# Molecular Phenotypes Associated with Heifer Fertility

Mackenzie Ann Marrella

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Animal and Poultry Science

Fernando H. Biase, Chair
Vitor R.G. Mercadante
Sarah E. Moorey
Gota Morota

April 22, 2024
Blacksburg, Virginia

Keywords: molecular marker, heifer fertility, genomic variants

Molecular Phenotypes Associated with Heifer Fertility

Mackenzie Ann Marrella

## ABSTRACT

Infertility and sub-fertility are some of the largest barriers to the profitability and sustainability of cattle production systems. There has been a push in the last several decades to decrease the burden faced by producers as a result of poor fertility. Typically, a producer would utilize traditional selection methods to improve fertility, however the complex, polygenic nature and low heritability of these traits makes this an inefficient method. Therefore, researchers have been investigating alternative strategies to improve female fertility in cattle. One strategy that has been a popular research subject in recent years is the identification of molecular markers that could be utilized for selection. An ideal biomarker should be able to be sampled non-invasively, however, many of the previous studies conducted in cattle have utilized biological fluids that require invasive collection methods. Additionally, there have been very few studies that have been focused on the identification of molecular markers associated with heifer fertility. Therefore, the aim of our work was to identify molecular markers in blood samples collected from beef and dairy heifers that could discriminate these heifers based on their fertility potential. Our investigations of the genome resulted in the identification of 3 single nucleotide polymorphisms (SNPs) that were associated with fertility, 16 SNPs that were associated with health, and 29 SNPs that were associated with both health and fertility. Some variants are able to indirectly impact the phenotype by affecting gene expression, termed expression quantitative trait loci (eQTLs), therefore we attempted to identify eQTLs associated with fertility. However, while we were unable to find significant connections between the SNP-gene pairs and fertility, we developed a new method for the identification of association between SNPs and transcript abundance. Using this method on data from purebred Angus and Holstein heifers, we identified two significant genes (*APMAP* and *DNAI7*) that were differentially expressed between fertile and sub-fertile heifers. That same study also identified one protein whose abundance differed between the fertile and sub-fertile heifers. Importantly, the integration of multiple levels of biological information (genome, transcriptome, proteome) yielded a biological profile that was able to correctly separate 21/22 heifers based on their fertility potential. Finally, investigation of the metabolome revealed one metabolite (2-dehydro-D-gluconate) that was differentially abundant between fertile and sub-fertile heifers. In conclusion, this work sheds light on the intricate molecular landscape underlying heifer fertility and provides several molecular markers that could potentially be utilized to select heifers with superior reproductive potential.

Molecular Phenotypes Associated with Heifer Fertility

Mackenzie Ann Marrella

## GENERAL AUDIENCE ABSTRACT

Before the early 2000s, many producers were heavily selecting for production traits without accounting for the negative relationship between many production traits and fertility. The large decrease in fertility that occurred as a result of these selection practices put a heavy burden on producers. Replacement heifer development is the third largest cost incurred by producers, behind feed and labor. Consequently, the failure of a heifer to conceive in her first breeding season translates directly into financial loss for the producer and lasting consequences on the animal's longevity and performance in subsequent breeding seasons. To make improvements in a trait, or traits, of interest, producers often selectively breed two animals with desirable characteristics. However, the complex nature of fertility traits limits the effectiveness of this method. As a result, researchers have been attempting to identify biological molecules whose abundance differs between cattle of differing fertility potential, termed molecular markers, that could be used to identify superior cattle earlier and more accurately. While a good amount of research has been conducted in mature cows and in a laboratory setting, very few studies have attempted to identify molecular markers of fertility in heifers. Therefore, the objective of this work was to identify different biological molecules (genomic variants, genes, proteins, and metabolites) whose abundance differed between heifers with varying fertility potential. Investigation into DNA variations led to the identification of three variants that were associated with fertility, 16 variants that were associated with health, and 29 variants that were associated with both health and fertility. Given that some variants can impact a trait by changing gene expression, we attempted to identify variations in RNA that were having this effect on heifer fertility. Although some variants were found to influence gene expression, we were unable to correlate these changes with fertility differences. However, in a different study, we were able to identify two genes (*APMAP* and *DNAI7*), as well as one protein (alpha-ketoglutarate-dependent-dioxygenase FTO), that differed significantly between fertile and sub-fertile heifers. Importantly, the results of this study allowed us to create a biological profile that was capable of accurately distinguishing 21/22 heifers based on their fertility potential. Finally, investigation of the metabolite profile revealed one metabolite (2-dehydro-D-gluconate) that was differentially abundant between fertile and sub-fertile heifers. Overall, this work sheds light on the complex nature of heifer fertility and provides several potential molecular that could be used to distinguish between heifers of varying reproductive potential.

**Acknowledgements**

First and foremost, I would like to thank my parents, Paul and Tracy Marrella, for their never-ending support and encouragement. Thank you for always allowing me to pursue what I love and for making sure I had everything I needed in order to do that.

To my first mentor, Dr. Heather Bradford, words cannot describe the instrumental role you played in me getting to where I am today. Thank you for giving me the opportunity to work with you on my first research project. I don't think I would have ever found my love of coding and analysis if it weren't for you. I know I wasn't always the easiest to deal with, but thank you for always being patient with me as I figured out research and coding and seemed to always be bringing you problem after problem. Lastly, thank you for always going above and beyond to help me accomplish my goals and become the best scientist I can be.

To my committee members, Dr. Vitor Mercadante, Dr. Gota Morota, and Dr. Sarah Moorey, thank you for pushing me outside of my comfort zone and challenging me to think about things in a different way.

To my advisor, Dr. Fernando Biase, thank you for always taking the time to teach me and talk things out until I understood them. Thank you for always putting up with me, I know I'm not the easiest to deal with at times. Thank you for always believing in me, even in the moments when I didn't believe in myself. I am my own harshest critic, but you always made sure to beat into my head that one bad moment doesn't define who I am.

To Jada, Gustavo, Chace, and Kaitlyn, my lab mates past and present, thank you for always being there when I needed to talk through things, rant, or just distract myself in the office. I also want to give a special thanks to Jada because I know I would not be here without your support and friendship. I wouldn't want to be trauma bonded with anybody else.

To all of the friends I have made throughout my time at Virginia Tech, your friendship means the world to me. Grad school is tough and trying at times, but you have made even the worst moments bearable. Lastly, thank you to Deano and Birdsong Farm for making the world's best chai and fueling me throughout grad school.

## Table of Contents

## LIST OF FIGURES

## LIST OF TABLES

# CHAPTER 1: Literature Review

## 1.1 ABSTRACT

Infertility and subfertility are some of the largest barriers to the profitability and sustainability of livestock production systems. Given that the majority of producers are raising their own replacement heifers, heifer fertility is of particular interest. Moreover, heifer fertility has been shown to be correlated with lifetime productivity and profitability. Therefore, heifers producing a calf in their first breeding season is crucial for producers. Efforts have been made to improve heifer fertility over the years; however, progress has been slow due to the low heritability of female reproductive traits, particularly in heifers. As a result, producers have been utilizing technologies such as antral follicle count, reproductive tract score, estrus synchronization, and expected progeny difference to select reproductively superior heifers and increase pregnancy rates. However, one limitation of these technologies is that most are utilized around the time of breeding, meaning producers have already invested considerable time and resources into raising these heifers. As a result, there has been an increase in attempts to develop technologies that could be used to identify heifers with superior reproductive potential earlier. Advances in sequencing and molecular profiling have allowed researchers to begin identifying differences in gene expression, microRNA, proteins, and metabolites that could be used to discriminate heifers based on their fertility potential much earlier. The aim of this review is to discuss the status of heifer fertility as well as the ongoing efforts to improve heifer selection for greater profitability and sustainable agriculture.

## 1.2 INTRODUCTION

The 2017 census reported by the U.S. Department of Agriculture's National Agricultural Statistics Service (NASS-USDA) indicated that 768,542 farms (729,046 and 54,599 farms with beef and milk cows, respectively) had cattle and heifers that calved in their inventory. In dairy and cow-calf beef production systems, farmers must make an essential decision of replacing cows in their herd, due to sale, diseases, mortality, reproductive deficiencies, or poor productivity [1-4]. For instance, the annual herd cull rate is in the range of 30-35% in dairy [5] and 15-20% in beef herds. To maintain continuous production, farmers often replace culled cows with heifers raised with the specific purpose of entering the breeding herd. Therefore, heifer replacement programs are one of the critical factors for sustainable cattle operations.

A compilation of data from NASS-USDA reported in the January cattle census, over the past ten years, indicates that around 10 million replacement heifers have been developed yearly since 2014, averaging 5.7 and 4.6 million beef and dairy heifers, respectively (Figure 1.1). The yearly stock of replacement heifers in the U.S. is approximately 20% and 50% of the number of mature cows in beef and dairy systems, respectively. Within the farming requirements for cow-calf or dairy production systems, virtually all heifers enter the replacement program using an equivalent amount of farm and natural resources. Still, on average (between 2013 and 2021), only 34.4% of the replacement heifers enter the breeding herd between 2013 and 2021 (ERS-USDA, 2022). Thus, farmers must conduct heifer development programs with a balance between careful management and selection of heifers to prevent economic losses.

Heifer development programs are an essential component for a successful cattle enterprise in both beef and dairy systems. In this review, our goal is to highlight the importance of heifer fertility for the U.S. cattle business and to discuss ongoing efforts to improve the early selection of heifers based on their fertility potential. We will focus on the following topics: 1-the economic impact of heifers that are raised as replacements but do not contribute as a cow in the breeding pool, 2-genetic aspects of heifer fertility, 3- the potential for molecular markers to aid in the decision-making process of which heifer should be raised in the replacement pool.

## 1.3 THE STATUS OF FERTILITY IN REPLACEMENT HEIFERS

The goal of raising heifers for replacement is to introduce them into a breeding program. Beyond being in good health, two major features for a heifer to enter a breeding herd are: 1- to be reproductively mature, having reached puberty and preferably having experienced one or more estrous cycles; and 2- have the appropriate body weight reached under controlled average daily gain. Careful heifer health, growth, and reproductive management maximize the heifer's potential for producing a calf within a desired window of time.

### 1.3.1 Dairy heifers

A national survey by the NASS-USDA, conducted in 2016, identified that 82.9% or more of dairy farms with a heard size equal to or greater than 50 cows use artificial insemination in their reproductive management. A recent analysis by Kuhn et al. [6] showed that overall conception rate

to artificial insemination for U.S. Holstein heifers is 57%. Between 50% and 65% of the Holstein heifers conceive and produce a calve from the first service [6-11]. By comparison, between 2.92% and 8.68% of the heifers do not conceive after multiple consecutive rounds of artificial insemination [7, 11] (Figure 1.2). An average of ~5% of heifers not becoming pregnant to multiple services translates to over 220,000 heifers being culled yearly (229.9 thousand on average, since 2010, NASS-USDA, Figure 1.1) after multiple consecutive artificial inseminations.

### 1.3.2 Beef heifers

A national survey of cow-calf operations that bred heifers for calving in 2017 indicated that most of the heifers were bred exclusively by bulls (76%), followed by heifers bred by artificial insemination and exposed to bulls (15.1%), or bred exclusively by artificial insemination (3.4%) [12]. A recent compilation of records of heifer pregnancy percentages showed that an average of 85% (range 64%-95%) [13-29] of heifers will become pregnant to artificial insemination or to a bull in their first breeding season. By comparison an average of 15% of beef heifers will not become pregnant in their first breeding season (range 4.8%-36%) (Figure 1.3). This proportion of heifers bred for replacement that do not calve by ~25 months of age is equivalent to an average of 850,700 heifers culled yearly due to fertility limitations.

## 1.4 ECONOMICS OF DELAYED HEIFER CULLING DUE TO REPRODUCTIVE INCAPACITY

Raising replacement heifers is one of the largest expenses incurred by producers. It has been estimated that raising a replacement heifer from birth to calving costs producers ~$1,400 for beef heifers [30] and ~$2,000 for dairy heifers [31]. Since the period from birth to calving is financially unproductive, producers need to implement proper management to avoid severe financial losses. With adequate body weight and reproductive maturity, dairy heifers will likely produce the greatest economic return when calving between 23 and 25 months of age [31]. Therefore, any heifer calving outside of this timeframe represents a potential barrier to profit maximization. One management strategy to minimize these losses is to identify and cull reproductively inefficient heifers earlier. It has been estimated that culling a heifer between 21.5-23.4 months of age results in a loss of $603/head. However, culling a heifer around the time of breeding (10.1-15.7 months of age) only results in a loss of $272/head [32]. Thus, the earlier a producer can identify and cull the inefficient cattle, the greater their potential net profit.

## 1.5 GENETICS OF HEIFER FERTILITY TRAITS

Reproduction is a complex biological function in animals. In females, it involves the coordination of several physiological and cellular processes, most of which have significant influence of genetic variation. For over a decade, researchers have identified several traits that have guided the selection of cattle based on their reproductive fitness. Specific to females, there are approximately 21 traits being measured to assess reproductive performance in heifers (Figure 1.4).

### 1.5.1 Heritability of heifer fertility traits

The comprehension of the additive component of economically important traits is key for artificial selection. The estimates of heritability for fertility related traits in heifers range from 0.007 to 0.324 ( [33-51] Figure 1.4). Altogether, heritability estimates for fertility traits in heifers are low (< 0.2), and only few examples show moderate heritability (>0.2). Most of these examples of moderate heritability are related to the age that the heifer may attain reproductive maturity. In beef heifers, these examples include age of puberty (0.221 [43]) and reproductive tract score (0.32). Whereas in dairy heifers, these examples include age at conception (0.312 [48]), age at first calving (0.296 [42]), and age at first insemination (0.227 [41], 0.324 [42]). As reported for other complex traits [52, 53], there is a small additive genetic component to most fertility traits in heifers.

As indicated in the breeders' equation, heritability is directly related to genetic progress. Although, genetic gains can be achieved in traits with low heritability [54], improving heifer fertility traits through traditional use of breeding values remains a critical challenge in cattle production systems world-wide. To that end, researchers have sought to understand the impact of variations in the genomic and molecular profiles on the reproductive outcome in heifers.

### 1.5.2 Genomic variants

Genetic selection for fertility in cattle is hindered by both the low heritability, especially for heifer fertility traits, and the highly polygenic nature of reproductive traits [55, 56]. As a result, improvements to cattle fertility have been slow. Efforts to improve fertility over the last decade have focused on identifying genomic variants that could potentially be used to select females with greater reproductive efficiency. The development of arrays for high throughput genotyping was a critical step in this search for genomic variations that influence economically important traits in livestock [57]. This technology allowed researchers to be able to identify genomic variants associated with cattle fertility traits using genome wide association studies (GWAS).

While several reports have identified quantitative trait loci with markers flanking regions associated with fertility traits in heifers (see table 1 in [58]), recent efforts have focused on the genome wide identification of single nucleotide polymorphisms (SNPs) associated with similar or the same phenotypes in heifers. We compiled 1183 SNPs that have been associated with heifer fertility (Figure 1.5A). Most SNPs have been associated with heifer conception rate [7, 59-61], followed by times bred to achieve pregnancy [7, 61]. Other traits related with heifer fertility (i.e.: pregnancy outcome after several artificial inseminations or after a breeding season) [62], antral follicle count [63], age at first calving [64], reproductive tract score [63], and pregnancy rate [64] have also been the focus of some studies. Only 87.5% of those SNPs are currently annotated in the European Variant Archive ("9903_GCA_002263795.2_current_ids") and can be identified as broadly distributed across the genome (Figure 1.5B). Most notably, there is no overlap between SNPs significantly associated traits involving heifer fertility (Figure 1.5C). Collectively these data show the polygenic nature of fertility in heifers, and the complexity for developing strategies for more efficient selection of heifers based on their fertility potential.

Additionally, deCamargo et al. was specifically interested in polymorphisms in *JY-1* and their effects on cattle reproductive traits. Importantly, this gene codes for the protein JY-1, which has been shown to be associated with folliculogenesis and early embryo development in cattle [65].

Four significant SNPs identified in *JY-1* were found to be associated with age at first calving. It is important to note that this study named significant SNPs based on their location in relation to the first exon of *JY-1*, with the first nucleotide of the first exon being named "1". The four significant SNPs in this study were located at positions -91, 392, 13043, and 13084. Of these SNPs, one is located in exon 2 (13043), two are located in an intron (392, 13084), and one is located in the promoter region of the gene (-91) [66].

## 1.6 POTENTIAL CAUSES OF INFERTILITY

### 1.6.1 Metabolism

Pregnancy is an energetically demanding process for the mother, so meeting energy requirements is essential for reproductive success. Maintaining adequate nutritional status is of particular importance in heifers and other young mammals, as they are still growing. Maternal growth and pregnancy are both energetically demanding processes, therefore, if a heifer has inadequate energy reserves, she will preferentially put the energy towards maintenance and growth rather than pregnancy [53]. This metabolic control of reproduction is regulated by several key hormones and metabolites including growth hormone, leptin, and insulin-like growth factor 1.

Leptin is a protein hormone that has shown to be an important factor linking metabolism and reproduction [67-69], as it is involved in feed intake, energy partitioning, and metabolism [70]. Moreover, genetic variants in the leptin gene have been associated with fertility traits in cattle. One study in Holstein Friesian heifers identified several single nucleotide polymorphisms in the leptin gene that were associated with fertility traits in cattle. The leptin SNP A1457G (G>A) was associated with total number of AI services in nulliparous heifers, where heifers with the genotype GG required 0.13 fewer services compared to the heterozygous animals. This SNP was also associated with days to conception. Interestingly, the animals with the AA genotype conceived 29 days earlier than the heterozygous animals. Another SNP, UASMS2 (C>T), was found to have the greatest effect on heifer fertility in this study. Animals with the TT genotype at this position had more services per conception (increase of 1.1) and received a greater number of total services (increase of 1.3). The final SNP associated with heifer fertility in this study was A59V (C>T). Heifers with a CC genotype were younger both at their first service (by 16 days) and first calving (by 15 days) [71]. A different study in *Bos taurus* x *Bos indicus* crosses showed that a different SNP in the leptin gene (g.92450765, G>A) was associated with services per conception in heifers. They found that animals with the GG and AG genotype required more services for a successful conception (2.11 and 0.69 times more, respectively) compared to animals with the AA genotype [72]. However, it is unknown if these SNPs are having a causative effect or if they are in linkage disequilibrium with the causative SNP. Therefore, further research would be needed before these findings could be utilized for selection.

Insulin-like Growth Factor 1 (IGF-1) is a peptide produced by important reproductive organs, such as the hypothalamus, ovaries, oviducts, and uterus [73-75], that plays an essential role in reproduction by influencing production of GnRH, secretion of pituitary gonadotropin, and the susceptibility of the ovary to FSH and LH [76]. IGF-1 has been associated with several important reproductive traits, including age at first calving (AFC) [77, 78] and conception rate to first service [79]. Since IGF-1 is thought to be an important metabolic signal for the initiation of puberty [80],

it is unsurprising that reduced IGF-1 serum concentrations have been associated with an increase in AFC [81]. There have been numerous studies that have identified variants related to IGF-1, however, they have conflicting results about the physiological effect of those variants on IGF-1 serum concentrations and fertility. One study in dairy cows have found that animals with the CC genotype at the *Sna*BI site in *IGF-1* (T>C) have higher IGF-1 serum concentrations and better fertility [82]. However, other studies have found that the presence of the T allele at the same position is associated with higher serum IGF-1 concentration and better fertility [83, 84]. Therefore, further studies need to be conducted in order to elucidate the true effect of this variant on serum IGF-1 concentrations. It is important to note that these SNPs have been identified in cows, therefore these results may not translate to heifers, as the correlation between cow and heifer fertility is quite low [85, 86]. In addition, these cows may also be in a negative energy balance as a result of lactation, which could be triggering the negative effects of these variants. Furthermore, polymorphisms in the gene growth hormone receptor (*GHR*) have been associated with differences in IGF-1 concentrations in cattle. One study in pre-pubertal Holstein heifers showed that heifers with the genotype GG at the NsiI site of *GHR* (A>G) had increased blood IGF-1 concentrations [87]. While these results are promising, more research is required to understand the role IGF-1 and its variants on heifer fertility.

### 1.6.2 Immune related causes

Over the years, the connection between reproduction and the immune system has become well established in the literature. The immune system plays an important role in numerous reproductive processes including follicle differentiation [88, 89], ovulation [88, 90], corpus luteum formation and regression [91-93], and maternal recognition and establishment of pregnancy [94-96]. However, it also plays a role in fertility by defending the uterus and endometrium from pathogens. Diseases of the reproductive tract have been shown to contribute to infertility and reproductive issues by leading to abnormal follicle development and function as well as impaired oocyte development and maturation [97]. Therefore, proper function of the immune system is essential for reproductive success in cattle.

Toll like receptors (TLR) are innate immune receptors that are responsible for recognizing microbes [98, 99], and they play an important role in the initial defense of the endometrium against pathogens [100]. Activation of TLRs leads to the production of cytokines and chemokines, which function to eliminate pathogens from tissues such as the endometrium [101-103]. Several studies conducted in cows have identified SNPs that are associated with increased risk for disease [104-106]. One study conducted in Holstein cows found that three SNPs in the *TLR9* gene (*TLR*9_A945G, *TLR*9_G1187A, and *TLR*9_C2788T), where the minor allele (G, A, and T, respectively) was associated with increased risk for metritis. The same study also identified one SNP in the gene *TLR4* (*TLR4*_T610C) and one SNP in the gene *TLR6* (*TLR6*_G14578A), where the minor allele (T and G, respectively) was associated with lower risk for clinical endometritis. Lastly, one significant SNP was identified in the gene *TLR2* (*TLR2*_C9564T), and the major allele (C) was associated with an increased risk for cytologic endometritis [107]. It is important to note that the effects of these SNPs may not be observed in heifers due to the metabolic deficits incurred by cows after parturition. These deficits may be influencing the immunity of these cows as well as their ability to clear infections [108].

### 1.6.3 Embryo survival

Embryo mortality is one of the greatest barriers to reproductive efficiency and profitability of cattle production systems [109, 110]. Embryo development and the successful establishment of pregnancy are dependent on a variety of maternal, paternal, and embryonic factors [111-113], making this process incredibly complex. Proper coordination and interaction between these factors are crucial, as disruptions in essential pathways have been shown to negatively affect embryo survival [114-116]. Early embryo development is also an energetically demanding process, and, as a result, mitochondria have to increase their ATP output to meet this demand [117]. However, if the oocyte doesn't have sufficient mitochondria and ATP content, it can lead to fertilization failure and abnormal embryo development [118, 119]. This increased mitochondrial output is necessary for embryo development, but it also leads to increased reactive oxygen species (ROS). ROS accumulation can be toxic and can lead to a state of oxidative stress [120]. Early embryos are highly susceptible to the effects of both oxidative stress and heat stress, and exposure to these stressors leads to lower embryo survival due to their effects on fertilization and development [121-123]. Thus, embryos have to have mechanisms to mitigate these effects and remain in a state of homeostasis.

Signal transducer and activator proteins (STAT) are transcription factors that are involved in cytokine signaling [124]. *STAT5A* is a member of the INF-τ signal transduction pathway, and a study in mice showed that these proteins are essential for the development of a functional corpus luteum [125]. Additionally, mutations in the gene *STAT5A* have shown to be associated with embryonic survival in cattle. One study showed that the G allele in SNP12195(G>C) of *STAT5A* is associated with lower fertilization and embryonic survival [115]. Similarly, another STAT gene, *STAT3*, has also been shown to be important for embryonic survival. *STAT3* is involved in cumulus expansion and oocyte maturation, and the inhibition of *STAT3* resulted in the inhibition of cumulus expansion [126]. A study by Khatib et al. in 2009 showed that oocytes collected from ovaries with the genotype AA at SNP25402(A>C) in *STAT3* are associated with higher fertilization rate. This study also looked at SNP interactions, and they found that when the genotype of SNP19069 (A>G) in *STAT1* was AA, embryos from ovaries that had the genotype AC at SNP25402 in *STAT3* had higher survival rates [114].

Mitochondria are responsible for producing the majority (>90%) of the ATP necessary for cellular functions [118], and they are one of the only sources of energy for the early stages of embryo development [127-129]. Coenzyme Q9 (COQ9) is a protein that is critical for cellular energy metabolism due to its role in the biosynthesis of COQ10 [130, 131]. COQ10 is a member of the electron transport system that plays an important role in ATP synthesis. Mutations in the gene *COQ9* have been shown to play an important role in embryo survival. A study by Ortega et al. showed that a mutation (rs109301586, G>A) in *COQ9* results in a modification of mitochondrial respiratory function and oxidative phosphorylation. The A allele in this position has been associated with higher oocyte mitochondrial content, which could lead to increased embryo survival. Moreover, the AA genotype in this SNP has also shown to be beneficial for embryo survival through its effects on cellular energy requirements [132]. Oocytes with this genotype also had increased mitochondrial DNA content, which is associated with successful oocyte maturation

and fertilization [133-135]. Importantly, the potential benefit of this mutation on oocyte maturation and fertilization and embryo survival was exemplified by the fact that cows with the AA genotype in this study had increased pregnancy rate and required fewer services per conception. It is also important to note, however, that this study was conducted in cows, therefore these effects may not carry over to heifers. Additionally, it is unknown if this SNP is having a causative effect or if it is in linkage disequilibrium with the causative mutation, thus further research is required to determine that this SNP is causative.

Heat shock proteins (Hsp) are a family of proteins that act as molecular chaperones that protect the cell against heat stress and are involved in maintaining protein structure under stress conditions [136]. Among Hsp family members, Hsp70 is one of the most abundant [137] and highly studied in terms of polymorphisms affecting fertility and embryo survival in cattle [138-141]. However, one study found SNPs in the genes DnaJ Heat Shock Protein Family (Hsp40) Member C27 (*DNAJC27)* and *DNAJC15* that were associated with blastocyst and fertilization rate. Hsp40 assists Hsp70 in its protein folding activity [142], and Hsp40 and Hsp70 work together to protect cells from apoptosis [143]. For the SNP (SNP36016, C > G) in *DNAJC27*, oocytes collected from ovaries with the genotype GG had higher fertilization rates compared to the other genotypes. Conversely, embryos that were produced from dams with the genotype GG at SNP85146 (G>A) in *DNAJC15* showed to have higher blastocyst rates compared to embryos from dams with the AA or AG genotype [144].

## 1.7 TRADITIONAL SOLUTIONS TO IMPROVING HEIFER FERTILITY

### 1.7.1 Genetic based strategies

Being able to select cattle with the highest genetic merit is of great economic importance to producers. Expected progeny differences (EPDs) allow for producers to compare the genetic merit of different sires and dams in order to make selection decisions. EPDs are available for a wide variety of traits in cattle, including carcass, growth, and milk traits [145-150]. However, EPDs for fertility in cattle are limited. Currently, heifer pregnancy EPDs are the only EPD available that are a direct measure of heifer fertility. Before the inclusion of heifer pregnancy EPDs into breeding programs, producers would use scrotal circumference EPDs as an indicator trait of age at puberty. However, once heifer pregnancy EPDs were created, scrotal circumference EPDs were no longer as useful as they once were [151, 152]. Having a heifer pregnancy EPD is particularly useful in breeding programs for several reasons. First, the heritability of female reproductive traits in cattle are quite low, typically ranging from <0.1 to 0.3 [55], however, heifer pregnancy is a moderately heritable trait (Figure 1.4). Second, measurement of this trait comes at no extra cost to the producer, as they are already measuring heifer pregnancy rate. Finally, the methods of pregnancy diagnosis are already well defined [153]. However, while the development of a heifer fertility EPD has been beneficial, only having one EPD that is a direct measure of heifer fertility limits the amount of progress that can be made in a generation. Further development of EPDs that are both direct and indirect measurements of heifer fertility could increase the rate of genetic gain and lead to greater improvements over time.

### 1.7.2 Management related strategies

Given the low heritability of most reproductive traits in cattle, improving fertility only using genetic selection is challenging. Therefore, implementing different management strategies, such as estrus synchronization or utilizing antral follicle count and/or reproductive tract score as predictors of fertility, allows producers to increase their probability of reproductive success. Antral follicle count (AFC) is the total number of antral follicles in the ovary at any given time [154], and it is used as a predictor of fertility in cattle. AFC has shown to be highly variable among different animals and highly repeatable within the same animal [155-157]. Due to this high repeatability, it is possible to classify an animal by their AFC after a single ultrasound [158]. Therefore, this method is also cost effective for the producer. AFC is also correlated with several measures of fertility such as calving day and pregnancy rate in heifers. *Bos taurus* heifers that have a higher AFC gave birth earlier in the calving season and had increased fertility compared to their low AFC counterparts [159, 160]. Additionally, AFC is correlated with the number of healthy follicles and oocytes in the ovary. Cattle that have a high AFC have a greater number of healthy oocytes, however, they also have a greater rate of antral follicle growth and atresia. Importantly, while these cattle have higher rates of follicular atresia, it is thought that they will still have a greater reproductive lifespan compared to their low AFC counterparts due to having a sufficient number of healthy oocytes [161].

