LOCALIZING GENETIC ELEMENTS ASSOCIATED WITH REPRODUCTIVE PERFORMANCE IN MINK USING A SEQUENCE-BASED SNP PANEL

by

Chaini Konwar

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia

December 2014

© Copyright by Chaini Konwar, 2014

TABLE OF CONTENTS

LIST OF FIGURES.......iv
LIST OF TABLES.......v
ABSTRACT.......vi
LIST OF ABBREVIATIONS USED.......vii
ACKNOWLEDGEMENTS.......viii
CHAPTER 1: INTRODUCTION.......1
CHAPTER 2: LITERATURE REVIEW.......3
2.1 Mink reproductive physiology.......3
2.1.1 Physiological factors of reproductive performance in female mink.......3
2.1.1.1 Follicle development and ovulation.......4
2.1.1.2 Fertilization and implantation.......5
2.1.1.3 Whelping, lactation and weaning.......7
2.1.2 Effect of feeding intensity on reproductive performance in female mink.......8
2.1.2.1 Obesity and its consequences.......9
2.1.2.2 Obesity and reproductive performance.......10
2.2 From linkage maps to association studies.......12
2.3 Genetic association studies.......14
2.3.1 Genome-wide association (GWA) studies.......14
2.3.2 Candidate gene approach.......15
2.3.2.1 Selection of candidate genes.......16
2.3.2.2 Prioritizing SNPs for genotyping.......18
2.3.2.3 Attributes of samples and markers' integrity.......20
2.3.2.4 Study designs for association analysis.......21
2.3.2.5 Tests for association analysis.......22
2.3.2.6 Validation studies.......23
2.3.2.7 Farm animal genomics.......25
CHAPTER 3: OBJECTIVES AND HYPOTHESIS.......27
CHAPTER 4: MATERIALS AND METHODS.......28
4.1 Resource population.......28
4.2 Response variables.......29
4.3 Blood sampling.......29
4.4 Laboratory Analyses.......30
4.4.1 DNA extraction.......30

4.4.2 Quality check 31
4.5 Reference genome and SNP discovery 32
4.6 Candidate gene SNP panel 35
4.6.1 Gene categories based on signalling pathways 35
4.6.1 Genotyping 38
4.6.1.1 Marker selection 38
4.7 Association analysis 40
CHAPTER 5: RESULTS 43
5.1.1 MDR group 43
5.1.2 CTRL group 46
5.2 Window analysis 48
5.2.1 MDR group 48
5.2.2 CTRL group 50
5.5 Summary of tail and window analysis 52
5.6 Reproductive performance of sister mink 53
5.7 SNPs distribution in the significant genes 54
5.8 Linkage disequilibrium (LD) analysis 55
CHAPTER 6: DISCUSSION 57
6.1 Approaches for detecting association between performance and genetic elements 57
6.1.1 Tail analysis 57
6.1.2 Window analysis 58
6.2 Comparison between MDR and CTRL groups 60
6.3 Model 1 61
6.3.1 Folliculogenesis and ovulation 63
6.3.2 Ovarian tissue remodelling and angiogenesis 64
6.3.3 Oviductal transport of embryos 65
6.3.4 Establishment and maintenance of pregnancy 65
6.3.5 Embryonic diapause and embryo reactivation 67
6.3.6 Lactation and maternal behaviour 67
6.4 Genes associated with reproduction in other (livestock) species 68
CHAPTER 7: CONCLUSION 70
REFERENCES 71
APPENDIX 1 89
APPENDIX 2 105

LIST OF FIGURES

Figure 1	Flowchart outlining the strategies for identification of putative candidate genes	17
Figure 2	Flowchart outlining the steps for prioritizing SNPs for genotyping	19
Figure 3	Flowchart outlining the steps for identification of SNPs for the candidate marker panel	34
Figure 4	Categories of the candidate genes in the marker panel based on their inferred signalling pathways	36
Figure 5	Flowchart outlining the various stages of the project	42
Figure 6	Graph of the reproductive performance of sister pairs in CTRL and MDR groups	53
Figure 7	Virtual mapping of mink co-ordinates with A3 gene in *Canis lupus*	54
Figure 8	Representation of the proposed mechanisms by which circadian rhythm, PPAR and MAPK pathways influence the critical reproductive processes in female mink. The model also shows the functional roles of the C4, MSX2, NR5A2, O1, SPARC and T1 genes in mammalian fertility and are highlighted in green	62

LIST OF TABLES

Table 1	Primers used for generating mink-specific GAPDH sequences	31
Table 2	Sequential steps of thermal cycling conditions for RT-PCR	32
Table 3	Candidate marker panel and primary quality control	39
Table 4	Phenotypic performance of females selected for tail analysis in CTRL and MDR	43
Table 5	SNPAssoc results for the tail analysis in the MDR group	44
Table 6	SNPAssoc results for the tail analysis in the CTRL group	46
Table 7	Phenotypic performance of females selected for window analysis in CTRL and MDR	48
Table 8	SNPAssoc results for the window analysis in the MDR group	49
Table 9	SNPAssoc results for the window analysis in the MDR group	51
Table 10	Overlap of significant genes in tail and window analysis	52

ABSTRACT

Reproductive performance is a multifactorial trait which is affected by genetic, physiological and environmental factors. This study aimed at identification of genetic elements associated with fertility in mink females using the candidate gene approach, and was performed in a population of Standard Black mink maintained between 2009 and 2013. Based on dietary treatment, this population was analysed independently as two categories: Control (CTRL) and Moderate Diet Restriction (MDR). Irrespective of the analytical approaches used, twelve genes in the MDR group and ten genes in the CTRL group showed significant associations with fertility, at probabilities of less than or equal to 0.05. The results of this research allowed a better understanding of the role of genetic elements in fine-tuning the seasonally regulated reproductive functions in female mink.

LIST OF ABBREVIATIONS USED

ARCAG	Atlantic Research Centre for Agricultural Genomics
BAC	Bacterial artificial chromosome
BWA	Burrows-wheeler alignment
CCFAR	Canadian Centre for Fur Animal Research
COC	Cumulus-oocyte-complex
CTRL	Control
Cx43	Connexin43
DNA	Deoxyribonucleic acid
EMBL	European Molecular Biology Laboratory
ERK	Extracellular signal-regulated kinase
FSHR	Follicle-stimulating hormone receptor
FOXL2	Forkhead box L2
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GnRH	Gonadotropin-releasing hormone
GWA	Genome-wide association
HMGN1	High-mobility group nucleosome binding domain 1
HWE	Hardy Weinberg Equilibrium
IGFBPs	Insulin like growth factor proteins
JNK	Jun-N-terminal kinases
LD	Linkage disequilibrium
LH	Luteinizing hormone
Lrh	Liver receptor homolog
MAF	Minor allele frequency
MAP	Mitogen-activated protein
MDR	Moderate diet restriction
MSX	Muscle segment homeobox
NBCEC	National Beef Cattle Evaluation Consortium
NR5A2	Nuclear receptor subfamily 5
PCOS	Polycystic ovary syndrome
PCR	Polymerase chain reaction
PPARs	Peroxisome proliferator-activated receptors
PR	Progesterone receptor
PTX	Pentraxin
QC	Quality control
RFLP	Restriction fragment length polymorphism
RT	Room temperature
RT-PCR	Real-time polymerase chain reaction
SCN	Suprachiasmatic nucleus
SHBG	Sex-hormone binding globulin
SNPs	Single nucleotide polymorphisms
SPARC	Secreted protein acidic and cystein-rich
TAG	Triacylglycerol
TNFα	Tumour necrosis factor alpha
VEGF	Vascular endothelial growth factor

ACKNOWLEDGEMENTS

Firstly, I would like to express my heartfelt gratitude to my supervisor Dr. Kirsti-Rouvinen Watt and my co-supervisor Dr. Bernhard Benkel. Their constant support and guidance have been my greatest motivation for the successful completion of the project. It has been a pleasure working and discussing research with them. I would like to present a token of thanks to my committee members Dr. Bruce Murphy and Dr. Robert Beiko for their insightful inputs to my research and assisting me whenever needed.

A special thanks to Katherine Rutherford for helping me on numerous times and sharing her valuable comments in understanding the details of the project. She has been a reliable academic support for me. I would also like to thank the staff at the Canadian Center for Fur Animal Research (Cindy Crossman, Annette Murphy, and Rae MacInnis) for their immense help during sampling. Carnivore and Nutrition Physiology Lab team also deserves a vote of thanks.

To my supportive partner, friend, guide, philosopher Ronak Desai, thank you for bearing me all this while. I couldn't have done this without you. Thank you Basanti Bandekar, you were always there for me to share a laugh and also provide me a shoulder to shed my tears. To the most stable, rational person I have ever met in my life, Ramesh Eerpina – thank you for reminding me life's not always golden. Thanks to Maria Caraza Salas for always supporting and believing in me no matter what. I would also thank two little members of my family here in Truro Mitesh Patel, Punya Raja for your selfless warmth and crazy Friday nights.

Finally, I dedicate my thesis to my ever supportive and loving family (Mom, Dad, Bro) for your immense confidence in me and infusing me with positivity whenever I needed. Even though you were miles away in a different country your love and warmth were my support system all this while. I love you.

CHAPTER 1: INTRODUCTION

The mink *(Neovison vison)* is a carnivorous, semi-aquatic mammal which belongs to the *Mustelidae* family (Dunstone, 1993). For a profitable mink industry, an important production criterion is optimal reproductive performance which affects the number of marketable offspring (Lagerkvist *et al.*, 1993). Reproductive performance depends on numerous factors such as folliculogenesis, ovulation rate, hormonal profiles, ovarian tissue remodelling, embryo implantation and development, lactation, weaning and feeding management practices. Litter size is also largely limited by the high mortality rate of the kits (Hansen *et al.*, 2010). Selection for increased litter size does not have any adverse effect on fur quality (Lagerkvist, 1997). Selection for large body size, on the other hand, can lead to small litter size and also a decline in fur quality (Lagerkvist *et al.*, 1994).

The candidate gene approach is a successful strategy to identify genetic elements underlying complex, economically important traits on the genome (Youngerman *et al.,* 2004; Ponsuksili *et al*., 2011). In the present study, a comparative genomics approach was used to identify the potential candidates where knowledge is derived from information-rich species including humans and mice. Sixty six (66) candidate genes were identified for this study. Based on their involvement in signalling pathways the genes were grouped into six categories including mitogen activated kinase (MAPK), circadian rhythm, peroxisome proliferator activated receptors (PPARS), DNA repair, cytokine, and others. Once the candidate genes were selected, single nucleotide polymorphisms (SNPs) for genotyping were prioritized by incorporating different parameters including genomic location, allele frequency and other criteria. After genotyping, a battery of control procedures including call rate, monomorphism and missing data were implemented to identify the appropriate

markers for the association analysis. With the advent of high throughput and low cost methods for genotyping, significant progress in candidate gene studies has been observed. Such studies provide useful information to understand the genetic basis of variation in performance among individuals that can improve existing animal breeding strategies.

Reproductive performance is affected by both biological (genetic) factors and environmental factors like diet management. However, the present study focussed on localization of genetic variation underlying fertility in mink females. This project was conducted to identify genetic elements underlying reproductive performance in female mink and hence, demonstrate the efficacy of genome-based selection for mink. The study involved the utilisation of tools for identification of possible association of polymorphic sites with the complex physiological fertility traits. It is anticipated that this research will help in improving the economic return for the ranchers via the integration of deoxyribonucleic acid (DNA) markers for fertility traits into their conventional selection schemes. Although this project aimed at improving female fertility, the technology developed as part of this study will be useful for the betterment of any other economically important trait in ranched mink such as feed efficiency and pelt quality.

CHAPTER 2: LITERATURE REVIEW

2.1 Mink reproductive physiology

Mink is a seasonal breeder (Pilbeam *et al.*, 1979). Late February to early March is considered as the breeding period for mink followed by whelping in late April to early May (Sundqvist *et al.,* 1988). Weaning occurs in June approximately 6-8 weeks after birth. Once weaning has taken place, mink kits continue to grow and mature quickly. This period of growth until pelting is divided into two main phases: early growth which lasts from June to late August and late growth which lasts from late August till pelting season (Rouvinen-Watt *et al.,* 2005). November marks the onset of the selection process for a future robust breeding stock followed by pelting of culled mink in December (Murphy, 1996). The first breeding season for the yearlings starts at the same time as that of the adult mink (Gulevich *et al.*, 1995). However, the lifespan of a mink dam is relatively short as litter size generally decreases after the second productive season (Lagerkvist *et al.*, 1994).

2.1.1 Physiological factors of reproductive performance in female mink

Reproductive performance is a multifactorial trait controlled by genetic, environmental and physiological factors (Abegaz *et al.*, 2002; Chebel *et al.*, 2007; Castellini *et al.*, 2010). It is a lowly heritable trait and hence, direct selection for litter size may not result in significant improvement in fertility (Lagerkvist *et al.*, 1993; Hansen *et al*., 2010). Lefèvre and Murphy (2008) outlined a number of physiological parameters which contribute to the size of litter in mink including follicle development, ovulation, fertilization, preimplantation embryo loss and postimplantation embryo loss. For the fur industry, the number of kits per dam surviving from birth until pelting is an important economic determinant.

2.1.1.1 Follicle development and ovulation

Mammalian female fertility is primarily controlled by ovarian folliculogenesis and ovulation which are in turn regulated by the hormonal profiles of the pituitary gonadotropins and other growth factors (Duggavathi *et al.*, 2008). Douglas *et al.* (1994) identified follicles of >0.7mm in diameter capable of ovulating in pastel mink. It was also observed that ovaries from unmated mink contained large, luteinized, unruptured follicles (Douglas *et al.*, 1994). These observations indicate that ovarian folliculogenesis is critical in determining fecundity in female mink. In the past few years, genes that regulate the complex intraovarian mechanisms controlling follicullogenesis have been identified in mice and cattle. These include the progesterone receptor (PR) gene, the liver receptor homolog (Lrh1) gene, the forkhead box L2 (FOXL2) gene and the follicle-stimulating hormone receptor (FSHR) gene (Lydon *et al.*, 1996; Duggavathi *et al.,* 2008; Uhlenhaut *et al.*, 2009; Yang *et al*., 2010). Little is known regarding the genetics behind ovarian follicular dynamics in mink although it seems reasonable to suggest that an increase in the number of preovulatory follicles can contribute to the size of the litter.

Mink is an induced ovulator (Pilbeam *et al.,* 1979), with ovulation occurring 36-52 hours after coital stimulus (Hansson, 1947; Enders, 1952). A synchronized wave of follicle development has been reported in mink following ovulation (Douglas *et al.*, 1994). Subsequent copulations induce successive ovulations if there is an appropriate interval between matings (Douglas *et al.,* 1998). Generally, a seven-day interval between matings results in a second ovulation (Hansson, 1947). However, another study reported that, irrespective of the number of matings, the highest number of weaned kits is achieved in the group of females with shortest interval (1-4 days) between the first and last matings (Ślaska

and Rozempolska-Rucińska, 2011). In the same study, equal numbers of kits born and weaned were observed for females who were mated two times (1+1), three times (1+1+1) and four times (1+1+1+1), which would suggest that mating mink more than twice is economically futile.

Socha and Markiewicz (2002) studied the relationship between the dates of first mating and related reproductive capacities in female mink. Mink females were classified based on their date of first mating. The first group comprised of females that mated until the 5th of March, and in the second group matings occurred between the 6th and the 9th of March. The third group had females that mated between the 10th and the 15th of March and females in the fourth group mated on the 16th of March and later. The study reported that the mean numbers of kits born and weaned were highest for females mated until the 5th of March and lowest for females mated after the 15th of March. These findings support the conclusion that both matting pattern and mating date can be a constraint on litter size.

2.1.1.2 Fertilization and implantation

In mink, fertilization occurs in the oviduct where the fertilized egg undergoes development until the blastocyst stage and remains in the uterus in a state of arrested development until some days before implantation (Enders, 1952). The newly fertilized eggs from the second mating develop until the blastocyst stage and join the first group of fertilized eggs in the uterus, a phenomenon recognised as superfetation (Enders, 1952). Uterine flushings have often found unfertilized ova, indicating that fertilization was incomplete (Lefèvre and Murphy, 2008) which therefore can be a cause of low reproductive performance in mink.

Delayed implantation is a typical feature of mink females which is characterised by a reversible arrest in embryo development known as embryonic diapause and is observed in every breeding season (Lefèvre and Murphy, 2009). The emergence, maintenance and termination of embryonic diapause is regulated by endogenous influences (maternal control, pituitary gland and uterine factors including polyamines), external environment (photoperiod) and cellular events (Lopes *et al.,* 2004; Murphy, 2012*)*. Owing to delayed implantation, the mean length of gestation in mink varies from 45 to more than 70 days (Hansson, 1947; Enders, 1952; Bowness, 1968). It has been observed that the length of gestation shortens as the mating date approaches the end of March (Hansson, 1947; Bowness, 1968). In other litter-bearing species, large litters are associated with a higher incidence of mummified foetuses, which suggests uterine space could be a restrictive factor for a successful gestation (Wu *et al.,* 1988). No such studies in mink have been published so far.

Enders (1952) determined a relationship between litter size and the length of gestation showing that smaller litters are associated with extended gestation periods. Another study also reported reduced litter size with increased gestation length (Hansen *et al.*, 2010). It could, therefore, be hypothesised that increased length of diapause correlates with embryonic losses, and that this could be a physiological limitation on litter size. Studies have revealed that differentially expressed genes including ornithine decarboxylase, high-mobility group nucleosome binding domain 1 (HMGN1), secreted protein acidic and cystein-rich (SPARC), are critical for the regulation of embryonic diapause in mink (Lefèvre *et al.*, 2008; Lefèvre *et al.*, 2011; Murphy, 2012). Identification of genetic elements responsible for the termination of embryonic diapause would improve the

understanding of mechanisms involved in its regulation and would decrease the risk of losing fertilized eggs, hence improving reproductive performance.

2.1.1.3 Whelping, lactation and weaning

Maximizing offspring survival rate at parturition is an important factor for a successful production season. In pigs, an extended and difficult parturition adversely influence the survivability of piglets and may also negatively affect maternal behaviour towards piglets (Malmkvist *et al.,* 2006). In agreement with this study, similar observations were made in mink where females with a prolonged parturition spent less time in kit-directed behaviour, e.g. licking kits and moving kits close to teats, and lost more than 70% of their kits from day of birth until day seven (Malmkvist *et al.,* 2007). During the latter part of gestation, feed consumption by mink females usually decreases leading to mobilization of body fat reserves. This implies that the dams enter the lactation period in a negative energy balance (Tauson, 1994; Tauson *et al.,* 1994). In most mink ranches, this period records the highest mortality for adult mink (Murphy, 1996).

During the first days after birth, new-born mink kits, being physiologically immature with poorly developed thermoregulation (Rouvinen-Watt and Harri, 2001), depend entirely on their mothers for warmth, protection, and nourishment (Tauson, 1994). Lactation is a critical determinant for reproductive success in mammals. In mink, a positive relationship exists between activated teats at two days postpartum and litter size (Korhonen, 1992). However, the number of active teats can be a potential constraint for large litters as competition for teats among kits may arise, as a result of which smaller kits may suffer from long periods of milk unavailability and may eventually die (Martino and Villar, 1990).

Mortality rates of 20-30% between birth and weaning have been reported in mink, of which 60-90% has been found to occur within first few days of birth (Martino and Villar, 1990; Schneider and Hunter, 1993). Similar mortality rates have been observed in other litter bearing species including pigs and cats (Scott *et al.,* 1978; Grandinson *et al.*, 2003). Furthermore, 11-50% of the dead kits in mink are stillborn and occur mostly in large litters with more than seven kits (Martino and Villar, 1990; Schneider and Hunter, 1993; Malmkvist *et al.,* 2007). Septicaemia is another cause of mortality in kits and occurs mostly within the first week after birth (Martino and Villar, 1990). During the stressful weaning period there is an upsurge in the activity of the kit's digestive enzymes and a simultaneous increase in the weight of digestive organs, which suggests the readiness of the mink's body for transition from milk to solid food (Tauson *et al.,* 1994). In mink, a higher percentage of kits weaned leads to a higher survival percentage at age of six months or pelting (Hansen *et al.,* 2010) and therefore improvement in number of kits weaned results in an economic advantage.

It is likely that identification of genetic elements affecting the complex physiological processes of reproduction would help in understanding the interplay of these genetic elements in fine-tuning reproductive functions. Achieving optimal reproductive performance would in turn increase the profitability of mink farming operations.

2.1.2 Effect of feeding intensity on reproductive performance in female mink

To maximize profitability, breeding for large body size in mink is a common practice as large pelts are obtained from large mink (Lagerkvist, 1997). High feeding intensity to promote maximum body weight gain in the fall leads to higher fat deposition (Korhonen and Niemelä, 1998) and consequently has adverse effects on reproductive performance. In

mink, high feeding intensity in fall leads to high female pre-mating weights (Tauson and Aldén, 1984) and necessitates severe slimming before breeding to achieve appropriate body condition. In studies by Tauson and Aldén (1984 & 1985), drastic weight loss of over 300g for the over-conditioned yearling females caused a higher percentage of barren females and increased kit mortality, compared to females in moderate condition who lost less than 30g and achieved optimal breeding success with lower kit losses.

Adjustment of feed availability to achieve better reproductive success is evidenced in a recent study by Boudreau *et al.* (2014). Dams on a restricted diet regime showed superior live litter sizes compared to the females in a control group which were fed *ad libitum* (Boudreau *et al.,* 2014). In the same study, it was also determined that females in the restricted group did not lose weight during the first three weeks of lactation and therefore, suggests improved metabolic health of the restricted females despite the additional nursing burden to maintain their larger litters.

2.1.2.1 Obesity and its consequences

Reviews indicate that obesity is a multifactorial, chronic disease which results from excessive fat accumulation in the adipose tissue due to energy imbalance between intake and expenditure (Kopelman, 2000; Nammi *et al.,* 2004). Obesity causes increased lipolysis which leads to an increase in free fatty acids in circulation and fatty acid overload in tissues including the liver, skeletal muscles and pancreas (reviewed by Grundy, 1998; Ye, 2007). Hepatic uptake of free fatty acids contributes to the synthesis and accumulation of triglycerides, causing hyperglycemia. In an attempt to stimulate glucose uptake, β-cell function in the pancreas is increased and consequently hyperinsulinemia occurs followed by insulin resistance (Frayn, 2001). Elevated insulin levels stimulate leptin production, and

in obese individuals hypothalamic resistance to these adiposity signals is observed (Wabitsch *et al.,* 1996). Furthermore, obesity is a state of inflammation characterised by high plasma concentrations of biomarkers including C-reactive protein, fibrinogen, tumour necrosis factor-alpha and interleukin-6 (Dandona *et al.,* 2004; Nguyen *et al.,* 2009), and these inflammatory parameters positively correlate with insulin resistance (Garanty-Bogacka *et al.,* 2011).

2.1.2.2 Obesity and reproductive performance

In mink industry, selection for a large body size is a common practice in order to maximize pelt revenue. Unfortunately, the obese phenotype in females is associated with poor reproductive performance. Hansen *et al.* (2010) reported a negative correlation between a dam's body weight and litter size and kit survival rate in mink. This is in agreement with the findings of Clausen *et al*. (2007) who also noted fewer live born kits for mink females who were overweight. Another study also determined that an increase in the dam's juvenile body weight is associated with an increase in the number of stillborn kits and increased kit mortality until 3 weeks of age (Lagerkvist *et al.,* 1994). A recent study by Boudreau *et al*. (2014) has demonstrated superior litter size in females with limited dietary allowance in the fall compared to a control group which was fed *ad libitum* and hence, were over-conditioned during the fall. In other litter bearing species, increased neonatal mortality rate, smaller litter size, and birthing difficulties were found in obese female cats and rats (Lawler and Monti, 1984; Rasmussen, 1998), suggesting that obesity has negative impacts on reproductive performance in litter bearing mammals.

Increased insulin levels during obesity stimulate androgen production, and sex-hormone binding globulin (SHBG) transports androgens to their target tissues. In obese women, low

SHBG concentrations have been reported as insulin acts as an inhibitory factor for hepatic SHBG synthesis (Plymate *et al.,* 1988). In the absence of circulating SHBG, an increase in free androgen is found (Pasquali, 2006), which results in a hyperandrogenic state associated with menstrual cycle abnormalities and chronic infertility (Pasquali and Gambineri, 2006). Also, insulin exerts inhibitory effects on insulin-like growth factor-binding protein 1, which binds to sex steroids and insulin-like growth factors thereby affecting regulation of ovarian growth and cyst formation (Poretsky *et al.,* 1999).

