ABSTRACT

SISK, DANA MARIE STANLEY. Reproductive Gene Expression in Male Sus scrofa: An examination of the Differential Gene Expression of Divergent Testosterone selection and development of a Ribonucleic Acid extraction protocol from whole Porcine Spermatozoa. (Under the direction of Melissa S. Ashwell.)

The ability to characterize and enhance market traits in livestock has facilitated a greater interest in determining the genetic tool kit available for manipulation. In swine, using new approaches in genomics, such as microarray analysis and biological pathway analysis, we can show genes up and down regulated in a variety of processes and conditions. To this end, we identify the genes, pathways, and disease biomarkers affected by divergent selection of testosterone in boars. Testicular samples were taken from boars at 0, 30, 120, 150, and 180 days of age for lines of high (HT) and low testosterone (LT). Evidence that many of the differences in gene expression were at the pubertal period of 150 days led to a subsequent microarray study of the 150 day HT and LT animals. Microarray studies were followed by validation with real-time RT-PCR of 11 genes and extensive GeneGo pathway analysis (Metacore) of differentially expressed genes. While increased testosterone has long been associated with increased growth rates, we now have supporting genomic evidence of the genes and pathways up-regulated and down-regulated in these lines. To this end, this study has identified several disease biomarkers that may require further investigation and biological pathways associated with growth and metabolism that allow the recommendation of selective breeding for high testosterone to increase lean growth traits.

The genetic blueprint contained in the spermatozoan transcriptome can also illuminate key issues in swine reproduction. By developing a procedure for effective RNA
extraction of boar spermatozoa we are one step closer to elucidating the porcine sperm transcriptome and the genes implicated in growth and fertility. A viable protocol was developed to handle the complexities of large scale extraction of RNA from porcine semen utilizing an RNA carrier and Dnase treatment. This protocol was validated with PCR amplification of *Sus scrofa* *prm1* in order to provide evidence of a successful RNA extraction from sperm.
Reproductive Gene Expression in Male *Sus scrofa*: An examination of the Differential Gene Expression of Divergent Testosterone selection and development of a Ribonucleic Acid extraction protocol from whole Porcine Spermatozoa

by

Dana Marie Stanley Sisk

A thesis submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Master of Science

Animal Science

Raleigh, North Carolina

2009

APPROVED BY:

_____________________________  __________________ ____________
Dr. Melissa S. Ashwell    Dr. William L. Flowers
Committee Chair

_____________________________
Dr. Alison Anne Motsinger-Reif
BIOGRAPHY

Dana Marie Stanley Sisk was born on January 12, 1983 in Wilmington, NC to John Dane and Mary Duncan Stanley. She grew up in rural Shallotte, NC and in 2001 graduated valedictorian of her class at West Brunswick High School. She accepted the prestigious NCSU Park Scholarship and attended North Carolina State University where she graduated with a B.S. in Biological Sciences and a minor in Sociology in 2005. Following graduation Dana accepted a position at CIIT Centers for Health Research at the Hamner Institutes in Research Triangle Park, NC where she worked in animal necropsy and subsequently served as a genomics research assistant under the direction of Dr. Russell Thomas. After her two year employment at the institute, Dana returned to NCSU in 2007 to pursue a M.S. in Animal Science with a Genomics concentration and a minor in Biotechnology which she will earn in December 2009. In July 2009 Dana married David A. Sisk and resides in Raleigh, NC.
ACKNOWLEDGEMENTS

It is with great appreciation that I would like to thank my Masters thesis committee for their insight and guidance. Specifically I would like to offer my sincere thanks to Dr. Alison Motsinger-Reif for her incredibly insightful statistical advice and Dr. Billy Flowers for his great willingness to share his boar fertility and swine reproduction expertise and for providing the boars for the semen RNA protocol. I would also like to thank my M.S. advisor Dr. Melissa S. Ashwell for assistance with endless revisions that eventually led to this finished work. In addition, I would like to acknowledge Dr. J.P. Cassady and Dr. O.W. Robinson for the use of their divergently selected Duroc boars that were central to the differentially selected testosterone study. Finally, my gratitude goes to my husband and family for the support that allowed me to pursue my research endeavors.
# TABLE OF CONTENTS

LIST OF TABLES ................................................................................................................ vi
LIST OF FIGURES .............................................................................................................. vii
LITERATURE REVIEW ....................................................................................................... 1
Statement of Purpose ......................................................................................................... 1
Testis ................................................................................................................................... 2
Testis Components ............................................................................................................. 3
Testosterone and its effects ............................................................................................... 5
Testosterone and Behavior ............................................................................................... 8
Testosterone and Growth ................................................................................................. 9
Health effects of Testosterone ......................................................................................... 11
Testosterone studies in Livestock ................................................................................. 14
Testosterone and Market Quality ................................................................................. 16
Boar Taint ....................................................................................................................... 18
Initiation of Testosterone Study .................................................................................... 20
GnRH ............................................................................................................................. 21
GnRH Challenge ............................................................................................................. 24
Selection of Divergent Lines ......................................................................................... 25
Characteristics of Lines ................................................................................................. 27
QTL Studies and Fertility ............................................................................................... 29
Techniques Utilized in Testosterone Study ................................................................ 30
Pig Oligo Microarrays ..................................................................................................... 31
PowerAtlas ...................................................................................................................... 33
Real-time RT-PCR .......................................................................................................... 34
JMP Genomics, SAS, Cary, NC ..................................................................................... 38
Pathway Analysis ........................................................................................................... 39
RNA Isolation from Spermatozoa ................................................................................. 40
Ribonucleic Acid Activity in Spermatozoa ................................................................. 40
The Sperm transcriptome as an Indicator of Fertility ................................................ 43
Formation of Spermatozoa ............................................................................................ 44
Techniques used for Semen RNA ................................................................................ 45
Separation of Viable Sperm Utilizing a Discontinuous Percoll Density Gradient .... 45
Semen RNA Isolation techniques ............................................................................... 47
Ribogreen Quantification ............................................................................................. 52
Conclusion .................................................................................................................... 53
References Cited ............................................................................................................ 69

TESTICULAR DIFFERENTIAL GENE EXPRESSION IN MALE SUS SCROFA SELECTED FOR HIGH AND LOW TESTOSTERONE ................................................. 79
Abstract ....................................................................................................................... 79
Introduction ................................................................................................................... 81
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Materials and Methods</td>
<td>83</td>
</tr>
<tr>
<td>Selective Mating of Boars</td>
<td>83</td>
</tr>
<tr>
<td>Sample Collection</td>
<td>84</td>
</tr>
<tr>
<td>RNA Isolation and Microarray Hybridization</td>
<td>85</td>
</tr>
<tr>
<td>Microarray Analysis</td>
<td>87</td>
</tr>
<tr>
<td>Follow-up Microarray Analysis of Individual Animals</td>
<td>89</td>
</tr>
<tr>
<td>Real-time RT-PCR</td>
<td>90</td>
</tr>
<tr>
<td>Pathway Analysis</td>
<td>91</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>93</td>
</tr>
<tr>
<td>Summary</td>
<td>107</td>
</tr>
<tr>
<td>References Cited</td>
<td>125</td>
</tr>
<tr>
<td>ISOLATION AND EXTRACTION OF PORCINE SPERMATOZOA</td>
<td></td>
</tr>
<tr>
<td>RIBONUCLEIC ACID FOR GENETIC APPLICATIONS AND ANALYSIS</td>
<td>129</td>
</tr>
<tr>
<td>Abstract</td>
<td>129</td>
</tr>
<tr>
<td>Introduction</td>
<td>130</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>133</td>
</tr>
<tr>
<td>Semen Collection</td>
<td>133</td>
</tr>
<tr>
<td>Isolation of Viable Sperm from Whole Semen</td>
<td>134</td>
</tr>
<tr>
<td>Large Scale RNA Extraction Protocol</td>
<td>135</td>
</tr>
<tr>
<td>RNA Quantification</td>
<td>137</td>
</tr>
<tr>
<td>RNA Extraction Validation</td>
<td>138</td>
</tr>
<tr>
<td>Iscript Select cDNA Synthesis</td>
<td>138</td>
</tr>
<tr>
<td>RT-PCR Platinum Taq Protocol</td>
<td>139</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>140</td>
</tr>
<tr>
<td>References Cited</td>
<td>146</td>
</tr>
<tr>
<td>Project Summary and Future Directions</td>
<td>149</td>
</tr>
</tbody>
</table>
LIST OF TABLES

TESTICULAR DIFFERENTIAL GENE EXPRESSION IN MALE SUS SCROFA
SELECTED FOR HIGH AND LOW TESTOSTERONE

Table 1. REST Relative Expression analysis of 150 day animals with expression fold changes of the HT line compared to LT .......................................................... 111
Table 2. REST Relative Expression analysis of 180 day animals with expression fold changes of the HT line compared to LT temperatures ............................................. 112
Table 3. JMP analysis of Real-time RT-PCR results of genes selected for line by age comparisons of 150LT, 150HT, 180LT, and 180HT by expression differences along with primer sequences ................................................................. 113
Table 4. Networks identified with gene expression from 150 day animals ....................... 117
Table 5. Biological Pathways represented in gene expression in 150 day HT and LT animals along with associated p-values ........................................................................... 123
Table 6. Disease Biomarkers identified in 150 day HT and LT animals along with corresponding p-values ................................................................................................. 124
LIST OF FIGURES

LITERATURE REVIEW
Figure 1. Reproductive Tract of the Male Pig ................................................................. 54
Figure 2. Histological Section through testicular parenchyma of a boar ......................... 55
Figure 3. Sertoli Cell ........................................................................................................ 56
Figure 4. Leydig Cells ................................................................................................... 57
Figure 5. Structure of Testosterone ............................................................................. 58
Figure 6. Testosterone and GnRH negative feedback loop ........................................... 59
Figure 7. Swine Oligonucleotide Arrays as visualized in ScanAnalyze ....................... 60
Figure 8. Real-time RT-PCR gene expression ............................................................... 61
Figure 9. 150 day volcano plot ..................................................................................... 62
Figure 10. One way ANOVA of real-time RT-PCR results generated by JMP ............. 63
Figure 11a. TIMP1, ESE3, HMG2, MAD2, TRAM1 network identified by GeneGo from experiments ................................................................................................................. 64
Figure 11b. MetaCore™ GeneGo Network Objects Legend ........................................ 65
Figure 12. Spermatogenesis overview where type A stem cells lie on the basal lamina of the convoluted seminiferous tubules .............................................................. 66
Figure 13. Seminiferous Tubule ................................................................................... 67
Figure 14. Structure of Spermatozoa ........................................................................... 68

TESTICULAR DIFFERENTIAL GENE EXPRESSION IN MALE SUS SCROFA SELECTED FOR HIGH AND LOW TESTOSTERONE
Figure 1a. Microarray pilot study experimental design with pooled samples at all time points ................................................................. 108
Figure 1b. Microarray design with 150 and 180 day individual animals ....................... 109
Figure 2a-d. Volcano plots of JMP analysis of microarray line by age comparisons ...... 110
Figure 3. Top ten rated GO Processes in each line by p-value based on GeneGo analysis .................................................................................................................................. 114
Figure 4. Top ten rated biological pathways in each line by p-value based on GeneGo analysis .................................................................................................................................. 115
Figure 5. Top ten rated disease biomarkers in each line by p-value based on GeneGo analysis .................................................................................................................................. 116
Figure 6a. F8c, mis12, tfpt, clecsf2, tom22 network ..................................................... 118
Figure 6b. Timp1, ese3, hmg2, mad2a, tram1 network ................................................. 119
Figure 6c. Egf, cd44, timp1, p200rhogap network ....................................................... 120
Figure 6d. Ndn, tfpt, hmg2, appbp2, cks1 network ..................................................... 121
Figure 6e. MetaCore™ GeneGo Network Object key ................................................ 122

ISOLATION AND EXTRACTION OF PORCINE SPERMATOZOA
RIBONUCLEIC ACID FOR GENETIC APPLICATIONS AND ANALYSIS
Figure 1. Semen Collection of Boars by ID ................................................................. 143
Figure 2. RNA yield (ng) as measured by Ribogreen assay per 9x10^8 cells .......... 144
Figure 3. RT-PCR with 100kb ladder and visible cDNA and genomic bands at 234bp and 334bp
LITERATURE REVIEW

Statement of Purpose

In the swine industry, the field of genetics is rapidly adjusting to produce pigs that accommodate the demands of the human population. In the state of North Carolina, this is particularly true where the swine industry is vital to the state’s economy and provides nourishment to people throughout the country. The purpose of this study is to investigate the means of increasing fertility and growth rates in pigs by examination and application of improvements in male genetics. Specifically, this study seeks to investigate genomically the benefits of testosterone in swine given the current research of increased growth rates, meat quality, and litter sizes. Through microarray analysis, quantitative RT-PCR validation, as well as network pathway analysis boars divergently selected for testosterone production will be examined to identify the differential gene expression and its biological consequences.

Additionally, this project develops a new protocol for the extraction of RNA from boar spermatozoa that can be used for additional gene studies. Due to the small quantities of RNA available in sperm, this procedure takes particular care to provide a large scale extraction that can successfully amplify gene transcripts and provide a foundation for possible fertility profiling and selection to further swine genetics.
Testis

It is difficult to discuss the effect of paternal genetics in the adult animal without first examining the origination of that input. The testis is the male organ responsible for the production of the male genetic contribution: sperm, and the fundamental control of many parts of the body through sex hormones (Figure 1). The release of sex hormones is controlled by the anterior pituitary and accomplished through the release of luteinizing hormone and follicle-stimulating hormone. Luteinizing hormone stimulates the Leydig cells to produce testosterone, while follicle-stimulating hormone prompts Sertoli cells to produce androgen-binding protein that binds and concentrates testosterone for spermatogenesis (Franca 2000).
Testis Components

The testis is dominated by two main cell types, Sertoli and Leydig cells (Figure 2). Sertoli cells are known as the “mother cell” and serve as secretory and structural support for developing sperm. They are column-shaped cells activated by follicle-stimulating hormone (FSH) and control the entry and exit of nutrients and hormones into the testis. Sertoli cells also control the differentiation of spermatogonia to mature spermatozoa as well as the phagocytosis of extra cytoplasm of sperm during the maturation phase. Finally, Sertoli cells secrete androgen-binding protein which binds to testosterone making it less lipophilic and concentrating it in the seminiferous tubule near the developing gametes to increase fertility (Franca 2000).

Sertoli cell differentiation ceases before puberty and interestingly the number of Sertoli cells during the prepubertal period determines testicular size and number of sperm produced by the mature boar (Lunstra 2002). In addition, periods of Sertoli cell proliferation correlate with increased levels of FSH and elongation of the seminiferous tubules. Sertoli cells also endure two phases of development. The first phase begins at birth (Figure 3) and continues to an age of approximately thirty days where the number of Sertoli cells per testis increases six fold. Next, when the animal is between three to four months of age, the number of Sertoli cells doubles and the cells begin to mature. After four months of age, the cell number begins to decrease, with stabilization in number occurring between seven and sixteen months (Franca 2000).
Leydig cells create the main component of the intertubular space in pig testis (Figure 4). Leydig cells are stimulated by leuteinizing hormone (LH) to release androgens such as testosterone, androstenedione, and dehydroepiandrosterone (DHEA). FSH increases the number of LH receptors on Leydig cells, thus increasing the response to LH. In contrast to most mammals, boars have three phases of Leydig cell development. Leydig cells develop during the early fetal period as they increase 100%, the perinatal period when they have an increase to 160%, and from three months to sixteen months when they are 80% of original volume. From birth to two months of age Leydig cells occupy thirty to forty percent of the parenchyma. After two months, the Leydig cells are only 7% of the total parenchyma volume and then from six to sixteen months their volume increases to ten to fifteen percent (Franca 2000).
Testosterone and its effects

Testosterone (Figure 5), an androgen of interest in many species due to its association with increased growth rates, can be traced to the testis. The hypothalamus secretes gonadotropin-releasing hormone (GnRH), which in turn stimulates the pituitary to release follicle stimulating hormone (FSH) and luteinizing hormone (LH) (Schally 1971). FSH then stimulates Sertoli cells which are responsible for spermatogenesis and LH stimulates Leydig cells to produce testosterone (Schally 1971). In the normal human male, the Leydig cells produce approximately 7mg of testosterone daily (Luke 1994). Testosterone circulating in the body is found mostly in a bound inactive form, bound to sex hormone-binding globulin and is not available to tissues (Manni 1985). Only one to two percent of testosterone is unbound and active (Dunn 1981). Clinical measurements can be taken as total serum testosterone which includes bound and free circulating testosterone, or free or unbound testosterone (Bagatell 1996).

While testosterone can act directly on cells, it can also be converted to dihydroxytestosterone (DHT) by 5α-reductase or to estradiol by aromatase. While testosterone and DHT can both bind to androgen receptors, DHT has a higher affinity for the receptor and is more potent (Luke 1994). The distribution of these androgen forms is tissue specific. Aromatase is most frequently found in the liver, adipose tissue, and central nervous system nuclei (Mooradian 1987). 5α-reductase is generally found in skin, prostate, and reproductive tissues and there are two isoforms of this enzyme which differ by tissue
specificity and susceptibility to inhibition. Activity is highest in the primordium of the prostate and external genitalia prior to virilization but is very low in Wolffian duct structures. Mutation of 5α-reductase results in an inability to correctly produce the male phenotype during embryogenesis caused by a lack of conversion of testosterone to DHT (Russell 1994).

