submitted to Livestock Science for publication*

# CHAPTER SIX

## Metabolite bio-markers associated with feed efficiency in beef cattle with divergent residual feed intake[4]

### 6.1. INTRODUCTION

RFI is a measure of feed efficiency defined as the difference between an individual's actual feed intake and the predicted feed intake (Koch et al. 1963). The predicted feed intake in beef cattle is estimated based on the individual's maintenance of body weight (BW) and growth (average daily gain-ADG). Therefore RFI is phenotypically independent of both BW and ADG. Selecting cattle for RFI could be implemented using phenotypic measurements. However, estimation of RFI requires measurement of individual animal feed intake, which is expensive and limited to the capacity of the recording equipment (Williams, 2010). RFI is also moderately heritable (Arthur et al. 2001) making it a good candidate for marker assisted selection with DNA or other predictive markers could be used in selection schemes (Moore et al. 2009). In reference to genetic markers, variation in several genes have been reported to influence RFI variation but their effects are either breed or population specific and are not always reproducible in genetically diverse populations (Sherman et al. 2010).

[4] A version of this chapter has been submitted to Livestock Science for publication (2013)

The discovery of metabolites associated with traits of importance in livestock such as plasma levels of carnitine and body weight in cattle (Weikard et al. 2010), indicates a potential for using metabolites as biomarkers to predict RFI. Metabolite marker assisted selection may offer an approach to be used as an alternative or complementary to genetic marker assisted selection. The significant metabolites may also be used to discover novel pathways influencing the variation in feed efficiency. Metabolite based selection techniques have the potential to be refined making them relatively easy and fast to use.

In addition, combining metabolite association studies with genetic association studies to develop triple association analysis between genotypes, metabolites and phenotypes as used by Weikard et al. (2010) to identify genotypes and metabolites associated with growth and lipid metabolism in cattle, may result in the identification of even more important physiological pathways involved in the regulation of RFI. This approach was also used to identify metabolites associated with specific genotypes in humans (Kettunen et al. 2012). As a selection tool, metabolite profiles have also been used to predict levels of phenotypes such as body mass in birds (Jenni-Eiermann and Jenni, 1994), bird growth rates (Albano et al. 2011), growth and body composition in sheep (Hegarty et al. 2006) and diseases such as diabetes (Kulkarni, 2012).

In this study, we used the nuclear magnetic resonance (NMR) technique to assess the levels of metabolites in plasma followed by association analysis to identify the metabolites associated with RFI in beef cattle. The significant metabolites were then used to reconstruct metabolic networks using IPA to analyze the interaction

between the significant metabolites. We also identified the biological pathways associated with the metabolites and made some hypotheses on the role they play in the regulation of feed efficiency.

## 6.2. MATERIALS AND METHODS

### 6.2.1. The animal resources

The composition and management of the steers used as the discovery population this study was described by Montanholi et al. (2010). The steers belonged to the University of Guelph in Guelph, Ontario, Canada and were cared for according to the guidelines of the Canadian Council of Animal Care, (1993). The project was approved by the University of Guelph animal care committee. The breed composition of this herd was described in detail by Mader et al. (2009) as crossbreeds composed of Angus, Limousin, Charolais, and Simmental as the major contributing breeds. The entire population entered into the feeding test consisted of 91 steers (46 in year 1 and 45 in year 2) and the average initial weight for the steers was 313±6.2kg.

Blood samples were collected through jugular venipuncture using a 10 ml blood collection tube (VacuntainerR; BD Inc., Franklin Lakes, NJ, USA) containing sodium heparin mounted with a 20 ga needle. Blood samples were immediately stored on ice until centrifugation (3,0003g for 20 min) to separate the blood plasma, which was stored at -80$^{o}$C until RFI was estimated.

### 6.2.2. RFI estimation

The statistical procedures used to estimate RFI in the discovery population were reported by Montanholi et al. (2010). In summary, the feeding tests were conducted for 140 days during which individual average daily feed intakes were recorded using an automated feeding system. Daily dry matter intake (DMI) over the testing period was calculated as a product of the daily intake in as fed basis and the percent dry matter in the feed. Average daily gain (ADG) was determined by a regression of body weight (BW) on days on trial. Mid-trial BW was calculated by computing the animals' intercept plus the ADG times 70 (half of the experimental period).

RFI was estimated for each steer using the formula described by Koch et al. (1963) (Equation 1).

$$RFI = DMI - \beta_0 + \beta_1 (ADG) + \beta_2 (BW) \quad \text{.....Equation 1}$$

Where $\beta_0$ was the intercept, $\beta_1$ and $\beta_2$ were the regression coefficients on ADG and on mid-trial body weight (BW) respectively.

The $R^2$ observed for this regression was 0.59 and 0.72 for year 1 and 2, respectively (Montanholi et al. 2010).

When the RFI values for each steer were available, the plasma samples were initially sorted into 2 batches according to the RFI value of the steers tested so that one batch consisted of steers with high RFI and the other batch consisted of steers with low RFI. The plasma samples were then sorted to establish 3 sets of samples from each batch (making a total of 6 sets) according to the period when the blood samples were collected. The three time periods considered in this study

included period 1, which corresponded to the second week after the beginning of the feeding test period, and period 2 and 3, corresponding to 6 and 9 weeks into the feeding period respectively.

A subset of 16 plasma samples collected from steers with the highest RFI values in the high RFI batch and 16 samples from steers with the lowest RFI values from the low RFI batch were sampled in each period for metabolomics analysis.

The size of the sample used in this study was estimated using the formula:

$n = \frac{2}{d^2} * C_p, power$ (Whotley and Ball, 2002)

Where $n$ is the number of subjects required in each group, $d$ is the standardized difference and $C_p, power$ is a constant defining the values chosen for p value and power. In this study the constant was estimated using a p value of 0.05 and power of 80% and was equal to 7.9 (Whitley and Ball, 2002).

The standardized difference ($d$) was estimated as:

$d = \frac{Target\ difference}{Standard\ deviation}$. The value of $d$ was set at 1 such that the target/detectable difference would be equal to standard deviation of RFI in the population. Therefore the sample size was estimated as $n = \frac{2}{1^2} * 7.9 = 15.8$ rounded off to 16 steers. A total of 96 samples were used in the metabolomic analysis.

### 6.2.3. Metabolite assays using NMR in beef cattle

Serum global metabolite concentrations were determined using Nuclear Magnetic Resonance (NMR) at the Chenomx® NMR facility at the University of Alberta,

Canada. This procedure is novel because, although it has been used and validated more frequently in humans and laboratory animals, this study is one of the first to apply it in samples obtained from cattle. The NMR analysis involved three steps namely sample preparation, spectra acquisition and sample fitting.

*Sample preparation*

Samples were filtered through 3kDa molecular weight cut-off filters (Nanosep 3K Omega microcentrifuge filter tubes; Pall Corporation, NY, USA) to remove macromolecules, including lipids and proteins. As the filter tube manufacturer treats the filter membranes with glycerol as a preservative, filters were either soaked overnight in distilled water, or rinsed with distilled water and centrifuged three times before use. Samples comprising of less than 530 µl after filtration were diluted with 30mM KH2PO4 solution (pH7) to ensure adequate volume for NMR acquisition.

5mm NMR tube (New Era Enterprises Inc., NJ, USA) contained a total of 600µl of liquid with the contents as follows: 60µl of Chenomx internal standard solution IS-2 (contains 5.3945 mM DSS-d6, 0.2 %w/v NaN3 in D2O) which is used for metabolite quantification, 10µl of Formate (pH7.0, added for lineshape correction) and 530µl of sample + phosphate buffer. This mixture was vortexed for 30s before it was transferred to an NMR tube for data acquisition. All metabolite concentrations obtained were adjusted by appropriate factors to account for the above dilutions, and represent the contents of the filtered samples, not the contents of the NMR tube.

*Spectra acquisition*

Spectra were acquired on a 600 MHz VNMRS spectrometer equipped with a 5mm HX probe (Agilent Technologies, CA, USA). The pulse sequence used was a 1D-tnnoesy with a 990 ms presaturation on water and a 4 s acquisition period. Spectra were collected with 256 transients and 4 steady-state scans at 298 K.

*Sample fitting*

Spectra were processed using the Processor module in Chenomx NMR Suite 7.5 software (Chenomx, Edmonton, Alberta, Canada). Spectra were zero filled to 64k points and Fourier transformed. Spectral phasing was performed on the spectra along with baseline correction (Chang et. al, 2007). Metabolites were identified and quantified with a targeted profiling approach using the Profiler and Library Manager modules in the same software which contains 304 total metabolites (Weljie et. al, 2006). Each spectrum was reviewed by at least two different analysts. A final review pass was done on all of the spectra before exporting concentration results. Concentration measurements were adjusted to report metabolite concentrations after filtration of the samples.

### 6.2.4. Statistical analysis

Association analysis between RFI and the metabolite levels was performed using SAS 9.1 (SAS Institute Inc. Cary, NC, USA) for each period separately using a multiple regression model and for the combined data at the three periods using a mixed model for repeated measures analysis.

*Metabolic networks*

To assess the molecular interactions between significant metabolites in each period, metabolic networks were reconstructed for each period using IPA software (Ingenuity Systems, www.ingenuity.com). To reconstruct the metabolic networks, a list of significant metabolites and their corresponding P-values were imported into the IPA software and the parameters were set to allow the network to include indirect relationships between the imported metabolites and metabolites in the knowledge base. Indirect relationships would assist in the identification of other metabolites that were not among the ones analyzed but may be associated with RFI. The IPA algorithm generated metabolic networks by mapping each metabolite identifier to its corresponding metabolite in the IPA knowledge base as described in details by Calvano et al., (2005). The metabolites are then overlaid onto a global molecular network developed from information contained in the knowledge base. Each network is assigned a significance score, which represents the likelihood that the imported metabolites within the network are found therein by random chance (Calvano et al., 2005). A high number of imported metabolites within a dataset lead to a higher network score. The network score is calculated as the negative of the exponent of the P-value such that a score of 25 will be equal to a P-value of $10^{-25}$ (Calvano et al., 2005) and therefore larger scores correspond to increased significance.

### 6.2.5. Validation analysis

To validate the results obtained in the discovery population, blood was collected from 10 Angus steers with high RFI and 10 Angus steers with low RFI from the University of Alberta beef ranch at Kinsella, Canada. The blood samples were collected at two time points corresponding to period 1 and 2 in the analysis conducted on the discovery population. Blood samples were not collected from the time period corresponding to period 3 because it was only one week to the end of the feeding test period, and, period 3 would be considered too late into the feeding test period and metabolites identified at this period would not offer significant additional information in predicting RFI in comparison to period 2.

The metabolite concentrations were determined using Nuclear Magnetic Resonance, NMR, at the Chenomx® NMR facility at the University of Alberta, Canada, as described above.

A multiple regression analysis was performed using SAS 9.1. (SAS Institute Inc. Cary, NC, USA) to identify the metabolites significantly associated with RFI in each of the two periods.

## 6.3. RESULTS

### 6.3.1. Discovery analysis

Multiple regression analysis showed that at period 1, two metabolites (creatine and glycine) were significantly ($P < 0.05$) associated with RFI and accounted for 36.3% of the phenotypic variation in RFI (Table 6.1, Fig 6.1A). In period 2, 10 metabolites were significantly ($P < 0.05$) associated with RFI and accounted for 74.2% of the phenotypic variation in RFI (Table 6.1, Fig 6.1B) and at period 3, 3

metabolites were significantly ($P < 0.05$) associated with RFI and accounted for 52.1% of the phenotypic variation (Table 6.1, Fig 6.1C). Creatine was the only metabolite that was significant in more than one period; being significant in both period 1 and 2.

In a mixed model for repeated measures, 2 metabolites (citrate and creatine) were associated with RFI across all periods and accounted for 33.2% of the phenotypic variation in RFI (Table 6.1, Fig 6.1D).

The concentration of all the significant metabolites fluctuated between periods (Figure 6.2 indicating an interaction between metabolite levels and the animal's developmental stage. The significant metabolites in the discovery population accounted for the highest phenotypic variation during period 2 although this R-square may be biased due to the usage of a larger number of metabolites in period 2 than the rest of the periods as shown in Table 6.1.

### 6.3.2. Validation analysis

The results of the validation analysis indicated that 8 metabolites were significantly ($P < 0.05$) associated with RFI in period 1 with a cumulative R-square of 75.3% (Table 6.2). Another metabolite; Trimethylamine was not significant ($P = 0.06$) but explained 23% of the variation therefore was considered relevant. None of the metabolites that were significant at period 1 in the discovery population was also significant in the validation population. However, creatinine, which was significant in the validation population, is a product of the metabolism of creatine, which was significant in the discovery population. Table 6.2 and Fig

6.2 show a summary of the results of association analysis in the validation population in period 1.

In period 2, 12 metabolites were significantly associated with RFI in the validation population explaining 98.4% of the phenotypic variation (Table 6.3). Three of these metabolites (creatine, hippurate and carnitine) were also significant in the discovery population. Together, these 3 metabolites accounted for 32.8% of the phenotypic variation. An additional metabolite glutamine was significantly associated with RFI in the validation population while its metabolic precursor, glutamate, was associated with RFI in the discovery population.

### 6.3.3. Metabolic networks and biological pathways

Using the results of the analysis of the discovery population, at period 1 only two metabolites, creatine and glycine were used and several links were created between these metabolites and other metabolites (Fig 6.3). The metabolic network indicated additional metabolites that may have biological relevance in the regulation of RFI including proinsulin, calpain, cholesterol, amylase and NAD+. The biological processes associated with the metabolites in period 1 included calcium/calmodulin signaling pathways, AMPK signaling, methane metabolism, bile acid biosynthesis, glucose metabolism and cholesterol metabolism.

At period 2, a metabolic network was reconstructed for the 10 metabolites that were significantly associated with RFI. Similar to the network in period 1, proinsulin was also involved in these interactions (Fig 6.4). Other metabolites and genes were the neuropeptide Y receptor Y5 (NPY5R), the MAPK pathway and

kynurenic acid which is a product of normal metabolism of L-tryptophan. The biological processes associated with the metabolites in period 2 were related to amino acyl tRNA biosynthesis, metabolism of serine, glutamate, glycine and threonine, nitrogen metabolism, urea cycle and GABA receptor signaling.

At period 3, the network only retrieved molecules interacting with Tyrosine as shown in figure 6.5. The metabolites formate and hydroxyisobutyrate were not associated with any records in the databases. The biological processes significantly associated with tyrosine at period 3 included the metabolism of nitrogen, pyruvate, methane, tyrosine, amino acid and glyoxylate, protein kinase signaling and dopamine receptor signaling.

### 6.3.4. Predicting RFI using metabolite assays

According to this study, the prediction of RFI levels in steers using blood metabolites is possible and can achieve the highest accuracy if performed at period 2 (6 weeks into the feeding period). The three metabolites that were significant in both discovery and validation population only explained 32% of the variation and the predicted values of RFI based on these metabolites had an average accuracy of prediction of 48% (Fig 6.6). The accuracy of prediction of RFI could be increased by using more metabolites as shown in fig 6.6, where 12 significant metabolites achieved an accuracy of prediction of 99.5% in the validation population.

## 6.4. DISCUSSION

In this study, the NMR assay approach was used to identify blood metabolites associated with RFI. We also investigated the metabolic networks and biological pathways associated with the significant metabolites and proposed the metabolites that could be used to predict RFI with high accuracy.

The results indicated that there are metabolites significantly associated with RFI at different stages during the feeding test period. The concentrations of metabolites fluctuate from one period to the next possibly depending on the underlying biological processes occurring in the animal at that stage of development. Of the greatest importance were the metabolites observed in period 2, which corresponded with week 6 of the feeding period. A total of 10 metabolites were significantly associated with RFI in the discovery population accounting for 74.2% of the phenotypic variation in RFI and three of them (creatine, hippurate and carnitine) were also significant in the validation population accounting for 32% of the phenotypic variation in RFI. In period 2 in the validation population, 9 metabolites accounted for 98.5% of the phenotypic variation in RFI. The R square values for each period may be biased due to the small sample size used in this study and requires to be re-estimated using larger populations of beef steers.

Carnitine was one of the metabolites significantly associated with RFI in period 2. In a similar study, carnitine was identified to be associated with body weight in cross-bred cattle with high carnitine levels associated with increased body weight (Weikard et al., 2010). In this study, the high RFI steers had relatively higher levels of carnitine than the low RFI steers and may indicate a relatively higher

body weight. This is contrary to the expectations of phenotypic RFI being expected to be independent of its component traits (Kennedy et al., 1993). However, the biological mechanisms that influence variation in RFI may also influence body weight resulting in genetic correlations between these traits. The lack of genetic dependence in RFI was reported by Kennedy et al., (1993). Carnitine plays a key role in energy metabolism of cells (Di Lisa et al., 1995; Scholte et al., 1996; Siliprandi et al., 1994) mainly, by transferring acyl groups from cytoplasm to mitochondrial matrix for β-oxidation. In addition, L-carnitine regulates coenzyme-A concentrations in cytosol and mitochondria, glucose and lipid metabolism (Arslan, 2006). These functions render the utilization of energy in feed and body stores more efficient (Arslan, 2006). The role of carnitine, therefore, may be influencing body composition in cattle and as a result influencing body weight. The relationship between RFI and body composition traits was discussed previously. Arthur et al., (2001) and Shenkel et al., (2004) showed that subcutaneous fat depth over the 12th and 13th ribs and rump had a positive genetic correlation with RFI in beef cattle.

Another metabolite which functions as part of the cell's energy shuttle is creatine, which was also significantly associated with RFI in both populations. During creatine metabolism, the high energy phosphate group of ATP is transferred to creatine to form phosphocreatine. Phosphocreatine will then be used to resynthesize ATP from ADP during increased energy demands. A detailed review of the creatine and phosphocreatine system can be obtained from Wallimann et al., (2011). Creatinine, a by-product of creatine metabolism, was reported to be

significantly associated with muscle mass and negatively associated with fat depth in sheep (Clarke et al., 1996). In this study, creatine levels were high in high RFI steers across all periods and in both discovery and validation populations. These levels may indicate an increased need for energy among the less feed efficient steers.

Other metabolites associated with energy metabolism were involved in the TCA cycle. These included the levels of acetate in blood, which were significantly associated with RFI in the discovery population. Acetate levels were higher in low RFI (efficient) steers than high RFI (inefficient) steers in the discovery population. In contrast, acetate levels were lower in the low RFI steers in the validation population. Acetate comprises the majority (70%) of volatile fatty acids produced in the rumen of cattle and is used for lipid metabolism and energy production for skeletal muscles, the heart and kidneys. Acetate is used to produce energy when individuals are in low energy balance and it forms fat when the animal is in high energy balance (Preston and Leng, 1987). Acetate reacts with Coenzyme-A to form acetyl-CoA which enters the TCA cycle (in a step which utilizes ATP) and reacts with oxaloacetate to form citrate (Preston and Leng, 1987). The citrate levels in serum were also significantly associated with RFI in the discovery population; steers with low RFI had less citrate than those with high RFI in both populations.

To explain the levels of acetate and citrate and their correlation with RFI, we suggest two hypotheses; firstly, low citrate levels in the low RFI steers may be a mechanism to reduce the amount of ATP produced by the TCA cycle. Secondly,

the lower citrate levels in low RFI steers may have resulted from reduced levels of acetyl-CoA in an alternative mechanism to reduce ATP production. The second hypothesis cannot be pursued further because the levels of acetate were inconsistent between the discovery and the validation populations.

Hydroxyisobutyrate, another metabolite associated with RFI in the discovery population, is associated with gluconeogenesis from valine and has been shown to be an inter-organ metabolite that preserves the gluconeogenic potential of valine (Letto et al., 1986). It has been shown that hydroxyisobutyrate inhibits key enzymes in energy metabolism in young rats by inhibiting the functions of the respiratory chain complex 1 to 3 and mitochondrial creatine kinase (Viegas et al., 2008). Hydroxyisobutyrate was also associated with high levels of circulating free fatty acids, increased intramyocellular lipid content, impaired insulin-mediated glucose uptake, diminished mitochondrial functioning and an overall weakened metabolic flexibility (Mullen and Ohlendieck, 2010). In this study, high RFI steers (feed inefficient) had a higher concentration of hydroxyisobutyrate than low RFI steers, which may indicate that the efficient steers had less inhibition to the respiratory chain, better energy production, less circulating free fatty acids and reduced intramuscular lipid content.

Some significant metabolites may play a role in energy production through gluconeogenesis such as the gluconeogenic amino acids glycine in period 1, glutamate, phenylalanine, threonine and lysine in period 2 and tyrosine in period 3. In the validation population, the significant amino acids that are gluconeogenic included glutamine, histidine and proline. In mature cattle, most of the amino

acids are deaminated and contribute to energy production (Preston and Leng, 1987). Each amino acid enters the TCA cycle at a different step for example glutamate enters the TCA to form alpha ketoglutarate while threonine forms succinyl CoA. In this study there was no consistent trend in the levels of gluconeogenic amino acids between the high and low RFI steers. The low RFI steers had higher concentration of glutamate, threonine and glycine but a lower concentration of tyrosine, lysine and phenylalanine compared to the high RFI steers.