AFC has also been correlated with differences in hormone concentrations. For example, differences in AFC are associated with progesterone concentration in cattle. *Bos taurus* females that have a low AFC have shown to have low concentrations of progesterone during their estrous cycle [162, 163]. Importantly, low concentrations of progesterone have been linked to high rates of embryonic mortality and less healthy oocytes [164, 165]. This further supports the conclusion that heifers with a low AFC have decreased reproductive potential compared to the high AFC heifers. Moreover, several studies have found that there is an inverse correlation between AFC and follicle stimulating hormone (FSH) concentration [155, 157, 161, 166]. It has been well established that circulating concentrations of FSH are higher in cattle with lower AFC [167, 168], which could lead to the follicular reserve diminishing too quickly over time [155]. Lastly, circulating anti-Mullerian hormone (AMH) concentration has been associated with AFC. Due to the high correlation ($r = 0.88$, *Bos taurus*) between AMH concentration and AFC, AMH can actually serve as a endocrine biomarker of AFC [169]. It should be noted, however, that a single ultrasound is easier and cheaper ($3.5 vs. $9.7, [170] ) than determining AMH concentration to predict reproductive potential [171]. Therefore, while it can be done, producers may be less likely to use AMH as a predictor of reproductive potential.

Another method that has been utilized by producers for many years for replacement heifer selection is reproductive tract scoring (RTS). RTS is a moderately heritable trait ($h^2 = 0.32$) that has been correlated with age at puberty, response to synchronization, and pregnancy rate. Heifers are scored on a scale of 1 to 5, where a score of 1 means the heifer is immature and non-cycling and a score of 5 means the heifer is mature and cycling [48]. Importantly, selecting heifers based on RTS has both long term and short-term benefits for the producer. In the short-term, RTS scores of 1 and 2 have been associated with longer days to calving, longer days to first AI, and deceased pregnancy rates compared to heifers with RTS scores of 4 and 5 [48, 172, 173]. One study by Dickinson et al. showed that heifers with an RTS of 4 and 5 had a 16% and 12% greater pregnancy rate to AI,

respectively, compared to heifers with an RTS of 3 [15]. Additionally, LeFever and Odde found that heifers with an RTS of 4 or more had a 13% greater pregnancy per AI rate compared to heifers with an RTS of 3 or less [174]. Finally, heifers with an RTS of 3-5 have been found to conceive 10 days earlier than their RTS or 1 or 2 counterparts [48]. These results, combined with the moderate heritability of this trait, highlight why reproductive tract scoring has become a popular method for replacement heifer selection.

However, the benefits of utilizing RTS for replacement heifer selection goes beyond that first breeding season, as it also impacts her longevity and performance in subsequent breeding seasons. As mentioned previously, lower RTS scores are associated with increased days to calving. Ensuring that cattle calve at the optimal time (~24 months of age) is of great importance to producers, as heifers that calve earlier are more profitable for producers. When a heifer calves early, she is more likely to remain in the herd and wean heavier calves through the first six parturitions, which increases the profitability of that heifer [175, 176]. A study by Holm and others in 2009 showed that a greater proportion of heifers with an RTS of 4 or 5 (77% vs. 54%) remained in the herd until their second breeding season [177]. However, it is important to note that while the utilization of RTS to improve heifer fertility has been beneficial, it is not without its limitations. For example, it is a very subjective measure and many heifers do not fit perfectly into one of the five scores. Therefore, while the accuracy is good, the repeatability of this measure is less than desired [177, 178].

While the other strategies mentioned in this review aim to assist in selection, estrus synchronization protocols aim to improve pregnancy rate at the end of the breeding season. The goal is to highly synchronize estrus and ovulation in order to improve the timing of insemination relative to onset of estrus [179, 180]. Estrus synchronization is widely utilized in cattle production systems due to its benefits to producers. One of the major benefits of estrus synchronization protocols is their ability to initiate estrus cyclicity in prepubertal heifers [181, 182]. This allows producers to maximize the number of heifers that are pregnant early in the breeding season, therefore also maximizing their potential profitability and productivity. Importantly, these protocols also reduce the amount of time and labor required for estrus detection, which makes it easier for producers to utilize artificial insemination [183, 184]. They do this by shortening the period of time where estrus detection would be required [184]. This is of great importance to producers because one of the major reasons for reduced conception rates in heifers is failure to detect estrus [185]. Therefore, by shortening the time frame where estrus would be occurring, it allows producers to focus their efforts and increase rates of estrus detection.

However, one issue with estrus synchronization is reduced pregnancy rates to fixed-time artificial insemination (FTAI) in heifers. This is thought to be attributed to inadequate synchronization of follicular waves in heifers compared to cows [186]. In order to improve synchronization rates, and subsequent pregnancy rates to FTAI, producers can use presynchronization protocols. Utilization of these protocols have been shown to increase estrous response, synchrony of estrus, and pregnancy rate to FTAI [183, 187, 188]. However, even when presychronization and estrus synchronization are used, some heifers will not conceive to the first insemination service. If a heifer does not conceive to the first service, producers can utilize resynchronization protocols to increase the number of heifers pregnant early in the breeding season [189-191]. While the

application of these different management strategies has helped to improve reproductive efficiency, infertility and sub-fertility still remain a major problem for producers.

## 1.8 NEW TECHNOLOGIES TO UNDERSTAND AND IMPROVE HEIFER FERTILITY

### 1.8.1 Transcriptomics

Measurements such as reproductive tract score, antral follicle count, and pelvic measurements have been used by producers for many years in order to select heifers with greater reproductive potential. However, infertility and sub-fertility are still a major barrier to the profitability and sustainability of the cattle industry. This has prompted researchers to investigate new ways to identify reproductively superior heifers as early and accurately as possible. One strategy that has gained popularity over the last decade is the identification of molecular markers associated with fertility through the use of different omics technologies. Of these omics technologies, transcriptomics has been the most widely utilized to investigate cattle fertility. Transcriptomics is the study of the transcriptome, which is the entire RNA content (coding and non-coding) of a cell, tissue, or organism. The major goal of many of these transcriptomics studies is to identify candidate genes and transcriptional profiles that could be used as molecular markers of cattle fertility.

Embryo mortality is one of the greatest barriers to reproductive efficiency in cattle, and much of the observed embryo loss has been attributed to insufficient uterine receptivity [192, 193]. As a result, the majority of transcriptomic studies investigating heifer fertility have been conducted to investigate transcriptome changes in the endometrium related to uterine receptivity and pregnancy success. Transcriptome differences in several key biological process have been found between fertile, sub-fertile, and infertile heifers as well as between receptive and non-receptive endometrium. For example, genes associated with metabolism [62, 192, 194], extracellular matrix remodeling [96, 194, 195], the immune system [96, 194-196], cell growth and proliferation [62, 96, 192, 194, 195], and signaling and transport [62, 96, 192, 194, 195] have been found to be differentially expressed between heifers of varying fertility potential.

Given the strong connection between the immune system and reproduction [197], there have also been a handful of studies aimed at identifying transcriptomic differences in the peripheral white blood cells (PWBCs) of heifers with varying pregnancy outcomes. The first study to investigate PWBCs as a potential medium for the identification of molecular markers of heifer fertility was conducted by Dickinson et al. in 2018. They identified six genes that were differentially expressed between pregnant and not-pregnant heifers at the time of insemination. Of these genes, they found that five (*ALAS2, CNKSR3, LOC522763, TAC3, AND TFF2*) had increased abundance in the not-pregnant heifers and one (*SAXO2*) that had decreased abundance in the non-pregnant heifers. Overall, the results of this study showed that the transcriptome of PWBCs had the potential to classify heifers based on their fertility potential [198].

The same group did a follow-up study where they investigated if differential co-expression of microRNA (miRNA) and mRNA profiles in the PWBCs of heifers at the time of breeding could be used to distinguish heifers based on pregnancy outcome. Interestingly, they found that the non-pregnant heifers not only had the majority of new mRNA:mRNA connections but they also had

inverted connections. These major changes in co-expression could be a result of changes in regulatory mechanisms that have the potential to explain the fertility differences observed in these heifers. Similarly, in the miRNA:mRNA coexpression analysis, they found that the non-pregnant heifers had significantly more connections than the pregnant heifers. Overall, the results from this study also supported the hypothesis that transcriptome profiles in PWBCs are able to discriminate heifers based on fertility potential, thus, they could potentially be used to identify molecular markers useful for selection [199].

A third study was conducted in 2023 by a different group that aimed to identify transcriptome differences that could be used to discriminate heifers based on fertility potential at the time of weaning. They identified 92 differentially expressed genes between the fertile and sub-fertile heifers. Of those, *GATM, MORN4*, *ANKRD35*, *TFF2*, and *RAMP3* were the top five upregulated and *CLEC4D*, *IGSF6*, *KCNK17*, *SLC13A5*, and *SLC11A1* were the top five downregulated genes in the sub-fertile heifers. Interestingly, when they conducted a gene network analysis, the sub-fertile heifer network showed to have increased connectivity [200], which is similar to what was observed by Moorey et al. in 2020 [199]. The top five differentially expressed genes that gained connectivity in the sub-fertile heifers included *VCAN*, *SLC11A1*, *CAMTA2*, *XDH*, and *GDA*. Conversely, *TFF2*, *OLR1*, *ENSBTAG00000052659*, *ENSBTAG00000039132*, and *ENSBTAG00000051464* gained connectivity in the fertile heifers. Notably, *TFF2* was found to be upregulated in the sub-fertile/ not pregnant heifers in both this study and the study by Dickinson et al. Given that these studies were conducted at different time points (weaning and time of breeding), this overlap between studies provides compelling evidence for *TFF2* as a candidate gene to select heifers based on their fertility potential. However, further validation would be needed before it could be utilized by producers.

### 1.8.2 MicroRNA

MicroRNA are short noncoding RNA that regulate gene expression and are involved in almost all biological processes [201, 202]. The identification and use of miRNA as biomarkers of fertility has gained popularity in recent years for several reasons. First, miRNAs are found in most body fluids, particularly in serum [203], and, ideally, biomarkers should be able to be sampled noninvasively. Thus, biomarkers that can be collected from body fluids, such as serum, are highly desirable [204]. They are also remarkably stable and have high tissue specificity [205-207]. Differences in miRNA expression have also been associated with reproduction in several species, including humans [208-211], mice [212-214], goats [215-217], pigs [218-220], and cattle [221-224]. Lastly, miRNA are high conserved between species, and the function of an miRNA is usually consistent from one species to another [225].

Unsurprisingly, most of the research regarding the use of miRNA as biomarkers of fertility has been conducted in humans. Differences in miRNA expression in humans have been associated with reproductive parameters such as the success of fertilization [226-229] and oocyte quality [228-230]. Some examples of miRNA that have shown to be associated with success of fertilization in humans include: miR-122-5p, miR-1260a, miR-486-5p, miR-132-3p, miR-130b, miR-92a, miR-16-1-3p, miR-1244, miR-206, miR-202-5p, miR-16-5p, miR-222-3p, miR-425-3p, miR-454-5p, miR-382-5p, and miR-127-3p [226-229]. While examples of miRNA associated with oocyte quality in humans include: miR-214, miR-454, miR-888, miR-320, miR-197, miR-663b,

miR-766-3p, miR-132-3p, and miR-16-5p [228-230]. Another parameter in both human and bovine that has shown to be associated with differences in miRNA expression is embryo developmental competence. Surprising, most of the research on miRNA biomarkers of developmental competence has been conducted in bovine, with some of the potential biomarkers being miR-30c, miR-10b, miR-novel-113, miR-novel-44, miR-novel-45, miR-novel-139, bta-miR-100, bta-miR-103, bta-miR-1, bta-miR-502a, bta-miR-140, bta-miR-92a, bta-miR-222, bta-miR-2285, miR-150, miR-145, miR-342, miR-450b, miR-380, and miR-10a [231-233]. However, studies in humans have also identified miRNA that are associated with developmental competence in humans including miR-30c, miR-20a, hsa-miR-372-3p, hsa-miR-373-3p, hsa-miR-451a, hsa-miR-200c-3p, hsa-miR-27a-3p, hsa-miR-26b-5p, hsa-miR-21-5p, hsa-miR-10a-5p, hsa-miR-26a-5p, hsa-miR-27b-3p, and let7b [234-236]. Notably, miR-30c has been associated with decreased developmental competence in both humans and cattle [231, 234], which further demonstrates the highly conserved nature of miRNAs across species. This shared function also provides very promising evidence of miR-30c's potential as a biomarker of developmental competence in both cattle and humans.

Only two studies, however, have been conducted to identify differences in the miRNA profile of heifers with different pregnancy outcomes. The first was conducted by Moorey and others in 2020 [199]. In this study, they investigated the differential coexpression between circulating miRNA and mRNA associated with pregnancy outcome in the peripheral white blood cells (PWBCs) of *Bos taurus* heifers. Interestingly, they found that there was a loss of positive and negative correlations in the not pregnant heifers when compared to their pregnant counterparts. Notably, one of the miRNA identified in this study, miR-130b, has been previously associated with reproduction. Decreased expression of this miRNA has been previously associated with decreased developmental potential in mice and humans [230]. However, while the study by Moorey et al. did not find any associations between differential expression of miRNA and pregnancy outcome, the results of these two studies provides encouraging evidence of miR-130b as a potential biomarker for heifer fertility.

The second study was conducted by Banerjee and others in 2023 [237]. Their study aimed to determine if the miRNA profiles in the PWBCs of heifers at weaning could be used to predict her reproductive outcome. They identified 16 miRNA that were differentially expressed between fertile and sub-fertile heifers, with 15 of the miRNAs being differentially connected between the fertile and sub-fertile network. In agreement with the study conducted by Moorey et al. in 2020, their study showed that there was a loss of miRNA:mRNA connectivity in the sub-fertile group. While there was no overlap between the significant miRNA detected between this study and the study by Moorey et al., Banerjee et al. identified several miRNAs that stand out as potential biomarkers due to their previous association with reproduction. One such example is let-7b, which was found to be downregulated in the sub-fertile heifers in this study. Lower expression of let-7b has also been correlated with decreased blastulation and embryo quality in humans [230, 236] and impaired corpus luteum angiogenesis in mice [238]. While these results point to a negative effect of the decreased expression of let-7b on fertility, more research is required to elucidate the role of let-7b in cattle and its potential as a biomarker for heifer fertility. Additionally, miR-92b was found to be downregulated both in the sub-fertile heifers in this study and in the heifers with low endometrial receptivity in a study conducted by Ponsuksili and others [239]. Of the potential

biomarkers presented in this review, these results provide some of the most encouraging evidence for a biomarker of heifer fertility due to the potential connection between the two results.

### 1.8.3 Proteomics

Proteomics is the large-scale characterization of the entire protein content of cells, tissues, or organisms. The use of proteomics, either by itself or in combination with transcriptomics, is beneficial to the study of biology for several reasons. First, cellular phenotypes are a result of proteins, not genes. Therefore, most of a gene's functional information would be found in the proteome, not the genome or transcriptome. Second, mRNA expression levels cannot be used to inform researchers about protein expression in a cell due to the poor correlation between the two measures [240-242]. Thus, the amount of information that can be obtained by only studying genes is limited [241]. Given the biological relevance of proteomics to the study of biology, it has become a very important tool for biomarker discovery.

While proteomics has been utilized to identify molecular markers for other production traits in cattle [243-246], very few studies have been conducted to identify proteomic differences that could potentially explain the fertility differences observed in heifers. One study by Marrella and Biase compared the protein abundance between fertile and sub-fertile heifers of Angus and Holstein genetic background. They found that the protein Alpha-ketoglutarate-dependent dioxygenase FTO had differential abundance between heifers classified based on their fertility group (fertile vs. sub-fertile), with greater abundance in the fertile heifers [247]. Importantly, variants in alpha-ketoglutarate-dependent dioxygenase FTO were associated with symptoms of metabolic disorders [248]. Therefore, the decreased abundance of FTO in the sub-fertile heifers in this study could indicate that these heifers are experiencing a metabolic imbalance that is contributing to their observed fertility differences.

Another study conducted by Gegenfurtner et al. aimed to identify proteomic differences that could lead to heifers being genetically predisposed to have higher/lower fertility. Of the proteins related to metabolic processes, legumain and folate receptor alpha have increased abundance in the low fertility heifers. Conversely, sorbitol dehydrogenase and glutamine-fructose-6-phosphate aminotransferase 1 were found to have increased abundance in high fertility heifers compared to their low fertility counterparts. Since both of these proteins are involved in glucose metabolism and they were more abundant in the high fertility heifers, it was suggested that the low fertility heifers could have an impaired metabolism, resulting in their lower genetic merit for fertility. This study also identified proteins related to the immune system that were differentially abundant between high and low fertility heifers. Interestingly, a large number of the proteins they identified were more abundant in the low fertility heifers, including lactoferrin, chromogranin A, and tubulointerstitial nephritis antigen-like 1 [249]. Given the importance of the immune system in the establishment of pregnancy, the authors proposed these results could be indicative of a dysregulation of the immune system that could be contributing to the reduced genetic merit of the low fertility heifers. Based on the results of these two studies, it is possible that proteomic differences could be leading to a metabolic imbalance or a dysregulation of the immune system that could be resulting in heifers having lower fertility.

Similarly, a study conducted by Moraes et al. also investigated proteomic differences in heifers that differed based on their fertility [250]. They identified 103 proteins that were differentially abundant in the uterine lumen fluid between pregnant and open heifers. Of the 103 differentially expressed proteins, the top 10 that had increased abundance in the pregnant heifers included pregnancy associated glycoprotein 11, trophoblast kunitz domain protein 1 precursor, mitochondrial acetyl-coenzyme A acetyltransferase 1, dihydrolipoamide dehydrogenase, mitochondrial 3-ketoacetyl-CoA thiolase, glutathione synthase, bovine glutamate dehydrogenase, heat shock 70 kDa protein 9, lamin A/C, and heat shock protein 60 kDa mitochondrial isoform X1. The pathways analysis of the proteins with increased abundance in the pregnant heifers showed that these proteins were involved in several processes that were important for early pregnancy such as amino acid biosynthesis and metabolism. Similar results were observed when they compared the high fertility pregnant heifers to the sub-fertile pregnant heifers. The proteins that had increased abundance in the uterine lumen fluid of the high fertility pregnant heifers were also found to be involved in amino acid biosynthesis and energy metabolism. Conversely, the top 10 proteins that had increased abundance in the uterine lumen fluid of the open heifers included PAS-6 and PAS-7 proteins, guanine nucleotide-binding protein G(q) subunit alpha, guanine nucleotide-binding protein G(i) subunit alpha-2, factor V, retinoic acid receptor responder (tazarotene induced) 1, FAM234A, MYO1B protein, bovine tubulin (1jff), glypican 1, and Ezrin. These results, led them to hypothesize that the observed differences in fertility were the result of improved uterine receptivity in the high fertility heifers [250]. Overall, the results of these three studies point to an important role for proteins involved in metabolism that could be predisposing these heifers to have lower fertility. However, while these results are promising, our knowledge about proteomic differences associated with heifer fertility is incredibly limited. Therefore, more research is required to investigate these differences and identify biomarkers that can be utilized to distinguish between heifers of varying fertility potential.

### 1.8.4 Metabolomics

Metabolomics is the study of the metabolome, which is made up of low molecular weight metabolites, such as lipids, amino acids, and vitamins, that are present in a cell, tissue, or organism. The metabolome is often considered the closest level of biological information to the phenotype, and it is influenced by both genotype and environment [251-253]. While the flow of biological information is not linear and every level of information is interacting, the metabolome is considered downstream of the genome, transcriptome, and proteome. Importantly, metabolites are not only considered the end product of these internal interactions, they are also the end product of interactions occurring outside the cell and outside the organism. Therefore, metabolomics gives researchers a more comprehensive measure of how the interactions between genotype and the environment could be impacting phenotype. It is this property that makes metabolomics such a powerful technology for the identification of biomarkers.

The majority of metabolomics studies in cattle have focused on the follicular fluid metabolome and its relationship with oocyte developmental competence. It is well established that there are cycle specific differences in the metabolome [254, 255] and that the metabolomic composition of follicular fluid is highly important for oocyte development and quality [256-259]. One study by Matoba et al. in 2013 explored how different markers in the follicular fluid and reproductive tissues relate to oocyte developmental competence. They first looked at the fatty acid profile of the

follicular fluid, and they found that less competent oocytes had increased concentrations of palmitic acid and decreased concentrations of linolenic acid. Then, they investigated differences in amino acid profiles and how that related to developmental competence. L-alanine, glycine, and L-glutamine were found to be beneficial for oocyte developmental competence [260]. These results are consistent with other studies that have shown the beneficial effects of alanine [259] as well as the negative effects of increased levels of palmitic acid on developmental competence [256]. Additionally, urea was found to have negative effects on competence both in this study and a study conduction in 2001 by De Wit and others [260, 261].

However, very few studies have been conducted to investigate metabolomic differences associated with heifer pregnancy outcome. A study by Phillips et al. in 2018 aimed to identify differences in plasma metabolite profiles of heifers that became pregnant after AI or remained open at the end of the breeding season. In this study, they identified 15 metabolites that differed between the two groups, seven of which had a greater than 2-fold difference. Interestingly, all 7 metabolites (asparagine, lysine, glutamine, histidine, tryptophan, cysteine, and ornithine) were lower in the interfile heifer group. They also utilized a logistic regression model and the list of differentially expressed metabolites to predict the pregnancy outcome of different samples. Importantly, they found that glutamine and histidine were able to correctly classify 90% of the samples [262].

Another study by Moraes et al. in 2020 aimed to identify protein and metabolite differences in the uterine lumen between fertile, sub-fertile, and infertile heifers. From their analyses, they identified 13 metabolites that differed between pregnant fertile and sub-fertile heifers, however, of the 13, only L-methionine was known. L-methionine was shown to be increased in the uterine lumen of pregnant fertile heifers compared to the sub-fertile heifers (pregnant or open). When comparing the metabolite profiles between pregnant and open heifers, they identified 315 differentially expressed metabolites. Of the 315 differential metabolites, the greatest differences were seen in L-methionine, Hypoxanthine, glutathione disulfide, Isovalerylcarnitine, R-malate, 2-Hydroxyphenylalanine, L-histidine, Methionine sulfoxide, N-alpha-acetyl-L-lysine, Phenylalanine, N-methyl-L-glutamate, 3-Hydroxy-3-methylglutarate, Inosine, 6-Keto prostaglandin G1, N-acetylneuraminate, Fumaric acid, Malate, L-tyrosine, Threonine/homoserine, and glutarate [250]. Notably, tryptophan and glutamine were also found to have increased concentrations in the uterine lumen of the pregnant compared to the open heifers. While this study and the study by Philips et al. [262] used different biological fluids, they both found that histidine, tryptophan, and glutamine had increased concentrations in the pregnant heifers. These results are promising; however, more studies would need to be conducted to validate these metabolites as biomarkers of heifer pregnancy outcome. One other thing to note is that most of the metabolomic studies here utilized biological fluids that have to be collected using invasive methods, which is not ideal for a biomarker. Therefore, identifying biomarkers present in samples that can be collected non-invasively would be largely beneficial for producers.

## 1.9 CONCLUSIONS

The low heritability and highly polygenic nature of female reproductive traits in cattle makes traditional selection an inefficient method of improving reproduction. This has prompted producers to utilize different management strategies and selection methods in order to improve their reproductive efficiency. Over the last two decades, the adoption of these methods by beef and

dairy producers has led to an overall increase in fertility, with most of the progress being attributed to the adoption of artificial insemination and genetic selection. However, despite seeing improvements in fertility in recent years, infertility and sub-fertility still remain a major issue in the cattle industry.

As a result, there has been a large push for the identification of molecular markers for cattle fertility as predictors of reproductive potential. While the number of studies conducted in heifers is limited, the results of these studies have identified several potential biomarkers that could be utilized to discriminate heifers of different fertility potential. However, there is a need for validation studies to confirm the predictive potential of these markers before they could be utilized by producers. Additionally, an ideal biomarker should be able to be sampled non-invasively, but many of the molecular markers identified in previous studies require invasive collection methods. That is not to say that these markers shouldn't be utilized by producers, but it is important to consider that ease of collection can be a major factor in a producer's decision to utilize genomic selection. Therefore, validating these biomarkers in biological fluids, such as serum, that can be more easily collected could increase the number of producers willing to adopt these methods.

Current research shows promise for the use of molecular markers as a tool to assist producers in identifying heifers with superior reproductive potential, but there is still a long way to go before this technology is ready to be made available to producers. As mentioned previously, validation of current results is an essential first step to reaching this goal. Furthermore, the predictive potential of these biomarkers needs to be investigated to confirm which will actually be useful to producers. The recent advancements in sequencing technologies have drastically reduced the cost of sequencing from what it once was, but the development of methods that make it as cheap and easy as possible for producers to utilize this technology will be critical in ensuring its success. While molecular markers provide a strong opportunity to improve fertility, genomic selection alone cannot improve fertility to the degree that is needed, therefore, application of this technology in addition to improved management strategies will be required increase fertility.

## 1.10 FIGURES

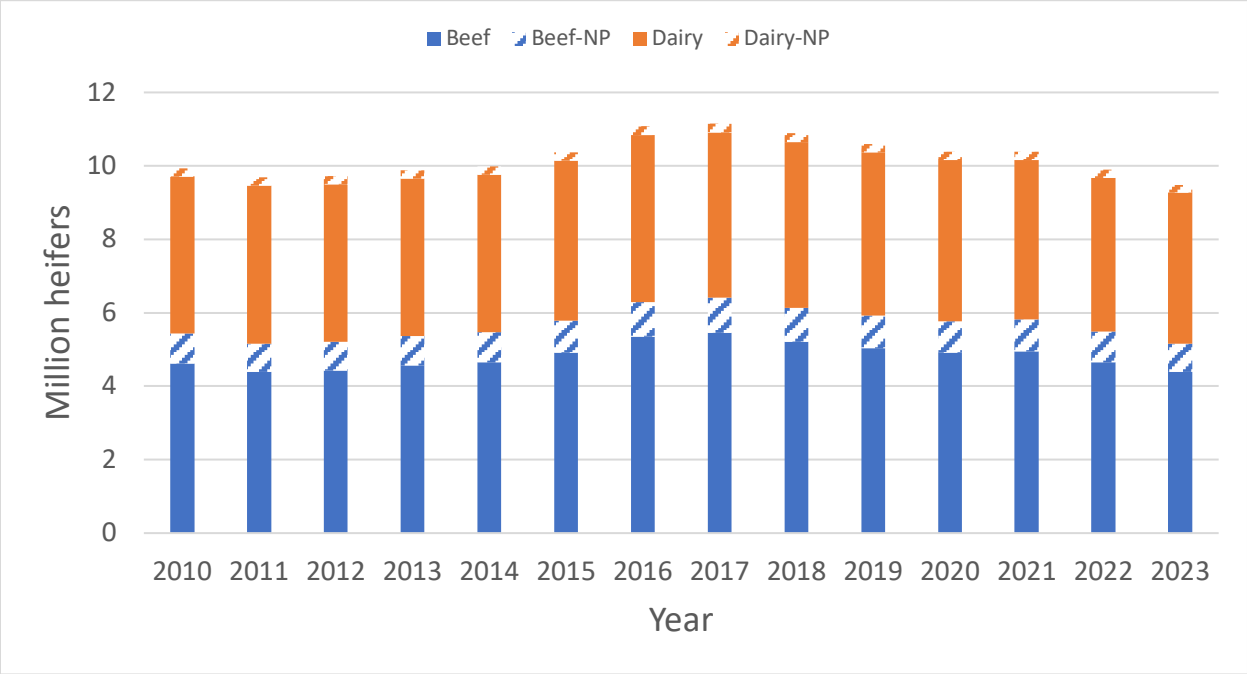


Figure 1.1. Census of heifers over 226.8 kg (500 lb) raised for cow replacement as of January of that year (NASS-USDA). Solid bars indicate the estimated proportion of heifers that become pregnant and diagonal stripes indicate estimated proportion of heifers that do not become pregnant (NP).

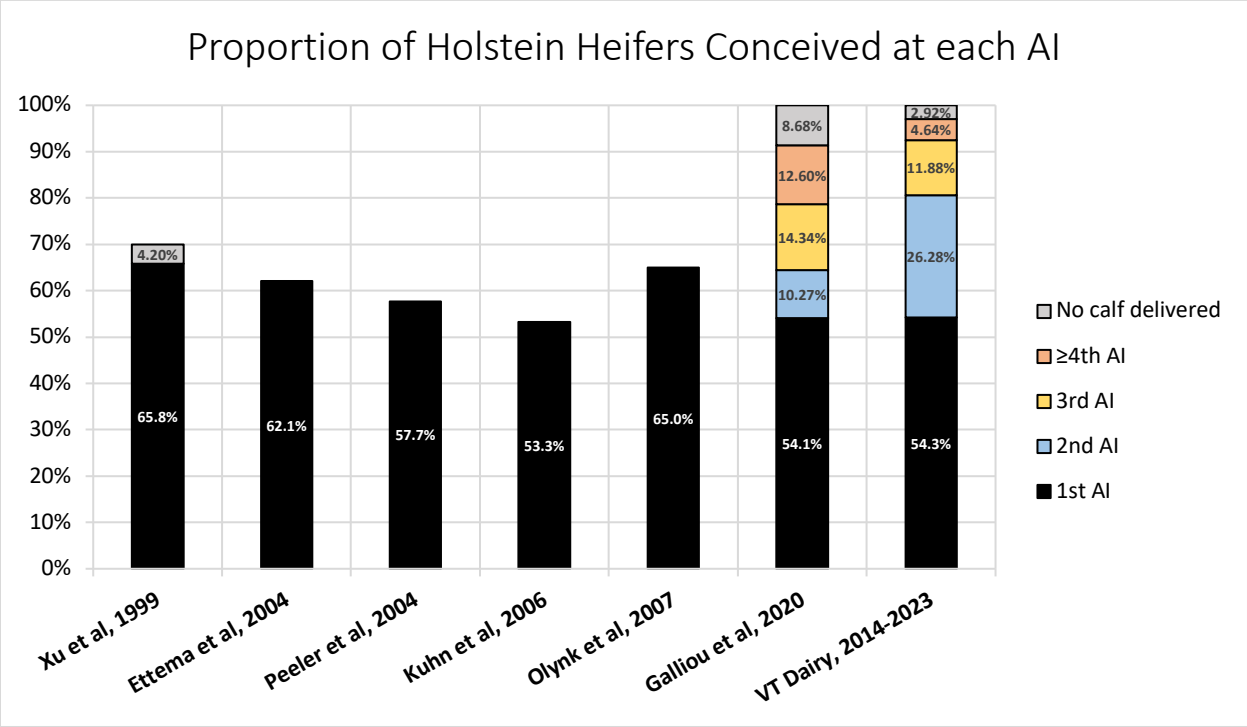


Figure 1.2. Conception rate in dairy heifers (Holstein).