During obesity, elevated leptin levels accelerate gonadotropin-releasing hormone (GnRH) pulse frequency and may affect the reproductive-axis (reviewed by Moschos *et al*., 2002). With increasing GnRH, the release of luteinizing hormone (LH) is stimulated causing hyperplasia of ovarian theca cells and further contributing to androgen production. Hyperandrogenism is a typical characteristic of polycystic ovary syndrome (PCOS), a common cause of infertility in women (Puurunen *et al.,* 2009). More than 50% of PCOS affected women have an obese phenotype suggesting that obesity-related hyperandrogenism favours PCOS (Gambineri *et al.,* 2002). Studies have also indicated that increased leptin concentrations in the ovary may impair ovulation by interfering with folliculogenesis and oocyte maturation (Duggal *et al.,* 2000). Leptin is also thought to act as a growth factor and play a key role in the regulation of the energy balance for nutrient availability between the mother and the fetus (Mostyn *et al.,* 2001). In pregnant mink, hyperleptinemia is reported during the last 20 days of gestation (Tauson *et al.,* 2004). The study demonstrated the anorexigenic effects of leptin on pregnant females which caused a decrease in their body weights during late gestation.

Offspring from overweight mothers have a greater risk of developing an obese phenotype later in life independent of postweaning diet. This shows that maternal obesity has a long-term adverse effect on the offspring (Howie *et al.,* 2009). Studies have also showed that maternal obesity in humans is linked with an increased risk of stillbirths and neonatal mortality (Kristensen *et al.,* 2005), similar to what occurs in mink (Lagerkvist *et al.,* 1994).

Based on the above, it is apparent that obesity has a negative impact on reproductive performance in mink. Intense selection for a large body size is unfavourable to achieve optimal reproductive success. Moreover, minimizing extreme fluctuations in body weight throughout the production cycle may improve mink health and fecundity. However, genetic variants associated with obesity in mink have not been identified and opportunities exist to investigate in this regard.

2.2 From linkage maps to association studies

In the early 1900s, Bateson and colleagues identified that their crosses between purple flowers with long pollen grains and red flowers with round pollen grains did not follow the "independent assortment" ratios predicted by Mendel (Bateson *et al.,* 1909). Later, Morgan (1911) working with *Drosophila melanogaster* first suggested the concept of linkage where two genes are closely linked on the same chromosome and do not assort independently. In 1913, Sturtevant, an undergraduate working with *Drosophila* constructed the first genetic map and also laid the foundation for genetic mapping. His research helped geneticists to develop genetic maps for other model organisms and allowed mapping of genes to chromosomal regions using controlled crosses. Construction of the first genetic linkage map using restriction fragment length polymorphism (RFLP) markers led to an interest in tracing inheritance patterns in human pedigrees (Botstein *et al.,* 1980). Thereafter, linkage

analysis in humans which is based on co-segregation of the marker and the trait became a widely studied area of genetics. With the advent of the polymerase chain reaction (PCR), RFLP markers were soon replaced by microsatellites primarily due to ease of amplification by PCR followed by allele sizing on gels (reviewed by Vignal *et al.,* 2002). Inconsistencies in allele size determination can be a technical concern in using microsatellites. Currently, SNPs are widely used as markers for linkage analysis due to their high abundance, widespread distribution across the genome and ease of genotyping. With technological advancement, SNP genotyping has become simple and cost effective, although identifying SNPs requires substantial efforts. Since SNPs are biallelic, relative to microsatellites a larger number of SNPs needs to be included in the analysis (Aitken *et al.,* 2004).

Traditionally, linkage analysis was used as an alternative method for identification of genetic elements in single-trait disorders. Linkage studies have been successful in the identification of genetic elements for single-trait disorders (Kerem *et al.,* 1989). Such studies, however, have had limited success when identifying genetic elements for complex traits where phenotype is determined by multiple factors. On the other hand, association analysis has been successful in identifying genetic variants with small effects on complex trait both in humans and livestock and has more power relative to linkage analysis (Johnson and O'Donnell, 2009; Rempel *et al.,* 2010). In association analysis, a genetic variant is genotyped in a population of unrelated, affected individuals for which phenotype and, ideally pedigree records are available. An allele is said to be associated with a trait of interest if it occurs at a significantly higher frequency in the affected group compared to the unaffected control population (reviewed by Balding, 2006).

2.3 Genetic association studies

Association studies have become an essential tool in determining the genetic basis of diseases and inherited quantitative traits. Genetic association studies can be conducted through a candidate gene or genome-wide association (GWA) approach (reviewed by Collins *et al*., 1997). In candidate gene association studies, the frequency of the causative allele in the affected individuals and non-affected individuals is examined (Cargill *et al*., 1999). This approach is used to identify genetic elements that strongly affect susceptibility to diseases and other quantitative traits (reviewed by Amos *et al*., 2011). In the GWA approach, studies involve the phenomenon of linkage disequilibrium (LD) using multiple SNPs to screen for genetic variants associated with the trait of interest (Cargill *et al*., 1999). Such studies are comprehensive in nature as they permit interrogation of the complete genome rather than focussing on small candidate regions and no prior assumptions are made about the genetic associations of the causal variants (reviewed by Pearson and Manolio, 2008).

2.3.1 Genome-wide association (GWA) studies

A review by Hirschhorn and Daly (2005) shows that GWA studies involve utilisation of genetic variants distributed throughout the whole genome along with integration of phenotypic and, ideally pedigree records to perform association analysis. Contrary to the candidate gene approach, a GWA study requires no prior information regarding gene function which renders it an unbiased approach. Minimizing cases of false positives is important for a GWA study as it involves a larger set of markers. False positive is the rejection of a null hypothesis (no significant association) when it is true. False positive rate is the proportion of significant associations that are false positives (Pearson and Manolia,

2008). Statistical fluctuations owing to a liberal p-value threshold, systematic bias due to population stratification and technical artefacts can be few possible sources of false positive associations (Hirschhorn and Daly, 2005). The most common way to reduce the FDR is by applying the Bonferroni correction where the conventional P value is divided by the number of tests performed in the association analysis (Yang *et al.,* 2005). Other approaches are estimation of Bayes factor probability and false positive report probability (Hochberg and Benjamini, 1990).

A multistage approach is suggested as a preferred GWA study design (Hirschhorn and Daly, 2005). The first stage begins with genotyping of a full set of SNPs in a small population at a modest p-value threshold to identify putative SNP associations. In the next stage, SNPs identified from the previous screen are re-tested in a larger, independent population.

GWA analysis is a novel yet productive method for the identification of genetic elements controlling phenotypic traits in domestic animals. The dairy industry has reported the use of GWA studies for many economically important traits such as fertility, milk yield and growth (Jiang *et al*., 2010; Mai *et al*., 2010). Currently, there are numerous reports of GWA studies in domestic animals using dense SNP marker panels for efficient identification of genetic elements associated with a complex trait and this information can be exploited to improve the existing breeding schemes (Sahana *et al*., 2010; Ponsuksili *et al*., 2011).

2.3.2 Candidate gene approach

As reviewed by Tabor *et al*. (2002), in candidate gene studies, researchers begin with the selection of a suitable candidate gene on the basis of any prior evidence indicating a critical role of the gene in the trait of interest within related species. Further, the validity of this

"educated guess" is investigated (Kwon and Goate, 2000). To recognise the candidate genes from larger lists, computationally efficient "gene prioritization tools" are available which saves time and resources invested in candidate gene approach (Tranchevent *et al.,* 2011). Once candidates are selected, SNP selection must be done based on their nature and location in the genome. Additionally, the gene variant should also occur in considerable frequency to allow detection of allelic differences between cases and controls with respect to the disease or trait under study. Since candidate gene studies involve smaller number of polymorphisms there is less occurrence of false positives. A concern, however, with candidate studies is their inability to discover any new genes beyond those selected as putative candidates.

With the creation of SNP marker databases (Krawczak *et al.,* 1997; Brookes *et al.*, 2000; Sherry *et al.,* 2000) and the advent of high-throughput methods for genotyping, the scope of the candidate gene approach has evidently changed. This strategy has been successful in the identification of genetic variants affecting economically important traits in livestock (Youngerman *et al.,* 2004; McNatty *et al.,* 2005).

2.3.2.1 Selection of candidate genes

Selection of candidate genes is the foremost step in candidate gene approach. Few strategies for identification of candidates are position-dependent, function-dependent, comparative-genomics and combined approach (reviewed by Zhu and Zhao, 2007) (Figure 1). In the present study, a comparative genomics approach is used to identify the potential candidates.

Selection of candidates

Position-dependent

- Based on genetic maps.
- Successful in identification of candidates in livestock for economically important traits (Grisart *et al.*, 2002).
- Refining candidates in targeted region is difficult.

Comparative genomics

- Based on cross-species approach.
- Studies in human and mice successful in identification of candidates in livestock (Rothschild *et al.*, 1996).
- Genetic architecture for complex traits may be different across species.

Function-dependent

- Based on functional significance (signalling, metabolic and developmental pathways).
- Successful in identification of candidates in livestock using knock-outs, cellular models (Schwerin *et al.*, 2003).
- *A priori* knowledge of genes required.

Combination of the above approaches

Gene prioritization tools to identify the most potential candidates

Figure 1. Flowchart outlining the strategies for identification of putative candidate genes

In comparative genomics approach, knowledge is derived from the available resources of information-rich species including humans and mouse to identify potential candidates underlying economically important traits in agricultural species'. This approach is based on the concept that candidate genes are "functionally conserved or are structurally homologous genes" identified in other species (reviewed by Zhu and Zhao, 2007). Genome comparisons between humans and mustelids (mink), bovines (cow) and suids (domestic pigs) have revealed conservation of multiple chromosomal segments across species owing to slow rate of global genomic shuffles (Johansson *et al*., 1995; Chowdhary *et al.,* 1996; Hameister *et al.,* 1997). However, it is also known that although same genes are retained in the chromosomes but the order of gene homologs between species is rarely identical. Lack of sufficient functional information for the majority of genes in model species limits the utility of this approach. Occasionally, due to genetic heterogeneity comparative mapping between species becomes inefficient. Nevertheless, this approach has been successful in identification of candidate genes affecting phenotypes of agricultural relevance including reproductive performance, growth and disease resistance in livestock (Liu *et al.,* 2001; Smith *et al.,* 2001).

2.3.2.2 Prioritizing SNPs for genotyping

In the candidate gene approach, it is important to select a limited number of SNPs for genotyping in order to save resources and perform association analysis in a statistically feasible manner. It is a challenging yet crucial step in a candidate gene study. Knowledge about the location and nature of SNPs and evaluating LD patterns would be helpful to determine the best subset of SNPs to be included for the association analysis (Figure 2).

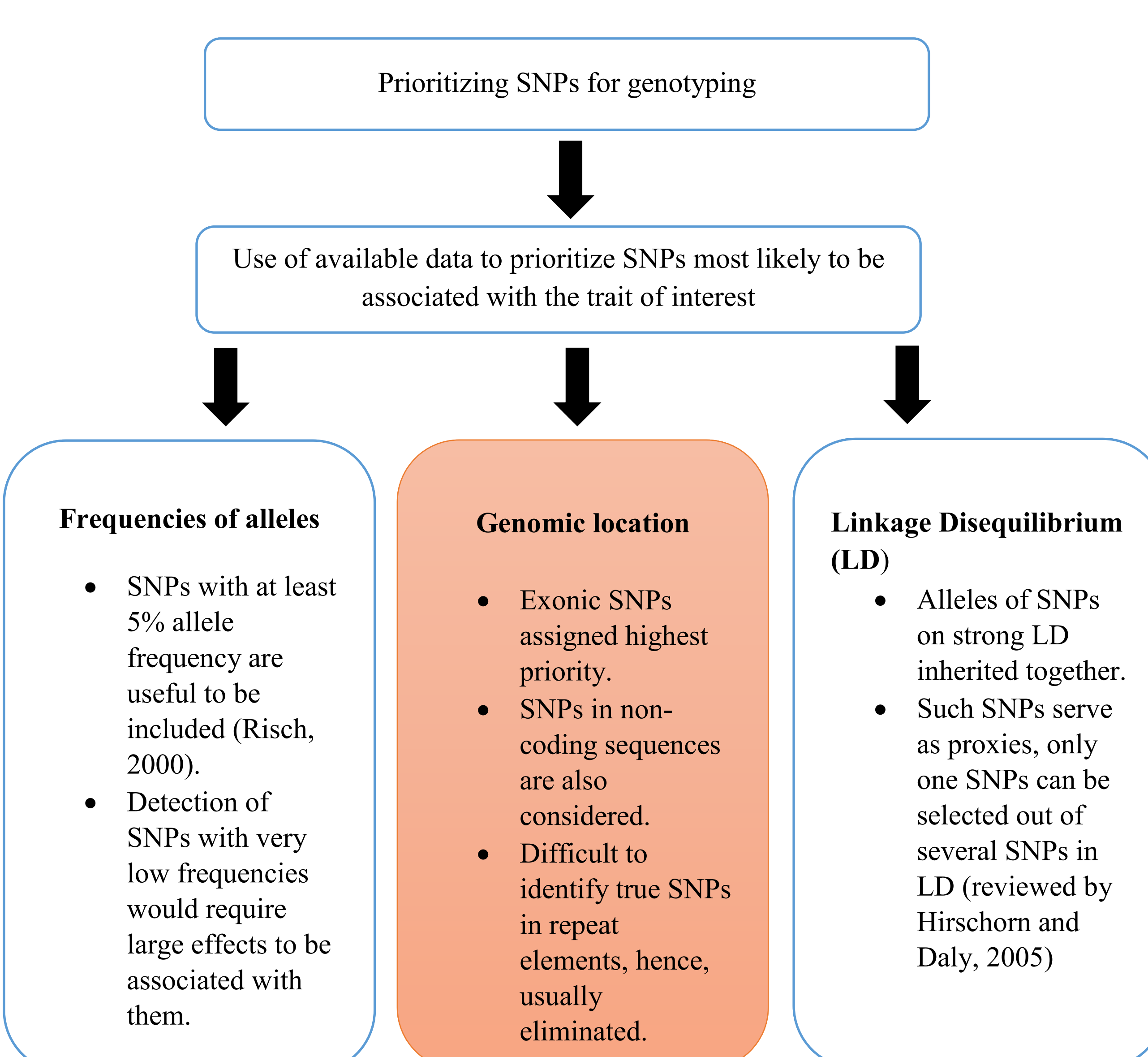


Figure 2. Flowchart outlining the steps for prioritizing SNPs for genotyping.

2.3.2.3 Attributes of samples and markers' integrity

Before the samples are sent for genotyping, checks on sample identity must be performed to avoid sample mix-ups, inadvertent sample duplication and cross contamination. Samples with low DNA concentration are also eliminated prior to genotyping as such samples usually fail to amplify and result in higher number of missing genotypes (Fu *et al.,* 2009). When evaluating the integrity of the markers, a filtering process must be performed by incorporating genotyping control procedures for different criteria. As a primary quality control (QC) procedure, cluster analysis is implemented by genotyping service providers. Each marker is analyzed independently to identify the three genotype clusters. To ensure the accuracy of the genotype clusters, standard cluster files with predefined cluster positions are compared with the newly defined marker cluster positions for the test population. If clustering is ambiguous where the cluster boundaries appear to be vague or exhibit considerable overlaps, the genotypes aren't reliable (Tindall *et al.,* 2010). Such markers are usually excluded from the analysis. On the other hand, a marker is retained if the genotype clusters are well separated and the clustering is unambiguous. DNA collection and processing procedures, plate to plate variability and environment variability are some of the sources of experimental variation leading to unwanted noise in signal intensities. To reduce the uncertainty associated with genotype clusters and minimise batch-related artifacts sophisticated algorithms like BEAGLECALL are freely available and widely used (Browning and Yu, 2009).

After primary QC procedures, a battery of internal genotyping control procedures is implemented. These include, genotyping call rate, missing data, deviations from Hardy Weinberg Equilibrium (HWE), minor allele frequency (MAF), and monomorphism of

SNPs (Chan *et al*., 2009). Missing data and call rate are the most universal indicators of markers' integrity (Di *et al*., 2005). If a stringent threshold is adopted for calling genotypes some true signals might be discarded. On the other hand, if a liberal threshold is set, call rates are maintained but accuracy is compromised (reviewed by McCarthy *et al*., 2008). HWE is another important criterion of determining SNP usefulness. HWE describes that within large populations, the allele and genotype frequencies remain constant from generation to generation unless disturbing forces of mutation, recombination, selection, genetic drift or population structure are introduced to misbalance the equilibrium (Benarie, 1981). SNPs showing extreme deviations from HWE are usually discarded (Wellcome Trust Case Control Consortium, 2007). Minor allele frequency (MAF) is an important criterion to be assessed for SNP selection. In a population the minor allele is the less frequent allele at a variable site (Kim *et al.*, 2011). A SNP with a very low MAF (<2%) is usually excluded as such an SNP is expected to have negligible effect on genomic evaluation (Wiggans *et al.*, 2009). Unfortunately, exclusion of SNPs due to MAF can cause a considerable loss of data and might affect the ability to detect rare polymorphisms (Gorlov *et al*., 2008). Monomorphic SNPs are uninformative and are eliminated. The establishment of appropriate threshold for different parameters is important in association studies. Set of thresholds, however, vary according to sample size and study design.

2.3.2.4 Study designs for association analysis

Two principal study designs for performing an association analysis are: population based designs (case-control study or cohort) and family based designs. In a case-control study, allele frequencies in cases with the disease, or ascertained for a specific phenotype, are compared to controls from a disease-free group (reviewed by McCarthy *et al*., 2008). Case

and control participants are taken from the same resource population and trait measurements are collected similarly for both groups in order to minimize the effects of confounding or population stratification (Zondervan, 2011). When cases and controls are selected from such a population, false positive associations are detected due to sampling differences. Another population-based study design is a cohort study. This involves collection of information from a larger group of individuals who are then evaluated for the incidence of disease or trait in subgroups categorized by genetic variants (reviewed by Pearson and Manolio, 2008). A drawback of cohort studies is that they require a lengthy follow-up period which makes them expensive. Family based designs range from simple parents-offspring trio designs to multigenerational pedigrees (Benyamin *et al.,* 2009). The trio study design includes the disease affected case participant and both of his/her parents. Selection of an appropriate study design for the association analysis depends on the resource population and the objectives of study.

2.3.2.5 Tests for association analysis

Data obtained for each SNP with major allele A and minor allele a is represented as a contingency table of counts of disease or phenotype status by either genotype count or allele count (Chanock *et al*., 2007). Associations with two alleles of a SNP are tested directly by comparing the frequency of each allele in cases and controls, and the frequency of each of three possible genotypes can also be compared. Under the null hypothesis of no association with the disease or trait of interest, it is expected that the allele or genotype frequencies will be equivalent in both case and control groups. A simple test of association can be done by a simple chi-squared (χ^2) test for independence (Chanock *et al*., 2007), with stringent statistical thresholds to determine significance of associations. The Cochran-

Armitage trend test is another commonly used test in evaluating associations where a set of scores is assigned to genotypes (Armitage, 1955). For quantitative traits, linear regression based methods are recommended and for categorical phenotypes, multinomial regression based approaches are used (reviewed by Balding, 2006). Currently, different machine learning methods are available to determine and predict associations between genetic variants and distinct phenotypes including Support Vector Machines (SVM), Artificial Neural Networks and Naïve Bayes (Ban *et al*., 2010; Wei *et al*., 2011). There are several publicly available programming packages which can be used to perform the association analysis such as PLINK and R, comprising of all the tools needed to assess genetic association for traits and also to perform genetic calculations.

2.3.2.6 Validation studies

To determine the credibility of a genetic association a validation study is always recommended. This is done in an independent, yet large, sample drawn from another population of unrelated individuals from the same breed, in order to validate the previous genetic association results and the phenotype measured should be the same with the initial study (Clarke *et al*., 2011; König, 2011). It has been reported that genetic variants that have a positive association with disease or the trait of interest in one population may not necessarily have the same consistent association in another population due to population diversity and/or bias leading to overestimation or a spurious association in the first study (Ioannidis *et al*., 2001). In a study conducted by Lohmueller *et al*. (2003), 301 published studies which identified 25 disease loci were re-analyzed by meta-analysis and only 11 loci were reported to have yielded significant association. Lack of reproducibility has been

often reported due to selection bias, population stratification, genotyping errors and others (Chanock *et al*., 2007).

Before moving markers from a discovery resource population to commercialization, independent validation studies are important to confirm previously established associations between a marker and a phenotype. For example, one of the primary roles of the National Beef Cattle Evaluation Consortium (NBCEC) is to validate associations between commercially available DNA-based tests and beef cattle production traits as claimed by the genotyping companies. The validation process at NBCEC takes place through a partnership between the breed associations and the commercial genotyping companies. The companies' request the validation and also perform the genotyping of the DNA samples provided by the consortium. Finally, NBCEC carries out the association analysis and re-confirms the claimed associations between the results of the DNA-based test and the phenotype. Several associations between commercially available DNA-based tests and beef production traits have been validated by NBCEC (Van Eenennaam *et al.,* 2007). The GeneSTAR® Tenderness is one of the validated DNA marker panel test which comprises of three markers (Calpastatin CAST-T1, Calpain-T2, and Calpain-T3). Calpastatin and calpain are naturally occurring proteolytic enzymes responsible for post-mortem meat tenderness (Koohmaraie, 1996). Company trials and published findings demonstrated an improvement in meat tenderness to be associated with the favourable forms of the three markers: CAST-T1, Calpain-T2 and Calpain-T3 (Barendse, 2002; Page *et al.,* 2002). These significant associations were successfully validated by NBCEC (Van Eenennaam *et al.,* 2007). An unbiased, third-party verification of the commercial DNA-based tests helps the producers to believe in the marker technology. One of the major challenges for conducting

a validation study in livestock is the paucity of larger (validation) populations with sufficient performance data.

As far as the current study is concerned, due to time constraints and in the absence of a larger population of mink with phenotypic records, validation of the marker panel was not possible. In the future, validation of the marker panel in as large a population as possible is encouraged. Such an initiative would need a coordinated effort among mink ranchers for collection of phenotypic records and DNA, and also require the cooperation of provincial and national producer organizations.

2.3.2.7 Farm animal genomics

Much progress in farm animal genomics has been made in the past decade, from rudimentary linkage maps to whole-genome sequencing. With the release of the first draft of the chicken genome (International Chicken Genome Sequencing Consortium, 2004), interest in chicken genome research increased significantly. In the same year, the first draft of the bovine genome was completed, closely followed by porcine genome draft assembly (Schook *et al.,* 2005). Similar sequencing efforts have been reported in sheep and domestic goat (Archibald *et al.,* 2010; Dong *et al.,* 2013). Besides the reference genome, SNP panels have been designed for farm animal species to detect genetic variability and to gain biological insights into the production and functional traits.

The American mink has lagged behind in the field of genomics and studies to develop molecular tools for improvement of genetic resources for *Neovison vison* have been limited. However, the first linkage map for the American mink has generated interest towards genetic research in this mammal (Anistoroaei *et al.,* 2007). Also substantial improvement in the microsatellite-based map has been reported (Anistoroaei *et al.,* 2012).

Currently, a draft genome sequence of the American mink is nearing completion. A database of approximately 246,000 contigs for the mink genome, generated by NextGen sequencing, is already available. Recently, the first transcriptome analysis derived from pool of four different tissues of the wild-type American mink was published (Christensen and Anistoroaei, 2014). The study identified 16,111 annotated coding sequences in mink deposited in European Molecular Biology Laboratory (EMBL) database (PRJEB1260). This data set is the only available data set in mink to be used for any genetic based analysis. In recent years, tools for genome analysis have been developed for the localization of genetic elements underlying genetic variation for complex traits including reproductive performance in livestock. Maintaining optimal reproductive success is crucial for animal production. Improvement of reproductive efficiency through traditional phenotypic selection is difficult due to its low heritability. Use of fertility-related markers to define the animal's performance in early life has refined and improved breeding strategies for most, if not all, livestock species. Association analysis has been successful in identifying genetic variants affecting reproductive performance in cattle, pigs, sheep and rabbits (McNatty et al., 2005; Argente *et al.,* 2010; Sahana *et al.,* 2010; Ponsuksili *et al*., 2011). Currently, apart from the present project no other studies for identification of genetic elements associated with fecundity in mink are underway. This research is expected to identify genes underlying reproductive performance in mink and hence, demonstrate the efficacy of genome-based selection for mink. It is anticipated that this research will open up opportunities for genetic evaluation which will in turn ensure significant economic gain to the fur industry.