Several other metabolites were included in the metabolic networks and may play a role in feed efficiency. The neuropeptide Y (NPY) and its receptors (NPYRs) in the metabolic network in period 2 have been implicated in regulation of appetite in humans (Minor et al., 2009) and cattle (Matteri, 2001). A previous genetic association analysis of SNPs located in the Neuropeptide Y gene indicated significant association between these SNPs and ADG, BW, feed efficiency and carcass marbling (Sherman et al., 2008). In this study, NPY5R created a network module with L-glutamic acid and proinsulin, two metabolites which were previously shown to be associated with regulation of appetite (Hermanussen et al., 2006) with glutamic acid down-regulating hypothalamic suppression of appetite and potentially leading to obesity. However, the effect of proinsulin and insulin on regulating appetite has been inconsistent across studies. Lele et al., (2006) showed that proinsulin and leptin were both significantly associated with obesity in humans with obese subjects having higher levels of proinsulin. Other studies have also reported the significant association between Leptin and RFI (Hoque et al.,

2009; Kelly et al., 2009). The biological processes associated with leptin and proinsulin included energy and lipid metabolism.

Another metabolite included into the networks was cholesterol. The metabolism of cholesterol is similar to that of other lipids (Preston and Leng, 1987) and may partially explain the reported negative correlation between RFI and fat composition in steers (Basarab et al., 2003).

The metabolic networks also included calpain, which is involved in post-mortem tenderization of meat. The involvement of calpain may partially explain the previously reported correlation between RFI and beef tenderness (Casas et al., 2006).

The biological processes associated with RFI included methane and nitrogen metabolism. This is consistent with the previously reported correlation between RFI and methane production, where low RFI steers were shown to produce less methane than high RFI steers (Nkrumah et al., 2006). Other biological processes associated with RFI in the metabolic network included AMPK signaling, glucose and cholesterol metabolism. The AMPK signaling pathway is involved in several metabolic processes including glycogen metabolism, glycolysis, gluconeogenesis, sterol synthesis, fatty acid oxidation and lipolysis.

## 6.5. SUMMARY

The metabolomics approach was used to identify metabolites associated with RFI in a discovery and validation population of beef steers with divergent RFI levels. Three metabolites were successfully validated and accounted for 32% of the

phenotypic variation in RFI. An additional set of 9 metabolites were also significantly associated with RFI in the validation population and together with the three metabolites, accounted for 98.5% of the variation in RFI. The 12 metabolites significantly associated with RFI may be used to develop a biomarker based selection tool for RFI in beef cattle. In this study, this tool was estimated to achieve an accuracy of RFI prediction of 99.5%. However, this estimate may be biased due to the small sample size used in this study and requires to be validated using a larger population of beef steers.

**Table 6.1: Metabolites associated with RFI in the discovery population**

| Period | Metabolite | P-value | Partial R-square | Cumulative R-square |
|---|---|---|---|---|
| 1 | Creatine | 0.010 | 0.087 | 0.087 |
| | Glycine | 0.001 | 0.276 | 0.363 |
| 2 | Threonine | 0.044 | 0.035 | 0.035 |
| | Carnitine | 0.008 | 0.042 | 0.077 |
| | Acetate | 0.032 | 0.044 | 0.121 |
| | Creatine | 0.002 | 0.057 | 0.178 |
| | Phenylalanine | 0.003 | 0.067 | 0.245 |
| | Lysine | 0.009 | 0.074 | 0.319 |
| | Citrate | 0.002 | 0.081 | 0.400 |
| | Betaine | 0.036 | 0.087 | 0.487 |
| | Glutamate | 0.001 | 0.102 | 0.589 |
| | Hippurate | 0.026 | 0.153 | 0.742 |
| 3 | Hydroxyisobutyrate | 0.0002 | 0.332 | 0.332 |
| | Tyrosine | 0.014 | 0.047 | 0.379 |
| | Formate | 0.007 | 0.141 | 0.521 |
| Entire | Citrate | 0.006 | 0.193 | 0.193 |
| | Creatine | 0.002 | 0.139 | 0.332 |

**Table 6.2: Metabolites significantly associated with RFI in the validation population in period 1**

| Metabolite | Partial R square | P - value |
|---|---|---|
| Creatinine | 0.2635 | 0.0206 |
| Valine | 0.2993 | 0.0033 |
| Choline | 0.1561 | 0.0088 |
| Histidine | 0.0821 | 0.0251 |
| Uridine | 0.0884 | 0.0048 |
| Dimethylamine | 0.0257 | 0.0686 |
| Trimethylamine | 0.0307 | 0.0230 |
| 2-Hydroxybutyrate | 0.0173 | 0.0445 |
| 3-HydroxybutyrateB | 0.0204 | 0.0056 |
| TOTAL | 0.9835 | |

**Table 6.3: Metabolites significantly associated with RFI in the validation population in period 2.**

| Metabolite | Partial R-Square | P-value |
|---|---|---|
| Creatine | 0.2568 | 0.0226 |
| Histidine | 0.1573 | 0.0475 |
| Succinate | 0.1542 | 0.0295 |
| Oxobutyrate | 0.1297 | 0.0227 |
| 4-Hydroxybutyrate | 0.1006 | 0.0192 |
| Hippurate | 0.0770 | 0.0140 |
| Trans- 4-Hydroxy- L-proline | 0.0483 | 0.0172 |
| Proline | 0.0411 | 0.0042 |
| Allantoin | 0.0094 | 0.0407 |
| Glutamine | 0.0045 | 0.0513 |
| Uridine | 0.0039 | 0.0118 |
| Carnitine | 0.0017 | 0.0002 |
| Total | 0.9845 | |

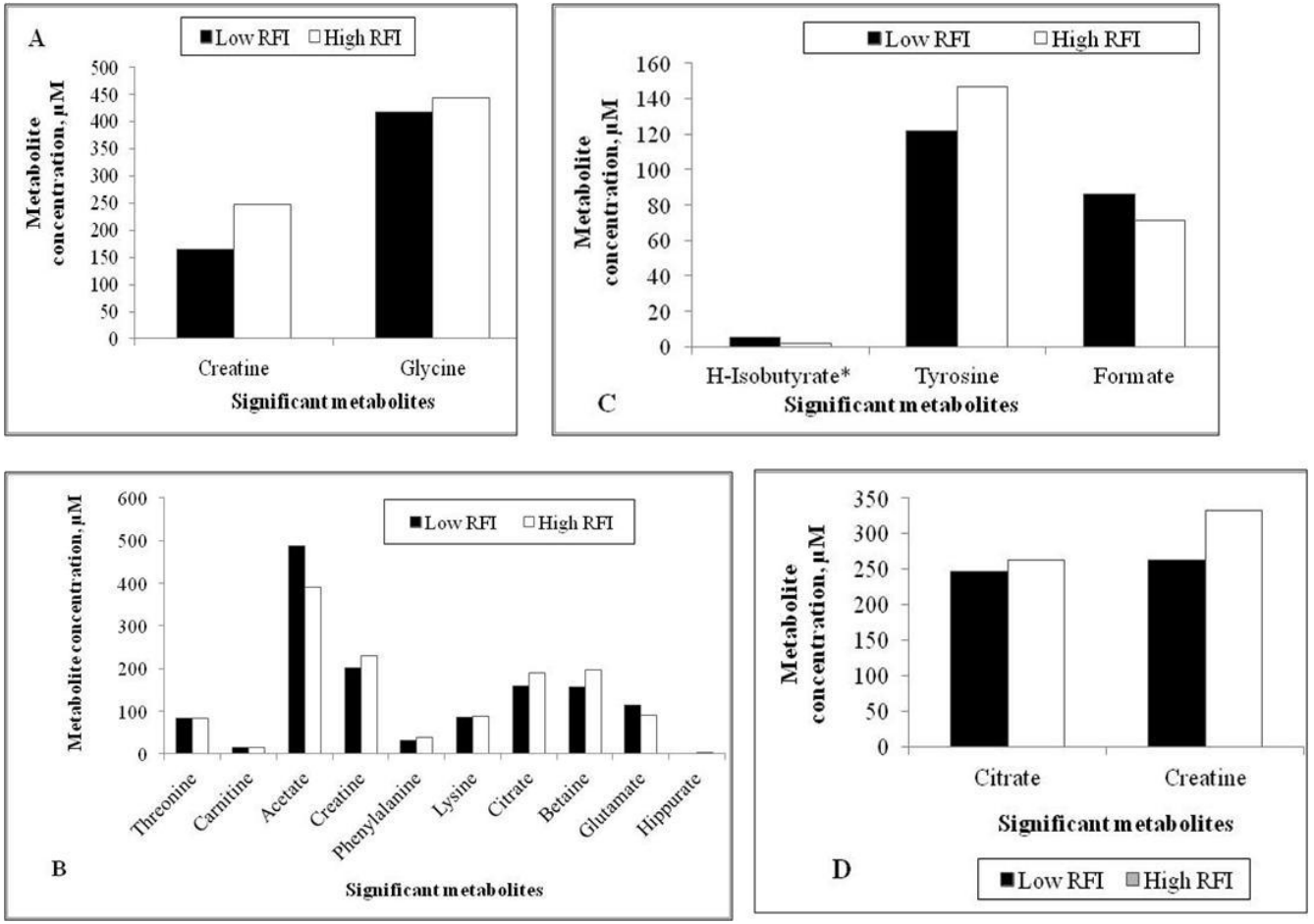



*Figure 6.1: Average concentrations of significant metabolites in µM, in high and low RFI steers in period 1 (A), period 2 (B) period 3 (C) and in the entire period (D) in the discovery population. *Hydroxyisobutyrate*

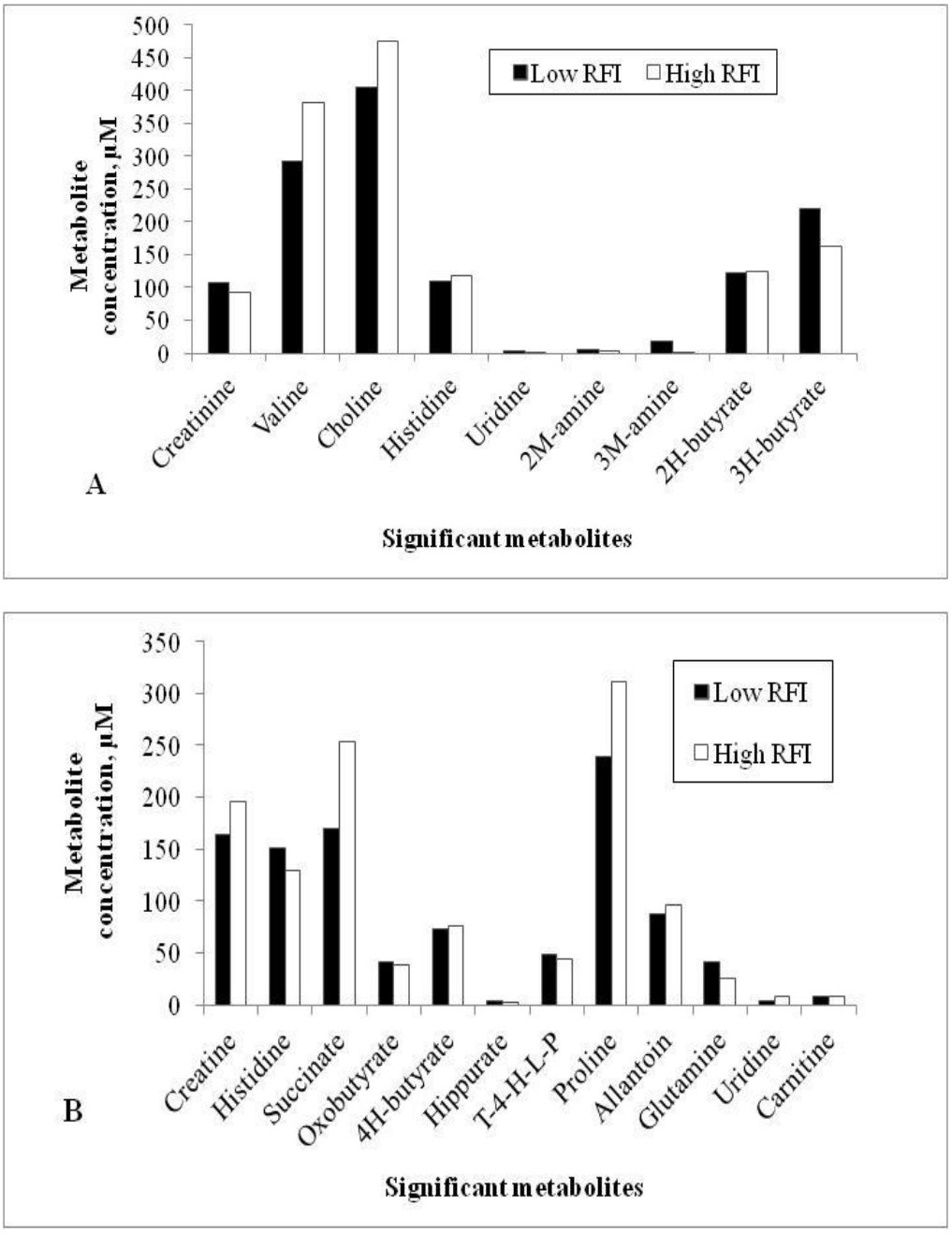



*Figure 6.2: Average concentrations of significant metabolites in high and low RFI steers in period 1 (A*) and period 2(B[#]) in the validation population.*

**2M-amine; Dimethylamine, 3M-amine; Trimethylamine, 2H-butyrate; 2-Hydroxybutyrate, 3H-butyrate; 3-Hydroxybutyrate.*

*[#]4H-butyrate; 4-Hydroxybutyrate, T-4-H-L-P; Trans-4-Hydroxy-L-Proline.*

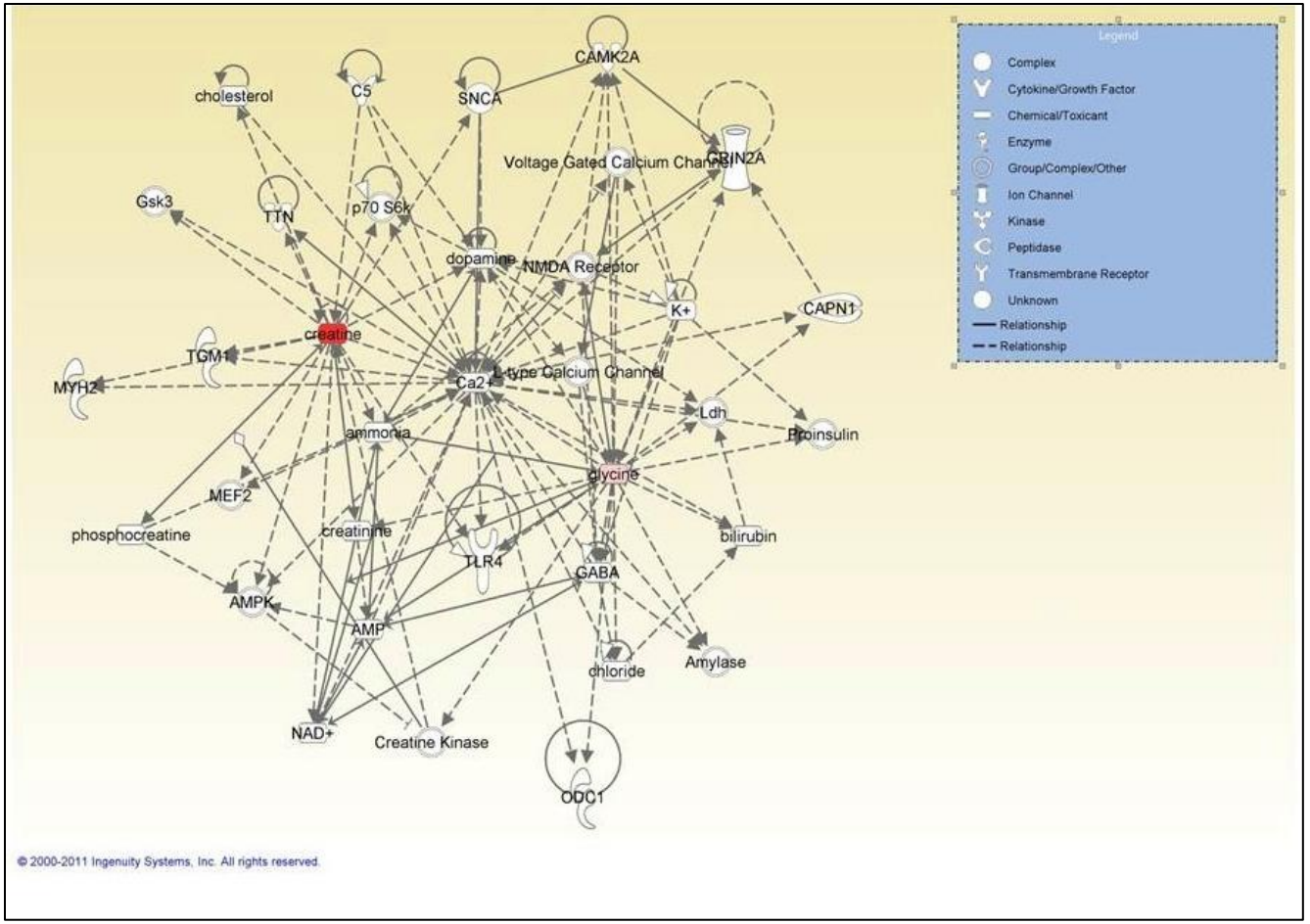



*Figure 6.3: Metabolic network for period 1 in the discovery population*

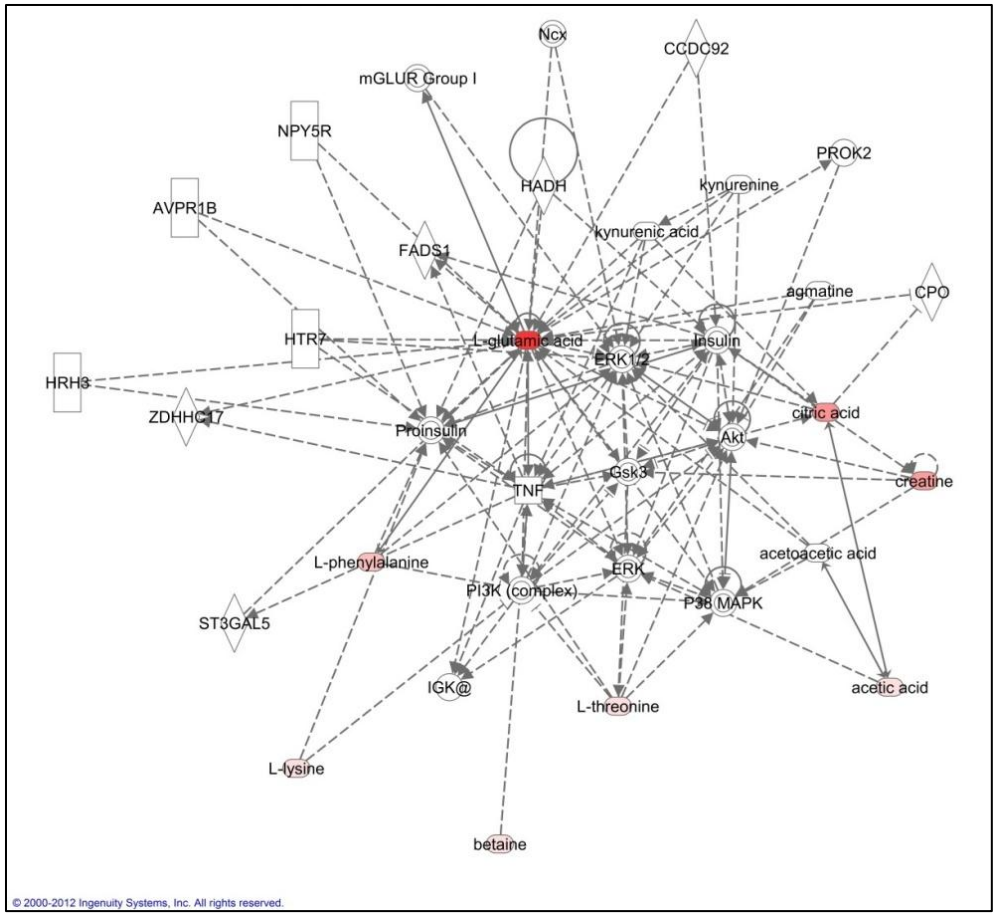



*Figure 6.4: Metabolic network for metabolites significant in period 2 in the discovery population*