Testosterone has long been implicated in many bodily processes within the male. It stimulates prenatal differentiation, development of male tissues such as the penis, testes, epididymis, seminal vesicles, and the prostate, as well as the maintenance of these tissues in the adult (Mooradian 1987). Testosterone is responsible for the initiation and maintenance of spermatogenesis and studies have been undertaken to ascertain its effects in this area. It appears that testosterone maintains spermatogenesis in a concentration dependent manner. A study in intact male rats examined the concentration of testosterone within the seminiferous tubules that is required for normal sperm production (Zirkin 1989). Testosterone release was experimentally clamped in the testes and the resulting levels of intratesticular testosterone were measured and related to the numbers of advanced spermatogenic cells maintained. At testosterone levels at 13ng/ml in the seminiferous tubule, despite this being an 80% decrease, spermatogenesis occurred at full capacity. However a further decrease to 9ng/ml, an approximate 85% reduction, reduced the number of spermatids maintained by the testes to approximately one hundred million with reduction to 4ng/ml (95% reduction) dropping the number of spermatids by an additional one hundred million. From this data it appeared that maintenance of spermatogenesis in the testes is
dependent on the amount of testosterone present, however, the testes produce far more testosterone than is required to maintain the maximum level of spermatogenesis (Zirkin 1989).
Testosterone and Behavior

Testosterone has been credited with effects on behavior. In humans, androgens stimulate and maintain sexual behavior (Davidson 1982). In cases of hypogonadism, testosterone replacement is used to induce sexual interest and improve sexual performance (Davidson 1979). It is unclear whether testosterone supplementation is useful in increasing the frequency or quality of erections in men (Davidson 1982). One study showed that suppressing the levels of testosterone to a level near that of castration reduced sexual fantasies, sexual desire, and number of erections. It also reported that there is a threshold in humans varying from person to person at which sexual activities are impaired (Bagatell 1994).

Aggressive behavior has been linked to increased testosterone levels. This has been most notably studied in nonhuman primates. It has been shown that aggression in some primate species has been directly correlated with serum testosterone levels (Michael 1978). No such link has been shown in humans based on self-reported aggression surveys when men were asked to rank the levels of aggression they felt and the answers were compared to actual testosterone levels (Bagatell 1994). With administration of exogenous testosterone, the form seems to dictate effect. Short-term administration of methyltestosterone has been shown to increase violent feelings, hostility, and irritability when given to normal men (Su 1993), while long-term administration of testosterone enanthate did not seem to affect aggression or irritability (Bagatell 1994).
Testosterone and Growth

Abuse of androgen administration has become a social issue, especially among athletes, because they have been shown to increase nitrogen retention and cause an increase in lean body mass and weight. The increase in muscle mass is caused by testosterone binding to the androgen receptor to increase the size of muscle cells but there does not appear to be an increase in the number of muscle cells (Mooradian 1987). Androgens have both anabolic and androgenic effects. The androgenic characteristics produce the male phenotype and secondary sex characteristics (Wilson 1980).

Testosterone affects many parts of the body. Dihydroxytestosterone is considered to be the active hormone in sebaceous glands and is responsible for sebum production. Hair growth in certain areas is dependent on testosterone levels. In humans for instance, the body will grow axillary hair and lower pubic hair with only small levels of testosterone, while hair growth on the chest, face, and upper pubic area require higher concentrations for growth. Temporal balding can be consequence of hair follicles converting testosterone to dihydroxytestosterone (Mooradian 1987).

Androgens have been shown to stimulate bone growth in vitro (Kasperk 1989) but it is not known what concentration is required to stimulate or maintain this growth. It is also not known if it is the testosterone itself that is responsible for the effects on bone or the estrogen that is derived from it (Bagatell 1996). Hypogonadism has been shown to be a risk factor for osteoporosis in men (Finkelstein 1990) and bone density increases during
puberty, reaches its peak during the mid twenties, and declines linearly with age (Gilsanz 1988).
Health effects of Testosterone

Administration of androgens, especially those that cannot be aromatized to estrogens, have been shown to decrease HDL (high-density lipoprotein) cholesterol levels (Bagatell 1994). In men treated with androgens for hypogonadism (Sorva 1988) or boys with delayed puberty, decreased levels of HDL were observed (Kirkland 1987). Although the cause is not clear, men generally have lower levels of HDL cholesterol and higher levels of LDL (low-density lipoprotein) than women (Heiss 1980).

In the liver androgens cause the production of salic acid and $\alpha_1$-antitrypsin, haptoglobin, hepatic triglyceride lipase, and sialic acid (Dickinson 1969). They cause a decrease in some binding proteins including sex-hormone binding globulin, as well as fibrinogen, and transferrin (Bagatell 1996). In the kidneys, testosterone stimulates the production of erythropoietin (Gardner 1968) which increases hemoglobin (Bagatell 1993). Declines in hemoglobin levels have been detected when androgens are inhibited and testosterone has been used in treatment of anemia associated with chronic renal failure, Fanconi’s anemia, and aplastic anemia (Ammus 1989).

In males, testosterone deficiency results in hypogonadism which could be manifested in two ways. Hypogonadism presents as both primary and secondary hypogonadism. Primary hypogonadism is presented with low serum testosterone levels and high serum luteinizing hormone and follicle-stimulating hormone. These patients are unable to synthesize androgens, and although androgen dependent processes can be maintained by
exogenous hormone replacement, these men remain infertile. Secondary hypogonadism is characterized by low serum testosterone and luteinizing hormone levels and low or subnormal levels of follicle-stimulating hormone. These patients are potentially fertile and exogenous hormone replacement can stimulate this and other androgen-dependent processes. In some cases, spermatogenesis and endogenous androgen production can be stimulated with exogenous testosterone replacement therapy (Bagatell 1996).

Side effects of androgen replacement therapy are dependent on the dose level. At replacement levels of testosterone, side effects include an increase in lean body mass and fluid retention as well as increased weight gain. Sometimes mild gynecomastia may occur as a result of the aromatization of testosterone to estradiol, especially in adolescents. Other effects include acne, sleep apnea, lower HDL levels, and erythrocytosis (Bagatell 1996).

Supraphysiological doses come with additional side effects including decreased testicular size, azoospermia (Wilson 1988), increased or decreased libido, increased aggression, and various psychotic symptoms (Uzych 1992). Adverse vascular effects include an increased risk of cardiovascular disease, likely associated with the decrease in HDL cholesterol levels (Jacobs 1990). Platelet counts increase and platelet aggregation is more likely (Ferenchick 1992). Risks are increased for strokes (Frankle 1988) and myocardial infarctions (McNutt 1988).

It has been suggested that androgens may also play a role in support of cardiovascular health. Postmenopausal females show an increased risk for coronary heart
disease, yet they still have a lower incidence of coronary heart disease than males in the same age group. Testosterone is converted to 17β-estradiol by aromatase and the effects of testosterone could influence both sexes. Although it has been proposed that testosterone can cause an increased risk of coronary artery disease, there has been no direct evidence. In fact, there has been shown an inverse correlation between plasma testosterone levels and the extent of coronary artery disease in men. In addition, studies show that physiological doses of testosterone may be useful in treatment of coronary artery disease. In a study of males with coronary artery disease, 3 minute intracoronary infusions of testosterone increased coronary artery diameter as well as coronary blood flow compared to controls (Webb 1999). A canine study had similar results, demonstrating that small doses of testosterone could theoretically be used as a treatment for coronary artery disease (Chou 1996).
Testosterone studies in Livestock

The effects of testosterone are not limited to humans, as testosterone injections have also been shown to increase gains in body weight in rats (Rubenstein and Soloman 1940), and increase growth rate of rabbits (Bergetrand 1950). Studies with immature guinea pigs showed that administration of testosterone propionate caused hypertrophy of the temporal muscle as well as other muscular enlargements (Papanicolaou 1938). It is this type of muscular increase which led to investigation of testosterone to improve livestock. Subcutaneous implantation of testosterone was found to elicit changes in lambs. Lambs administered 10mg or 20mg of testosterone subcutaneously for 68 days were shown to have qualitatively improved market quality, based on visual inspection and observation of a leaner body by experienced sheep buyers. Quantitatively, both the 20mg testosterone treatment group (P<0.05) and the 10mg testosterone treatment group (P<0.01) showed improved feed efficiency by requiring significantly less feed per pound of gain than control animals (Andrews 1949).

The study of testosterone in cattle has been centered on improvements for human consumption. In the cattle industry, testosterone has long been implicated in growth effects. Testosterone has been shown to cause an increased rate of gain and feed efficiency in steer and heifer calves (O'Mary 1952). A study in beef cattle examined daily gain, feed efficiency as well as carcass characteristics in animals receiving weekly intramuscular injections of testosterone at the rate of 1mg per kg of body weight. The experiment was
performed on three heifers and three steers beginning at a body weight of five hundred pounds and ending at a body weight of eight hundred pounds. Six heifers and six steers were maintained as control animals. Heifers treated with testosterone gained 0.53 pounds more per day than the control heifers, and testosterone treated steers gained 0.29 pounds more per day than control steers. Testosterone treated animals also required less feed per unit of gain per controls, with treated heifers requiring only 3.93 pounds of TDN (total digestible nutrients) per pound gained compared to control heifers that required 5.01 pounds of TDN. Similarly, the testosterone treated steers required 3.65 TDN per pound of gain, while the control steers required 4.30 pounds of TDN. Because the study focused on beef cattle, the meat quality and carcass effects were also examined. Testosterone treatment increased round weight by 2.8% for both steers and heifers and treated heifers also had 2.3% more chuck than control heifers. Untrimmed loin was found to be reduced by 2.77% in treated cattle, meaning a higher percentage of meat went to other choice cuts (Burris 1953).
The effects of testosterone in swine are usually documented with regard to affected market characteristics. These traits are usually body weight, rate of growth, feed efficiency, and amount of back fat, however testosterone can affect other characteristics as well. Most studies have examined the effects of testosterone through castration studies because it was found that intact boars have a higher percent of lean in the rough loin, less back fat, longer body length, higher live weight and higher carcass preferred cut yield weight (Bratzler 1954). It was also found that accessory sex glands and kidney weights are higher on noncastrates than castrates. However, boars produce poor quality meat due to boar taint (Bratzler 1954). Hunt found that meat quality was affected by castration, as rams had a higher percentage of lean rib meat than wethers but he found no significant difference in weight gain between the two (Hunt 1938). Bratzler implanted castrated boars with 193mg testosterone pellets and found no significant difference between rate of gain per pound of feed, or in sex accessory gland and kidney weights between normal castrated pigs and pigs implanted with testosterone (Bratzler 1954).

Intramuscular injection of testosterone in Duroc and Hampshire barrows and gilts showed interesting effects. Although semi-weekly injection of 0.5mg per kg of body weight of testosterone propionate showed no significant effects on rate of gain, carcass length, carcass grade, or thickness of backfat, effects were seen mostly in behavior, as the animals receiving injections mounted other and exhibited boar-like behavior. In addition, it was
reported that they appeared more trim than their control counterparts. Internal effects included enlarged and cystic ovaries with hyperemic and hyperplastic genital tracts from the gilts as well as enlarged and hardened prostate glands and seminal vesicles from the barrows (Sleeth 1953).
Boar Taint

It is estimated that the boar industry could improve profits by 30% with the use of intact boars because of the increase in lean growth (De Lange 1995). Testosterone levels are greatly decreased by castration, but raising intact males in the pork industry is discouraged because of boar taint, an unpleasant odor and taste in the meat of intact males (Quintanilla 2003). Boar taint is caused by high levels of androstenone and skatole. Androstenone is a steroid produced in the testis when an animal reaches sexual maturity and accumulates in the adipose tissue. Skatole is produced in the hindgut of the pig and is then released into the blood stream. Most of the skatole is metabolized by the liver and then secreted, however some is stored in the fat. Intact males are said to have increased metabolism and clearance by the liver which may contribute to the clearance of skatole. Skatole and androstenone accumulate in the fat and create the odor of boar taint when the meat is heated (Squires 1999).

There have been few advances in the elimination of androstenone, except with methods that also reduce the growth rate of the pig such as somatotropin (Bidanel 1991). Skatole can be reduced by manipulation of the microflora of the gut (Jensen 1995) and can be controlled by diet and environmental factors. Maintaining a clean environment for the pigs by removing manure and using slatted floors keeps the animals from absorbing skatole from the environment. Some antibiotics can be used to reduce lactobacillus, the bacteria that produce skatole. Diets high in fermentable carbohydrate or bicarbonate can reduce
skatole levels. Also, management decisions such as slaughter at younger ages, or increased water consumption and restricting feed forty eight hours before and withholding food twelve hours before slaughter, reduces skatole substrates. The half life of skatole is only ten hours so treatments are only necessary the week before slaughter (Squires 1999). Instead of implementing these management and treatment practices, the United States boar industry uses castration to combat boar taint.

Current research in this area is focused on identifying candidate genes as markers for the identification of low boar taint animals. Two variants of cytochrome b5 have been identified in the testis, one of high molecular weight, and one of low molecular weight. The low molecular weight variant was linked to a higher incidence of androstenone, and therefore boar taint, and could possibly be used as a marker to begin reduction of boar taint (Squires 1999). Skatole clearance has been associated with the enzymes CYP2E1 and CYP2A6. It has been found that boars with low levels of these enzymes have higher concentrations of skatole due to processing of the compound in the liver (Squires 1999). It would be valuable to identify genes involved in boar taint and determine if there is an effect of testosterone on these genes.
Initiation of Testosterone Study

In order to determine the heritability of testosterone traits, a North Carolina State University study of Duroc boars was initiated for the divergent selection of lines of high and low testosterone between 1982 and 1992. In the initial study, boars were weaned at 42 days and then creep feed was made available after 21 days. They were housed five or six per pen in confinement from other pigs (Lubritz 1991). Each year 25 female and 5 male Durocs were mated to each other to produce the next generation. From each litter one boar was chosen at random and subjected to GnRH challenge. The five boars with the highest and lowest post-challenge testosterone levels were selected to sire the subsequent generation (Robison 1994).
GnRH

Gonadotropin releasing hormone (GnRH) is a neurohormone synthesized in the preoptic area of the hypothalamus that is released by the GnRH-secreting neurons into the hypophysial portal bloodstream at the median eminence where it travels through portal blood to the pituitary gland. The pituitary is stimulated to release FSH and LH based on the GnRH pulses. High frequency pulses release LH and low frequency pulses release FSH. Release of prepubertal GnRH is very low. GnRH pulses in mature males are released consistently, but pulses in females are consistent with cyclic changes in the estrous cycle. GnRH is responsible for maintaining ovulation, folliculogenesis, and the maintenance of the corpus luteum in females. In males, GnRH controls spermatogenesis. Males have pulsatile secretion of GnRH and testosterone has a negative feedback effect on its release (Figure 6). Castration or the introduction of LH or FSH causes an increase in the pulsatile secretions (Brown 1994). Testosterone’s negative feedback effect most likely takes place in neural sites in the medial preoptic area and the anterior hypothalamus, as well as the median eminence and the arcuate nucleus in the medial basal hypothalamus (Sar 1973). Based on studies with ovariectomized rats, in females estrogen causes increases and decreases of GnRH, LH, and FSH depending on treatment and dosage conditions. Baseline GnRH secretion is controlled by neurons in the medial basal hypothalamus. In primates, this area facilitates the pre-ovulatory GnRH surge as well. In contrast, in the rat, preoptic neurons provoke this surge. Estrogen injections into these sites cause the estrogen positive feedback
and the release of LH. Injections of estrogen into the rat medial basal hypothalamus inhibit LH. Likewise, in primates the anterior pituitary is also affected. Introduction of estrogen into the anterior pituitary of female primates causes an increase in the number of GnRH receptors and a release of LH. In contrast, in the male primate introduction of testosterone in the anterior pituitary inhibits GnRH and LH release (Brown 1994).

The most acknowledged role of GnRH in vertebrates is stimulation and release of the gonadotropins LH and FSH from the pituitary gland. It has now been shown that there are approximately two to three different forms of GnRH in each vertebrate species and they exist in multiple tissues. Currently there are fourteen known variants of GnRH and ten of those variants have been found in mammals. Variants have not only been found in the pituitary, but have also been uncovered in the brain, retina, and gonads and have been highly associated with reproduction (Lethimonier 2004). Kakar et. al used Reverse Transcription Polymerase Chain Reaction (RT-PCR), Southern blot analysis, and DNA sequencing to identify the genes coding GnRH hormone and its receptor in non-reproductive organs such as human liver, heart, skeletal muscle, kidney, placenta, and pituitary (Kakar 1995). There are three different types of GnRH receptors which are known as GnRH-R1 through GNRH-R3 and are expressed spatio-temporally (Lethimonier 2004). Signal transduction occurs when GnRH binds to GnRH receptors (GnRH-R) and, through activation of 7 transmembrane G-proteins and phospholipases, there is an increase in cystolic diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP3) that releases calcium. It has also been found that gonadotropin release was decreased in cells depleted in intracellular calcium
(Klausen 2002). GnRH variants and their functions are still being explored. Fish are a popular model to study the variants, their receptors, and their roles, especially since teleost fish express 8 of the 10 variants (Lethimonier 2004).
GnRH Challenge

A GnRH challenge is a technique used to measure sensitivity to GnRH. Blood samples are taken before and after an injection of GnRH or an analog. Often this is used to measure gonadotropins but can be used to ascertain other responses affected by GnRH, like cortisol or testosterone (Wise 2000). Usually samples are taken at a fixed set of intervals before and after the introduction of GnRH. Pre-injection levels are used to define basal levels and are a good indicator of actual performance. Post-injection levels show what is released or produced in response to GnRH and shows an animal’s potential for response if sufficiently stimulated (Lubritz 1991). In a study using Meishan and white composite boars, both lines showed an increase in LH, FSH, and testosterone after GnRH stimulation with a dose of 1500ng/kg of body weight. When levels were decreased to 15ng/kg, only LH and testosterone were affected (Wise 1996). In the NCSU lines, the pre- and post-injection levels were positively correlated (Walker 2004).
Selection of Divergent Lines

In the North Carolina State University study, one boar was randomly selected from each litter and subjected to the GnRH challenge consisting of five pre-GnRH blood draws every 30 minutes prior to GnRH injection. For the next two hours, blood was drawn every 15 minutes to constitute 12 blood draws. Then blood was taken every 30 minutes for an additional two hours (Robison 1994). Validated through a previous study, the technique of RadioImmunoAssay was utilized to determine the levels of testosterone in duplicate in the samples. Anhydrous ethyl ether was used to extract testosterone from the blood which was subsequently measured using standard curves of recrystallized 17β-hydroxy-4-androsten-3-one (Juniewicz 1984). Samples were 10 ml each (Lubritz 1991). The five boars with the highest and lowest post-GnRH testosterone levels were selected to sire the next generation (Robison 1994).