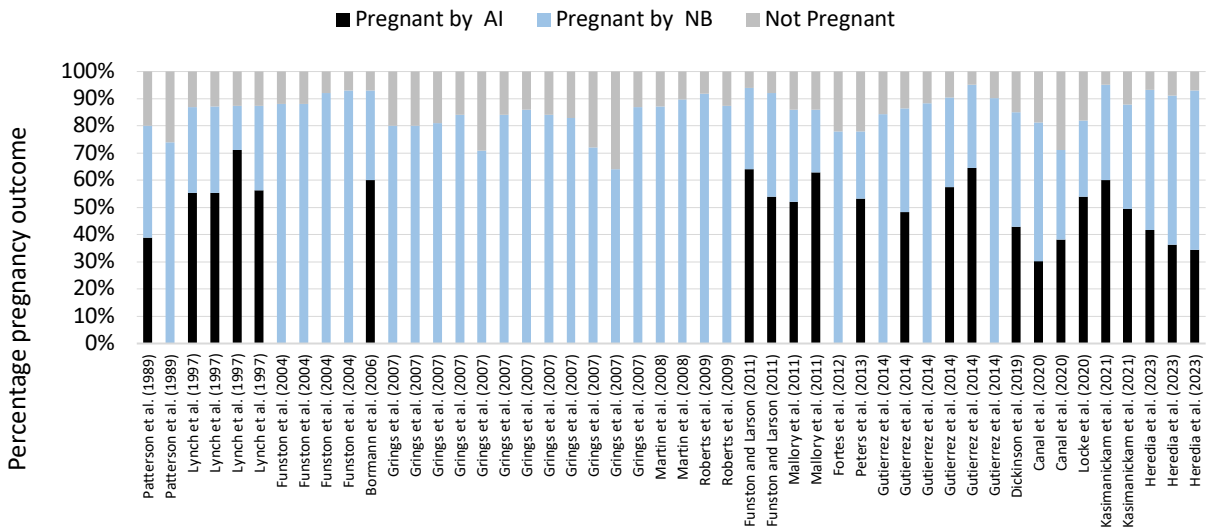


Figure 1.3. Pregnancy rates in beef heifers. Multiple bars from the same reference indicate different experimental treatment within the same study. Modified from reference [58]

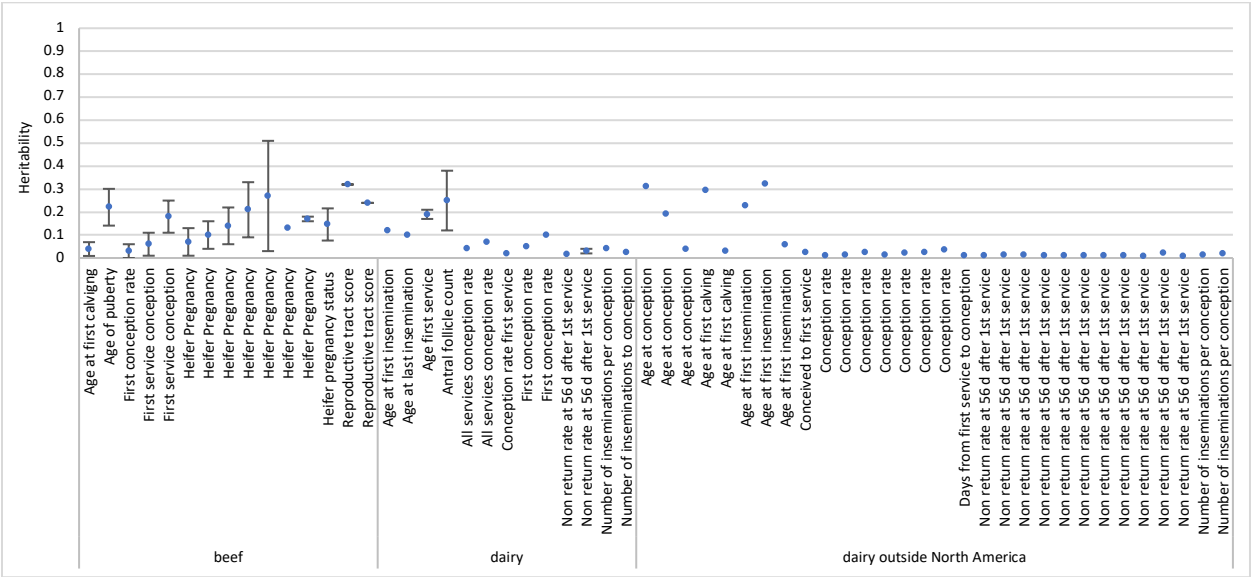


Figure 1.4. Heritability estimates of heifer reproductive traits.

A

| Trait | N SNPs compiled | annotated |
|---|---|---|
| heifer conception rate | 836 | 753 |
| times bred to achieve pregnancy | 292 | 254 |
| heifer fertility | 24 | 18 |
| antral follicle count | 14 | 0 |
| age at first calving | 8 | 8 |
| reproductive tract score | 6 | 0 |
| pregnancy rate | 3 | 2 |
| Total | 1183 | 1035 |

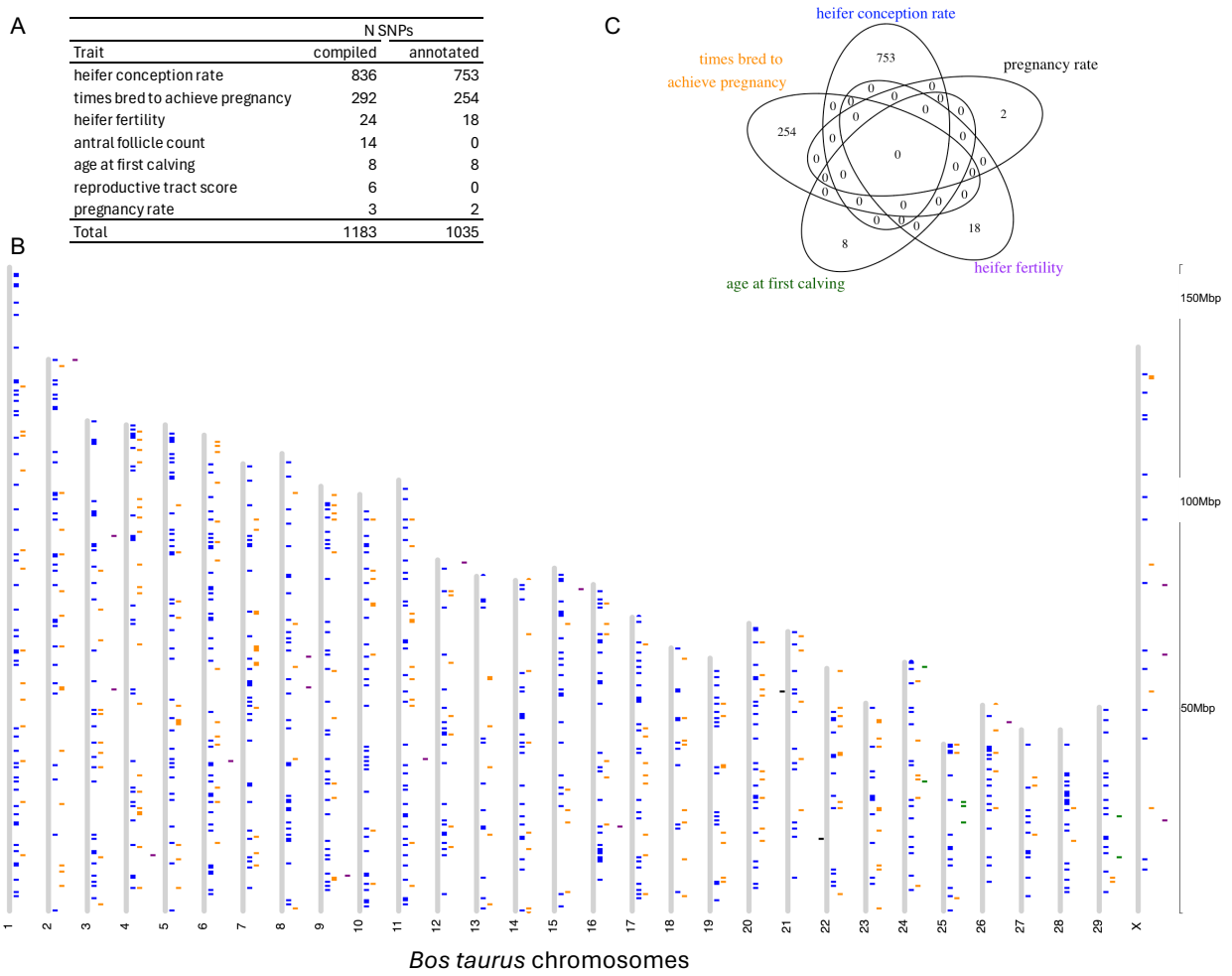


Figure 1.5. Overview of single nucleotide polymorphisms associated with traits related to heifer fertility. (A) Number summary of SNPs associated with different traits. (B) Distribution of significant SNPs across the cattle (*Bos taurus*) genome. (C) Venn diagram of the annotated SNPs associated with traits related to heifer fertility. SNPs were compiled from original reports with no sub setting based on P values.

CHAPTER 2:
Identification of SNPs Associated with Health and Fertility in Dairy Heifers

## 2.1 Abstract

**Background:** Health and infertility represent two of the biggest reasons for culling in dairy production systems. Heifers leaving the herd before breeding and those who do not produce a calf in their first breeding season represent a large source of loss for producers. The objectives of this study were to identify genomic variants that are associated with heifer fertility as well as heifers leaving the herd before 13 months of age due to health reasons. Genotypic records consisting of 78,964 SNPs from 746 heifers from the Virginia Tech Dairy center were obtained from the Council on Dairy Cattle Breeding and production records were obtained from PCDART. Association analyses were conducted in PLINK using the Fisher's exact test as well as dominance and recessive tests. Significance was inferred if $P < 1x10^{-5}$.
**Results:** No significant SNPs were identified under the allelic, dominance, or recessive tests in both the infertility analysis and the analysis comparing heifers pregnant after one AI to those who received five AIs. However, one SNP was found to be significant between heifers pregnant after one AI and those who received four AIs. Then, in the health analysis five SNPs were significant under the allelic test, 10 SNPs were significant under the dominance test, and five SNPs were significant under the recessive test. Finally, in the health + fertility analysis, 17 SNPs were significant under the allelic test, 17 were significant under the dominance test, and eight were significant under the recessive test.
**Conclusions:** Differences in genomic variation were associated with a heifer's fertility potential as well as the likelihood of her leaving the herd before 13 months of age due to health reasons.

## 2.2 Background

The sustainability of dairy farms is critically dependent on having as many healthy cows in lactation as possible, therefore replacing cows that leave the herd due to culling or death represent a significant barrier to dairy cattle production systems. Raising replacement heifers is one of the largest sources of expenses for dairy farms, and the majority of dairy farmers in the US raise their own replacement heifers. Therefore, producers need to implement careful management strategies to avoid severe financial losses that are unable to be recovered.

Infertility, or subfertility, and disease represent some of the largest barriers to the profitability and sustainability of dairy cattle production in the United States [263]. Heifer fertility is of particular interest to producers since heifers that calve at the appropriate time have shown to have increased productivity and longevity in the herd [264-267]. Disease in heifers has also been associated with decreased productivity. Heifers that are affected by disease as a calf have shown to have decreased milk production in their first lactation [268], increased calving intervals [269], and increased age at first calving [270]. Therefore, being able to identify these heifers earlier in their life will allow producers to limit their losses, better allocate their resources, and improve sustainability overall.

Utilizing molecular phenotyping is a promising approach to identify the heifers with decreased production potential earlier [271]. A popular method to identify genetic markers, especially in the early days of biomarker discovery, is to conduct genome-wide association studies (GWAS).

Previous GWAS in dairy cattle have identified biomarkers related to milk production [272-274], reproduction [7, 275, 276], as well as disease [277-279]. However, many of these studies were conducted in cows or were done later in that heifer's life, such as around the time of breeding or calving. By these time points, a producer has already invested considerable resources into raising that heifer, thus limiting the potential benefit of using this technology to inform culling decisions [280]. However, while progress has been made in biomarker identification, more research is still required to identify genetic markers that can be utilized by producers.

Our study was carried out to test the hypothesis that there will be differences in genetic variations between 1) heifers of different fertility potential and 2) heifers that remain in the herd as replacements or leave the herd before 13 months of age due to health reasons. The objectives of this study were to identify variants associated with: 1) heifer fertility in dairy cattle; and 2) dairy heifers leaving the herd before 13 months of age due to health reasons. This study identified variants that could serve as potential biomarkers for health and fertility in dairy heifers.

## 2.3 Results

In our first study, we were interested in identifying significant nucleotide polymorphisms (SNPs) that were related to heifer fertility. To do this, heifers were separated into four groups for analysis: (i) pregnant to the first artificial insemination (AI) (N=271); (ii) artificially inseminated and failed to become pregnant after ≥4 services or were bred and left for reproductive reasons (N=11); (iii) artificially inseminated four times (N= 28); and (iv) artificially inseminated five times (N=8). Before filtering and quality control, data consisted of 78,964 SNPs from 746 Holstein heifers.

In order to identify SNPs that were associated with infertility in heifers, we compared the heifers who were pregnant to the first artificial insemination vs. the heifers who were either artificially inseminated and failed to become pregnant after ≥4 services or were bred but were culled due to reproductive reasons. After quality control and filtering, 73,749 single nucleotide polymorphisms (SNPs) from 282 individuals were retained for the infertility analysis. Under all three tests (allelic, dominance, recessive), there were no SNPs that were significantly associated with fertility ($P < 1x10^{-5}$). Following that analysis, we went on to determine if any SNPs were associated with sub-fertility in heifers. To test this, we ran two separate analyses. The first analysis compared the heifers who got pregnant after one AI to those who were artificially inseminated four times. After quality control and filtering, 73,925 SNPs from 299 heifers were utilized for analysis. No significant SNPs were identified in the allelic or the recessive test ($P < 1x10^{-5}$). However, under the dominance test, one SNP (rs41610326), located in an intergenic region, was significantly associated with fertility in heifers ($P < 1x10^{-5}$) (Table 2.1). Interestingly, this SNP was located in a QTL region that has been previously associated with non-return rate in cattle. The second analysis investigating sub-fertility compared the heifers who got pregnant after one AI to those who were artificially inseminated five times. The data for this analysis consisted of 73,417 SNPs from 279 heifers. However, similar to the infertility analysis, no significant SNPs were identified under the allelic, dominance, or recessive test.

While looking into the information contained in the performance records utilized for the fertility analysis, we noticed that many of the heifers who were culled before breeding were culled due to health reasons. Given that these heifers represent a large source of economic loss for producers,

we wondered if there were SNPs that were associated with these heifers leaving the herd. Using fertility as a proxy for health, we compared the heifers who got pregnant after one AI service (N=271) to those who left the herd before 13 months of age due to health reasons (N=14). After quality control and filtering, 73,673 SNPs from 285 heifers were retained for this analysis. From the allelic association test, we identified five SNPs (rs42250917, rs41566637, rs3423252931, rs110940666, rs133955943) that were significantly associated (Fisher's exact test, $P < 1x10^{-5}$) with a heifer leaving the herd before 13 months of age due to health reasons (Table 2.1). Four out of five of the SNPs were located in intergenic regions, however, one SNP was located in an intron of the gene *PAM*. Additionally, three of these SNPS were located in QTL regions previously associated with clinical mastitis, immunoglobulin G level, and somatic cell score.

Then, under the dominance test, ten SNPs (rs42250917, rs110940666, rs43242536, rs3423111611, rs3423111627, rs29019349, rs110438172, rs41570419, rs135592668, rs43767288) were found to be significantly associated ($P < 1x10^{-5}$ ) with heifer health (Table 2.1). Two SNPs were located in introns of the genes *KPNA1* and *TFDP2*, one SNP was located downstream of *FAM162A*, and the rest were located in intergenic regions. Moreover, several SNPs were located in QTL regions previously associated with health traits including, clinical mastitis, somatic cell score, M. paratuberculosis susceptibility, initial packed red blood cell volume, bovine respiratory disease susceptibility, and bovine tuberculosis susceptibility. Finally, under the recessive test, five SNPs (rs41566637, rs3423252931, rs136701038, rs42794065, and a SNP with no rs at the position 9:16857647) were found to be significantly associated ($P < 1x10^{-5}$) with heifer health (Table 2.1). Two SNPs were located in introns of *PAM* and *FGF18*, respectively, while the rest were located in intergenic regions. Similar to the results from the other two tests, several SNPs were located in QTL regions previously associated with immunoglobulin G level, somatic cell score, infectious bovine keratoconjunctivitis susceptibility, M. paratuberculosis susceptibility, and clinical mastitis.

Interestingly, however, several of the significant SNPs from the health and fertility analyses was also located in QTL regions previously associated with both health and fertility traits. This finding, along with the previously established connections between health and fertility, made us wonder if we could identify SNPs that were associated with both heifer fertility and heifer health. To investigate this, we compared the heifers who were pregnant after one AI service (N= 271) to those who left the herd before 13 months due to health reasons and those who were bred and never got pregnant (N=22), and after quality control and filtering 73,520 SNPs from 293 heifers were utilized for this analysis. From the allelic association analysis, we identified 17 SNPs (rs41566637, rs42250917, rs135592668, rs29023180, rs41618427, rs3423252931, rs132682865, rs111010098, rs41581784, rs110072536, rs41667474, rs135999525, rs41570419, rs42554556, rs42517435, rs109626146, rs110519737) that were significantly associated (Fisher's exact test, $P < 1x10^{-5}$) with health and fertility in heifers (Table 2.1). Seven of these SNPs were located in introns of the genes *TFDP2, PLXDC2, PAM, LIMK2, PLXDC2, EPHB1, and ENSBTAG00000081391*, respectively, while the rest were located in intergenic regions. Moreover, the significant SNPs identified in this analysis were also located in QTL regions previously associated with 23 health and fertility traits, notably, inseminations per conception, first service conception, heat intensity, veterinary treatments, general disease susceptibility, gestation length, infectious bovine keratoconjunctivitis susceptibility, abomasum displacement, and stillbirth. We were also able to identify 17 SNPs (rs42250917, rs135592668, rs29023180, rs41667474, rs135999525, rs41570419, rs42517435, rs109626146, rs42165802, rs132987419, rs43767288, rs110433832, rs3423437717,

rs110513671, rs137774454, rs42262483, rs42263251) that were significantly associated ($P < 1x10^{-5}$) with health and fertility under the dominance test (Table 2.1). Eight SNPs were located in introns of *TFDP2*, *PLXDC2*, *EPHB1*, *ENSBTAG00000081391*, *OPCML*, *NEURL1B*, *METTL24*, and *CPEB4*, respectively, one was located upstream of the gene *SMIM17*, and the rest were located in intergenic regions. Several of the significant SNPs were also located in QTL regions previously associated with 27 health and fertility traits such as, inseminations per conception, bovine spongiform encephalopathy, first service conception, conception rate, heat intensity, general disease susceptibility, stillbirth, gestation length, age at puberty, M. paratuberculosis susceptibility, dystocia, parasite detection rate, fertility treatments, infectious bovine keratoconjunctivitis susceptibility, and stayability. Finally, under the recessive test, we identified five SNPs (rs41566637, rs41618427, rs136701038, rs133955943, rs42794065) that were significantly associated with health and fertility (Table 2.1). rs42794065 was located in an intron of *FGF18*, while the rest were located in intergenic regions. Finally, four out of the five SNPs were located in QTL regions previously associated with immunoglobulin G level, somatic cell score, non-return rate, calving ease, infectious bovine keratoconjunctivitis susceptibility, and M. paratuberculosis susceptibility**.**

## 2.4 Discussion

Health and fertility are two of the largest reasons why cattle leave the herd, making them traits of great economic importance in the cattle industry. Genome-wide association studies have become a very popular approach to identify genomic variants that are associated with economically important traits. In this study, we aimed to identify variants that were associated with health and fertility in Holstein heifers. It should be noted that one major limitation of the current study is the low sample size, which greatly impacts the power of this study. However, even with the low sample size, we were still able to identify several variants that were associated with health and fertility in heifers.

One of the top SNPs from the health analysis is located in an intron of the gene Karyopherin subunit alpha 1 (*KPNA1*), also called integrin α5. KPNA1 is a protein that mediates the transport of other proteins into the nucleus. Importantly, this protein is responsible for the nuclear localization of signal transducer and activator of transcription 1 (STAT1) and STAT3 [281, 282], which are critical for health and reproduction. KPNA1, STAT1, and STAT3 all play an important role in the host immune response. Some viruses target KPNA1 in order to prevent the nuclear translocation of STAT1 and STAT3, therefore modifying the immune response to allow for infection [283-286]. Additionally, STAT1 and STAT3 have been shown to be associated with several respiratory diseases, including Bovine Respiratory Disease and Bovine Tuberculosis [287-289]. This connection to respiratory disease in cattle was particularly interesting as many of the heifers who left the herd before 13 months of age left due to respiratory problems. If there were a mutation in *KPNA1* that inhibited the transport of STAT1 or STAT3 into the nucleus, it could have negative impacts on the immune system, leading to an increased susceptibility or increased severity of the disease that caused these heifers to be culled.

Given the important roles of the immune system in cattle reproduction, we investigated if there were any SNPs that were associated with both health and fertility in heifers. From this analysis, we were able to identify 29 SNPs that were significantly associated with both health and fertility.

Two of the significant SNPs were located in an intron of Plexin domain-containing 2 (*PLXDC2*). *PLXDC2* encodes a transmembrane protein, PLXDC2, that serves as an activating ligand for Adhesion G-protein coupled receptor D1 (ADGRD1). Importantly, a study in mice found that ADGRD1 plays a critical role in embryo transport by influencing oviductal fluid secretions, and that mice lacking functional ADGRD1 were sterile. In order for proper embryo transport to occur, ADGRD1 needs to be activated by PLXDC2 located on the surface of the cumulus cells [290]. Therefore, if a variant in *PLXDC2* were to inhibit the binding of PLXDC2 to ADGRD1, it could potentially impair embryo transport to the uterus. Interestingly, another study done in humans found that *PLXDC2* also has immunoregulatory functions. They showed that loss of *PLXDC2* altered the macrophage phenotype, leading to an increased pro-inflammatory response and downregulation of tissue healing and anti-inflammatory genes [291]. A significant SNP in this gene could possibly be connected to an inflammatory disorder that is causing these heifers to leave the herd before 13 months of age.

Another significant SNP from the health and fertility analysis was located in an intron of the gene LIM domain kinase 2 (*LIMK2).* LIMK2 is a serine/threonine kinase that is involved in actin cytoskeletal organization as well as microtubule organization. Given these functions, LIMK2, along with LIMK1, has also been shown to be important for embryo cleavage and development in mice [292-294] and pigs [295]. One study by Duan and others in 2018 showed that inhibition of LIMK1 and 2 in mouse oocytes led to failure of embryo cleavage and blastocyst development [293]. Another study in pigs showed similar results. They found that embryos with the LIMK1/2 knockdown showed abnormal cell division and blastocyst formation. Interestingly, they also found that the majority of the embryos in the LIMK1/2 knockdown group arrested at the 1-cell stage [295]. Therefore, a significant SNP in *LIMK2* could be leading to impaired embryo development in these heifers, thus contributing to their lower fertility.

Unsurprisingly, several SNPs were significant in both the health and the health and fertility analysis. One of those SNPs was located in the gene Transcription factor DP2 (*TFDP2*). TFDP2 is a cofactor that, through interaction with E2F family members, plays a role in both health and reproduction through its regulatory role in cell cycle progression [296, 297]. Interestingly, one study found that TFDP2 contributes to the proliferation of the Porcine Reproductive and Respiratory syndrome virus (PRRSV), and that TFDP2 knockdown led to a suppression of PRRSV [298]., Additionally, TFDP2 binds with E2F family members to form a E2F/DP heterodimer, and this interaction has been shown to be essential for E2F to become functional [299-301]. Of these E2F family members, TFDP2 has been shown to form heterodimers with E2F1, E2F2, E2F3, and E2F4 [302, 303], though it has been suggested that it preferentially associates with E2F4 [302]. Importantly, E2F4 plays roles in both health and reproduction in a wide variety of species including mice, pigs, humans, and cattle [298, 304-311]. Knockdown of E2F4 in mice has been shown to lead to fetal anemia, neonatal lethality, increased susceptibility to bacterial infection, male and female sterility, fetal growth retardation, chronic rhinitis, abnormal development of airway epithelium, and abnormal RBC production [304, 307]. Another study showed that E2F4 is also important for embryo development in pigs, as knockdown of E2F4 led to decreased blastocyst rate and total cell number [308]. If there was a mutation in *TFDP2* that disrupted its interaction with E2F family members, heifers carrying this mutation could have an impaired immune and/or reproductive function as a result. Therefore, this mutation could be contributing to the phenotypic differences observed between the heifers in this study.

Fibroblast growth factor 18 (FGF18) is a highly conserved pleiotropic growth factor that plays key roles in follicle development, steroidogenesis, and development. Several studies found that FGF18 has an inhibitory role in steroidogenesis in cattle, as FGF18 was able to inhibit the production of both progesterone and estradiol [312, 313]. Given the importance of these two hormones for follicle development, it is not surprising that FGF18 has also been linked to follicular atresia. The presence of FGF18 leads to an increase in granulosa cell apoptosis, which has been associated with follicular atresia in cattle [313-316]. Moreover, direct injection of FGF18 into a growing follicle caused that follicle to regress [315]. FGF18 also plays key roles in lung development, skeletal development, limb development, and neural development [317-322]. Therefore, it is possible that a significant SNP in this gene could be disrupting normal reproductive function in the heifers in this study.

## 2.5 Methods

### Data collection

The Council on Dairy Cattle Breeding (CDCB) provided us with genomic data from the Virginia Tech Dairy Center. This dataset contained 78,964 genotypes from 746 Holstein heifers. We also obtained reproductive and health data from the Virginia Tech Dairy Center using the herd management software PCDART. The records encompass the years 2007-2023, and included heifers. Data were filtered to retain only Holstein heifers that had complete records of phenotype and health-related information.

### Processing of genotypes for analysis

Imputed genotypes needed to be recoded in R (v.4.2.3) [323] as missing (00), homozygous reference (11), heterozygous (12), and homozygous alternate (22) for PLINK analysis. Then, the data were filtered to only include records for animals that had an ID, phenotype, and genotype. After filtering, 283 animals were retained for further analysis. From these data, MAP and PED files were created for analysis in PLINK ( v1.90b6.18) [324]. The PED file was converted to a BED file in order to save time and space. SNPs were then removed if they had a minor allelic frequency < 0.01, a genotypic frequency < 0.05, or deviated from Hardy-Weinberg Equilibrium (HWE) ($P < 0.00001$). Finally, in order to identify population stratification in our study population, we conducted a principal component analysis (PCA) in PLINK. Results were visualized using the package “ggplot2” [325] in R.

#### *Fertility*

We wanted to investigate both infertility and subfertility in our study, and in order to do that we identified groups of heifers who were pregnant after one AI (N=271), were artificially inseminated ≥4 times and failed to become pregnant or were bred but left the herd for reproductive reasons (N=11), were artificially inseminated four times (N=28), or were artificially inseminated five times (N = 8). For the infertility analysis, we compared the heifers who got pregnant after one AI to the heifers who were artificially inseminated ≥4 times. Following quality control, 73,749 SNPs were retained for further analysis. Then, to investigate SNPs associated with subfertility, we ran two

separate analyses. The first analysis compared the heifers who were pregnant after one AI to the heifers who were artificially inseminated four times, and, after quality control, 73,925 SNPs were retained for this analysis. We then ran another analysis comparing heifers who were pregnant after one AI to the heifers who were artificially inseminated five times, and 73,417 SNPs were utilized for this analysis. All reported SNP coordinates are relative to the btau9 assembly converted with the LiftOver tool [326].

### *Health*

While analyzing data during the fertility study, we noticed that heifers were leaving the herd before they were bred due to health and reproductive reasons. This led us to question if there were genomic variants that were associated with these heifers leaving the herd before 13 months of age. Using fertility as a proxy for health, we compared the heifers that got pregnant after first AI (N=271) to heifers that left the herd before 13 months of age due to health reasons (N=22). Data from the culled group were made up of production and genotypic data from cull heifers that were subset to include those who were in the herd for less than 390 days and left due to health reasons. This data was then merged with the data from the control group (i.e. heifers that got pregnant after one AI). The resulting dataset was then subset to include only heifers that had an ID, phenotype, and genotype. Once this was done, we followed the same procedures used for the fertility study. Following filtering, 73,572 SNPs from 293 individuals were retained for further analysis.

### *Health and Fertility*

To test if any SNPs were associated with both health and fertility, we compared the heifers who got pregnant after one AI service (N=271) to those who either did not become pregnant or left the herd due to health reasons before 13 months of age (N=29). Using the subset data from the previous two analysis, production and genotypic data from the three groups of heifers were merged to create one data set. We then followed the same procedures used in the previous two analyses. Finally, after quality control and filtering, 73,738 SNPs from 300 heifers were retained for further analysis.