CHAPTER 3: OBJECTIVES AND HYPOTHESIS

The objectives of the proposed research are:

1. To screen a set of informative SNPs/polymorphic sites in a mink resource population.
2. To assess the association of the identified polymorphic sites with reproductive performance trait in mink.

The hypothesis of the proposed research is:

We expect the variation in allele frequency will influence reproductive performance in mink resource population.

CHAPTER 4: MATERIALS AND METHODS

4.1 Resource population

The discovery data set in which the association analysis was performed is a population of Standard Black female mink maintained since 2009 to 2013 at the Canadian Centre for Fur Animal Research (CCFAR), Faculty of Agriculture, Dalhousie University. The population was analysed independently as two distinct categories: Control (CTRL) group and Moderate Diet Restriction (MDR) group to mitigate the confounding effects of dietary treatment. Every year, one hundred (100) females were in the CTRL group while the other one hundred (100) full sister females were in the MDR group. Females in the CTRL group were fed *ad libitum* according to normal farm feeding practice (Rouvinen-Watt *et al.*, 2005) and females placed in the MDR group were fed ~20% less than the CTRL group from September to December to maintain a body condition score of 3 (Hynes *et al.,* 2004). The diet consisted of commercially produced standard wet mink feed and water was available *ad libitum*. Experimental procedures were performed in accordance with the guidelines of the Canadian Council on Animal Care (Olfert *et al.*, 1993).

For this project, females were selected based on their primiparous production. Each female was housed individually in a standard sized cage in a multi-row barn at CCFAR except for breeding. Full-sister pairs were mated to the same male in March and the number of mating attempts made was recorded. A sister-pair was eliminated from breeding if one of the females and/or both became ill, did not wean at least one kit or died. All mortalities were sent to the Nova Scotia Department of Agriculture Veterinary Services Pathology Laboratory, Truro for post-mortem examination. Cases of reproductive complications such as mastitis, dystocia and nursing sickness were recorded.

4.2 Response variables

Litter size is the primary response variable in the study. It is the measure of total number of born kits both dead and alive (TB). It also involves the estimation of number of kits alive at birth (D1), 21 days (D21), 42 days (D42) for each dam. Based on litter size, survivability (%) was calculated from the formula: Kit survivability $=\frac{Kits\ at\ D42}{Total\ kits\ at\ birth} * 100$. Rouvinen-Watt and Armstrong developed a body condition scoring system which was used to score the females every month except during breeding and lactation (Hynes *et al.,* 2004). Body weight was measured to the nearest gram on a monthly basis except in March and during lactation where weights were recorded 1 day post-partum and every three weeks after that until weaning when the kits were six weeks of age (Rouvinen-Watt *et al.*, 2005). During whelping, kits were individually weighed along with the dams, and the sex of kits, number of dead kits, and group body weights of male and female kits were recorded. All kits were also weighed along with the dams in the beginning of September to get an estimate of lean body mass, and again during pelting season for assessing mature body size.

4.3 Blood sampling

Mink dams were sampled for blood by clipping a toe nail of a hind limb at the start of the season for the yearling females, annually at 42 d post-partum, and at the end of the productive season for the 4-year-old females or when culled from breeding for the 1-3 year old females. A drop of blood from each female was collected on the Whatman™ FTA™ cards, dried, and stored for further analysis. In 2009, blood samples were collected on the Whatman® FTA Bloodstain cards, which were later found to be suitable for short term sample storage only. Hence, Whatman® FTA Classic cards intended for longer term storage were used thereafter.

4.4 Laboratory Analyses

4.4.1 DNA extraction

The captured nucleic acids were purified according to the Whatman™ FTA™ nucleic acid purification protocol with minor modifications (Whatman™, Kent, United Kingdom) using 1.2 mm Harris micropunch. To sterilize the micropunch a core from a blank FTA card soaked with a solution of TE (Tris-EDTA) buffer was taken followed by punching another core from a 70% ethanol soaked card. To prevent contamination, this step was repeated in between samples. After this step, 200 µL of Whatman® FTA purification reagent was used to wash each core followed by five minutes incubation at room temperature (RT) with occasional mixing. Thereafter reagent was drained while retaining the core in the microcentrifuge tube. This washing step was repeated three times. After this, 200 µL of TE buffer was added, vortexed and incubated for another three minutes at RT. Again reagent was discared and the core was retained. This step was repeated for twice. Following this, 25 µL of nuclease-free water (QIAGEN, Hilden, Germany) was added to each tube and incubated for twenty minutes in a heat block (AccuBlock™ Digital Dry Bath) set at 90°C. Liquid was carefully removed and transferred into a new microcentrifuge tube using a pipette. Samples were stored at 4°C until a full set of samples was extracted.

To determine DNA quality, gel electrophoresis (1% agarose gel) was attempted but the gels showed no bands indicating that the concentration of DNA was below the sensitivity of the assay. To confirm this, PCR was done with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mink-specific primers shown in Table 1 followed by gel electrophoresis.

4.4.2 Quality check

Once a complete sample set was extracted, samples were transferred to 96-well Progene PCR® microplate and sealed with an adhesive film. Since the sample number was large, analysing the samples by standard PCR wasn't considered reasonable as running all the amplified products on agarose gels is time-consuming and laborious. Hence, it was determined real-time polymerase chain reaction (RT-PCR) using the LightCycler® 480 II (Roche Diagnostics) was a better alternative. GAPDH is a commonly used reference gene for gene expression studies (Barber *et al.*, 2005), so mink-specific GAPDH primers were used as shown in Table 1. For the RT-PCR mix, a 10 µl reaction was prepared. This mix comprised of: 1µL of template DNA of unknown concentration, 600 nM of each forward and reverse primer, 1X SsoFast™ EvaGreen® Supermix and nuclease-free water (QIAGEN, Hilden, Germany).

Table 1: Primers used for generating mink-specific GAPDH sequences

Gene	**Primer**	**Primer sequence**	**Amplicon size**
GAPDH	*Forward*	TGACAAAGTGGTCATTGAGAGCAA	177 bp
	Reverse	AGAAAGCTGCCAAATACGATGACA	

The thermal cycling conditions used for this assay are shown in Table 2 and are based on Bio-Rad's recommended conditions. Melt curve profiles of the samples with a distinct single peak of ensured successful amplification and hence the presence of DNA was confirmed. Curves which exhibited a low signal plateau compared to the others were suggestive of poor amplification. These samples were re-extracted and the quality check procedure was repeated. To confirm the RT-PCR results, a few random samples from each

complete set were taken and gel electrophoresis was performed. Sharp bright bands were observed which ensured good quality DNA for genotyping.

Table 2: Sequential steps of thermal cycling conditions for RT-PCR

Step	Cycles	Temperature (°C)	Time (seconds)
Enzyme inactivation		98	120
Denaturation	40	98	10
Annealing and extension		55	20

4.5 Reference genome and SNP discovery

The genome sequence database was derived from a Nova Scotia Jet Black mink which was the most inbred mink individual available from within a closed ranch population housed at CCFAR. DNA extracted from the liver of donor mink was sheared to different fragment sizes and construction of libraries for Next Generation Sequencing was performed at the Atlantic Research Centre for Agricultural Genomics (ARCAG), Faculty of Agriculture, Dalhousie University. Bacterial Artificial Chromosome (BAC) libraries with large inserts of mink genomic DNA were constructed at ARCAG and by collaborators at the University of Copenhagen. Draft assembly construction for the American mink involved *de novo* assembly of contigs, generation of scaffolds from contigs by a comparative genomics strategy using ferret scaffold (http://genome.ucsc.edu/) and canine genomes (Lindblad-Toh *et al.,* 2005). Both ferret and dog are considered to be close relatives of mink (Anistoroaei and Christensen, 2006; Anistoroaei *et al.,* 2009). Currently, refining of the reference genome is underway and the assembly will be soon deposited in a public domain.

Genome sequencing was performed on a Standard black male mink and a Standard black female mink. Sequence reads from the standard black mink were compared to the Jet Black donor mink using Burrows-Wheeler Alignment Tool (BWA) software followed by variants calling using Samtools (Li *et al.,* 2009). The objective was to bias the SNP discovery initiative towards variation that is present in the Standard black type mink used in this study.

Sixty-six candidate gene sequences were retrieved from the dog genome and ferret scaffold.

Using BLAST, exon portions of the gene sequences were compared with the mink contig database. The exons in the contig of interest were identified.

Assembler was asked to provide VCF (variant call format) and flanking information for contigs of interest for standard black female and standard black male.

From the BLAST file, the region of interest on the contig (exon hits +/- 5000 bp) was identified.
From the VCF, SNPs which are in the gene of interest were determined.

Repeatmasking (Institute for Systems Biology, Seattle, Washington) was performed on the contigs and SNPs within repeat elements eliminated

SNPs that were scattered throughout the gene were selected. If possible, at least one SNP from promoter region and one downstream were selected. Exonic SNPs were given priority.

SNPs which were common to the standard black female and male were considered important and highlighted.

SNPs which were homozygous for the alternate allele in both standard black female and male were eliminated.

For some genes, no SNPs were identified on the basis of the filtering criteria. So, SNPs were identified from the neighbouring gene.

Figure 3. Flowchart outlining the steps for identification of SNPs for the candidate marker panel.

4.6 Candidate gene SNP panel

The candidate gene panel in this study originated from a genome-wide scan for genes underlying fertility and reproductive longevity in the mouse, which was previously also successfully used in Holstein cattle as well as in layer hens population (Benkel, personal communication, 2014). It was therefore anticipated that the panel would capture a significant amount of the genetic variation associated with fertility in mink. A number of genes including muscle segment homeobox loci (MSX1 and MSX2), nuclear receptor subfamily 5 (NR5A2), PR, pentraxin (PTX3) and SPARC were added to the panel based on the published findings implicating these genes in reproductive physiology in mink and other species (Lydon *et al.*, 1995; Varani *et al.*, 2002; Lefèvre *et al.*, 2011; Cha *et al.*, 2013; Zhang 2013; Bertolin *et al.*, 2014). The resulting marker panel consisted of 316 SNPs distributed over 66 candidate genes. On proprietary grounds, the mouse-sourced genes are encoded A2 to T7. Of the number of SNPs attempted in each gene is shown in Table 3. Based on their involvement in signalling pathways, the 66 genes were grouped into six categories namely mitogen-activated protein (MAP) kinase, circadian rhythm, peroxisome proliferator-activated receptors (PPARs), DNA repair, cytokine, and other. Figure 3 shows the sequential steps involved in the identification of SNPs for the candidate gene-based marker panel.

4.6.1 Gene categories based on signalling pathways

The candidate gene-based marker panel for fertility in mink covered 66 genes. Based on their involvement in signalling pathways, the 66 genes were grouped into six categories namely MAP, circadian rhythm, PPARs, DNA repair, cytokine and others (see Figure 4).

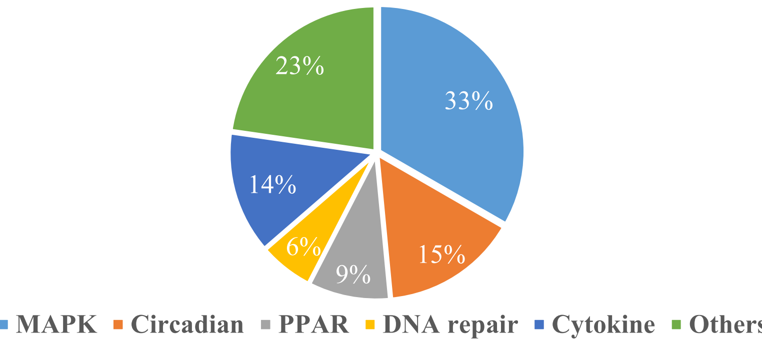


Figure 4. Categories of the candidate genes in the marker panel based on their inferred signalling pathways.

The mitogen-activated protein (MAP) kinase signalling pathways are a conserved set of signalling pathways that control many cellular functions including cell differentiation, cell proliferation and cell death (reviewed by Nishida and Gotoh, 1993). In mammals, there are four signalling cascades of the MAPK family: p38 MAPK protein kinases (Han *et al.*, 1994), c-Jun-N-terminal kinases (JNKs) (Woodgett *et al.*, 1995), extracellular signal-regulated kinases 1 and 2 (ERKs) (Boulton and Cobb, 1991) and ERK5 (Lee *et al.*, 1995). The mammalian circadian timing system coordinates a wide range of complex physiological processes including reproduction. Suprachiasmatic nucleus (SCN) in the hypothalamus is the central rhythm generator in the circadian regulatory network (Hastings *et al.*, 2008). The cellular rhythmicity of the SCN is created by the positive and negative feedback loops which are controlled by a core of oscillator genes and their protein products (Yang, 2010).

PPARs are ligand-activated transcription factors which belong to the nuclear receptor family (Mangelsdorf *et al.*, 1995). Among the three isoforms of PPARs, PPARγ is highly expressed in the ovarian tissues of the rat, pig, sheep and human (Lambe and Tugwood 1996; Komar *et al.*, 2001; Schoppee *et al.*, 2002; Froment *et al.*, 2003), in uterine tissue of

mink (Desmarais *et al.,* 2007*)* and has also been detected in the adipose tissue (Chawla *et al.*, 1994). PPARs regulate metabolic homeostasis in many metabolically active tissues including muscle and adipose tissue (Lee *et al.,* 2006). Apart from controlling metabolic processes, PPARs are also involved in key reproductive functions such as ovarian folliculogenesis, angiogenesis and steroidogenesis (Dupont *et al.,* 2012).

Mammalian cells utilize DNA repair pathways to repair DNA lesions either caused by exogenous agents including UV-light, ionizing radiation and chemicals or cellular endogenous processes including oxidation, alkylation and hydrolysis of bases (reviewed by De Bont and Van Larebeke, 2004). The four major DNA repair pathways and mechanisms are mismatch repair, base excision repair, nucleotide excision repair and double strand break repair (Dexheimer, 2013). These repair pathways are essential to maintain the integrity of the genome as accumulation of large number of lesions can cause tumor growth or programmed cell death.

Cytokines and their specific receptors enable cells to respond to different stimuli and thus, initiate key cellular processes. Signalling through cytokine receptors commonly occurs by a mechanism known as janus kinase (JAK) – signal transducers and activators of transcription (STAT) signalling (O'Shea *et al.,* 2011). Inactive JAK enzymes are attached to the cytoplasmic domains of cytokine receptors. Binding of the cytokine molecules to the specific receptors causes activation of the associated JAKs and also phosphorylates the intracellular tyrosine creating sites for STATs. Further the STAT dimers migrate to the nucleus where they bind to specific promoters and activate transcription of target genes (Rawlings *et al.,* 2004).

The "others" category comprises of candidate genes which do not specifically belong to the above signalling pathways and is a collection of various other pathways and processes.

4.6.1 Genotyping

The animals were typed using the candidate SNP marker panel. Genotyping was performed by Sequenom® Bioscience (now acquired by Agena Bioscience, San Diego, USA). Data was assembled into different genotype clusters (homozygote major, homozygote minor and heterozygote). Genotype and allele frequencies were calculated by direct count.

4.6.1.1 Marker selection

As a part of the primary (1°) QC procedures, Sequenom® Bioscience utilized a SNP genotype clustering algorithm to discriminate between the genotype clusters and identify the appropriate markers. Total number of SNPs attempted was 316. For number of SNPs returned by Sequenom® Bioscience after 1° QC see Table 3.

Once the genotype results were sent by Sequenom® Bioscience, secondary QC procedures were implemented. Call rate was chosen as a filtering criterion to assess markers' integrity. It is a common indicator to determine optimal genotypes for association studies (Moorhead *et al.,* 2006). Markers with call rate less than 70% were eliminated. Monomorphic SNPs were also discarded as they yielded no genotypic differences. Total number of SNPs which passed the secondary QC procedures and were included for the association analysis were 255.

Finally, LD analysis was performed in SNPStats©, Catalan Institute of Oncology, Barcelona, Spain (Solé *et al.,* 2006). If two SNPs are in LD, only one SNP is to be included in the analysis as they serve as proxies for each other (reviewed by Hirschorn and Daly, 2005). For the development of a second generation marker panel, HWE analysis is

encouraged. However, in the resource population, some of the assumptions of a classic Hardy-Weinberg model were not met including infinitely large population, random mating and no migration. SNPs deviating from HWE may also reflect genotyping error (Salanti *et al.,* 2005).

Table 3: Candidate marker panel and primary quality control

Gene category	Code	SNPs attempted	SNPs after 1° QC
MAPK	A2	10	10
	B1	4	4
	C2	3	3
	EGF	8	8
	FGF2	5	4
	FGF9	4	3
	H1	2	2
	M1	5	5
	N1	3	3
	N2	2	2
	N5	16	16
	N8	3	3
	O1	5	5
	P1	3	3
	PR	5	5
	P7	2	2
	P8	5	5
	R1	2	2
	S7	6	4
	SOX5	20	20
	S9	3	3
	S11	6	6
Circadian	A3	7	7
	C3	5	5
	CSNK1E	4	4
	N3	3	2
	N4	3	3
	N6	7	6
	P2	4	3
	PER1	5	3
	P4	5	5
	R3	3	3
PPAR	A1	4	4
	C6	4	3
	P6	5	5
	S2	6	6
	S3	8	8

Gene category	Code	SNPs attempted	SNPs after 1° QC
	S12	3	3
Cytokine	I1	4	2
	I2	4	4
	M2	1	1
	S4	4	4
	S5	6	5
	S6	4	4
	T1	4	4
	TNF	4	3
	T4	7	7
DNA repair	F1	1	0
	M4	3	2
	R2	4	2
	S1	5	4
Others	B2	6	5
	C1	6	4
	C4	3	3
	L1	7	6
	M3	6	6
	MSX1	3	2
	MSX2	4	2
	NR5A2	9	9
	PTX3	3	3
	R4	4	2
	SPARC	6	6
	T3	2	2
	T5	2	2
	T6	2	2
	T7	4	4
Total		316	288

4.7 Association analysis

In this study, SNPAssoc (1.9-2): R package was used to perform the association analysis (González *et al*., 2007). The R package has a greater flexibility for handling data and is an open-source software package (R Development Core Team, 2008). P-value obtained in the descriptive statistics was used as an evaluating criterion to determine significant associations. The package has all the tools to assess genetic association for traits and also perform genetic calculations.

Two approaches were carried out for the association analysis. For the first approach, females were separated into 'low' (0) and 'high' (1) performance groups based on primiparous production (litter size and kit survivability percentage). Based on dietary regimes, MDR and control (CTRL) were analysed as separate groups as a marginally significant effect of diet on kits born was observed ($p = 0.085$). High performers ($n = 98$ controls, $n = 89$ MDR) were the females with 5-9 kits at weaning and kit survivability of at least 75% at weaning, which is considered optimal for economic sustainability in the mink industry. The low performers ($n = 105$ controls, $n = 93$ MDR) weaned either 1-4 kits or 10 or more kits and had kit survivability of less than 75% at weaning. The other approach focused on mink females from the high and low tails of the phenotypic distribution across the population. Generally, for a moderate sized population, 20-30% of the entire resource population is considered as the ideal size for the tails (Gallais *et al.,* 2007). In this study, the contrasting groups of females were identified based on primiparous production (litter size). Based on treatment effect, MDR and CTRL groups were analysed as separate groups and females with contrasting phenotypes were identified from the respective tails. The high performance group comprised of females with 7-9 kits at weaning and the low performers weaned 1-3 kits. For MDR, the total number of females in the high and low tails comprised of 94 females and for CTRL it was 95 females.

Hence, a total of four (window analysis for MDR and CTRL; tail analysis for MDR and CTRL) types of analysis was performed. To increase the confidence in our results and eliminate spurious associations, only SNPs which showed a significant association in at least two of the analyses were considered to be truly significant.

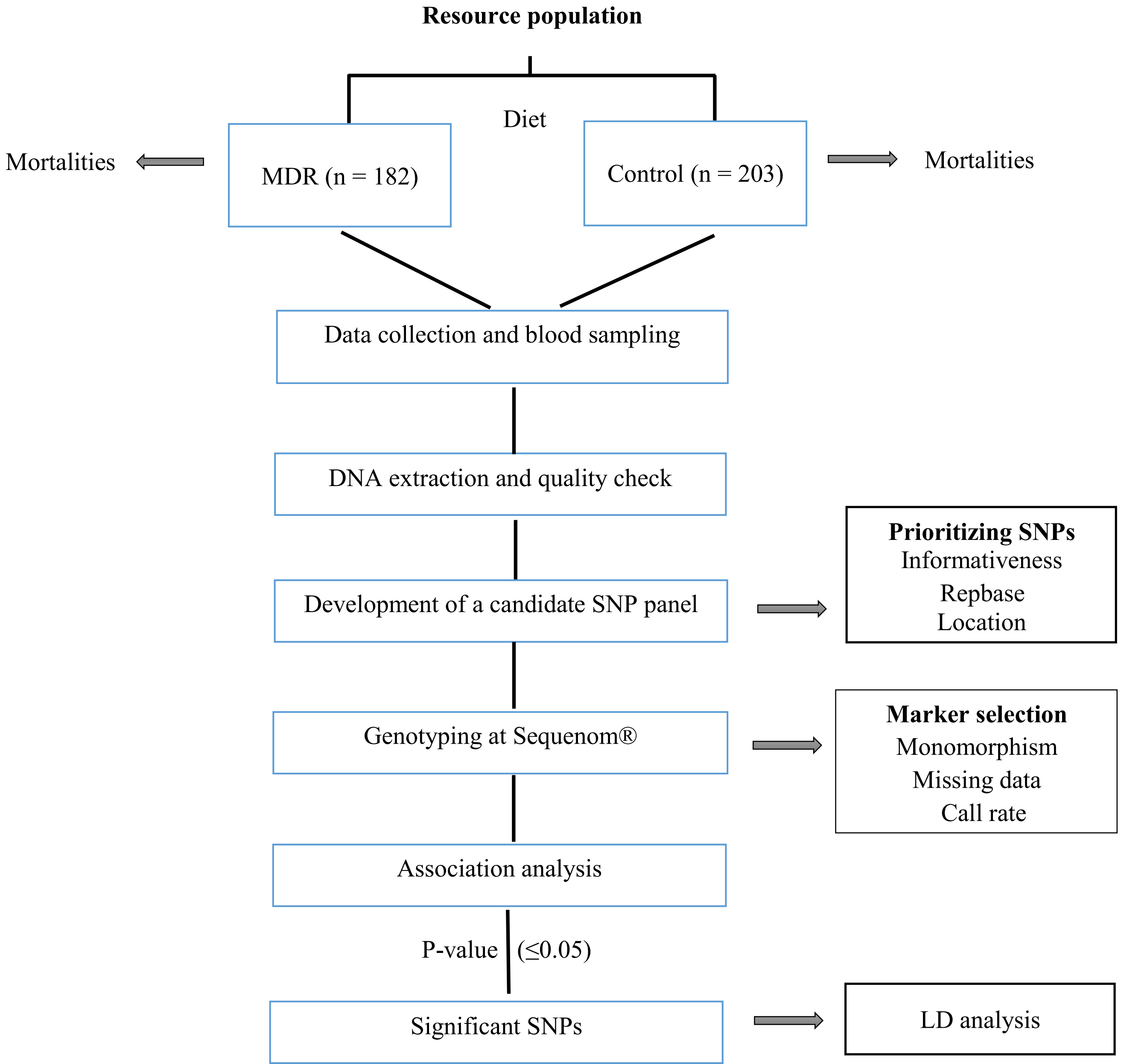


Figure 5. Flowchart outlining the various stages of the project

CHAPTER 5: RESULTS

5.1 Tail analysis

In this approach, mink females were selected from the extremes of the phenotypic distribution. Females in the high and low tails showed performance differential with respect to mean number of kits at birth and weaning across both CTRL and MDR populations (Table 4).