Data was maintained on testes length and width on these boars, as well as estimated testes volume at 168 days. Backfat depth was measured by ultrasound and adjusted to 104kg of body weight. Growth data were also maintained as days to 104kg.

After selection of the tenth generation, testosterone levels as indicated by pre- and post-challenge values, were approximately three times higher in the high testosterone line than in the low testosterone line (Robison 1994). From this point, selection continued through random mating within each line. At generation 21, the testosterone levels were
once again evaluated to validate that the lines remained divergent. Testosterone levels averaged 49.0ng/mL in the high line and 27.8ng/mL in the low line (P<0.01) (Walker 2004).
Characteristics of Lines

In previous studies completed on these boars, certain line characteristics were observed. In the 1991 study, Lubritz found that testosterone levels were related to testes size, where boars of the high testosterone line tended to have larger testes. Also, the data suggested that increased levels of testosterone contributed to better growth rates (Lubritz 1991).

Robison and coworkers also calculated heritability estimates showing that selection differentials were generally lower for the low testosterone line. This was attributed to the biological limits reached by the later generations of the low line where boars with very low testosterone began to experience delays in puberty and poor libido. This caused a difficulty in selecting for the lower testosterone extremes in the low testosterone line. The study also showed that the high testosterone line had larger testes ($P<0.05$) and required fewer days to reach 104kg ($P<0.05$). However, it was the first study to show that the female offspring of the boars from the high line had larger litter sizes measured in number born alive ($P<0.05$) (Robison 1994).

In 2004, a follow-up study looked at these lines in relation to sperm production and testicular morphology. Using histological techniques as well as modeling, morphological characteristics were examined at generation 20 and sperm production characteristics were measured at generation 21. It was found that total paired testes weight was higher in the low testosterone line ($P<0.01$) but that epididymal weight was higher in the high testosterone
boars ($P<0.01$). Leydig cell volume density ($P<0.01$) and estimated mass ($P<0.01$) were higher in the high testosterone line but seminiferous tubule volume density ($P<0.07$) and estimated structure mass ($P<0.10$) tended to be higher for the low testosterone line. Also, there were no significant differences in the volume density ($P \leq 0.27$) or estimated mass ($P \leq 0.26$) of the Sertoli cells between lines. Blood serum tests on generation 21 also showed that there was not a significant difference in FSH plasma concentrations ($P \leq 0.30$). Measured by daily sperm production per gram of testis ($P<0.55$), total daily sperm production ($P<0.34$), sperm per gram of testis ($P<0.54$), and total testicular sperm ($P \leq 0.34$), morphological examination of generation 20 boars showed that sperm characteristics were not significantly different between lines (Walker 2004).
QTL Studies and Fertility

Another technique often used to observe traits in livestock is quantitative trait loci (QTL) mapping. This technique involves assigning a genetic function to a physical location along the chromosome which is often done by the mapping of single nucleotide polymorphisms (SNPs). The effect on a certain trait can be calculated for each QTL or SNP to determine if they may be used as markers for the trait. In this way, QTL studies are able to identify the actual loci associated with a particular phenotype. Through this method QTLs have been identified for many characteristics including meat quality and fertility. A recent study looked at several polymorphisms suspected of affecting reproduction and found that some affect litter size. Based on its proximity to an identified QTL affecting litter size Neuropilin (nrp2) was found to be a candidate gene with the possibility of increasing litters (Du 2009). In another study, several QTLs were mapped that had significant effects on percentage of lean meat and carcass length in white landrace/wild boar crosses (Andersson-Eklund 1998). While this study did not identify specific genes associated with these QTLs, it is often a useful first step to identify the area of the chromosome affected and these types of studies have laid the groundwork for additional genomic analysis.
Techniques utilized in Testosterone study

To further the study of these divergently selected high and low testosterone boars we will examine them genomically. Utilizing several techniques to be discussed further, this study will identify the gene expression associated with the observed characteristics of these lines and determine its biological significance.
Pig Oligo Microarrays

Several techniques facilitated the genomic examination of these lines. The Swine Protein-Annotated Oligonucleotide Microarray or PigOligoArray (Figure 7) was crucial in evaluation of differences in gene expression within and across lines. They were developed by individual investigators and institutions in an open source collaboration to benefit swine research. Those with an interest in swine physiology can reasonably purchase the arrays and download the annotations, the sequences of the oligonucleotides on the arrays, and the consensus sequences those oligos represent at no cost (Consortium 2009).

The collaboration is currently on its second generation of 70-mer oligonucleotide spotted arrays. These arrays represent 20,400 oligos that were designed with pig expressed sequence tags (ESTs) matched to other phylogenetically related vertebrate proteins. To maximize probe efficiency, distinguish between orthologs and paralogs, and to exclude chimeric ESTs, probes were aligned to protein clusters using Megablast. Porcine and bovine sequences were also aligned. RefSeq was used to confirm the correct orientation of the probes. Controls were included on the arrays to indicate hybridization stringency. There are six mismatch probes with either 1, 2, 3, 5, 7, or 10 mismatches for 60 contigs. They also contain 60 negative controls which are scrambled sequences without any homology to the swine or bovine genome. Two hundred and fourteen cross species positive controls were included with one hundred percent homology to bovine sequences. Finally, the probes were
annotated with Gene Ontology to include descriptions of homologous proteins (Consortium 2009).
Often with microarray studies it becomes difficult to determine sample sizes to accommodate such large numbers of multiple comparisons. Such a problem arose with this study, so PowerAtlas was employed. PowerAtlas is a free internet based program to allow scientists to estimate required sample sizes based on previous studies and was used in this study to evaluate sample sizes for sufficient statistical power for the microarray experiments. Good statistical power is needed to detect differences in experiments. Often each gene may have its own effect size and variance so that it is difficult to address microarray experiments with multiple hypotheses. The user may input pilot data or estimate statistical power by building on previous studies from the Gene Expression Omnibus. Based on a desired level of power, investigators can estimate required sample sizes to most efficiently use microarrays in their research (Page 2006).
Real-Time RT-PCR

Quantitative reverse transcription Polymerase Chain Reaction (RT-PCR) is a process that is used for quantification of mRNA and is often used to validate microarray results. In RT-PCR, RNA must be reverse transcribed into cDNA which will serve as the template for the PCR assay. During the PCR, the DNA is exponentially amplified. These processes required RNA and DNA-dependent DNA polymerases. They can function simultaneously in two enzymes/one tube reactions, or the reactions can be performed separately in two enzymes/two-tube reactions. Some enzymes may function in both roles and be used for one enzyme/one-tube reactions (Bustin 2000).

The reverse transcriptase (RT) used in the reaction can be a very important consideration when undertaking PCR. Different reverse transcriptases can cause variations in the quantification of mRNA. Two of the most common RTs are avian myeloblastosis virus reverse transcriptase (AMV-RT) and Moloney murine leukaemia virus reverse-transcriptase (MMLV-RT). While both RTs have their merits, AMV-RT is more robust (Brooks 1995) and can be stable at temperatures up to 55°C (Freeman 1996). It is also better at overcoming problems with template secondary structures (Bustin 2000). In comparison, MMLV-RT is valuable because it has less RNase-H activity (Kotewicz 1988), which degrades DNA:RNA hybrids, so it is better when amplifying longer transcripts like full length cDNA molecules (Zhang 1999).
To amplify transcripts, random hexamer primers, oligo-dT primers, or mRNA-specific primers can be used. While random hexamer and oligo-dT primers maximize the number of transcripts amplified, they also can overestimate copy numbers by up to nineteen fold (Zhang 1999). The use of mRNA specific primers decreases background in the form of non-specific amplification (Zhang 1999).

Real-time RT-PCR or Quantitative RT-PCR (qRT-PCR) refers to the ability to visualize and quantify the DNA amplification in RT-PCR in real time. While there are currently four chemistries available for real-time PCR, the North Carolina State University Domestic Animal and Genomics Lab utilizes SYBR Green fluorescent dye assays. SYBR dye is designed to bind to doubled stranded DNA. During the process, the template is denatured and the SYBR dye exhibits little fluorescence. During the primer and probe annealing stage, the SYBR starts to bind to the double stranded primer/target and gives a light emission when excited by the appropriate light source. During polymerization, additional SYBR molecules bind to the new double stranded DNA, which causes an increase in fluorescence that can be seen in real time. During denaturation, the SYBR dissociates from the DNA and the fluorescence returns to background levels. The amount of fluorescence is proportional to the amount of double stranded DNA present in the sample during each cycle (Bustin 2000).

Quantification is achieved through examination of a sample’s Ct, the point at which sample fluorescence crosses a common threshold (Figure 8). Values of fluorescence are recorded during every amplification cycle to be representative of the amount of product
amplified. A greater amount of template in the reaction will result in shorter time period until the fluorescent signal is above background (Gibson 1996). This means that greater template will result in smaller Cts because it will take less cycles to reach the threshold. The Ct is measured during the amplification phase and not in the plateau phase so that it is not reagent limited (Bustin 2000).

Traditionally housekeeping genes are used as a means of normalization between different tissues and genes. Two popular housekeeping genes are Beta-Actin and Glyceraldehyde 3-phosphate dehydrogenase (GAPDH). Housekeeping genes have been utilized for normalization procedures because they are genes that are usually involved with the basic day to day functions of the cell and their expression levels are expected to remain constant despite treatment (Bustin 2000). Beta-Actin encodes a ubiquitous cytoskeleton protein and was one of the first genes to be used as a housekeeper. It has been shown to be relatively resistant to changes caused by diseases (Kreuzer 1999). Although it is widely used, its expression has been shown to vary in some experiments in tissues such as human breast epithelial cells, blastomeres, porcine alveolar macrophages (Foss 1998) and canine myocardium (Carlyle 1996). It also shares more than ninety percent homology with its pseudogene (Kreuzer 1999). GAPDH also shares a high level of homology with pseudogenes. Its levels also vary within individual (Bustin 1999), during pregnancy (Cale 1997), during various parts of the cell cycle (Mansur 1993), and with development (Puissant 1994). Bustin and colleagues showed that there was even variation between transcription of GAPDH in samples taken from the same individual over different time points (Bustin 1999).
While there are many ways to calculate the quantification of real-time RT-PCR, the most popular method used with SYBR detection is the Pfaffl method (Pfaffl 2001). The Pfaffl method relies on calculating a relative expression level of a target gene versus a reference gene. This expression level is based on the real-time PCR efficiencies of each gene and the threshold crossing point (CP or CT) deviation of the unknown sample versus a control. Theoretically, product should double every cycle and if this is the case, the efficiency would be 100%. However, this is not always the case in practice and the slope of the log of the fluorescence intensity of the signal over the cycle number should be equal to the log of the efficiency. The Pfaffl relative expression ratio is $\text{Ratio} = \frac{(E_{\text{target}})^{\Delta CP_{\text{target}} (\text{control-sample})}}{(E_{\text{ref}})^{\Delta CP_{\text{ref}} (\text{control-sample})}}$, where $E_{\text{target}}$ is the real-time PCR efficiency of the target gene, $E_{\text{ref}}$ is the real-time efficiency of the reference gene, $\Delta CP_{\text{target}}$ is the Cp (Ct) deviation of the control minus sample of the target gene, and $\Delta CP_{\text{ref}}$ is the CP (Ct) deviation of the control minus sample of the reference gene (Pfaffl 2001).
In order to analyze the vast amount of data obtained in this experiment from microarray and real-time RT-PCR analysis, this lab employed the use of JMP Genomics. SAS developed JMP Genomics to deal with the large volumes of complex data generated from advancing scientific methods. Instruments such as high-throughput sequencers, mass spectrometers, and microarray processors create data sets typically several megabytes per experiment. JMP Genomics contains input engines and customizable analytical processes to deal with these types of data. The JMP Genomics program uses SAS code. It allows users an interface to manipulate data with statistical calculations and generate outputs including data sets, lists, charts, and graphs. This program provides sophisticated statistical analysis to accommodate the volume and density of genomics information and data sets generated by today’s high-throughput technologies (Wolfinger 2003). For this study, the JMP program was used to analyze microarray results with one way ANOVA (Analysis of Variance) (Figure 9). The program also contains a feature to allow Lowess smoothing (Cleveland 1979) which was utilized to normalize data within and between arrays. Another feature allows the creation of volcano plots (Figure 10) from the data which graphically show the fold changes of genes between ages and lines compared to their p-values. This aspect facilitated the selection of genes for further analysis.
Pathway Analysis

Metacore’s GeneGo pathway analysis is an internet based program that allows the input of experimental data in order to generate network maps of disease and biological processes. Given gene expression data including expression intensity and significance from microarray analysis programs, GeneGo can determine the association of genes in the experimental data with known biological pathways and processes (Figures 11a-b) as well as identify which of those genes are putative disease biomarkers. GeneGo determines associations from Metabase, a manually curated database containing over six million experiments in mammalian biology and chemistry.
RNA Isolation from Spermatozoa

Ribonucleic Acid Activity in Spermatozoa

The discovery that spermatozoa contain not only DNA, but also RNA has prompted many to investigate its potential role in developing sperm and the post-zygotic contribution of the male genome. There are many genes which are expressed exclusively in the developing male gamete. Protamine genes, \textit{prm1} and \textit{prm2}, are expressed in round spermatids and are under transcriptional and translational control (Kleene 2000). It has been shown that protein repressors bind to the poly-A tails or specific RNA sequences in the 3’-UTR (untranslated region) to cause translational regulation (Steger 1999). The protamine mRNAs are transcribed in the round spermatid and then stored as ribonucleoprotein complexes until they are translated in the elongating spermatid (Sinclair 1982). It has been suggested that alternative sperm chromatin structure is intended for imprinting purposes or that it is to help with the repackaging and activation of the male contribution (Kramer 1997). In mice it was shown that premature expression of these protamines stopped spermatid differentiation (Lee 1995).

The initial detection of RNA in sperm came in 1988 in the sperm nucleus of the fern \textit{Scolopendrium} (Rejon 1988). Next, RNA was visualized in rat spermatozoa using an RNase-colloidal gold procedure (Pessot 1989). It was not until 1993 that specific RNAs were identified in the sperm of the pollen grain. These RNAs were U1 and U2 snRNAs.
(Concha 1995). The first mRNAs were identified in human sperm cells in 1993 and were the proto-oncogene *c-myc* (Kumar 1993). Reverse transcription PCR was first used in the detection of sperm RNA in 1994 and detected human leukocyte antigens (Chiang 1994). Differential display was used to find that β-actin and PRM2 mRNAs were present in normal semen while azoospermic semen lacked PRM2 and only had β-actin. This discovery led investigators to hypothesize that RNA fingerprinting could be used as indicators of male fertility (Miller 1994).

Sperm not only contribute DNA to the ovum, but upon fertilization, they also deliver many novel RNAs (Krawetz 2005). They also release signaling molecules and transcription factors. STAT4, which activates transcription, and PT32 which participates in the early events of zygotic development are also passed along. Sperm-specific RNAs have been found in the fertilized oocyte (Ostermeier 2004). Using hamster sperm penetration assays (Johnson 1995), several genes were identified in sperm that were absent in human or mouse oocyte Serial Analysis of Gene Expression (SAGE) or cDNA libraries and were not found in unfertilized oocytes (Ostermeier 2004). One gene of interest discovered in this method was *foxl1b* which is a transcription factor that is vital to embryo patterning but is only found in the adult testis or the developing fetal brain (Murphy 1994). Another gene, *wnt5a*, is a proto-oncogenic signaling molecule that is involved in cellular differentiation and morphology development (Moon 1997). Short half lives of most of the RNAs discovered indicates that they may be important in embryogenesis (Krawetz 2005).
The birth of a parthenogenetically derived mouse indicates that sperm are not absolutely required for male derived factors. The successful birth was achieved through imitation of the *igf2-h19* gene cluster that is normally contributed paternally. While the 0.3% success rate and eighty percent of the animals failing to be resuscitated at birth (Kono 2004) shows that a developmental switch may be lacking, the experiment showed that it is possible to bypass some of the paternal control points (Krawetz 2005). Studies revealed that a miRNA complementary to the *igf2* receptor is present in human sperm and it may be a mechanism of control (Ostermeier 2005).
The Sperm transcriptome as an Indicator of Fertility

Transcriptional profiling of the spermatozoa of normal fertile men can give insight into fertility issues. In the most comprehensive study of sperm RNA transcripts, 2,780 non-poly(A)-enriched transcripts from one individual was compared to a pool of 3,281 poly(A)-enriched transcripts from nine other men. The study found that all but four transcripts from the individual were found in the pool. The pool was also compared to a pool of testes transcripts and all of the semen transcripts were found in the semen pool. This suggests that the gene expression of the sperm can be viewed as a window to past events in spermatogenesis in the testis (Ostermeier 2002).