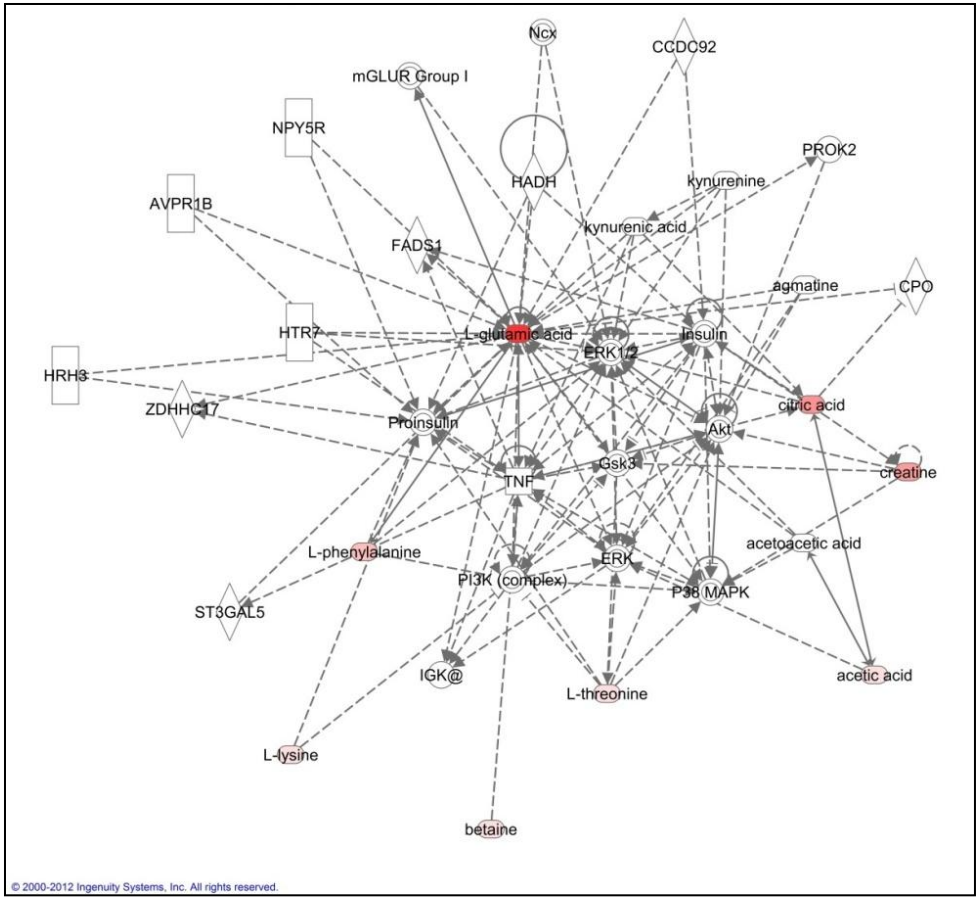
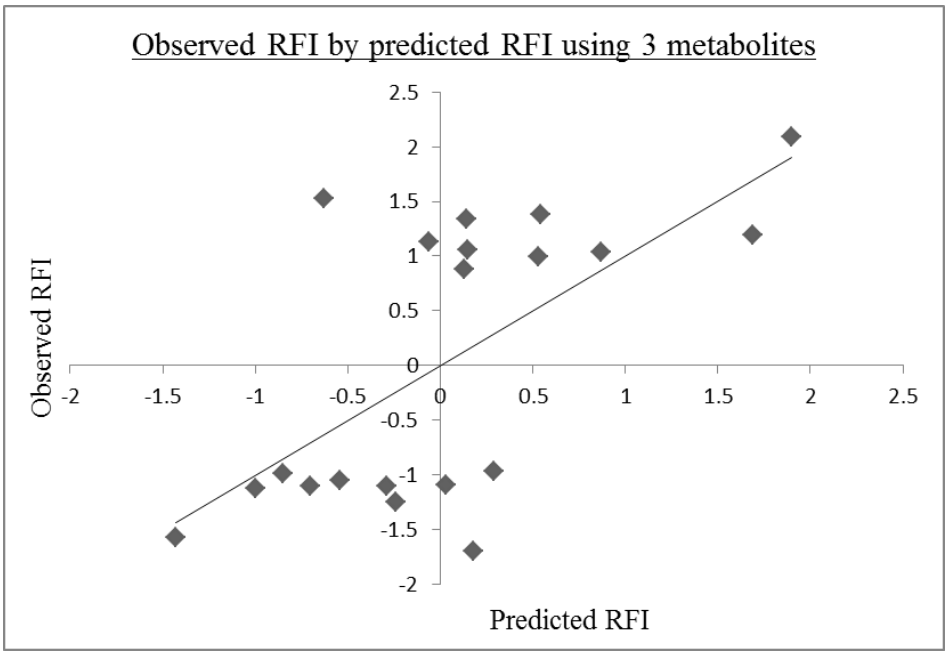



*Figure 6.5: Metabolic network for metabolites significant in period 3 in the discovery population*



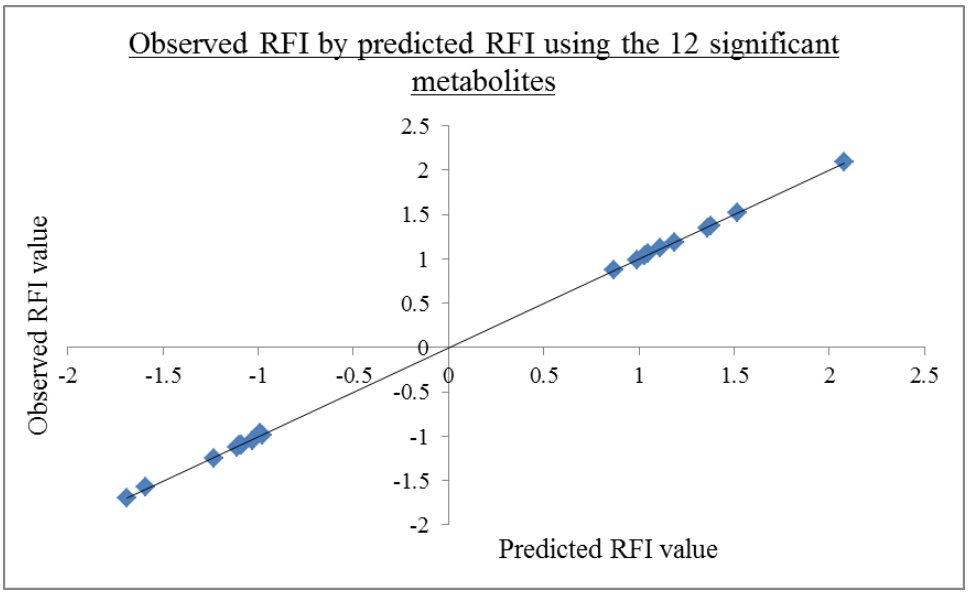




*Figure 6.6: Relationship between predicted values and observed values of RFI using the 3 metabolites that were significant in both populations (A) and the 12 metabolites that were significant in the validation population (B) in period 2.*

## 6.6. REFERENCES


Albano, N., Masero, J. A., Villegas, A., Abad-Gómez, J. M., and Sánchez-Guzmán J. M. 2011. Plasma metabolite levels predict bird growth rates: A field test of model predictive ability. Comp. Biochem. Physiol. A. Mol. Integr. Physiol. 160 (1), 9-15.

Arslan, C. 2006. L-Carnitine and its use as a feed additive in poultry feeding a review. Revue Méd. Vét. 157 (3), 134-142.

Arthur, P. F., Archer, J. A., Johnston, D. J., Herd, R. M., Richardson, E. C. and Parnell, P. F. 2001. Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. J. Anim. Sci. 79, 2805–2811.

Basarab, J. A., Price, M. A., Aalhus, J. L., Okine, E. K., Snelling, W. M., and Lyle. K. L. 2003. Residual feed intake and body composition in young growing cattle. Can. J. Anim. Sci. 83,189-204.

Canadian Council on Animal Care. 1993. Guide to the care and use of experimental animals. Vol. 1. E. D. Olfert, B. M. Cross, and A. A McWilliams, eds. 2nd ed. CCAC, Ottawa, ON.

Casas, E., White, S. N., Wheeler, T. L., Shackelford, S. D., Koohmaraie, M., Riley, D. G., Chase Jr., C. C., Johnson, D. D., and Smith, T. P. L. 2006. Effects of calpastatin and μ-calpain markers in beef cattle on tenderness traits. J. Anim. Sci. 84, 520-525.

Clarke, J. N., Binnie, D. B., Jones, K. R., Mowat, C. M., Purchas, R. W., and Uljee, A. E. 1996. Repeatabilities of blood plasma metabolites and their association with leanness in genotypes showing a wide divergence in carcass composition. Proc. N.Z. Soc. Anim. Prod. 56,180–183.

Di Lisa, F., Barbato, R., Menabo, R., Siliprandi, N. 1995: Carnitine and carnitine esters in mitochondrial metabolism and function. Dev. Cardiovasc. Med. 162, 21-38.

Hegarty, R. S., McFarlane, J R., Banks, R., Harden. S. 2006. Association of plasma metabolites and hormones with the growth and composition of lambs as affected by nutrition and sire genetics. Australian Journal of Agricultural Research, 57, 683–690.

Herd, R. M. and Arthur, P. F. 2009. Physiological basis for residual feed intake. J Anim Sci. 87 (14), 64-71.

Hermanussen, M., Garcia, A. P., Sunder, M., Voigt, M., Salazar, V., Tresguerres, J. A. 2006. Obesity, voracity, and short stature: the impact of glutamate on the regulation of appetite. Eur. J. Clin. Nutr. 60 (1), 25–31.

Jenni-Eiermann, S. and Jenni, L. 1994. Plasma Metabolite Levels Predict Individual Body-Mass Changes in a Small Long-Distance Migrant, the Garden Warbler. The Auk, 111 (8), 888 - 899.

Kennedy, B. W., Werf, J. H. J. and Meuwissen, T. H. E. 1993. Genetic and statistical properties of residual feed intake. J. Anim. Sci. 71, 3239-3250.

Kettunen, J., Tukiainen, T., Antti-Pekka, S., Ortega-Alonso, A., Tikkanen, E., Leo-Pekka, L., Kangas, A. J., Soininen, P., Würtz, P., Silander, K., Dick, M.D., Rose, R. J., Savolainen, M. J., Viikari, J., Kähönen, M., Lehtimäki, T., Pietiläinen, K.H., Inouye, M., McCarthy, M. I., Jula, A., Eriksson, J., Raitakari, O. T., Salomaa, V., Kaprio, J., Marjo-Riitta, J., Peltonen, L., Perola, M., Freimer, N. B., Ala-Korpela, M., Palotie, A., and Ripatti, S. 2012. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nature Genetics. 44, 269–276.

Koch, R. M., Swiger, L. A., Chambers, D., and Gregory. K. E. 1963. Efficiency of feed use in beef cattle. J. Anim. Sci. 22, 486–494.

Kulkarni, R. N. 2012. Identifying Biomarkers of Subclinical diabetes. Diabetes. 61, 1925-1926

Lele, R. D., Joshi, R. S., and Gupte, A. 2006. Association of Adipocytokines (Leptin, Adiponectin TNF-alpha), Insulin and Proinsulin with Diabetes - The Mumbai Obesity Project [MOP]. J Assoc Physicians India; 54:689-96.

Letto, J., Brosnan, M. E., and Brosnan, J. T. 1986. Valine metabolism. Gluconeogenesis from 3-hydroxyisobutyrate. Biochem. J. 240, 909-912

Mao. F., Chen, L., Vinsky, M., Okine, E., Wang, Z., Basarab, J., Crews, D. H. Jr., Li, C. 2013. Phenotypic and genetic relationships of feed efficiency with growth performance, ultrasound, and carcass merit traits in Angus and Charolais steers. J. Anim. Sci. 91(5):2067-76.

Matteri, R. L. 2001. Overview of central targets for appetite regulation. J. Anim. Sci. 79, 148-158.

Minor, R. K., Chang, J. W., and DeCabo, R. 2009. Hungry for Life: How the arcuate nucleus and neuropeptide Y may play a critical role in mediating the benefits of calorie restriction. Mol Cell Endocrinol. 299 (1), 79–88.

Montanholi, Y. R., Swanson, K. C., Schenkel, F. S., McBride, B. W., Caldwell, T. R., and Miller, S. P. 2009. On the determination of residual feed intake and associations of infrared thermography with efficiency and ultrasound traits in beef bulls. Livestock Science. 125, 22–30.

Montanholi, Y. R., Swanson, K. C., Palme, R., Schenkel, F. S., McBride, B. W., Lu, D. and Miller, S. P. 2010. Assessing feed efficiency in beef steers through feeding behavior, infrared thermography and glucocorticoids. Animal. 4(5), 692 – 701.

Mullen, E., and Ohlendieck, K. 2010. Proteomic profiling of non-obese type 2 diabetic skeletal muscle. Int. J. Mol. Med. 25 (3), 445-58.

Nkrumah, J. D., Okine, E. K., Mathison, G. W., Schmid, K., Li, C., Basarab, J. A., Price, M. A., Wang, Z., and Moore, S. S. 2006. Relationship of residual feed intake with metabolic rate, methane production and energy partitioning in beef cattle. J. Anim. Sci. 84, 145–153

Preston, T. R., and Leng. R. A. 1987. Matching ruminant production systems with available resources in the tropics and sub-tropics. Penambul Books. Armidale, N.S.W.

Schenkel, F. S., Miller, S. P., and Wilton, J. W. 2004. Genetic parameters and breed differences for feed efficiency, growth and body composition traits of young beef bulls. Can. J. Anim. Sci. 84,177–185.

Schenkel, F. S., Miller, S. P., Jiang, Z., Mandell, I. B., Ye, X., Li, H., and Wilton, J.W. 2006. Association of a single nucleotide polymorphism in the calpastatin gene with carcass and meat quality traits of beef cattle. J. Anim. Sci. 84, 291–299.

Sherman, E. L., Nkrumah, J. D., Murdoch, B. M., Li, C., Wang, Z., Fu, A. and Moore, S. S. 2008. Polymorphisms and haplotypes in the bovine neuropeptide Y, growth hormone receptor, ghrelin, insulin-like growth factor 2, and uncoupling proteins 2 and 3 genes and their associations with measures of growth, performance, feed efficiency, and carcass merit in beef cattle. J Anim. Sci. 86 (1), 1-16.

Sherman, E. L., Nkrumah, J. D., Moore, S. S. 2010. Whole genome single nucleotide polymorphism associations with feed intake and feed efficiency in beef cattle. J. Anim. Sci. 88,16-22.

Scholte, H. R., Boonman, A. M. C., Hussaarts-Odijk, L. M., Ross, J. D., Van Oudheusden, L. J., Pereira, R. R., Wallenburg, H. C. S. 1996: New aspects of the biochemical regulation of the carnitine system and mitochondrial

fatty acid oxidation. In: Carnitine - Pathochemical Basics and Clinical Applications. Eds Seim H, Löster H, Ponte Press Bochum, 11-31.

Siliprandi, N., Venerando, R., and Tassani, V. 1994: The "carnitine system": recent aspects. Adv. Exp. Med. Biol. 368, 161-164.

Viegas, C. M., Da Costa, F. G., Schuck, P. F., Tonin, A. M., Zanatta A., de Souza Wyse, A. T., Dutra-Filho, C. S., Wannmacher, C. M., Wajner, M. 2008. Evidence that 3-hydroxyisobutyric acid inhibits key enzymes of energy metabolism in cerebral cortex of young rats. Int. J. Dev. Neurosci. 26 (3-4), 293-9.

Wallimann, T., Tokarska-Schlattner, M., Schlattner, U. 2011. The creatine kinase system and pleiotropic effects of creatine. Amino Acids. 40 (5), 1271 – 1296.

Weikard, R., Altmaier, E., Suhre, K., Weinberger, K. M., Hammon, H. M., Albrecht, E., Setoguchi, K., Takasuga, A., and Kühn, C. 2010. Metabolomic profiles indicate distinct physiological pathways affected by two loci with major divergent effect on Bos taurus growth and lipid deposition. Physiol Genomics 42A, 79–88.

Williams, C. B. 2010. Application of biological simulation models in estimating feed efficiency of finishing steers. J Anim Sci. 88, 2523-2529.

Wishart, D. S., Knox, C., and Guo, A. C. 2009. HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res. 37 (Database issue), D603-610.

## *PREFACE TO CHAPTER SEVEN*

*Genes and metabolites do not function independently, in most cases their functions are correlated and result in interactions between them. Interactions between genes are implemented by biological processes (pathways) while interactions between metabolites are implemented by genes (usually enzymes).*

*Chapter Seven is aimed at analysing the interactions between genes and metabolites and the biological processes involved in these interactions.*

*This chapter has been submitted for publication in Animal Science Journal, 2013.*

# CHAPTER SEVEN

## Analysis of biological networks and biological pathways associated with residual feed intake in beef cattle[5]

### 7.1. INTRODUCTION

Improving profitability in livestock production can be achieved by reducing the costs of production (inputs) while either maintaining or improving the quantity and/or quality of the products (outputs) (MacNeil et al., 1997). In beef production, the cost of feed is the largest single expense (input), accounting for approximately 60- 70% of the total cost of production (Herd et al., 2003; Arthur et al., 2004). Therefore, there is need to increase the efficiency of feed utilization by reducing the amount of feed consumed for the same level of production. Most measures of feed efficiency are correlated with production traits except residual feed intake (RFI). RFI is defined as the difference between actual feed intake and predicted feed intake based on the individual's body weight (maintenance) and average daily gain (growth/production) (Koch et al., 1963). RFI is phenotypically independent of the production traits used to estimate it (Archer et al., 1999) indicating that RFI represents inherent differences in basic metabolic processes that determine the efficiency of feed utilization (Korver, 1988).

[5] A version of this chapter was submitted to Animal Science Journal for publication, 2013.

The genetic basis of RFI has been studied and several genes were reported to be significantly associated with RFI (Arthur and Herd, 2006; Moore et al., 2006; Barendse et al., 2007; Nkrumah et al., 2007a; Sherman et al., 2008 a, b). However, the amount of variation explained so far remains relatively small, for example Sherman et al., (2008a) identified SNPs accounting for 6% of the phenotypic variation in RFI. The missing heritability may arise from several factors (*reviewed in detail in Chapter 3*) including association analysis approaches that rely on LD between genotyped SNPs and unknown functional DNA variants. These approaches are more effective when the population consists of genetically related individuals (Snelling et al., 2013). The missing heritability may also result from the association analyses taking a reductionist approach by identifying discrete molecules associated with RFI. Although the reductionist approach has had several successes, (see above), most complex traits, including RFI, are characterized by complex interactions between cellular constituents such as DNA, RNA and proteins, and are usually influenced by multiple biological processes simultaneously (Barabasi and Oltvai, 2004).

To understand the biological processes associated with RFI there is need to address the mechanisms by which an associated variant gives rise to the phenotypic differences observed. In addition, molecular variants can be analyzed in the framework of pathways and networks (Han, 2008). The knowledge of gene functions and interactions may provide greater insight into the genes and genomic mechanisms affecting polygenic traits, and facilitate functional genomic selection for economically important traits (Snelling et al., 2013). The biological networks

could also be used to identify additional genes and metabolites involved in the biological processes and that may have an effect on RFI.

The objectives of this study were to utilize molecular data obtained from association analyses using genes and plasma metabolites to reconstruct biological interaction networks for RFI and analyze the networks to offer a holistic view of the regulation of RFI and identify the specific biological processes associated with RFI.

In reference to the gene networks, we hypothesized that the genes associated with RFI significantly interact with one another connected by the biological processes they are involved in. Each biological process would result in metabolites as components of either the substrates or the products of that process.

In reference to the metabolic networks, we hypothesized that the metabolites associated with RFI also significantly interact with one another connected by the protein products of genes, which utilize the metabolites as substrates or produce them as products of the specific biological reactions in which they are involved.

## 7.2. MATERIALS AND METHODS

### 7.2.1. Summary of association analysis

Phenotypic and genotypic data were obtained from 531 beef steers at the University of Alberta ranch at Kinsella, Canada. The breed composition of this herd was described by Nkrumah et al., (2007b).

The steers were managed and tested under feedlot conditions using the GrowSafe automated feeding system (GrowSafe Systems Ltd., Airdrie, Alberta, Canada) as described by Nkrumah et al., (2004) and all the animals were managed and cared for according to the guidelines of the Canadian Council on Animal Care (CCAC, 1993).

### 7.2.2. Estimation of RFI

Collection of feed intake data and calculation of RFI was discussed in detail by Basarab et al., (2003). In summary, the actual feed intake for each steer was measured using the GrowSafe automated feeding system (GrowSafe Systems Ltd., Airdrie, Alberta, Canada). This feed intake was converted to dry matter intake (DMI) by multiplying it by the dry matter content of the diet. DMI was then standardized across the different years to 10 MJ of metabolizable energy (ME)/kg of dry matter. The predicted feed intake was estimated based on metabolic mid-weight and average daily gain. Average daily gain (ADG) was calculated as the slope from the regression of body weight (BW) on test day. Metabolic mid-weight was obtained as the mid-weight on test to the power of 0.75.

### 7.2.3. Identification of genes associated with RFI

About 1100 positional candidate genes were identified within the genomic positions of previously reported QTL from the bovine QTL database (cattleQTLdb) release 10 (Hu et al., 2007). These QTLs were associated with traits related to RFI such as feed / dry matter intake, growth, FCR, average daily

gain and energy balance. An additional set of positional candidate genes consisted of 1018 genes positioned within a range of 500kbp on both sides of 310 SNPs previously reported to be significantly associated with RFI in a GWAS using the 50K bovine SNP chip (Mujibi et al., 2011). Therefore a total of 2118 positional candidate genes were used in this analysis.

The positional candidate genes were then prioritized into functional candidate genes according to their functions inferred from gene ontology terms (Ashburner et al., 2000) in relation to the physiological processes associated with feed efficiency. The physiological processes associated with RFI were reviewed by Richardson and Herd (2004) and Herd and Arthur (2009) indicating that metabolism (anabolism and catabolism) may account for about 42% of the variation in RFI. Other processes that account for the variation in RFI include body composition (5%), digestion (10%), physical activity (9%), thermoregulation and heat increment of feeding (9%) and unknown factors (25%).