Table 4: Phenotypic performance of females selected for tail analysis in CTRL and MDR

Tail analysis	Mean # kits at birth ± S.E.	Mean # kits at weaning ± S.E.	Percentage survival (%)
CTRL			
Low	3.78 ± 0.39	2.02 ± 0.10	72.93 ± 4.25
High	8.74 ± 0.19	7.70 ± 0.12	89.22 ± 1.69
MDR			
Low	4.15 ± 0.57	2.18 ± 0.12	72.93 ± 5.33
High	9.26 ± 0.29	7.85 ± 0.11	87.72 ± 1.95

5.1.1 MDR group

A total of eleven genes showed significant associations in the MDR group, identified from the high and low tails across the MDR population, at probabilities of less than or equal to 0.05. Four genes showed associations with at least one SNP in each gene at highly significant probabilities of less than or equal to 0.01.

Table 5: SNPAssoc results for the tail analysis in the MDR group

Gene	Gene category	#SNP	SNP Id	P-value
N5	MAPK	6	1	0.019
			3	0.035
			4	0.024
			5	0.043
			7	0.038
			8	0.027
O1	MAPK	1	1	0.036
SOX5	MAPK	1	1	0.006
S11	MAPK	1	1	0.046
T1	Cytokine	1	1	0.023
T4	Cytokine	1	1	0.032
A3	Circadian	2	1	0.008
			2	0.044
C3	Circadian	2	1	0.003
			2	0.006
R2	DNA repair	1	1	0.007
MSX2	Others	1	1	0.028
SPARC	Others	1	1	0.034

#SNP refers to the number of SNPs with significant associations ($p \leq 0.05$) for a gene.

Among the "MAPK pathway" candidate genes, six SNPs in the N5 gene, one SNP in the O1 gene, one SNP in the SOX5 and one SNP in the S11 gene showed significant association at probabilities of less than 0.05 (see Appendix 1: Tables 1, 2, 3, 4, 5, 6, 7, 8). For SNP 3 and SNP 5 in the N5 gene, the T allele was associated with an increase in mean number of kits at weaning and hence, is the favourable allele for fertility. For SNP 7 and SNP 4 in the N5 gene, the G allele was associated with an increase in the mean number of kits at weaning and hence, is the favourable allele for fertility. For SNP 8 and SNP 1 in the N5 gene, the A allele increased the mean number of kits at weaning and hence, is favourably associated with fertility. For SNP 1 in the O1 gene and SNP 1 in the SOX 5 gene, the G allele and the C allele respectively increased the mean number of kits at

weaning and hence, are favourably associated with fertility. For SNP 1 in the S11 gene, the homozygous TT-CC genotype improved the mean number of kits at weaning and hence, homozygous genotypes are favourable for fertility.

Among the "cytokine receptor signalling pathway" candidate genes, one SNP in the T1 gene and one SNP in the T4 gene showed significant associations at probabilities of less than 0.05 (see Appendix 1: Tables 9, 10). For SNP 1 in the T4 gene, the homozygous CC-GG genotype improved the mean number of kits at weaning and hence, homozygous genotypes are favourable for fertility. For SNP 1 in the T1 gene, the T allele was associated with an increase in mean number of kits and hence, is the favourable allele for fertility.

Among the "circadian rhythm signalling pathway" candidate genes, two SNPs in the A3 gene and two SNPs in the C3 gene showed significant associations at probabilities of less than 0.05 (see Appendix 1: Tables 11, 12, 13, 14). For both SNP 1 and SNP 2 in the A3 gene, the T allele increased the mean number of kits at weaning and hence, is favourably associated with fertility. For SNP 1 and SNP 2 in the C3 gene, the G allele and the A alleles respectively were associated with an increase in the mean number of kits at weaning and hence, are the favourable alleles for fertility.

Among the "DNA repair signalling pathway" candidate genes, one SNP in the R2 gene showed significant association at a probability of less than 0.01 (see Appendix 1: Table 15). For SNP 1 in the R2 gene, the homozygous AA-GG genotype improved the mean number of kits at weaning and hence, homozygous genotypes are favourable for fertility.

Among the "others" category of candidate genes, one SNP in the MSX2 gene and one SNP in the SPARC gene showed significant associations at probabilities of less than 0.05 (see

Appendix 1: Tables 16, 17). For SNP 1 in the MSX2 gene and SNP 1 in the SPARC gene, the T allele and the C allele respectively were associated with an increase in the mean number of kits at weaning and hence, are the favourable alleles for fertility.

5.1.2 CTRL group

A total of seven genes showed significant associations in the CTRL group identified from the high and low tails across the control population at a probability of less than or equal to 0.05. Two genes showed association with atleast one SNP in each gene at highly significant probabilities of less than or equal to 0.01.

Table 6: SNPAssoc results for the tail analysis in the CTRL group

Gene	Gene category	#SNP	SNP Id	P-value
N5	MAPK	8	1	0.038
			2	0.012
			4	0.016
			5	0.0002
			6	0.033
			7	0.002
			8	0.014
				0.016
O1	MAPK	1	1	0.042
S11	MAPK	1	1	0.044
P6	PPAR	3	1	0.048
			2	0.022
			3	0.026
C3	Circadian	1	1	0.009
NR5A2	Others	1	1	0.018
SPARC	Others	1	1	0.031

#SNP refers to the number of SNPs with significant associations ($p \leq 0.05$) for a gene.

Among the “MAPK signalling pathway” candidate genes, eight SNPs in the N5 gene, one SNP in the O1 gene and one SNP in the S11 gene showed significant associations at probabilities of less than 0.05 (see Appendix 1: Tables 18, 19, 20, 21, 22, 23, 24, 25, 26,

27). For SNP 5 and SNP 3 in the N5 gene, the T allele increased the mean number of kits at weaning and hence, is favourably associated with fertility. For SNP 1 and SNP 2 in the N5 gene, the C allele were associated with an increase in mean number of kits at weaning and hence, is the favourable allele for fertility in the contrasting group of mink females. For SNP 4 and SNP 6 in N5 gene, the G allele increased the mean number of kits at weaning and hence, is the favourable allele for fertility. For SNP 8 in the N5 gene, the A allele increased in the mean number of kits at weaning and hence, is favourably associated with fertility. For SNP 7 in the N5 gene, the homozygous AA-GG genotype improved the mean number of kits at weaning and hence, homozygous genotypes are favourable for fertility. For SNP 1 in the O1 gene and SNP 1 in the S11 gene, the homozygous AA-GG and TT-CC genotypes respectively, improved the mean number of kits at weaning and hence, homozygous genotypes are favourable for fertility.

Among the "PPAR signalling pathway" candidate genes, three SNPs in the P6 gene showed significant associations at probabilities of less than 0.05 (see Appendix 1: Tables 28, 29, 30). For SNP 1, SNP 3 and SNP 2 in the P6 gene, heterozygous GT, CT and AT respectively, were associated with an increase in mean number of kits at weaning. Therefore, heterozygous advantage for fertility is evident here.

Among the "circadian rhythm signalling pathway" candidate genes, one SNP in the C3 gene showed significant association at a probability of less than 0.01 (see Appendix 1: Table 31). For SNP 1 in the C3 gene, the G allele was associated with an increase in mean number of kits at weaning and hence, is the favourable allele for fertility.

Among the "others" category of candidate genes, one SNP in the NR5A2 gene and one SNP in the SPARC gene showed significant associations at probabilities of less than 0.05

(see Appendix 1: Tables 32, 33). For SNP 1 in the NR5A2 gene, the heterozygous CT genotype is associated with an increase in mean number of kits at weaning. Therefore, heterozygous advantage for fertility is evident here. For SNP 1 in the SPARC gene, the C allele increased the mean number of kits at weaning and hence, is the favourable allele associated with fertility.

5.2 Window analysis

The mean number of kits at birth are almost identical for high and low performers across both CTRL and MDR populations. However, the mean number of kits at weaning between the groups exhibit performance differential (Table 7).

Table 7: Phenotypic performance of females selected for window analysis in CTRL and MDR

Window analysis	Mean # kits at birth ± S.E.	Mean # kits at weaning ± S.E.	Percentage survival (%)
CTRL			
Low	6.37 ± 0.46	2.49 ± 0.17	42.98 ± 2.75
High	7.35 ± 0.14	6.65 ± 0.11	91.40 ± 0.86
MDR			
Low	7.44 ± 0.48	3.09 ± 0.17	44.17 ± 2.42
High	7.45 ± 0.16	6.71 ± 0.12	91.35 ± 0.93

5.2.1 MDR group

A total of seven genes showed significant associations in the MDR group at probabilities of less than or equal to 0.05.

Table 8: SNPAssoc results for the window analysis in the MDR group

Gene	Gene category	#SNP	SNP Id	P-value
N5	MAPK	3	2	0.040
			5	0.049
			6	0.033
SOX5	MAPK	1	1	0.019
T1	Cytokine	1	1	0.047
T4	Cytokine	1	1	0.026
C4	Others	1	1	0.041
MSX2	Others	1	1	0.050
SPARC	Others	1	1	0.019

#SNP refers to the number of SNPs with significant associations ($p \leq 0.05$) for a gene.

Among the "MAPK signalling pathway" candidate genes, three SNPs in the N5 gene and one SNP in the SOX5 gene showed significant associations at probabilities of less than 0.05 (see Appendix 1: Tables 34, 35, 36, 37). For SNP 5 and SNP 6 in the N5 gene, the T allele and the G allele respectively, increased the likelihood of becoming a high performer and hence, are the favourable alleles for fertility. For SNP 2 in the N5 gene, the C allele increased the likelihood of becoming a high performer and hence, is favourably associated with fertility. For SNP 1 in the SOX5 gene, the homozygous CC-TT genotype increased the likelihood of becoming a high performer and hence, homozygous genotype is favourably associated with fertility.

Among the "cytokine receptor signalling pathway" candidate genes, one SNP in the T1 gene and one SNP in the T4 gene showed significant associations at probabilities of less than 0.05 (see Appendix 1: Tables 38, 39). For SNP 1 in the T1 gene, the T allele increased the likelihood of becoming a high performer and hence, is the favourable allele for fertility. For SNP 1 in T4 gene, the homozygous CC-GG genotype increased the likelihood of

becoming a high performer and hence, homozygous genotypes are favourably associated with fertility.

Among the "others" category of candidate genes, one SNP in the C4 gene, one SNP in the MSX2 gene and one SNP in the SPARC gene showed significant associations at probabilities of less than 0.05 (see Appendix 1: Tables 40, 41, 42). For SNP 1 in the C4 gene, a single copy of T allele increased the likelihood of becoming a high performer and hence, is the favourable allele for fertility. For SNP 1 in the MSX2 gene, the T allele increased the likelihood of becoming a high performer and hence, is the favourable allele for fertility in the MDR diet group of mink females. For SNP 1 in the SPARC gene, the homozygous CC-TT genotype increased the likelihood of becoming a high performer and hence, homozygous genotype is favourably associated with fertility.

5.2.2 CTRL group

A total of six genes showed significant associations in the CTRL group at probabilities of less than 0.05. Two genes showed associations with atleast one SNP in each gene at highly significant probabilities of less than or equal to 0.01.

Table 9: SNPAssoc results for window analysis in the CTRL group

Gene	Gene category	#SNP	SNP Id	P-value
P6	PPAR	3	1	0.007
			2	0.042
			3	0.017
A3	Circadian	2	1	0.022
			2	0.011
C3	Circadian	1	2	0.050
R2	DNA repair	1	1	0.025
C4	Others	1	1	0.039
NR5A2	Others	1	1	0.050

#SNP refers to the number of SNPs with significant associations ($p \leq 0.05$) for a gene.

Among the “PPAR signalling pathway” candidate genes, three SNPs in the P6 gene showed significant associations at probabilities of less than 0.05 (see Appendix 1: Tables 43, 44, 45). For SNP 1, SNP 2 and SNP 3 in the P6 genes, the heterozygote CT, AT and CT genotypes respectively, increased the likelihood of becoming a high performer. Therefore, heterozygous advantage for fertility is evident here.

Among the “circadian rhythm signalling pathway” candidate genes, two SNPs in the A3 gene and one SNP in the C3 gene showed significant associations at probabilities of less than 0.01 (see Appendix 1: Tables 46, 47, 48). For SNP 1 and SNP 2 in the A3 gene, the heterozygous CT genotype increased the likelihood of becoming a high performer. Therefore, heterozygous advantage for fertility is evident here. For SNP 2 in the C3 gene, the A allele was associated with an increase in the mean number of kits at weaning and hence, is the favourable allele for fertility.

Among the “DNA repair signalling pathway” candidate genes, one SNP in the R2 gene showed significant association at a probability of less than 0.05 (see Appendix 1: Tables

49). For SNP 1 in the R2 gene, the homozygous AA-GG genotype increased the likelihood of becoming a high performer and hence, homozygous genotypes are favourably associated with fertility.

Among the “others” category of candidate genes, one SNP in the C4 gene and one SNP in the NR5A2 gene showed significant associations at probabilities of less than 0.05 (see Appendix 1: Tables 50, 51). For SNP 1 in the C4 gene, the C allele increased the likelihood of becoming a high performer and hence, is the favourable allele for fertility. For SNP 1 in the NR5A2 gene, the heterozygote CT genotype increased the likelihood of becoming a high performer. Therefore, heterozygous advantage for fertility is evident here.

5.5 Summary of tail and window analysis

Table 10: Overlap of significant genes in tail and window analysis

Tail analysis		**Window analysis**	
MDR	CTRL	MDR	CTRL
N5	N5	N5	P6
O1	O1	SOX5	A3
SOX5	S11	T1	C3
S11	P6	T4	R2
T1	C3	C4	C4
T4	NR5A2	MSX2	NR5A2
A3	SPARC	SPARC	
C3			
R2			
MSX2			
SPARC			

In the tail analysis, five genes (N5, O1, S11, C3 and SPARC) are significantly associated in both MDR and CTRL groups and these genes fall into three signalling pathway categories as follows: MAPK (3 genes); circadian rhythm (1 gene); and others (1 gene). In

the window analysis, only one gene (C4) is significantly associated in both MDR and CTRL groups and this gene fall into the others category. Irrespective of the analytical approached used, a core of three genes (C3, N5 and SPARC) is implicated in three out of the four analyses, forming a robust set of markers.

5.6 Reproductive performance of sister mink

Irrespective of the analytical approaches, there were one hundred and three (103) sister pairs involved in the analysis. One of the females in the sister pair was fed *ad libitum,* hence, she was placed in the CTRL group. While the other female was fed ~20% less than the CTRL from September to December, hence, she was placed in the MDR group. It can be inferred from the graph that in spite of being fed two different diets, females in the sister pair exhibited almost similar reproductive performance (Figure 6). This trend is consistent across many of the sister pairs.

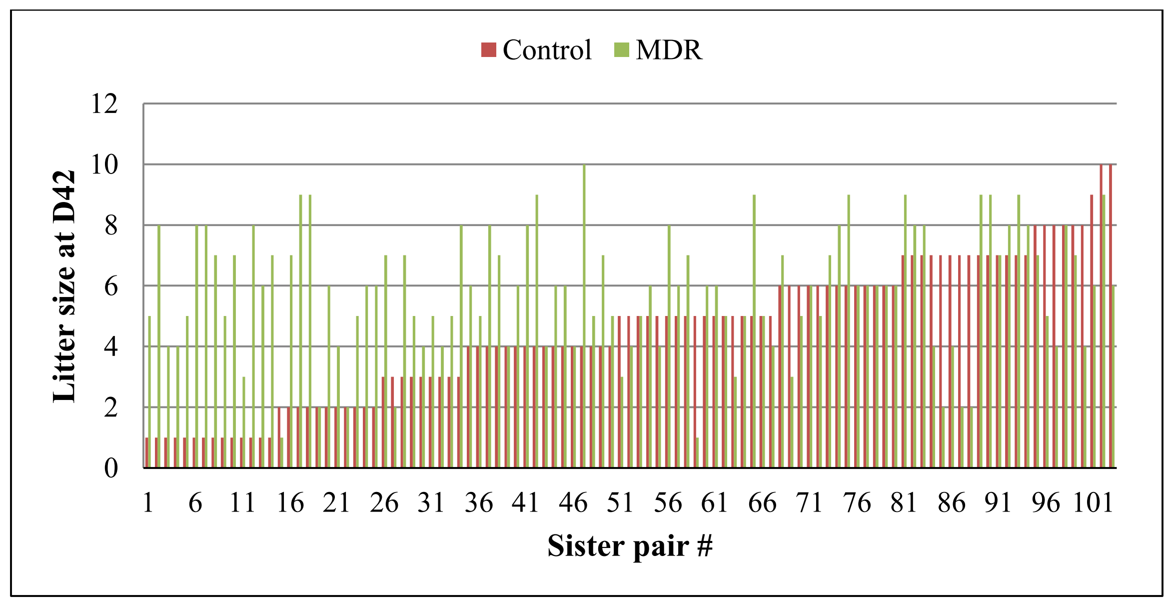
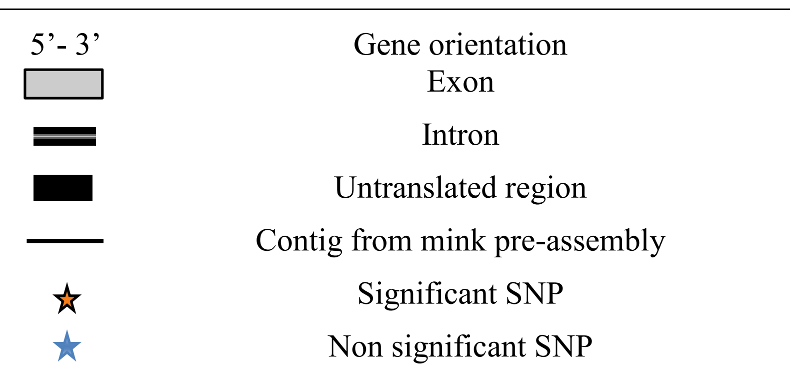


Figure 6. Graph of the reproductive performance of sister pairs in CTRL and MDR diet groups

5.7 SNPs distribution in the significant genes

Gene sequences of *Canis lupus* or *Mustela putorius furo* were retrieved from UCSC genome browser as both dog and ferret are considered to be close relatives of mink (Anistoroaei and Christensen, 2006; Anistoroaei *et al.,* 2009). Using BLAST tool in NCBI+, the sequences were compared with the mink contig database.



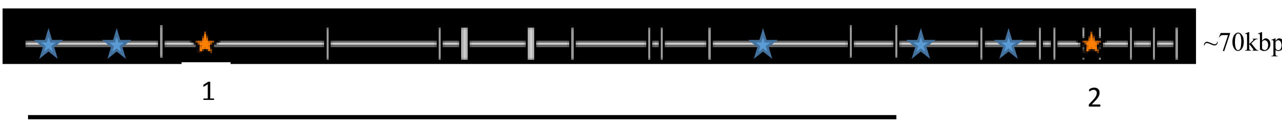


Figure 7. Virtual mapping of mink co-ordinates with A3 gene in *Canis lupus*.

The figure indicates that A3 gene in *Canis lupus* comprises of twenty exons. There are seven SNPs identified in this gene. Among them, there are two intronic SNPs which showed a significant association at a probability of less than 0.05 in both window and tail analysis for the CTRL and MDR population.

In the above Figure 6 the significant SNPs are interspersed among the non-significant SNPs. A possible explanation can be that the recent DNA variants are superimposed on the core group of ancient DNA variants. Hence, distinct blocks aren't evident and SNPs are inherited as one unit instead. In some genes (see Appendix 2: Figure 1, 2, 3, 4 and 5), the

SNPs are inherited as two distinct blocks. The first block comprises of the significant SNPs and the second block comprises of the non-significant SNPs. For other genes (see Appendix 2: Figure 9, 10, 11, 12 and 13) among the total number of SNPs (≥5) identified in the respective genes, one SNP showed significant association. This might be a spurious association and/or at a stringent p-value threshold it might no longer exhibit significance. Hence, validation of such findings on a larger population is encouraged.

5.8 Linkage disequilibrium (LD) analysis

Linkage disquilibrium patterns could be influenced by many factors including natural selection, recombination, mutation and genetic drift. Natural selection is a non-random process where the traits become either frequent or rare in a population due to differential reproduction leading to inflated LD. Recombination leads to breakdown of LD patterns and therefore the extent of LD is expected to be inversely proportional to the local recombination rate (Ardlie *et al*., 2002). Similarly high mutation rates are correlated with little or no LD (Ardlie *et al*., 2002). Genetic drift describes the uneven distribution of allele frequency due to random sampling of gametes over the successive generations and increased drift tends to increase LD patterns (Ardlie *et al*., 2002).

In this study, LD analysis was performed on the significant set of genes only if more than one SNP in the same gene showed significant association at a probability of less than or equal to 0.05. Based on this criteria, four such genes were identified: N5 (8 SNPs), P6 (3 SNPs) A3 (2 SNPs) and C3 (2 SNPs). Correlation between a pair of SNPs was determined by R-square which is a quantitative measure of LD (Shifman *et al*., 2003). It ranges from 0 when loci are not in complete LD to 1 which is a case of perfect LD. In case of perfect LD, information obtained from one marker is exactly similar to the other and hence, one

of the markers is eliminated. Usually, R squared-values above 1/3 indicate sufficiently strong LD to be useful for association mapping. Therefore, R square values of >0.75 were considered to be in strong LD. Based on R squared values, SNPs in A3 and C3 genes were found to be not in LD. For N5 gene, SNP 5 (see Appendix 2: Figure 6) can be eliminated on the basis of LD analysis. For P6 gene, SNP 1 (see Appendix 2: Figure 7) can be eliminated on the basis of LD analysis. Information on LD patterns could be useful in prioritizing SNPs for genotyping. Such data will be helpful for the development of the second generation marker panel.

CHAPTER 6: DISCUSSION

6.1 Approaches for detecting association between performance and genetic elements

6.1.1 Tail analysis

A typical approach for a pathway candidate gene marker study would involve splitting the resource population into three groups based on performance, i.e. a 'high' performance group consisting of those individuals in top 10-20% of the performance distribution (the 'high' tail), a 'low' performance group consisting of those individuals in bottom 10-20% of the performance distribution (the 'low' tail) (Gallais *et al.,* 2007). The remaining 80-60% of the individuals occupying the intermediate space in the performance distribution curve. Assessment of allele frequency differences is initially limited to a comparison of the 'high' and 'low' tail animals in order to maximise the allele frequency differences between the groups if there is a genetic component associated with the trait.

Such a comparison was carried as part of this study where the high performance group comprised of females with 7-9 kits at weaning and the low performers weaned 1-3 kits. This selection of individuals from the extremes of the phenotypic distribution increases the power of the analysis to identify the genetic elements significantly associated with the trait of interest. Table 4 shows the performance differential among the mink females in the high and low tails with respect to the mean number of kits at birth and weaning across both CTRL and MDR populations.

A total of eleven genes (N5, O1, SOX5, S11, T1, T4, A3, C3, R2, MSX2, and SPARC) showed significant association with fertility, identified from the high and low tails across the MDR population. Also, a total of seven genes (N5, O1, S11, P6, C3, NR5A2, and SPARC) showed significant associations in the CTRL group identified from the high and

low tails across the CTRL population. The genetic elements identified from this approach represents the genes which are critical during weaning stage as measurements at D42 (weaning point) were included in the analysis.

Reproductive performance is a complex multifactorial trait (Abegaz *et al.*, 2002; Chebel *et al.*, 2007; Castellini *et al.*, 2010) and involves a number of genes. It is therefore, expected that certain set of genes would be significant at a specific stage of the reproductive cycle. The various stages include folliculogenesis, ovulation, fertilization, pregnancy establishment, embryogenesis, whelping, lactation, and weaning. Specific reproductive features exhibited by mink females including, delayed implantation and embryonic diapause further contributes towards the complexity of this trait. Therefore, it can be suggested that a better methodology would be to reanalyse the resource population using the tail approach only, but fragmented into at least two phases, i.e. number of kits at birth and number of kits at weaning. The expectation is that such an approach will allow the identification of genetic elements essential during the birth-to-weaning phase separately, at least in part, from genes that are more important prior to parturition.

6.1.2 Window analysis

In addition to the high tail/low tail comparison, a second analysis was also performed, hereafter known as the 'window' approach, which involved splitting the resource population into 'low' (0) and 'high' (1) performance groups based on primiparous production (litter size and kit survivability percentage). High performers weaned 5-9 kits and had kit survivability of at least 75% at weaning. The low performers were the females with either 1-4 kits or 10 or more kits at weaning and kit survivability of less than 75% at weaning.