This also represents the ability to create a fertility profile from the genes expressed in sperm. In a follow-up study, gene expression of fertile and infertile men was evaluated in an effort to establish spermatozoal RNAs as a validation of markers of male fertility. Samples were taken from men of ideal, medium, and poor quality semen and hybridized to microarrays. The study revealed 228 total transcripts from all of the samples. There was a 98% homology between the ideal and medium quality samples and a 39% homology between the medium and poor samples. The amount of variation among normal fertile men is very low while the variation in gene expression in non-fertile men is high, suggesting that it will be much more practical to create a transcriptional profile as a predictor of good fertility rather than of infertility, which comes with much more variability (Ostermeier 2005).
There are three phases of spermatogenesis (Figure 12) that contribute to the establishment of the male genetic component. They are spermatocytogenesis, spermatidogenesis, and spermiogenesis (Figure 13). In spermatocytogenesis, stem cells called spermatogonia undergo mitosis to form diploid primary spermatocytes, which then undergo meiosis to form haploid secondary spermatocytes. After spermatocytogenesis, the secondary spermatocytes undergo meiosis II to produce haploid spermatids in a very quick process known as spermatidogenesis. Spermiogenesis is the final phase of development where spermatids mature into spermatozoa (Figure 14) and is itself divided into four phases. In the Golgi phase the sperm head is formed and the Golgi apparatus creates the enzymes used in the acrosome. During the cap phase the Golgi apparatus surrounds the nucleus to form the acrosomal cap. Next, in the acrosomal phase the tail is formed and finally, during the maturation phase residual bodies made of excess cytoplasm are phagocytosed by Sertoli cells. Sertoli cells in turn release the sperm in secretory fluid into the seminiferous tubule from which point they travel to the epididymis to mature (Campbell 1995).
Techniques used for Semen RNA

Separation of Viable Sperm Utilizing a Discontinuous Percoll Density Gradient

Several techniques were employed in the investigation of the sperm transcriptome. This study first proceeded to procure viable sperm. The ideal separation of viable sperm from an entire ejaculate should select for the highest number of competent sperm without contamination from somatic cells or seminal plasma. The two most popular methods for isolation of viable sperm from whole ejaculates are the swim up method and separation by Percoll gradient (Moohan 1995). In the swim up method, sperm are placed in a tube with modified Earle’s Balance Salt Solution- 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic Acid (EBSS-HEPES) buffer at a 45° angle followed by centrifugation of the viable phase for collection. In the Percoll Density Gradient procedure, layers of decreasing Percoll density are placed in a tube and upon centrifugation the viable sperm will be in the highest density phase, while degenerate and dead sperm along with somatic cells will remain in the low density Percoll phase (Moohan 1995).

The direct swim up method and Percoll density gradient separation methods were scientifically evaluated by previous studies (Moohan 1995) to determine which method produced the greatest number of viable spermatozoa. The Percoll technique isolated on average 90% of the total sperm in the ejaculate, while the swim up method produced only 73% of the total sperm number. While sperm isolated by swim up exhibited greater
motility, those isolated with the Percoll density gradient had higher curvilinear velocity (total speed and distance in all directions) and lateral head displacement (the measurement of the three dimensional movement of the sperm head displayed by hyperactive sperm), with straight line velocity being the same for both groups. The sperm isolated by the Percoll gradient also exhibited a higher level of hyperactivation for up to six hours after isolation, a trait necessary for fertilization. To examine the effects on sperm longevity, progressive motility in both groups was monitored and a decrease was found in sperm isolated by the swim up method and not the Percoll gradient method. Therefore, it was concluded that the Discontinuous Percoll density gradient is the most effective method for selecting viable spermatozoa, especially when used for sensitive applications such as reproduction (Moohan 1995). Consequently, this was the technique selected for this study.
Semen RNA Isolation Techniques

There are several proposed techniques for extracting RNA from semen samples and many were examined in order to determine a suitable protocol for this study. Several labs have published protocols for the extraction of RNA from various species’ semen samples. The standard for RNA extraction from human semen samples was published by Ostermeier and co-workers of the Krawetz lab. This procedure claims to yield pure spermatozoa (Ostermeier 2002). The protocol begins with fresh human ejaculates allowed to liquefy at room temperature for approximately thirty minutes before they are washed in a frozen storage buffer consisting of HEPES, NaCl, Magnesium Acetate, and glycerol. Samples are then frozen at -80°C and thawed, washed twice in PBS, and stored in Somatic cell lysis buffer containing Sodium Dodecyl Sulfate (SDS), Triton-X, and water. Samples are then extracted using the Qiagen RNeasy kit with modifications to manufacturer’s instructions using a twenty six gauge needle for homogenization and heating the lysate for thirty minutes at 65°C. Homogenization was repeated and then the sample was placed on the Qiagen spin column. Then the sample was processed according to Qiagen’s instructions. After elution, dithiothreitol (DTT) and RNAse-block was added before a twenty minute incubation with RNAse-free DNase-I. Quantification was achieved with Ribogreen RNA quantifying kit. In order to determine the presence of RNA, RT-PCR was performed with intron-spanning primer prm2 and a 149bp fragment was observed. The human genomic DNA band would have been 310bp and corresponded to DNA contamination (Ostermeier 2005).
The Goodrich lab adopted a similar protocol for human sperm. This protocol is very similar to the one described above except that samples are homogenized with a Pro 200 homogenizer and samples are not heated at any point in the procedure. Samples can also be measured by Ribogreen (Molecular Probes), plate scanning using a Typhoon Imager, or a Nanodrop. The prm2 primer was also used to detect RNA using RT-PCR and the lack of DNA contamination (Goodrich 2007).

While examining voltage-dependent calcium channels in spermatozoa, Goodwin et al devised a protocol to extract RNA from fresh sperm obtained through the swim-up method and through percoll density gradient centrifugation. This lab used guanidinium isothiocyanate as well as a modified Purescript RNA Isolation Kit (Gentra Systems). Modifications included the addition of dithiothreitol to 40mmol/l and a four hour incubation at 55°C before the Protein-DNA purification step (Goodwin 2000).

In 2004, Wasilk compared the phenol-chloroform RNA extraction method with Qiagen’s RNeasy protocol for extraction of RNA from boar semen. This lab was interested in obtaining viral RNA from the semen. Whole semen was centrifuged with the seminal plasma removed and then resuspended in an equal volume of sterile phosphate buffered saline. It was found that Qiagen’s RNeasy kit was more effective at RNA extraction from boar semen, however the procedure was modified from manufacturer’s instructions. To solubilize the semen, equal volumes of proteinase K-HIRT buffer was added to the semen sample and heated at 37°C for ten minutes, followed by a centrifugation. The supernatant was added to Guanidinium isothiocyanate (GITC) lysis buffer in place of Qiagen’s RLT
buffer with 70% ethanol and the remainder of the manufacturer’s instructions were followed. Real-time PCR was used to validate the presence of viral RNA (Wasilk 2004).

In 2006, a SAGE library of human spermatozoa was constructed from RNA extracted by using Trizol RNA reagent (Invitrogen) following the manufacturer’s protocol. From a single ejaculate containing $4.6-9.2 \times 10^7$ sperm were isolated to yield about 0.7-1.4µg of RNA. RNA was pooled from ten individuals and a SAGE library was constructed from one of the individuals and from the pooled ejaculate RNAs. The sequenced individual SAGE library produced 20,237 raw tags which contained 2,459 unique tags, while the pooled sample library yielded 21,052 raw tags containing 2,712 unique tags. Of these, 638 tags were unique to the individual sample, while 682 tags were unique to the pooled ejaculate sample. There were 564 overlapping genes that were identified to be involved in transcription regulation and protein synthesis (Zhao 2006).

Studies to characterize the role of atrial natriuretic peptide in the acrosome reaction of sperm led to a protocol to isolate RNA from boar spermatozoa in order to ascertain its expression level. Semen was collected from boars using the gloved hand method and then subjected to a Percoll gradient of 90% and 45%. Two and a half milliliters was added each of 90% and 45% Percoll respectively, followed by 1ml of fresh semen in a 15ml tube. Next, the tube was centrifuged at 700g for ten minutes and the pellet was washed with PBS three times. RNA was extracted using Sigma’s Tri Reagent according to manufacturer’s instructions and dissolved in 20µl nuclease-free water. Semi-quantitative PCR was used to verify the presence of the ANP transcript (Zhang 2006).
The Environmental Protection Agency (EPA)’s National Health and Environmental Effects Research Laboratory (NHEERL) developed protocols for isolation of RNA from both human and rodent sperm. This protocol is initiated with fresh sperm washed with PBS. The novel technique of this protocol is the addition of the Somatic Cell Lysis Buffer (SCLB) which is used to remove contaminating somatic cells. After an incubation in SCLB on ice, samples are examined under the microscope to affirm the absence of somatic cell contamination. If somatic cells are still present, the SCLB procedure is repeated. Following isolation of pure sperm, the protocol recommends the use of Sigma’s Tri Reagent to isolate RNA (Thompson 2003).

From North Carolina State University, the Farin lab has developed a protocol for the extraction of spermatozoa RNA from extended equine semen (Wrench 2008). Briefly, extended semen is divided into 50ml tubes, centrifuged, and the seminal pellet is divided into smaller tubes before Trizol Reagent (Invitrogen) is used to isolate RNA. The procedure is verified when an mRNA smear is visible on an agarose gel.

In 2007 Lalancette et. al used suppressive-subtractive hybridization to study markers for bull fertility. During this process she examined the transcriptomes of bulls of high fertility versus bulls of low fertility. They also developed a protocol for the isolation of RNA from frozen bull semen. Using two frozen straws per extraction, semen was thawed at 30°C for two minutes and then layered on a 40%/70% Percoll (Sigma) gradient. After centrifugation at 700g for 45 minutes at room temperature the pellets were washed in TL-SPERM (Sigma). RNA Extraction was performed using Trizol Reagent (Invitrogen)
according to manufacturer’s instructions except for a few modifications. Cells were resuspended in Trizol and incubated at 60°C for thirty minutes with vortexing every ten minutes to lyse cells. The ethanol precipitation in the Trizol protocol was carried out with the addition of a linear acrylamide carrier (Ambion/Biosystems). Samples received DNase I (Ambion) treatment and precipitation with ammonium acetate also included the linear acrylamide carrier. Sample RNA was quantified using Ribogreen Reagent (Molecular Probes) and a Nanodrop ultraviolet/visible spectrophotometer. *Dazl* and *prm1* primers were used in RT-PCR to verify the presence of cDNA. Next, amplification of RNA was accomplished using DB Supersmart cDNA synthesis kit and then each cDNA library was constructed. These libraries through suppressive-subtractive hybridization led to profiles of bull fertility and identified new transcripts previously unknown in the bull transcriptome (Lalancette 2008).
Ribogreen Quantification

Accurate quantification of nucleic acids is an essential starting point in many assays. There are a variety of methods to quantify RNA. Ribogreen is a very sensitive fluorescence based RNA quantification assay and was the quantification method of choice for this study. When bound to nucleic acids it fluoresces but the unbound form is non-fluorescent. It can detect RNA in concentrations as little as 1.0ng/ml. When quantification was in range of 1.0 to 50ng/ml, the linear correlation of fluorescence detection versus RNA concentration was \( r^2 = 0.999 \) and this linearity was consistent even in the presence of ethanol, urea, protein, chloroform, agarose, salts, and even some detergents. Its fluorescence is relatively independent in fragment sizes from 500 bases to 9kB. However, at smaller sizes, such as 100 base fragments, fluorescence signal intensity is decreased twenty eight percent. The protocol consists of creating a reference standard curve of increasing amounts of RNA with the reagent. Fluorescence is measured using a microplate reader at an optimal excitation of 485nm and emission at 530nm and samples values are compared with standards. The assay is two hundred times more sensitive than ethidium bromide (Jones 1998).
Conclusion

While the swine industry appears to be one of the last to rely on male genetics for selection, male genomic analysis could lead to improvements in health, growth, and production of food animals. Through this study and the use of microarray analysis, real time RT-PCR technology, and sophisticated data analysis, the effects of testosterone level on individual animals as well as their offspring will be highlighted by examining the resulting gene expression in divergently selected high and low testosterone lines. In addition, we hope to be one step closer to unlocking the genetic secrets contained in the spermatozoa transcriptome which is said to hold the key to past spermatogenesis and a large genetic contribution to future generations, by the development of a viable sperm RNA isolation protocol. Using many existing techniques, and a few novel ones, we hope to contribute to the knowledge of porcine genomics to equip the swine industry with the tools necessary to feed the world’s population, as well as provide insight into health issues facing humans and other species.
Figures

Figure 1. Reproductive Tract of Male Pig

Copyright © www.ThePigSite.com - Reproduced with Permission
Figure 2. Histological section through testicular parenchyma of a boar.
Figure 3. Sertoli Cell. Sertoli cell (S) nuclear morphology in pigs at birth (A), puberty (B), and sexual maturity (C). Gomori trichrome, x1090 (Franca 2000)
Figure 4. Leydig cells. Leydig cells in the interstitial tissue of the boar testis.

Copyright Dharam M. Ramnani, MD, courtesy of the Biology Image Library
http://biologyimagelibrary.com/imageID=37082
Figure 5. Structure of Testosterone

(www.worldofmolecules.com)
Figure 6. Testosterone and GnRH negative feedback loop (Amory 2007).
Figure 7. Swine Oligonucleotide Arrays as visualized in ScanAnalyze
Figure 8. Real-time RT-PCR gene expression. Each sample Ct indicated by sample crossing red threshold line.
Figure 9. 150 day volcano plot. Volcano plot generated by JMP Genomics showing differential gene expression in 150 day animals. Each point represents the expression of one gene.
Figure 10. One way ANOVA of real-time RT-PCR results generated by JMP. Indicates differential expression of ADFP by treatment group 1. 150Low 2. 150High 3. 180Low 4. 180High
**Figure 11a.** TIMP1, ESE3, HMG2, MAD2a, TRAM1 network identified by GeneGo from experiments. Genes up-regulated in 150High animals are marked with red circles; genes up-regulated in 150Low animals with blue circles. The 'checkerboard' color indicates mixed expression for the gene between multiple tags for the same gene.
Figure 11b. Metacore GeneGo Network Objects Legend
**Figure 12.** Spermatogenesis overview where type A stem cells lie on the basal lamina of the convoluted seminiferous tubules. These divide and their daughter cells migrate towards the lumen. In humans in roughly 64 days they differentiate themselves into sperm cells up to the outer surface of the epithelium.

(http://www.embryology.ch/dutch/cgametogen/spermato03.html)
Figure 13. Seminiferous Tubule. Gray’s anatomy cross-sectional diagram of the seminiferous tubule showing spermatozoan with tails in spermiogenesis during the acrosomal phase (http://education.yahoo.com/reference/gray/subjects/subject/258)
Figure 14. Structure of Spermatozoa
References Cited


Concha, II (1995). "U1 snRNP components are present in the vegetative and generative nuclei of the pollen grain." Sexual plant reproduction 8(6).


Moohan, J. M., Lindsay, Kevin S. (1995). "Spermatozoa selected by a discontinuous Percoll density gradient exhibit better motion characteristics, more hyperactivation, and longer survival than direct swim-up." Fertility and Sterility 64(1): 160-165.


Testicular Differential Gene Expression in Male *Sus scrofa*

Selected for High and Low testosterone

Abstract

Testosterone influences growth and reproductive characteristics in domestic animals. In food animals, differences have been noted in meat quality, litter size, growth rate and other economically important traits. To investigate effects of testosterone, Duroc boars were divergently selected for 10 generations for both high and low levels of testosterone and then maintained until generation 21 by random mating within lines. Boars in the high testosterone (HT) lines exhibited greater epididymal weight and higher Sertoli cell volume density and mass, while females had a greater number of piglets born alive. The low testosterone (LT) displayed greater seminiferous tubule volume and mass. In a pilot study, swine 70-mer oligonucleotide microarrays were used to identify differential gene expression in the testes samples of boars from these lines at 1, 30, 120, 150, and 180 days of age pooled by age group (n=5). Based on the preliminary study, gene expression differences were most pronounced in the 150 and 180 day samples, which were used for follow-up microarray analysis using the 5 individuals within each group and age. Approximately 100 genes with divergent expression patterns in the individuals were identified and selected for further investigation. Additional analysis included validation by Real-time RT-PCR and GeneGo
pathway analysis to identify biological pathways associated with testosterone. As expected, affected biological processes include growth pathways, but several inflammation and adipocyte differentiation pathways also appear to be affected. Several of the differentially expressed genes in high testosterone animals are considered biomarkers in diseases including many types of cancer, aortic diseases, and other genetic disorders, but also represent markers for increased growth. In light of this study, we have been able to identify specific genes affected by testosterone that may be able to improve health status, fertility, meat quality, and maturation in boars, and have gained a broader insight into the gene pathways affected by testosterone.
Introduction

Testosterone levels have been linked to the rate of growth of boars, even so much as to prompt the study of exogenous testosterone supplements. These studies have shown that the addition of testosterone promotes the deposition of muscle and nitrogen balance in castrated animals (Lund-Larson, 1977). However, boars can be genetically selected in order to produce high or low levels of testosterone, and effects can be seen in subsequent generations. The female offspring of boars with high testosterone have been shown to produce larger litter sizes (Robison, 1994). Male selection intensity is also much higher than female selection, so that selection of male traits over female traits can lead to greater rates of genetic progress (Robison, 1994). To this end, a study on Duroc boars over ten generations showed that selection for either large testes or high testosterone could be used to develop breeding stock enhanced for growth, a valuable market characteristic (Lubritz, 1991). A follow-up study, looking at the same lines after twenty-one generations indicated that boars high in testosterone had higher volume densities of Leydig cells than the line selected for low testosterone, but the density of Sertoli cells did not change. Subsequently, sperm production was also relatively equal between the two lines (Walker, 2004).

The reproductive maturation of boars occurs in very distinct phases, beginning shortly after birth. This development has been mapped through the proliferation of specialized cells in the testes and changes in hormone levels. Between birth and one month of age, the number of Sertoli cells, cells responsible for sperm development, gains a six-fold
increase in each testis. Also during this time, Leydig cells, cells that secrete testosterone, also show a marked increase in number. At the onset of puberty at approximately three to four months of age, until its completion at approximately four to five months of age, the number of Sertoli cells doubles and Leydig cells also dramatically increase in number. Simultaneously, during both of these stages of proliferation, FSH plasma levels are high (Franca, 2000). These physiological changes led us to further investigate genetic differences present between the two lines.

Now that microarray technology is available to interrogate the whole genome at once, the following study has examined differential gene expression exhibited by boars with high or low testosterone at different developmental time points. Samples from these boars have been used in microarray analysis and subsequent validation with quantitative real-time RT-PCR to determine differentially expressed genes. Determination of the effect of testosterone on fertility, growth, development, health, and other important characteristics will allow the expansion of marketable applications for the use of genetics to further examine and advance swine characteristics.
Materials and Methods

Selective Mating of Boars

A line of Duroc pigs maintained at North Carolina State University according to industry standards was divergently selected for ten generations from 1982 to 1992 for high and low testosterone levels in response to a GnRH challenge as previously described (Robison 1994). A RadioImmunoAssay was utilized to determine the levels of testosterone in the samples (Juniewicz 1984). The five boars with the highest and lowest post-GnRH challenge testosterone levels were selected to sire the next generation (Robison 1994).