The gene ontology terms considered in this study were related to energy acquisition through carbohydrate, fat and protein metabolism and energy utilization through growth, body temperature regulation, muscle activity and immune response. There were 116 genes whose functions were associated with RFI and were considered the functional candidate genes. The remaining 2002 positional candidate genes that were not functionally relevant were not included in any further analysis.

SNPs located in the functional candidate genes were identified by comparing cDNA sequences with reference sequences from Ensembl version 57 (Hubbard et

al., 2009). To generate the cDNA library, RNA was prepared from pooled liver samples using TRIzol® reagent (Invitrogen, Carlsbad, CA, USA) using the protocol described in the TRIzol® reagent user manual (Connolly et al., 2006). The liver cDNA library was constructed according to the TruSeq® RNA and DNA sample preparation kit (Illumina, USA, v2 published by Illumina Inc. 2011, publication number 970-2009-039) and cDNA sequencing was performed on the Genome Analyzer II using the TruSeq® RNA and DNA sample preparation kit (Illumina, USA, v2 published by Illumina inc 2011, publication number 970-2009-039).

To generate a list of putative SNPs, Maq (version 0.7.1) (Li et al., 2008) was used to map DNA sequencing reads and the bovine transcript sequences from Ensembl version 57 (Hubbard et al., 2009) were used as reference sequences. The SNPs returned by Maq's SNPfilter command were annotated using NGS-SNP (Grant et al., 2011) by assigning a functional class to each SNP (e.g. nonsynonymous) and then providing NCBI, Ensembl identifiers and gene ontology terms for the affected genes, when applicable. Only the non-synonymous SNPs were considered functional candidate SNPs and were retained if the minor allele frequency was greater than 10%.

The non-synonymous SNPs were prioritized based on the expected effect of the amino acid change: a non-conservative amino acid change was considered to have a greater chance of being associated with a trait variation than a conservative amino acid change in the same gene. A total of 117 SNPs were selected within 116 genes (with 2 SNPs in the CAST gene) and genotyped at GeneSeek® using

DNA samples obtained from steers at the University of Alberta ranch at Kinsella, Canada.

A multiple marker association analysis was performed as shown in equation 1 using the RFI data obtained from 531 steers from the University of Alberta ranch at Kinsella, Canada, using ASReml 3 (Gilmour et al., 2009).

$$Y_{ij} = \mu + X_{1i}\beta + \sum_{j=1}^{n} X_{2j} g_j + Za + e \qquad \text{(Equation 1)}$$

$Y_{ij}$ represented the RFI value for animal *i* which has *j* SNPs

$\mu$ was the population mean, β was a vector of fixed effects associated with animal *i*, in this case the breed of the sire and the batch (Mujibi et al., 2010), $\sum_{j=1}^{n} X_{2j} g_j$ was the sum of the SNP effects *(g)*, $X_2j$ was a design matrix relating an observation *(Y)* to one of the genotypes 0, 1 or 2 at the *j*th SNP, *a* was a vector of random additive effects (exclusive of SNP *g* effect(s) in the model) constructed using each animal's pedigree information to construct the relationship matrix A. It was assumed to be normally distributed with a mean of 0 and variance of $A\sigma^2_{a.}$

*e* was the vector of random residual effects assumed to be normally distributed with a mean of 0 and a variance of $I\sigma^2$ where I was an identity matrix

### 7.2.4. Analysis of blood metabolites

The steers used in the analysis of metabolites belonged to the University of Guelph ranch in Ontario. The breed composition of the population was described by Montanholi et al., (2009). In summary, the total number of steers entering the feeding test was 91 with 46 from year 1 and 45 from year 2. The average weight

of the steers was 313 ± 6.2 kg and once the test begun, steers had a period of approximately 2 weeks to acclimatize to the facilities and feeding system. About 5ml of blood was collected through jugular venipuncture from beef steers in the year 2 group every week during the feeding test period into collection tubes containing sodium heparin and immediately stored on ice until centrifugation.

Estimation of RFI for individual steers was described by Montanholi et al., (2009). In summary, individual average daily feed intakes for the feeding test period were recorded by an automated feeding system. Daily dry matter intake (DMI) over the testing period was calculated as a product of the daily intake in as fed basis and the percent dry matter in the feed. Average daily gain (ADG) was determined by a regression of body weight on days on test. RFI was estimated using the equation described by Montanholi et al., (2009, 2010) as;

$$RFI = DMI - \beta 0 + \beta 1\ (ADG) + \beta 2\ (BW) \quad \text{…..Equation 1}$$

Where β0 was the intercept, β1 and β2 were the regression coefficients on ADG and on mid-trial body weight (BW) respectively. The ADG was determined by a regression of BW on days on trial, with six observations per animal. Mid-trial BW was calculated by computing the animals' intercept plus the ADG times 70 (half of the experimental period).

At the end of the feeding period when RFI values were available for all the animals, the plasma samples were sorted into batches according to the period when they were collected and the RFI value of the steers tested. Three time periods were considered in this study; period 1 corresponded to the second week

after the beginning of the feeding test period, period 2 and 3, corresponded to 6 weeks and 9 weeks into the feeding period respectively. From each of the three periods, 16 plasma samples were selected from the steers with the highest RFI value and 16 plasma samples from the steers with the lowest RFI value establishing two groups with divergent RFI values. A total of 96 samples were used in the metabolomic analysis.

Association analysis between RFI and the metabolite levels was performed using SAS 9.1 (SAS Institute Inc. Cary, NC, USA) for each period separately using a multiple regression model accounting for the year of birth of the steers and for the combined data at the three periods using a mixed model for repeated measures analysis.

### 7.2.5. Reconstruction of biological networks

Biological networks were reconstructed using IPA software (Ingenuity Systems, www.ingenuity.com) using the genes and metabolites significantly associated with RFI. The IPA software was selected because it offered a large knowledge base and can model relationships between genes, proteins, metabolites and can be used to identify the biological processes involved (Ingenuity knowledge base). However, the IPA knowledge base is built using data obtained from studies in mice and humans. Although these species can be utilized as model animals and most of the biological processes between them and other animal species may be similar, caution should be practiced when dealing with biological processes that differ widely between humans/mice and cattle. For example, biological processes

related to immune response may be relatively similar between humans, mice and cattle while the major processes related to feed digestion may differ between the species.

To reconstruct the biological networks, a list of significant molecules (genes or metabolites) and their corresponding P-values were imported into the IPA software and the parameters were set to allow the network to include indirect relationships between the imported molecules and the corresponding molecules in the knowledge base. Indirect relationships would assist in the identification of other genes or metabolites that were not among the ones analyzed but may be associated with RFI. The IPA algorithm generated biological networks by mapping each gene/metabolite identifier to its corresponding gene/metabolite in the IPA knowledge base as described in detail by Calvano et al., (2005). The molecules are then overlaid onto a global network developed from information contained in the knowledge base. Each network is assigned a significance score, which represents the likelihood that the imported molecules within the network are found therein by random chance (Calvano et al., 2005). A high number of imported molecules within a dataset lead to a higher network score. The network score is calculated as the negative of the exponent of the P-value such that a score of 25 will be equal to a P-value of $1.0^{-25}$ (Calvano et al., 2005) and therefore larger scores correspond to higher significance.

The gene networks were reconstructed from the genes significantly associated with RFI and metabolic networks were reconstructed for each period from the metabolites significantly associated with RFI in that period. However, at period

3, only 1 of the 3 significant metabolites was mapped in the IPA knowledge base and used to reconstruct the network therefore this network will not be discussed in detail. Further analysis of the interaction networks was performed using the canonical pathways tool within the IPA software.

## 7.3. RESULTS AND DISCUSSION

In summary, 25 SNPs in 24 genes were significantly associated with RFI ($P < 0.05$) accounting for 19.7% of the phenotypic variation in RFI (Table 4.3 and also shown in Table 7.1 ).

The metabolites significantly associated with RFI were 2 in period 1, 10 in period 2 and 3 in period 3 (Table 6.1 and again in Table 7.2 .

### 7.3.1. Analysis of gene network

The gene network reconstructed using IPA is shown in figure 7.1 with the corresponding locations of the genes within the cell. The additional genes which had not been analyzed in this study but were present in the network will be discussed as components of the biological pathways associated with RFI and potential candidate genes associated with RFI.

There were 2 major hubs in the gene network; the first hub was associated with the Ubiquitin C (UBC) gene and the second hub was associated with the insulin induced gene 1 (INSIG1). There was also a minor hub associated with the ERK1/2 gene.

The hub associated with the UBC gene had multiple interactions with other genes including PLEKHA7, PARP14, SMARCAL1, UBA5, LRP5, CAST and INSIG1 (Fig 7.1). UBC also interacted with ERK1/2 through LRP5, UBA5, GHR, OSMR or LIFR.

The UBC gene encodes a polyubiquitin precursor which can be conjugated into Ubiquitin monomers or polymers and can result in different effects in the cell depending on the residues to which the ubiquitin is conjugated. Generally the UBC gene is involved in protein degradation, DNA repair, cell cycle regulation, kinase modification, endocytosis and regulation of other cell signaling pathways (Rebhan et al., 1997). These processes are generally classified and described as post translational regulation of gene expression and may cause variation in phenotypes by influencing the levels of proteins (such as enzymes) present and their activity in performing the biochemical processes. The UBC gene interacts with all the genes that were significant in this study indicating a possible role in influencing the activity of their protein products.

The INSIG1 gene hub had interactions with other molecules including Lysophosphatidylcholine acyltransferase 3 (LPCAT3), Acyl-CoA synthetase short chain family member 2 (ACSS2), ELOVL fatty acid elongase, Acetoacetyl CoA synthetase (AACS), endoplasmic reticulum lipid raft associated protein (ERLIN2), 3-hydroxyl-3-methylglutaryl-CoA synthase 2 (HMGCS2) and STAR related lipid transfer protein 4 (STARD4). All these genes relate to metabolism and specifically this hub represents biological processes related to energy, lipid and steroid metabolism. Polymorphisms that may be present in these genes may

cause variation in lipid metabolism and may influence body composition in beef cattle steers. The importance of lipid metabolism and body composition in influencing the levels of RFI was previously discussed by Naik et al., (2007) and Richardson et al., (2004). Additional evidence on the role of lipid metabolism on RFI was provided by the observed genetic correlation of 0.17 between RFI and subcutaneous fat depth at the $12^{th}$ and $13^{th}$ ribs and the rump (Arthur et al., 2001; Shenkel et al., 2004).

The ERK1/2 hub (which also interacts with the UBC gene through UBA5) was closely linked to GHR, LIFR, JAK, OSMR and Glycoprotein 130 (GP130-LIFR). Other genes linked to this hub included CLCF1, STAT5a/b, STAT1/3/5 and IL31RA. ERK1 is a protein kinase involved in regulating transcription, translation and cytoskeleton rearrangements. It also causes phosphorylation of several transcription factors resulting in regulation of meiosis and mitosis and post mitotic functions of cells (Rebhan et al., 1997). The details of these processes will be discussed in a subsequent section.

**7.3.2. Canonical pathway analysis:**

a. The growth hormone (GH) signaling pathway

The growth hormone receptor precursor (GHR) was identified to be associated with RFI in another study by Sherman et al., (2008). In this analysis, GHR interacts closely with JAK, ERK1/2, STAT5 a/b and STAT1/3/5 as shown in figure 7.2a. A detailed representation of the GH signaling pathway is shown in figure 7.2b and fig. 7.7.

Growth hormone is a peptide hormone; therefore, it cannot directly cross the plasma membrane. Its effect is initiated when it binds to the GH receptor (GHR) located on the plasma membrane. This signal can have an effect on several pathways (fig. 7.2b) but only three of these pathways will be discussed further due to their role in metabolism.

In the first instance, the signal stimulates the JAK2 pathway which is capable of stimulating either STAT1, 3 and 5 or the ERK1/2. The signal that stimulates the ERK1/2 will then stimulate CEBPA. CEBPA (CCAAT/enhancer binding protein) is a transcription factor protein that binds to certain promoters and enhancers and modulates the expression of these genes. The most important genes modulated by this protein are leptin and growth hormone. Several studies have reported the significant association between leptin and RFI (Hoque et al., 2009, Kelly et al., 2009). The fact that CEBPA modulates the expression of the genes by binding to the promoter site indicates that polymorphisms located in the promoter sites may affect the binding to CEBPA thereby affecting the levels of gene expression. Additional evidence of this effect was provided by previous association analysis of SNPs located in the promoter region of the leptin gene, which indicated significant association between some of the SNPs with RFI (Nkrumah et al., 2005).

In the second instance, JAK2 stimulates STAT1, 3 and 5. Then the phosphorylated (activated) STAT5 act as a transcription regulator for genes such as Insulin like growth factor 1 and 2 (IGF1, IGF2), Insulin like growth factor binding protein 3 (IGFBP3) and Insulin growth factor binding protein acid labile subunit (ALS) (fig

7.2b and Fig 7.7). This pathway can also be activated through PLCG (phospholipase C gamma) which then stimulates Diacylglyceride (DAG). DAG has an effect on Protein kinase C (PKC), which stimulates STAT5 transcription activation proteins.

IGF-1 is a hormone that regulates growth and cellular anabolism and was already reported to be associated with feed efficiency (Bishop et al., 1989; Stick et al., 1998). In fact, IGF-1 has been used as a marker for indirect selection for RFI (Davis and Simmen, 2006). However, these authors also reported a positive correlation between the levels of IGF-1 and some measures of growth indicating that using IGF-1 to select for RFI may also be selecting for growth traits simultaneously. Both IGF-1 and 2, through the IGF-1 receptor, are involved in pathways associated with protein synthesis and cell survival (Fig 7.2b).

In the third instance, JAK2 stimulates P1-3K (1-phosphatidyl Inositol 3 kinase) either directly or indirectly through IRS1 (Insulin receptor substrate 1). The P1-3K is involved in the expression of GLUT4 on the plasma membrane (Fig 7.2b). GLUT4 allows the facilitated diffusion of glucose into muscles and fat cells aided by insulin (Fig 7.7). It can be seen, therefore, that growth hormone and its receptor are involved in biological processes associated with protein synthesis, cell survival, lipid metabolism and glucose transport across the cell membrane.

b. Oncostatin M receptor signaling

Oncostatin M receptor signaling is similar to growth hormone signaling. Oncostatin M binds to the oncostatin M receptor (OSMR) stimulating Janus

kinase (JAK) which then stimulates STAT1, 3 and 5. The activated STAT1, 3 and 5 proteins enter the nucleus and act as transcription activators for several genes. This pathway differs from the growth hormone signaling pathway, which activates the transcription of different genes. In oncostatin M signaling pathway the expressed genes include; Matrix metallopeptidase 1 (MMP1), 3 (MMP3), 13 (MMP13) and TIMP metallopeptidase inhibitor 3 (TIMP-3) (Fig 7.3). Most of these genes are associated with cell proliferation, migration, differentiation, angiogenesis and apoptosis (Rebhan et al., 1997).

### 7.3.3. Metabolic networks

At period 1 only two metabolites, *creatine* and *glycine* were significantly associated with RFI (Table 7.2) and used to reconstruct the metabolic network using IPA. Several links were created between these metabolites and other metabolites that were not included in the analysis in this study (Fig 7.4). Creatine interacts closely with AMP activated protein kinase (AMPK), which is an enzyme that plays a role in cellular energy homeostasis. The net effect of AMPK activation is stimulation of hepatic fatty acid oxidation and ketogenesis, inhibition of cholesterol synthesis, lipogenesis, and triglyceride synthesis, inhibition of adipocyte lipolysis and lipogenesis, stimulation of skeletal muscle fatty acid oxidation and muscle glucose uptake, and modulation of insulin secretion by pancreatic beta-cells (Winder and Hardie, 1999). Creatine also interacts with SLC2A4 (GLUT4) which is responsible for glucose transport across the cell membrane.

A review of the creatine canonical pathways (figure 7.5) indicated that additional biological processes involved included steroid biosynthesis, chloride transport and glycogen synthesis. In the mitochondrion, biological processes affected included triacylglycerol synthesis and β-oxidation of fatty acids to acetyl-CoA. Other molecules that would initiate similar signals to AMPK are Leptin, Adiponectin and Insulin (fig 7.5). The importance of lipid, cholesterol and glucose metabolism in influencing feed efficiency in farm animals was reviewed in detail by Rafael et al., (2007).

AMPK has also been associated with increased blood supply to trained muscle cells by promoting angiogenesis (Ouchi et al., 2005) and may be a response to muscle activity. Increased muscle activity also requires increased supply of energy as indicated by the involvement of the Glucose transporter (SLC2A4/GLUT4) and Glycogen synthetase kinase (Fig 7.4). Muscle activity also requires the involvement of muscle contractile tissues as indicated in the network (Fig. 7.4) by the interactions with Myosin heavy chain 2 (MYH2) and Titin (TTN). The correlation between some measures of physical activity and the variation in RFI has been reported previously in poultry (Luiting et al., 1991), pigs (DeHaer et al., 1993) and in beef cattle (Herd and Arthur, 2009; Nkrumah et al., 2003). These genes identified here may (partially) indicate the underlying biological mechanism for this correlation.

At period 2, 10 metabolites were significantly associated with RFI. These included Acetic acid, Carnitine, Betaine, Citric acid, Creatine, Hippuric acid, Glutamic acid, Lysine, Phenylalanine and Threonine. Eight of these metabolites

were mapped in IPA, and used in the metabolic network analysis. The presence of creatine in this network (Fig. 7.6) resulted in an overlap of some of the metabolites with some metabolites in the network reconstructed for period 1 (fig 7.4). In addition, this network also showed several similarities with the gene network for RFI (Fig. 7.1); such as the involvement of the ERK1/2, MAPK and P13K described in a preceding section. One of the major canonical pathways in the metabolic network in this period was the citric acid (TCA) cycle and figure 7.6 shows the interaction between acetate, acetoacetate and citrate in a module that is made complete by ERK1/2. These metabolites are important in the TCA cycle and may be used to identify additional candidate metabolites associated with RFI. Acetate is used to produce energy when in low energy balance and it forms fat when the individual is in high energy balance (Preston and Leng, 1987). Acetate reacts with Coenzyme A to form acetyl CoA which enters the TCA cycle (in a step which utilizes ATP) and reacts with oxaloacetate to form citrate.

In this study, citrate levels in plasma were significantly associated with RFI; steers with low RFI had less citrate and higher acetate than those with high RFI. This may indicate the importance of the TCA cycle in regulating feed efficiency. The interaction between acetic acid and citric acid is indicated in the network (Fig 7.6) as a bidirectional relationship where the metabolites can influence each other in both directions. Although the specific mechanisms remain just hypotheses, it is worth noting that regulating the levels of acetate and/or citrate may affect the rate of the processes occurring in the TCA cycle and therefore the amount of energy produced. In addition, the conversion of acetate into acetyl CoA requires ATP,

therefore reducing the rate of this step may save the individual's ATP and may explain why low RFI steers had higher levels of acetate than high RFI steers yet both high and low RFI steers had similar levels of acetyl CoA. An association between RFI and mitochondrial ATP production was reported by Kelly et al., (2009). They reported significant associations between UCP-3, PGC-1α and COX II genes with RFI in beef cattle and indicated the importance of cellular energetic efficiency in influencing RFI.

At period 3, three metabolites; tyrosine, hydroxyisobutyrate and formate were significantly associated with RFI. In the network reconstruction, only tyrosine was used and therefore this network will not be discussed further. Hydroxyisobutyrate (HB) was not identified in the IPA knowledgebase so there was no network associated with it. HB is associated with gluconeogenesis from valine (Letto et al., 1986). It has been shown that hydroxyisobutyrate inhibits key enzymes in energy metabolism in young rats by inhibiting the functions of the respiratory chain complex 1 to 3 and mitochondrial creatine kinase (Viegas et al., 2008). Hydroxyisobutyrate was also associated with high levels of circulating free fatty acids, increased intramyocellular lipid content, impaired insulin-mediated glucose uptake, diminished mitochondrial functioning and an overall weakened metabolic flexibility (Mullen and Ohlendieck, 2010).

### 7.4. Biological networks indicate a relationship between feed efficiency and cholesterol metabolism

The regulation of cholesterol synthesis depends on the levels of cholesterol in the cell. In humans, a high intake of cholesterol from food results in reduced

production in the cell (Espenshade and Hughes, 2007). Livestock, on the other hand, consume almost similar feed for extended periods of time therefore the amount of cholesterol obtained from the feed remains relatively constant. In livestock, variation in levels of cholesterol in the cell is largely due to variation in cellular regulation of cholesterol biosynthesis.

A detailed account of cholesterol regulatory mechanisms was provided by Espenshade and Hughes (2007). In summary, it relies on detecting the levels of intracellular cholesterol in the endoplasmic reticulum by the sterol regulatory element-binding proteins (SREBP1 / 2). When cholesterol levels in the cell are high, the SREBP protein is bound to two other proteins; SREBP-cleavage-activating protein (SCAP) and Insulin induced gene 1 (INSIG1). INSIG1 was one of the genes that were significantly associated with RFI in this study. When cellular cholesterol levels are low, INSIG1 is unbound from the complex and SREBP is cleaved. The cleaved SREBP enters the nucleus and bind to the sterol regulatory element (SRE) acting as a transcription factor, which stimulates the transcription of several genes including the low-density lipoprotein (LDL) receptor and HMG-CoA reductase. One of the low density lipoprotein receptors, LRP5, was significantly associated with RFI in this study. In general, the SREBP pathway regulates the expression of several genes that are involved in lipid synthesis and metabolism.