This classification was done for consistency with industry practices and recognition that large litters impose unsustainable lactation burdens on mother due to their limited number of teats and their limited capacity to nurture. For this study, litter size at weaning measures the reproductive performance in the mink females. Ideally, a mink female that weaned a larger litter (above a certain threshold) would be classified as a high performer. Contrary to this rationale, based on the window approach, mink females which weaned 10 or more kits at weaning were classified as the low performers and hence, this could be misleading and inappropriate. Based on this approach, mink females with high, low and mediocre reproductive performance were selected which would in turn, dilute the performance differential. Hence, comparison of allele frequency differences among the groups seems to be conceptually incorrect.

Table 7 shows the means for the number of kits at birth are almost identical for the high and low performers across both CTRL and MDR populations. The mean number of kits at weaning between the performance groups exhibited some separation which justifies our rationale behind using this approach. However, there were only few mink females that weaned 10 or more kits which can be a plausible reason for the performance differential observed at weaning.

The window analysis identified a total of seven genes (N5, SOX5, T1, T4, C4, MSX2, and SPARC) which showed significant associations in the MDR group and six genes (P6, A3, C3, R2, C4, and NR5A2) showed significant associations in the CTRL group at probabilities of less than 0.05. Based on the performance differential in the numbers of kits at weaning, the window analysis identified genes important during the birth-to-weaning phase.

6.2 Comparison between MDR and CTRL groups

In the tail analysis, an overlap of five significant genes was observed between the MDR and CTRL group. A total of six MDR specific genes and two CTRL specific genes were identified. A considerable overlap of significant genes between the two dietary treatment groups was observed. Additionally, Figure 6 shows the individual performance of sister mink on the two diets. It was observed from the scatter plot that irrespective of the dietary regime, the sister pairs perform in unison. This is a trend consistent across many of the sister pairs. Based on the overlap of genes among the two dietary groups and exhibition of similar reproductive performance among the sister pairs, it can be suggested that the underlying genetics contribute significantly towards the variability of the trait. Although, the effect of diet on reproductive performance cannot be ignored completely.

6.3 Model 1

Our results identified fourteen genes that were significantly associated with reproductive performance in mink females. Based on the available literature, nine out of the fourteen genes are involved in at least one of the three signalling pathways namely, the circadian rhythm pathway, the PPARs pathway and the MAPK pathway.

The A3 and C3 genes are involved in the circadian rhythm pathway. The P6 and S11 genes are involved in the PPAR signalling pathway. The N5, SOX5 and T4 genes are involved in the MAPK signalling pathway. The remaining five genes (C4, MSX2, NR5A2, O1, SPARC and T1) do not belong to the pathways mentioned above but studies report that these genes regulate key processes involved in mammalian female fertility. The role of the R2 gene in mammalian fertility was not found in literature. However, the R2 gene is involved in the DNA repair pathway and reviews have concluded that DNA repair mechanisms are involved in mammalian gametogenesis (Baarends *et al.,* 2001). Therefore, it could be suggested that the R2 gene plays an important role in repair of DNA damage in the developing germline cells and hence, could affect normal reproductive functions.

Model 1 is a representation of the proposed mechanisms by which the identified three key pathways (circadian rhythm, PPAR and MAPK) influence the critical reproductive processes in female mink. The model also highlights the role of the signalling pathways in controlling the reproductive functions of the uterus. Apart from the pathways, the individual functional roles of the significant genes (C4, MSX2, NR5A2, O1, SPARC and T1) in mammalian fertility is also illustrated as shown in Figure 8.

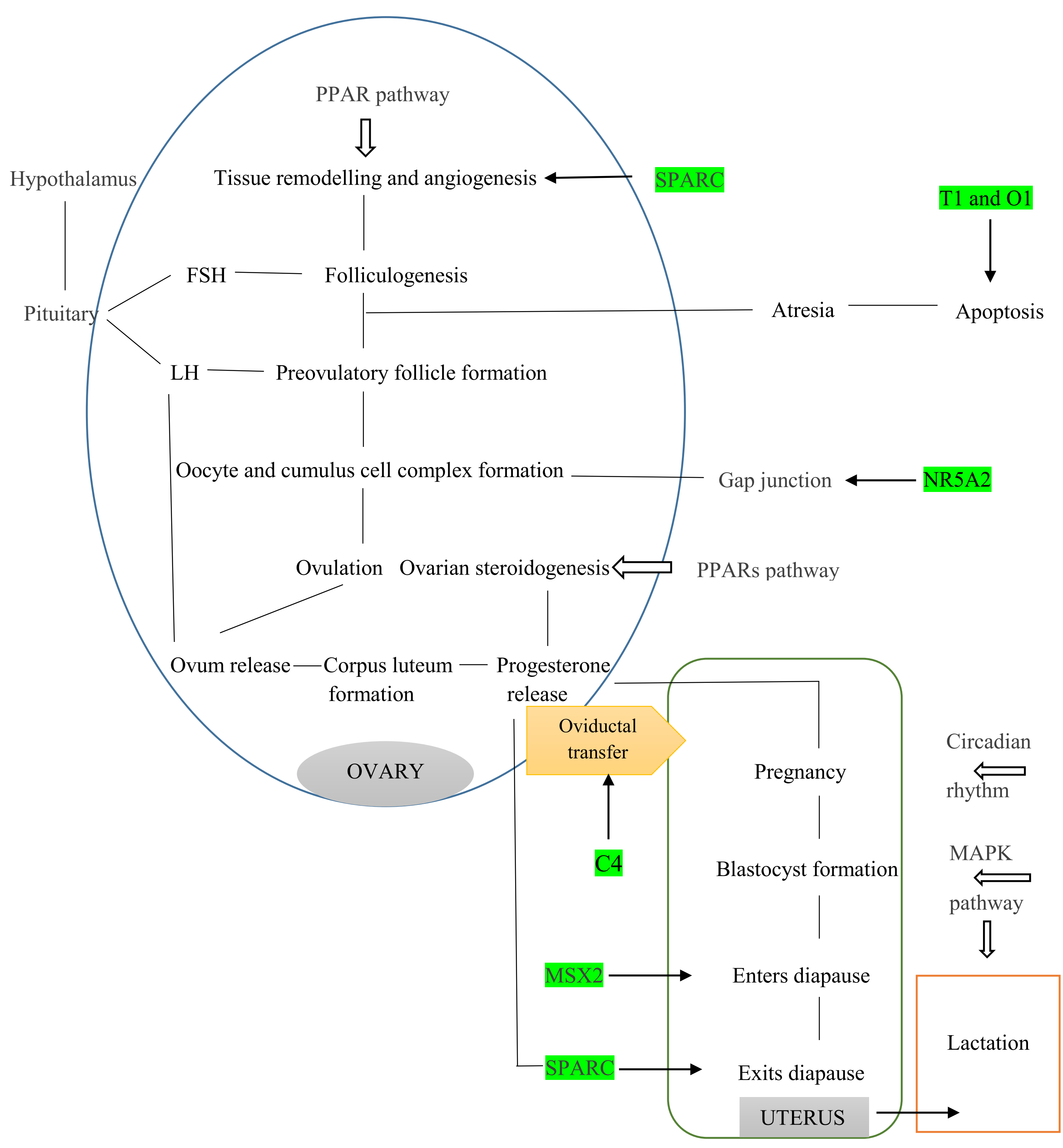


Figure 8: Representation of the proposed mechanisms by which PPAR, MAPK and circadian rhythm pathways influence the critical reproductive processes in female mink. The model also shows the functional roles of the C4, MSX2, NR5A2, O1, SPARC and T1 genes in mammalian fertility and are highlighted in green.

6.3.1 Folliculogenesis and ovulation

In mink, gonadotrophs in the anterior pituitary produce the two gonadotropins: FSH and LH (Murphy and James, 1976). At the beginning of the ovulatory cycle, FSH secretion from the pituitary stimulates a cohort of follicles to grow and enter the preovulatory stage. The FSH signal also stimulates the granulosa cells of the preovulatory follicle to synthesize the LH receptors (reviewed by Chappel and Howles, 1991) which are functionally coupled with aromatase, the enzyme which converts androgens to estrogen. Studies have reported that activation of the PPARs can disturb the ovarian follicular dynamics by reducing the aromatase activity as observed in the mouse ovarian granulosa cells (Toda *et al.*, 2003). LH released from the pituitary acts on the receptors present on the thecal cells and granulosa cells to direct the release of estrogen. This estrogen acts as a negative feedback for the hypothalamus-pituitary axis and supresses the release of FSH (reviewed by Hillier, 1994). Hence, only one dominant follicle continues to mature and ovulate from the ovarian follicle pool. The remaining non-ovulatory follicles degenerate and undergo atresia by a process called apoptosis (Hughes and Gorospe, 1991; Tilly *et al.,* 1991). Two of the identified significant genes (O1 and T1) are associated with apoptosis and hence, it could be hypothesised that they play a role in follicular atresia.

NR5A2 (orphan nuclear receptor subfamily 5, group A, member 2) is another significant gene identified in the analysis. It is a critical regulator of biological mechanisms essential for maintenance of female fertility (Duggavathi *et al.,* 2008). Highest levels of NR5A2 are expressed in the granulosa and luteal cells of the ovary (Hinshelwood *et al.,* 2005). Within the preovulatory follicle, the cumulus granulosa cells are tightly connected to each other and to the oocyte (Pedersen and Peters, 1968). Following the upsurge of LH, cumulus cells

produce an extracellular matrix which causes cumulus expansion and results in the expulsion of the cumulus-oocyte-complex (COC) (Rajkovic *et al.,* 2006). Prostaglandins and hyaluronic acid are critical for this cumulus expansion (Camaioni *et al.,* 1993; Lim *et al.,* 1997). In the absence of NR5A2, the prostaglandin and hyaluronan pathways are compromised which leads to defective cumulus expansion (Duggavathi *et al.,* 2008) as well as, failure to ovulate.

6.3.2 Ovarian tissue remodelling and angiogenesis

The ovarian tissue, particularly the surface epithelium, the granulosa and thecal cell layers continuously undergoes remodelling to accommodate the varying sizes of the growing follicles. These processes are critical for follicular development, ovulation and follicular-luteal transition. PPARs regulate the transcriptional activity of proteases including matrix metalloproteinase and plasminogen activator which are responsible for tissue remodelling and angiogenesis in the ovary (Sang, 1998; Liu, 1999). Apart from proteinases, PPARs also modulate the expression of angiogenic factors such as vascular endothelial growth factor (VEGF) which regulates the ovarian vasculature for formation of new blood vessels during folliculogenesis (Yamakawa *et al.*, 2000).

Along with PPARs, studies suggest SPARC as a primary regulatory gene involved in ovarian tissue remodelling (Joseph *et al.*, 2004). Expression of SPARC is regulated by the angiogenic factor VEGF in bovine luteal cells (Joseph *et al.*, 2004), which has been reported to be controlled by PPARs as stated above. Knowledge of the working of PPARs and SPARC in the reproductive processes is sparse in livestock and is therefore, an area that warrants further investigation to gain better insights into mammalian fertility.

6.3.3 Oviductal transport of embryos

Once fertilisation has taken place in the ampulla of the oviduct, the fertilized ovum travels through the oviduct to reach the uterine lumen for implantation. Smooth muscle contractions and oscillations of the cilia present on the oviductal epithelium facilitates this passage of embryos through the oviduct-uterine junction (Lyons *et al.,* 2006). Timely oviductal transport is crucial for a normal implantation process as failure of this transport results in embryo retention within the oviduct. Genes such as C4 play an important role in the transport of embryos through the oviduct into the uterine cavity and hence, influence reproductive success in females (reviewed by Sun and Dey, 2012).

6.3.4 Establishment and maintenance of pregnancy

The MAPK signalling pathway is important for the establishment of pregnancy. The role of MAPK pathway in determining reproductive success is evidenced in a study by Madan *et al.* (2005) who found that this pathway is involved in preimplantation embryogenesis in bovine species (Madan *et al.*, 2005). The study reported that blockage of the MAPK signalling pathway during bovine embryo development inhibited blastocyst formation, which is an indicator of embryonic health.

Following the release of the ovum, a corpus luteum develops from the follicular luteal cells. Stimulation by LH is essential for progesterone biosynthesis from cholesterol (Niswender, 2002). NR5A2 deficiency in mice causes luteal insufficiency and reduced progesterone synthesis is observed which causes implantation failure (Bertolin *et al.,* 2014). Although administration of exogenous progesterone restores implantation, embryo overcrowding, compromised placental formation, and fetal growth retardation are frequently observed (Zhang *et al.,* 2013). A study by Schoonjans et al. (2002) reported that reduction in the

ovarian progesterone level is a result of disruption in the ovarian steroidogenesis which requires cholesterol as a substrate. One of the major sources of cholesterol for this process is supplied by importation via by genes such as receptor-B1 (SR-B1). The same study also found that NR5A2 controls the expression of SR-B1 which governs cholesterol uptake by ovarian cells and hence, influences ovarian steroidogenesis.

Apart from NR5A2 and MAPK pathway, circadian rhythm also coordinates multiple processes that are involved in successful establishment of pregnancy. Miller *et al.* (2004) found that pregnant mice with a mutation in one of the core circadian genes displayed increased fetal reabsorption rate and higher rate of full-term pregnancy failure compared to the wild type females. Extended or non-productive labor and failure to enter labor were common complications in delivery which were observed in the mutant females (Miller *et al.,* 2004). Summa *et al.* (2012) demonstrated that environmental disruption in the circadian rhythm caused a substantial decrease in pregnancy success in mice. What is prominent in the study is that a 6-hour phase advance in the light cycle led to a significant decrease in the number of productive matings. In another study, targeted disruption of one of the core circadian genes in mice caused multiple reproductive insufficiencies including reduced corpora lutea formation, abnormal blastocyst formation and failure to successfully implant (Boden, 2008). Such studies indicate that disruptions in the endogenous timing system lead to multiple reproductive complications. However, further research is encouraged to identify the molecular mechanisms and genes involved in maintenance of circadian rhythm in mink.

6.3.5 Embryonic diapause and embryo reactivation

The MSX2 gene is one of the identified significant genes associated with fertility in female mink. MSX genes have been reported to be crucial for establishment and maintenance of embryonic diapause (Cha *et al.,* 2013). Uterine inactivation of MSX genes in mice failed to initiate diapause and also resume blastocyst reactivation (Cha *et al.,* 2013). In other unrelated diapausing mammals, including the American mink, similar results were observed (Cha *et al.,* 2013).

Uterine modifications are associated with the re-emergence of the blastocyst from embryonic diapause. Study by Lefèvre *et al.* (2011) reported SPARC as a potential stimulator of uterine modifications associated with blastocyst reactivation following embryonic diapause in mink. Their study revealed that the expression of SPARC was significantly elevated in the uterus at blastocyst reactivation compared to the diapause period. It was suggested that the upregulation of SPARC resulted from progesterone secretion due to reactivation of the corpus luteum (Lefèvre *et al.,* 2011). SPARC is also one of the significant genes identified in this research project and compelling evidence establishes SPARC as a key gene involved in embryo reactivation and hence, determining female reproductive fitness in mink.

6.3.6 Lactation and maternal behaviour

Alveolus is the functional unit of the mammary gland which is responsible for production and secretion of milk. Genes involved in the MAPK signalling pathway have been implicated with alveolar cell maintenance during pregnancy and lactation (Buono *et al.,* 2006). Targeted disruption of such genes in the mammary gland of lactating mice impairs alveolar mammary gland development and such mice are incapable of feeding their pups

(Raafat *et al.,* 2011). A significant increase in MAPK controlled genes expression has been observed in the early months of lactation in bovines (Bionaz and Loor, 2008). Apart from lactation, studies in mice have also revealed the role of MAPK signalling cascade in controlling maternal behaviour towards her pups (Kuroda *et al.,* 2007; Satoh *et al.,* 2011). These studies suggest that MAPK signalling pathway is crucial for mammary gland development, milk synthesis and kit-directed behaviour in female mammals. However, such studies have not been attempted in mink so far.

6.4 Genes associated with reproduction in other (livestock) species

In any agricultural species, reproductive performance is a determinant of breeding herd efficiency. In the past few years, genes and signalling pathways that regulate the complex reproductive mechanisms have been identified.in livestock using the candidate gene approach.

In a study conducted in Holstein dairy cattle, polymorphism in the fibroblast growth factor (FGF2) gene was significantly associated with the survival rate of embryos (Khatib *et al*., 2008). In the present study, FGF2 was incorporated in the candidate gene marker panel due to its known role in embryonic survival and fertilization rate. However, association analysis did not reveal any significant associations of the genetic variants in FGF2 with fertility in the mink resource population. Another study determined the association of polymorphisms in the STAT proteins with embryonic survival and fertilization success in cattle (Khatib *et al*., 2008). In agreement with the findings, our study also revealed significant associations of genetic variants in STAT proteins with mink fertility. Lonergan et al. (1996) reported the role of epidermal growth factor (EGF) in bovine oocyte maturation by affecting the cumulus-oocyte-complex formation. Hence, it was expected that genetic variants in EGF

might be associated with fertility in mink as well. In contrast, our research failed to identify any such associations with EGF in the mink experimental population. Studies in cattle populations have also established the association of signalling pathways including circadian rhythm with reproductive success (Cochran *et al*., 2013). This confirms our findings where genes involved in circadian rhythm pathway have shown significant association with fertility in mink resource population.

Association analysis has also identified growth differentiation factor (GDF-9) and bone morphogenetic proteins (BMPs) genes as key candidates associated with fertility in different breeds of sheep including Cambridge and Belclare (Hanrahan *et al*., 2004). Candidate gene analysis has identified associations of polymorphisms in estrogen receptor and pre-implantation protein genes with litter size in pigs (Rothschild *et al*., 1996; Niu *et al*., 2006). These fertility associated genes were not a part of the candidate gene marker panel used in this study. Hence, the effect of polymorphisms in these genes was not investigated in the mink resource population. It is, however, recommended that these genes could be potential candidates for the development of a second generation marker panel for mink fertility in future.

CHAPTER 7: CONCLUSION

In summary, the study resulted in the development of a candidate gene marker panel, the development of a database of fertility performance records for the CCFAR experimental population and identification of genetic elements underlying reproductive performance in the CCFAR Standard Black mink herd. The study identified SNPs in fourteen genes found to be significantly associated with fecundity in mink females. Published literature documents the involvement of these genes in female reproductive physiology which further strengthens the findings of this study.

As far as the current study is concerned, the association analysis provides preliminary information required for the calculation of genomic estimated breeding values (gEBVs). However, due to time constraints sample collection from a "validation" population couldn't be performed, which is necessary to assess the power of the marker panel in predicting performance of individuals. The gEBVs can be potentially incorporated as a selection tool to identify animals with better reproductive performance, resulting in substantial economic return for the ranchers. As a successful example, the dairy industry has been using (gEBVs) as a selection tool for the Holstein dairy cattle.

Although this study focused on female reproductive performance in mink, the technology developed as a part of this project is expected to be useful for the improvement of any economically important trait including feed efficiency and/or pelt quality in ranched mink. The availability of a large resource population with both DNA and reliable performance records currently limits the prospects of genomic-based selection in the fur industry. However, the cost of genomic testing is steadily decreasing and it is expected that an economically feasible cost level will be achieved in near future.

REFERENCES

Abegaz, S., Duguma, G., Negussie, E., Gelmesa, U., Terefe, F. and Rege, J. (2002). Factors affecting reproductive performance and estimates of genetic parameters of litter size in horro sheep. *The Journal of Agricultural Science*. 139, 79-85.

Aitken, N., Smith, S., Schwarz, C. and Morin, P. A. (2004). Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. *Molecular Ecology*. 13, 1423-1431.

Amos, W., Driscoll, E. and Hoffman, J. I. (2011). Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? *Proceedings of the Royal Society B: Biological Sciences.* 278, 1183-1188.

Anistoroaei, R. and Christensen, K. (2006). A test of mink microsatellite markers in the ferret: amplification and sequence comparisons. *Hereditas.* 143, 198-201.

Anistoroaei, R., Ansari, S., Farid, A., Benkel, B., Karlskov-Mortensen, P. and Christensen, K. (2009). An extended anchored linkage map and virtual mapping for the American mink genome based on homology to human and dog. *Genomics*. 94, 204-210.

Anistoroaei, R., Menzorov, A., Serov, O., Farid, A. and Christensen, K. (2007). The first linkage map of the American mink (*Mustela vison*). *Animal Genetics*. 38, 384-388.

Anistoroaei, R., Nielsen, V., Markakis, M. N., Karlskov-Mortensen, P., Jørgensen, C. B., Christensen, K. and Fredholm, M. (2012). A re-assigned American mink (*Neovison vison*) map optimal for genome-wide studies. *Gene*. 511, 66-72.

Archibald, A. L., Cockett, N. E., Dalrymple, B. P., Faraut, T., Kijas, J. W., Maddox, J. F., McEwan, J., Hutton Oddy, V., Raadsma, H., Wade, C., Wang, J., Wang, W. and Xun, X. (2010). The sheep genome reference sequence: a work in progress. *Animal Genetics*. 41, 449-453.

Ardlie K. G., Kruglyak, L. and Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics* 3, 299-309.

Argente, M. J., Merchán, M., Peiró, R., García, M. L., Santacreu, M. A., Folch, J. M. and Blasco, A. (2010). Candidate gene analysis for reproductive traits in two lines of rabbits divergently selected for uterine capacity. *Journal of Animal Science*. 88, 828-836.

Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics*. 11, 375-386.

Balding, D. J. (2006). A tutorial on statistical methods for population association studies. *Nature Reviews Genetics*. 7, 781-791.

Ban, H. J., Heo, J. Y., Oh, K. S. and Park, K. J. (2010). Identification of type 2 diabetes-associated combination of SNPs using support vector machine. *Biomed Central Genetics.* 11, 26.

Barber, R. D., Harmer, D. W., Coleman, R. A. and Clark, B. J. (2005). GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. *Physiological Genomics*. 21, 389-395.

Barendse, W. J. 2002. DNA markers for meat tenderness. International patent application PCT/AU02/00122. [International patent publication WO 02/064820 A1].

Bateson, W., Waunders, E. R. and Punnett, R. C. (1909). Experimental studies in the physiology of heredity. *Molecular and General Genetics*. 2, 17-19.

Benarie, M. (1981). Mathematical population genetics. *Science of the Total Environment.* 19, 198-199.

Benyamin, B., Visscher, P. M. and McRae, A. F. (2009). Family-based genome-wide association studies. *Pharmacogenomics.* 10, 181-190.

Bertolin, K., Gossen, J., Schoonjans, K. and Murphy, B. D. (2014). The orphan nuclear receptor Nr5a2 is essential for luteinization in the female mouse ovary. *Endocrinology*. 155, 1931-1943.

Bionaz, M. and Loor, J. J. (2008). Gene networks driving bovine milk fat synthesis during the lactation cycle. *BioMed Central Genomics*. 9, 366.

Boden, M. J. (2008). The role of circadian rhythms in reproduction: development and fertility in the bmal1 null mouse (doctoral dissertation), University of Adelaide.

Botstein, D., White, R. L., Skolnick, M. and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics.* 32, 314-331.

Boudreau, L., Benkel, B., Astatkie, T. and Rouvinen-Watt, K. (2014). Ideal body condition improves reproductive performance and influences genetic health in female mink. *Animal Reproduction Science*. 145, 86-98.

Boulton, T. G. and Cobb, M. H. (1991). Identification of multiple extracellular signal-regulated kinases (ERKs) with antipeptide antibodies. *Cell Regulation*. 2, 357-371.

Bowness, E. R. (1968). A survey of the gestation period and litter size in ranch mink. *The Canadian Veterinary Journal.* 9, 103-106.

Brookes, A. J., Lehväslaiho, H., Siegfried, M., Boehm, J. G., Yuan, Y. P., Sarkar, C. M., Bork, P. and Ortigao, F. (2000). HGBASE: a database of SNPs and other variations in and around human genes. *Nucleic Acids Research*. 28, 356-360.

Browning, B. L. and Yu, Z. (2009). Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. *The American Journal of Human Genetics*. 85, 847-861.

Buono, K. D., Robinson, G. W., Martin, C., Shi, S., Stanley, P., Tanigaki, K., Honju, T. and Hennighausen, L. (2006). The canonical Notch/RBP-J signaling pathway controls the balance of cell lineages in mammary epithelium during pregnancy. *Developmental Biology*. 293, 565-580.

Butler, W. R. (1998). Effect of protein nutrition on ovarian and uterine physiology in dairy cattle. *Journal of Dairy Science*. 81, 2533-2539.