Data was maintained on testes length and width on these boars, as well as estimated testes volume at 168 days. Backfat depth was measured by ultrasound and adjusted to 104kg of body weight. Growth data were also maintained as days to 104kg.

After selection of the tenth generation, testosterone levels as indicated by pre- and post-challenge values, were approximately three times higher in the high testosterone line than in the low testosterone line (Robison 1994). From this point, lines were maintained through random mating within each line. At generation 21, the testosterone levels were once again evaluated to validate that the lines remained divergent. Testosterone levels averaged 49.0ng/mL in the high line and 27.8ng/mL in the low line (P<0.01) (Walker 2004).
Sample Collection

At generation 21, both whole testes were collected from each line from five boars at each time point (1 day, 30 days, 120 days, 150 days, and 180 days). Boar housing and maintenance, as well as sampling, were completed according to NCSU IACUC standards by trained technicians. Intact samples were flash-frozen and stored at -80°C.
RNA Isolation and Microarray Hybridization

RNA was extracted from the parenchyma of one testis per boar using a protocol combining the use of TriReagent (Molecular Research Center) and Qiagen’s RNeasy kit. RNA was isolated from each individual by first obtaining an approximately 1cm² section of frozen parenchyma by shattering one whole frozen testis with an RNase-free mallet. The testis sample was combined in a screw top microcentrifuge tube with 0.1mm glass beads, 1ml RLT Buffer (Qiagen) and 10µl β-mercaptoethanol (BME). Homogenization was completed by beating in a Beadbeater homogenizer (Biospec) until each sample was fully homogenized. Samples were cooled on ice after homogenization and in between sessions in the beadbeater if more than one homogenization was required. Next, samples were centrifuged at (16,000g) for 3 minutes and the supernatant was transferred to a new tube. After the addition of one volume of 70% ethanol mixed by pipetting, the sample was transferred 700µl at a time to the RNeasy spin column and placed in a 2ml collection tube. Each sample was then centrifuged 15 seconds at 8,000g. The flow through was discarded and the collection tube was reused. Any remaining sample was then transferred to the spin column, centrifuged, and the flow through was once again discarded, preserving the collection tube for reuse. Next, 700ul of Buffer RM1 was added to the RNeasy spin column and the remainder of the manufacturer’s instructions (Qiagen) were followed, until samples were eluted with 30ul RNase-free water. Quality and concentration of samples were quantified using a Nanodrop spectrophotometer and samples were stored at -80°C.
Five micrograms of each of the five RNA samples was pooled to create treatment groups of 150Low, 150High, 180Low, and 180High representing age in days and genetic line (n=5). From the 25ug pool, 5ug of RNA was used according to manufacturer’s instructions (Promega) to create fluorescently labeled cDNA using the ChipShot Indirect labeling and Clean-up system. Briefly, aminoallyl cDNA was synthesized, RNase H treated, and then purified. Cy3 and Cy5 dye (Promega) was conjugated to the aminoallyl cDNA. Samples were quantified using a Nanodrop spectrophotometer. Pooled samples were then hybridized using the using the Pronto Universal Hybridization Kit (Corning) following all manufacturer’s instructions for twelve hours to twenty swine protein-annotated oligonucleotide microarray slides (Consortium 2009) in the design shown in Figure 1a.
Microarray Analysis

Fluorescence spot analysis was completed using Scanalyze (Eisen Lab) which provides a method for gridding microarray scanned images, evaluating their signal intensities and background, and providing an output of numerical signal data which was then evaluated by JMP Genomics Version 3.2 (SAS, Cary, NC). Using JMP, data were normalized between and within lines with Lowess smoothing (Cleveland 1979) and then line by age comparisons were made with one way analysis of variance (ANOVA), while compensating for multiple comparisons using FDR ($\alpha \leq 0.05$). ANOVA is a statistical test to determine if the means of separate groups are equal given the assumptions that the data groups are independent, normally distributed, and that they share homoscedasticity, meaning equal variances (Lindman 1974). The sum of squares are then partitioned into effects components and compared using an f-test. Here, the F-statistic can be described as the between group variability divided by the within group variability and is compared to the critical value in an F-distribution represented at some value of $\alpha$ (Lomax 2007). The False Discovery Rate (FDR) adjusts large amounts of data based on the proportion of false positives in the significant results (Benjamini 1995) and has become a widely accepted method for evaluation of multiple hypothesis testing to correct for multiple comparisons in the thousands of statistical tests involved in the evaluation of microarray data (Pawitan 2005). Subsequent volcano plots of each line by age comparison of the data were used to determine genes for further investigation of differential expression. Approximately one
hundred genes were selected on the basis of the greatest expression fold changes with the most significance and were briefly analyzed with Metacore™ Geneo pathway analysis to determine if further investigation should be pursued.
Follow-up Microarray Analysis of Individual Animals

The two age groups appearing to have the greatest differential expression between lines from the initial microarray experiment were chosen for follow-up microarray analysis of individuals. RNA from the 150 and 180 day animals was labeled, hybridized individually to arrays using the same Corning Pronto Universal Hybridization kit as described above with some modifications to minimize background interference on the microarrays. Instead of 42°C incubations, all incubations in the hybridization and post hybridization protocols were performed at 48°C. In addition, hybridization incubation time was increased from twelve hours to eighteen hours. Furthermore, a new experimental design was devised to maximize comparisons between lines (Figure 1b). The microarray data were analyzed with JMP Genomics in the same method as that of the initial experiment with one way ANOVA and an FDR cutoff of $\alpha \leq 0.05$. When this method did not provide significant results, the data were further evaluated with a $-\log_{10}(p\text{-value})=3$ cutoff ($p\text{-value} \leq 0.48$) to select genes for further exploration. This set of data containing approximately fifty genes was then evaluated through real-time RT-PCR and pathway analysis.
Real-time RT-PCR

Using Beacon Designer 7.0 (Premier Biosoft), primer sets were designed specifically for each identified gene for SYBR green probes to avoid cross homology and template structure issues as well as to span an intron-exon boundary whenever possible. Primers were designed from porcine sequences whenever possible or from human sequences in the absence of viable porcine primers. cDNA for real-time PCR was created from each boar’s RNA sample via ABI’s High Capacity cDNA Archive kit using random primers according to manufacturer’s instructions. Quantitative real-time RT-PCR was used to validate genes with differential expression seen by microarray analysis between boars in divergent lines at 150 or 180 days of age. Data was analyzed using the Pfaffl method (Pfaffl 2001) using β-Actin as a reference gene. The Pfaffl relative expression ratio was calculated for each target gene as Ratio = \[
\left(\frac{\Delta CP_{target}}{\Delta CP_{ref}}\right)_{control-sample} / \left(\frac{\Delta CP_{ref}}{\Delta CP_{target}}\right)_{control-sample}
\]
where $E_{target}$ is the real-time PCR efficiency of the target gene, $E_{ref}$ is the real-time efficiency of the reference gene, $\Delta CP_{target}$ is the Cp (Crossing point or Ct) deviation of the control minus sample of the target gene, and $\Delta CP_{ref}$ is the CP (Ct) deviation of the control minus sample of the reference gene (Pfaffl 2001). Next, relative expression ratio data was analyzed in JMP using one way ANOVA. Following this analysis, gene Cts and efficiencies were analyzed in Relative Expression Software Tool (REST) which calculates not only significance but converts expression levels to fold changes based on gene Cts and PCR efficiencies (Pfaffl 2002).
Pathway Analysis

By comparing results from both the microarray analysis and quantitative real-time PCR, approximately 100 differentially expressed genes from the initial microarray experiment were examined using GeneGo MetaCore\textsuperscript{TM} version 5.4 pathway analysis software to identify pathways and biological processes potentially different in all of the original treatment groups examined. Additionally, due to observed differences in expression in the 150 day animals in the follow-up microarray analysis, the selected fifty genes along with their expression differences were examined in more detail for biological significance using GeneGo. Specifically, an excel file containing a list of genes determined to be differentially expressed in each comparison of HT vs. LT, and 150 day vs. 180 day from the microarray experiment was uploaded to Metacore’s Data Manager. The uploaded excel file also contained the differences in signal intensity in the form of high testosterone signal intensity minus low testosterone signal intensity or 150 day intensity minus 180 day intensity for each differentially expressed gene in the data set. The examination of the data was accomplished through the “Analyze Single Experiment” function with a defined threshold of 0.001 (p-value ≤ 0.05). When prompted, application of the up regulation option produced a report of the genes differentially up-regulated in high testosterone animals. Likewise, selection of the down regulation option produced a report of the genes differentially up-regulated in the low testosterone animals. Additional analysis to see differences in the 150 day age group was initiated using the EZstart function to complete network analysis on the
experiment by designating the program to build biological networks from the differentially expressed gene list. From the generated report, the top twenty-five biologically networks were identified. Affected pathway and disease biomarkers were also determined for the 150HT and 150LT animals in the same manner as described above. These results were then analyzed to develop a concept of the pubertal gene expression created by differing levels of testosterone.
Results and Discussion

Many genes in the original pooled microarray study were up-regulated in the 150 day animals, especially in the high testosterone line. In the subsequent study of 150 and 180 day old animals the distribution of up-regulated gene expression was approximately equal in both lines (HT: 53%) and both age groups (180 day: 51%). Volcano plots were produced to show the fold changes between treatments and ages and corresponding p-values (Figures 2a-d). Since this experiment had reduced statistical Power due to the small sample size, results lacked significance (p-value \( \leq 0.05 \)). In order to utilize the microarray results as an exploratory tool, stringency was reduced to a reasonable level (p-value \( \leq 0.48 \)). From this criterion, approximately 100 of the 27,000 genes deemed biologically relevant and having the highest fold changes between lines and most significant p-values were selected for pathway analysis and 11 were selected from volcano plots for validation by Real-time PCR. Approximately 65% of genes in this 100 gene set were up-regulated in the 150 day animals presumably because this is nearing the age of maturity for swine when many physiological changes are expected. Distribution of up- and down-regulated gene expression was approximately equal between lines and between and within lines comparisons.

Real-time RT-PCR results were analyzed with JMP and REST for the 11 genes and were consistent in expression direction (up or down regulation) with the microarray analysis, though not all expression levels were proportionate and not all were significant at a p-value of 0.05 (Tables 1-3). It is important to note that when comparing microarray analysis results
to real-time RT-PCR results, that microarray analysis is based on signal intensity which increases in proportion to expression. In contrast, real-time RT-PCR is measured in a sample ratio calculated as a ratio between the Ct (threshold crossing point) of the housekeeping gene and the Ct of the experimental gene, so that a greater ratio is indicative of increased Ct value, and decreased expression, which can also be converted to fold changes in expression. ANOVA of real-time RT-PCR was performed using JMP Genomics in line by age comparisons, line comparisons, and age comparisons (Table 3) on 11 genes from the microarray results. In comparison REST was used to calculate the fold changes in gene expression of the HT in relation to the LT line in both the 150 day (Table 1) and 180 day animals (Table 2).

Many of the genes identified through the microarray analysis were also discovered in biological pathways by GeneGo. These pathways included regulation of lipid metabolism, angiotensin signaling, and signal transduction. Processes affected include progesterone signaling, inflammation, and cellular adhesion and proliferation. Disease biomarkers were identified for coronary arteriosclerosis, myocardial ischemia, arterial occlusive diseases, kidney diseases, hyperaldosteronism, and hypotension (Data not shown).

In order to determine the biological significance of the genes indicated by the greater expression disparities found in the 150 day microarray analysis, a more extensive pathway analysis was once again accomplished through Metacore’s GeneGo. Through the use of gene signal intensities from the microarray analysis, this program was utilized to create maps of relevant biological networks and disease processes indicated by differences in the
microarray results for the 150HT and 150LT animals. In this way, the vast data obtained from examination of 20,000+ genes become biologically relevant and used to predict practical application of the knowledge gained from them.

If given microarray signal intensity input, GeneGo evaluates and ranks the placement of genes in specific networks through a variety of methods. In this analysis, special consideration was given to two ranking functions. The Z-score and p-value were used to rank the genes according to their relevance in biological networks. The Z-score takes into consideration the number of nodes and objects in a subnetwork and ranks the subnetwork according to its level of saturation of the genes expressed in the experiment (Table 4). A subnetwork given a high Z-score would be considered highly saturated in genes from the experiment and is given more biological consideration.

The p-value given to the networks in GeneGo represents the probability of mapping an experiment to a network map by chance considering the number of genes in the experiment versus the number of genes in the network map. Its calculation is essentially the same p-value used in most hypergeometric distributions. The lower the p-value, the less likely the experiment was erroneously mapped to a network.

Examination of GeneGo pathway analysis, highlighted through the identification of networks, Gene Ontology (GO) processes, and disease biomarkers associated with experimental expression levels provides a snapshot of the biological impact of the data. Genes up-regulated in the High testosterone line point to a role in the GO processes ATP synthesis and energy transport as well as a player in immune response through interleukins
IL-10 and IL-12. Genes up-regulated in the Low testosterone line are associated with the DNA repair process and the synthesis of cellular components such as nucleosome, chromosome, and chromatin assembly (Figure 3).

When examining the biological pathways involved with high testosterone, some of the most significant pathways are the cellular adhesion pathways through cell matrix glycoconjugates and extracellular matrix remodeling (Figure 4). This corresponds to the RTq-PCR findings that \textit{cd44}, a gene important in cell adhesion, selected from the microarray data, has highest expression in the 150 day and HT line but results lacked significance. Still, \textit{cd44} is a large component of the extracellular matrix and its many isoforms have been implicated in the metastasis of several forms of cancer. Current research is focused on identifying those variants and creating anti-\textit{cd44} drugs as cancer treatments (Naor 1997). While this finding raises concerns about the health of the HT line, no cancer has been reported in these lines, and perhaps it is the increased action of cellular adhesion that has this gene up-regulated. The regulation of cellular apoptosis and lipid metabolism were also identified as active pathways in the HT line (Figure 4). The inclusion of these pathways offers little surprise after real-time RT-PCR investigation of \textit{rnase12} and \textit{adfp}.

While RnaseAs are required for normal cellular functions in the degradation of specific RNA sequences, \textit{rnase12} treatment is also known to induce apoptosis in aggressive anaplastic thyroid cancers (Kotchetkov 2001). After a 20 day injection of \textit{rnase12} thyroid tumors in nude mice began to decrease in size and were completely absent in 30 days while noncancerous cells were unaffected (Kotchetkov 2001). This anti-cancer effect may make it
useful to investigate the potential health benefits of RNases. The 150HT animals had
greater expression than the 150LT animals in JMP analysis of real-time RT-PCR results (p-
value ≤0.0428) and expression was significantly higher in the HT vs. LT and in 150 day
animals vs. 180 day animals. REST confirmed these results showing a 5.9 fold change in
eexpression between the 150LT and 150HT lines (p-value ≤0.049) (Table 1).

Although it has not been directly shown to cause an increase in adipose tissue,
adipose differentiation protein (adfp) or the adipophilin gene is associated with increased
adipocyte differentiation. It is involved with lipid droplet homeostasis (Chang 2007) and is
a potential concern because the industry prefers lean growth. In this study, JMP analysis
found adfp expression to be significantly higher in 150HT lines than the 180HT lines (p-
value ≤0.0146) and in the 150HT lines than the 150LT (p-value ≤0.0088). REST results
were concurrent, as 150HT adfp expression was found to be upregulated 2.88 over 150LT
animals (p-value ≤0.020) (Table 1). We would expect fat composition to be higher in the LT
lines, compared to lean weight, because increased testosterone levels are associated with
increased lean growth (Lubritz 1991), however a study on growth, carcass, and meat quality
traits with this line showed that the HT animals had a slightly smaller percentage of lean and
a larger backfat depth that LT animals (Bender 2006) so the identification of this gene and
involvement of the lipid metabolism pathway could also be rationalized.

Another pathway identified in the HT line was the EGF signaling pathway. This
corresponds with the selection of the epididymal growth factor (egf) gene from the
microarray data because of its significance and relationship with growth and weight gain.
*Egf* has also been shown to be responsible for the weight gain in ovariectomized rats. This weight gain was decreased in rats that were administered anti-*egf* rabbit antiserum (Kurachi 1993). JMP analysis showed that this gene was up-regulated in the 150HT over the 180HT, consistent with microarray results, but values were insignificant (p-value ≤ 0.05). Once data was analyzed with REST and reaction efficiencies were considered, *egf* showed the largest up-regulation observed in the 150 day HT animals over the 150LT animals of a 6.9 fold change (p-value ≤ 0.043) (Table 1). This upregulation in the HT line was anticipated, considering the line was originally sought for its increased growth rates.

Other pathways identified by pathway analysis (Table 3) included cAMP signaling in signal transduction (HT), various immune response pathways (HT, LT), as well as protein folding (LT), gluconeogenesis (LT), and transcription regulation (HT, LT). One of those regulators of transcription, *eef1a1* is an elongation transcription factor implicated in many disease processes. It has been found upregulated in many cancers (Goncalves 2005), found to be downregulated in Alzheimer’s patients (Brooks 2007), and is being explored as a possible target for male contraception. Gamendazole, one medication currently being examined is believed to bind to *eef1a1* and stop the bundling of F-actin which disrupts the integrity of the Sertoli cell-germ cell junction causing infertility (Tash 2008). This elongation factor maintained higher expression in the 150HT than the 180HT (p-value ≤0.0081) and in the 180LT than the 180HT (p-value ≤0.0439) (Table 3) which could indicate greater fertility. REST analysis showed higher expression in both the 150HT and 180LT groups but results lacked significance (Table 1, Table 2).
One of the most readily applicable features in the analysis of this experiment is the identification of disease biomarkers associated with divergent selection of testosterone (Figure 5). Through analysis with GeneGo low testosterone was shown to be associated with vitamin E deficiency, many types of infections, and types of ataxias. High testosterone was linked with several types of cancer, aortic aneurisms, and arthritis. Associated cancers included papilloma, erythroblastic leukemia, lymphangiomyoma, and lymphatic vessel and smooth muscle tumors. Some of the genes selected for RTq-PCR analysis had high associations with cancer.