Cholesterol synthesis begins with the conversion of Acetyl CoA and acetoacetyl CoA to hydroxymethylglutaryl-CoA (HMG-CoA) (Liscum, 2002). Acetyl CoA is formed from the reaction between acetate and coenzyme A. (Note that the plasma

levels of acetate were significantly associated with variation in RFI in this study). The HMG-CoA is reduced to mevalonate in a rate-limiting step where the HMG-CoA reductase enzyme is highly regulated. This is followed by a series of reactions (details provided by Liscum, 2002) resulting in lanosterol, which is then converted to cholesterol by 19 additional reactions. The reactions involved in the conversion of lanosterol to cholesterol are catalyzed by members of the cytochrome $P_{450}$ enzyme superfamily (Liscum, 2002). In this study, Cytochrome $P_{450}$ subfamily 2B (CYP2B) gene was significantly associated with variation in RFI.

Some of the other metabolites that were indirectly associated with cholesterol metabolism included creatine and citrate, which would influence the levels of AMPK and as a result may have an impact on cholesterol metabolism (fig 7.5).

Using the biological network analysis, we hypothesize that feed efficiency in beef cattle is associated with cholesterol levels in tissues. In addition, we hypothesize that steers with high RFI (inefficient) will also have high cholesterol levels compared to low RFI (efficient) steers. We propose this positive correlation because the pattern of regulation of cholesterol biosynthesis is similar to the regulation of lipid metabolism in the cell. Using the trend observed in previous studies which reported a positive correlation between RFI and intermuscular, intramuscular, subcutaneous and average back fat in beef carcasses (Arthur et al., 2001; Basarab et al., 2003; Shenkel et al., 2004) it also appears that low RFI steers may have low cholesterol levels.

## 7.5. SUMMARY and CONCLUSION

Residual feed intake is a complex trait regulated by many genes and biological pathways (Herd and Arthur, 2009). Several metabolites can also act as intermediate phenotypes to indicate the substrates and/or products of the biological reactions associated with RFI. Figure 7.7 summarizes the interaction between the different genes and metabolites that were significantly associated with RFI and discussed in this study. The biological processes associated with variation in RFI begin from glucose uptake into the cell through the GLUT4 protein. The GHR gene, Insulin, creatine-AMPK and Leptin have an impact on the efficiency of the glucose transporter. Once glucose enters the cell, it could undergo glycolysis into pyruvate and later into acetyl CoA. Gluconeogenic pathways and β oxidation of fatty acids also results in the generation of acetyl CoA. In this study, metabolites such as acetate and citrate were significantly associated with RFI and may indicate an influence on the rate of formation of acetyl CoA. The formation of acetyl CoA is also influenced by AMPK (figure 7.5). Some of the amino acids that are glucogenic were also significantly associated with RFI, and lipid metabolism was influenced by multiple genes including GHR, INSIG1 and LRP5. The role of lipid metabolism in influencing RFI was reported previously by Naik et al., (2007) and Richardson et al., (2004).

Biological networks such as those identified in this study may be used to identify the biological processes associated with a phenotype and to give an insight on the interactions between the molecules. The biological processes associated with the trait and the canonical pathways associated with the molecules may be used to

predict the relationship between the trait and other traits of importance in beef productivity.

**Table 7.1: Candidate genes significantly associated with residual feed intake in beef cattle steers**

| Gene name | BTA: Position (bp) | P-value | SNP position | SNP allele[§] | Amino acid change |
|---|---|---|---|---|---|
| **PARP14** | 1:68144657 | 0.028 | 5002 | G/A | Gly -> Arg |
| **UMPS** | 1:70328819 | <0.001 | 572 | C/T | Arg -> Cys |
| **UBA5** | 1:139111130 | 0.048 | 231 | A/T | Glu -> Val |
| **ACAD11** | 1:138045480 | 0.004 | 816 | C/T | Ser -> Leu |
| **BIN1** | 2:5642793 | 0.026 | 916 | G/A | Ala -> Thr |
| **ASNSD1** | 2:6949248 | <0.001 | 499 | G/A | Ala -> Thr |
| **MKI67IP** | 2:76998684 | 0.036 | 900 | A/T | Arg -> Stop |
| **AOX1** | 2:89545687 | 0.0001 | 1732 | A/G | Ser -> Gly |
| **SMARCAL1** | 2:105138600 | 0.034 | 265 | T/C | Ser -> Pro |
| **PQLC2** | 2:137970404 | 0.023 | 691 | A/G (C/G) | Arg -> Gly |
| **NECAP2** | 2:140340622 | 0.013 | 568 | A/G | Thr ->Ala |
| **INSIG1** | 4:121360120 | 0.006 | 308 | A/G | Ser -> Gly |
| **CAST** | 7:97480120 | 0.026 | 271 | G/A | Gly -> Asp |
| **CAST** | 7:97526153 | 0.007 | 672 | A/G | Thr ->Ala |

| | | | | | |
|---|---|---|---|---|---|
| **PLEKHA7** | 15:34411065 | 0.04 | 2490 | C/T | Pro -> Ser |
| **APIP** | 15:65118633 | <0.001 | 737 | G/A | Val -> Met |
| **CYP2B** | 18:49958396 | 0.014 | 16 | G/A | Val -> Met |
| **OCLN** | 20:10849769 | 0.008 | 1051 | T/C (A/G) | Cys -> Arg |
| **GHR** | 20:33897128 | 0.026 | 1643 | G/A | Ala -> Thr |
| **OSMR** | 20:37772898 | 0.006 | 620 | G/T | Arg -> Met |
| **LIFR** | 20:38170739 | 0.005 | 1162 | G/A | Ala -> Thr |
| **UGT3A1** | 20:40434540 | 0.023 | 1597 | T/C | Met -> Thr |
| **SLC45A2** | 20:42286376 | <0.0001 | 718 | G/A | Ala -> Thr |
| **MYO10** | 20:59933885 | 0.0001 | 2375 | G/T | Arg -> Leu |
| **LRP5** | 29:47717873 | 0.018 | 3166 | A/G | Met -> Val |

[§]*For some genes, the SNP alleles reported in the SNP database were different from the alleles observed in the population used in this study but the amino acid and nucleotide positions were the same. The alleles reported in the database are shown in parentheses.*

**Table 7.2: Metabolites associated with RFI in beef cattle steers in three periods**

| Period | Metabolite | P-value |
|---|---|---|
| 1 | Creatine | 0.010 |
| | Glycine | 0.001 |
| 2 | Threonine | 0.044 |
| | Carnitine | 0.008 |
| | Acetate | 0.032 |
| | Creatine | 0.002 |
| | Phenylalanine | 0.003 |
| | Lysine | 0.009 |
| | Citrate | 0.002 |
| | Betaine | 0.036 |
| | Glutamate | 0.001 |
| | Hippurate | 0.026 |
| 3 | Hydroxyisobutyrate | 0.0002 |
| | Tyrosine | 0.014 |
| | Formate | 0.007 |

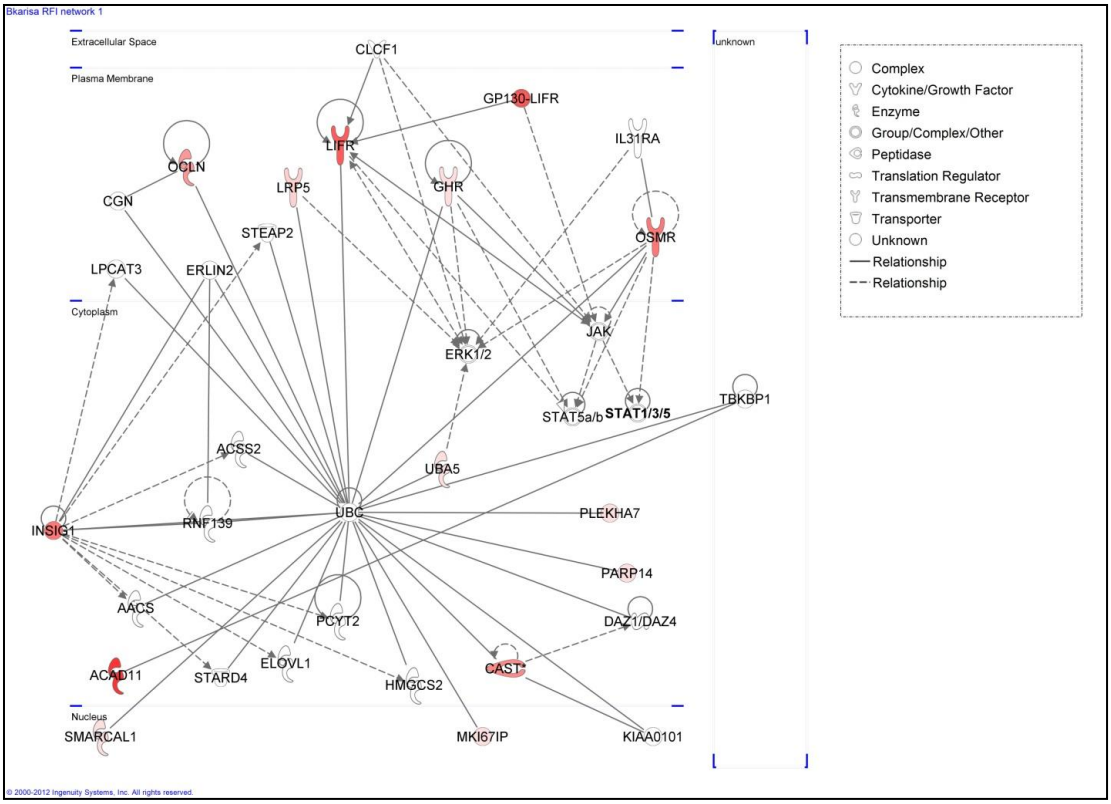



*Figure 7.1: RFI gene network showing the sub-cellular layout where each gene exerts its functions. About 3 hubs can be identified in this network; the UBC hub, the INSIG1 hub and the hub associated with ERK1/2 and STAT genes. Most of the genes are located in the plasma membrane acting as receptors, transport molecules or signaling molecules. Signals are then passed to molecules in the cytoplasm, for example, from GHR to JAK and then to the STAT genes.*

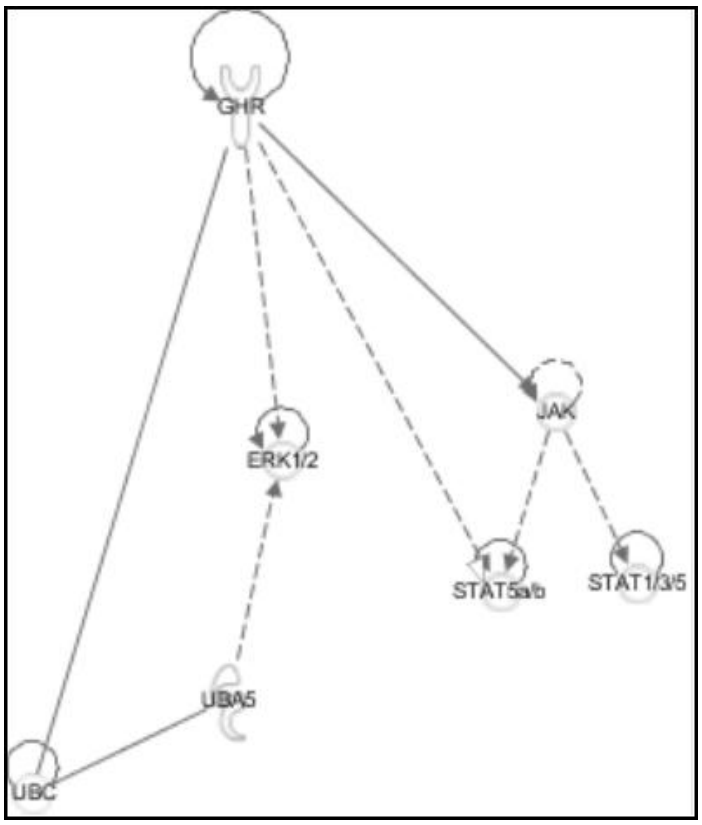



*Figure 7.2a: The GHR sub-network as shown in the gene network analysis. The GHR gene module shown here is an incomplete representation of the signals passed from the growth hormone through the growth hormone receptor (GHR) which then stimulates either the JAK or the ERK1/2. The interaction between GHR and UBC has no direction therefore it is impossible to predict which molecule would have an effect on the other.*

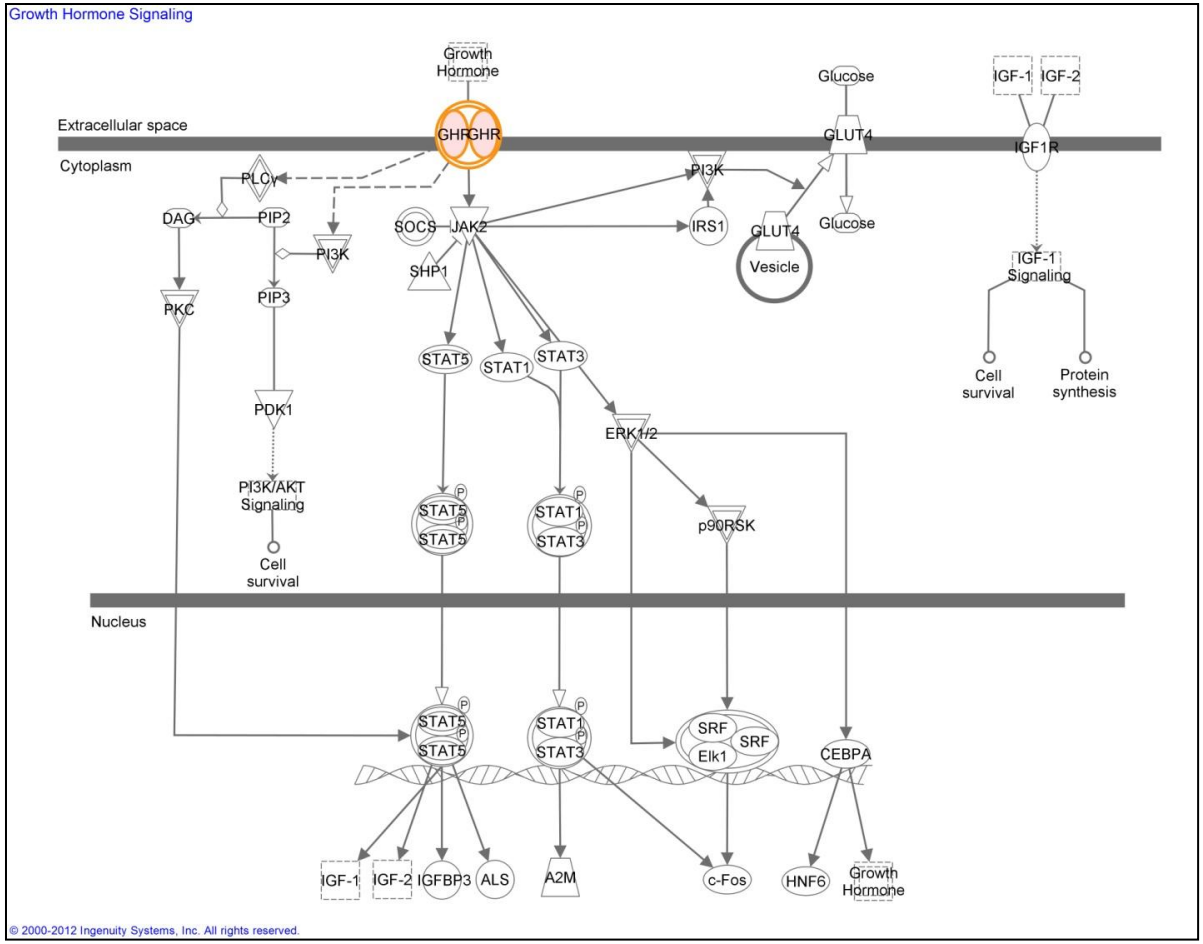



*Figure 7.2b: A detailed GH signaling pathway, showing additional candidate genes that may be associated with RFI. This canonical pathway was identified in IPA® (ingenuity systems). The STAT proteins enter the nucleus and act as transcription factors for IGF1, 2 and their binding proteins. The ERK1/2 stimulates CEBPA which acts as transcription factor for genes such as Leptin and Growth hormone.*

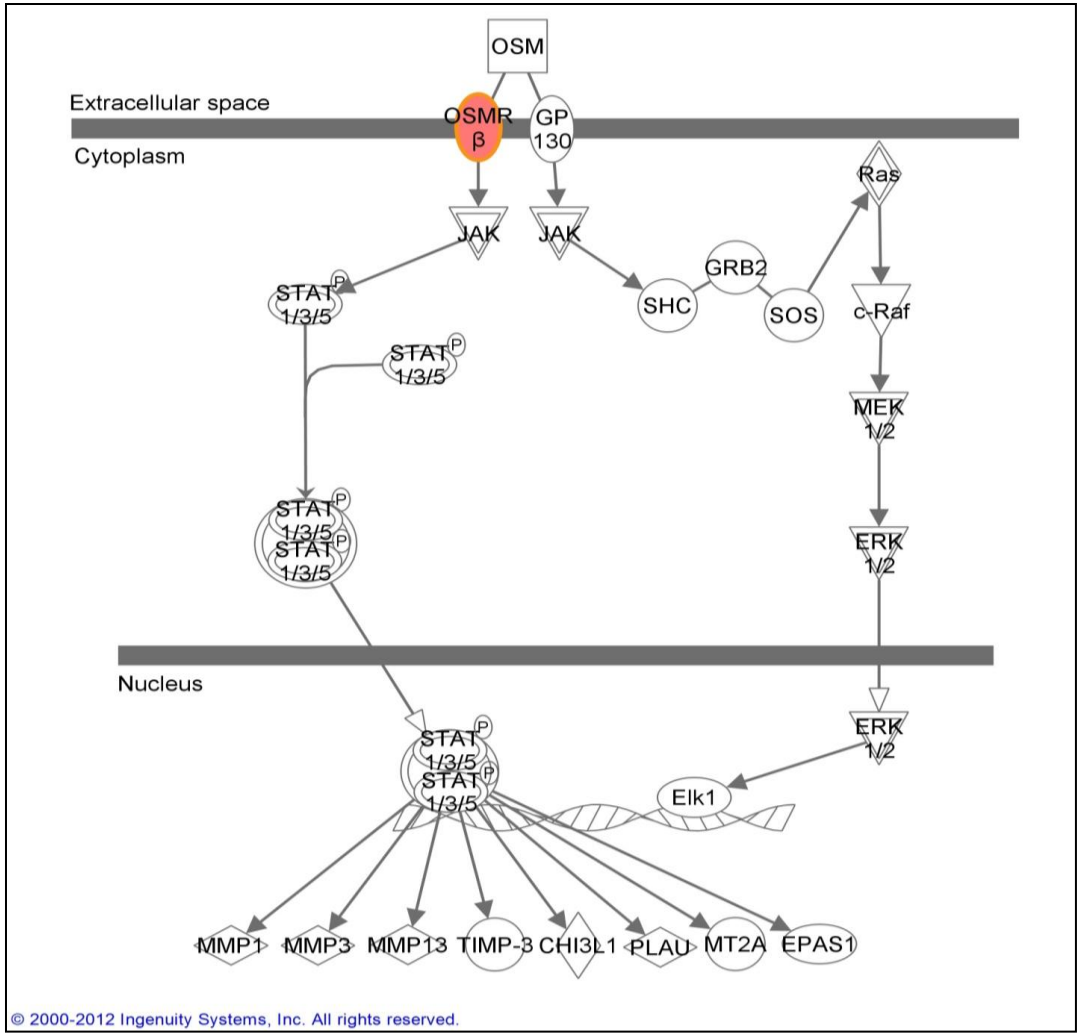



*Figure 7.3: The Oncostatin M signaling pathway and the Oncostatin M receptor. The Oncostatin M signaling pathway also uses one of the Janus kinases (JAK), which also stimulates STAT5. This pathway differs from the growth hormone signaling pathway by the genes whose expression has been affected. While the GH signaling pathways influence the IGF genes, the Oncostatin M signaling pathway influences the genes that encode the Matrix metallopeptidases.*