## Statistical analysis

### *Association analysis*

Association analyses were conducted in PLINK using the Fisher's exact test as well as dominance and recessive tests. Associations were considered significant at a level of alpha = $1x10^{-5}$, as previously reported for case control studies [327] and GWAS for reproductive traits done in cows and heifers [7, 328, 329]. Results were visualized using the package "ggmahn" [330] in R.

## 2.6 Tables

| SNP | Variant identifier | Position | P-value | | | Gene |
|---|---|---|---|---|---|---|
| | | | Fisher's exact test | Dominance | Recessive | |
| *pregnant to the 1st AI vs those who were artificially inseminated four times* | | | | | | |
| BTA-86812-no-rs | rs41610326 | 12:23151956 | ns | ns | $8.673x10^{-6}$ | - |
| *pregnant to the 1st AI vs those left the herd before 13 months* | | | | | | |
| BovineHD0900005670 | rs42250917 | 9:20753589 | $8.949x10^{-8}$ | $5.169x10^{-8}$ | ns | - |
| Hapmap47842-BTA-115522 | rs41566637 | 15:36724876 | $1.362x10^{-6}$ | ns | $5.421x10^{-7}$ | - |
| BovineHD0700030481 | rs3423252931 | 7:101956846 | $1.552x10^{-6}$ | ns | $3.243x10^{-6}$ | *PAM*[abi] |
| BovineHD2800009326 | rs110940666 | 28:34273514 | $5.998x10^{-6}$ | $5.914x10^{-6}$ | ns | - |
| BovineHD2800008462 | rs133955943 | 28:31910046 | $7.571x10^{-6}$ | ns | ns | - |
| BTB-00028704 | rs43242536 | 1:66890930 | ns | $5.523x10^{-8}$ | ns | *KPNA1*[abi] |
| ARS-BFGL-NGS-91410 | rs3423111611 | 1:61817159 | ns | $9.189x10^{-7}$ | ns | - |
| BovineHD0100017684 | rs3423111627 | 1:61774550 | ns | $1.464x10^{-6}$ | ns | - |
| BTA-05186-no-rs | rs29019349 | 1:66862192 | ns | $2.377x10^{-6}$ | ns | *FAM162A*[abd] |
| ARS-BFGL-NGS-101934 | rs110438172 | 1:61629350 | ns | $2.630x10^{-6}$ | ns | - |
| BTA-86792-no-rs | rs41570419 | 9:20808486 | ns | $5.624x10^{-6}$ | ns | - |
| BovineHD0100036130 | rs135592668 | 1:126950096 | ns | $6.649x10^{-6}$ | ns | *TFDP2*[abi] |
| BovineHD0100018165 | rs43767288 | 1:63716985 | ns | $6.993x10^{-6}$ | ns | - |
| BovineHD0700030639 | rs136701038 | 7:102430585 | ns | ns | $7.963x10^{-7}$ | - |
| BovineHD2000001052 | rs42794065 | 20:3241584 | ns | ns | $4.600x10^{-6}$ | *FGF18*[abi] |
| BovineHD0900004616 | - | 9:16857647 | ns | ns | $8.664x10^{-6}$ | - |
| *pregnant to the 1st AI vs those left the herd before 13 months or did not become pregnant* | | | | | | |
| Hapmap47842-BTA-115522 | rs41566637 | 15:36724876 | $2.810x10^{-7}$ | ns | $1.364x10^{-7}$ | - |
| BovineHD0900005670 | rs42250917 | 9:20753589 | $1.203x10^{-6}$ | $2.397x10^{-6}$ | ns | - |
| BovineHD0100036130 | rs135592668 | 1:126950096 | $1.241x10^{-6}$ | $6.353x10^{-7}$ | ns | *TFDP2*[abi] |
| Hapmap56093-rs29023180 | rs29023180 | 13:21473454 | $2.259x10^{-6}$ | $5.974x10^{-6}$ | ns | *PLXDC2*[abi] |
| Hapmap48197-BTA-117083 | rs41618427 | 12:15108440 | $2.580x10^{-6}$ | ns | $2.857x10^{-7}$ | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| BovineHD0700030481 | rs3423252931 | 7:101956846 | $2.689x10^{-6}$ | ns | ns | *PAM*[abi] |
| BovineHD1700021006 | rs132682865 | 17:70047720 | $2.855x10^{-6}$ | ns | ns | *LIMK2*[abi] |
| BovineHD1300006435 | rs111010098 | 13:21720773 | $2.873x10^{-6}$ | ns | ns | *PLXDC2*[abi] |
| Hapmap43230-BTA-31437 | rs41581784 | 12:11470110 | $3.039x10^{-6}$ | ns | ns | - |
| BovineHD2500011541 | rs110072536 | 25:40344012 | $3.250x10^{-6}$ | ns | ns | - |
| Hapmap44603-BTA-51543 | rs41667474 | 20:12136194 | $4.280x10^{-6}$ | $5.600x10^{-6}$ | ns | - |
| BovineHD0100038446 | rs135999525 | 1:134179453 | $6.232x10^{-6}$ | $5.344x10^{-6}$ | ns | *EPHB1*[abi] |
| BTA-86792-no-rs | rs41570419 | 9:20808486 | $6.333x10^{-6}$ | $4.455x10^{-6}$ | ns | - |
| BovineHD2200007132 | rs42554556 | 22:24041674 | $6.845x10^{-6}$ | ns | ns | - |
| ARS-BFGL-NGS-57426 | rs42517435 | 29:23276174 | $6.880x10^{-6}$ | $9.561x10^{-6}$ | ns | *ENSBTAG00000081391*[bi] |
| BovineHD1200024526 | rs109626146 | 12:80645474 | $8.835x10^{-6}$ | $5.856x10^{-6}$ | ns | - |
| BovineHD2700000235 | rs110519737 | 27:1908266 | $9.330x10^{-6}$ | ns | ns | - |
| Hapmap38768-BTA-66476 | rs42165802 | 29:34208123 | ns | $2.631x10^{-6}$ | ns | *OPCML*[abi] |
| BovineHD1300010872 | rs132987419 | 13:37335920 | ns | $3.169x10^{-6}$ | ns | - |
| BovineHD0100018165 | rs43767288 | 1:63716985 | ns | $4.270x10^{-6}$ | ns | - |
| ARS-BFGL-NGS-108087 | rs110433832 | 20:4422634 | ns | $7.230x10^{-6}$ | ns | *NEURL1B*[abi] |
| BovineHD1800018570 | rs3423437717 | 18:63845221 | ns | $8.645x10^{-6}$ | ns | *SMIM17*[abu] |
| ARS-BFGL-NGS-28479 | rs110513671 | 20:4516902 | ns | $8.723x10^{-6}$ | ns | - |
| BovineHD0900011194 | rs137774454 | 9:39892919 | ns | $9.347x10^{-6}$ | ns | *METTL24*[abi] |
| BTB-01104181 | rs42262483 | 20:5633232 | ns | $9.886x10^{-6}$ | ns | *CPEB4*[abi] |
| ARS-BFGL-NGS-18943 | rs42263251 | 20:5695030 | ns | $9.886x10^{-6}$ | ns | - |
| BovineHD0700030639 | rs136701038 | 7:102430585 | ns | ns | $2.826x10^{-6}$ | - |
| BovineHD2800008462 | rs133955943 | 28:31910046 | ns | ns | $4.734x10^{-6}$ | - |
| BovineHD2000001052 | rs42794065 | 20:3241584 | ns | ns | $4.837x10^{-6}$ | *FGF18*[abi] |

Table 2.1. Summary of association analysis results. ns = not significant; [a] annotated in the NCBI database; [b] annotated in the Ensemble database; [i] intron; [u] upstream genetic variant; [d] downstream genetic variant; - intergenic region.

CHAPTER 3:
Robust identification of regulatory variants (eQTLs) using a differential expression framework developed for RNA-sequencing

## 3.1 Abstract

**Background:** A gap currently exists between genetic variants and the underlying cell and tissue biology of a trait, and expression quantitative trait loci (eQTL) studies provide important information to help close that gap. However, two concerns that arise with eQTL analyses using RNA-sequencing data are normalization of data across samples and the data not following a normal distribution. Multiple pipelines have been suggested to address this. For instance, the most recent analysis of the human and farm Genotype-Tissue Expression (GTEx) project proposes using trimmed means of M-values (TMM) to normalize the data followed by an inverse normal transformation.

**Results:** In this study, we reasoned that eQTL analysis could be carried out using the same framework used for differential gene expression (DGE), which uses a negative binomial model, a statistical test feasible for count data. Using the GTEx framework, we identified 35 significant eQTLs ($P<5x10^{-8}$) following the ANOVA model and 39 significant eQTLs ($P<5x10^{-8}$) following the additive model. Using a differential gene expression framework, we identified 930 and six significant eQTLs ($P<5x10^{-8}$) following an analytical framework equivalent to the ANOVA and additive model, respectively. When we compared the two approaches, there was no overlap of significant eQTLs between the two frameworks. Because we defined specific contrasts, we identified trans eQTLs that more closely resembled what we expect from genetic variants showing complete dominance between alleles. Yet, these were not identified by the GTEx framework.

**Conclusions:** Our results show that transforming RNA-sequencing data to fit a normal distribution prior to eQTL analysis is not required when the DGE framework is employed. Our proposed approach detected biologically relevant variants that otherwise would not have been identified due to data transformation to fit a normal distribution.

## 3.2 Introduction

A large body of studies has demonstrated that genetic variations have a direct or indirect impact on the development of phenotypic variation [331-335]. Such studies advanced our understanding of the genetic architecture of complex traits. More recently, the integration of large-scale genetic studies with transcriptome data has also identified genetic variants that explain variance in transcript abundance of specific genes (reviewed in [336]). The integration of multiple omics datasets, including genotypes, is an important step toward closing the biological gap that exists between genotypes and phenotypes [337].

Recent publications from the human Genotype-Tissue Expression (GTEx) [338, 339] and the cattle GTEx [340] projects have shed light on the genetic control of gene expression in large mammals. The recent findings indicate that genomic variants have a greater impact on gene expression than

previously anticipated [341]. These studies have provided valuable information which will help close the critical gap between genomic variants and phenotypic variation [342, 343], especially those associated with health in humans and livestock.

Given the importance of identifying expression quantitative trait loci (eQTL) [344] to understand cell or tissue biology, several statistical approaches have emerged to allow the coordinated analysis of genomic variants and transcript abundance (reviewed by Nica and Dermitzakis [344]). While the first eQTL studies used microarray data [345], most of the analyses carried out in recent years use RNA-sequencing data. One emerging concern is the normalization of the data across samples. To that end, several methods have been used for data normalization across samples such as the trimmed mean of M-values (TMM) [346], fragments per kilobase per million reads (FPKM) [347], and transcript per million reads (TPM) [348]. These and other methods have been evaluated, and TMM might have an advantage over other methods [349]. Another concern related to eQTL analysis is that RNA-sequencing data do not follow a normal distribution, however, all statistical approaches currently employed assume that the inputted data will follow a normal distribution. Researchers have addressed this by transforming the data using the variance stabilization [350-352], $Log_2$ transformation [353, 354], or the inverse normal transformation [338, 340, 355, 356].

Because the principle of eQTL analysis is to identify differences in transcript abundance between genotypes [345], we reasoned that the analysis of eQTLs using transcript abundance estimated from RNA-sequencing could be carried out using the same framework used for differential gene expression. A major benefit of using such a framework is that differences in transcript abundance are tested and estimated using a negative binomial model [350, 357, 358], which is suitable for sequence count data [359, 360]. Thus, we hypothesized that biologically meaningful eQTLs would be identified without transforming RNA-sequencing data to fit a normal distribution. Here, our objective was to identify eQTLs in cattle peripheral white blood cells (PWBCs) using RNA-sequencing data and the Bioconductor [361] package "edgeR" [357, 362], which was designed for DGE analysis using the general linear model framework.

## 3.3 Methods

### Data processing for variant detection, and variant filtering

We analyzed RNA-sequencing data from 42 heifers (*Bos taurus*, Angus x Simmental) publicly available in the GEO database: GSE103628 [198, 363] and GSE146041 [199]. First, we trimmed sequencing adapters and retained reads with an average quality score equal to or greater than 30 using Trimmomatic (v. 0.39) [364]. Then, we used Hisat2 (v.2.2.0) [365] to align the pair-end short reads to the cattle genome [366, 367] (Bos_taurus.ARS-UCD1.2.99), obtained from the Ensembl database [368]. Next, we used Samtools (v.1.10) [369] to filter reads that did not map, secondary alignments, alignments from reads that failed platform/vendor quality checks, and were PCR or optical duplicates. Duplicates were removed using the function "bammarkduplicates" from biobambam2 (2.0.95) [370]. The function "SplitNCigarReads" from GATK (v.4.2.2.0) [371] was then used to separate sequences with a CIGAR string, which resulted from sequencing exon-exon boundaries. Variants were then called in our data by using the functions "bcftools mpileup" and "bcftools call" from Samtools [369].

We filtered the variants with the function "bcftools view" from Samtools to select sites where 20 or more reads were used to identify a variant. Next, in R software (4.0.3) [372], we retained variant sites that were identified as single nucleotide polymorphisms and retained variants with genotypes called in at least 20 samples (Fig. 1A).

**Variant annotation**

After the list of significant SNP-gene pairs was generated from the eQTL analysis, attributes were read in from the Ensembl genome database. The attribute list was merged with the output from the eQTL analysis as well as the nucleotide genotypic data for all samples. Ensembl Variant Effect Predictor [373] was used to compare our data to the cattle genome (Bos taurus, ARS-UCD1.2) to identify the functional consequences of the SNPs.

**Quantification of transcript abundance**

For the expression dataset, we obtained the raw read counts from our previous work [199]. First, we eliminated one sample that had less than a million reads mapped to the annotation; second, we calculated counts per million reads (CPM) [357]; third, we retained protein-coding genes that had CPM greater than two in five or more samples. Next, we calculated TPM [374], which was used in all plots with transcript abundance.

**eQTL analysis**

First, we tested whether the samples presented a genetic stratification using plink [324] to calculate the eigenvectors [375]. Given the sample elimination due to low mapping to the annotation, we carried out an eQTL analysis with 41 samples. To prevent overinflation of effects when working with variants with low allelic frequencies [376] and conduct a robust analysis with enough samples in each group of genotypes, we further retained those single nucleotide polymorphisms that had at least five animals in each of the two homozygotes and heterozygote genotypes, had a minor allelic frequency > 0.15, and followed Hardy-Weinberg equilibrium (false discovery rate=0.05), which was tested with the R package "HardyWeinberg" [377]. In both approaches described below, eQTLs that overlapped between the ANOVA and additive model are only reported in the ANOVA model.

*Approach 1: TMM normalized and normal-transformed RNA-seq data*

In line with standard procedures adopted for eQTL analysis [338, 355, 356], we normalized expression abundance for 10,332 genes using the TMM method [346]. First, we used the function "calcNormFactors" from the R package "edgeR" [357, 362] to calculate the normalization factors then we multiplied the normalization factors by the respective library size. Next, we used the function "cpm" with the normalized library size to obtain TMM normalized counts per million. Next, we carried out an inverse normal transformation [338, 355, 356] using the "RankNorm" function from the R package "RNOmni". Additive and ANOVA analyses were carried out independently for eQTL analysis with the R package "MatrixEQTL" [378] using 6,216 SNPs. In both models, we used genotypes as a fixed effect. We inferred a significant eQTL when the nominal *P*-value was less than $5 \times 10^{-8}$, which is a threshold commonly applied to genome-wide

association studies [379-383], and corresponded to a false discovery rate [384] of 4% and 12% for the ANOVA and additive model, respectively.

*Approach 2: Using a differential gene expression framework*

We analyzed the RNA-sequencing data with a general linear model in "edgeR" and tested for differential gene expression using the quasi-likelihood F-test [385, 386]. We note that the normalization adopted by default in "edgeR" adjusts for library sequencing depth, but we added the TMM normalization factors calculated by the function "calcNormFactors" to the procedure for identification of eQTLs.

As part of our proposed approach, we also eliminated genes that had outlier values of transcript abundance, which reduced the transcriptome data to 4149 genes. For these analyses, gene expression data were used as the dependent variable. Genotypes and collection sites were included in the model as independent variables (fixed effects). For additive analysis, the genotypes were input as numerical variables. For ANOVA-like analysis, we carried out a two-tier analysis. First, we tested the association between SNP and gene transcript abundance using all three genotypes as a factor variable. Next, we subset SNPs that were significantly associated with gene transcript abundance and pseudo-coded the genotypes to establish two contrasts [387]. The first contrast compared the homozygote genotype from the reference allele versus the heterozygote and the homozygote genotype from the alternate allele (i.e. AA versus AB, BB). The second contrast compared the homozygote genotype from the alternate allele versus the heterozygote and the homozygote genotype from the reference allele (i.e. AA, AB versus BB). We also inferred a significant eQTL when the nominal *P*-value was less than $5 \times 10^{-8}$ [379-383].

**Visualization of the results**

We used the R packages "ggplot2", "cowplot" [388], or "plotly" [389] for plotting [390] and used Cytoscape [391] to visualize eQTLs in network style.

**Analysis of gene ontology enrichment**

We tested several lists of genes for the enrichment of gene ontology using the R package "GOseq" [392]. In order to account for multiple hypothesis testing, P-values were adjusted by family wise error rate (FWER) [393]. Results were maintained if they had FWER <0.05.

**3.4 Results**

**Overview of SNP identification**

We compiled genotype data at 23,506,613 nucleotide positions. Not surprisingly, 99.6% of the genomic positions were homozygous for the reference allele and 2167 positions were homozygous for the alternate allele. Our pipeline identified 91,006 nucleotide positions showing polymorphisms in our samples. After testing for the deviation of Hardy-Weinberg equilibrium (Fig. 3.1B), we retained 6207 SNPs further analysis (Fig. 3.1C).

Notably, 96% (n=5964) of the SNPs have been previously identified and are recorded in the Ensembl variant database [394, 395], which includes the dbSNP ( [396] version 150), while 243 SNPs were not identified in Ensembl variant database (Additional file 2). Most of the SNPs are in 3 prime UTRs (n=1553), and a smaller proportion (n=483) were annotated as missense variants (Additional file 2). We observed no genetic substructure of the individuals based on the SNPs analyzed here (Additional file 3: Fig. S1).

**eQTL analyses**

For eQTL analysis, we obtained the matrix with raw counts from a previous study [199] from our group. After filtering for lowly expressed genes, we quantified the transcript abundance for 10,332 protein-coding genes. We then analyzed the transcriptome and the SNP data following the two frameworks.

*Approach 1: TMM normalized and normal-transformed RNA-seq data*

The inverse normal transformation within a gene and across samples [356] indeed normalized the RNA-sequencing data (Additional file 3: Fig.3.2). Using the R package "MatrixEQTL" [397], the ANOVA and additive analyses concluded in 4.699 and 2.473 seconds respectively using one core processor (2.60GHz).

We identified 35 significant eQTLs ($P < 5x10^{-8}$) following the ANOVA model (Fig. 3.3). Annotated SNPs mapped to the genes: *ASCC1*, *BOLA-DQB*, *FAF2*, *IARS2*, *MGST2*, *MRPS9*, *NECAP2*, *TRIP11* (Additional file 4). We also identified 39 significant eQTLs ($P < 5x10^{-8}$) following the additive model (Fig. 3.4). Annotated SNPs mapped to the genes *AHNAK*, *GLB1*, *TRIP11* (Additional file 5), and most of the SNPs on the gene *TRIP11* composed the majority of the eQTLs.

*Approach 2: Using a differential gene expression framework*

Using the R package "edgeR" [357], all tests to determine dominance and additive models were completed in 36 and nine minutes respectively using 34 core processors (2.60GHz). We identified 936 significant eQTLs ($P < 5x10^{-8}$). These eQTLs were formed by 16 SNPs present in the dbSNP and one SNP that is a putatively new variant (Additional file 2) influencing the transcript abundance of 445 genes. The majority (98.6%) of the eQTLs were formed by SNPs on the gene TATA-Box binding protein associated factor 15 (*TAF15*), followed by six eQTLs formed by SNPs on the gene SMG6 nonsense-mediated mRNA decay factor (*SMG6*). The other annotated genes with SNPs forming significant eQTLs were *TRIP11*, *PI4KA*, *LMBR1L*, and *ZNF175*. There was no overlap of significant eQTL between both approaches (Additional file 6, , Additional file 3: Fig. S3).

It was also possible to separate the eQTLs into dominance or additive allelic interaction. We determined that six of the eQTLs followed the pattern of an additive allelic relationship (Fig. 3.5A, Additional file 7). Two SNPs (rs41892216 and rs135008768) impacting the expression of the gene sialic acid-binding Ig-like lectin 14 are also present in the region containing the sialic acid-binding Ig-like lectin gene family on chromosome 18. One SNP is a missense mutation (18:57565792, Fig.

3.5B) on the gene *SIGLEC5* and the SNP on nucleotide 18:57498163 is a variant downstream to *SIGLEC6*. Two other SNPs were annotated to the genes *PI4KA* (17:72208968, rs133672368), *TRIP11* (21:56676553, rs479089277) and *ZNF175* (18_57538713, rs109161398).

We also identified 930 significant eQTLs following a dominance allelic relationship (Additional file 8). Eight annotated SNPs mapped to the genes (*LMBR1L*, *SMG6*, *TAF15*, and *TRIP11*)). Of notice, four intronic variants on the gene *TAF15* (19:14551828, 19:14554927, 19:14554403, and19:14553701, Fig. 3.6A) were collectively associated with the expression of 427 genes, with some examples depicted in Fig. 3.6B.

Given the number of genes expressed in PWBCs that were influenced by SNPs, we asked if there would be an enrichment of gene ontology [398] biological processes among these 427 genes. We observed that by setting a more stringent threshold of significance for the eQTLs ($P < 5x10^{-10}$), we subset 196 genes, which are enriched for two biological processes (FWER < 0.05: regulation of catalytic activity (fold-enrichment: 3.54; genes: *APBA3*, *ARHGDIA ARHGEF1*, *CAPN1*, *DENND1C*, *EEF1D*, *EIF2B3*, *RAB3IP*, *RALGDS*, *RING1*, Additional file 9), and endocytic recycling (fold-enrichment: 7.81, genes: *CCDC22*, *DENND1C*, *PTPN23*, *SNX12*).

## 3.5 Discussion

The major goal of our work was to identify genes expressed in PWBCs of crossbred beef heifers whose transcript abundance is impacted by genetic variants. We used a gold standard approach presented by the GTEx consortium, but also analyzed the RNA-sequencing data without a transformation to force a Gaussian distribution of the counts. The framework for eQTL analysis presented here is motivated by the following rationale: (i) the vast majority of eQTL analyses carried out currently use RNA-sequencing data; (ii) by the nature of the procedures, RNA-sequencing data is count data, which is not normally distributed [399, 400]; and (iii) in principle, an eQTL analysis is an expansion of a differential gene expression (DGE) analysis, where samples are grouped by their genotypes, which is analogous to groups or treatments typically used in DGE analysis. Compared to the latest GTEx framework, our analysis of RNA-sequencing data from cattle PWBCs using the DGE framework identified more eQTLs under the dominance model and an equivalent number of eQTLs under the additive model of allele interaction when compared to the framework used in the human or farm GTEx consortia.

Our study has a few limitations, but they do not hinder the validity of our findings. First, we identified SNPs using the RNA-sequencing data, thus we are not accounting for genomic variants in promoters or distal cis-regulatory elements. This is likely to have impacted the limited number of cis-eQTLs reported here. Second, our transcriptome data represents a mixture of white cells identified in the blood. The proportion of different cells that compose the mixture of white cells was not accounted in our model. A genetic factor contributing to a potential greater abundance of one specific cell type [401] is thus a confounding factor in our study. However, these two limitations do not directly impact our main take home message that there is no need for researchers to normalize RNA-sequencing data in eQTL studies.

*Variant genotyping using RNA-sequencing data*

RNA-sequencing data is feasible for the identification of genomic variants in a wide range of organisms, including livestock [402-405], and multiple pipelines have been developed for variant discovery and genotype calling [402-405]. Here we opted for a hybrid approach, which utilized the "SplitNCigarReads" function of GATK [406] followed by the functions "mpileup" and "call" from BCFtools [407]. The reason for using BCFtools was that it calls genotypes at every nucleotide position by default so that individuals were genotyped regardless of the homozygote or heterozygote makeup.

Prior research showed that the efficacy of genotype calling using RNA-sequencing data is high [408]. Although we did not assess the specificity of genotype calling with an orthogonal method, we employed a stringent requirement for coverage equal to or greater than 20x, which is higher than the previously suggested 10x [403, 408] for high confidence genotype calling. In addition, 96% of the variants identified in our pipeline are present in the dbSNP ( [396] version 150), and the variants have the same allelic composition reported in the dbSNP. Our hybrid pipeline efficiently genotyped individuals at homozygote and heterozygote genomic positions, although further confirmation is required for the variants called in our work that are not reported in the dbSNP.

*eQTL analysis using RNA-sequencing with and without forcing the data into a Gaussian distribution*

Current statistical approaches employed for eQTL analysis [409] assume that the data is normally distributed, and the transformation of RNA-sequencing data to enforce a normal distribution is employed in nearly all major eQTL studies. Our comparison of the RNA-sequencing data prior to and after transforming the data ( Additional file 3: Fig. S1) does confirm that the inverse normal transformation [356] is highly effective in reducing skewness and shrinking the variance to reduce the impact of extreme values in the analysis [400], and thus making the data suitable for statistics tests requiring normally distributed data.

We first analyzed our data following the GTEx framework [338], transforming the data to achieve a normal distribution. Our analysis yielded less significant associations between genotype and gene transcript abundance relative to previously published studies that worked with genes expressed in blood samples [410-413] and the recent results from the cattle GTEx consortium [340]. This large difference was expected because we only utilized 6,207 SNPs in our analysis, which yields less genotypic data as compared to high-throughput genotyping platforms or imputation of SNPs from reference populations. Another difference between our procedure and other reports was the stringent threshold to infer significance ($P = 5x10^{-8}$, $-Log_{10}(5x10^{-8}) = 7.3$).

We noted, however, that visual inspection of the data with significant eQTLs identified with the ANOVA model (see examples in Fig. 3.3C) does not clearly indicate patterns of data distribution that resemble the definition of allelic interaction characterized as complete dominance [414, 415]. The dispersion of the data with significant eQTLs identified with the additive model (see examples in Fig. 3.4C) does indicate patterns of data distribution that resemble alleles interacting in additive mode [414, 415]. However, the distribution of heterozygotes showed two groups of samples with district profiles.

The graph profiles obtained from significant eQTLs using the GTEx framework prompted us to analyze the data using a DGE framework. To that end, we carried out an analysis using one of the commonly used statistical algorithms coded in the R package "edgeR" [357, 362, 416]. The comparison of our eQTL analysis using "edgeR" showed a striking contrast with the analysis using the GTEx framework and "MatrixEQTL" in many important aspects. First, there was no overlap of significant eQTLs obtained between the two approaches within this study. Here, we point out that identifying which eQTL is true is virtually impossible without further mechanistic experiments that confirm the influence of allelic variants on gene expression [417, 418]. Our findings add to previous observations that the type of statistical analysis carried out is a critical contributor to the lack of replicability observed across eQTL studies [419, 420]. Second, working with specific contrasts, we were able to identify trans eQTLs that more closely resemble complete dominance, which were not identified by the standard framework. Our results are evidence that the number of genes whose expression are under genetic control and follow patterns of complete dominance [421, 422] is probably more common than previously expected [338]. The identification of groups of genes enriched for specific biological processes strongly supports that this genetic control under the dominance model may have a biological role in the function of PWBCs.

We identified two important aspects that show a contrast between the ANOVA framework and the DEG framework we propose here. First, the functions in "MatrixEQTL" require less computational resources and time to conclude the analysis relative to the calculations carried out using the DGE framework in "edgeR". Our proposed approach is inherently more complex, as we carried out multiple tests to provide robust and valuable information about dominance interaction between alleles. It is also very important to note that our study is not about the tools ("MatrixEQTL" or "edgeR"), because researchers can use other tools for the standard analysis of eQTL such as "FastQTL" [423] or DESeq2 [350] for the DGE framework. Second, the transformation of the data to fit a normal distribution clearly shrinks the variance (Additional file 3, Fig. S4), reducing the differences in transcript abundance among genotypes thus reducing the likelihood of these eQTLs to be inferred as significant. In the end, the most critical choice researchers need to make is between (i) forcing data that is not normally distributed and has many outlier data points [399, 400] into normality or (ii) utilizing a framework that employs a statistical test appropriate for count data.

### 3.6 Conclusions

In summary, different types of data normalization and analytical procedures lead to a variety of combinations that can be used for eQTL analysis using RNA-sequencing. Most of these approaches also transform the data to fit a normal distribution. Our analysis showed that it is possible to carry out eQTL studies using the concepts and analytical framework developed for differential gene expression that does not require data transformation to fit a normal distribution, thus it is likely more suitable for RNA-sequencing. The approach proposed here can uncover genetic control of gene expression that is biologically relevant for the tissue studied that otherwise may not be detected through data transformation and linear models.