Tissue inhibitor of metalloproteinase I (timp1) was no exception in its involvement with both cancer and arthritis and was the gene with the highest fold change in expression between high and low lines. Its expression was highest in the 150 day HT animals when compared to the 180HT (p-value ≤ 0.0432) (Table 3). It was also highest in the 150HT compared to the 150 Low Testosterone (LT) line, and in the high lines in general, but qRT-PCR differences were insignificant in this respect. REST analysis showed that there was a 2.6 fold change in expression between the 150HT and 150LT (Table 1) compared to only a .594 fold change in the 180 day animals (Table 2). This gene is best known for its role in tissue remodeling and has been implicated in processes from tissue degradation during arthritic changes (Martel-Pelletier 1994) to metastasis of breast cell carcinomas (Ree 1997). It is also active in normal processes such as the ovulatory growth of follicles, breakdown of the follicular wall, and the progression of the follicle to the corpus luteum to its regression in the estrous cycle (Curry 2003), (Curry 2001). Of particular interest to this study, is the role
of \textit{timp1} in the regulation of the extracellular matrix turnover during testicular development in prepubertal Sertoli cells in rats where it is stimulated by FSH (Ulisse 1994). In the 2004 study of these lines, FSH was elevated in the HT line and could explain the increase in \textit{timp1} expression.

\textit{Twistnb} was also investigated because of its connection with cancer and its importance in human health. While \textit{twistnb} has been associated with Saethre-Chotzen Syndrome, a syndrome comprised of fused cranial bones and abnormalities of the extremities, it was recently discovered that mutations in this gene are autosomal dominant and represent an increased chance of breast cancer in humans. In a Swedish study, 52\% of female Saethre-Chotzen patients eventually developed breast cancer (Sahlin 2007). In swine, a connection to Saethre-Chotzen syndrome or breast cancer has not been reported but such health issues if experienced by swine would lead to a market loss. No such problems were reported in these lines and in this study, there were no significant differences found in real-time PCR results, although expression tended to be higher in the HT line.

Calcyclin (\textit{s100a6}) is associated with calcium binding and cellular proliferation. Like many of the genes previously discussed, it is also routinely upregulated in many forms of cancer (Breen 2003). In some studies, this upregulation seems to be during invasion or metastasis followed by a downregulation (Komatsu 2000). Calcyclin had significantly higher expression in the 150HT animals than in the 150LT animals in the real-time analysis with a fold change of 5.9 (p-value $\leq 0.041$) (Table 1). The increased growth of these animals at puberty also makes this result characteristic. Interestingly, there was not a
significant difference in the 180 day old animals, where perhaps maturation and growth have started to slow.

Other genes from the microarray experiment were selected because of their relationships with infertility or other problems in swine. Spag7 was a natural choice for further examination as it codes an antisperm antibody that impairs the functions of sperm and reproduction capabilities are vital to maintenance of lines and value in industry. The reduced fertilization capacity of sperm in the presence of these spag7 antibodies are responsible for immunological infertility (Bohring 2001) and although results are insignificant (p-value $\leq 0.05$) in both JMP and REST analysis, expression tended to be higher in the 150 and HT animals, which is noteworthy despite the absence of recorded breeding issues with these animals.

Shippo1 or Outer dense fiber of sperm tails 3 (odf3) was examined because of its relationship to fertility as it is the dense fiber found in sperm tails. It helps with elastic recoil and protects against shearing. Transcription occurs in haploid germ cells and it is observed in the testis, as well as in the flagella of spermatids, and along the entirety of mature sperm. It is believed that shippo1 developed as a protective mechanism for the sperm rather than a causative agent for flagellar movement (Egydio de Carvalho 2002) but lack of this gene could cause a functional infertility. No significant differences were found between lines or age groups for shippo1 by any method of analysis.

Necdin (ndn) is expressed in post mitotic neurons and is important for muscle development. Necdin was also cited as a biomarker for maternal aggression in pigs resulting
in them attacking and killing their own offspring within twenty-four hours after birth (Claire 2007). Such a tendency would prove detrimental to a porcine line. People lacking the paternally derived \textit{ndn} allele develop Prader-Willi Syndrome, resulting in obesity, hypotonia, hypogonadism, and mental retardation (Jay 1997). Not only does necdin contribute to muscle development in neonates, it is also responsible for regeneration of damaged muscles in adult animals. Necdin null mice sustaining muscle injuries were unable to heal after muscle damage (Deponti 2007). The importance of muscle development and regeneration in any animal is without debate. While expression was generally higher in the LT line and the 180 day old animals, results lacked significance (p-value ≤ 0.05).

Examination of the larger picture shows that most of the up-regulation of gene expression in the 150 day old HT line represent increased cellular proliferation and growth rates and is consistent with the growth and pubertal maturation of boars of that life stage. To further examine the differences between lines at this pivotal age of maturation, additional network analysis was undertaken.

Approximately twenty-five networks of interest were reconstructed by GeneGo to show differences in gene expression between 150HT and 150LT animals from the microarray data (Table 4, Figure 6a.-e.). They were prioritized based on the number of canonical pathways they contain. Pathways were identified by the core genes involved and the GO processes they represent.

The network earning the highest Z-score is the network responsible for wound healing (16.3%), heme oxidation (4.7%), and fatty acid oxidation (7.0%). It is also involved
with blood coagulation and porphyrin catabolic processes (4.7%). Among the 22 experimental genes identified in this network are f8c (↑HT), mis12 (↑HT), tfpt (↑LT), clecsf2 (↑HT), and tom22 (↑LT) and the pathway was given the high Z-score of 72.58 (p-value 3.95e-48) (Figure 6a.). About half of the genes have higher expression in the HT line and half in the LT line. The identification of this pathway is a positive indication for these animals, as this pathway is critical for health maintenance and is especially important in wound healing. Because of its association with wound healing, testosterone has been investigated as a supplement to other treatments to improve healing in burn victims. Studies have shown that when a testosterone analog was used in addition to traditional burn treatments, 40% and 70% burn victims showed marked improvement over patients treated with traditional and placebo treatment. Patients receiving the testosterone analog showed significantly reduced weight and nitrogen loss and significantly increased wound healing. Recovery time was also reduced from thirteen days to nine days, without any observed negative side effects (Robert 2000).

Many of the recognized networks in the 150HT and 150LT lines contain genes highlighted through microarray validation. For example, the timp1(↑HT), ese3, hmg2(↑LT), mad2a (↑LT), and tram1(↑LT) network (Figure 6b.) was the network with the eighth highest Z-score (22.89) responsible for the GO processes of positive regulation of cellular process (49.0%), positive regulation of biological process (51.0%), regulation of molecular function (34.7%), regulation of catalytic activity (32.7%), and collagen catabolic process (10.2%) (p-value ≤5.13e-12). Timp1 was also present in the egf (↑HT), cd44(↑HT),
timp1(↑HT), p200rhogap(↑LT), hmg2(↑LT) network (Figure 6c.) (Z-score 19.57, p-value ≤4.75e-10), which is responsible for the positive regulation of cellular process (64.0%), positive regulation of biological process (66.0%), regulation of developmental process (54.0%), organ development (60.0%), and intracellular signaling cascade (54.0%) (Figure 6c).

Necdin is found in the ndn (↑LT), tfpt (↑LT), hmg2 (↑LT), appbp2 (↑LT), cks1 (↑LT) network (Z-score 19.57, p-value ≤4.75e-10) which is involved in regulation of apoptosis (42.0%), regulation of programmed cell death (42.0%), the negative regulation of biological process (56.0%), the positive regulation of biological process (56.0%), as well as the negative regulation of cellular process (52.0%) (Figure 6d.). The fact that this pathway was associated with the low testosterone line is not a surprise since ablation of testosterone to the testis has been shown to result in programmed cell death and studies have shown that replacement testosterone administration to hypophysectomized rats reduced DNA fragmentation by 75% in whole testis tissue and 64% in seminiferous tubules (Tapanainen 1993).

The biological pathways with the greatest differential expression between the 150HT and 150LT lines are highlighted in Table 5. These pathways included many that are regulatory in nature and could perhaps represent the mechanisms to homeostasis during the pubertal and growth changes occurring in this group of animals. These pathways are related to cellular adhesion and ECM remodeling (p-value ≤1.827E-3) in the HT line and cell cycle regulation (p-value ≤1.386E-3) and cAMP signal transduction (p-value ≤1.953E-3) in the
LT line which correspond to processes common during the maturation of boars at this age. The high association of the HT line with growth processes including EGFR signaling (p-value ≤ 2.735E-2) and leptin signaling (p-value ≤ 3.103E-2), provides additional evidence for selection of HT as a method for increasing growth rates. The association of leptin with high testosterone could be indicative of the relationship between leptin receptors on the surface of porcine sperm and fertility. Although the role of these receptors is not completely clear, they are believed to have a role in sperm motility (De Ambrogi 2007). The increased immune response pathways (p-value ≤ 1.5E-2) in this line may still be an issue of debate which may be further explained by the identified disease biomarkers in these animals.

Disease biomarkers were identified for 150HT and 150LT lines for a variety of diseases, with most gene markers solely up-regulated in the HT lines (Table 6). These HT diseases included arthritis (p-value ≤ 5.629E-8) and aortic aneurysms (p-value ≤ 5.652E-7). In addition, biomarkers were identified for several types of cancer like intraductal papilloma (p-value ≤ 8.028E-11), erythroblastic leukemia (p-value ≤ 1.374E-8), lymphangiomyoma (p-value ≤ 1.664E-8), and others; however, current research suggests that high levels of testosterone are inversely related to mortality from cancer and cardiovascular disease and that low testosterone is a biomarker for cancer, especially prostate cancer (Khaw 2007). While the high number of disease biomarkers identified provides grounds for further examination, it is possible that these genes are upregulated in the high testosterone animals simply due to increased growth rates. Biomarkers identified in the LT line include MERRF syndrome (p-value ≤ 2.132E-5), and cerebellar ataxia (p-value ≤ 1.2E-4). Yet, although
there has not been an increased incidence of tumors or health problems reported in these lines, it is unclear at what age these problems might develop, and whether animals in these lines are culled before this age.
Summary

Through this microarray investigation and the associated pathway analysis of the effects of testosterone, it has been determined that the genes upregulated in high testosterone animals during puberty are indeed links in such processes as growth regulation and metabolism, consistent with the increased growth rates previously associated with the high testosterone line. The disease biomarkers for disorders such as cancer identified within the HT line contrast sharply with the lack of disease incidence in the line and the existing evidence that numerous cancers and aortic diseases may accompany a decrease in testosterone, and lead us to suggest a further investigation of health effects before deeming that selective breeding for high testosterone may pose any health risks. Through this genomic examination and the implications of the biological pathways involved, the natural increase in testosterone gained through selective breeding can continue to be recommended as a management technique in an effort to increase growth rates and meat quality in the swine industry.
Figures

Figure 1a. Microarray pilot study experimental design with pooled samples at all time points. Numbers 1-20 represent array number while H1-H5 and L1-L5 correspond to pooled treatment group as shown in the legend to the right. Arrows between treatment groups represent the dye labeling of each array as the forward head of the arrow corresponds to Cy3 dye labeling and the back end of the arrow is Cy5. For example, the Array 1 consists of the Low Testosterone 1 day old treatment group labeled with Cy3 and the High Testosterone 1 day old treatment group labeled with Cy5.
Figure 1b. Microarray experimental design with 150 and 180 day individual animals. Numbers 1-20 represent array number while 150H, 150L, 180H, and 180L correspond to an individual animal in each treatment group as shown in the legend to the right. Arrows between treatments represent the dye labeling of each array as the forward head of the arrow corresponds to Cy3 dye labeling and the back end of the arrow is Cy5. For example, Array 1 consists of one 150 day High Testosterone animal sample labeled with Cy3 and one 150 day High Testosterone animal sample labeled with Cy5.
Figure 2a-d. Volcano plots of JMP analysis output of microarray line by age comparisons. Each gene is represented by a dot on the graph while the x-axis measures the fold change in expression between the two groups and the y-axis corresponds to the level of significance. 2a. Comparison of differential gene expression in 150 day and 180 day age groups regardless of testosterone level. 2b. Comparison of differential gene expression in the HT and LT lines ignoring age. 2c. Comparison of differential gene expression in the 150HT and 150LT treatment groups. 2d. Comparison of differential gene expression in the 180HT and 180LT treatment groups.
Table 1. REST Relative Expression analysis of 150 day animals with expression fold changes of HT line compared to LT. Genes are color coded based on the line with the greatest expression.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Reaction Efficiency</th>
<th>150HT:150LT expression</th>
<th>Std. Error</th>
<th>95% C.I.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIMP1</td>
<td>0.58</td>
<td>2.63</td>
<td>0.597 - 8.382</td>
<td>0.506 - 14.213</td>
<td>0.151</td>
</tr>
<tr>
<td>EGF*</td>
<td>0.76</td>
<td>6.91</td>
<td>1.153 - 50.233</td>
<td>0.126 - 85.344</td>
<td>0.043</td>
</tr>
<tr>
<td>SPAG7</td>
<td>0.68</td>
<td>1.29</td>
<td>0.709 - 2.244</td>
<td>0.465 - 2.964</td>
<td>0.375</td>
</tr>
<tr>
<td>ADFP*</td>
<td>0.73</td>
<td>2.88</td>
<td>1.733 - 6.753</td>
<td>0.730 - 9.401</td>
<td>0.02</td>
</tr>
<tr>
<td>SHIPPO1</td>
<td>0.75</td>
<td>1.86</td>
<td>1.145 - 3.179</td>
<td>0.566 - 4.627</td>
<td>0.062</td>
</tr>
<tr>
<td>TWISTNB</td>
<td>0.75</td>
<td>1.66</td>
<td>0.975 - 2.828</td>
<td>0.609 - 4.504</td>
<td>0.115</td>
</tr>
<tr>
<td>NDN</td>
<td>0.74</td>
<td>0.90</td>
<td>0.430 - 1.680</td>
<td>0.323 - 2.285</td>
<td>0.737</td>
</tr>
<tr>
<td>CD44</td>
<td>0.72</td>
<td>2.24</td>
<td>0.510 - 9.094</td>
<td>0.273 - 19.851</td>
<td>0.235</td>
</tr>
<tr>
<td>EEF1A1</td>
<td>0.63</td>
<td>1.51</td>
<td>1.009 - 2.141</td>
<td>0.882 - 2.980</td>
<td>0.07</td>
</tr>
<tr>
<td>RNASE12*</td>
<td>0.70</td>
<td>5.91</td>
<td>1.037 - 18.603</td>
<td>0.681 - 30.206</td>
<td>0.049</td>
</tr>
<tr>
<td>s100a6*</td>
<td>0.74</td>
<td>4.92</td>
<td>1.274 - 24.411</td>
<td>0.394 - 57.857</td>
<td>0.041</td>
</tr>
</tbody>
</table>