Camaioni, A., Hascall, V. C., Yanagishita, M. and Salustri, A. (1993). Effects of exogenous hyaluronic acid and serum on matrix organization and stability in the mouse cumulus cell-oocyte complex. *Journal of Biological Chemistry*. 268, 20473-20481.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C. P., Lim, E. P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G. Q. and Lander E. S. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genetics.* 22, 231-238.

Castellini, C., Dal, B. A., Arias-Álvarez, M., Lorenzo, P. L., Cardinali, R. and Rebollar, P. G. (2010). The main factors affecting the reproductive performance of rabbit does: a review. *Animal Reproduction Science.* 122, 3-4.

Cha, J., Sun, X., Bartos, A., Fenelon, J., Lefèvre, P., Daikoku, T., Shaw, G., Maxson, R., Murphy, B. D., Renfree, M. B and Dey, S. K. (2013). A new role for muscle segment homeobox genes in mammalian embryonic diapause. *Open Biology*. 3, 130035-130045.

Chan, E. K. F., Hawken, R. and Reverter, A. (2009). The combined effect of SNP-marker and phenotype attributes in genome-wide association studies. *Animal Genetics.* 40, 149-156.

Chanock, S. J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D. J., Thomas, G., Hirschhorn, J. N., Abecasis, G., Altshuler, D., Bailey-Wilson, J. E., Brooks, L. D., Cardon, L. D., Daly, M., Donnelly, P., Fraumeni, J., Freimer, N. B., Gerhard, D. S., Gunter, C., Guttmacher, A. E., Guyer, M. S., Harris, E. L., Hoh, J., Hoover, R., Kong, C. A., Merikangas, K. R., Morton, C. C., Palmer, L. J., Phimister, E. G., Rice, J. P., Roberts, J., Rotimi, C., Tucker, M. A., Vogan, K. J., Wacholder, S., Wijsman, E. M., Winn, D. M., Collins, F. S. and NCI-NHGRI Working Group on Replication in Association Studies. (2007). Replicating genotype-phenotype associations. *Nature.* 447, 655-660.

Chappel, S. C. and Howles, C. (1991). Reevaluation of the roles of luteinizing hormone and follicle-stimulating hormone in the ovulatory process. *Human Reproduction*. 6, 1206-1212.

Chawla, A., Schwarz, E. J., Dimaculangan, D. D. and Lazar, M. A. (1994). Peroxisome proliferator-activated receptor (PPAR) gamma: adipose-predominant expression and induction early in adipocyte differentiation. *Endocrinology.* 135, 798-800.

Chebel, R. C., Braga, F. A. and Dalton, J. C. (2007). Factors affecting reproductive performance of Holstein heifers. *Animal Reproduction Science.* 101, 208-224.

Chowdhary, B. P., Frönicke, L., Gustavsson, I. and Scherthan, H. (1996). Comparative analysis of the cattle and human genomes: detection of ZOO-FISH and gene mapping-based chromosomal homologies. *Mammalian Genome*. 7, 297-302.

Christensen, K. and Anistoroaei, R. (2014). An American mink (*Neovison vison*) transcriptome. *Animal Genetics*. 45, 301-303.

Clarke, G. M., Anderson, C. A., Pettersson, F. H., Cardon, L. R., Morris, A. P. and Zondervan, K. T. (2011). Basic statistical analysis in genetic case-control studies. *Nature Protocols*. 6, 121-133.

Clausen, T. N., Sandbøl, P. and Hejlesen, C. (2007). Body score of females in the winter and breeding period. Annual Report 2006, Danish Fur Breeders Research Center, Holstebro, Denmark, 181-184.

Cochran, S. D., Cole, J. B., Null, D. J. and Hansen, P. J. (2013). Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. *BMC Genetics*. 14, 49.

Collins, F. S., Guyer, M. S. and Charkravarti, A. (1997). Variations on a theme: cataloging human DNA sequence variation. *Science.* 278, 1580-1581.

Dandona, P., Aljada, A. and Bandyopadhyay, A. (2004). Inflammation: the link between insulin resistance, obesity and diabetes. *Trends in Immunology*. 25, 4-7.

De Bont, R. and Van Larebeke, N. (2004). Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis.* 19, 169-185.

Desmarais, J. A., Lopes, F. L., Zhang, H., Das, S. K. and Murphy, B. D. (2007). The peroxisome proliferator-activated receptor gamma regulates trophoblast cell differentiation in mink (*Mustela vison*). *Biology of Reproduction*. 77, 829-839.

Dexheimer, T. S. (2013). DNA repair pathways and mechanisms. In Mathews, L., Cabarcas, S. and Hurt, E. DNA repair of cancer stem cells. Springer. Netherlands.

Di, X., Matsuzaki, H., Webster, T. A., Hubbell, E., Liu, G., Dong, S., Bartell, D., Huang, J., Chiles, R., Yang, G., Shen, M. M., Kulp, D., Kennedy, G. C., Mei, R., Jones, K. W. and Cawley, S. (2005). Dynamic model based algorithms for screening and genotyping over 100K SNPs on oligonucleotide microarrays. *Bioinformatics.* 21, 1958-1963.

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., Tosser-Klopp, G., Wang, J., Yang, S., Liang, J., Chen, W., Chen, J., Zeng, P., Hou, Y., Bian, C., Pan, S., Li, Y., Liu, X., Wang, W., Servin,B., Sayre, B., Zhu, B., Sweeney, D., Moore, R., Nie, W., Yongyi Shen, Zhao, R., Zhang, G., Li, J., Faraut, T., Womack, J., Zhang, Y., Kijas, J., Cockett, N., Xu, X., Zhao, S., Wang, J. and Wang, W. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nature Biotechnology*. 31, 135-141.

Douglas, D. A., Pierson, R. A. and Murphy, B. D. (1994). Ovarian follicular development in mink (*Mustela vison*). *Reproduction.* 100, 583-590.

Douglas, D. A., Song, J. H., More, G. M. and Murphy, B. D. (1998). Differentiation of the corpus luteum of the mink (*Mustela vison*): mitogenic and steroidogenic potential of luteal cells from embryonic diapause and postimplantation gestation. *Biology of Reproduction.* 58, 1163-1169.

Duggal, P. S., Van der Hoek, K. H., Milner, C. R., Ryan, N. K., Armstrong, D. T., Magoffin, D. A. and Norman, R. J. (2000). The *in vivo* and *in vitro* effects of exogenous leptin on ovulation in the rat. *Endocrinology*. 141, 1971-1976.

Duggavathi, R., Volle, D. H., Mataki, C., Antal, C. M., Messaddeq, N., Auwerx, J., Murphy, B. D. and Schoonjans, K. (2008). Liver receptor homolog 1 is essential for ovulation. *Genes and Development*. 22, 1871-1876.

Dunstone. (1993). The Mink. T & AD Poyser Ltd, London. UK.

Dupont, J., Reverchon, M., Cloix, L., Froment, P. and Rame, C. (2012). Involvement of adipokines, AMPK, PI3K and the PPAR signaling pathways in ovarian follicle development and cancer. *International Journal of Developmental Biology*. 56, 959-967.

Enders, R. K. (1952). Reproduction in the mink (*Mustela vison*). *Proceedings of the American Philosophical Society*, 691-755.

Frayn, K. N. (2001). Adipose tissue and the insulin resistance syndrome. *Proceedings of the Nutrition Society.* 60, 375-380.

Froment, P., Fabre, S., Dupont, J., Pisselet, C., Chesneau, D., Staels, B. and Monget, P. (2003). Expression and functional role of peroxisome proliferator-activated receptor-γ in ovarian folliculogenesis in the sheep. *Biology of Reproduction*. 69*,* 1665-1674.

Fu, W., Wang, Y., Wang, Y., Li, R., Lin, R. and Jin, L. (2009). Missing call bias in high-throughput genotyping. *Biomed Central Genomics*. 10, 106-120.

Gallais, A., Moreau, L. and Charcosset, A. (2007). Detection of marker–QTL associations by studying change in marker frequencies with selection. *Theoretical and Applied Genetics*. 114, 669-681.

Gambineri, A., Pelusi, C., Vicennati, V., Pagotto, U. and Pasquali, R. (2002). Obesity and the polycystic ovary syndrome. *International Journal of Obesity*. 26, 883-896.

Garanty-Bogacka, B., Syrenicz, M., Goral, J., Krupa, B., Syrenicz, J., Walczak, M. and Syrenicz, A. (2011). Changes in inflammatory biomarkers after successful lifestyle intervention in obese children. *Endokrynologia Polska*. 62, 499-505.

Gavrilova, O., Haluzik, M., Matsusue, K., Cutson, J. J., Johnson, L., Dietz, K. R., Nicol, C.J., Vinson, C., Gonzalez, F. J. and Reitman, M. L. (2003). Liver peroxisome proliferator-activated receptor γ contributes to hepatic steatosis, triglyceride clearance, and regulation of body fat mass. *Journal of Biological Chemistry*. 278, 34268-34276.

Gorlov, I. P., Gorlova, O. Y., Sunyaev, S. R., Spitz, M. R. and Amos, C. I. (2008). Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *The American Journal of Human Genetics*. 82, 100-112.

Grandinson, K., Rydhmer, L., Strandberg, E. and Thodberg, K. (2003). Genetic analysis of on-farm tests of maternal behaviour in sows. *Livestock Production Science*. 83, 141-151.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N, Mni,M., Reid, S., Simon, P., Spelman, R., Georges, M. and Snell, R. (2002). Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research.* 12, 222-231.

Grundy, S. M. (1998). Multifactorial causation of obesity: implications for prevention. *The American Journal of Clinical Nutrition*. 67, 563-572.

Gulevich, R. G., Klotchkov, D. V., Ivanova, L. N. and Osadchuk, L. V. (1995). Gonadal function in mink under artificial photoperiods. *Journal of Reproduction and Fertility.* 103, 147-152.

Hameister, H., Klett, C., Bruch, J., Dixkens, C., Vogel, W. and Christensen, K. (1997). Zoo-FISH analysis: the American mink (*Mustela vison*) closely resembles the cat karyotype. *Chromosome Research*. 5, 5-11.

Han, J., Lee, J. D., Bibbs, L. and Ulevitch, R. J. (1994). A MAP kinase targeted by endotoxin and hyperosmolarity in mammalian cells. *Science*. 265, 808-811.

Hanrahan, J. P., Gregan, S. M., Mulsant, P., Mullen, M., Davis, G. H., Powell, R. and Galloway, S. M. (2004). Mutations in the genes for oocyte-derived growth factors GDF9 and BMP15 are associated with both increased ovulation rate and sterility in Cambridge and Belclare sheep (*Ovis aries*). *Biology of Reproduction*. 70, 900-909.

Hansen, B. K., Su, G. and Berg, P. (2010). Genetic variation in litter size and kit survival of mink (*Neovison vison*). *Journal of Animal Breeding and Genetics*. 127, 442-451.

Hansson, A. (1947). The physiology of reproduction in mink *(Mustela vison*) with special reference to delayed implantation. *Acta Zoologica*. 28, 1-136.

Hastings, M. H., Maywood, E. S. and Reddy, A. B. (2008). Two decades of circadian time. *Journal of Neuroendocrinology*. 20, 812-819.

Hillier, S. G. (1994). Current concepts of the roles of follicle stimulating hormone and luteinizing hormone in folliculogenesis. *Human Reproduction*. 9, 188-191.

Hinshelwood, M. M., Shelton, J. M., Richardson, J. A. and Mendelson, C. R. (2005). Temporal and spatial expression of liver receptor homologue-1 (LRH-1) during embryogenesis suggests a potential role in gonadal development. *Developmental Dynamics*. 234, 159-168.

Hirschhorn, J. N. and Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics.* 6, 95-108.

Hochberg, Y. and Benjamini, Y. (1990). More powerful procedures for multiple significance testing. *Statistics in Medicine*. 9, 811-818.

Howie, G. J., Sloboda, D. M., Kamal, T. and Vickers, M. H. (2009). Maternal nutritional history predicts obesity in adult offspring independent of postnatal diet. *The Journal of Physiology*. 587, 905-915.

Hughes, J. R, Francis, M. and Gorospe, W. C. (1991). Biochemical identification of apoptosis (programmed cell death) in granulosa cells: evidence for a potential mechanism underlying follicular atresia. *Endocrinology*. 129, 2415-2422.

Hynes A. M., Rouvinen-Watt, K. and Armstrong, D. (2004). Body condition and glycemic control in mink females during reproduction and lactation. *Scientifur.* 28, 79-86.

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

Ioannidis, J. P., Ntzani, E. E., Trikalinos, T. A. and Contopoulos-Ioannidis, D. G. (2001). Replication validity of genetic association studies. *Nature Genetics*. 29, 306-309.

Jiang, L., Liu, J., Sun, D., Ma, P., Ding, X., Yu, Y. and Zhang, Q. (2010). Genome wide association studies for milk production traits in Chinese Holstein population. *Public Library of Science One.* 5, 1-12.

Johansson, M., Ellegren, H. and Andersson, L. (1995). Comparative mapping reveals extensive linkage conservation but with gene order rearrangements between the pig and the human genomes. *Genomics*. 25, 682-690.

Johnson, A. D. and O'Donnell, C. J. (2009). An open access database of genome-wide association results. *Biomed Central Medical Genetics.* 10, 6.

Joseph, C., Hunter, M. G., Sinclair, K. D. and Robinson, R. S. (2012). The expression, regulation and function of secreted protein, acidic, cysteine-rich in the follicle–luteal transition. *Reproduction*. 144, 361-372.

Kerem, B., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K., Chakravarti, A., Buchwald, M. and Tsui, L. C. (1989). Identification of the cystic fibrosis gene: genetic analysis. *Science.* 245, 1073-1080.

Khatib, H., Maltecca, C., Monson, R. L., Schutzkus, V., Wang, X. and Rutledge, J. J. (2008). The fibroblast growth factor 2 gene is associated with embryonic mortality in cattle. *Journal of Animal Science*. 86, 2063-2067.

Kim, S. Y., Lohmueller, K. E., Albrechtsen, A., Li, Y., Korneliussen, T., Tian, G., Grarup, N., Jiang, T., Andersen, G., Witte, D., Jorgensen, T., Hansen, T., Pedersen, O., Wang, J. and Nielsen, R. (2011). Estimation of allele frequency and association mapping using next-generation sequencing data. *Biomed Central Bioinformatics.* 12, 231.

Komar, C. M., Braissant, O., Wahli, W. and Curry, T. E. (2001). Expression and localization of PPARs in the rat ovary during follicular development and the periovulatory period. *Endocrinology*. 142, 4831-4838.

König, I. R. (2011). Validation in genetic association studies. *Briefings in Bioinformatics*. 12, 253-238.

Koohmaraie, M. (1996). Biochemical factors regulating the toughening and tenderization processes of meat. *Meat Science*. 43, 193-201.

Kopelman, P. G. (2000). Obesity as a medical problem. *Nature*. 404, 635-643.

Korhonen, H. (1992). Activated mammary number and litter size in the mink. *Reproduction, Nutrition, Development.* 32, 67-71.

Korhonen, H. and Niemelä, P. (1998). Effect of *ad libitum* and restrictive feeding on seasonal weight changes in captive minks (*Mustela vison*). *Journal of Animal Physiology and Animal Nutrition*. 79, 269-280.

Krawczak, M. and Cooper, D. N. (1997). The human gene mutation database. *Trends in Genetics*. 13, 121-122.

Kristensen, J., Vestergaard, M., Wisborg, K., Kesmodel, U. and Secher, N. J. (2005). Pre-pregnancy weight and the risk of stillbirth and neonatal death. *BJOG: An International Journal of Obstetrics and Gynaecology*. 112, 403-408.

Kuroda, K. O., Meaney, M. J., Uetani, N., Fortin, Y., Ponton, A. and Kato, T. (2007). ERK-FosB signaling in dorsal MPOA neurons plays a major role in the initiation of parental behavior in mice. *Molecular and Cellular Neuroscience*. 36, 121-131.

Kwon, J. M. and Goate, A. M. (2000). The candidate gene approach. *Alcohol Research and Health*. 24, 164-168.

Lagerkvist, G. (1997). Economic profit from increased litter size, body weight and pelt quality in mink (*Mustela vison*). *Acta Agriculturae Scandinavica*. 47, 57-63.

Lagerkvist, G., Johansson, K. and Lundeheim, N. (1993). Selection for litter size, body weight, and pelt quality in mink *(Mustela vison)*: experimental design and direct response of each trait. *Journal of Animal Science* 71, 3261-3272.

Lagerkvist, G., Johansson, K. and Lundeheim N. (1994). Selection for litter size, body weight, and pelt quality in mink (*Mustela vison*): correlated responses. *Journal of Animal Science*. 72, 1126-1137.

Lambe, K. G. and Tugwood, J. D. (1996). A human peroxisome-proliferator-activated receptor-γ is activated by inducers of adipogenesis, including thiazolidinedione drugs. *European Journal of Biochemistry*. 239, 1-7.

Lawler, D. F. and Monti, K. L. (1984). Morbidity and mortality in neonatal kittens. *American Journal of Veterinary Research*. 45, 1455-1459.

Lee, C. H., Olson, P., Hevener, A., Mehl, I., Chong, L. W., Olefsky, J. M., Gonzalez, F. J., Ham, J., Kang, H., Peters, J. M. and Evans, R. M. (2006). PPAR regulates glucose metabolism and insulin sensitivity. *Proceedings of the National Academy of Sciences*. 103, 3444-3449.

Lee, J. D., Ulevitch, R. J. and Han, J. H. (1995). Primary structure of BMK1: a new mammalian map kinase. *Biochemical and Biophysical Research Communications.* 213, 715-724.

Lefèvre, P. L. and Murphy, B. D. (2008). Physiological constraints on litter size in mink. *Scientifur*. 32, 13-14.

Lefèvre, P. L. and Murphy, B. D. (2009) Differential gene expression in the uterus and blastocyst during the reactivation of embryo development in a model of delayed implantation. *Methods in Molecular Biology.* 550, 11-61.

Lefevre, P. L., Beaudry, D., Palin, M. F. and Murphy, B. D. (2008). Uterine global gene expression at embryo reactivation after embryonic diapause in the American Mink. *Scientifur.* 32, 111-112.

Lefèvre, P. L., Palin, M. F., Beaudry, D., Dobias-Goff, M., Desmarais, J. A., Llerena V. E. M. and Murphy, B. D. (2011). Uterine signaling at the emergence of the embryo from obligate diapause. *American Journal of Physiology - Endocrinology and Metabolism*. 300, 800-808.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*. 25, 2078-2079.

Lim, H., Paria, B. C., Das, S. K., Dinchuk, J. E., Langenbach, R., Trzaskos, J. M. and Dey, S. K. (1997). Multiple female reproductive failures in cyclooxygenase 2–deficient mice. *Cell*. 91, 197-208.

Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., Clamp M., Chang, J. L., Kulbokas, E. J., Zody, M.C., Mauceli, E., Xie, X., Breen, M., Wayne, R. K., Ostrander, E. A., Ponting, C. P., Galibert, F., Smith, D. R., DeJong, P. J., Kirkness, E., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C. W., Cook, A., Cuff, J., Daly, M. J., DeCaprio, D., Gnerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K. P., Parker, H. G., Pollinger, J. P., Searle, S. M., Sutter, N. B., Thomas, R., Webber, C., Baldwin, J., Abebe, A., Abouelleil, A., Aftuck, L., Ait-Zahra, M., Aldredge, T., Allen, N., An, P., Anderson, S., Antoine, C., Arachchi, H., Aslam, A., Ayotte, L., Bachantsang, P., Barry, A., Bayul, T., Benamara, M., Berlin, A., Bessette, D., Blitshteyn, B., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Brown, A., Cahill, P., Calixte, N., Camarata, J., Cheshatsang, Y., Chu, J., Citroen, M., Collymore, A., Cooke, P., Dawoe, T., Daza, R., Decktor, K., DeGray, S., Dhargay, N., Dooley, K., Dooley, K., Dorje, P., Dorjee, K., Dorris, L., Duffey, N., Dupes, A., Egbiremolen, O., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Ferreira, P., Fisher, S., FitzGerald, M., Foley, K., Foley, C., Franke, A., Friedrich, D., Gage, D., Garber, M., Gearin, G., Giannoukos, G., Goode, T., Goyette, A., Graham, J., Grandbois, E., Gyaltsen, K., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Healy, C., Hegarty, R., Honan, T., Horn, A., Houde, N., Hughes, L., Hunnicutt, L., Husby, M., Jester, B., Jones, C., Kamat, A., Kanga, B., Kells, C., Khazanovich, D., Kieu, A. C., Kisner, P., Kumar, M., Lance, K., Landers, T., Lara, M., Lee, W., Leger, J. P., Lennon, N., Leuper, L., LeVine, S., Liu, J., Liu, X., Lokyitsang, Y., Lokyitsang, T., Lui, A., Macdonald, J., Major, J., Marabella, R., Maru, K., Matthews, C., McDonough, S., Mehta, T., Meldrim, J., Melnikov, A., Meneus, L., Mihalev, A., Mihova, T., Miller, K., Mittelman, R., Mlenga, V., Mulrain, L., Munson, G., Navidi, A., Naylor, J., Nguyen, T., Nguyen, N., Nguyen, C., Nguyen, T., Nicol, R., Norbu, N., Norbu, C., Novod, N., Nyima, T., Olandt, P., O'Neill, B., O'Neill, K., Osman, S., Oyono, L., Patti, C., Perrin, D., Phunkhang, P., Pierre, F., Priest, M., Rachupka, A., Raghuraman, S., Rameau, R., Ray, V., Raymond, C., Rege, F., Rise, C., Rogers, J., Rogov, P., Sahalie, J., Settipalli, S., Sharpe, T., Shea, T., Sheehan, M., Sherpa, N., Shi, J., Shih, D., Sloan, J., Smith, C., Sparrow, T., Stalker, J., Stange-Thomann, N., Stavropoulos, S., Stone, C., Stone, S., Sykes, S., Tchuinga, P., Tenzing, P., Tesfaye, S., Thoulutsang, D., Thoulutsang, Y., Topham, K., Topping, I., Tsamla, T., Vassiliev, H., Venkataraman, V., Vo, A., Wangchuk, T., Wangdi, T., Weiand, M., Wilkinson, J., Wilson, A., Yadav, S., Yang, S., Yang, X., Young, G., Yu, Q., Zainoun, J., Zembek, L., Zimmer, A. and Lander, E. S. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature.* 438, 803-819.

Liu, H. C., Cheng, H. H., Tirunagaru, V., Sofer, L. and Burnside, J. (2001). A strategy to identify positional candidate genes conferring Marek's disease resistance by integrating DNA microarrays and genetic mapping. *Animal Genetics*. 32, 351-359.

Liu, Y. X. (1999). Regulation of the plasminogen activator system in the ovary. *Neurosignals*. 8, 160-177.

Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S. and Hirschhorn, J. N. (2003). Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genetics*. 33, 177-182.

Lonergan, P., Carolan, C., Van Langendonckt, A., Donnay, I., Khatir, H. and Mermillod, P. (1996). Role of epidermal growth factor in bovine oocyte maturation and preimplantation embryo development in vitro. *Biology of Reproduction*. 54, 1420-1429.

Lopes, F. L., Desmarais, J. A. and Murphy, B. D. (2004). Embryonic diapause and its regulation. *Reproduction.* 128, 669-678.

Lydon, J. P., DeMayo, F. J., Conneely, O. M. and O'Malley, B. W. (1996). Reproductive phenotypes of the progesterone receptor null mutant mouse. *The Journal of Steroid Biochemistry and Molecular Biology*. 56, 1-6.

Lydon, J. P., DeMayo, F. J., Funk, C. R., Mani, S. K., Hughes, A. R., Montgomery, C. A., Shyamala, G., Conneely, O. and O'Malley, B. W. (1995). Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities. *Genes and Development*. 9, 2266-2278.

Lyons, R. A., Saridogan, E. and Djahanbakhch, O. (2006). The reproductive significance of human fallopian tube cilia. *Human Reproduction Update*. 12, 363-372.

Madan, P., Calder, M. D. and Watson, A. J. (2005). Mitogen-activated protein kinase (MAPK) blockade of bovine preimplantation embryogenesis requires inhibition of both p38 and extracellular signal-regulated kinase (ERK) pathways. *Reproduction*. 130, 41-51.

Mai, M. D., Christiansen, F. B., Sahana, G. and Guldbrandtsen, B. (2010). A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. *Journal of Animal Science.* 88, 3522-3528.