### 3.7 Figures

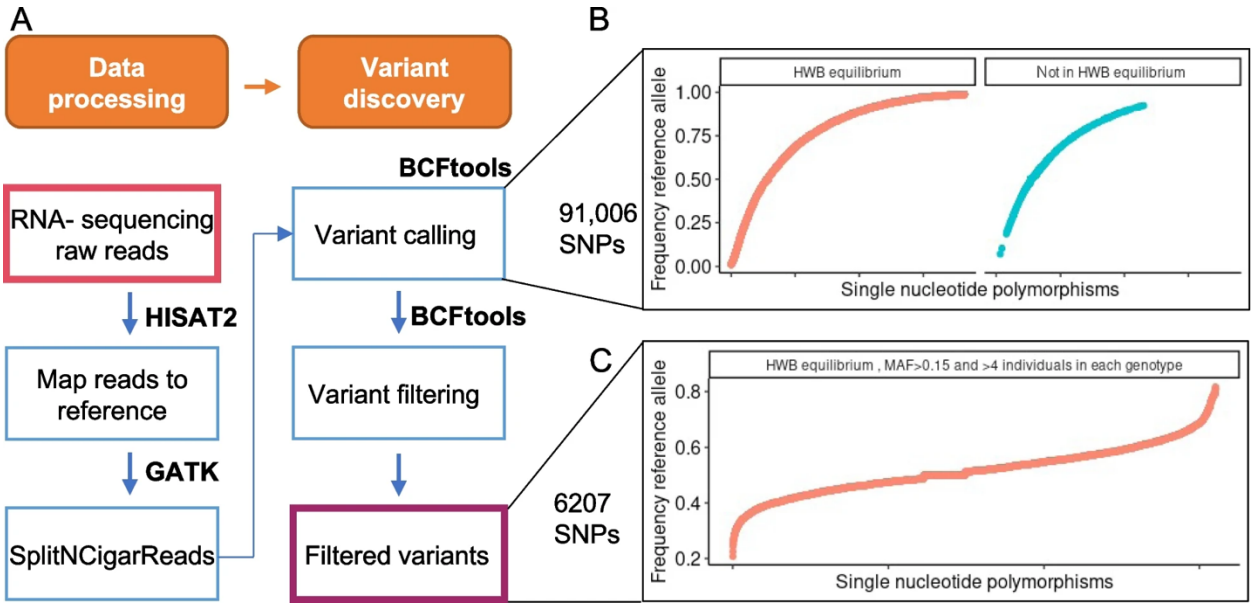


Figure 3.1. Overview of genotyping and variant discovery using RNA-sequencing data from PWBCs. **A** Schematics of bioinformatics procedures. **B** Distribution of allelic frequency of all variants genotyped in at least 26 samples. **C** Distribution of allelic frequency of all variants genotyped in at least 26 samples followed by filtering to retain 6,207 SNPs. (HW: Hardy–Weinberg; MAF: minimum allele frequency).

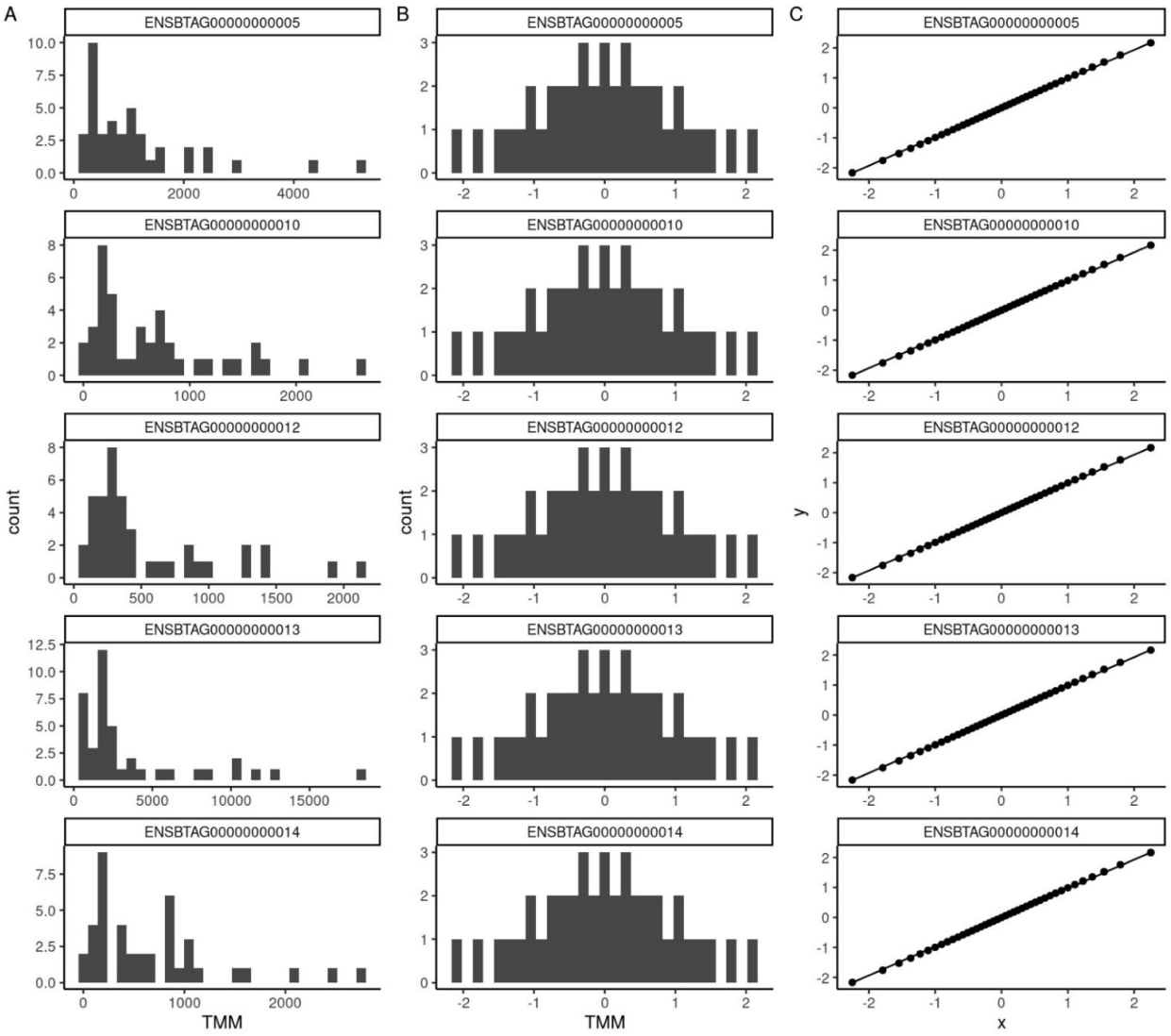


Figure 3.2. Representation of RNA-Sequencing data after (A) normalization of the count data with the TMM method and adjustment per million reads, and normalization as demonstrated by the (B) histogram and (C) qqplot.

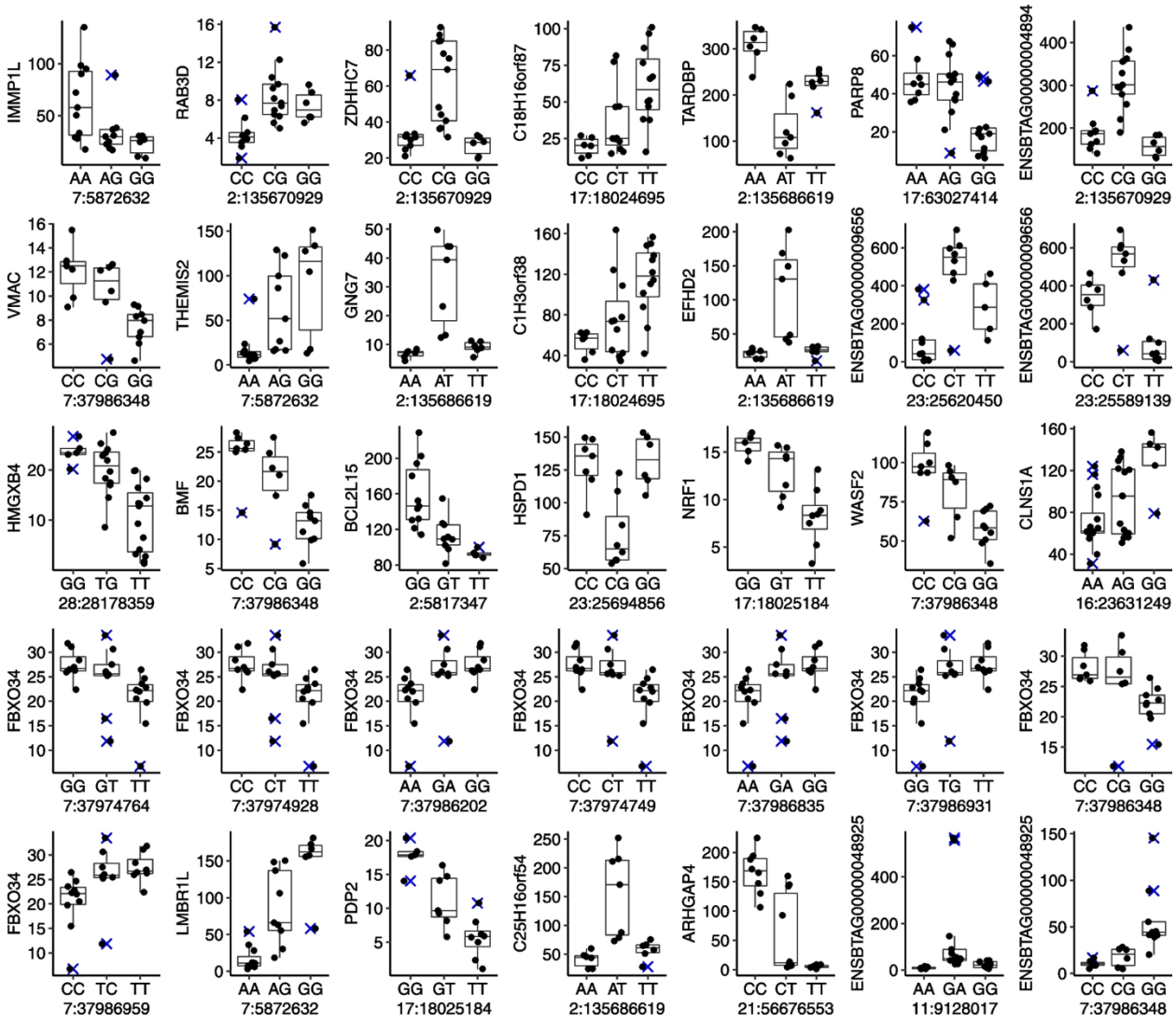


Figure 3.3. eQTLs identified using ANOVA model on TMM normalized counts per million and normal-transformed RNA-seq data. *Y* axis for all graphs is TMM normalized transcripts per million.

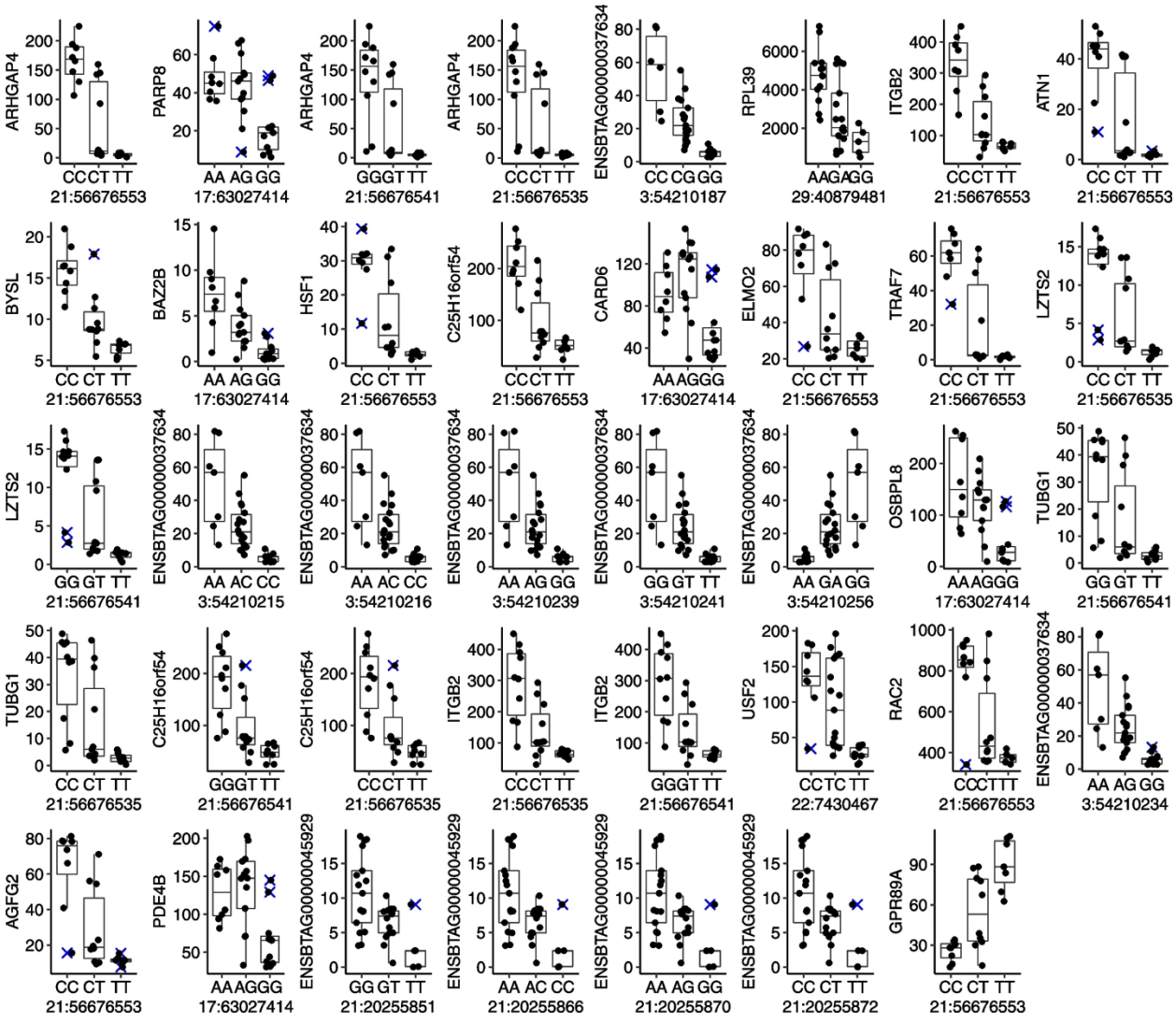


Figure 3.4. eQTLs identified using additive model on TMM normalized counts per million and normal-transformed RNA-seq data. *Y* axis for all graphs is TMM normalized transcripts per million.

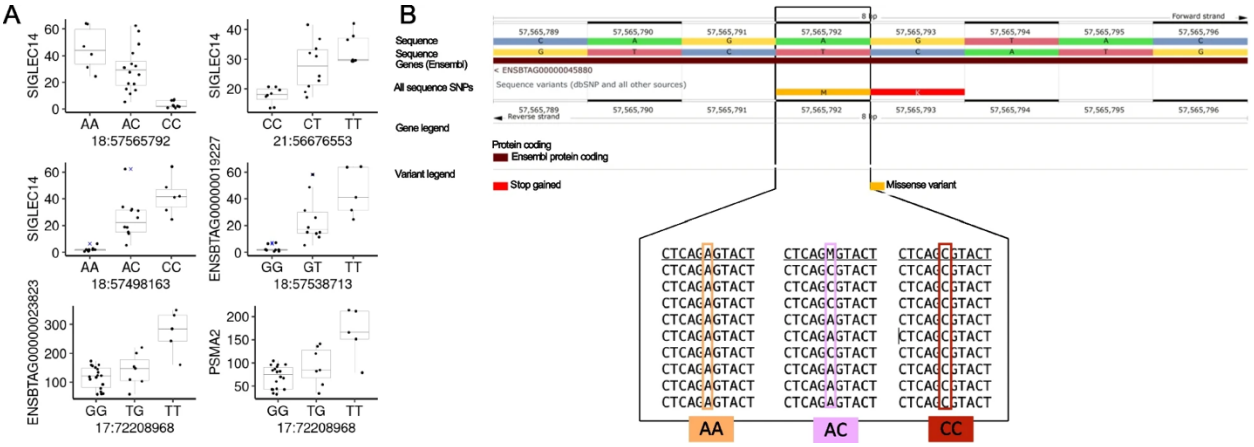


Figure 3.5. Significant eQTLs were identified using the differential gene expression framework. **A** Network depicting the connectivity between SNPs and the genes whose genotypes are influencing their transcript abundance. **B** Bar plot of the frequency of genes containing SNPs forming eQTLs. Only SNPs that were annotated to genes with a symbol (within a gene model, or within 1,000 nucleotides on each side) are depicted in this figure.

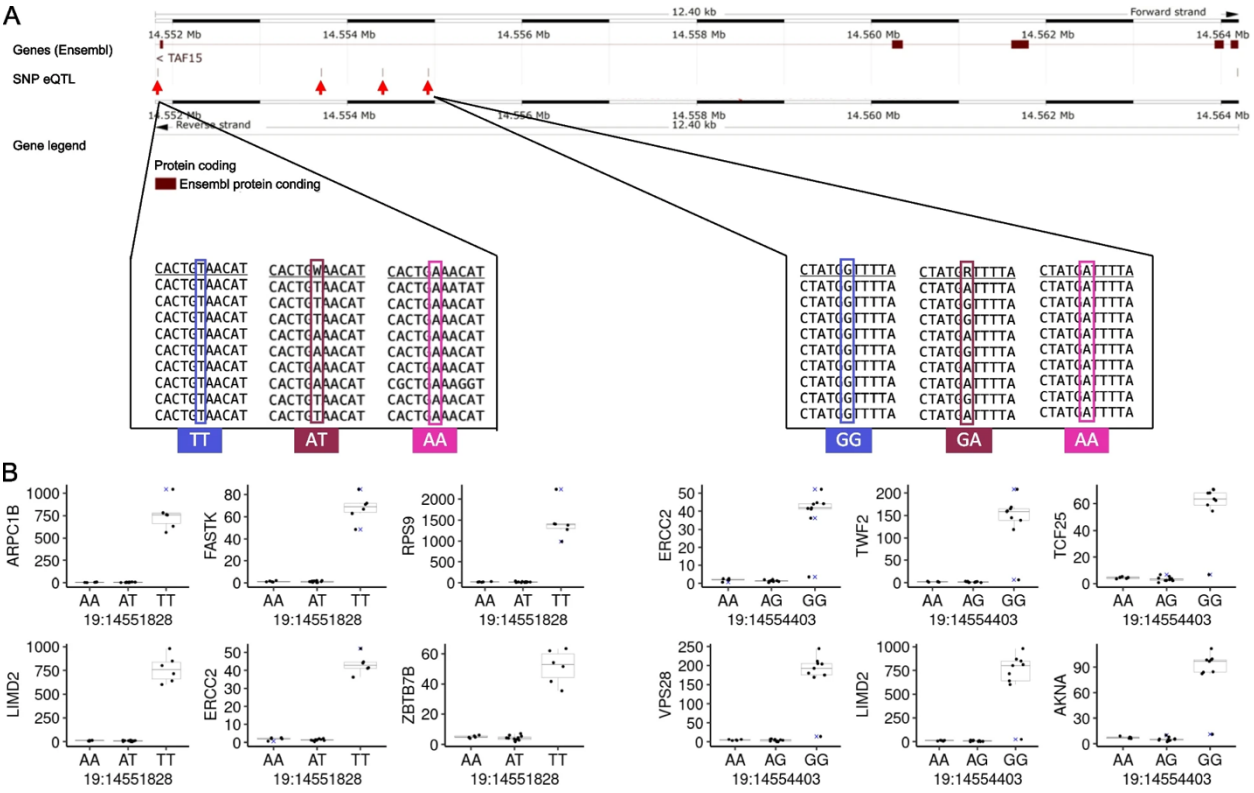


Figure 3.6. Significant eQTLs were inferred using the differential gene expression framework following the additive relationship between alleles. **A** Eight eQTLs following the additive model determined by edgeR. *Y* axis for all graphs is TMM normalized transcripts per million. **B** Ensembl genome browser indicating the SNP position and examples of raw data used for the SNP's identification.

CHAPTER 4:

# A multi-omics analysis identifies molecular features associated with fertility in heifers (*Bos taurus*)

## 4.1 Abstract

**Background***:* Infertility or subfertility is a critical barrier to sustainable cattle production, including in heifers. The development of heifers that do not produce a calf within an optimum window of time is a critical factor for the profitability and sustainability of the cattle industry. In parallel, heifers are an excellent biomedical model for understanding the underlying etiology of infertility because well-nourished heifers can still be infertile, mostly because of inherent physiological and genetic causes.

**Methods:** Using a high-density single nucleotide polymorphism (SNP) chip, we collected genotypic data, which were analyzed using an association analysis in PLINK with Fisher's exact test. We also produced quantitative transcriptome data and proteome data. Transcriptome data were analyzed using the quasi-likelihood test, Wald's test, and the likelihood test and proteome data were analyzed using a generalized mixed model and Student's t-test.

**Results:** We identified two SNPs significantly associated with heifer fertility (rs110918927, chr12: 85648422, $P = 6.7x10^{-7}$; and rs109366560, chr11:37666527, $P = 2.6x10^{-5}$). We identified two genes with differential transcript abundance (eFDR $\leq 0.002$) between the two groups (Fertile and Sub-Fertile): Adipocyte Plasma Membrane Associated Protein (*APMAP*, 1.16 greater abundance in the Fertile group) and Dynein Axonemal Intermediate Chain 7 (*DNAI7*, 1.23 greater abundance in the Sub-Fertile group). Our analysis revealed that the protein Alpha-ketoglutarate-dependent dioxygenase FTO was more abundant in the plasma collected from Fertile heifers relative to their Sub-Fertile counterparts (FDR < 0.05). Lastly, an integrative analysis of the three datasets identified a series of molecular features (SNPs, gene transcripts, and proteins) that discriminated 21 out of 22 heifers correctly based on their fertility category.

**Conclusions:** Our multi-omics analyses confirm the complex nature of female fertility. Very importantly, our results also highlight differences in the molecular profile of heifers associated with fertility that transcend the constraints of breed-specific genetic background.

## 4.2 Introduction

The latest data from the Food and Agriculture Organization show that in 2020 more than 46% of the daily protein supply in the world was from animal-based foods (FAO-STATS). Bovine meat and milk accounted for 12.8% of the total protein supply in the world in 2020 (FAO-STATS). These numbers underscore the importance of cattle production to sustain a growing demand for protein globally [424]. Infertility or subfertility is a critical barrier to sustainable cattle production [263], including in heifers. For example, approximately 15% [58] and 5% [7] of beef and dairy heifers, respectively, do not calve at 24 months of age. Heifers that calve at an optimum age have

greater productivity and longevity in the herd [265, 267, 425-428]. Therefore, identifying heifers with optimum fertility is a promising approach to improving sustainability in cattle production. The heritability of breeding values for heifer fertility is often low for beef [43, 44, 47, 429-432] and dairy [6, 42, 49-51, 433] heifers, which indicate that there are multiple genetic factors impacting this complex trait beyond additive genetic effects. Another potential avenue for the understanding of infertility is the use of molecular phenotyping [271]. The pioneering efforts focused on genome-wide association studies (GWAS) to identify genetic markers associated with heifer fertility [7, 275, 328, 329, 429, 434-444], but only a few seem to be reproducible across populations [443]. More recent efforts have also focused on transcriptome [198, 199, 363] and metabolome [445] datasets characterizing these molecules in blood samples. Again, limited genes have been identified with differential transcript abundance across datasets [199]. Much research is needed for the identification of molecular features that can help explain fertility fitness.

Altogether, approximately 5% of heifers are infertile [7, 13], and this cohort is a great biological model for studying the genetic bases of infertility for several reasons. First, neither dairy nor beef heifers are under the challenging metabolic demand required for milk production [446-448]. Second, post-partum cows need to undergo a critical period of physiological and anatomical recovery before the next breeding [449-451]. Third, there are several postpartum diseases with negative consequences on reproduction success [452-454]. Reproductive problems in well-managed heifers are inherent to their physiology [433], most of which are also under genetic control [455], or directly related to mutations [456] that impair female reproductive functions.

Angus and Holstein heifers have similar frequencies of infertility or subfertility [7, 13] despite the selection pressures directed at beef or dairy production, and thus have distant genetic background. Most studies involving the identification of biological features associated with fertility in heifers have used either one group of purebred or crossbreed animals [7, 198, 199, 275, 328, 329, 363, 429, 434-445]. Here, we carried out a case-control [457] experiment to test the hypothesis that differences in genetic variants, gene transcript, and protein abundance due to fertility fitness would be shared between heifers of different genetic background. Our objective was to contrast genetic variants, gene transcripts, and protein abundance between Fertile and Sub-Fertile heifers from Angus and Holstein genetic backgrounds. We show that both the independent analysis and multi-omics approach identified molecular signatures that capable of discriminating heifers of differing fertility potential, and thus with an underlying biology associated with fertility that is shared between both breeds.

## 4.3 Methods

### Ethics Statement

Animal handling for this experiment was approved by the Institutional Animal Care and Use Committee (IACUC) at Virginia Polytechnic Institute and State University.

### Experimental Design

We collected blood samples from purebred Angus heifers (n=12), averaging 14 months in age, at the time of their first artificial insemination (AI) service. Heifers were subjected to a 7-Day Co-

Synch + CIDR estrus synchronization protocol prior to breeding. Briefly, heifers were administered an intramuscular (IM) injection of gonadotrophin-releasing hormone (GnRH, 100 μg; Factrel®; Zoetis Inc.) on Day 0, followed by the insertion of a controlled internal drug release (CIDR, 1.38 g Progesterone; Eazi-Breed™ CIDR®; Zoetis Inc.). On Day 7, the CIDR was removed and an injection of prostaglandin F2 alpha (PGF2α, 25 μg; Lutalyse®; Zoetis Inc.) was delivered. Fixed-time AI was performed 54±2 hours following CIDR removal alongside a second injection of GnRH.

Additionally, we collected blood samples from purebred Holstein (n=10) heifers, averaging 12 months in age, at the time of the first AI service. Heifers were enrolled in a 5-Day CIDR-Synch protocol before insemination. Briefly, an IM injection of GnRH was delivered on Day 0 with the insertion of a CIDR device. The CIDR device was removed on Day 5, followed by an IM injection of PGF2α. A second injection of PGF2α was administered 24 hours later. Then, timed AI was performed with a second GnRH injection on Day 8.

Heifers were identified as Fertile (Holstein, n=5; Angus, n=5) or Sub-Fertile (Holstein, n=5; Angus, n=7) based on their pregnancy outcome, following similar criteria used previously [198, 199]. Fertile animals were identified as those who became pregnant and subsequently delivered a calf following the first insemination service. Angus heifers were categorized as Sub-Fertile after failing to achieve pregnancy following two insemination services and exposure to a bull for natural breeding. Holstein heifers were identified as Sub-Fertile after needing four or more artificial inseminations.

Heifers were synchronized with protocols that have been identified by prior research to have high success for a heifer to become pregnant to AI [458, 459]. Hence the different protocols for beef and dairy heifers. The criteria for classification were different for each group due to differences in management that are inherent to beef and dairy replacement heifers. Most importantly, each heifer had multiple opportunities to become pregnant before being classified as sub-fertile. The heifers utilized in this study were not part of a nutritional experiment, and thus nutrition was not accounted as a variable nor was it a factor in the selection of heifers. All dairy heifers were raised with equivalent exposure to feed. Similarly, all beef heifers were raised with equivalent exposure to feed.

**Blood Sample Collection and White Blood Cell Isolation**

Fifty ml of blood were drawn from each animal by venipuncture of the jugular vein using 18 mg K2 EDTA vacutainers (Becton, Dickinson, and Company). The tubes were inverted for proper mixing with the anticoagulant and then immediately placed on ice until further processing.

We processed the blood samples following procedures described elsewhere [198, 199, 460] within three hours of sampling [460]. Tubes containing whole blood samples were centrifuged for 25 minutes (min) at 4°C and 2,000*xg* to separate the buffy coat. The buffy coat was then aspirated and mixed with 14 ml of red blood cell lysis buffer (1.55 M ammonium chloride, 0.12 M sodium bicarbonate, 1 mM EDTA (Cold Spring Harbor Protocols)). Then, the solution was centrifuged for 10 min at 4°C at 800x*g* and the supernatant was discarded. The remaining pellet was mixed with 200 μl TRIzol™ Reagent (Invitrogen™, Thermo Fisher Scientific, Waltham, MA) in a 2 ml

cryotube (Corning Inc., Corning, NY) prior to snap-freezing with liquid nitrogen. Samples were then stored at -80°C until further processing.

**Total RNA and DNA Extraction**

The buffy coat samples were thawed at room temperature in a total volume of 525 µl TRIzol™ Reagent. Then, total RNA was extracted from peripheral white blood cells using the Zymo Research Direct-zol™ DNA/RNA Miniprep kit (Zymo Research Corporation, Irvine, CA), according to the manufacturer's protocol. Next, we assessed the quality of the RNA by quantifying the RNA integrity number (RIN) for each sample using the Agilent RNA 6000 Pico kit (Agilent, Santa Clara, CA) on the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA).

**Genotyping and Data Processing**

We submitted 400 ng of DNA for each heifer to Neogen (Neogen Corporation, Lincoln, NE) for genotyping. The samples were genotyped using the Illumina BovineHD Beadchip (Illumina Inc., San Diego, CA) genotyping array (777K). We processed the data for quality control [461] using PLINK [324]. First, we removed SNPs that were preferentially called in one of the groups in the case and control. This was followed by the removal of samples with more than 10% of the genotypes missing, and removal of SNPs with a minor allelic frequency less than 1%, a missing rate greater than 10%, or deviation from the Hardy-Weinberg equilibrium ($P<0.00001$). Next, we carried out variant pruning. We considered a window size of 50 kilobases with five variants in each window at a correlation threshold of 0.2. After pruning, we calculated relatedness and inbreeding coefficients using the parameter '--make-rel' in PLINK (Additional file 2). All reported SNP coordinates are relative to btau9 assembly converted with the LiftOver tool [326].