150HT
150LT
Table 2. REST Relative Expression analysis of 180 day animals with expression fold changes of HT line compared to LT. Genes are color coded based on the line with the greatest expression.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Reaction Efficiency</th>
<th>180HT:180LT Expression</th>
<th>Std. Error</th>
<th>95% C.I.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIMP1</td>
<td>0.58</td>
<td>0.59</td>
<td>0.122 - 1.502</td>
<td>0.095 - 7.453</td>
<td>0.34</td>
</tr>
<tr>
<td>EGF</td>
<td>0.76</td>
<td>1.03</td>
<td>0.383 - 10.216</td>
<td>0.281 - 21.146</td>
<td>0.987</td>
</tr>
<tr>
<td>SPAG7</td>
<td>0.68</td>
<td>0.81</td>
<td>0.234 - 1.778</td>
<td>0.157 - 2.181</td>
<td>0.754</td>
</tr>
<tr>
<td>ADFP</td>
<td>0.73</td>
<td>0.71</td>
<td>0.383 - 2.427</td>
<td>0.321 - 4.090</td>
<td>0.442</td>
</tr>
<tr>
<td>SHIPPO1</td>
<td>0.75</td>
<td>1.22</td>
<td>0.190 - 3.831</td>
<td>0.083 - 7.250</td>
<td>0.816</td>
</tr>
<tr>
<td>TWISTNB</td>
<td>0.75</td>
<td>1.21</td>
<td>0.282 - 2.626</td>
<td>0.208 - 3.555</td>
<td>0.69</td>
</tr>
<tr>
<td>NDN</td>
<td>0.74</td>
<td>0.83</td>
<td>0.518 - 1.306</td>
<td>0.372 - 2.095</td>
<td>0.478</td>
</tr>
<tr>
<td>CD44</td>
<td>0.72</td>
<td>0.53</td>
<td>0.114 - 1.053</td>
<td>0.093 - 6.913</td>
<td>0.221</td>
</tr>
<tr>
<td>EEF1A1</td>
<td>0.63</td>
<td>0.56</td>
<td>0.300 - 0.967</td>
<td>0.257 - 1.439</td>
<td>0.092</td>
</tr>
<tr>
<td>RNASEA</td>
<td>0.70</td>
<td>2.66</td>
<td>0.354 - 23.821</td>
<td>0.149 - 221.256</td>
<td>0.395</td>
</tr>
<tr>
<td>s100a6</td>
<td>0.74</td>
<td>0.86</td>
<td>0.292 - 3.334</td>
<td>0.113 - 11.824</td>
<td>0.773</td>
</tr>
</tbody>
</table>
Table 3. JMP ANOVA of Real-time RT-PCR results of genes selected for line by age comparisons of 150LT, 150HT, 180LT, and 180HT groups by expression differences, along with primer sequences. The treatment group with the greatest gene expression is indicated in each line by age comparison. Genes highlighted in red indicate overall higher expression in HT line, while genes highlighted in green represent higher expression in the LT line.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Name</th>
<th>Description</th>
<th>Primer sequence</th>
<th>180High-150High p-value</th>
<th>150Low-150High p-value2</th>
<th>180High-180Low p-value3</th>
<th>150Low-180Low p-value4</th>
<th>180Low-180Low p-value5</th>
</tr>
</thead>
<tbody>
<tr>
<td>timp1</td>
<td>Tissue Inhibitor of Metalloproteinase I</td>
<td>Regulatory component of remodeling processes; involved in morphogenesis, cancer, arthritis, changes, potentiation of lymphocytes and neural production in the testis and ovaries</td>
<td>F-5’-CTTCGCAACAGCATGTATGAG R-5’-GCTGCCACACACTCTGCAAG</td>
<td>0.0432*</td>
<td>0.0781</td>
<td>0.2388</td>
<td>0.3786</td>
<td></td>
</tr>
<tr>
<td>egf</td>
<td>External Growth Factor</td>
<td>Activation of cell division and growth.促销细胞周期的激活，促进细胞生长</td>
<td>F-5’-CACTAGAGTTATGAACTTACG R-5’-ATGACTGAGAATACCTTCA</td>
<td>0.0566</td>
<td>0.0243*</td>
<td>0.7349</td>
<td>0.8407</td>
<td></td>
</tr>
<tr>
<td>spag7</td>
<td>Adenylate Associated Nucleotidase-7</td>
<td>Role in the regulation of the membrane surface. Indicated at increased adenylyl cyclase</td>
<td>F-5’-GATTACACGCGACGTGACATG R-5’-ATGCCGAGATGGGTCTGCA</td>
<td>0.8219</td>
<td>0.2476</td>
<td>0.3989</td>
<td>0.8477</td>
<td></td>
</tr>
<tr>
<td>adfp</td>
<td>Adipophilin; Adipose-differentiation related protein</td>
<td>Major component of the adipose tissue, involved in the growth and differentiation of adipocytes</td>
<td>F-5’-CTCTGCAACAGCATGTATGAG R-5’-GCTGCCACACACTCTGCAAG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shipo1</td>
<td>DFRS; Outer Dense Fiber of Sperm Tail 3</td>
<td>Component of the sperm tail, involved in spermatogenesis, and protection of spermatozoa</td>
<td>F-5’-CCAGAAGGAACTCACCATTG R-5’-AGGACATCGTTGTAGATCATAG</td>
<td>0.7918</td>
<td>0.1358</td>
<td>0.9439</td>
<td>0.2172</td>
<td></td>
</tr>
<tr>
<td>twistnb</td>
<td>TWIST neighbor</td>
<td>Mammary gland development. Regulated in breast determination.</td>
<td>F-5’-CCGCTACCTTAACAGGAAACG R-5’-GCTCTCCCACAAACTTTGAC</td>
<td>0.6815</td>
<td>0.0744</td>
<td>0.2698</td>
<td>0.1377</td>
<td></td>
</tr>
<tr>
<td>ndn</td>
<td>Necdin</td>
<td>Intronless gene located in the Prader-Willi syndrome deletion region; important for respiratory development</td>
<td>F-5’-CACAATGGGTTGATAGGAAATTACG R-5’-ACCAGTGAGAGGAACAGTCG</td>
<td>0.8583</td>
<td>0.7663</td>
<td>0.2898</td>
<td>0.2417</td>
<td></td>
</tr>
<tr>
<td>cd44</td>
<td>CD44 molecule; Indian Blood Group</td>
<td>Protein is a cell-surface glycoprotein involved in cell-cell interactions, cell adhesion and migration, receptor for hyaluronic acid, interacts with various ligands, involved in lymphocyte adhesion, homing, and tumor metastasis</td>
<td>F-5’-CCAACACCTCCCAGTATGAC R-5’-CTTCAGGATTCGTTCTGTTATTCTC</td>
<td>0.0818</td>
<td>0.1283</td>
<td>0.1871</td>
<td>0.2770</td>
<td></td>
</tr>
<tr>
<td>ref1a1</td>
<td>S100 Calcium-binding Protein A6; Calcyclin; Prolactin Receptor-associated protein</td>
<td>Cell cycle progression and differentiation, stimulation of Ca2+-dependent insulin release, stimulation of prolactin secretion, and exocytosis, implicated in melanoma</td>
<td>F-5’-GAGGACACTGCTCAATTG R-5’-AGGTGAGATTGCAACTGGG</td>
<td>0.0081*</td>
<td>0.2496</td>
<td>0.0439*</td>
<td>0.7255</td>
<td></td>
</tr>
<tr>
<td>mas20</td>
<td>Macrophage A</td>
<td>Initiates the development of macrophage-like cells, dendritic cells, and neutrophils, involved in phagocytosis, and anti-inflammatory cytokines, and anti-viral response.</td>
<td>F-5’-TCTCTCCCTGCTCTGCCTG R-5’-GCTGCCACACACTCTGCAAG</td>
<td>0.0518</td>
<td>0.0428*</td>
<td>0.4198</td>
<td>0.4881</td>
<td></td>
</tr>
<tr>
<td>100a6</td>
<td>S100 Calcium-binding Protein A6; Calcyclin; Prolactin Receptor-associated protein</td>
<td>Cell cycle progression and differentiation, stimulation of Ca2+-dependent insulin release, stimulation of prolactin secretion, and exocytosis, implicated in melanoma</td>
<td>F-5’-GAGGACACTGCTCAATTG R-5’-AGGTGAGATTGCAACTGGG</td>
<td>0.1053</td>
<td>0.0223*</td>
<td>0.6176</td>
<td>0.2050</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3. Top ten rated GO Processes in each line by p-value based on GeneGo analysis. For comparison purposes, HT is displayed on the left and LT is on the right.
Figure 4. Top ten rated biological pathways in each line by p-value based on GeneGo analysis. For comparison purposes, HT is displayed on the left and LT is on the right.
Figure 5. Top ten rated disease biomarkers in each line by p-value based on GeneGo analysis. For comparison purposes, HT is displayed on the left and LT is on the right.
Table 4. Networks identified with gene expression from 150 day animals. Total nodes represent the total number of network objects while root nodes is the number of objects (genes) present in the network present in the 150 day animals. The Z-score shows the level of saturation of experimental genes in the network based on the number of total network objects. The higher the Z-score, the larger percentage of the experimental genes are involved in the network and the increased probability that the network is affected by the experiment.

<table>
<thead>
<tr>
<th>No</th>
<th>Network</th>
<th>GO Processes</th>
<th>Total nodes</th>
<th>Root nodes</th>
<th>p-Value</th>
<th>zScore</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F8C, MIS12, TPPT, CLECSP2, TOM22</td>
<td>wound healing (16.3%; 3.298e-06), heme oxidation (4.7%; 6.874e-06), fatty acid beta-oxidation (7.0%; 5.797e-05), blood coagulation, extrinsic pathway (4.7%; 9.944e-05), porphyrin catabolic process (4.7%; 9.944e-05)</td>
<td>50</td>
<td>22</td>
<td>3.95E-48</td>
<td>72.58</td>
</tr>
<tr>
<td>2</td>
<td>KRT, GPD1, BAf60A, STRA13, Ataxin-10</td>
<td>monosaccharide biosynthetic process (11.5%; 2.417e-05), posttranscriptional regulation of gene expression (19.2%; 2.817e-05), alcohol biosynthetic process (11.5%; 4.446e-05), cellular catabolic process (34.6%; 7.971e-05), negative regulation of survival gene product expression (7.7%; 8.588e-05)</td>
<td>50</td>
<td>11</td>
<td>3.51E-23</td>
<td>46.04</td>
</tr>
<tr>
<td>3</td>
<td>PKC-delta, ESE3, MAD2a, HIP17, ZFP36L2</td>
<td>regulation of molecular function (50.0%; 4.483e-16), positive regulation of cellular process (62.5%; 1.353e-15), positive regulation of developmental process (43.8%; 5.446e-10), regulation of apoptosis (45.8%; 1.215e-14), regulation of programmed cell death (45.8%; 1.547e-14)</td>
<td>50</td>
<td>11</td>
<td>9.12E-21</td>
<td>36.9</td>
</tr>
<tr>
<td>4</td>
<td>MAD2a, APPBP2, MIS12, BAf60A, Podocalyxin-like 1</td>
<td>cell division (25.6%; 2.270e-10), mitosis (23.3%; 4.027e-10), M phase of mitotic cell cycle (23.3%; 4.935e-10), regulation of mitotic cell cycle (20.9%; 5.029e-10), mitotic cell cycle (27.9%; 9.995e-10)</td>
<td>50</td>
<td>9</td>
<td>3.02E-16</td>
<td>29.82</td>
</tr>
<tr>
<td>5</td>
<td>PKC-delta, MUC1, EGF, CD44 (ICD), Podocalyxin-like 1</td>
<td>organ development (69.7%; 1.706e-13), response to organic substance (39.4%; 6.660e-12), response to abiotic stimulus (39.4%; 1.055e-11), regulation of apoptosis (48.5%; 1.770e-11), regulation of programmed cell death (48.5%; 1.547e-11)</td>
<td>41</td>
<td>8</td>
<td>5.39E-13</td>
<td>26.69</td>
</tr>
<tr>
<td>6</td>
<td>F8C, Adipophilin, Podocalyxin-like 1, PCM121, Ataxin-10</td>
<td>regulation of biological quality (48.1%; 2.331e-06), regulation of neuronal synaptic plasticity (14.8%; 2.817e-05), regulation of localization (29.6%; 1.672e-05), regulation of synaptic transmission (18.5%; 1.892e-05), regulation of transport (25.9%; 1.933e-05)</td>
<td>50</td>
<td>7</td>
<td>6.60E-13</td>
<td>26.33</td>
</tr>
<tr>
<td>7</td>
<td>PTS, 6-(L-erythro-1,2-dihydroxypropyl-3-methyl phosphate)-7,8-dihydropterin + polyphosphate, L-ascorbic acid + polyphosphate = L-ascorbic acid 2-phosphate + polyphosphate, 6-pyruvoyl-5,6,7,8-tetrahydrobiopterin biosynthetic process (100.0%; 3.647e-04), tetrahydrobiopterin metabolic process (100.0%; 4.255e-04), L-phenylalanine catabolic process (100.0%; 5.471e-04), L-phenylalanine metabolic process (100.0%; 6.686e-04), ptomaine and derivative biosynthetic process (100.0%; 7.295e-04)</td>
<td>6</td>
<td>1</td>
<td>1.82E-03</td>
<td>23.37</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>TRAP1, ESE3, HM32, MAD2a, TRAM1</td>
<td>positive regulation of cellular process (49.0%; 1.071e-09), positive regulation of biological process (51.0%; 1.293e-09), regulation of molecular function (34.7%; 9.474e-09), regulation of catalytic activity (32.7%; 1.157e-08), collagen catabolic process (10.2%; 1.193e-08)</td>
<td>50</td>
<td>7</td>
<td>5.13E-12</td>
<td>22.89</td>
</tr>
<tr>
<td>9</td>
<td>BAf60A, SLC34A1, FUMH, Keratin 14, CGRP-RCP</td>
<td>anion transport (23.8%; 2.756e-06), inorganic anion transport (19.9%; 2.593e-05), DNA fragmentation during apoptosis (9.5%; 2.619e-04), cell structure disassembly during apoptosis (9.5%; 4.212e-04), homeostasis of number of cells (14.3%; 5.909e-04)</td>
<td>50</td>
<td>5</td>
<td>4.91E-09</td>
<td>19.83</td>
</tr>
</tbody>
</table>
Figure 6a. F8c, mis12, tfpt, clecf2, tom22 network. The top scored (by the number of pathways) network from the 150HT vs. 150LT analysis. Genes upregulated in the HT line are marked with red circles; LT with blue circles. Refer to figure 4e. for GeneGo network object key.
Figure 6b. *Timp1, ese3, hmg2, mad2a, tran1* network. The eighth scored (by the number of pathways) network from the 150HT vs. 150LT analysis. Genes upregulated in the HT line are marked with red circles; LT with blue circles. Refer to figure 4e. for GeneGo network object key.
Figure 6c. *Egf, cd44, timp1, p200rhogap, hmg2* network. The tenth scored (by the number of pathways) network from the 150HT vs. 150LT analysis. Genes upregulated in the HT line are marked with red circles; LT with blue circles. Refer to figure 4e. for GeneGo network object key.
Figure 6d. *Ndn*, *tfpt*, *hmg2*, *appbp2*, *cks1* network. The sixteenth scored (by the number of pathways) network from the 150HT vs. 150LT analysis. Genes upregulated in the HT line are marked with red circles; LT with blue circles. Refer to figure 4e. for GeneGo network object key.
Figure 6e. Metacore™ GeneGo Network object key.
Table 5. Biological pathways represented in gene expression in 150 day HT and LT animals along with associated p-values.

<table>
<thead>
<tr>
<th>Top GeneGo Pathway Maps in 150HT</th>
<th>pValue</th>
<th>Top GeneGo Pathway Maps in 150LT</th>
<th>pValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell adhesion_Cell-matrix glycoconjugates</td>
<td>9.775e-04</td>
<td>Cell cycle_Role of APC in cell cycle regulation</td>
<td>1.386e-03</td>
</tr>
<tr>
<td>Cell adhesion_ECM remodeling</td>
<td>1.827e-03</td>
<td>Signal transduction_cAMP signaling</td>
<td>1.853e-03</td>
</tr>
<tr>
<td>Immune response_Antigen presentation by MHC class II</td>
<td>1.500e-02</td>
<td>Regulation of lipid metabolism_Regulation of lipid metabolism by niacin and scopolamine</td>
<td>2.363e-02</td>
</tr>
<tr>
<td>Immune response_Sialic-acid receptors (Siglecs) signaling</td>
<td>1.500e-02</td>
<td>Transcription_Transcription factor Tubby signaling pathways</td>
<td>2.955e-02</td>
</tr>
<tr>
<td>Immune response_Oncostatin M signaling via JAK-Stat in mouse cells</td>
<td>2.243e-02</td>
<td>Protein folding_Membrane trafficking and signal transduction of G-alpha (i) heterotrimeric G-protein</td>
<td>3.297e-02</td>
</tr>
<tr>
<td>Immune response_Oncostatin M signaling via JAK-Stat in human cells</td>
<td>2.488e-02</td>
<td>Development_Delta- and kappa-type opioid receptors signaling via beta-arrestin</td>
<td>3.976e-02</td>
</tr>
<tr>
<td>Development_EGFR signaling via PIP3</td>
<td>2.735e-02</td>
<td>Transcription_Transcription regulation of aminoacid metabolism</td>
<td>4.318e-02</td>
</tr>
<tr>
<td>Transport_RAB5A regulation pathway</td>
<td>3.103e-02</td>
<td>Neurophysiological process_Dopamine D2 receptor transactivation of PDGFR in CNS</td>
<td>4.879e-02</td>
</tr>
<tr>
<td>Development_Leptin signaling via JAK/STAT and MAPK cascades</td>
<td>3.103e-02</td>
<td>Neurophysiological process_Role of CCK5 in presynaptic signaling</td>
<td>4.834e-02</td>
</tr>
<tr>
<td>Immune response_IL-10 signaling pathway</td>
<td>3.255e-02</td>
<td>Apoptosis and survival_p53-dependent apoptosis</td>
<td>4.993e-02</td>
</tr>
</tbody>
</table>
Table 6. Disease biomarkers identified in 150 day HT and LT animals along with corresponding p-values.

<table>
<thead>
<tr>
<th>Top GeneGo Diseases in 150HT (by Biomarkers)</th>
<th>Top GeneGo Diseases in 150LT (by Biomarkers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>name</td>
<td>pValue</td>
</tr>
<tr>
<td>Papilloma, Intraductal</td>
<td>8.028e-11</td>
</tr>
<tr>
<td>Leukemia, Erythroblastic, Acute</td>
<td>1.374e-08</td>
</tr>
<tr>
<td>Lymphangiomyoma</td>
<td>1.664e-08</td>
</tr>
<tr>
<td>Smooth Muscle Tumor</td>
<td>1.664e-08</td>
</tr>
<tr>
<td>Lymphangiomyomatosis</td>
<td>1.664e-08</td>
</tr>
<tr>
<td>Lymphatic Vessel Tumors</td>
<td>2.609e-08</td>
</tr>
<tr>
<td>Arthritis, Experimental</td>
<td>5.629e-08</td>
</tr>
<tr>
<td>Keratoacanthoma</td>
<td>1.423e-07</td>
</tr>
<tr>
<td>Leukemia, Monocytic, Acute</td>
<td>3.756e-07</td>
</tr>
<tr>
<td>Aortic Aneurysm, Abdominal</td>
<td>5.652e-07</td>
</tr>
</tbody>
</table>
References Cited


Isolation and Extraction of Porcine Spermatozoa Ribonucleic Acid for Genetic Applications and Analysis

Abstract

Within the tiny membranes of the vast paternal contribution of spermatozoa, lies at least half of the genetic blueprint to future generations as well as a showcase of the past events of spermatogenesis. In swine, there are many fertility and growth concerns that would benefit from a dependable and efficient procedure to isolate ribonucleic acid from spermatozoa such as sperm quality, parturition rates, litter sizes, and animal health and market quality. The discovery of the transcriptome available for manipulation could provide great advances for the ability to generate fertility profiles and alter the genetics of a herd. While protocols exist in many other species for the isolation of RNA from spermatozoa, Sus scrofa spermatozoa represent a more difficult case, due to its fragile nature, and special considerations must be made to ensure successful extraction. Here, we present a novel method of porcine spermatozoa RNA isolation, from the point of collection from the donor boar to validation of the sperm-specific genetic transcripts it contains.
Introduction

The discovery that spermatozoa contain RNA in addition to DNA has prompted many to investigate its potential role in developing sperm and the post-zygotic contribution of the male genome. The initial detection of RNA in sperm came in 1988 in the sperm nucleus of the fern *Scolopendrium* (Rejon 1988). Next, RNA was visualized in rat spermatozoa using a RNase-colloidal gold procedure (Pessot 1989). It was not until 1993 that specific RNAs were identified in the sperm of the pollen grain and these RNAs were U1 and U2 snRNAs (Concha 1995). Similarly, the first mRNAs were identified in human sperm cells in 1993 and these were the proto-oncogene *c-myc* (Kumar 1993). Reverse transcription PCR became a valuable tool used first used in the detection of sperm RNA in 1994 with its detection of human leukocyte antigens (Chiang 1994).