Malmkvist, J., Gade, M. and Damm, B. I. (2007). Parturient behaviour in farmed mink (*Mustela vison*) in relation to early kit mortality. *Applied Animal Behaviour Science*. 107, 120-132.

Malmkvist, J., Pedersen, L. J., Damgaard, B. M., Thodberg, K., Jørgensen, E. and Labouriau, R. (2006). Does floor heating around parturition affect the vitality of piglets born to loose housed sows? *Applied Animal Behaviour Science*. 99, 88-105.

Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schütz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P. and Evans, R. M. (1995). The nuclear receptor superfamily: the second decade. *Cell*. 83, 835-839.

Martino, P. E. and Villar, J. A. (1990). A survey on perinatal mortality in young mink. *Veterinary Research Communications.* 14, 199-205.

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. and Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics*. 9, 356-369.

McNatty, K. P., Galloway, S. M., Wilson, T., Smith, P., Hudson, N. L., O'Connell, A., Bibby, Heath, D. A., Davis, G. H., Hanrahan, J. P. and Juengel, J. L. (2005). Physiological effects of major genes affecting ovulation rate in sheep. *Genetics, Selection, Evolution*. 37, 25-38.

Miller, B. H., Olson, S. L., Turek, F. W., Levine, J. E., Horton, T. H. and Takahashi, J. S. (2004). Circadian clock mutation disrupts estrous cyclicity and maintenance of pregnancy. *Current Biology*. 14, 1367-1373.

Moorhead, M., Hardenbol, P., Siddiqui, F., Falkowski, M., Bruckner, C., Ireland, J., Hywel, B. J., Jain, M., Willis, T. D. and Faham, M. (2006). Optimal genotype determination in highly multiplexed SNP data. *European Journal of Human Genetics*. 14, 207–215.

Morgan, T. H. (1911). Random segregation versus coupling in Mendelian inheritance. *Science*. 34, 384-384.

Moschos, S., Chan, J. L. and Mantzoros, C. S. (2002). Leptin and reproduction: A review. *Fertility and Sterility*. 77, 433-444.

Mostyn, A., Keisler, D. H., Webb, R., Stephenson, T. and Symonds, M. E. (2001). The role of leptin in the transition from fetus to neonate. *Proceedings of the Nutrition Society*. 60, 187-194.

Murphy, B. D. (1996). Female Reproductive System. In: Hunter, D.B. and Lemieux, N. (Ed). Mink…biology, health and disease. Graphic and Print Services, Guelph, Ontario.

Murphy, B. D. (2012). Embryonic diapause: advances in understanding the enigma of seasonal delayed implantation. *Reproduction in Domestic Animals*. 47, 121-124.

Murphy, B. D. and James, D. A. (1976). Cells of the adenohypophysis of the mink (*Mustela vison*) identified by immunohistochemical and functional criteria. *Cells Tissues Organs*. 94, 184-203.

Nammi, S., Koka, S., Chinnala, K. M. and Boini, K. M. (2004). Obesity: An overview on its current perspectives and treatment options. *Nutrition Journal*. 3, 1-8.

Nguyen, X. M. T., Lane, J., Smith, B. R. and Nguyen, N. T. (2009). Changes in inflammatory biomarkers across weight classes in a representative US population: a link between obesity and inflammation. *Journal of Gastrointestinal Surgery*. 13, 1205-1212.

Nishida, E. and Gotoh, Y. (1993). The MAP kinase cascade is essential for diverse signal transduction pathways. *Trends in Biochemical Sciences*. 18, 128-131.

Niswender, G. D. (2002). Molecular control of luteal secretion of progesterone. *Reproduction*. 123, 333-339.

Niu, B. Y., Xiong, Y. Z., Li, F. E., Deng, C. Y., Jiang, S. W., Ye, L. Z. and Guo, W. H. (2006). Polymorphism of the pig pre-implantation protein 3 (prei3) gene and its association with litter size traits. *South African Journal of Animal Science*. 36, 209.

Olfert, E. D., Cross, B. M. and McWilliam, A. A. (1993). In: Guide to the Care and Use of Experimental Animals. Canadian Council on Animal Care. Ottawa, Ontario.

O'Shea, J. J., Gadina, M. and Kanno, Y. (2011). Cytokine signaling: birth of a pathway. *The Journal of Immunology.* 187, 5475-5478.

Page, B. T., Casas, E., Heaton, M. P., Cullen, N. G., Hyndman, D. L., Morris, C. A., Crawford, A. M., Wheeler, T. L., Koohmaraie, M., Keele, J. W. and Smith, T. P. (2002). Evaluation of single-nucleotide polymorphisms in CAPN1 for association with meat tenderness in cattle. *Journal of Animal Science*. 80, 3077-3085.

Pasquali, R. (2006). Obesity and androgens: facts and perspectives. *Fertility and Sterility*. 85, 1319-1340.

Pasquali, R. and Gambineri, A. (2006). Metabolic effects of obesity on reproduction. *Reproductive Biomedicine Online*. 12, 542-551.

Pearson, T. A. and Manolio, T. A. (2008). How to interpret a genome-wide association study. *Journal of the American Medical Assocition.* 299, 1335-1344.

Pedersen, T. and Peters, H. (1968). Proposal for a classification of oocytes and follicles in the mouse ovary. *Journal of Reproduction and Fertility*. 17, 555-557.

Pilbeam, T. E, Concannon, P. W. and Travis, H. F. (1979). The annual reproductive cycle of mink (*Mustela vison*). *Journal of Animal Science.* 48, 578-584.

Plymate, S. R., Matej, L. A., Jones, R. E. and Friedl, K. E. (1988). Inhibition of sex hormone-binding globulin production in the human hepatoma (Hep G2) cell line by insulin and prolactin. *The Journal of Clinical Endocrinology & Metabolism*. 67, 460-464.

Ponsuksili, S., Brand, B., Schwerin, M., Murani, E. and Wimmers, K. (2011). Integrating expression profiling and whole-genome association for dissection of fat traits in a porcine model. *Journal of Lipid Research.* 52, 668-678.

Poretsky, L., Cataldo, N. A., Rosenwaks, Z. and Giudice, L. C. (1999). The insulin-related ovarian regulatory system in health and disease. *Endocrine Reviews*. 20, 535-582.

Puurunen, J., Piltonen, T., Jaakkola, P., Ruokonen, A., Morin-Papunen, L. and Tapanainen, J. S. (2009). Adrenal androgen production capacity remains high up to menopause in women with polycystic ovary syndrome. *Journal of Clinical Endocrinology & Metabolism*. *94*, 1973-1978.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Raafat, A., Goldhar, A. S., Klauzinska, M., Xu, K., Amirjazil, I., McCurdy, D., Lashin, K., Salomon, D., Vonderhaar, B., Egan, S. and Callahan, R. (2011). Expression of Notch receptors, ligands, and target genes during development of the mouse mammary gland. *Journal of Cellular Physiology*. 226, 1940-1952.

Rajkovic, A., Pangas, S. A. and Matzuk, M. M. (2006). Follicular development: mouse, sheep, and human models. In: Knobil, E and Neill, J. D. (3rd Ed). Knobil and Neill's Physiology of Reproduction. Elsevier, Amsterdam.

Rasmussen, K. M. (1998). Effects of under and overnutrition on lactation in laboratory rats. *Journal of Nutrition*. 128, 390-393.

Rawlings, J. S., Rosler, K. M. and Harrison, D. A. (2004). The JAK/STAT signaling pathway. *Journal of Cell Science*. 117, 1281-1283.

Rempel, L. A., Nonneman, D. J., Wise, T. H., Erkens, T., Peelman, L. J. and Rohrer, G. A. (2010). Association analyses of candidate single nucleotide polymorphisms on reproductive traits in swine. *Journal of Animal Science*. 88, 1-15.

Risch, N. J. (2000). Searching for genetic determinants in the new millennium. *Nature.* 405, 847–856.

Rothschild, M., Jacobson, C., Vaske, D., Tuggle, C., Wang, L., Short, T., Eckardt, G., Sasaki, S., Vincent, A., McLaren D., Southwoodt, O., Van Der Steent, H., Mileham, A. and Plastow, G. (1996). The estrogen receptor locus is associated with a major gene influencing litter size in pigs. *Proceedings of the National Academy of Sciences of the United States of America*. 93, 201-205.

Rouvinen-Watt, K. and Harri, M. (2001). Observations on thermoregulatory ontogeny of mink (*Mustela vison*). *Journal of Thermal Biology.* 26, 9-14.

Rouvinen-Watt, K., White, M. B. and Campbell, R. (2005). Mink Feeds and Feeding. Ontario Ministry of Agriculture and Food, through the Agricultural Research Institute of Ontario; Nova Scotia Agricultural College, Truro, NS, Canada.

Sahana, G., Guldbrandtsen, B., Bendixen, C. and Lund, M. S. (2010). Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. *Animal Genetics.* 41, 579-588.

Salanti, G., Amountza, G., Ntzani, E. E. and Ioannidis, J. P. (2005). Hardy–Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power. *European Journal of Human Genetics.* 13, 840–848.

Sang, Q. X. (1998). Complex role of matrix metalloproteinases in angiogenesis. *Cell Research*. 8, 171-177.

Satoh, Y., Endo, S., Nakata, T., Kobayashi, Y., Yamada, K., Ikeda, T., Takeuchi, A., Hiramoto, T., Watanabe, Y. and Kazama, T. (2011). ERK2 contributes to the control of social behaviors in mice. *The Journal of Neuroscience*. 31, 11953-11967.

Schneider, R. R. and Hunter, D. B. (1993). Mortality in mink kits from birth to weaning. *The Canadian Veterinary Journal*. 34, 159-163.

Schook, L. B., Beever, J. E., Rogers, J., Humphray, S., Archibald, A., Chardon, P., Milan, D., Rohrer, G. and Eversole, K. (2005). Swine Genome Sequencing Consortium (SGSC): a strategic roadmap for sequencing the pig genome. *Comparative and Functional Genomics*. 6, 251-255.

Schoonjans, K., Annicotte, J. S., Huby, T., Botrugno, O. A., Fayard, E., Ueda, Y., Chapman, J. and Auwerx, J. (2002). Liver receptor homolog 1 controls the expression of the scavenger receptor class B type I. *European Molecular Biology Organisation Reports*. 3, 1181-1187.

Schoppee, P. D., Garmey, J. C. and Veldhuis, J. D. (2002). Putative activation of the peroxisome proliferator-activated receptor γ impairs androgen and enhances progesterone biosynthesis in primary cultures of porcine theca cells. *Biology of Reproduction*. 66, 190-198.

Schwerin, M., Czernek-Schäfer, D., Goldammer, T., Kata, S. R., Womack, J. E., Pareek, R., Pareek, C., Walawski, K. and Brunner, R. M. (2003). Application of disease-associated differentially expressed genes - mining for functional candidate genes for mastitis resistance in cattle. *Genetics, Selection, Evolution*. 35, 19-34.

Scott, F. W., Geissinger, C. and Peltz, R. (1978). Kitten mortality survey. *Feline Practice*. 8, 334-364.

Sherry, S. T., Ward, M. and Sirotkin, K. (2000). Use of molecular variation in the NCBI dbSNP database. *Human Mutation*. 15, 68-75.

Shifman, S., Kuypers, J., Kokoris, M., Yakir, B. and Darvasi, A. (2003). Linkage disequilibrium patterns of the human genome across populations. *Human Molecular Genetics*. 12, 771-776.

Ślaska, B. and Rozempolska-Rucińska, I. (2011). Mating system and level of reproductive performance in mink *(Neovison vison). Annals of Animal Science.* 11, 105-113.

Smith, T. P. L., Showalter, A. D., Sloop, K. W., Rohrer, G. A., Fahrenkrug, S. C., Meier, B. C. and Rhodes, S. J. (2001). Identification of porcine Lhx3 and SF1 as candidate genes for QTL affecting growth and reproduction traits in swine. *Animal Genetics*. 32, 344-350.

Socha, S. and Markiewicz, D. (2002). Effect of mating and whelping dates on the number of pups in mink. *Electronic Journal of Polish Agricultural Universities, Animal Husbandry*, *5*.

Solé, X., Guinó, E., Valls, J., Iniesta, R. and Moreno, V. (2006). SNPStats: a web tool for the analysis of association studies. *Bioinformatics*. 22, 1928-1929.

Sturtevant, A. H. (1913). The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. *Journal of Experimental Zoology*. 14, 43-59.

Summa, K. C., Vitaterna, M. H., and Turek, F. W. (2012). Environmental perturbation of the circadian clock disrupts pregnancy in the mouse. *Public Library of Science One*. 7, 37668.

Sun, X. and Dey, S. K. (2012). Endocannabinoid signaling in female reproduction. *ACS Chemical Neuroscience*. 3, 349-355.

Sundqvist, C., Ellis, L. C. and Bartke, A. (1988). Reproductive endocrinology of the mink (*Mustela vison*). *Endocrinology Reviews.* 9, 247-266.

Tabor, H. K., Risch, N. J. and Myers, R. M. (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Reviews Genetics*. 3, 391-397.

Tauson, A. H. (1994). Postnatal development in mink kits. *Acta Agriculturae Scandinavica.-Animal Science*. 44, 177-184.

Tauson, A. H. and Aldén, E. (1984). Pre-mating bodyweight changes and reproductive performance in female mink. *Acta* Agriculturae *Scandinavica*. 34, 177-187.

Tauson, A. H. and Aldén, E. (1985). Different feeding intensity levels to mink. 2. Effects on female reproductive performance, pre-weaning kit growth and longevity of females. *Swedish Journal of Agricultural Research*. 15, 97-107.

Tauson, A. H., Elnif, J. and Hansen, N. E. (1994). Energy metabolism and nutrient oxidation in the pregnant mink (*Mustela vison*) as a model for other carnivores. *The Journal of Nutrition* 124, 2609-2613.

Tauson, A. H., Forsberg, M. and Chwalibog, A. (2004). High leptin in pregnant mink (*Mustela vison*) may exert anorexigenic effects: a permissive factor for rapid increase in food intake during lactation. *British Journal of Nutrition*. 91, 411-421.

Tilly, J. L., Kowalski, K. I., Johnson, A. L. and Hsueh, A. J. (1991). Involvement of apoptosis in ovarian follicular atresia and postovulatory regression. *Endocrinology*. 129, 2799-2801.

Tindall, E. A., Petersen, D. C., Nikolaysen, S., Miller, W., Schuster, S. C. and Hayes, V. M. (2010). Interpretation of custom designed Illumina genotype cluster plots for targeted

association studies and next-generation sequence validation. *BioMed Central Research Notes*. 3, 39.

Toda, K., Okada, T., Miyaura, C. and Saibara, T. (2003). Fenofibrate, a ligand for PPARα, inhibits aromatase cytochrome P450 expression in the ovary of mouse. *Journal of Lipid Research.* 44, 265-270.

Tranchevent, L. C., Capdevila, F. B., Nitsch, D., De, M. B., De, C. P. and Moreau, Y. (2011). A guide to web tools to prioritize candidate genes. *Briefings in Bioinformatics*. 12, 22-32.

Uhlenhaut, N. H., Jakob, S., Anlag, K., Eisenberger, T., Sekido, R., Kress, J., Treier, A. C., Klugmann, C., Klasen, C. and Holter, N. I. (2009). Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell.* 139, 1130-1142.

Van Eenennaam, A. L., Li, J., Thallman, R. M., Quaas, R. L., Dikeman, M. E., Gill, C. A., Franke, D. E. and Thomas, M. G. (2007). Validation of commercial DNA tests for quantitative beef quality traits. *Journal of Animal Science*. 85, 891-900.

Varani, S., Elvin, J. A., Yan, C., DeMayo, J., DeMayo, F. J., Horton, H. F., Byrne, M. C. and Matzuk, M. M. (2002). Knockout of pentraxin 3, a downstream target of growth differentiation factor-9, causes female subfertility. *Molecular Endocrinology*. 16, 1154-1167.

Vignal, A., Milan, D., SanCristobal, M. and Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*. 34, 275-306.

Wabitsch, M., Jensen, P. B., Blum, W. F., Christoffersen, C. T., Englaro, P., Heinze, E., Rascher, W., Teller, W., Tornqvist, H. and Hauner, H. (1996). Insulin and cortisol promote leptin production in cultured human fat cells. *Diabetes.* 45, 1435-1438.

Wei, W., Visweswaran, S. and Cooper, G. (2011). The application of Naïve Bayes model averaging to predict Alzheimer's disease from genome-wide data. *Journal of the American Medical Informatics Association.* 18, 370-375.

Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 447, 661-678.

Wiggans, G. R., Sonstegard, T. S., VanRaden, P. M., Matukumalli, L. K., Tassell, C. P., Schnabel, R. D., Taylor, J. F. and Schenkel, F. S. (2009). Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. *Journal of Dairy Science.* 92, 3431-3436.

Woodgett, J. R., Avruch, J. and Kyriakis, J. (1995). The stress activated protein kinase pathway. *Cancer Surveys*. 27, 127-138.

Wu, M. C., Hentzel, M. D. and Dziuk, P. J. (1988). Effect of stage of gestation, litter size and uterine space on the incidence of mummified fetuses in pigs. *Journal of Animal Science.* 66, 3202-3207.

Yamakawa, K., Hosoi, M., Koyama, H., Tanaka, S., Fukumoto, S., Morii, H. and Nishizawa, Y. (2000). Peroxisome proliferator-activated receptor-γ agonists increase vascular endothelial growth factor expression in human vascular smooth muscle cells. *Biochemical and Biophysical Research Communications*. 271, 571-574.

Yang, Q., Cui, J., Chazaro, I., Cupples, L. A. and Demissie, S. (2005). Power and type I error rate of false discovery rate approaches in genome-wide association studies. *BioMed Central Genetics*. 6, 134.

Yang, W. C., Li, S. J., Tang, K. Q., Hua, G. H., Zhang, C. Y., Yu, J. N., Han, L. and Yang, L. G. (2010). Polymorphisms in the 52 upstream region of the FSH receptor gene and their association with superovulation traits in Chinese Holstein cows. *Animal Reproduction Science.* 119, 172-177.

Yang, X. (2010). A wheel of time: the circadian clock, nuclear receptors, and physiology. *Genes and Development*. 24, 741-747.

Ye, J. (2007). Role of insulin in the pathogenesis of free fatty acid-induced insulin resistance in skeletal muscle. *Endocrine, Metabolic & Immune Disorders - Drug Targets*. 7, 65-74.

Youngerman, S. M., Saxton, A. M., Oliver, S. P. and Pighetti, G. M. (2004). Association of CXCR2 polymorphisms with subclinical and clinical mastitis in dairy cattle. *Journal of Dairy Science.* 87, 2442-2448.

Zhang, C., Large, M. J., Duggavathi, R., DeMayo, F. J., Lydon, J. P., Schoonjans, K., Kovanci, E. and Murphy, B. D. (2013). Liver receptor homolog-1 is essential for pregnancy. *Nature Medicine*. 19, 1061-1066.

Zhu, M. and Zhao, S. (2007). Candidate gene identification approach: progress and challenges. *International Journal of Biological Sciences*. 3, 420-427.

Zondervan, K. T. (2011). Genetic association study design. In Zeggini, E. and Morris, A. P. Analysis of complex disease association studies: A practical guide. London.

APPENDIX 1

SNPAssoc output for the tail and window analysis in MDR and CTRL groups.

SNPAssoc returns a set of results in the form of four inheritance models (codominant, dominant, recessive and overdominant) which corresponds to different groups of genotypes. When only one copy of the allele is required to induce a positive effect on the phenotype, the mode of inheritance is called dominant. When two copies of the allele are required to induce a positive effect on the phenotype, the mode of inheritance is called recessive. When both the alleles are individually expressed in the presence of each other and induce a positive effect on the phenotype, the mode of inheritance is called codominant. When the heterozygote genotype have a stronger positive effect on the phenotype, the mode of inheritance is called overdominant or heterozygote advantage. Tables 1-33 correspond to the SNPAssoc results for the tail analysis. Tables 34-51 correspond to the SNPAssoc results for the window analysis.

Table 1: SNPAssoc results for SNP 3 in gene N5 in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	T/T	40	6.400	0.090
	G/T	36	4.889	
	G/G	8	5.625	
Dominant	T/T	40	6.400	***0.035***
	G/T-G/G	44	5.023	
Recessive	T/T-G/T	76	5.684	0.958
	G/G	8	5.625	
Overdominant	T/T-G/G	48	6.271	0.036
	G/T	36	4.889	

Table 2: SNPAssoc results for SNP 5 in gene N5 in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	T/T	39	6.359	0.126
	C/T	44	5.023	
	C/C	8	5.375	
Dominant	T/T	39	6.359	***0.043***
	C/T-C/C	52	5.077	
Recessive	T/T-C/T	83	5.651	0.806
	C/C	8	5.375	
Overdominant	T/T-C/C	47	6.191	0.063
	C/T	44	5.023	

Table 3: SNPAssoc results for SNP 7 in gene N5 in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	A/A	25	5.720	0.072
	A/G	45	4.978	
	G/G	17	6.941	
Dominant	A/A	25	5.720	0.779
	A/G-A/A	62	5.516	
Recessive	A/A-A/G	70	5.243	***0.038***
	G/G	17	6.941	
Overdominant	A/A-G/G	42	6.214	0.058
	A/G	45	4.978	

Table 4: SNPAssoc results for SNP 4 in gene N5 in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	G/G	38	6.316	0.080
	C/G	48	4.896	
	C/C	6	4.833	
Dominant	G/G	38	6.316	***0.024***
	C/G-C/C	54	4.889	
Recessive	G/G-C/G	86	5.523	0.590
	C/G	6	4.833	
Overdominant	G/G-C/C	44	6.114	0.052
	C/G	48	4.896	

Table 5: SNPAssoc results for SNP 8 in gene N5 in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	A/A	39	6.359	0.075
	A/G	45	4.867	
	G/G	8	5.500	
Dominant	A/A	39	6.359	***0.027***
	A/G-A/A	53	4.962	
Recessive	A/A-A/G	84	5.560	0.957
	G/G	8	5.500	
Overdominant	A/A-G/G	47	6.213	0.031
	A/G	45	4.867	

Table 6: SNPAssoc results for SNP 1 in gene N5 in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	A/A	38	6.447	0.067
	A/G	45	4.956	
	G/G	8	5.125	
Dominant	A/A	38	6.447	***0.019***
	A/G-G/G	53	4.981	
Recessive	A/A-A/G	70	5.639	0.043
	G/G	17	5.125	
Overdominant	A/A-G/G	47	6.217	0.042
	A/G	45	4.956	

Table 7: SNPAssoc results for SNP 1 in gene O1 in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	A/A	31	5.290	0.077
	A/G	28	4.607	
	G/G	22	6.545	
Dominant	A/A	31	5.290	0.808
	A/G-G/G	50	5.460	
Recessive	A/A-A/G	59	4.966	***0.036***
	G/G	22	6.545	
Overdominant	A/A-G/G	53	5.811	0.090
	A/G	28	4.607	

Table 8: SNPAssoc results for SNP 1 in gene S11 in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	T/T	52	6.019	0.171
	C/T	36	4.861	
	C/C	3	6.667	
Dominant	T/T	52	6.019	0.114
	C/T-C/C	39	5.000	
Recessive	T/T-C/T	88	5.545	0.529
	C/C	3	6.667	
Overdominant	T/T-C/C	55	6.055	***0.046***
	C/T	36	4.861	

Table 9: SNPAssoc results for SNP 1 in gene T4 in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	C/C	34	6.412	0.081
	C/G	43	4.884	
	G/G	15	5.800	
Dominant	C/C	34	6.412	0.046
	C/G-G/G	58	5.121	
Recessive	C/C-C/G	77	5.558	0.777
	G/G	15	5.800	
Overdominant	C/C-G/G	49	6.224	***0.032***
	C/G	43	4.884	

Table 10: SNPAssoc results for SNP 1 in gene T1 in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	G/G	54	4.944	0.023
	G/T	25	6.040	
	T/T	5	8.400	
Dominant	G/G	54	4.944	0.028
	G/T-T/T	30	6.433	
Recessive	G/G-G/T	79	5.291	0.024
	T/T	5	8.400	
Overdominant	G/G-C/C	59	5.237	***0.008***
	C/G	25	6.040	

Table 11: SNPAssoc results for SNP 1 in gene A3 in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	T/T	26	6.692	0.018
	C/T	31	5.097	
	C/C	20	4.200	
Dominant	T/T	26	6.692	***0.008***
	C/T-C/C	51	4.745	
Recessive	T/T-C/T	57	5.825	0.044
	C/C	20	4.200	
Overdominant	T/T-C/C	46	5.609	0.483
	C/T	31	5.097	