**Library Preparation and Sequencing**

For sequencing library construction, 900 ng of total RNA was diluted into 25 µl of nuclease-free water, and RNA quantity was confirmed using the Qubit™ RNA High Sensitivity Assay kit (Invitrogen™, Thermo Fisher Scientific, Waltham, MA) on the Qubit™ 4 Fluorometer (Invitrogen™, Thermo Fisher Scientific, Waltham, MA). Libraries were prepared for next-generation sequencing using the Illumina Stranded mRNA Prep kit (Illumina, Inc., San Diego, CA) and the IDT® for Illumina RNA UD indexes (Illumina, Inc., San Diego, CA) according to the manufacturer's instructions. Sequencing was conducted on the NovaSeq 6000 sequencing system (Illumina, Inc., San Diego, CA) using the NovaSeq 6000 SP Reagent kit v1.5 (Illumina, Inc., San Diego,CA) to produce paired-end reads 150 nucleotides in length. Sequencing was performed by the VANTAGE laboratory at Vanderbilt University Medical Center (Nashville, TN).

**Sequence Alignment and Filtering**

We aligned the sequences to the cattle reference genome (Bos_taurus.ARS-UCD1.2.105) in the Ensembl [462] database with hisat2 [365, 463, 464] using the -very-sensitive parameter. Then, we used Samtools [407, 465] to filter sequences and remove secondary alignments, duplicates, and unmapped reads. Next, we used biobambam2 [370] to mark and remove duplicates.

**Transcript Quantification and Gene Filtering**

The number of fragments that matched to the Ensembl [462] cow gene annotation (Bos_taurus.ARS-UCD1.2.105) was quantified using featureCounts [466], and we preserved genes annotated as protein-coding, pseudogenes, or long non-coding RNA. Genes were then retained for further analysis if counts per million (CPM) and fragments per kilobase per million (FPKM) were >1 in at least five samples.

**Proteomics Data and Processing**

One hundred µl of plasma per sample was submitted to the Virginia Tech Mass Spectrometry Incubator (VT-MSI) facility at the Fralin Life Sciences Institute, Virginia Tech, for protein extraction and data collection.

Plasma samples (100 µl) were acidified by the addition of 11.1 µl 12% (v/v) o-phosphoric acid (MilliporeSigma, St. Louis, MO), then proteins were precipitated by the addition of 725 µl LC/MS grade methanol and incubated at -80°C overnight. Precipitated protein was collected by centrifugation and solubilized in S-trap lysis buffer (10% (w/v) SDS in 100 mM triethylammonium bicarbonate ( MilliporeSigma, St. Louis, MO, pH 8.5)). Protein concentration was determined by measuring the absorbance at 280 nm, then 150 µg of protein for each sample was reduced using DTT (4.5 mM) then alkylated with iodoacetamide (10 mM, MilliporeSigma, St. Louis, MO). Unreacted I iodoacetamide was quenched with DTT (10 mM, MilliporeSigma, St. Louis, MO) and samples were acidified using o-phosphoric acid (MilliporeSigma, St. Louis, MO). Protein was again precipitated using methanol and incubated at -80ºC overnight as above. Precipitated protein was loaded onto a micro S-trap and washed with methanol then digested overnight with trypsin. Peptides were recovered and five µg, as determined by measuring the absorbance at 215 nm using a DS-11 FX+ spectrophotometer/fluorometer (DeNovix, Wilmington, DE), of each sample was analyzed twice (duplicates) using ESI-MS/MS Orbitrap Fusion Lumos (Thermo Fisher Scientific (Waltham, MA)).

Samples were first loaded onto a precolumn (Acclaim PepMap 100 (Thermo Scientific, Waltham, MA), 100 µm x 2 cm) after which flow was diverted to an analytical column (50 cm µPAC (PharmaFluidics, Woburn, MA). The UPLC/autosampler utilized was an Easy-nLC 1200 (Thermo Scientific, Waltham, MA). Flow rate was maintained at 150 nl/min and peptides were eluted utilizing a 2 to 45% gradient of solvent B in solvent A over 88 minutes. Spray voltage on the µPAC compatible Easy-Spray emitter (PharmaFluidics, Woburn, MA) was 1300 volts, the ion transfer tube was maintained at 275°C, the RF lens was set to 30% and the default charge state was set to 3.

MS data for the m/z range of 400-1500 was collected using the orbitrap at 120000 resolution in positive profile mode with an AGC target of 4.0e5 and a maximum injection time of 50 ms. Peaks were filtered for MS/MS analysis based on having isotopic peak distribution expected of a peptide with an intensity above 2.0e4 and a charge state of 2-5. Peaks were excluded dynamically for 15 seconds after 1 scan with the MS/MS set to be collected at 45% of a chromatographic peak width with an expected peak width (FWHM) of 15 seconds. MS/MS data starting at m/z of 150 was collected using the orbitrap at 15000 resolution in positive centroid mode with an AGC target of 1.0e5 and a maximum injection time of 200 ms. Activation type was HCD stepped from 27 to 33.

Data were analyzed utilizing Proteome Discoverer 2.5 (Thermo Scientific, Waltham, MA) combining a Sequest HT and Mascot 2.7 (Matrix Science, Boston, MA) search into one result summary for each sample. Both searches utilized the UniProt reference *Bos taurus* proteome database and a common protein contaminant database provided with the Proteome Discoverer (PD) software package [467]. Each search assumed trypsin-specific peptides with the possibility of 2 missed cleavages, a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.1 Da. Sequest HT searches also included the PD software precursor detector node to identify MS/MS spectra containing peaks from more than one precursor. Sequest HT searches included a fixed modification of carbamidomethyl at Cys and the variable modifications of oxidation at Met and loss of Met at the N-terminus of a protein (required for using the INFERYS rescoring node). Peptide matches identified by Sequest HT were subjected to INFERYS rescoring to further optimize the number of peptides identified with high confidence.
Mascot searches included the following dynamic modifications in addition to the fixed modification of Cys alkylated by iodoacetamide (carbamidomethylated): oxidation of Met, acetylation of the protein N-terminus, cyclization of a peptide N-terminal Gln to pyro-Glu, and deamidation of Asn/Gln residues.

Protein identifications were reported at a 1% false discovery rate (high confidence) or at 5% false discovery rate (medium confidence) based on searches of decoy databases utilizing the same parameters as above. The software matched peptide peaks across all runs, and protein quantities are the sum of all peptide intensities associated with the protein.

**Principal component analysis**

We carried out principal component analysis for the genotypes after pruning using the parameter '--pca' in PLINK. The eigenvectors were used for plotting. For the transcriptome data, first we obtained the variant stabilized data using the function 'vst' from the R package 'DESeq2'. Next we calculated the components using the function 'plotPCA' in R. For the protein data, we averaged the values for each technical duplicate and used these values as input for the function 'prcomp' in R.

**Statistical Analyses**

*SNP Association Analysis*

After filtering, 575,053 genotypes from 22 animals were used for association analysis conducted in PLINK [324] using Fisher's exact test. We adjusted the nominal P values to correct for multiple hypothesis testing using the adaptative permutation procedure [468] in PLINK [324]. Locus association was inferred at alpha = $1x10^{-5}$, as reported by The Wellcome Trust Case Control Consortium [327] for case-control studies and previous GWAS analyses of reproductive traits in cows or heifers [7, 328, 329], which corresponded to an adjusted P value <0.005.

*Differential Transcript Abundance*

We compared transcript abundance between samples from each breed and each fertility group. The R packages 'edgeR' [357, 362], with the quasi-likelihood test, and 'DEseq2' [350], using the Wald's and likelihood test, were utilized to conduct the analyses. We adjusted the raw *P* values for multiple hypothesis testing by calculating the empirical false discovery rate (eFDR, [469]) with 10,000 permutations. Differences in transcript abundance were deemed statistically significant when eFDR <0.002 in the results obtained from the three tests.

*Differential Protein Abundance*

To identify differential protein abundance that is robust to the algorithm utilized, we analyzed the protein data using two different algorithms. First, we transformed the protein data using natural logarithm ($Log_e(x)$). We analyzed the transformed data using a generalized mixed model [470] using the R package 'lme4', which included the fertility group (Fertile or Sub-Fertile), breed (Angus or Holstein), and the random effect of the subject. Random effect was included in this analysis as samples were assayed twice to provide a more robust estimate of differential protein abundance. Then, we used the function "emmeans", which tests the significance of the difference ($H_0$:$\mu_1=\mu_2$, $H_1$:$\mu_1\neq\mu_2$) with the Student's t test [471], to calculate the estimated differences in protein abundance between fertility groups within each breed. We also analyzed the log-transformed data using the R package 'limma' [472]. We accounted for the same independent variables mentioned above (fertility group and breed), in addition to accounting for the correlation between the duplicated data for each individual with the function 'duplicateCorrelation'. We tested for a differential abundance of the identified proteins using the empirical Bayes Statistics implemented in the function 'eBayes' [473, 474]. In both analyses, we adjusted the nominal P values using FDR [384]. Significance was assumed if FDR< 0.05 in both approaches.

*Multi-omics Factor Analysis*

We analyzed the multimodal multi-omics datasets (genome, transcriptome, and proteome) interactively using Multi-omics Factor Analysis approach [475, 476]. We subset the genotypes, transcriptome, and proteome data to reduce the global profiling. We retained SNPs with a P value < 0.001 for the Fisher's test, genes with a P value < 0.01 for all three statistical tests employed, and proteins with a P value < 0.05 in both statistical tests used. We conducted the analysis using the R package "MOFA2" [475, 476], accounting for the breed as a group.

## 4.4 Results

### Overview of the data produced

We selected 22 *Bos taurus* heifers of Angus (n=12) and Holstein (n=10) breeds based on their fertility fitness (Fig. 4.1A). We isolated total RNA from circulating white blood cells, averaging 16.3 µg±4.0, and quality, measured by the RIN, averaging 9.4±0.4. The extraction of genomic DNA yielded 1.1 µg±0.4. We produced RNA-sequencing data (Fig. 4.1B) and quantified the transcript abundance of 12,445 genes (12,105 protein-coding genes, 228 long non-coding RNAs, and 112 pseudogenes). We also analyzed 575,053 nucleotide positions across the bovine genome (Fig. 4.1B). Lastly, we produced untargeted proteomics data from plasma that resulted in the relative quantification of 213 proteins. As expected, the genotypic and proteomic data clustered

the heifers of different genetic background separately (Fig. 4.1A). Conversely, there was no clustering of the samples based on the transcriptome data of the peripheral white blood cells (Fig. 4.1D,E).

**GWAS identifies SNPs associated with fertility in Angus and Holsteins heifers**

Our analysis identified two SNPs significantly associated with heifer fertility (rs110918927, chr12: 85648422, $P = 6.7x10^{-7}$; and rs109366560, chr11:37666527, $P = 2.6x10^{-5}$, Fig. 4.2A, Additional file 3). For the SNP rs110918927, all heifers that delivered a calf after one artificial insemination presented the genotype AA ($f_{(A)}=1$, $f_{(G)}=0$), whereas 11 out of 12 heifers classified as sub-fertile presented at least one copy of the allele G ($f_{(A)}=0.29$, $f_{(G)}=0.71$). This polymorphism sits in an intergenic region of the genome with the closest gene located > 73 kilobases downstream relative to the SNP. For the SNP rs109366560, none of the heifers classified as sub-fertile were homozygous for the allele G ($f_{(G)}=0.12$, $f_{(A)}=0.88$), and five out of the nine fertile heifers genotyped were homozygous GG ($f_{(G)}=0.78$, $f_{(A)}=0.22$). This SNP is located on intron 22 of the gene Echinoderm microtubule associated protein like 6 (*EML6*).

**Transcriptome analysis identifies differential transcript abundance between Fertile and Sub-Fertile heifers**

Next, we sought to determine if there were differences in transcript abundance from circulating white blood cells between the Fertile and Sub-Fertile heifer groups, accounting for their genetic background. We identified two genes whose transcript abundance differed (eFDR $\leq$ 0.002) between the two groups (Fertile and Sub-Fertile), namely Adipocyte Plasma Membrane Associated Protein (*APMAP*, 1.16 greater abundance in the Fertile group) and Dynein Axonemal Intermediate Chain 7 (*DNAI7*, 1.23 greater abundance in the Sub-Fertile group) (Fig. 4.3A, Additional file 4).

**Proteomic analysis identifies differential protein abundance between Fertile and Sub-Fertile heifers**

We also tested if there were differential abundance in proteins present in the plasma of heifers classified based on their fertility groups in both genetic backgrounds. The protein Alpha-ketoglutarate-dependent dioxygenase FTO was more abundant in the plasma collected from Fertile heifers relative to their Sub-Fertile counterparts (FDR < 0.05, Fig.4.3B, Additional file 5).

**Integrative multi-omics analysis identifies molecular features that classify heifers based on their fertility potential**

When each data were evaluated independently, the quantification of 22 and 23 gene and protein relative abundances accounted for 44.1% and 16.6% of the variance associated with fertility classification, respectively, and the genotypic information of 59 SNPs explained 70.1% of the variance associated with fertility classification. Overall, there were four factors identified in the analysis with the potential to distinguish the samples based on their fertility status, out of which

three were most representative with Factors one, two, and three being mostly dominated by genotype, transcript, and protein data, respectively (Fig. 4.4A). Factors one, two, and three separated most of the samples based on their fertility classification except two, three, and five samples, respectively (Fig. 4.4B).

Notably, the top nine SNPs that explained most of the variance related to Factor one are located in a window on chromosome 5 spanning from nucleotide 118332762 to 118345383. The tenth SNP was the top significant polymorphism identified on chromosome 12 nucleotide 85648422 according to our Fisher's exact test contrasting heifers of different fertility potential (Fig. 4.4C). Among the genes whose transcript abundance explained the variance related to Factor two, we identified the following annotated genes: ArfGAP with SH3 domain, ankyrin repeat and PH domain 3 (*ASAP3*), ATP synthase membrane subunit c locus 1 (*ATP5MC1*), Centrosomal protein 170 (*CEP170*), Myeloid derived growth factor (*MYDGF*), Coiled-coil domain containing 34 (*CCDC34*), RAD51 associated protein 1 (*RAD51AP1*), and Ubiquinol-cytochrome c reductase complex III subunit VII (*UQCRQ*) (Fig. 4.4C). Among the proteins whose abundance explained the variance related to Factor three, the following were annotated to known genes: Apolipoprotein C-II (*APOC2*), Lymphocyte cytosolic protein 1 (*LCP1*), Vitamin K-dependent protein Z (*PROZ*), Albumin (*ALB*), Serotransferrin-like (LOC525947), Complement component C8 beta chain (*C8B*), Pigment epithelium-derived factor (*SERPINF1*), Phosphatidylinositol-glycan-specific phospholipase D (*GPLD1*), Alpha-ketoglutarate-dependent dioxygenase FTO (*FTO*) (Fig. 4.4C). Collectively, data from the genotypes, transcriptome, and proteome clustered 21 out of 22 heifers correctly based on their fertility status, with only one Fertile heifer clustering with the group of Sub-Fertile heifers (Fig. 4.4D).

## 4.5 Discussion

Reproduction is a multidimensional biological function in mammals that can be partitioned into multiple components or traits [477], and as a consequence, infertility is a complex phenotype with multifactorial origins, including a strong genetic component [455, 456]. Our study was not designed to identify molecular markers for future use in selection programs. Rather, our work addressed two critical questions regarding the underlying biology of infertility: a) whether multiple layers of molecular information, present in the circulatory system, would differ based on female fertility fitness; and b) whether the integrative analysis of multiple layers of molecular information would be a better predictor of the causes of infertility. Our analysis identified molecular signatures in the genome, transcriptome, and proteome that provide important insights about the root causes of infertility.

Neither one of the significant SNPs were located in a region previously associated with female reproductive traits [478]. These SNPs have also not been previously reported to be associated with fertility traits in previous investigations that focused on sire-centric models [7, 275, 440-443], nor on studies that focused on genotyped heifers only [329, 444]. However, it is notable that the polymorphism rs110918927 is in the gene *EML6*, which produces a protein that participates in the function of spindle microtubules in oocytes [479]. Knockdown of this protein in mice oocytes at the germinal vesicle stage impairs spindle morphology and increases aneuploidy [479] in oocytes that progress to the metaphase II stage in the absence of EML6 [480]. The gene *EML6* also produces transcripts in bovine oocytes [481], and the significant SNP in this gene is a strong

indication of a functional connection to reduced oocyte developmental competence in the Sub-Fertile group of heifers.

Genes differentially expressed in the peripheral white blood cells have been associated with fertility in heifers [198, 199, 363]. The protein APMAP exhibits arylesterase activity, which is known to protect lipoproteins from oxidation [482]. Importantly, the APMAP protein regulates adipose composition and metabolic health, and the disruption of the *APMAP* gene in mice leads to an increase in visceral adipose tissue expansion [483]. This protein was also shown to be less abundant in the omental tissue of women diagnosed with polycystic ovary syndrome [484]. Therefore, lower expression of *APMAP* in the peripheral white blood cells of Sub-Fertile heifers is possibly connected with a metabolic, hormonal or inflammatory disorder that disrupts fertility in heifers.

The Protein DNAI7 composes the axonemal dynein complex and participates in beta-tubulin binding activity and microtubule binding activity, and thus contributes to ciliary beating [485]. Variants that impair the function of DNAI7 are associated with Primary Ciliary Dyskinesia, with one potential consequence being the abnormal function of cilia and possible impaired transport of the cleaving embryos into the uterus [486]. DNAI7 may also function as a cell cycle regulator, and dysregulated transcript abundance of DNAI7 was associated with nasopharyngeal neoplasm in mice [487] and lung adenocarcinoma in humans [488]. Since Sub-Fertile heifers have greater abundance of *DNAI7* transcripts in their circulating white blood cells, it is possible that dysregulation in the cell cycle has a biological link with subfertility. Further research is required, however, to evaluate whether a dysregulation in the cell cycle linked to upregulation of *DNAI7* is connected with increased inflammation [483] associated with less transcripts from *APMAP*.

The protein Alpha-ketoglutarate-dependent dioxygenase FTO has oxidative demethylation activity of abundant N6-methyladenosine ($m^6A$) residues in RNA [489]. The protein FTO preferentially demethylates N6,2′-O-dimethyladenosine ($m^6A_m$) rather than $m^6A$ and contributes to a reduced stability of $m^6A_m$ mRNAs [490]. On a systemic level, genomic variants in FTO were associated with symptoms of metabolic disorders [248], although the effects observed in humans, such elevated body mass index [491, 492], and mice [493] may be contradictory. Also worth noting, a variant on the FTO gene was associated with polycystic ovary syndrome [494]. Interestingly, in mice, the FTO gene is downregulated due to a deficiency in essential amino-acids [495], and deficiency in the FTO protein causes postnatal growth retardation and a significant reduction in adipose tissue and lean body mass [496]. Our observation of the FTO abundance in heifers of different fertility potential is an indication that Sub-Fertile heifers could be experiencing a metabolic imbalance, contributing to their lower fertility. We note that the heifers utilized in this experiment were not nutritionally challenged and thus, our observations are a consequence of their intrinsic biological system and how it may utilize nutrients.

The next step was to interrogate the data we produced in a comprehensive manner. Interestingly, the largest source of variability was observed in the genomic data. Nine of the top ten SNPs that were assigned to Factor one were located in an intron of the TAFA chemokine-like family member 5 (*TAFA5*) gene. These SNPs are within a quantitative trait loci for milk yield [497], a trait negatively correlated with reproductive traits [498], however, no relationship between genetic variants in this gene and female fertility has been reported previously. None of the top ten genes

with transcript abundance relevant for the modeling of the variance were identified as differentially expressed when analyzed independently. This result is not surprising because the identification of significant features using standard statistical approaches for association analysis is not necessarily the best approach for identifying predictive genes associated with complex traits [499, 500]. It was surprising that three out of nine annotated proteins, which composed the top ten proteins that explained most of the variance in factor three, were also identified in our analyses using general linear mixed models. The most interesting result, however, was that all three data modalities were able to separate 21 out of 22 heifers correctly based on their fertility potential. Our results show that molecular differences have strong signals linked to fertility fitness that surpasses their differing genetic background.

### 4.6 Conclusions

Our interrogation of multiple levels of biological information (genome, transcriptome, and proteome) at a systemic level in heifers highlighted the molecular complexity of female fertility. While the genomic data pointed to a disruption of oocyte developmental competence, the transcriptome and proteomic data point to metabolic dysregulation contributing to subfertility or infertility. Although the differences in molecular profiles identified in our study need to be further validated by mechanistic studies, our results, supported by the current literature, highlight differences in the molecular profile associated with female fertility that transcend the constraints of breed-specific genetic background.

## 4.7 Figures

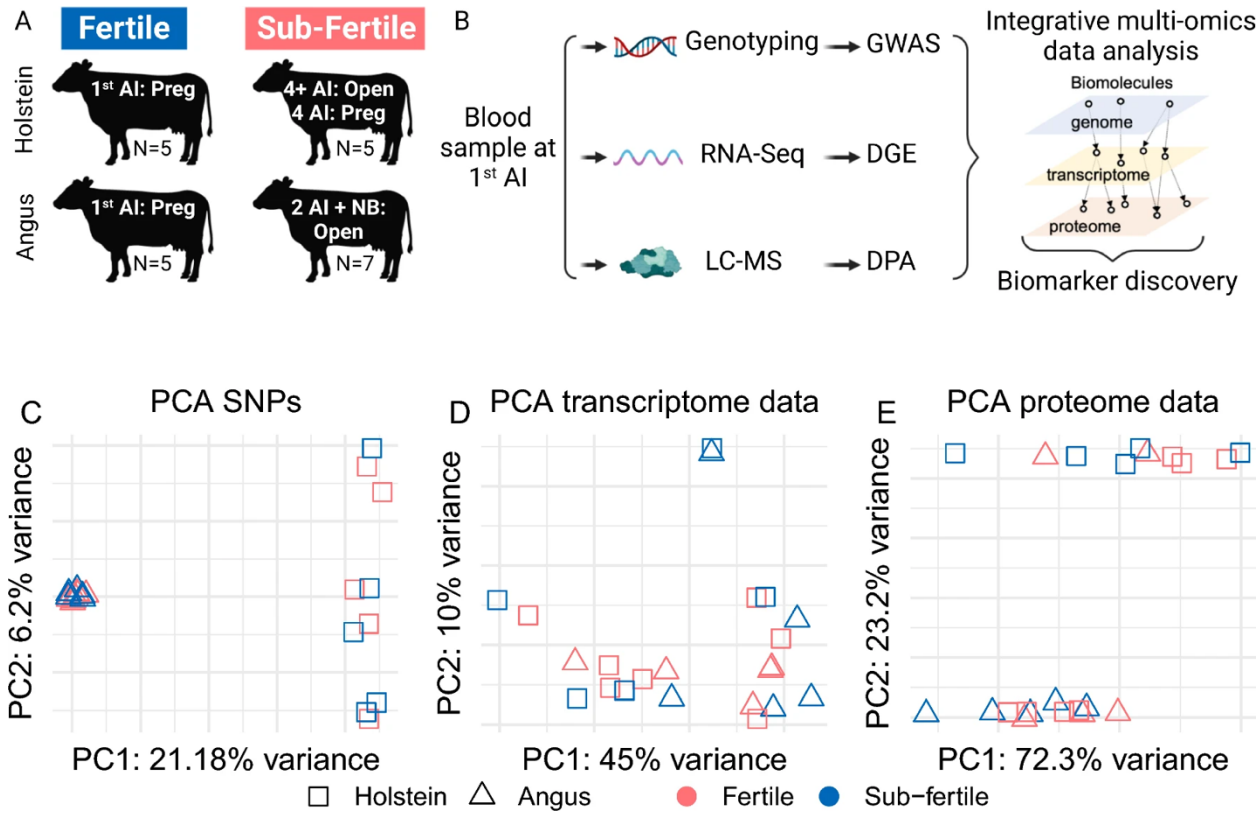


Figure 4.1. Overview of data produced. (**A**) Breeds and classification used in this study, including sample size. (**B**) Schematics of the data produced, and analysis undertaken. Principal component analysis of the genome-wide single nucleotide polymorphisms (**C**), transcriptome (**D**) and (**E**) proteome data. GWAS: genome-wide association analysis, DGE: differential gene expression, DPA: differential protein abundance.

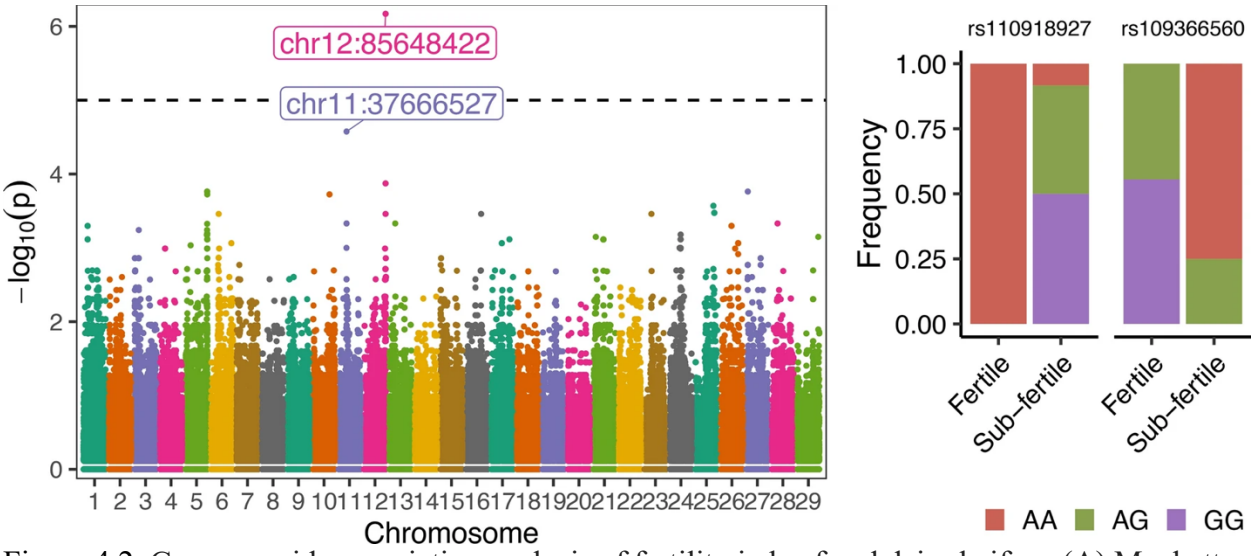


Figure 4.2. Genome-wide association analysis of fertility in beef and dairy heifers. (**A**) Manhattan plot with the distribution of SNPs across their genome and their *P* values from Fisher's exact association test. (**B**) Genetic frequencies of the two SNPs that are putatively linked to fertility in heifers.

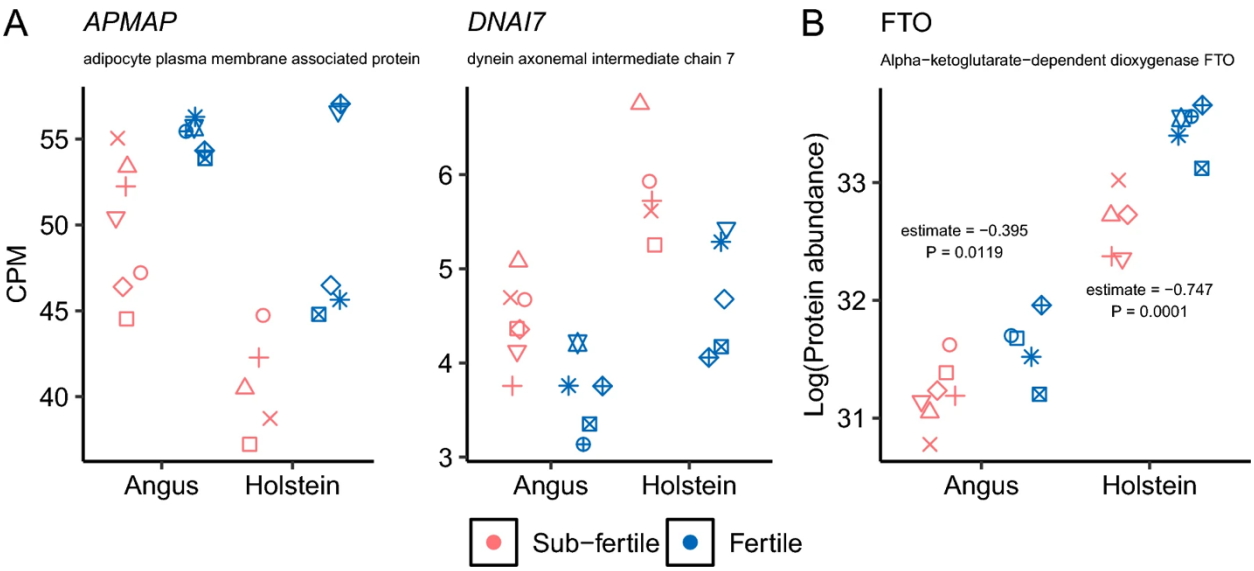


Figure 4.3. Differential transcript and protein abundance associated with fertility. (**A**) Transcript abundance. (**B**) Protein abundance. In each plot, within each breed, shapes indicate the same animal.