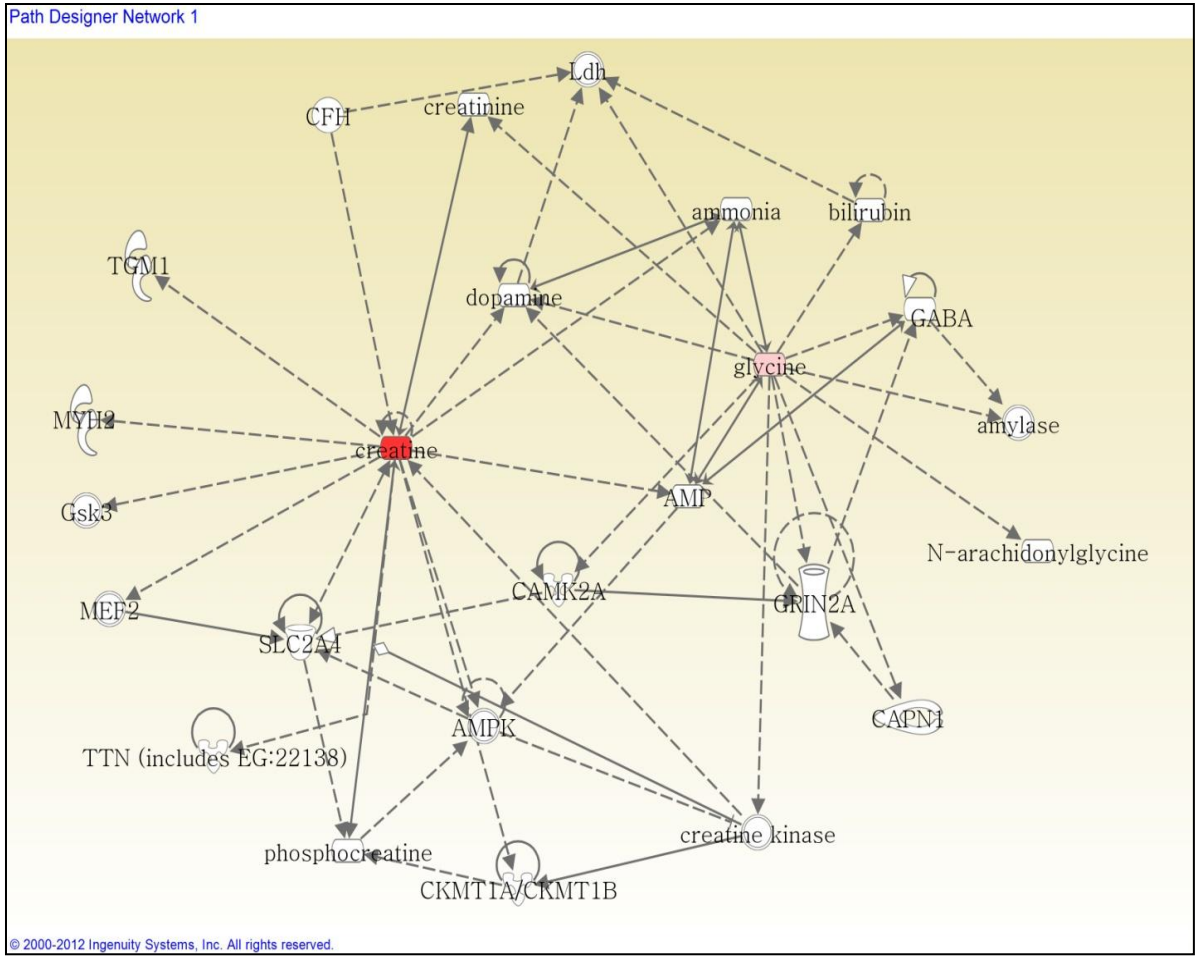



*Figure 7.4: Metabolic network reconstructed from metabolites significantly associated with RFI in period 1. Creatine interacts with AMPK, which plays a major role in energy metabolism including processes such as lipolysis, glycogen metabolism and glucose transport. The SLC2A4 (GLUT4) is the transporter protein involved in glucose transport. Other molecules such as TTN and MYH2 may be involved in muscle contraction thereby utilizing glucose and increasing the levels of AMP which stimulates the AMPK signaling*

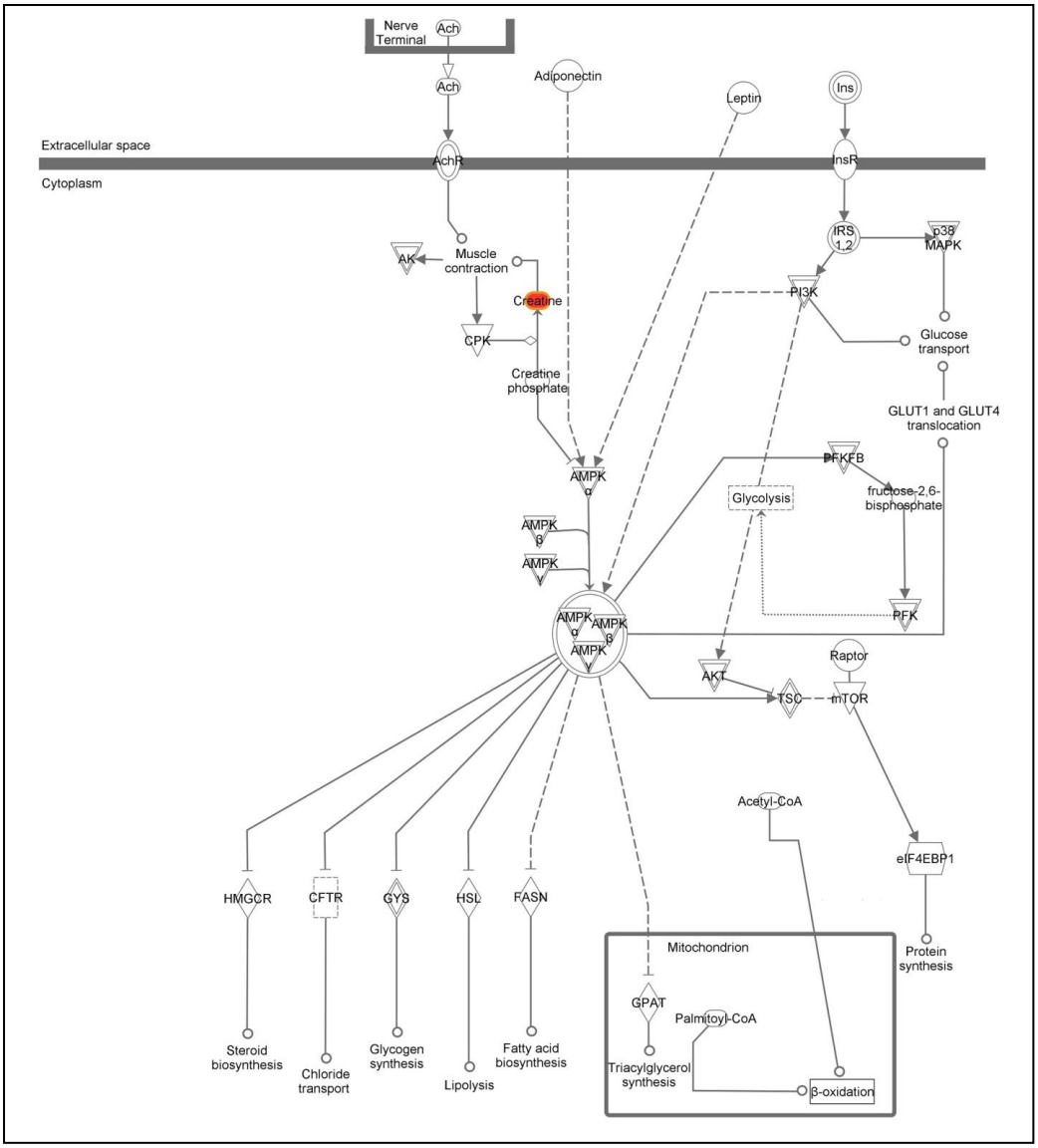



*Figure 7.5: The AMPK signaling pathway, one of the canonical pathways for metabolites significantly associated with RFI in period 1. Several signals can initiate the AMPK pathway including Leptin, Adiponectin and levels of AMP in the cells. AMPK has an effect on glycogen synthesis, steroid biosynthesis, lipolysis and fatty acid biosynthesis. In the mitochondria, AMPK signaling pathway stimulates triacylglycerol synthesis. It also results in β-oxidation which forms Acetyl CoA. The AMPK signaling pathway can also result in protein synthesis and glucose transport.*

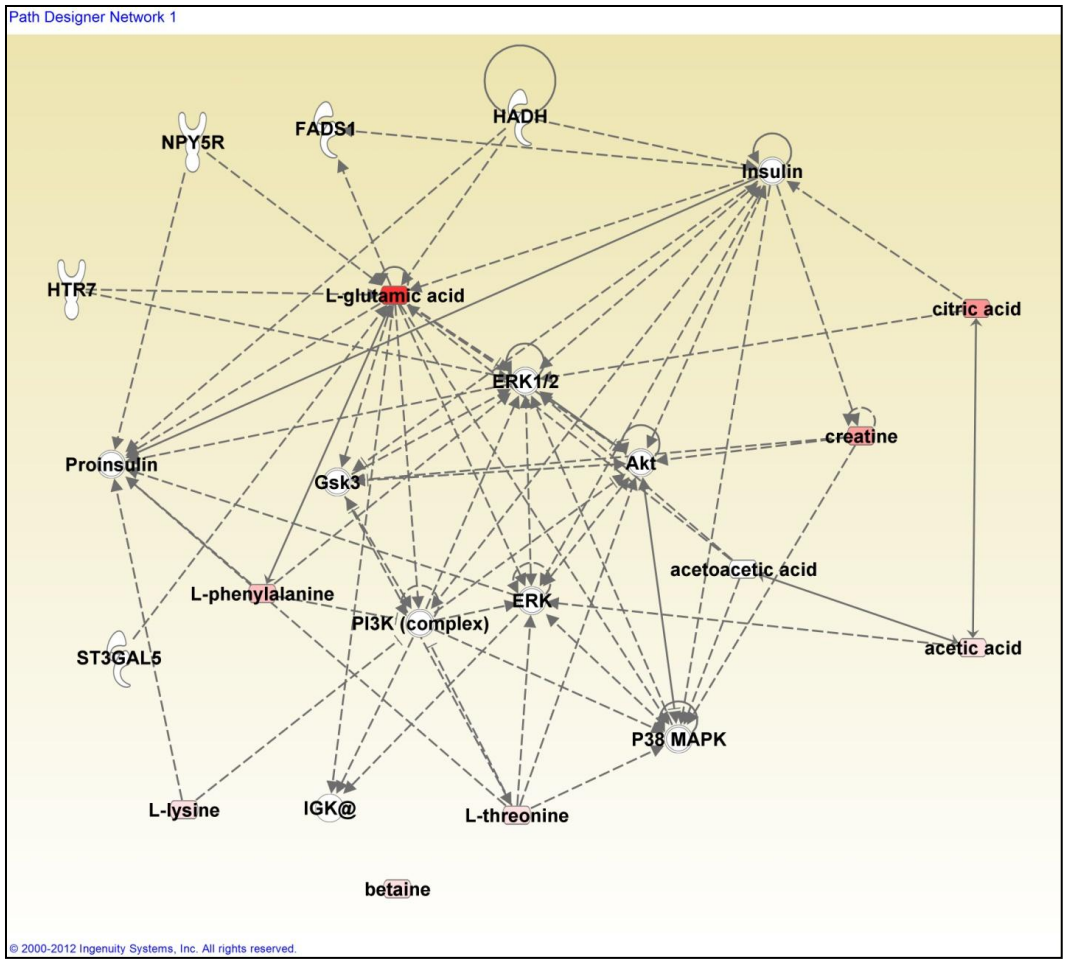



*Figure 7.6: Metabolic network reconstructed from metabolites significantly associated with RFI in period 2. Some of the metabolites associated with RFI in period 2 shown in this network indicate the role of the TCA cycle in regulating RFI. Levels of Acetate, which may influence levels of Acetyl CoA, were associated with RFI and interacted closely with levels of citrate. Citrate being the product of the reaction between Acetyl CoA and oxaloacetate also indicates the effects of the TCA cycle on RFI. These reactions may indicate an effect on the amount of ATP produced by steers with differing RFI levels.*

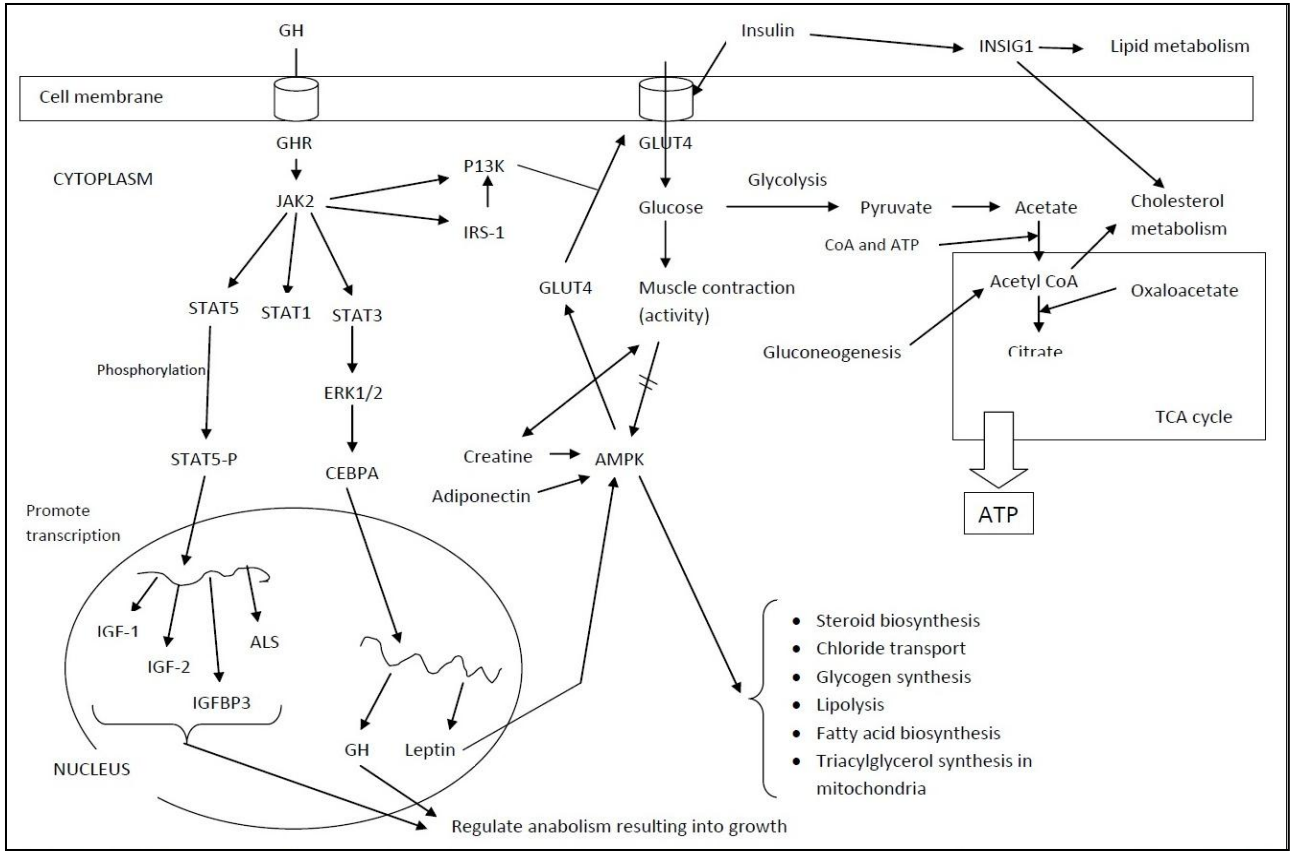



*Figure 7.7: A summary of the interaction between some genes, metabolites and biological processes associated with RFI.*

## 7.6. REFERENCES


Archer JA, Richardson EC, Herd RM, Arthur PF. 1999. Potential for selection to improve efficiency of feed use in beef cattle: A review. Australian Journal of Agricultural Research **50**, 147–161.

Arthur PF, Archer JA, Johnston DJ, Herd RM, Richardson EC, Parnell PF. 2001. Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. Journal of Animal Science **79**, 2805–2811.

Arthur PF, Archer JA, Herd RM. 2004. Feed intake and efficiency in beef cattle: Overview of recent Australian research and challenges for future. Australian Journal of Experimental Agriculture **44**, 361–369.

Arthur PF, Herd RM. 2006. Selection for growth and feed efficiency—The Australian experience. Journal of Integrated Field Science **3,** 59–66.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry MJ, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, and Sherlock G. 2000. Gene ontology: tool for the unification of biology. Nature Genetics. **25**:25-29.

Barabasi A, Oltvai ZN. 2004. Network biology: Understanding the cell's functional organization. Nature reviews. Genetics **5**, 101 - 113.

Barendse W, Reverter A, Bunch RJ, Harrison BE, Barris W, Thomas MB. 2007. A validated whole-genome association study of efficient food conversion in cattle. Genetics **176,** 1893–1905.

Basarab JA, Price MA, Aalhus JL, Okine EK, Snelling WM, Lyle KL. 2003. Residual feed intake and body composition in young growing cattle. Canadian Journal of Animal Science **83**, 189–204.

Bishop MD, Simmen RCM, Simmen FA, Davis ME. 1989. The relationship of insulin-like growth factor-I with post-weaning performance in Angus beef cattle. Journal of Animal Science **67**, 2872–2880.

Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, Chen RO, Brownstein BH, Cobb JP, Tschoeke SK, Miller-Graziano C, Moldawer LL, Mindrinos MN, Davis RW, Tompkins RG, Lowry SF. 2005. A network-based analysis of systemic inflammation in humans. Nature 437(7061), 1032-1037.

Canadian Council on Animal Care. 1993. Guide to the care and use of experimental animals. Vol. 1. Olfert ED, Cross BM, McWilliams AA. eds. 2nd ed. CCAC, Ottawa, ON.

Connolly MA, Clausen PA and Lazar JG. 2006. Purification of RNA from Animal Cells Using Trizol. Cold Spring Harb Protoc, pdb.prot4104.

Davis ME, Simmen RCM. 2006. Genetic parameter estimates for serum insulin-like growth factor I concentrations, and body weight and weight gains in

Angus beef cattle divergently selected for serum insulin-like growth factor I concentration. Journal of Animal Science **84,** 2299–2308.

De Haer LCM, Luiting P, Aarts HLM. 1993. Relations among individual (residual) feed intake, growth performance and feed intake pattern of growing pigs in group housing. Livestock Production Science **36,** 233-253.

Espenshade PJ, Hughes AL. 2007. Regulation of sterol synthesis in eukaryotes. Annual Review of Genetics **41**, 401–427.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2009. ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, UK.

Grant JR, Arantes AS, Xiaoping L. and Stothard P. 2011. In-depth annotation of SNPs arising from resequencing projects using NGS-SNP. Bioinformatics. **27**:2300-2301.

Han JD. 2008. Understanding biological functions through molecular networks Cell Research **18**, 224–237.

Herd RM, Archer JA, Arthur PF. 2003. Reducing the cost of beef production through genetic improvement of feed intake: Opportunity and challenges to application. Journal of Animal Science **81**, E9–E17.

Herd RM, Arthur PF. 2009. Physiological basis for residual feed intake. Journal of Animal Science 87 **(14), E64-E71.**

Hoque MA, Katoh K, Suzuki K. 2009. Genetic associations of residual feed intake with serum insulin-like growth factor-I and leptin concentrations,

meat quality, and carcass cross sectional fat area ratios in Duroc pigs. Journal of Animal Science **87**(10), 3069-75.

Hu Z, Park CA, Fritz ER, Reecy JM. 2010. QTLdb: A Comprehensive Database Tool Building Bridges between Genotypes and Phenotypes. Invited Lecture with full paper published electronically on The 9th World Congress on Genetics Applied to Livestock Production. Leipzig, Germany August 1-6, 2010.

Hubbard, T. J. 2009. Ensembl 2009. Nucleic Acids Research. **37,** D690-D697.

Kelly AK, McGee M, Crews Jr. DH, Fahey AG, Wylie AR. Kenny DA.. 2009. Effect of divergence in residual feed intake on feeding behavior, blood metabolic variables, and body composition traits in growing beef heifers. Journal of Animal Science **88,** 109-123.

Korver S. 1988. Genetic aspects of feed intake and feed efficiency in dairy cattle: A review. Livestock Production Science **20**, 1-13.

Letto J, Brosnan ME, Brosnan JT. 1986. Valine metabolism. Gluconeogenesis from 3-hydroxyisobutyrate. Biochemistry Journal **240**, 909-912.

Li, H, Ruan J, and Durbin R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Research. **18,** 1851-1858.

Liscum L. 2002. Cholesterol biosynthesis. Chapter 5 in D.E. Vance and J.E. Vance (Eds.) Biochemistry of Lipids, lipoproteins and Membranes 4th Edn Elsevier Science B.V.

MacNeil MD, Nugent RA, Snelling WM. 1997. Breeding for Profit: an Introduction to Selection Index Concepts. Range Beef Cow Symposium. Paper 142.

Montanholi YR, Swanson KC, Palme R, Schenkel FS, McBride BW, Lu D. and Miller SP. 2010. Assessing feed efficiency in beef steers through feeding behavior, infrared thermography and glucocorticoids. Animal. **4**(5), 692 – 701.

Moore SS, Crews Jr DH, Nkrumah JD. 2006. Multiple and candidate gene approaches to genetic evaluation of feed efficiency in beef cattle. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production. Brazilian Society of Animal Breeding, Sao Paolo, Brazil. No. 14–01, Belo Horizonte, Brazil.

Mujibi FD, Moore SS, Nkrumah JD, Wang Z and Basarab JA. 2010. Season of testing and its effect on feed intake and efficiency in growing beef cattle. Journal of Animal Science. **88**, 3789-99.