Table 12: SNPAssoc results for SNP 2 in gene A3 in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	T/T	40	6.250	0.134
	C/T	37	5.000	
	C/C	12	4.917	
Dominant	T/T	40	6.250	***0.044***
	C/T-C/C	49	4.980	
Recessive	T/T-C/T	77	5.649	0.430
	C/C	12	4.917	
Overdominant	T/T-C/C	52	5.942	0.142
	C/T	37	5.000	

Table 13: SNPAssoc results for SNP 1 in gene C3 in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	A/A	31	4.194	0.013
	A/G	45	6.022	
	G/G	12	6.500	
Dominant	A/A	31	4.194	***0.003***
	A/G-G/G	57	6.123	
Recessive	A/A-A/G	76	5.276	0.195
	G/G	12	6.500	
Overdominant	A/A-G/G	43	4.837	0.066
	A/G	45	6.022	

Table 14: SNPAssoc results for SNP 2 in gene C3 in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	G/G	33	4.394	0.023
	A/G	42	6.048	
	A/A	16	6.438	
Dominant	G/G	33	4.394	***0.006***
	A/G-A/A	58	6.155	
Recessive	G/G-A/G	75	5.320	0.179
	A/A	16	6.438	
Overdominant	G/G-A/A	49	5.061	0.120
	A/G	42	6.048	

Table 15: SNPAssoc results for SNP 1 in gene R2 in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	A/A	44	6.136	0.028
	A/G	20	4.050	
	G/G	16	6.062	
Dominant	A/A	44	6.136	0.080
	A/G-G/G	36	4.944	
Recessive	A/A-A/G	64	5.484	0.498
	G/G	16	6.062	
Overdominant	A/A-G/G	60	6.117	***0.007***
	A/G	20	4.050	

Table 16: SNPAssoc results for SNP 1 in gene MSX2 in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	T/T	37	6.405	0.092
	C/T	47	5.021	
	C/C	6	4.833	
Dominant	T/T	37	6.405	***0.028***
	C/T-C/C	53	5.000	
Recessive	T/T-C/T	84	5.631	0.534
	C/C	6	4.833	
Overdominant	T/T-C/C	43	6.186	0.067
	C/T	47	5.021	

Table 17: SNPAssoc results for SNP 1 in gene SPARC in the MDR group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	C/C	38	6.316	0.104
	C/T	38	4.842	
	T/T	7	5.143	
Dominant	C/C	38	6.316	***0.034***
	C/T-T/T	45	4.889	
Recessive	C/C-C/T	76	5.579	0.721
	T/T	7	5.143	
Overdominant	C/C-T/T	45	6.133	0.055
	C/T	38	4.842	

Table 18: SNPAssoc results for SNP 5 in gene N5 in the CTRL group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	T/T	43	5.535	0.000
	C/T	31	4.806	
	C/C	15	2.267	
Dominant	T/T	43	5.535	0.012
	C/T-C/C	46	3.978	
Recessive	T/T-C/T	74	5.230	***0.0002***
	C/C	15	2.267	
Overdominant	T/T-C/C	58	4.690	0.860
	C/T	31	4.806	

Table 19: SNPAssoc results for SNP 3 in gene N5 in the CTRL group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	T/T	44	5.477	0.004
	G/T	34	4.559	
	G/G	11	2.273	
Dominant	T/T	44	5.477	0.017
	G/T-G/G	45	4.000	
Recessive	T/T-G/T	78	5.077	***0.002***
	G/G	11	2.273	
Overdominant	T/T-G/G	55	4.836	0.669
	G/T	34	4.559	

Table 20: SNPAssoc results for SNP 1 in gene N5 in the CTRL group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	A/A	42	5.452	0.047
	A/G	32	4.531	
	G/G	13	3.231	
Dominant	A/A	42	5.452	0.039
	A/G-G/G	45	4.156	
Recessive	A/A-A/G	74	5.054	***0.038***
	G/G	13	3.231	
Overdominant	A/A-G/G	55	4.927	0.548
	A/G	32	4.531	

Table 21: SNPAssoc results for SNP 2 in gene N5 in the CTRL group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	C/C	46	5.522	0.013
	C/G	34	4.353	
	G/G	11	2.818	
Dominant	C/C	46	5.522	***0.012***
	C/G-G/G	45	3.978	
Recessive	C/C-C/G	80	5.025	0.019
	G/G	11	2.818	
Overdominant	C/C-G/G	57	5.000	0.316
	C/G	34	4.353	

Table 22: SNPAssoc results for SNP 4 in gene N5 in the CTRL group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	G/G	44	5.477	0.027
	C/G	38	4.263	
	C/C	10	3.000	
Dominant	G/G	44	5.477	***0.016***
	C/G-C/C	48	4.000	
Recessive	G/G-C/G	82	4.915	0.054
	C/C	10	3.000	
Overdominant	G/G-C/C	54	5.019	0.233
	C/G	38	4.263	

Table 23: SNPAssoc results for SNP 6 in gene N5 in the CTRL group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	G/G	47	5.362	0.044
	A/G	37	4.351	
	A/A	10	3.000	
Dominant	G/G	47	5.362	***0.033***
	A/G-A/A	47	4.064	
Recessive	G/G-A/G	84	4.917	0.052
	A/A	10	3.000	
Overdominant	G/G-A/A	57	4.947	0.343
	A/G	37	4.351	

Table 24: SNPAssoc results for SNP 8 in gene N5 in the CTRL group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	A/A	46	5.457	0.024
	A/G	39	4.231	
	G/G	10	3.000	
Dominant	A/A	46	5.457	***0.014***
	A/G-G/G	49	3.980	
Recessive	A/A-A/G	85	4.894	0.054
	G/G	10	3.000	
Overdominant	A/A-G/G	56	5.018	0.203
	A/G	39	4.231	

Table 25: SNPAssoc results for SNP 7 in gene N5 in the CTRL group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	A/A	29	5.241	0.005
	A/G	40	3.675	
	G/G	21	6.000	
Dominant	A/A	29	5.241	0.250
	A/G-G/G	61	4.475	
Recessive	A/A-A/G	69	4.333	0.022
	G/G	21	6.000	
Overdominant	A/A-G/G	50	5.560	***0.002***
	A/G	40	3.675	

Table 26: SNPAssoc results for SNP 1 in gene O1 in the CTRL group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	A/A	21	5.143	0.046
	A/G	39	4.000	
	G/G	17	6.059	
Dominant	A/A	21	5.143	0.503
	A/G-G/G	56	4.625	
Recessive	A/A-A/G	60	4.400	***0.043***
	G/G	17	6.059	
Overdominant	A/A-G/G	38	5.553	0.022
	A/G	39	4.000	

Table 27: SNPAssoc results for SNP 1 in gene S11 in the CTRL group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	T/T	60	4.900	0.081
	C/T	20	3.808	
	C/C	6	6.500	
Dominant	T/T	60	4.900	0.360
	C/T-C/C	32	4.312	
Recessive	T/T-C/T	86	4.570	0.117
	C/C	6	6.500	
Overdominant	T/T-C/C	66	5.045	***0.044***
	C/T	26	3.808	

Table 28: SNPAssoc results for SNP 1 in gene P6 in the CTRL group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	T/T	60	4.350	0.099
	G/T	27	5.593	
	G/G	5	3.200	
Dominant	T/T	60	4.350	0.179
	G/T-G/G	32	5.219	
Recessive	T/T-G/T	87	4.736	0.259
	G/G	5	3.200	
Overdominant	T/T-G/G	65	4.262	***0.048***
	G/T	27	5.593	

Table 29: SNPAssoc results for SNP 3 in gene P6 in the CTRL group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	C/C	60	4.300	0.073
	C/T	27	5.815	
	T/T	7	4.143	
Dominant	C/C	60	4.300	0.064
	C/T-T/T	34	5.471	
Recessive	C/C-C/T	87	4.770	0.591
	T/T	7	4.143	
Overdominant	C/C-T/T	67	4.284	***0.022***
	C/T	27	5.815	

Table 30: SNPAssoc results for SNP 2 in gene P6 in the CTRL group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	A/A	61	4.361	0.059
	A/T	28	5.750	
	T/T	5	3.200	
Dominant	A/A	61	4.361	0.117
	A/T-T/T	33	5.364	
Recessive	A/A-A/T	89	4.798	0.242
	T/T	5	3.200	
Overdominant	A/A-T/T	66	4.273	***0.026***
	A/T	28	5.75	

Table 31: SNPAssoc results for SNP 1 in gene C3 in the CTRL group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	A/A	28	3.464	0.034
	A/G	44	5.182	
	G/G	7	5.571	
Dominant	A/A	28	3.464	***0.009***
	A/G-G/G	51	5.235	
Recessive	A/A-A/G	72	4.514	0.367
	G/G	7	5.571	
Overdominant	A/A-G/G	35	3.886	0.051
	A/G	44	5.182	

Table 32: SNPAssoc results for SNP 1 in gene NR5A2 in the CTRL group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	C/C	74	4.405	0.046
	C/T	16	6.250	
	T/T	4	3.250	
Dominant	C/C	74	4.405	0.095
	C/T-T/T	20	5.650	
Recessive	C/C-C/T	90	4.733	0.329
	T/T	4	3.250	
Overdominant	C/C-T/T	78	4.346	***0.018***
	C/T	16	6.250	

Table 33: SNPAssoc results for SNP 1 in gene SPARC in the CTRL group.

Model	Genotype	N	ME (mean number of kits)	P-value
Codominant	C/C	45	5.244	0.057
	C/T	30	4.533	
	T/T	8	2.625	
Dominant	C/C	45	5.244	0.084
	C/T-T/T	38	4.132	
Recessive	C/C-C/T	75	4.960	***0.031***
	T/T	8	2.625	
Overdominant	C/C-T/T	53	4.849	0.064
	C/T	30	4.533	

Table 34: SNPAssoc results for SNP 5 in gene N5 in the MDR group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	T/T	37 (41.6%)	49 (56.3%)	0.132
	C/T	43 (48.3%)	30 (34.5%)	
	C/C	9 (10.1%)	8 (9.2%)	
Dominant	T/T	37 (41.6%)	49 (56.3%)	***0.049***
	C/T-C/C	52 (58.4%)	38 (43.7%)	
Recessive	T/T-C/T	80 (89.9%)	79 (90.8%)	0.836
	C/C	9 (10.1%)	8 (9.2%)	
Overdominant	T/T-C/C	46 (51.7%)	57 (65.5%)	0.062
	C/T	43 (48.3%)	30 (34.5%)	

Table 35: SNPAssoc results for SNP 6 in gene N5 in the MDR group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	G/G	36 (40.9%)	49 (57.0%)	0.098
	A/G	45 (51.1%)	31 (36.0%)	
	A/A	7 (8.0%)	6 (7.0%)	
Dominant	G/G	36 (40.9%)	49 (57.0%)	***0.033***
	A/G-A/A	52 (59.1%)	37 (43.0%)	
Recessive	G/G-A/G	81 (92.0%)	80 (93.0%)	0.806
	A/A	7 (8.0%)	6 (7.0%)	
Overdominant	G/G-A/A	43 (48.9%)	55 (64.0%)	0.044
	A/G	45 (51.1%)	31 (36.0%)	

Table 36: SNPAssoc results for SNP 2 in gene N5 in the MDR group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	C/C	39 (43.3%)	50 (58.8%)	0.110
	C/G	44 (48.9%)	29 (34.1%)	
	G/G	7 (7.8%)	6 (7.1%)	
Dominant	C/C	39 (43.3%)	50 (58.8%)	***0.040***
	C/G-G/G	51 (56.7%)	35 (41.2%)	
Recessive	C/C-C/G	83 (92.2%)	79 (92.9%)	0.856
	G/G	7 (7.8%)	6 (7.1%)	
Overdominant	C/C-G/G	46 (51.1%)	56 (65.9%)	0.047
	C/G	44 (48.9%)	29 (34.1%)	

Table 37: SNPAssoc results for SNP 1 gene SOX5 in the MDR group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	C/C	35 (40.7%)	44 (57.9%)	0.051
	C/T	39 (45.3%)	21 (27.6%)	
	T/T	12 (14.0%)	11 (14.5%)	
Dominant	C/C	35 (40.7%)	44 (57.9%)	0.028
	C/T-T/T	51 (59.3%)	32 (42.1%)	
Recessive	C/C-C/T	74 (86.0%)	65 (85.5%)	0.924
	T/T	12 (14.0%)	11 (14.5%)	
Overdominant	C/C-T/T	47 (54.7%)	55 (72.4%)	***0.019***
	C/T	39 (45.3%)	21 (27.6%)	

Table 38: SNPAssoc results for SNP 1 in gene T1 in the MDR group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	G/G	58 (70.7%)	44 (55.7%)	0.108
	G/T	22 (26.8%)	30 (38.0%)	
	T/T	2 (2.4%)	5 (6.3%)	
Dominant	G/G	58 (70.7%)	44 (55.7%)	***0.047***
	G/T-T/T	24 (29.3%)	35 (44.3%)	
Recessive	G/G-G/T	80 (97.6%)	74 (93.7%)	0.219
	T/T	2 (2.4%)	5 (6.3%)	
Overdominant	G/G-T/T	60 (73.2%)	49 (62.0%)	0.130
	G/T	22 (26.8%)	30 (38.0%)	

Table 39: SNPAssoc results for SNP 1 in gene T4 in the MDR group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	C/C	24 (26.4%)	34 (39.5%)	0.073
	C/G	50 (54.9%)	33 (38.4%)	
	G/G	17 (18.7%)	19 (22.1%)	
Dominant	C/C	24 (26.4%)	34 (39.5%)	0.061
	C/G-G/G	67 (73.6%)	52 (60.5%)	
Recessive	C/C-C/G	74 (81.3%)	67 (77.9%)	0.573
	G/G	17 (18.7%)	19 (22.1%)	
Overdominant	C/C-G/G	41 (45.1%)	53 (61.6%)	***0.026***
	C/G	50 (54.9%)	33 (38.4%)	

Table 40: SNPAssoc results for SNP 1 in gene C4 in the MDR group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	C/C	75 (83.3%)	62 (73.8%)	***0.041***
	C/T	12 (13.3%)	22 (26.2%)	

Table 41: SNPAssoc results for SNP 1 in gene MSX2 in the MDR group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	T/T	37 (41.6%)	47 (54.7%)	0.085
	C/T	45 (50.6%)	33 (38.4%)	
	C/C	7 (7.9%)	6 (7.0%)	
Dominant	T/T	37 (41.6%)	47 (54.7%)	***0.050***
	C/T-C/C	52 (58.4%)	39 (45.3%)	
Recessive	T/T-C/T	82 (92.1%)	80 (93.0%)	0.822
	C/C	7 (7.9%)	6 (7.0%)	
Overdominant	T/T-C/C	44 (49.4%)	53 (61.6%)	0.104
	C/T	45 (50.6%)	33 (38.4%)	

Table 42: SNPAssoc results for SNP 1 in gene SPARC in the MDR group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	C/C	34 (42.5%)	45 (60.0%)	0.058
	C/T	38 (47.5%)	22 (29.3%)	
	T/T	8 (10.0%)	8 (10.7%)	
Dominant	C/C	34 (42.5%)	45 (60.0%)	0.029
	C/T-T/T	46 (57.5%)	30 (40.0%)	
Recessive	C/C-C/T	72 (90.0%)	67 (89.3%)	0.891
	T/T	8 (10.0%)	8 (10.7%)	
Overdominant	C/C-T/T	42 (52.5%)	53 (70.7%)	***0.019***
	C/T	38 (47.5%)	22 (29.3%)	

Table 43: SNPAssoc results for SNP 1 in gene P6 in the CTRL group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	T/T	85 (69.7%)	41 (56.2%)	0.017
	G/T	28 (23%)	30 (41.1%)	
	G/G	9 (7.4%)	2 (2.7%)	
Dominant	T/T	85 (69.7%)	41 (56.2%)	0.057
	G/T-G/G	37 (30.3%)	32 (43.8%)	
Recessive	T/T-G/T	113 (92.6%)	71 (97.3%)	0.152
	G/G	9 (7.4%)	2 (2.7%)	
Overdominant	T/T-G/G	94 (77.0%)	43 (58.9%)	***0.007***
	G/T	28 (23.0%)	30 (41.1%)	

Table 44: SNPAssoc results for SNP 2 in gene P6 in the CTRL group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	A/A	85 (68.5%)	43 (58.1%)	0.106
	A/T	30 (24.2%)	28 (37.8%)	
	T/T	9 (7.3%)	3 (4.1%)	
Dominant	A/A	85 (68.5%)	43 (58.1%)	0.138
	A/T-T/T	39 (31.5%)	31 (41.9%)	
Recessive	A/A-A/T	115 (92.7%)	71 (95.9%)	0.347
	T/T	9 (7.3%)	3 (4.1%)	
Overdominant	A/A-T/T	94 (75.8%)	46 (62.2%)	***0.042***
	A/T	30 (24.2%)	28 (37.8%)	

Table 45: SNPAssoc results for SNP 3 in gene P6 in the CTRL group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	C/C	84 (67.2%)	43 (57.3%)	0.021
	C/T	30 (24.0%)	30 (40.0%)	
	T/T	11 (8.8%)	2 (2.7%)	
Dominant	C/C	84 (67.2%)	43 (57.3%)	0.162
	C/T-T/T	41 (32.8%)	32 (42.7%)	
Recessive	C/C-C/T	114 (91.2%)	73 (97.3%)	0.069
	T/T	11 (8.8%)	2 (2.7%)	
Overdominant	C/C-T/T	95 (76.0%)	45 (60.0%)	***0.017***
	C/T	30 (24.0%)	30 (40.0%)	

Table 46: SNPAssoc results for SNP 1 in gene A3 in the CTRL group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	T/T	36 (34.3%)	14 (22.2%)	0.072
	C/T	46 (43.8%)	39 (61.9%)	
	C/C	23 (21.9%)	10 (15.9%)	
Dominant	T/T	36 (34.3%)	14 (22.2%)	0.093
	C/T-C/C	69 (65.7%)	49 (77.8%)	
Recessive	T/T-C/T	82 (78.1%)	53 (84.1%)	0.335
	C/C	23 (21.9%)	10 (15.9%)	
Overdominant	T/T-C/C	59 (56.2%)	24 (38.1%)	***0.022***
	C/T	46 (43.8%)	39 (61.9%)	

Table 47: SNPAssoc results for SNP 2 in gene A3 in the CTRL group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	T/T	65 (53.3%)	32 (42.1%)	0.023
	C/T	42 (34.4%)	40 (52.6%)	
	C/C	15 (12.3%)	4 (5.3%)	
Dominant	T/T	65 (53.3%)	32 (42.1%)	0.125
	C/T-C/C	57 (46.7%)	44 (57.9%)	
Recessive	T/T-C/T	107 (87.7%)	72 (94.7%)	0.089
	C/C	15 (12.3%)	4 (5.3%)	
Overdominant	T/T-C/C	80 (65.6%)	36 (47.4%)	***0.011***
	C/T	42 (34.4%)	40 (52.6%)	

Table 48: SNPAssoc results for SNP 2 in gene C3 in the CTRL group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	G/G	87 (69.6%)	43 (56.6%)	0.170
	A/G	33 (26.4%)	28 (36.8%)	
	A/A	5 (4.0%)	5 (6.6%)	
Dominant	G/G	87 (69.6%)	43 (56.6%)	***0.050***
	A/G-A/A	38 (30.4%)	33 (43.4%)	
Recessive	G/G-A/G	120 (96.0%)	71 (93.4%)	0.421
	A/A	5 (4.0%)	5 (6.6%)	
Overdominant	G/G-A/A	92 (73.6%)	48 (63.2%)	0.120
	A/G	33 (26.4%)	28 (36.8%)	

Table 49: SNPAssoc results for SNP 1 in gene R2 in the CTRL group.

Model	Genotype	0 = High performers	1 = Low performers	P-value
Codominant	A/A	48 (44%)	36 (52.9%)	0.071
	A/G	45 (41.3%)	17 (25.0%)	
	G/G	16 (14.7%)	15 (22.1%)	
Dominant	A/A	48 (44.0%)	36 (52.9%)	0.248
	A/G-G/G	61 (56.0%)	32 (47.1%)	
Recessive	A/A-A/G	93 (85.3%)	53 (77.9%)	0.213
	G/G	16 (14.7%)	15 (22.1%)	
Overdominant	A/A-G/G	64 (58.7%)	51 (75.0%)	***0.025***
	A/G	45 (41.3%)	17 (25.0%)	

Table 50: SNPAssoc results for SNP 1 in gene C4 in the CTRL group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	C/C	87 (70.7%)	61 (83.6%)	0.119
	C/T	33 (26.8%)	11 (15.1%)	
	T/T	3 (2.4%)	1 (1.4%)	
Dominant	C/C	87 (70.7%)	61 (83.6%)	***0.039***
	C/T-T/T	36 (29.3%)	12 (16.4%)	
Recessive	C/C-C/T	120 (97.6%)	72 (98.6%)	0.598
	T/T	3 (2.4%)	1 (1.4%)	
Overdominant	C/C-T/T	90 (73.2%)	62 (84.9%)	0.051
	C/T	33 (26.8%)	11 (15.1%)	

Table 51: SNPAssoc results for SNP 1 in gene NR5A2 in the CTRL group.

Model	Genotype	0 = Low performers	1 = High performers	P-value
Codominant	C/C	102 (81.0%)	54 (72.0%)	0.145
	C/T	20 (15.9%)	20 (26.7%)	
	T/T	4 (3.2%)	1 (1.3%)	
Dominant	C/C	102 (81%)	54 (72.0%)	0.144
	C/T-T/T	24 (19%)	21 (28.0%)	
Recessive	C/C-C/T	122 (96.8%)	74 (98.7%)	0.396
	T/T	4 (3.2%)	1 (1.3%)	
Overdominant	C/C-T/T	106 (84.1%)	55 (73.3%)	***0.050***
	C/T	20 (15.9%)	20 (26.7%)	

APPENDIX 2

SNPs distribution in genes

Gene sequences of *Canis lupus* or *Mustela putorius furo* were retrieved from UCSC genome browser as both dog and ferret are considered to be close relatives of mink (Anistoroaei and Christensen, 2006; Anistoroaei *et al.,* 2009). Using BLAST tool in NCBI+, the sequences were compared with the mink contig database.

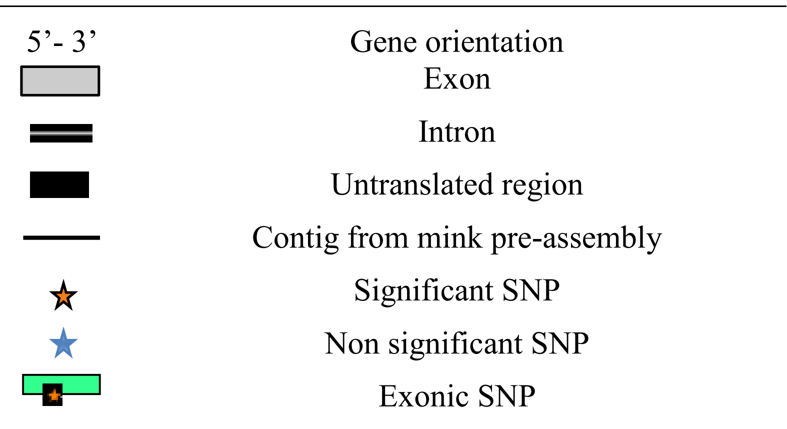


Table 1: Gene maps showing the distribution of SNPs

Fig	Gene	Map	Size (kbps)	#of exons	Significance ($p<0.05$)
1	C3		60	20	Tail and window analysis
2	C4		2	2	Window analysis

Fig	Gene	Map	Size (kbps)	#of exons	Significance (p<0.05)
3	O1		30	20	Tail analysis
4	R2		90	24	Tail and window analysis
5	T1		60	10	Tail and window analysis
6	N5		20	29	Tail and window analysis
7	P6		40	12	Tail and window analysis
8	MSX 2		6	2	Tail and window analysis

Fig	Gene	Map	Size (kbps)	#of exons	Significance ($p<0.05$)
9	NR5 A2		100	7	Tail and window analysis
10	SOX 5		400	15	Tail and window analysis
11	SPA RC		20	10	Tail and window analysis
12	S11		70	17	Tail analysis
13	T4		20	12	Tail and window analysis