Sperm not only contribute DNA to the ovum; upon fertilization they deliver many novel RNAs as well (Krawetz 2005). They also release signaling molecules and transcription factors. Sperm RNA represent half of the RNA found in the testis and sperm-specific RNAs have been found in the fertilized oocyte (Ostermeier 2004). Using hamster sperm penetration assays (Johnson 1995), several genes were identified in sperm that were absent in human or mouse oocyte SAGE (Serial Analysis of Gene Expression (Velculescu 1995)) or cDNA libraries and were not found in unfertilized oocytes (Ostermeier 2004). One gene of interest discovered in this method was *foxl1b*, a transcription factor vital to embryo patterning only found in the adult testis or the developing fetal brain (Murphy
Another gene found in spermatozoa, *wnt5a*, is a proto-oncogenic signaling molecule that is involved in cellular differentiation and morphology development (Moon 1997). Interestingly, short half lives of most of the RNAs discovered indicate that they may be important in embryogenesis (Krawetz 2005).

Many genes are expressed exclusively in the developing male gamete. Protamine genes, *prm1* and *prm2*, are expressed in spermatids and are under transcriptional and translational control (Kleene 2000). It has been shown that protein repressors bind to the poly-A tails or specific RNA sequences in the 3′-UTR to cause translational regulation of these genes (Steger 1999). The protamine mRNAs are transcribed in the round spermatid and then stored as ribonucleoprotein complexes until they are translated in the elongating spermatid (Sinclair 1982). It is has been suggested that alternative sperm chromatin structure is intended for imprinting purposes or that it is to help with the repackaging and activation of the male contribution (Kramer 1997). In mice it was shown that premature expression of these protamines stopped spermatid differentiation (Lee 1995). In humans, differential display was used to find that *β-actin* and *prm2* mRNAs were present in normal semen while azoospermic semen lacked *prm2* and only had *β-actin*. This discovery led the investigators to hypothesize that RNA fingerprinting could be used as indicators of male fertility (Miller 1994).

Transcriptional profiling of the spermatozoa of fertile and infertile men can give insight into fertility issues. In the most comprehensive study of sperm RNA transcripts, 2,780 non-poly(A)-enriched transcripts from one individual were compared to a pool of
3,281 poly(A)-enriched transcripts from nine other men. The study found that all but four transcripts from the individual were found in the pool. The pool of semen transcripts was also compared to a pool of testes transcripts and all of the semen transcripts were found in the testes pool. This suggests that the gene expression of the sperm can serve as a window to past events in spermatogenesis in the testis (Ostermeier 2002).

Additionally, this discovery represents the ability to create a fertility profile from the genes expressed in sperm. In a follow-up study, semen gene expression of men with ideal, medium, and poor quality semen samples were compared to find gene markers of fertility. The study revealed that the amount of variation among normal fertile men is very low while the variation in gene expression in non-fertile men is high suggesting the creation of a transcriptional profile as a predictor of good fertility rather than of infertility, which comes with much more variability (Ostermeier 2005).

The purpose of this study is to create a protocol that efficiently and reliably isolates RNA from whole porcine sperm in quantities large enough for genomic applications in fertility studies. The RNA should be of high quality with the capability of cDNA synthesis and sperm-specific transcript amplification. To date, this is the first protocol in swine to meet these requirements, offering a high quantity, high quality RNA able to amplify the \textit{prm1} transcript.
Materials and Methods

Semen Collection

Semen samples were collected once weekly for three weeks from three cross-bred boars approximately two years old, housed according to industry standards. Collection was made through the gloved hand technique (Hancock 1959), by trained technicians. Volume, concentration, and motility estimates were obtained immediately after collection and prior to storage of samples at 37°C until processing.
Isolation of Viable Sperm from Whole Semen

Processing of samples included subjecting whole semen to a 70/30% Percoll gradient to remove dead sperm and cellular contaminants. Gradients were diluted using Beltsville Thawing Solution (BTS) resuspended according to manufacturer’s instructions. Freshly collected samples were layered in a 15 ml conical tube with 2 ml 70% Percoll, 2ml 30% Percoll, and 1ml whole semen, followed by centrifugation for 30 minutes at 1500 RPM in a fixed angle centrifuge at room temperature.

Following centrifugation, the first and second upper layers containing 1 ml of seminal plasma, and 2ml of the dead spermatozoa and somatic cells, respectively, were removed and discarded. The concentration of viable spermatozoa in the 2ml of the final layer was ascertained using a hemocytometer. These samples were then aliquotted into 3x10^8 cells into 1.5ml microcentrifuge tubes and centrifuged at 2,000 RPM at 4°C for 15 minutes. Supernatants were discarded and the clean spermatozoa were flash frozen in liquid nitrogen. Storage was at -80°C until RNA extraction was performed.
Large Scale RNA Extraction Protocol

It was found that each RNA extraction in a 50ml conical tube could accommodate a maximum of $9 \times 10^8$ cells. One to two volumes of Trizol (Invitrogen) was added to each of three microcentrifuge tubes containing the $3 \times 10^8$ sperm cells immediately after removal from -80°C. Samples were vortexed and washed repeatedly with Trizol until all were transferred to the large conical tube. Samples remaining in the microcentrifuge tubes were repeatedly washed with Trizol and vortexed until all sample was transferred into the conical. The entire sample was then vortexed for approximately 15 seconds, following a thirty minute incubation at 60°C, with vortexing every ten minutes as previously described (Lalancette 2008).

After centrifugation at 12,000g for fifteen minutes at 4°C, the supernatant was removed and placed in a new tube. In accordance with the Trizol manufacturer’s instructions, 200ul chloroform per each 1ml Trizol was added to the sample; samples were shaken for 15 seconds and incubated at room temperature for 3 minutes. Following a 15 minute centrifugation at 12,000g at 4°C, the aqueous upper phase was transferred to a new 50 ml conical tube.

Ethanol precipitation was performed by adding ammonium acetate to the sample to 0.5M. Linear acrylamide (Ambion) was used as an RNA carrier and added at approximately 10ug/ml with vortexing to ensure mixing. After this step, two volumes of 100% ethanol were added and the sample was chilled at -20°C for 15 minutes followed by a 20 minute
centrifugation at 10,500g the supernatant was carefully removed and discarded. Samples were resuspended in 300ul nuclease-free water and then DNAse I treatment (Ambion) was performed according to manufacturer’s instructions by adding DNase I buffer to a 1X concentration and 2 units of DNase I. After incubation at 37°C for 30 minutes, DNase I was inactivated by the addition of EDTA to a concentration of 5mM and heating samples to 75°C for 10 minutes.

One final precipitation was performed by adding ammonium acetate to a final concentration of 2.5M and once again adding linear acrylamide to samples at 10ug/ml mixing well. Two and a half volumes of 100% ethanol were added to samples before they were chilled at -20°C for 35 minutes and then centrifuged at maximum speed for 15 minutes. Supernatants were removed and discarded and the pellet was washed with cold 75% ethanol to remove residual salts. Samples were resuspended in 20ul nuclease-free water and stored at -80°C.
RNA Quantification

Because of the small quantity of RNA in sperm samples, quantification was completed using Ribogreen Reagent (Molecular Probes) in the Thermo Labsystems Fluroscan Ascent FL plate reader, with an excitation phase at 485nm and an emission phase at 538nm. Sample values were derived with comparison of a standard curve according to manufacturer’s instructions.
RNA Extraction Validation

Iscript select cDNA synthesis

cDNA synthesis was completed using 1.5ng of RNA with the iSCRIPT Select cDNA synthesis kit (BioRad) utilizing the RNase H+ iScript reverse transcriptase with an Oligo(dT)\textsubscript{20} primer according to manufacturer’s instructions. The program protocol consists of 5 minutes at 25°C, followed by 30 minutes at 42°C, 5 minutes at 85°C, and a 4°C hold.
RT-PCR Platinum Taq protocol

To validate that the material extracted from the semen samples was indeed RNA, PCR primers were designed for the *Sus scrofa* protamine 1 (*prm1*) gene, a gene known only to be present in porcine sperm (Maier 1988). The primers, Forward

5’CAGACGAAGGAGGAGATGTTG3’ and Reverse 5’TCAAGATGCGGCGAGG3’, were designed using Beacon Design software (Premier Biosoft, CA, USA) to span an intron/exon boundary and sequence validated for *prm1*. The Platinum Taq Polymerase High Fidelity Kit (Invitrogen) was used in RT-PCR to amplify the target gene from 1ul of the undiluted cDNA template. RT-PCR proceeded at 95°C for two minutes followed by twenty-seven cycles of thirty seconds at 95°C, one minute at 55°C, and one minute at 70°C, followed by a final five minutes at 70°C. The target was visualized on a 1.2% agarose gel (Figure 3).
Results and Discussion

This extraction protocol yields approximately 362ng of RNA per $9 \times 10^8$ cells (Figure 2) equivalent to approximately 0.0004pg/cell. In addition, evidence suggests that the genetic material isolated is in fact RNA, as it was effectively used to synthesize cDNA, and amplify the $prm1$ transcript by RT-PCR. The resulting 1.2% agarose gel showed that this procedure was able to amplify bands corresponding to 234bp and 334bp, respectively representative of the porcine $prm1$ cDNA and genomic DNA fragments (Figure 3).

While spermatozoa prove to be a difficult material for RNA extraction, porcine spermatozoa bring even more complexity. Despite our lab’s attempts to replicate spermatozoa protocols utilized in other species from many other laboratories (Goodwin 2000; Ostermeier 2002; Thompson 2003; Wasilk 2004; Zhang 2006; Goodrich 2007), this protocol, derived loosely from several others, was the only method that produced RNA capable of amplifying the $prm1$ gene. The amount of RNA extracted from each spermatozoa cell is consistent with other reported values.

Several problems arise when extracting RNA from semen. First, boar spermatozoa is, in itself, fragile. It must be kept fresh at $37^\circ$C or as extended semen near $17^\circ$C in order to ensure stability. It is very important that porcine semen be maintained at the correct temperature. Secondly, it is not uncommon to have malformed or degenerate sperm present in a normal boar ejaculate. In fact, in the normal boar ejaculate, only 40% of sperm is considered viable (Moohan 1995). There are also somatic cells present in the ejaculate. To
ensure that only viable sperm are selected for RNA extraction, a discontinuous Percoll gradient was chosen, which results in three to four times greater yield than the swim up method (Berger 1985). The swim up method is also another option to procure viable sperm used in conjunction with somatic cell lysis buffer to remove somatic cells but it was found that the Percoll gradient was most effective and efficient for our purposes since it is more likely to produce viable sperm capable of binding to the zona pellucida and fertilizing the oocyte (Moohan 1995) and it was consistently more reliable. In addition, the proportion of RNA to total cell nucleic acids and other components is very small. This facilitates the need for linear acrylamide as a carrier. Furthermore, the final quantity of RNA extracted is small, however Ribogreen Reagent (Molecular Probes) was found to be an effective means of measurement since it can accurately quantitate very small concentrations.

Another common obstacle is that methods intended for the extraction of RNA often yield DNA as well. Despite the DNase treatment included in this protocol, there was still a small amount of genomic DNA amplification as seen in the RT-PCR results (Figure 3). Additional DNase treatment may be necessary for applications sensitive to DNA contamination. This could be achieved through replication of the DNase treatment described in this protocol or an alternate treatment before amplification by RT-PCR.

The development of a successful porcine sperm extraction protocol is useful to identify genes present in the boar transcriptome. One protocol isolated the leptin receptor \textit{ob-r} in porcine semen to highlight the possible role of leptin in sperm formation and motility (De Ambrogi 2007). There have also been advances in other species that have led to
transcriptional profiling as an indicator of fertility, such as the human fertility profiles of fertile and nonfertile men (Ostermeier 2002), which could prove useful in porcine reproduction. Even in other livestock species, transcriptional analysis of such animals as the bull (Lalancette 2008), has led to the discovery of novel transcripts that may one day shed light on pathways to genetic improvement. Such an analysis of porcine sperm would undoubtedly contribute to improvements in swine genetics, particularly because traditional sperm evaluation parameters were found to have little correlation with actual fertilizing capabilities (Gadea 2004). This protocol because of its ability to isolate RNA at large quantities sufficient for large scale procedures offers a method to begin the type of fertility studies that would benefit swine genetics. Despite similar protocols in this and other species, it is the first to offer amplification of a sperm-specific gene transcript with primers spanning an intron and exon boundary to differentiate DNA contamination. It is our hope that this protocol will advance the understanding of swine and mammalian transcriptomes in order to propagate a greater utilization of genomics in porcine reproduction.
### Figures

<table>
<thead>
<tr>
<th>Boar</th>
<th>Date of Collection</th>
<th>Volume (ml)</th>
<th>Concentration</th>
<th>Motility</th>
</tr>
</thead>
<tbody>
<tr>
<td>00909</td>
<td>08/05/2008</td>
<td>204</td>
<td>338x10⁶</td>
<td>Superior</td>
</tr>
<tr>
<td>00909</td>
<td>08/12/2008</td>
<td>150</td>
<td>298x10⁶</td>
<td>Superior</td>
</tr>
<tr>
<td>00909</td>
<td>08/19/2008</td>
<td>249</td>
<td>276x10⁶</td>
<td>Superior</td>
</tr>
<tr>
<td>00913</td>
<td>08/05/2008</td>
<td>223</td>
<td>315x10⁶</td>
<td>Superior</td>
</tr>
<tr>
<td>00913</td>
<td>08/12/2008</td>
<td>273</td>
<td>301x10⁶</td>
<td>Superior</td>
</tr>
<tr>
<td>00913</td>
<td>08/19/2008</td>
<td>369</td>
<td>207x10⁶</td>
<td>Superior</td>
</tr>
<tr>
<td>28809</td>
<td>08/05/2008</td>
<td>197</td>
<td>254x10⁶</td>
<td>Superior</td>
</tr>
<tr>
<td>28809</td>
<td>08/12/2008</td>
<td>176</td>
<td>290x10⁶</td>
<td>Superior</td>
</tr>
<tr>
<td>28809</td>
<td>08/19/2008</td>
<td>172</td>
<td>263x10⁶</td>
<td>Superior</td>
</tr>
</tbody>
</table>

Figure 1. Semen Collection of Boars by ID. Volume and concentration recorded immediately after collection. Qualitative motility determined visually by technician.
Figure 2. RNA yield (ng) as measured by Ribogreen assay per $9 \times 10^6$ cells.
Figure 3. RT-PCR with 100kb ladder and visible cDNA and genomic DNA bands at 234bp and 334bp; Lane 1: Testis prm1 cDNA sample, Lane 2: Semen prm1 cDNA sample, Lane 3: Semen genomic DNA sample. Genomic DNA band also visible in cDNA samples.
References Cited


Concha, II (1995). "U1 snRNP components are present in the vegetative and generative nuclei of the pollen grain." Sexual plant reproduction 8(6).


Moohan, J. M., Lindsay, Kevin S. (1995). "Spermatozoa selected by a discontinuous Percoll density gradient exhibit better motion characteristics, more hyperactivation, and longer survival than direct swim-up." Fertility and Sterility 64(1): 160-165.


Project Summary and Future Directions

This research endeavor has not only provided a comprehensive look at many aspects of swine reproduction but has also answered some of the technical questions associated with applying the scholarly principles of genetics. We have examined the practical aspects of creating separate lines of swine by selectively breeding animals over many generations and what that means to the actual gene expression in individual animals. Animals selectively breed for high and low testosterone were initially bred to showcase the heritabilities associated with testosterone but also were shown to have vast differences in growth and litter sizes. This project took that evaluation many steps further through microarray analysis, quantitative RT-PCR, and genomic pathway analysis to show the gene expression associated with high and low testosterone is indicative of various metabolism pathways. The increased level of up-regulated gene expression in the 150 day HT animals showcases the high level of biological activity in those animals as they near puberty with advanced growth rates. The identification of the cellular proliferation and immune response pathways identified in this group may be explanatory of this principle, along with the detection of the numerous cancer biomarkers. While the identification of these biomarkers is concerning, the absence of disease reported in these lines thus far provides reason to believe that these biomarkers are merely up-regulated due to increased cellular proliferation associated with growth or that these animals are not maintained to an age sufficient to produce these health problems. Future directions could include additional genomic studies to examine the gene
expression of older animals. This could not only determine future health problems potentially associated with older adult animals but could also be useful in translational medicine to show testosterone health risks to humans. Also, the follow-up RT-qPCR analysis of additional genes shown to be differentially expressed in the microarray study could be useful in identifying specific genes associated with swine reproduction like those associated with fertility or disease resistance.

The development of the ribonucleic acid extraction protocol from porcine sperm marked not only a personal achievement but also one for the swine community. While other protocols have claimed isolation of RNA from pig semen, this protocol has proved to be viable through its reproducibility and the ability to amplify the \textit{prml} transcript from cDNA created from the RNA. The invention of this protocol provides a valuable starting point for future studies in swine reproduction. With large amounts of quality RNA comes the ability to perform many additional genomic studies. In other species such studies were in the form of differential SAGE libraries or fertility profiling utilizing microarrays. The swine industry could benefit from any of these approaches as well as smaller scale RT-PCR studies with additional fertility genes previously highlighted in other species. A reliable RNA extraction procedure is key to ensuring future study in porcine gene expression in semen transcripts which hold the male contribution in an industry dependent on fertility.

The completion of this M.S. thesis project is offered as one step forward in the advancement of swine genomics as well as an investment in education and the desire to learn all that affects our world.