Mujibi FDN, Nkrumah JD, Durunna ON, Grant JR, Mah J, Wang Z, Basarab JA, Plastow G, Crews Jr DH, Moore SS. 2011. Associations of marker panel scores with feed intake and efficiency traits in beef cattle using pre-selected

single nucleotide polymorphisms. Journal of Animal Science **89** (11), 3362-71.

Mullen E, Ohlendieck K. 2010. Proteomic profiling of non-obese type 2 diabetic skeletal muscle. International Journal of Molecular Medicine **25** (3), 445-458.

Naik M. 2007. Identification and characterization of genetic markers and metabolic pathways controlling net feed efficiency in beef cattle. PhD Thesis, University of Adelaide University, Australia

Nkrumah JD, Ammoura A, Guercio S, Lyle KL, Price MA, Okine EK, Basarab JA, Hansen C, Li C, Murdoch B Moore SS. 2003. Relationships between residual feed intake, feed intake, growth traits, ultrasound carcass measurements and feeding behavior in young hybrid beef calves. Canadian Society of Animal Science Abstract, June 10-13, Saskatoon, SK, Canada.

Nkrumah JD, Basarab JA, Price MA, Okine EK, Ammoura A, Guercio S, Hansen C, Li C, Benkel B, Murdoch B, Moore SS. 2004. Different measures of energetic efficiency and their phenotypic relationships with growth, feed intake, and ultrasound and carcass merit in hybrid cattle. Journal of Animal Science **82** (8), 2451-2459.

Nkrumah JD, Li C, Yu J, Hansen C, Keisler DH, Moore SS. 2005. Polymorphisms in the bovine leptin promoter associated with serum leptin concentration, growth, feed intake, feeding behavior, and measures of carcass merit. Journal of Animal Science **83,** 20-28.

Nkrumah JD, Sherman EL, Li C, Marques E, Crews Jr DH, Bartusiak R, Murdoch B, Wang Z, Basarab JA, Moore SS. 2007a. Primary genome scan to identify putative quantitative trait loci for feedlot growth rate, feed intake, and feed efficiency of beef cattle. Journal of Animal Science **85**, 3170–3181.

Nkrumah JD, Crews DH Jr, Basarab JA, Price MA, Okine EK, Wang Z, Li C, Moore SS. 2007b. Genetic and phenotypic relationships of feed intake and different measures of feed efficiency with growth and carcass merit of beef cattle. Journal of Animal Science **85**, 2711–2720.

Nafikov RA, Beitz DC. 2007. Carbohydrate and Lipid Metabolism in Farm Animals. Journal of Nutrition 137 (3), 702-705.

Ouchi N. Shibata R, Walsh K. 2005. AMP-activated protein kinase signaling stimulates VEGF expression and angiogenesis in skeletal muscle. Circulation Research **96** (8), 838–46.

Preston TR, Leng RA. 1987. Matching ruminant production systems with available resources in the tropics and sub-tropics. Penambul Books Armidale.

Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. 1997. GeneCards: integrating information about genes, proteins and diseases. Trends in Genetics **13** (4), 163.

Richardson EC, Herd RM. 2004. Biological basis for variation in residual feed intake in beef cattle. 2. Synthesis of results following divergent selection. Australian Journal of Experimental Agriculture **44**, 431–440.

Schenkel FS, Miller SP, Wilton JW. 2004. Genetic parameters and breed differences for feed efficiency, growth and body composition traits of young beef bulls. Canadian Journal Animal Science **84**, 177–185.

Sherman EL, Nkrumah JD, Murdoch BM, Li C, Wang Z, Fu A, Moore SS. 2008a. Polymorphisms and haplotypes in the bovine neuropeptide Y, growth hormone receptor, ghrelin, insulin-like growth factor 2, and uncoupling proteins 2 and 3 genes and their associations with measures of growth, performance, feed efficiency, and carcass merit in beef cattle. Journal of Animal Science **86**, 1–16.

Sherman EL, Nkrumah JD, Murdoch BM, Moore SS. 2008b. Identification of polymorphisms influencing feed intake and efficiency in beef cattle. Animal Genetics **39**, 225–231.

Snelling WM, Cushman RA, Keele JW, Maltecca C, Thomas MG, Fortes MRS, Reverter A (2013) Networks and pathways to guide genomic selection. Journal of Animal Science. **91**, 537-552.

Stick DA, Davis ME, Loerch SC, Simmen RCM. 1998. Relationship between blood serum insulin-like growth factor 1 concentration and postweaning feed efficiency of crossbred cattle at three levels of dietary intake. Journal of Animal Science **76**, 498–505.

Viegas CM, da Costa FG, Schuck PF, Tonin AM, Zanatta A, de Souza Wyse AT, Dutra-Filho C S, Wannmacher CM, Wajner M. 2008. Evidence that 3-hydroxyisobutyric acid inhibits key enzymes of energy metabolism in cerebral cortex of young rats. International Journal of Developmental Neuroscience **26** (3-4), 293-9.

Winder WW, Hardie DG. 1999. AMP-activated protein kinase, a metabolic master switch: possible roles in type 2 diabetes. American Journal of Physiology **277** (1 Pt 1), E1–10.

# CHAPTER 8

## 8.1. SUMMARY AND GENERAL DISCUSSION

Ideally, a holistic understanding of the regulation of feed efficiency would have to consider all the levels of regulation of phenotypes as shown in Figure 2.1. This ability to connect phenotypes to genotypes is one of the fundamental objectives in genetics research (Botstein and Risch, 2003). One of the most common techniques used to perform genotype-phenotype association studies is the candidate gene approach (Zhu and Zhao, 2007). This approach relies on prior knowledge of the physiology of the trait and the functions of the genes. In this study, the candidate genes to be considered for the analysis were selected based on prior knowledge obtained from QTL analyses performed and stored in the QTL database (QTLdb) (Hu et al., 2007) and the results of a genome wide association study (GWAS) reported by Mujibi et al., (2011). The QTLs considered were associated with feed intake/dry matter intake, average daily gain, feed conversion ratio, energy balance and body weight. The choice of these QTLs was based on the physiology of RFI, which was reviewed in detail by Herd and Arthur (2009). The candidate gene approach, discussed in Chapter 3, involved identifying positional candidate genes, which are all the genes located in the region of the QTL (or from the GWAS) about 500Kbp upstream and downstream of the position of the QTL or marker. These were defined as the positional candidate genes. The positional candidate genes were then prioritized according to their biological function such that genes whose functions were closely related to the

physiology of feed efficiency were retained for further analysis and were defined as functional candidate genes. We acknowledge the fact that there was a possibility to exclude genes whose known functions were not related to feed efficiency but would have had a role in the variation in feed efficiency. This would especially arise when the function of the gene was not totally understood, or the physiological basis of feed efficiency was not totally understood. A total of 117 functional SNPs were identified within the functional candidate genes in a process described in detail in Chapter 4. Steers from the University of Alberta ranch at Kinsella, were used for all the candidate genes studies in this project.

In the first study, Chapter 4, a multiple marker association analysis was performed to identify the SNPs significantly associated with RFI. The total population was sub-divided into a training set and a validation set based on the year the steers were born and pedigree information such that there were no common sires between the two subsets of the population. Twenty two SNPs were associated with RFI in the discovery population ($P < 0.05$) accounting for 18.1% of the variation in RFI. Seven of the 22 SNPs were also associated with RFI in the validation population ($P < 0.05$) and 25 SNPs were significantly associated with RFI in the pooled population. Using SIFT® software (Ng and Henikoff, 2003), 3 of the 22 significant SNPs were predicted to cause a significant effect on protein function ($P < 0.05$). One of the three SNPs was located in the GHR gene and was associated with a significant effect on the tertiary structure of the GHR protein ($P < 0.05$) as modeled using SWISSModel® software (Arnold et al., 2006). The GHR gene was reported to be associated with RFI by Sherman et al., (2008). However,

in contrast with this study, the SNPs reported by Sherman et al (2008) were located in the introns and we may not be able to explain how they function to affect RFI.

The other genes that were significantly associated with RFI included SLC45A2, AOX1, MYO10, NECAP2, OCLN, OSMR, UMPS, CAST, ACAD11, UGT3A1, CYP2B, LRP5. A complete list of the genes can be obtained in Chapter 4. Some genes such as NECAP2 and OSMR showed additive effect while other genes such as GHR and CAST showed a dominance effect.

This study resulted in a total of 25 genes which were associated with RFI and, if validated across different genetically diverse populations, can be used to develop a marker assisted selection tool to select beef cattle for RFI.

Although RFI is phenotypically independent of some production traits (Archer et al., 1999), studies have shown that it is genetically correlated with some of them (Kennedy et al., 1993), indicating that the biological processes that cause variation in RFI also cause variation in other production traits. Chapter 5 was therefore aimed at testing whether the functional candidate genes identified for RFI would be significantly associated with carcass quality traits in the same population of steers. A multiple marker association analysis was used to test this hypothesis. Seven (out of the 117) SNPs were significantly associated with various carcass quality traits ($p \leq 0.005$) and three of them were the same as those that were significantly associated with RFI ($p \leq 0.005$). The GHR, SLC45A2 and the ARHGAP1 genes were significantly associated with both RFI and at least one

of the carcass traits. The other genes associated with carcass traits included DGKH, HSD17B12, LRP4 and NDUFS3. The DGKH and HSD17B12 genes are involved in lipid and cholesterol metabolism among other functions and may partially explain the positive correlation reported between RFI and average back fat and intramuscular fat (Basarab et al., 2003). Therefore if these 3 pleiotropic markers are to be incorporated into a marker assisted selection panel for RFI, there will be need to estimate the effects that the favorable genotypes will have on the relevant carcass traits and if the effects are antagonistic, breeders should ensure that the economic benefits outweigh the possible losses.

Metabolites have the potential to be used as markers for selection of beef cattle. There are some studies reporting significant associations between a trait, a metabolite and a gene(s). For example; Weikard et al., (2010) reported significant association between mutations in the *NCAPG* and *GDF8* genes and average daily gain in two independent populations of cattle. They also observed a significant association between the mutation in NCAPG with the metabolites carnitine, arginine and total dimethylarginine levels in plasma. In a separate study, Pliakogiannis et al., (1993) reported a significant association between carnitine and triglyceride levels in plasma and body weight.

Based on these results and the potential for using blood metabolites to test for productivity traits in cattle, the objective of Chapter 6 was to identify blood metabolites associated with variation in residual feed intake (RFI) in beef cattle. The metabolites were assessed at three time points to identify metabolites associated with RFI in each period. In addition, there was need to use a mixed

model for repeated measures analysis to identify the metabolites significantly associated with RFI throughout the feeding period. To achieve these objectives, blood samples were collected through jugular venipuncture from 32 (16 high and 16 low RFI) steers at three time points (periods). Period 1, 2 and 3 corresponded to 2, 6 and 9 weeks into the feeding period respectively. The metabolite concentrations were determined using nuclear magnetic resonance (NMR). An analysis of variance was performed in two ways: for each period separately to identify metabolites associated with RFI in each period and using a mixed model for repeated measures to identify metabolites associated with RFI irrespective of period. The results indicated that at period 1, creatine and glycine were significant ($P < 0.05$) and accounted for 36% of the phenotypic variation in RFI. In period 2, 10 metabolites (details in Chapter 6) were significant ($P < 0.05$) and accounted for 74% of the variation in RFI and at period 3, hydroxyisobutyrate, fumarate and tyrosine were significant ($P < 0.05$) and accounted for 52% of the variation in RFI. In a mixed model for repeated measures, 2 metabolites were significantly ($P < 0.05$) associated with RFI across all periods. These results indicate that period 2 would be the best period to use the metabolites to test for RFI as it results in highest accuracy. However, this period required 10 metabolites to achieve that accuracy. Although period 3 had relatively low accuracy, it only required 3 metabolites to achieve it.

The metabolite creatine was associated with RFI in period 1 and 2, and in the entire feeding period. Creatine is a source of energy in cells especially in muscles and brain. It is formed from glycine (which was associated with RFI in period 1)

and arginine. A good review of the role of creatine in energy metabolism is provided by Wallimann et al., (2011). The creatine system connects the processes that produce ATP such as glycolysis and oxidative phosphorylation to the processes that utilize ATP. Metabolites which were closely involved in the processes that produce ATP were citrate and acetate, which were significantly associated with RFI in period 2. Carnitine also was significantly associated with RFI in period 2. Similar results indicated that carnitine was associated with average daily gain (Weikard et al., 2010). The other metabolites associated with RFI were amino acids, and we hypothesize that they may be involved in either protein synthesis or gluconeogenesis.

The analysis of metabolites was followed further with a validation analysis performed using steers at the University of Alberta in Kinsella. The results indicated that 3 metabolites were significant in both discovery and validation populations and accounted for 32% of the variation in RFI in the validation population in period 2. There were no metabolites significant in both populations in period 1. The differences observed between the populations may have resulted from among other reasons, the small number of steers used. Therefore there is need to identify more metabolites that may be associated with RFI using a larger population size and at an earlier age during the development of the cattle.

Chapter 7 was aimed at consolidating the results reported in the previous chapters into a condensed form focusing more on the interactions between the genes, metabolites and their biological functions. The reconstruction of gene and metabolic networks was made to address these interactions and to indicate the

possibility of using them to predict the effects of these molecules on other molecules and traits. The major canonical pathways included the growth hormone signaling (including protein anabolism), Insulin mediated pathways, AMPK signaling and cholesterol biosynthesis (Chapter 7). The biological processes identified in this study were similar to those identified by Richardson and Herd (2004) and Herd and Arthur (2009). In addition, the genes, metabolites and biological processes identified indicated that the biosynthesis of cholesterol may vary between beef cattle with varying RFI levels. An exposition of this relationship was provided in Chapter 7.

## 8.2. GENERAL CONCLUSIONS AND RECOMMENDATIONS

The results in this study indicate that RFI is a complex trait that is regulated by several genes and biological processes. Some of the genes significantly associated with RFI have been reported in this study. However, we recommend that these genes be tested for association in a different population consisting of diverse breeds of cattle before they can be used to develop marker assisted selection panels for those populations. A similar assessment was performed using population structure and genetic diversity measures in Chapter 9.

This study has also identified metabolites as an alternative in selection for RFI. This concept is still in its infancy but in this study, we show that it has a great potential for predicting RFI with relatively high accuracy. We recommend that the metabolites identified in this study be validated in other populations to assess their consistency across different populations and breeds. There is need to identify metabolites associated with RFI at an earlier age. This may be achieved by

collecting blood samples from calves at different ages and performing a metabolite profile for each calf. This profile could be used later to perform analysis of variance when the RFI values are available. However, caution should be taken to ensure a difference in the diet offered to the animals does not influence the metabolite levels. This may be done by ensuring animals are offered the same diet or by identifying metabolites which are significantly associated with RFI irrespective of the diet.

The details provided on the biological processes affected by the significant genes and metabolites were aimed at obtaining a greater understanding of the underlying factors that cause variation in RFI and giving insights on the possible pleiotropic effects of the genes and metabolites.

This pleiotropy was tested further using the genes that were associated with RFI to assess their effects on carcass traits. As expected, 3 of the 25 genes that were associated with RFI were also associated with some carcass traits. This was a small proportion but is still important as it indicates that selection for RFI using these genes may have an effect on carcass traits. We recommend that breeders who may wish to use these genes to cautiously look out for these pleiotropic effects.

### 8.3. OUTPUTS AND IMPLICATIONS IN THE BEEF INDUSTRY

From this study:

- We now have an additional 25 SNPs that can be tested further for use in marker assisted selection for RFI in beef cattle. These markers account for

a larger proportion of the phenotypic variance than most markers reported in the past.

- We also reported plasma metabolites that can be used to predict RFI in beef cattle at different stages during the feeding test period

- We tested the pleiotropic effects of the significant genes and found that about 12% (3 of 25) of the genes are also significantly associated with some carcass traits. The effects on carcass traits may be favorable depending on the breeding goals of the breeder therefore we caution that breeders who use the 3 genes should be aware of their possible effects on carcass traits as well as RFI.

- We have also provided a comprehensive analysis of the biological processes underlying the variation in RFI. This analysis is expected to give insights into the nature of this trait and predict the other processes and traits that may be affected if selection for RFI is practiced.

## 8.4. REFERENCES


Archer, J. A., E. C. Richardson, R. M. Herd, and P. F. Arthur. 1999. Potential for selection to improve efficiency of feed use in beef cattle: A review. Aust. J. Agric. Res. 50:147–161.

Arnold, K., B. Lorenza, K. Jürgen, and S. Torsten. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling Structural bioinformatics Vol. 22 no. 2 pages 195–201.

Basarab, J. A., M. A. Price, J. L. Aalhus, E. K. Okine, W. M. Snelling, and K. L. Lyle. 2003. Residual feed intake and body composition in young growing cattle. Can. J. Anim. Sci.; 83:189-204.

Botstein, D. and N. Risch. 2003. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet; 33 Suppl: 228-37.

Herd, R. M. and P. F. Arthur. 2009. Physiological basis for residual feed intake. Journal of Animal Science, vol. 87 no. 14 suppl E64-E71.



Herd, R. M. and P. F. Arthur. 2009. Physiological basis for residual feed intake. Journal of Animal Science, vol. 87 no. 14 suppl E64-E71.



Hu, Z., C. A. Park, E. R. Fritz, and J. M. Reecy. 2010. QTLdb: A Comprehensive Database Tool Building Bridges between Genotypes and Phenotypes. Invited Lecture with full paper published electronically on

The 9th World Congress on Genetics Applied to Livestock Production. Leipzig, Germany August 1-6, 2010.

Kennedy, B. W., J. H. J. van der Werf, and T. H. E. Meuwissen. 1993. Genetic and statistical properties of residual feed intake. J. Anim. Sci.; 71:3239–3250.

Mujibi, F. D. N., J. D. Nkrumah, O. N. Durunna, J. R. Grant, J. Mah, Z. Wang, J. Basarab, G. Plastow, D. H. Crews, Jr. and S. S. Moore. 2011. Associations of marker panel scores with feed intake and efficiency traits in beef cattle using pre-selected single nucleotide polymorphisms. J Anim Sci.; 89 (11): 3362-71.

Ng, P. C and S. Henikoff. 2003. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Research, Vol. 31, No. 13 3812-3814.

Pliakogiannis, T., Chatzidimitriou, C., Evangeliou, A., Böhles, H. J., Kalaitzidis, K. 1993. Serum carnitine levels, lipid profile, and metabolic status of patients on continuous ambulatory peritoneal dialysis. Perit. Dial. Int. 13 Suppl 2:S440-3.

Sherman, E. L., J. D. Nkrumah, B. M. Murdoch, and S. S. Moore. 2008. Identification of polymorphisms influencing feed intake and efficiency in beef cattle. Anim. Genet. 39:225–231.

Wallimann, T., M. Tokarska-Schlattner, and U. Schlattner. 2011. The creatine kinase system and pleiotropic effects of creatine. Amino acids; 40 (5): 1271 – 1296.

Weikard, R., E. Altmaier, K. Suhre, K.M. Weinberger, H.M. Hammon, E. Albrecht, K. Setoguchi, A. Takasuga, and C. Kühn. 2010. Metabolomic profiles indicate distinct physiological pathways affected by two loci with major divergent effect on Bos taurus growth and lipid deposition. Physiol Genomics 42A: 79–88.

Zhu M, Zhao S. 2007. Candidate Gene Identification Approach: Progress and Challenges. Int. J. Biol. Sci. 3(7):420-427

# CHAPTER 9

## SUPPLEMENTARY WORK

**A genetic diversity analysis and its effects on the results obtained from association studies in reference to the University of Alberta beef ranch at Kinsella, Canada**

### 9.1. INTRODUCTION

The province of Alberta, with about 4.95 million cattle, has the largest cattle population in Canada, which stands at about 12.46 million (Statistics Canada, 2012). The Canadian beef industry utilizes Bos taurus breeds, mainly Angus but also includes Charolais, Hereford, Simmental and Limousin breeds. Cross breeding is utilized to combine the best characteristics of several breeds into one animal and results in increased genetic diversity in the population (Amer et al., 1992). The genetic diversity of the beef cattle populations is key to the productivity and survival of the populations as well as the wellbeing of the producers. Genetic diversity also enables producers and breeders to select animals with preferred genotypes for breeding (Pariset et al., 2006). In addition to selection, the analysis of genetic diversity can be used to detect population structure existing in large populations.

Previously, studies focused on assessing genetic diversity and population structure in cattle were conducted using low-density microsatellites and Y-chromosome markers (Li et al., 2007). However, analysis of single nucleotide polymorphisms (SNP) is becoming the standard approach for population structure and genetic

diversity studies (Edea et al., 2013). SNPs are becoming more popular because they are abundant in the genome, they are genetically stable, and are amenable to high-throughput automated analysis (Vignal et al., 2002). The usefulness of SNPs in population structure and genetic diversity analysis has been demonstrated in several studies (McKay et al., 2008, The Bovine HapMap Consortium, 2009 and Lin et al., 2010).

With the advent of marker assisted selection (MAS), assessment of population structure and genetic diversity is important for several reasons:

Firstly, genetic markers found to be significantly associated with a trait in genetically diverse populations are expected to retain their significance across genetically stratified populations resulting in successful utilization across multiple breeds and populations. On the contrary, markers that show significant associations in specific populations of genetically related individuals may not be utilizable in genetically distant populations due to the decay of LD over several generations.