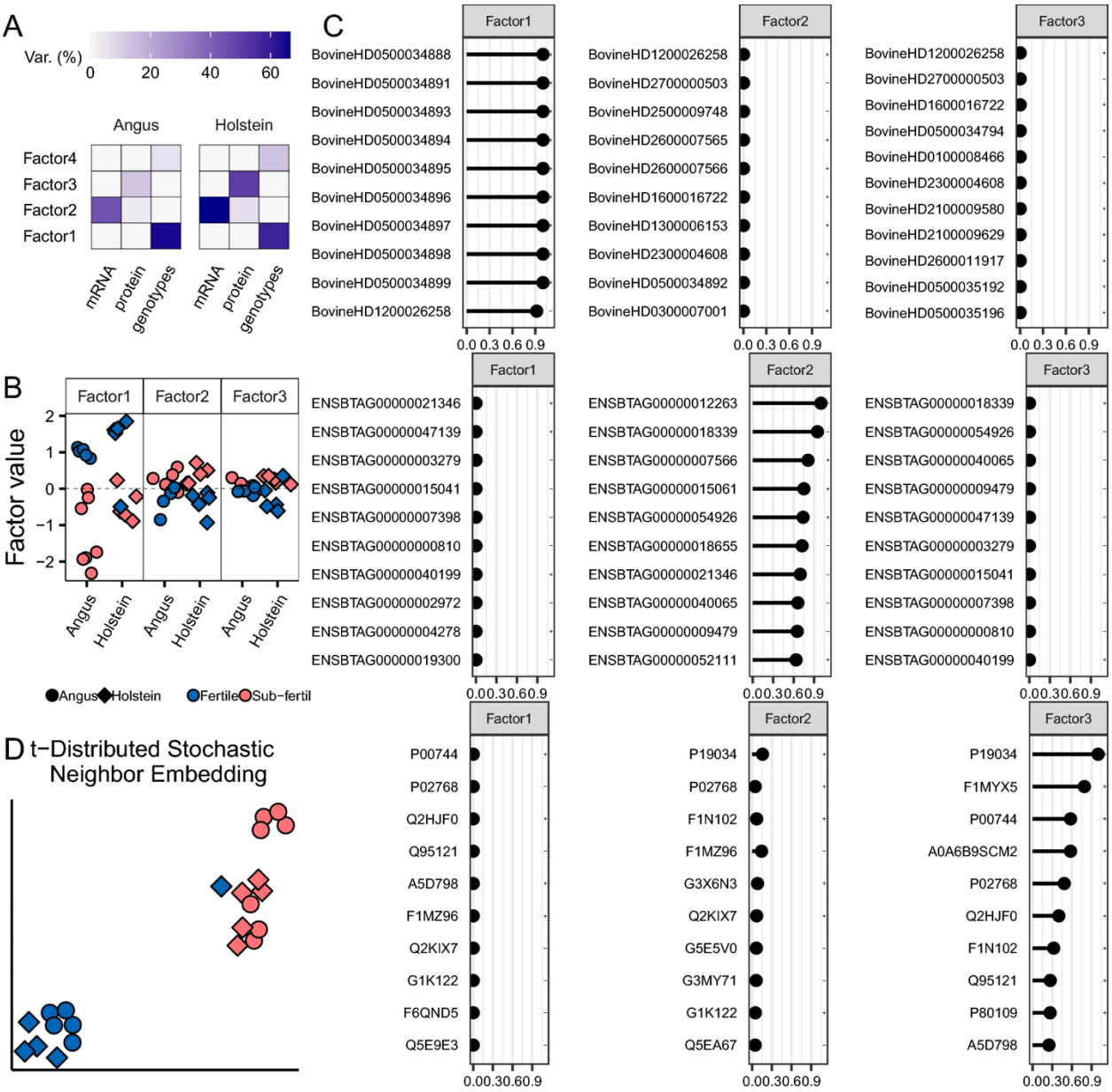


Figure 4.4. Multi-omics analysis of heifer fertility. (**A**) Percentage of variance explained by each factor within each data modality. (**B**) Relative separation of samples based on their phenotype for each factor plotted. (**C**) The relative weight of importance of the top ten features (SNP identifiers are from the array annotation, gene identifiers are from Ensembl, protein identifiers are from UniProt) in each data modality and factor plotted. (**D**) sample clustering using all three data modes.

# CHAPTER 5:
# Higher abundance of 2-dehydro-D-gluconate in the plasma of sub-fertile or infertile *Bos taurus* heifers

## 5.1 Abstract

Infertility or subfertility impacts approximately 5% and 15% of dairy and beef heifers (*Bos taurus*), respectively. Heifers that do not produce a calf within an optimum window of time have a significant negative impact on the profitability and sustainability of the cattle industry. Selection of heifers based on their fertility potential remains a challenge yet to be resolved. Here, we tested the hypothesis that heifers of different fertility potential have differing metabolome signatures in their plasma. We obtained blood from *Bos taurus* heifers at their first artificial insemination and processed the samples to separate the plasma. The heifers were classified based on their reproductive outcome as fertile (pregnant and delivered a calf after their first artificial insemination (AI)) or sub-fertile (Angus heifers: no pregnancy after two AI and exposure to a bull; Holstein heifers: no pregnancy by the third AI). We tested the relative abundance of 140 metabolites obtained from 22 heifers (Angus fertile n = 5, Angus sub-fertile n = 7, Holstein fertile N = 5, Holstein sub-fertile N = 5). The metabolite 2-Dehydro-D-gluconate ($C_6H_{10}O_7$) was significantly more abundant in the plasma of sub-fertile heifers in both breeds (1.4-fold, false discovery rate <0.1). In the context that a small proportion of circulating metabolites in the plasma were quantified in this study, the results show that the metabolomic profile in the blood stream may be associated with heifer fertility potential.

## 5.2 Introduction

Heifers that calve at an optimum age have greater productivity and longevity in the herd [265, 267, 425-428]. Infertility or subfertility is a critical barrier to sustainable cattle production [263], including in heifers. For example, approximately 15% [58] and 5% [7] of beef and dairy heifers, respectively, do not calve at 24 months of age. Therefore, identifying heifers with optimum fertility is a promising approach to improving sustainability in cattle production.

The broad detection and quantification of metabolites, named metabolomics [501], can provide critical biological insights into multiple aspects of the physiology in complex organisms [502]. Multiple studies have used metabolomics approaches to understand metabolic profiles associated with nutritional conditions [503-508] and female reproductive phenotypes [445, 509, 510] in cattle. The plasma component of the blood is a major carrier of metabolites produced in mammals, and 3126 of those metabolites have been quantified in the plasma of human blood (https://serummetabolome.ca/statistics, 12/29/2023, [511]). Because blood carries metabolites used, secreted or disposed by nearly all organs, this connective tissue is valuable for the discovery of molecular signatures, including metabolites, linked to complex phenotypes in a biological system [512].

Our group reported that transcript abundances of specific genes are altered in the circulating white blood cells of crossbred heifers of different fertility potential [198, 199, 363], and recently reported genotypes, transcript and protein abundances in the bloodstream associated with fertility in *Bos taurus* purebred heifers [513]. Here, we tested the hypothesis that specific metabolites would be

differentially abundant in plasma of Holstein and Angus heifers (*Bos taurus*) with different fertility potentials. The objective of this study was to determine the metabolite profiles of heifers at the time of artificial insemination and contrast metabolite abundances between heifers classified as fertile or sub-fertile.

## 5.3 Materials and Methods

Animal handling for this experiment was approved by the Institutional Animal Care and Use Committee (IACUC) at Virginia Polytechnic Institute and State University.

### Heifer classification

We collected blood samples from purebred Angus heifers (n=12), averaging 14 months in age, at their first artificial insemination service. Heifers were subjected to a 7-Day Co-Synch + CIDR estrus synchronization protocol prior to breeding. Briefly, heifers were administered an intramuscular (IM) injection of gonadotropin-releasing hormone (GnRH, 100 μg; Factrel®; Zoetis Inc.) on Day 0, followed by the insertion of a controlled internal drug release (CIDR, 1.38 g Progesterone; Eazi-Breed™ CIDR®; Zoetis Inc.). On Day 7, the CIDR was removed and an injection of prostaglandin F2 alpha (PGF2α, 25 μg; Lutalyse®; Zoetis Inc.) was delivered. Fixed-time artificial insemination was performed 54±2 hours following CIDR removal along a second injection of GnRH.

Additionally, we collected blood samples from purebred Holstein (n=10) heifers, averaging 12 months in age, at the time of the first artificial insemination service. Heifers were enrolled in a 5-Day CIDR-Synch protocol before artificial insemination. Briefly, an IM injection of GnRH was delivered on Day 0 with the insertion of a CIDR device. The CIDR device was removed on Day 5, followed by an IM injection of PGF2α. A second injection of PGF2α was administered 24 hours later. Then, timed AI was performed with a second GnRH injection on Day 8.

Heifers were identified as Fertile (Holstein, n=5; Angus, n=5) or Sub-Fertile (Holstein, n=5; Angus, n=7) based on their pregnancy outcome, following similar criteria used previously [198, 199]. Fertile animals were identified as those who became pregnant and subsequently delivered a calf following the first insemination service. Angus heifers were categorized as Sub-Fertile after failing to achieve pregnancy following two insemination services and exposure to a bull for natural breeding. Holstein heifers were identified as Sub-Fertile after needing four or more artificial inseminations.

Heifers were synchronized with protocols that have been identified by prior research to have high success for a heifer to become pregnant to artificial insemination [458, 459]. Hence the different protocols for beef and dairy heifers. The criteria for classification were different for each group due to differences in management that are inherent to beef and dairy replacement heifers. Most importantly, each heifer had multiple opportunities to become pregnant before being classified as sub-fertile. The heifers utilized in this study were not part of a nutritional experiment, and thus nutrition was not accounted as a variable nor was it a factor in the selection of heifers. All dairy heifers were raised with equivalent exposure to feed. Similarly, all beef heifers were raised with equivalent exposure to feed.

**Blood Sample Collection and plasma separation**

Ten ml of blood were drawn from each animal by venipuncture of the jugular vein using 18 mg K2 EDTA vacutainers (Becton, Dickinson, and Company). The tubes were inverted for proper mixing with the anticoagulant and then immediately placed on ice until further processing.

We processed the blood samples following procedures described elsewhere [198, 199, 460] within three hours of sampling [460]. Tubes containing whole blood samples were centrifuged for 25 minutes at 4°C and 2,000*xg*. Two ml of plasma were aspirated and centrifuged at 1,000*xg* for 10 minutes at 4°C to pellet any remaining cells. The supernatant was deposited in a cryotube and snap-frozen with liquid nitrogen. Samples were then stored at -80°C until further processing.

**Metabolome data collection**

Ultra-High-Performance Liquid Chromatography-High Resolution Mass Spectrometry (UHPLC-HRMS) analysis was performed on plasma samples at the Biological and Small Molecule Mass Spectrometry Core (RRID: SCR_021368) at the University of Tennessee, Knoxville, as previously described [506]. One hundred microliter aliquots of plasma samples were thawed at 4 °C for 30 minutes, and metabolites within each sample were extracted using a 40:40:20 methanol/acetonitrile/water solution with 0.1 M formic acid [514, 515]. Extraction solvent (1.3 mL) was added to the samples which underwent agitation and vortexing before samples were then chilled at −20°C for 20 minutes. Once samples were properly chilled, the tubes were centrifuged at 4 °C and 15000 rpm for 5 minutes to form pellets and remove debris from the sample. A second set of 2 mL microcentrifuge tubes were used to collect the supernatant from the first set of tubes. Extraction solvent (0.2 mL) was again dispensed into the first set of tubes, which contained the pellet, and the pipet tip was used to suspend the pellet in the extraction solvent. The tubes were then re-submitted to agitation and vortexing, chilling, and centrifugation as described above. The supernatant from the tube was added to the second set of tubes which contained the previous supernatant. The first tube set with the final pellet was then discarded. The second set of tubes containing the supernatant underwent drying using nitrogen gas. Once dried, tubes were filled with 300 µL of LCMS grade water for sample resuspension. The resuspended samples were vortexed and centrifuged at 4 °C and 15000 rpm for 5 minutes before transferring an aliquot to new autosampler vials for UHPLC-HRMS analysis. Metabolites present in samples were separated using chromatography column (Synergi Hydro RP, 2.5 µm, 100 mm × 2.0 mm column; Phenomenex, Torrance, CA, United States) which was maintained at 25°C. The mobile phase solvents used to elute metabolites were 1) 97:3 LCMS grade water: methanol with 15 mM acetic acid and 11 mM tributylamine and 2) 100% LCMS grade methanol. At a flow rate of 0.2 mL/minute, the solvent gradient was 1) 100% and 2) 0% from 0 to 5 minutes, 1) 80% and 2) 20% from 5 to 13 minutes, 1) 45% and 2) 55% from 13 to 15.5 minutes, 1) 5% and 2) 95% from 15.5 to 19 minutes, and 1) 100% and 2) 0% from 19 to 25 minutes. An Exactive Plus Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, United States) with an electrospray ionization probe attached was used for full scan mass analysis, operating in negative polarity mode with a scan range between 72 and 1,000 m/z, a resolution of 140,000, and an acquisition gain control of $3 \times 10^6$.

The tandem liquid chromatography and mass spectrometry analysis generated Xcalibur (RAW) files which were converted to an open source mzML format (msconvert software; ProteoWizard package) to enable data centroiding. The centroided data were processed using the Metabolomic Analysis and Visualization Engine (MAVEN; mzroll software, Princeton University) to identify metabolites that are present in the samples by comparing their retention time and mass-to-charge ratio with an in-house standard library. Peak identification for each metabolite is carried out using a variety of factors including peak shape, retention time, mass-to-charge ratio, signal-to-noise-ratio and this is validated using the natural abundance of isotopes in the compound The program then generates pre-processed peak data tables. Normalized data was obtained by dividing the the raw peak areas by the volume of sample used for metabolite extraction.

**Analysis of metabolome data**

First, we transformed the normalized metabolome data using logarithm ($Log_2(x)$). In order to increase the robustness of our findings we analyzed the metabolome data using two different univariate methodologies [516]. We analyzed the transformed data using a generalized linear model [470] using the "lm" function from the "stats" R package, which included the fertility group (Fertile or Sub-Fertile), breed (Angus or Holstein) as fixed effects. Then, we used the function "emmeans", to test the significance of the difference of adjusted means ($\mu$) ($H_0$:$\mu_1=\mu_2$, $H_1$:$\mu_1\neq\mu_2$) with the Student's t test [471], to calculate the estimated differences in metabolite abundance. We also analyzed the data using the R package 'limma' [472, 517]. We accounted for the same independent variables mentioned above (fertility group and breed) to fit the model and then tested for a differential abundance of the metabolites using the empirical Bayes Statistics implemented in the function "eBayes" [473, 474]. In both analyses, we adjusted the nominal P values using false discovery rate (FDR) [384]. Significance was inferred if FDR< 0.1 in both approaches.

**5.4 Results and Discussion**

We collected data for 140 metabolites present in the plasma of heifers. The majority of metabolites identified and quantified were "amino acids; amino acid precursors and derivatives" (n=40), followed by "nucleosides, nucleotides and analogues" (n=21). In addition, there were compounds classified as "amino acid metabolism" (n=12), "carbohydrates and carbohydrate conjugates" (n=11) and "lipids and lipid-like molecules" (n=11), among other functional classification. Interestingly, the top three most abundant compounds participate in energy related metabolic pathways (Citrate/isocitrate, TCA cycle; Lactate and Pyruvate, Glycolysis & Gluconeogenesis) (Figure 5.1A).

The contrast of metabolites abundance between heifers of different genetic background resulted in the identification of 2-Dehydro-D-gluconate ($C_6H_{10}O_7$) as differentially abundant between the two groups of fertility (fold change=1.41, FDR<0.1). Our analysis within breeds also confirmed that the metabolite 2-Dehydro-D-gluconate was more abundant in the plasma of sub-fertile Angus and Holstein heifers at a similar fold change (Figure 5.1B).

In a similar work [445], 15 metabolites were identified as significantly associated with fertility in heifers. Ten of those metabolites were also identified in our samples, but were not statistically different in our tests. One possible source of difference in the results is that Phillips et al sampled

angus crossed heifers whereas we worked with purebred Angus and Holstein heifers. In addition, it is unclear whether 2-Dehydro-D-gluconate was identified in the samples by Phillips et al for comparison with our results.

The abundance of the metabolite 2-Dehydro-D-gluconate has been connected with a fertility related phenotype. In beef cows (*Bos taurus*), there is a tendency of pregnancy loss following induced ovulation of dominant follicles < 12.1 mm in diameter as compared to when ovulation occurs when the dominant follicle is 14.7 mm in diameter [518]. Also in beef cows, the abundance of 2-Dehydro-D-gluconate in the follicular fluid is positively correlated with the follicle diameter at the time of administration of GnRH to induce ovulation [510]. Notably, in the same study, d-Gluconate also showed positive correlation with follicle diameter at the time of administration of GnRH to induce ovulation [510]. d-Gluconate had the second smallest P-values in our analysis, but did not reach the FDR for statistical inference of significant difference between fertility groups. Our results allied with the literature show an important connection of 2-Dehydro-D-gluconate with fertility in *Bos taurus* heifers.

The connection between greater quantities of 2-Dehydro--gluconate, a compound in the pentose phosphate pathway, in the blood of sub-fertile heifers is not trivial. One possible explanation is that fertile heifers could have a greater abundance of circulating estradiol [519] which would stimulate more activity in pentose phosphate pathway [520] in fertile heifers relative to sub-fertile heifers. Consequently, there would be less 2-Dehydro-D-gluconate in the bloodstream. We did not measure circulating estradiol in our samples to confirm this assumption, however. Another plausible explanation involves the fact that microorganisms in the gut microbiome produce and utilize 2-Dehydro-D-gluconate, and thus variations in the microbiome could also be a source differential abundance [521] of 2-Dehydro-D-gluconate in the plasma of heifers of different fertility potential.

In summary, we report the profile of 140 metabolites in plasma sampled from Angus and Holstein heifers (*Bos taurus*). A robust statistical analysis identified a greater abundance of the metabolite 2-Dehydro-D-gluconate in heifers that failed to produce a calf after multiple breeding attempts. Considering the highly variable nature of metabolites in the bloodstream [512], the differential abundance of 2-Dehydro-D-gluconate in two distinct groups of heifers enhance the robustness of this finding. This is the second report associating 2-Dehydro-D-gluconate with fertility phenotype in *Bos taurus*. Thus, further research may deepen our understanding of the connection of this metabolite and possibly the pentose phosphate pathway with female fertility.

## 5.5 Figures

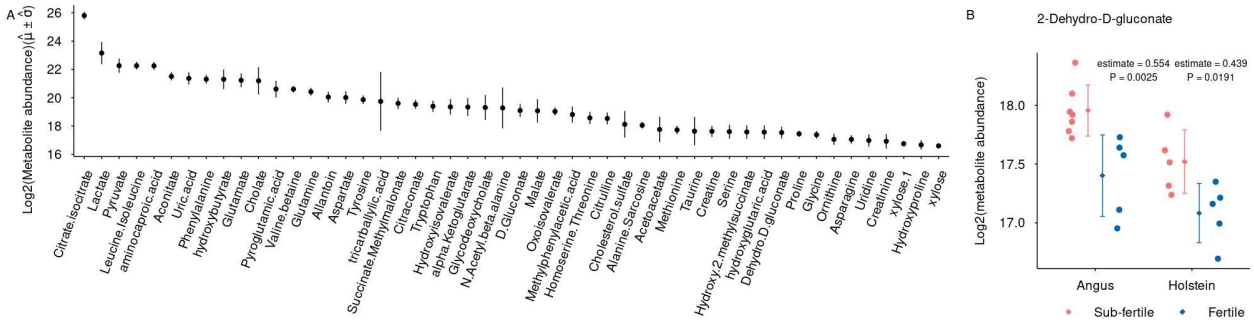


Figure 5.1. Metabolites in the plasma of heifers. (A) Forty most abundant metabolites detected in the plasma of heifers at breeding (mean ± standard deviation). (B) Differential abundance of 2-Dehydro-D-gluconate based on the reproductive outcome of Angus and Holstein heifers. Dots represent the data points; diamond represent the mean and error bars represent one standard deviation.

# CHAPTER 6:
# Conclusions

Over the last 20 years there has been a push to improve cattle fertility due to the financial burden it places on producers. Traditional selection has shown to be an inefficient method for the improvement of female fertility; therefore, research has begun to investigate the use of molecular markers to identify reproductively efficient cattle earlier and more accurately. The objective of this work was to identify molecular markers that could be used to differentiate between heifers of varying fertility potential. Based on the results in chapter 2, we were able to conclude that genotypic differences exist between heifers of different fertility potential as well as between heifers who left the herd before 13 months of age due to health reasons and those who remained in the herd as replacement heifers. Importantly, one SNP that tended to be associated with infertility in this study is located just upstream of the gene *NUFIP1* in cattle, and preliminary results from our lab show promising evidence that knockout of *NUFIP1* leads to embryonic mortality.

Most of the eQTL studies in recent years have been utilizing RNA-sequencing data to conduct their analyses, and RNA-seq data is count data, which doesn't follow a normal distribution. However, all of the statistical methods that are currently available for eQTL analysis assume that the data follows a normal distribution, therefore the data needs to be transformed in order to meet this assumption. Since forcing RNA-seq data to fit a normal distribution is not ideal, we developed a method to attempt to identify eQTLs without having to transform the data. The results from chapter 3 show that it is possible to identify biologically relevant eQTLs without forcing the data to fit a normal distribution.

Studies that only look at one level of biological information, such as the genome or transcriptome, are limited in the amount of information they can tell us about the underlying biology of complex traits. Therefore, studies that integrate several types of -omics data to see how their interactions are influencing the phenotype will be crucial in elucidating the biological mechanisms that underly the observed fertility differences. The separate analyses of the transcriptome, proteome, and metabolome described in chapters 4 and 5 were able to identify molecular markers that differed between fertile and sub-fertile heifers. Notably, the data from these analyses highlight the importance of metabolism on fertility and point to metabolic imbalance as a possible reason for the worse fertility in the sub-fertile heifers. Overall, the results from this work highlight the complex nature of female fertility traits in cattle and show that differences in a heifer's molecular profile can contribute to her fertility potential.

# REFERENCES


1. Fetrow J, Nordlund KV, Norman HD. Invited review: Culling: nomenclature, definitions, and recommendations. J Dairy Sci. 2006;89(6):1896-905. Epub 2006/05/17. doi: 10.3168/jds.S0022-0302(06)72257-3. PubMed PMID: 16702253.
2. McConnel CS, Lombard JE, Wagner BA, Garry FB. Evaluation of factors associated with increased dairy cow mortality on United States dairy operations. J Dairy Sci. 2008;91(4):1423-32. Epub 2008/03/20. doi: 10.3168/jds.2007-0440. PubMed PMID: 18349234.
3. Bascom SS, Young AJ. A summary of the reasons why farmers cull cows. J Dairy Sci. 1998;81(8):2299-305. Epub 1998/09/28. doi: 10.3168/jds.S0022-0302(98)75810-2. PubMed PMID: 9749397.
4. Mathews KH, Short SD. The beef cow replacement decision. J Agribus. 2001;19(2):191-211.
5. Moreira LC, Rosa GJM, Schaefer DM. Beef production from cull dairy cows: a review from culling to consumption. J Anim Sci. 2021;99(7). Epub 2021/06/15. doi: 10.1093/jas/skab192. PubMed PMID: 34125214; PubMed Central PMCID: PMCPMC8281100.
6. Kuhn MT, Hutchison JL, Wiggans GR. Characterization of Holstein heifer fertility in the United States. J Dairy Sci. 2006;89(12):4907-20. Epub 2006/11/16. doi: 10.3168/jds.S0022-0302(06)72541-3. PubMed PMID: 17106123.
7. Galliou JM, Kiser JN, Oliver KF, Seabury CM, Moraes JGN, Burns GW, et al. Identification of loci and pathways associated with heifer conception rate in U.S. Holsteins. Genes (Basel). 2020;11(7). Epub 2020/07/12. doi: 10.3390/genes11070767. PubMed PMID: 32650431; PubMed Central PMCID: PMCPMC7397161.
8. Olynk NJ, Wolf CA. Expected net present value of pure and mixed sexed semen artificial insemination strategies in dairy heifers. J Dairy Sci. 2007;90(5):2569-76. Epub 2007/04/14. doi: 10.3168/jds.2006-460. PubMed PMID: 17430962.
9. Peeler ID, Nebel RL, Pearson RE, Swecker WS, Garcia A. Pregnancy rates after timed AI of heifers following removal of intravaginal progesterone inserts. J Dairy Sci. 2004;87(9):2868-73. Epub 2004/09/18. doi: 10.3168/jds.S0022-0302(04)73416-5. PubMed PMID: 15375046.
10. Ettema JF, Santos JE. Impact of age at calving on lactation, reproduction, health, and income in first-parity Holsteins on commercial farms. J Dairy Sci. 2004;87(8):2730-42. Epub 2004/08/26. doi: 10.3168/jds.S0022-0302(04)73400-1. PubMed PMID: 15328299.
11. Xu ZZ, Burton LJ. Reproductive performance of dairy heifers after estrus synchronization and fixed-time artificial insemination. J Dairy Sci. 1999;82(5):910-7. Epub 1999/05/26. doi: 10.3168/jds.S0022-0302(99)75309-9. PubMed PMID: 10342228.
12. USDA. Beef 2017, "Beef cow-calf management practices in the United States, 2017, report1". Fort Collins, CO: 2020 Contract No.: #.782.0520.
13. Bormann JM, Totir LR, Kachman SD, Fernando RL, Wilson DE. Pregnancy rate and first-service conception rate in Angus heifers. J Anim Sci. 2006;84(8):2022-5. Epub 2006/07/26. doi: 10.2527/jas.2005-615. PubMed PMID: 16864860.
14. Canal LB, Fontes PLP, Sanford CD, Mercadante VRG, DiLorenzo N, Lamb GC, et al. Relationships between feed efficiency and puberty in Bos taurus and Bos indicus-influenced replacement beef heifers. J Anim Sci. 2020;98(10). doi: 10.1093/jas/skaa319. PubMed PMID: WOS:000593058200027.

15. Dickinson SE, Elmore MF, Kriese-Anderson L, Elmore JB, Walker BN, Dyce PW, et al. Evaluation of age, weaning weight, body condition score, and reproductive tract score in pre-selected beef heifers relative to reproductive potential. J Anim Sci Biotechnol. 2019;10:18. Epub 2019/03/21. doi: 10.1186/s40104-019-0329-6. PubMed PMID: 30891236; PubMed Central PMCID: PMCPMC6390375.
16. Fortes MR, Snelling WM, Reverter A, Nagaraj SH, Lehnert SA, Hawken RJ, et al. Gene network analyses of first service conception in Brangus heifers: use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors. J Anim Sci. 2012;90(9):2894-906. Epub 20120627. doi: 10.2527/jas.2011-4601. PubMed PMID: 22739780.
17. Funston RN, Deutscher GH. Comparison of target breeding weight and breeding date for replacement beef heifers and effects on subsequent reproduction and calf performance. J Anim Sci. 2004;82(10):3094-9. doi: 10.2527/2004.82103094x. PubMed PMID: 15484963.
18. Funston RN, Larson DM. Heifer development systems: dry-lot feeding compared with grazing dormant winter forage. J Anim Sci. 2011;89(5):1595-602. doi: 10.2527/jas.2010-3095. PubMed PMID: 21521820.
19. Grings EE, Geary TW, Short RE, MacNeil MD. Beef heifer development within three calving systems. J Anim Sci. 2007;85(8):2048-58. doi: 10.2527/jas.2006-758. PubMed PMID: WOS:000248043600024.
20. Gutierrez K, Kasimanickam R, Tibary A, Gay JM, Kastelic JP, Hall JB, et al. Effect of reproductive tract scoring on reproductive efficiency in beef heifers bred by timed insemination and natural service versus only natural service. Theriogenology. 2014;81(7):918-24. doi: 10.1016/j.theriogenology.2014.01.008. PubMed PMID: 24560451.
21. Heredia D, Ojeda-Rojas OA, Londono MC, Lasso SD, Bisinotto RS, Binelli M, et al. A single dose of FSH or hCG during a split-time AI program did not enhance follicular growth or pregnancy per artificial insemination in beef heifers. J Appl Anim Res. 2023;51(1):434-40. doi: 10.1080/09712119.2023.2218477. PubMed PMID: WOS:001000604800001.
22. Kasimanickam R, Kasimanickam V, Kappes A. Timed artificial insemination strategies with or without short-term natural service and pregnancy success in beef heifers. Theriogenology. 2021;166:97-103. doi: 10.1016/j.theriogenology.2021.02.023. PubMed PMID: WOS:000637206400012.
23. Locke JWC, Thomas JM, Knickmeyer ER, Ellersieck MR, Yelich JV, Poock SE, et al. Comparison of long-term progestin-based protocols to synchronize estrus prior to natural service or fixed-time artificial insemination in Bos indicus-influenced beef heifers. Anim Reprod Sci. 2020;218:106475. Epub 20200430. doi: 10.1016/j.anireprosci.2020.106475. PubMed PMID: 32507258.
24. Lynch JM, Lamb GC, Miller BL, Brandt RT, Jr., Cochran RC, Minton JE. Influence of timing of gain on growth and reproductive performance of beef replacement heifers. J Anim Sci. 1997;75(7):1715-22. doi: 10.2527/1997.7571715x. PubMed PMID: 9222826.
25. Mallory DA, Nash JM, Ellersieck MR, Smith MF, Patterson DJ. Comparison of long-term progestin-based protocols to synchronize estrus before fixed-time artificial insemination in beef heifers. J Anim Sci. 2011;89(5):1358-65. doi: 10.2527/jas.2010-3694. PubMed PMID: WOS:000289884900014.
26. Martin JL, Creighton KW, Musgrave JA, Klopfenstein TJ, Clark RT, Adams DC, et al. Effect of prebreeding body weight or progestin exposure before breeding on beef heifer performance through the second breeding season. J Anim Sci. 2008;86(2):451-9. doi: 10.2527/jas.2007-0233. PubMed PMID: WOS:000253443800024.