Secondly, genetically diverse populations could be split into distinct subpopulations containing more genetic uniformity within the sub-populations and genetically distant between the sub-populations. In association studies, one of the subpopulations could be used as a discovery population and the other, the most distant subpopulation, could be used as a validation population.

In this study, we used SNPs to analyse the genetic diversity and population structure of the beef cattle population at the University of Alberta beef ranch in

Kinsella, Canada. We then identified the probable number of populations within the genotyped individuals with the intention of determining whether a subset of the population was genetically distant enough to establish it as a genetically distinct population that could be used to validate markers identified as significant in the other subset of the population.

## 9.2. MATERIALS AND METHODS

### 9.2.1. Animal resources

The analysis was performed using genotypes obtained from 670 steers at the university of Alberta beef ranch in Kinsella, Canada. The composition of the population was described in detail by Goonewardene et al., (2003). In summary, the sires were mainly from 3 breeds; Angus, Charolais or Alberta hybrid bulls. Dams, on the other hand, were crosses developed from 3 composite cattle lines; Beef synthetic 1, Beef synthetic 2 and Dairy X Beef synthetic lines. The Beef synthetic lines were also composed of crosses of multiple breeds as described by Goonewardene et al., (2003).

### 9.2.2. Genes, genotyping and population structure analysis

The genes and SNPs used in this study were identified using the candidate gene approach as described in Chapter 4. A total of 117 SNPs were genotyped and used for the population structure analysis. The details of the identification of SNPs used in this analysis were also described in Chapter 4.

The specific objectives of this analysis included:

i. Determining the probable number of population clusters in the genotyped steers.

ii. Determine the relationship between the identified population clusters and the breed composition of the individuals in that cluster.

iii. Utilize the genetic diversity to divide the genotyped steers into a discovery and validation population by:

- Forcing the population into 2 clusters (*k = 2)* and determine the genetic diversity between the clusters, and whether the two clusters can be used as discovery and validation populations.

- Using the optimum number of clusters obtained (i) and select the subpopulations that are genetically diverse enough to be used as discovery and validation populations.

Genotype data was prepared and converted to the structure (.str) format using CONVERT program (Glaubitz, 2004).

Population structure analysis was performed using STRUCTURE software (Pritchard et al., 2000). The software assumes a model in which there are *k* populations (also called clusters), contributing to the genotype of each individual and each is characterized by a set of allele frequencies at each marker locus (Edea et al., 2013)

To determine the optimum number of clusters in the population, the *k* value was set between 2 and 12 with 10,000 iterations and a burn-in period of 10,000. We

performed 4 independent runs for each predefined number of population (*k = 2-12*). The optimum value of *k* was the one that resulted in the highest mean value of –log likelihood.

To assess the relationship between population clusters and breed composition, the *k* value was set to 3 to correspond with the number of sire breeds in this population (Angus, Charolais and Hybrid).

**9.3. RESULTS**

The number of clusters (*k*) that resulted in the highest mean value of –log likelihood was around *k = 10 - 11*, indicating that the population could be sub-divided into around 10 – 11 clusters (sub-populations). A graphical representation of the estimated membership coefficients (y-axis) to the 11 clusters is shown in Fig 9.1. Each individual is represented by a single vertical line broken into k colored segments whose lengths are proportional to each of the *k* inferred clusters. An analysis of the breed composition in each cluster indicated that all Angus steers were distributed across 4 clusters, Charolais steers belonged to 2 clusters and the Hybrid steers were located in 5 clusters.

When the population was forced to fit in 3 clusters (*k = 3),* the relationship between clustering and breed composition indicated that the three sire breeds occupied distinct clusters as shown in Figure 9.2. Cluster 1 corresponded to steers sired by Angus sires, cluster 2 corresponded to Charolais sires and cluster 3 corresponded to hybrid sires. Genetic heterogeneity was still observed (Figure

9.2) but the cluster-breed relationship was more defined than in the structure established by 11 clusters. Figure 9.3 is a triangle plot showing the distinct clusters of the three breeds in the population. The number of individuals in each breed/cluster was 205, 197 and 268 for cluster 1, 2 and 3 respectively.

When the population was forced into 2 clusters, in an attempt to establish a discovery and a validation population, the three breeds were distributed across both clusters. There were 382 and 288 individuals in cluster 1 and 2 respectively. The steers whose sires were hybrids seemed to have a higher probability of being in cluster 1. Most of the individuals with higher probability of belonging to cluster 2 seemed to have sires of the Charolais breed. Most individuals who had an almost equal probability of being in cluster 1 or 2 seemed to have sires in the Angus breed.

The allele frequency divergence between the two clusters was very low (0.0083) indicating that the two populations were genetically closely related. This frequency divergence was also low relative to the allele frequency divergence observed between the three breeds/clusters (0.016). The frequency divergence in the three clusters indicated that there was a relatively larger genetic distance between the individuals when clustered into three clusters than when clustered into two clusters.

When individuals were forced into 2 clusters, the average distances between individuals in the same cluster were 0.2 relative to the distance between individuals in the same cluster when clustered into 3 clusters (0.19). This

indicated that when individuals were forced into two clusters the individuals within the clusters were more genetically diverse. However, when the population was clustered into three clusters individuals within the same cluster were relatively more genetically similar.

When dividing the population into a discovery and validation population, the allele frequency divergence between clusters and the average distances between individuals within the same cluster indicated that the optimal clustering would be obtained if 3 clusters were used instead of 2; three clusters would achieve higher allele frequency divergence between clusters and lower genetic distance within clusters.

The greatest allele frequency divergence observed using the three clusters was between cluster 1 and 3 at 0.016. The allele frequency divergence between cluster 2 and 3, and 1 and 2 was relatively lower; 0.013 and 0.012 respectively. This indicated that cluster 3 was more closely related to cluster 2 than cluster 1 and cluster 2 was more closely related to cluster 1 compared to cluster 3.

Therefore, cluster 1 could be considered relatively genetically distinct from cluster 3 while cluster 2 was equally related to cluster 1 as it was to cluster 3. Cluster 1 may be used as a discovery population while cluster 3 could be used as the validation population, or vice versa. Because cluster 2 was equally related to clusters 1 and 3, it could be used as discovery or validation population. In this study we categorised cluster 2 as part of the discovery population to help increase the number of steers in the discovery population. This corresponded to steers sired

by Angus sires (cluster 1) belonging to the validation population and those sired by Charolais (cluster 2) and Hybrid sires (cluster 3) belonging to the discovery population.

## 9. 4. Application of population clustering in association analysis

### 9.4.1. Population composition and statistical analysis

The total number of steers that had both genotypic and corresponding phenotypic data was 531. These steers were born at the University of Alberta ranch in Kinsella, Canada, between 2003 and 2006. In each year, the feeding tests were conducted in spring and fall seasons, therefore each year had two batches of steers that were entered into the feeding test to evaluate their residual feed intake (RFI).

The population (n = 531) was divided into a discovery and a validation population such that the discovery population consisted all the individuals located in cluster 2 and 3 (sired by Charolais and Hybrid bulls respectively) while the validation population consisted of steers in cluster 1 (sired by Angus bulls). In total there were 348 steers in the discovery population and the validation population consisted of 183 steers.

A detailed description of the collection and calculations associated with the phenotypic data for RFI can be obtained in Chapter 4.

Association analysis for RFI was performed on the discovery population using multiple marker association analysis as described in Chapter 4. The SNPs significantly associated with RFI in the discovery population were validated in the

validation population using the multiple marker association analysis as described in Chapter 4.

### 9.4.2. Results and Discussion

Only 6 SNPs were significantly associated with RFI ($P < 0.05$) in the discovery population accounting for 6.4% of the phenotypic variation in RFI (Table 9.1)

One of the SNPs was located in the Occludin (OCLN) gene with genotypes CC, CT and TT. Genotypes CT and TT were both associated with negative (low) RFI values and thus were more desirable. The other SNPs were located in SLC30A5, MX1, SMPD1, DNAJC24 and OSMR genes (Table 9.1).

A comparison of the results obtained in this analysis and those obtained in Chapter 4 indicated that 2 genes (OCLN and OSMR) were significantly associated with RFI in both analyses. This may indicate that these SNPs retain significance across genetically diverse populations. To test this hypothesis a validation analysis was performed using the steers sired by Angus bulls (cluster 1). The results indicated that 3 of the 6 SNPs were still significantly associated ($P < 0.05$) with RFI and accounted for 2.9% of the variation in RFI in the validation population. These SNPs were located in the OCLN, SLC30A5 and OSMR genes. These results confirm that the 2 genes (OCLN and OSMR) remain significantly associated with RFI even in genetically diverse populations of beef cattle.

The SNPs significantly associated with RFI in this study accounted for a relatively small proportion of the total phenotypic variance, compared to the SNPs

associated with RFI in Chapter 4, which accounted for 19.7% of the phenotypic variance. This difference may have resulted from the difference in the number of significant SNPs associated with RFI; 25 SNPs accounted for 19.7% (in Chapter 4) while 6 SNPs accounted for 6.4%. Alternatively, the difference may have resulted from the clustering technique applied in the second analysis, which enabled the populations to consist of individuals with similar allele frequencies.

The results obtained from association analyses in beef cattle have frequently failed to be reproduced in populations composed of breeds that are different from the breed composition in the discovery population (Hayes and Goodard, 2010). This usually resulted from the differences in allele frequencies between the individuals in the discovery and validation populations. In addition, the failure to replicate association analysis results may result from the decay of linkage disequilibrium (LD) between markers and the causative genes from one breed to another (Hayes and Goodard, 2010). Because of the decay in LD, only a few SNPs will remain significant across different breeds as indicated in this analysis.

## 9.5. CONCLUSION

This study has shown that SNP genotypes can be used to determine the probable number of clusters (sub-populations) in a larger population consisting of genetically diverse individuals. The subpopulations differ in their allelic frequencies indicating the underlying genetic differences between them. In this study, the allelic divergence between clusters was relatively low (*around 0.01)* indicating that the individuals were largely genetically similar.

There was a consistent relationship between clusters and breed composition. Steers were located in specific clusters depending on the breed of their sire establishing three distinct clusters corresponding to the 3 breeds of sires present in the Kinsella population (fig 9.3).

Because of the small difference in allelic frequency divergence between cluster 2 and 3 and the high genetic distance between individuals within the same cluster in cluster 2 and 3, the 2 clusters were not sufficiently genetically different to allow them to be divided into 2 genetically distinct populations. However, despite this limitation, we have shown that it is possible to use the allelic divergence information obtained in this study to divide a population into a discovery and validation subset. This will especially be relevant if the population consists of more genetically diverse individuals and the two subsets of the population have larger differences in their allelic frequencies between them and lower differences between the individuals within the subpopulation.

The 3 SNPs shown to be significantly associated with RFI in both discovery and validation populations indicate that they may retain their significant association with RFI across genetically diverse populations of beef cattle including different breeds. However, these SNPs need to be tested in these genetically diverse breeds including *Bos indicus* to confirm this hypothesis.

**Table 9.1: SNPs significantly associated with RFI in the discovery population**

| **SNP** | Gene name | BTA | SNP alleles | SNP position | P-value |
|---|---|---|---|---|---|
| **MX1** | Interferon-regulated resistance GTP-binding protein MxA | 1 | C/T | 440 | 0.0090 |
| **SMPD1** | Sphingomyelin phosphodiesterase 1 | 15 | C/T | 1698 | 0.0119 |
| **DNAJC24** | DnaJ (Hsp40) homolog, subfamily C, member 2 | 15 | C/T | 143 | 0.0107 |
| **OCLN** | Occludin | 20 | C/T | 1051 | 0.0381 |
| **SLC30A5** | Solute carrier family 30) member 5 | 20 | A/G | 1666 | 0.0002 |
| **OSMR** | Oncostatin M receptor | 20 | G/T | 620 | 0.0301 |

Figure 9.1: Population structure of the Kinsella beef cattle population in 11 clusters

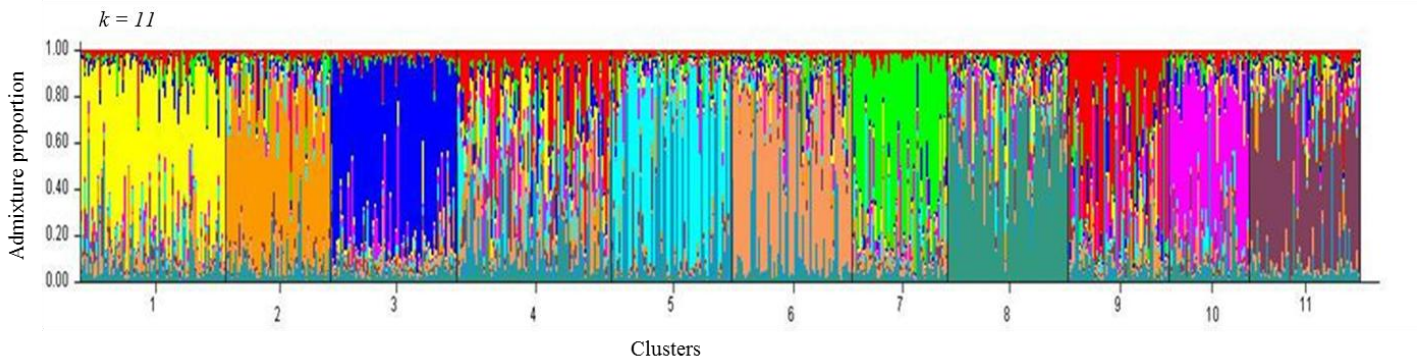
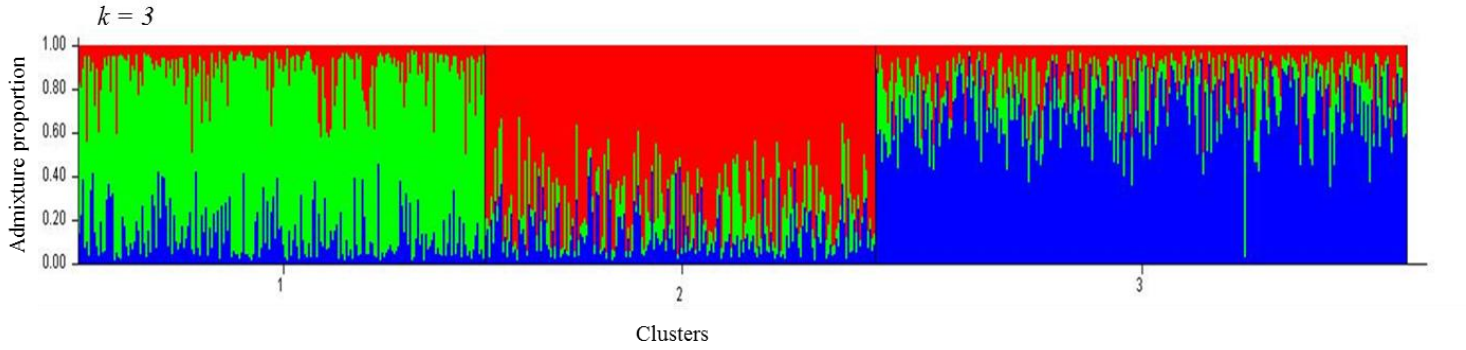



Figure 9.2: Population structure of the Kinsella beef cattle population in 3 clusters



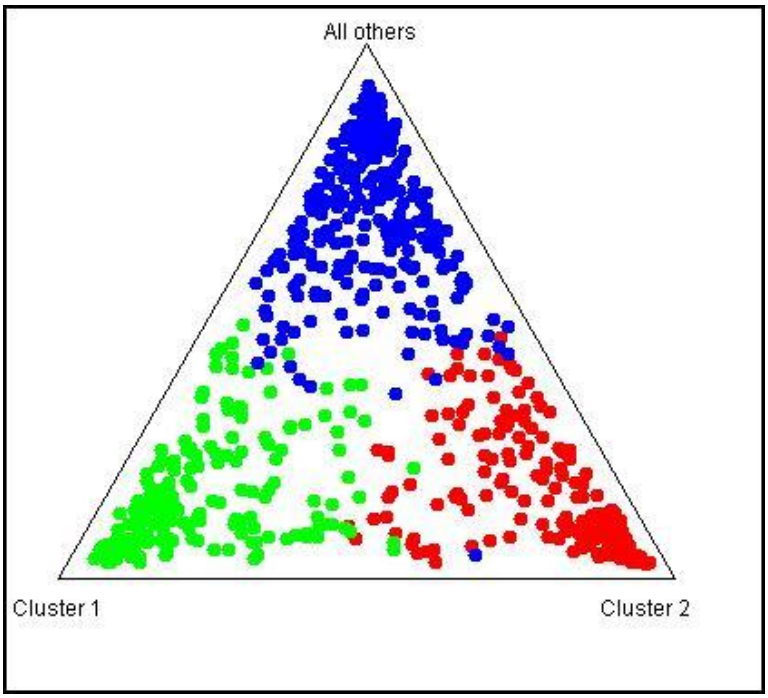



*Figure 9.3: Triangle plot of the population structure of the Kinsella population in 3 clusters. The individual, A, may have been placed in that cluster due to an error in the associated breed of the sire in the dataset. Each cluster corresponds to the breed of the sire such that cluster 1 corresponds to Angus sires, cluster 2 to Charolais sires and cluster 3 to Alberta hybrid sires.*

## 9.6. REFERENCES

Amer, P. R., Kemp, R. A. and Smith, C. 1992. Genetic differences among the predominant beef cattle breeds in Canada: An analysis of published results. Canadian Journal of Animal Science, 72(4): 759-771,

Edea, Z., Dadi, H., Kim, S., Dessie, T., Lee, T., Kim, H., Kim, J., Kim, K. 2013. Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. Front. Genet. doi: 10.3389.

Goonewardene, L. A., Z. Wang, M. A. Price, R.- C. Yang, R. T. Berg, and M. Makarechian. 2003. The effect of udder type and calving assistance on weaning traits of beef and dairy x Beef calves. Livest. Prod. Sci. 81: 47-56.

Glaubitz, J. C. 2004. CONVERT: a user friendly program to reformat diploid genotypic data for commonly used population genetic software packages. Molecular Ecology Notes 4: 309-310.

Li, M. H., Zerabruk, M., Vangen, O., Olsaker, I., and Kantanen, J. 2007. Reduced genetic structure of north Ethiopian cattle revealed by Y- chromosome analysis. Heredity 98, 214–221.

Lin, B. Z., Sasazaki, S., and Mannen, H. 2010. Genetic diversity and structure in Bos taurus and Bos indicus populations analyzed by SNP markers. Anim. Sci. J. 81, 281–289.

McKay, S. D., Schnabel, R. D., Murdoch, B. M., Matukumalli, L. K., Aerts, J., Coppieters, W. 2008. An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. BMC Genetics 9:37. doi:10.1186/1471-2156-9-37

Pariset, L., Cappuccio, I., Marsan, P. A., Dunner, S., Luikart, G., Obexer-Ruff, G., Peter, C., Marletta, D., Pilla, F., Valentini, A., 2006. Assessment of population structure by single nucleotide polymorphisms (SNPs) in goat breeds. Journal of Chromatography B, doi:10.1016/j.jchromb.2006.01.011

Pritchard, J. K., Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Statistics Canada – Cattle statistics, 2012. Catalogue no. 23-012-X. Vol 11, no. 1.

The HapMap Consortium, 2009. Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. Science. 324 (5926): 528-532.

Vignal, A., Milan, D., Sancristobal, M., and Eggen, A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. 34, 275–305.

## APPENDIX 1

## STATUS OF MANUSCRIPTS SUBMITTED FROM THE PhD PROJECT (AS OF 30 July 2013)

1. **Based on Chapter 3 of the thesis:**

Karisa, B. K, Moore, S. S, Plastow, G. S. Complicated relationships: A review of biological networks and pathways in animal science. **PUBLISHED in Springer Science Reviews, 2013, DOI: 10.1007/s40362-013-0005-8**

2. **Based on Chapter 4 of the thesis**

B. K. Karisa, J. Thomson, Z. Wang, P. Stothard, S. S. Moore, G. S. Plastow. Candidate genes and single nucleotide polymorphisms associated with variation in residual feed intake in beef cattle. **PUBLISHED in Journal of Animal Science; DOI:10.2527/jas.2012-6170.**

3. **Based on Chapter 5 of the thesis**

Karisa, B. K., Thomson, J., Wang, Z., Bruce, H. L., Plastow, G.S., Moore, S.S. Candidate genes and biological pathways associated with carcass quality traits in beef cattle. **Accepted for publication in Canadian Journal of Animal Science.**

4. **Based on Chapter 6 of the thesis**

B. K. Karisa, J. Thomson, Z. Wang, C. Li, S. P. Miller, S.S. Moore, G. S. Plastow. Metabolite bio-markers associated with feed efficiency in beef cattle with divergent residual feed intake. **Submitted to Livestock Science, 2013. The authors have received the reviewer's comments and are making corrections.**

5. **Based on Chapter 7 of the thesis**

Brian Karisa, Stephen Moore and Graham Plastow. Analysis of biological networks and biological pathways associated with residual feed intake in beef cattle. **Animal Science Journal, 2013. The authors received reviewers' comments and submitted a revised version. The chapter in the thesis reflects the corrections suggested by the reviewers.**