27. Patterson DJ, Corah LR, Kiracofe GH, Stevenson JS, Brethour JR. Conception rate in Bos taurus and Bos indicus crossbred heifers after postweaning energy manipulation and synchronization of estrus with melengestrol acetate and fenprostalene. J Anim Sci. 1989;67(5):1138-47. doi: 10.2527/jas1989.6751138x. PubMed PMID: 2737973.
28. Peters SO, Kizilkaya K, Garrick DJ, Fernando RL, Reecy JM, Weaber RL, et al. Heritability and Bayesian genome-wide association study of first service conception and pregnancy in Brangus heifers. J Anim Sci. 2013;91(2):605-12. doi: 10.2527/jas.2012-5580. PubMed PMID: WOS:000319687800010.
29. Roberts AJ, Geary TW, Grings EE, Waterman RC, MacNeil MD. Reproductive performance of heifers offered ad libitum or restricted access to feed for a one hundred forty-day period after weaning. J Anim Sci. 2009;87(9):3043-52. Epub 20090522. doi: 10.2527/jas.2008-1476. PubMed PMID: 19465497.
30. Hughes H. Raised replacement heifers some economic considerations. Vet Clin N Am-Food A. 2013;29(3):643-+. doi: 10.1016/j.cvfa.2013.07.013. PubMed PMID: WOS:000327680400013.
31. Tozer PR, Heinrichs AJ. What affects the costs of raising replacement dairy heifers: A multiple-component analysis. J Dairy Sci. 2001;84(8):1836-44. doi: DOI 10.3168/jds.S0022-0302(01)74623-1. PubMed PMID: WOS:000170470700010.
32. Overton MW, Dhuyvetter KC. Symposium review: An abundance of replacement heifers: What is the economic impact of raising more than are needed? J Dairy Sci. 2020;103(4):3828-37. doi: 10.3168/jds.2019-17143. PubMed PMID: WOS:000519992900077.
33. Weller JI, Ron M. Genetic analysis of fertility traits in Israeli Holsteins by linear and threshold models. J Dairy Sci. 1992;75(9):2541-8. Epub 1992/09/01. doi: 10.3168/jds.S0022-0302(92)78016-3. PubMed PMID: 1452858.
34. Brzakova M, Zavadilova L, Pribyl J, Pasek P, Kasna E, Kranjcevicova A. Estimation of genetic parameters for female fertility traits in the Czech Holstein population. Czech J Anim Sci. 2019;64(5):199-206. doi: 10.17221/51/2018-Cjas. PubMed PMID: WOS:000469022700002.
35. Andersen-Ranberg IM, Klemetsdal G, Heringstad B, Steine T. Heritabilities, genetic correlations, and genetic change for female fertility and protein yield in Norwegian Dairy Cattle. J Dairy Sci. 2005;88(1):348-55. Epub 2004/12/14. doi: 10.3168/jds.S0022-0302(05)72694-1. PubMed PMID: 15591399.
36. Janson L. Studies on fertility traits in Swedish dairy-cattle .2. genetic-parameters. Acta Agr Scand. 1980;30(4):427-36. doi: Doi 10.1080/00015128009435290. PubMed PMID: WOS:A1980KV89800010.
37. Liu Z, Jaitner J, Reinhardt F, Pasman E, Rensing S, Reents R. Genetic evaluation of fertility traits of dairy cattle using a multiple-trait animal model. J Dairy Sci. 2008;91(11):4333-43. doi: 10.3168/jds.2008-1029. PubMed PMID: WOS:000260277200027.
38. Jansen J. Direct and maternal genetic-parameters of fertility traits in Friesian cattle. Livest Prod Sci. 1986;15(2):153-64. doi: Doi 10.1016/0301-6226(86)90024-2. PubMed PMID: WOS:A1986E053900003.
39. Liu A, Lund MS, Wang Y, Guo G, Dong G, Madsen P, et al. Variance components and correlations of female fertility traits in Chinese Holstein population. J Anim Sci Biotechnol. 2017;8:56. Epub 2017/07/07. doi: 10.1186/s40104-017-0189-x. PubMed PMID: 28680590; PubMed Central PMCID: PMCPMC5493847.

40. Pryce JE, Veerkamp RF, Thompson R, Hill WG, Simm G. Genetic aspects of common health disorders and measures of fertility in Holstein Friesian dairy cattle. Anim Sci. 1997;65:353-60. doi: Doi 10.1017/S1357729800008559. PubMed PMID: WOS:000071342800004.
41. de Haer L. Estimation of genetic parameters of fertility traits, for virgin heifers in The Netherlands. Proceedings of the 2013 Interbull meeting 2013;(47).
42. Jagusiak W. Fertility measures in Polish Black-and-White cattle. 1. Genetic parameters of heifer fertility traits. J Anim Feed Sci. 2005;14(3):423-33. doi: DOI 10.22358/jafs/67036/2005. PubMed PMID: WOS:000231912700002.
43. Toghiani S, Hay E, Sumreddee P, Geary TW, Rekaya R, Roberts AJ. Genomic prediction of continuous and binary fertility traits of females in a composite beef cattle breed. J Anim Sci. 2017;95(11):4787-95. Epub 2018/01/03. doi: 10.2527/jas2017.1944. PubMed PMID: 29293708; PubMed Central PMCID: PMCPMC6292294.
44. Bormann JM, Totir LR, Kachman SD, Fernando RL, Wilson DE. Pregnancy rate and first-service conception rate in Angus heifers. J Anim Sci. 2006;84:2022-5. doi: 10.2527/jas.2005-615. PubMed PMID: 16864860.
45. Hawken RJ, Zhang YD, Fortes MRS, Collis E, Barris WC, J.Corbet N, et al. Genome-wide association studies of female reproduction in tropically adapted beef cattle. J Anim Sci. 2012;90:1398-410. doi: 10.2527/jas2011-4410.
46. Peters SO, Kizilkaya K, Garrick DJ, Fernando RL, Reecy JM, Weaber RL, et al. Heritability and Bayesian genome-wide association study of first service conception and pregnancy in Brangus heifers. J Anim Sci. 2013;91:605-12. doi: 10.2527/jas.2012-5580. PubMed PMID: 23148252.
47. Doyle SP, Golden BL, Green RD, Brinks JS. Additive genetic parameter estimates for heifer pregnancy and subsequent reproduction in Angus females. J Anim Sci. 2000;78(8):2091-8. doi: 10.2527/2000.7882091x.
48. Andersen KJ, Lefever DG, Brinks JS, Odde KG. The use of reproductive-tract scoring in beef heifers. Agri-Practice. 1991;12(4):19-&. PubMed PMID: WOS:A1991GB71200002.
49. Raheja KL, Burnside EB, Schaeffer LR. Heifer fertility and its relationship with cow fertility and production traits in Holstein dairy cattle. J Dairy Sci. 1989;72(10):2665-9. Epub 1989/10/01. doi: 10.3168/jds.S0022-0302(89)79407-8. PubMed PMID: 2600228.
50. Muir BL, Fatehi J, Schaeffer LR. Genetic relationships between persistency and reproductive performance in first-lactation Canadian Holsteins. J Dairy Sci. 2004;87(9):3029-37. doi: DOI 10.3168/jds.S0022-0302(04)73435-9. PubMed PMID: WOS:000223459800033.
51. Tiezzi F, Maltecca C, Cecchinato A, Penasa M, Bittante G. Genetic parameters for fertility of dairy heifers and cows at different parities and relationships with production traits in first lactation. J Dairy Sci. 2012;95(12):7355-62. doi: 10.3168/jds.2012-5775. PubMed PMID: WOS:000311192900054.
52. Arthur PF, Archer JA, Johnston DJ, Herd RM, Richardson EC, Parnell PF. Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. J Anim Sci. 2001;79(11):2805-11. doi: 10.2527/2001.79112805x. PubMed PMID: 11768108.
53. Zhang F, Wang Y, Mukiibi R, Chen L, Vinsky M, Plastow G, et al. Genetic architecture of quantitative traits in beef cattle revealed by genome wide association studies of imputed whole genome sequence variants: I: feed efficiency and component traits. BMC Genomics. 2020;21(1):36. Epub 20200113. doi: 10.1186/s12864-019-6362-1. PubMed PMID: 31931702; PubMed Central PMCID: PMCPMC6956504.

54. Garcia-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-Lopez FJ, Van Tassell CP. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. P Natl Acad Sci USA. 2016;113(28):E3995-4004. Epub 2016/06/30. doi: 10.1073/pnas.1519061113. PubMed PMID: 27354521; PubMed Central PMCID: PMCPMC4948329.
55. Cammack KM, Thomas MG, Enns RM. Reproductive traits and their heritabilities in beef cattle. The Professional Animal Scientist. 2009;25(5):517-28. doi: https://doi.org/10.15232/S1080-7446(15)30753-1.
56. Wiggans GR, VanRaden PM, Cooper TA. The genomic evaluation system in the United States: Past, present, future. J Dairy Sci. 2011;94(6):3202-11. doi: 10.3168/jds.2010-3866. PubMed PMID: WOS:000290777800054.
57. Gutierrez-Reinoso MA, Aponte PM, Garcia-Herreros M. Genomic analysis, progress and future perspectives in dairy cattle selection: A review. Animals (Basel). 2021;11(3). Epub 2021/03/07. doi: 10.3390/ani11030599. PubMed PMID: 33668747; PubMed Central PMCID: PMCPMC7996307.
58. Moorey SE, Biase FH. Beef heifer fertility: importance of management practices and technological advancements. J Anim Sci Biotechnol. 2020;11:97. Epub 2020/10/06. doi: 10.1186/s40104-020-00503-9. PubMed PMID: 33014361; PubMed Central PMCID: PMCPMC7528292.
59. Cochran SD, Cole JB, Null DJ, Hansen PJ. Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. BMC Genet. 2013;14. doi: Artn 4910.1186/1471-2156-14-49. PubMed PMID: WOS:000320617800001.
60. Jiang JC, Ma L, Prakapenka D, VanRaden PM, Cole JB, Da Y. A large-scale genome-wide association study in US Holstein cattle. Front Genet. 2019;10. doi: 10.3389/fgene.2019.00412. PubMed PMID: WOS:000467988100001.
61. Ortega MS, Denicol AC, Cole JB, Null DJ, Hansen PJ. Use of single nucleotide polymorphisms in candidate genes associated with daughter pregnancy rate for prediction of genetic merit for reproduction in Holstein cows. Animal Genetics. 2016;47(3):288-97. doi: 10.1111/age.12420. PubMed PMID: WOS:000374991500002.
62. Neupane M, Geary TW, Kiser JN, Burns GW, Hansen PJ, Spencer TE, et al. Loci and pathways associated with uterine capacity for pregnancy and fertility in beef cattle. Plos One. 2017;12(12). doi: 10.1371/journal.pone.0188997. PubMed PMID: WOS:000417648600026.
63. Stegemiller MR, Murdoch GK, Rowan TN, Davenport KM, Becker GM, Hall JB, et al. Genome-wide association analyses of fertility traits in beef heifers. Genes. 2021;12(2). doi: 10.3390/genes12020217. PubMed PMID: WOS:000622562500001.
64. Akanno EC, Plastow G, Fitzsimmons C, Miller SP, Baron V, Ominski K, et al. Genome-wide association for heifer reproduction and calf performance traits in beef cattle. Genome. 2015;58(12):549-57. Epub 20150805. doi: 10.1139/gen-2015-0031. PubMed PMID: 26484575.
65. Bettegowda A, Yao J, Sen A, Li Q, Lee KB, Kobayashi Y, et al. JY-1, an oocyte-specific gene, regulates granulosa cell function and early embryonic development in cattle. P Natl Acad Sci USA. 2007;104(45):17602-7. doi: 10.1073/pnas.0706383104. PubMed PMID: WOS:000250897600013.
66. de Camargo GMF, Costa RB, de Albuquerque LG, Regitano LCD, Baldi F, Tonhati H. Association between JY-1 gene polymorphisms and reproductive traits in beef cattle. Gene. 2014;533(2):477-80. doi: 10.1016/j.gene.2013.09.126. PubMed PMID: WOS:000327829100003.

67. Hausman GJ, Barb CR, Lents CA. Leptin and reproductive function. Biochimie. 2012;94(10):2075-81. Epub 20120302. doi: 10.1016/j.biochi.2012.02.022. PubMed PMID: 22980196.
68. Wathes DC. Mechanisms linking metabolic status and disease with reproductive outcome in the dairy cow. Reprod Domest Anim. 2012;47:304-12. doi: 10.1111/j.1439-0531.2012.02090.x. PubMed PMID: WOS:000306918700044.
69. Zieba DA, Amstalden M, Williams GL. Regulatory roles of leptin in reproduction and metabolism: A comparative review. Domest Anim Endocrin. 2005;29(1):166-85. doi: 10.1016/j.domaniend.2005.02.019. PubMed PMID: WOS:000229890000014.
70. Vineeth MR, Surya T, Sivalingam J, Kumar A, Niranjan SK, Dixit SP, et al. Genome-wide discovery of SNPs in candidate genes related to production and fertility traits in Sahiwal cattle. Trop Anim Health Pro. 2020;52(4):1707-15. doi: 10.1007/s11250-019-02180-x. PubMed PMID: WOS:000542689000021.
71. Clempson AM, Pollott GE, Brickell JS, Bourne NE, Munce N, Wathes DC. Evidence that leptin genotype is associated with fertility, growth, and milk production in Holstein cows. J Dairy Sci. 2011;94(7):3618-28. doi: 10.3168/jds.2010-3626. PubMed PMID: WOS:000291827600046.
72. Yadav T, Magotra A, Kumar R, Bangar YC, Garg AR, Kumar S, et al. Evaluation of candidate genotype of leptin gene associated with fertility and production traits in Hardhenu (Bos taurus x Bos indicus) cattle. Reprod Domest Anim. 2020;55(12):1698-705. doi: 10.1111/rda.13826. PubMed PMID: WOS:000577692100001.
73. Daftary SS, Gore AC. IGF-1 in the brain as a regulator of reproductive neuroendocrine function. Exp Biol Med. 2005;230(5):292-306. doi: Doi 10.1177/153537020523000503. PubMed PMID: WOS:000228786800003.
74. Spicer LJ, Echternkamp SE. The ovarian insulin and insulin-like growth-factor system with an emphasis on domestic-animals. Domest Anim Endocrin. 1995;12(3):223-45. doi: Doi 10.1016/0739-7240(95)00021-6. PubMed PMID: WOS:A1995RG84800001.
75. Watson AJ, Westhusin ME, Winger QA. IGF paracrine and autocrine interactions between conceptus and oviduct. J Reprod Fertil. 1999:303-15. PubMed PMID: WOS:000083605300024.
76. Lucy MC. Regulation of ovarian follicular growth by somatotropin and insulin-like growth factors in cattle. J Dairy Sci. 2000;83(7):1635-47. doi: DOI 10.3168/jds.S0022-0302(00)75032-6. PubMed PMID: WOS:000088124900021.
77. Brickell JS, Bourne N, Cheng Z, Wathes DC. Influence of plasma IGF-I concentrations and body weight at 6 months on age at first calving in dairy heifers on commercial farms. Reprod Domest Anim. 2007;42:118-9. PubMed PMID: WOS:000249245800227.
78. Yimaz A, Davis ME, Simmen RCM. Analysis of female reproductive traits in Angus beef cattle divergently selected for blood serum insulin-like growth factor I concentration. Theriogenology. 2006;65(6):1180-90. doi: 10.1016/j.theriogenology.2005.06.018. PubMed PMID: WOS:000236159100015.
79. Patton J, Kenny DA, McNamara S, Mee JF, O'Mara FP, Diskin MG, et al. Relationships among milk production, energy balance, plasma analytes, and reproduction in Holstein-Friesian cows. J Dairy Sci. 2007;90(2):649-58. doi: DOI 10.3168/jds.S0022-0302(07)71547-3. PubMed PMID: WOS:000243599200013.
80. Velazquez MA, Spicer LJ, Wathes DC. The role of endocrine insulin-like growth factor-I (IGF-I) in female bovine reproduction. Domest Anim Endocrinol. 2008;35(4):325-42. Epub 20080808. doi: 10.1016/j.domaniend.2008.07.002. PubMed PMID: 18703307.

81. Brickell JS, Bourne N, McGowan MM, Wathes DC. Effect of growth and development during the rearing period on the subsequent fertility of nulliparous Holstein-Friesian heifers. Theriogenology. 2009;72(3):408-16. Epub 20090529. doi: 10.1016/j.theriogenology.2009.03.015. PubMed PMID: 19481791.
82. Mirzaei A, Sharifiyazdi H, Ahmadi MR, Ararooti T, Ghasrodashti AR, Kadivar A. The effect of polymorphism in gene of insulin-like growth factor-I on the serum periparturient concentration in Holstein dairy cows. Asian Pac J Trop Biomed. 2012;2(10):765-9. doi: 10.1016/S2221-1691(12)60226-3. PubMed PMID: 23569844; PubMed Central PMCID: PMCPMC3609217.
83. Nicolini P, Carriquiry M, Meikle A. A polymorphism in the insulin-like growth factor 1 gene is associated with postpartum resumption of ovarian cyclicity in Holstein-Friesian cows under grazing conditions. Acta Veterinaria Scandinavica. 2013;55. doi: 10.1186/1751-0147-55-11. PubMed PMID: WOS:000317047200001.
84. Silva Silveira PA, Butler WR, da Silva TC, Barros CC, Correa MN, Schneider A. Association of polymorphisms in the IGF-I, GHR and STAT5A genes with serum IGF-I concentration and reproductive performance of Holstein dairy cows. Anim Reprod Sci. 2019;211:106206. Epub 20191021. doi: 10.1016/j.anireprosci.2019.106206. PubMed PMID: 31785637.
85. Jamrozik J, Fatehi J, Kistemaker GJ, Schaeffer LR. Estimates of genetic parameters for Canadian Holstein female reproduction traits. J Dairy Sci. 2005;88(6):2199-208. doi: DOI 10.3168/jds.S0022-0302(05)72895-2. PubMed PMID: WOS:000229224500028.
86. Junior GAO, Schenkel FS, Alcantara L, Houlahan K, Lynch C, Baes CF. Estimated genetic parameters for all genetically evaluated traits in Canadian Holsteins. J Dairy Sci. 2021;104(8):9002-15. doi: 10.3168/jds.2021-20227. PubMed PMID: WOS:000671943900014.
87. Kawashima C, Munakata M, Matsui M, Miyamoto A, Kida K, Shimizu T. Polymorphism in promoter region of growth hormone receptor is associated with potential production capacity of insulin-like growth factor-1 in pre-pubertal Holstein heifers. J Anim Physiol an N. 2016;100(6):1037-40. doi: 10.1111/jpn.12470. PubMed PMID: WOS:000387441100005.
88. Abdulrahman N, Fair T. Contribution of the immune system to follicle differentiation, ovulation and early corpus luteum formation. Anim Reprod. 2019;16(3):440-8. Epub 20191023. doi: 10.21451/1984-3143-AR2019-0087. PubMed PMID: 32435287; PubMed Central PMCID: PMCPMC7234072.
89. Walsh SW, Fair T, Browne JA, Evans AC, McGettigan PA. Physiological status alters immunological regulation of bovine follicle differentiation in dairy cattle. J Reprod Immunol. 2012;96(1-2):34-44. Epub 20120912. doi: 10.1016/j.jri.2012.07.002. PubMed PMID: 22980436.
90. Abdulrahman Alrabiah N, Evans ACO, Fahey AG, Cantwell N, Lonergan P, McCormack J, et al. Immunological aspects of ovarian follicle ovulation and corpus luteum formation in cattle. Reproduction. 2021;162(3):209-25. Epub 20210804. doi: 10.1530/REP-21-0165. PubMed PMID: 34255737.
91. Neuvians TP, Schams D, Berisha B, Pfaffl MW. Involvement of pro-inflammatory cytokines, mediators of inflammation, and basic fibroblast growth factor in prostaglandin F2alpha-induced luteolysis in bovine corpus luteum. Biol Reprod. 2004;70(2):473-80. Epub 20031015. doi: 10.1095/biolreprod.103.016154. PubMed PMID: 14561657.
92. Penny LA, Armstrong D, Bramley TA, Webb R, Collins RA, Watson ED. Immune cells and cytokine production in the bovine corpus luteum throughout the oestrous cycle and after

induced luteolysis. J Reprod Fertil. 1999;115(1):87-96. doi: 10.1530/jrf.0.1150087. PubMed PMID: 10341726.
93. Walusimbi SS, Pate JL. Physiology and Endocrinology Symposium: role of immune cells in the corpus luteum. J Anim Sci. 2013;91(4):1650-9. Epub 20130219. doi: 10.2527/jas.2012-6179. PubMed PMID: 23422006.
94. Mansouri-Attia N, Oliveira LJ, Forde N, Fahey AG, Browne JA, Roche JF, et al. Pivotal role for monocytes/macrophages and dendritic cells in maternal immune response to the developing embryo in cattle. Biol Reprod. 2012;87(5):123. Epub 20121129. doi: 10.1095/biolreprod.112.101121. PubMed PMID: 23034158.
95. Raghupathy R. Th1-type immunity is incompatible with successful pregnancy. Immunol Today. 1997;18(10):478-82. doi: 10.1016/s0167-5699(97)01127-4. PubMed PMID: WOS:A1997YD18000005.
96. Salilew-Wondim D, Holker M, Rings F, Ghanem N, Ulas-Cinar M, Peippo J, et al. Bovine pretransfer endometrium and embryo transcriptome fingerprints as predictors of pregnancy success after embryo transfer. Physiol Genomics. 2010;42(2):201-18. doi: 10.1152/physiolgenomics.00047.2010. PubMed PMID: WOS:000279586900006.
97. Gilbert RO. Symposium review: Mechanisms of disruption of fertility by infectious diseases of the reproductive tract. J Dairy Sci. 2019;102(4):3754-65. Epub 20190214. doi: 10.3168/jds.2018-15602. PubMed PMID: 30772031.
98. Kaisho T, Akira S. Toll-like receptor function and signaling. J Allergy Clin Immunol. 2006;117(5):979-87; quiz 88. Epub 20060403. doi: 10.1016/j.jaci.2006.02.023. PubMed PMID: 16675322.
99. Vasselon T, Detmers PA. Toll receptors: a central element in innate immune responses. Infect Immun. 2002;70(3):1033-41. doi: 10.1128/IAI.70.3.1033-1041.2002. PubMed PMID: 11854180; PubMed Central PMCID: PMCPMC127779.
100. Wira CR, Fahey JV. The innate immune system: gatekeeper to the female reproductive tract. Immunology. 2004;111(1):13-5. doi: 10.1111/j.1365-2567.2004.01796.x. PubMed PMID: 14678193; PubMed Central PMCID: PMCPMC1782397.
101. Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. Cell. 2006;124(4):783-801. doi: 10.1016/j.cell.2006.02.015. PubMed PMID: 16497588.
102. Beutler B. Inferences, questions and possibilities in Toll-like receptor signalling. Nature. 2004;430(6996):257-63. doi: 10.1038/nature02761. PubMed PMID: 15241424.
103. Janeway CA Jr., Travers P, Walport M, Shlomchik MJ. Infectious agents and how they cause disease. Immunobiology: The Immune Sysyem in Health and Disease. 5 ed: New York: Garland Science; 2001. p. 382-8.
104. Koets A, Santema W, Mertens H, Oostenrijk D, Keestra M, Overdijk M, et al. Susceptibility to paratuberculosis infection in cattle is associated with single nucleotide polymorphisms in Toll-like receptor 2 which modulate immune responses against Mycobacterium avium subspecies para tuberculosis. Prev Vet Med. 2010;93(4):305-15. doi: 10.1016/j.prevetmed.2009.11.008. PubMed PMID: WOS:000274956000008.
105. Song YP, Sun LP, Guo AZ, Yang LG. Effects of toll-like receptor 2 gene mutation on resistance to bovine brucellosis. Livest Sci. 2014;170:30-4. doi: 10.1016/j.livsci.2014.10.014. PubMed PMID: WOS:000347509700005.
106. Sun LP, Song YP, Riaz H, Yang HZ, Hua GH, Guo AZ, et al. Polymorphisms in toll-like receptor 1 and 9 genes and their association with tuberculosis susceptibility in Chinese Holstein

cattle. Vet Immunol Immunopathol. 2012;147(3-4):195-201. doi: 10.1016/j.vetimm.2012.04.016. PubMed PMID: WOS:000305435100010.
107. Pinedo PJ, Galvao KN, Seabury CM. Innate immune gene variation and differential susceptibility to uterine diseases in Holstein cows. Theriogenology. 2013;80(4):384-90. Epub 20130612. doi: 10.1016/j.theriogenology.2013.04.027. PubMed PMID: 23768650.
108. Wathes DC, Cheng Z, Chowdhury W, Fenwick MA, Fitzpatrick R, Morris DG, et al. Negative energy balance alters global gene expression and immune responses in the uterus of postpartum dairy cows. Physiol Genomics. 2009;39(1):1-13. Epub 20090630. doi: 10.1152/physiolgenomics.00064.2009. PubMed PMID: 19567787; PubMed Central PMCID: PMCPMC2747344.
109. Santos JE, Thatcher WW, Chebel RC, Cerri RL, Galvao KN. The effect of embryonic death rates in cattle on the efficacy of estrus synchronization programs. Anim Reprod Sci. 2004;82-83:513-35. doi: 10.1016/j.anireprosci.2004.04.015. PubMed PMID: 15271477.
110. Wiltbank MC, Baez GM, Garcia-Guerra A, Toledo MZ, Monteiro PL, Melo LF, et al. Pivotal periods for pregnancy loss during the first trimester of gestation in lactating dairy cows. Theriogenology. 2016;86(1):239-53. Epub 20160422. doi: 10.1016/j.theriogenology.2016.04.037. PubMed PMID: 27238438.
111. Hoelker M, Held E, Salilew-Wondim D, Schellander K, Tesfaye D. Molecular signatures of bovine embryo developmental competence. Reprod Fertil Dev. 2013;26(1):22-36. doi: 10.1071/RD13255. PubMed PMID: 24305174.
112. Miravet-Valenciano JA, Rincon-Bertolin A, Vilella F, Simon C. Understanding and improving endometrial receptivity. Curr Opin Obstet Gynecol. 2015;27(3):187-92. doi: 10.1097/GCO.0000000000000173. PubMed PMID: 25827647.
113. Spencer TE, Forde N, Lonergan P. Insights into conceptus elongation and establishment of pregnancy in ruminants. Reprod Fertil Dev. 2017;29(1):84-100. doi: 10.1071/rd16359. PubMed PMID: WOS:000390419100010.
114. Khatib H, Huang W, Mikheil D, Schutzkus V, Monson RL. Effects of signal transducer and activator of transcription (STAT) genes STAT1 and STAT3 genotypic combinations on fertilization and embryonic survival rates in Holstein cattle. J Dairy Sci. 2009;92(12):6186-91. doi: 10.3168/jds.2009-2439. PubMed PMID: 19923622.
115. Khatib H, Monson RL, Schutzkus V, Kohl DM, Rosa GJ, Rutledge JJ. Mutations in the STAT5A gene are associated with embryonic survival and milk composition in cattle. J Dairy Sci. 2008;91(2):784-93. doi: 10.3168/jds.2007-0669. PubMed PMID: 18218766.
116. Wang X, Schutzkus V, Huang W, Rosa GJ, Khatib H. Analysis of segregation distortion and association of the bovine FGF2 with fertilization rate and early embryonic survival. Anim Genet. 2009;40(5):722-8. Epub 20090506. doi: 10.1111/j.1365-2052.2009.01904.x. PubMed PMID: 19456315.
117. Johnson MT, Mahmood S, Patel MS. Intermediary metabolism and energetics during murine early embryogenesis. J Biol Chem. 2003;278(34):31457-60. Epub 20030604. doi: 10.1074/jbc.R300002200. PubMed PMID: 12788927.
118. May-Panloup P, Boguenet M, El Hachem H, Bouet PE, Reynier P. Embryo and its mitochondria. Antioxidants-Basel. 2021;10(2). doi: ARTN 13910.3390/antiox10020139. PubMed PMID: WOS:000622056800001.
119. Van Blerkom J, Davis PW, Lee J. ATP content of human oocytes and developmental potential and outcome after in-vitro fertilization and embryo transfer. Hum Reprod. 1995;10(2):415-24. doi: 10.1093/oxfordjournals.humrep.a135954. PubMed PMID: 7769073.

120. Dumollard R, Carroll J, Duchen MR, Campbell K, Swann K. Mitochondrial function and redox state in mammalian embryos. Semin Cell Dev Biol. 2009;20(3):346-53. doi: 10.1016/j.semcdb.2008.12.013. PubMed PMID: 19530278.
121. Bain NT, Madan P, Betts DH. The early embryo response to intracellular reactive oxygen species is developmentally regulated. Reprod Fertil Dev. 2011;23(4):561-75. doi: 10.1071/Rd10148. PubMed PMID: WOS:000290166300007.
122. Ealy AD, Drost M, Hansen PJ. Developmental changes in embryonic resistance to adverse effects of maternal heat stress in cows. J Dairy Sci. 1993;76(10):2899-905. doi: 10.3168/jds.S0022-0302(93)77629-8. PubMed PMID: 8227617.
123. Sartori R, Sartor-Bergfelt R, Mertens SA, Guenther JN, Parrish JJ, Wiltbank MC. Fertilization and early embryonic development in heifers and lactating cows in summer and lactating and dry cows in winter. J Dairy Sci. 2002;85(11):2803-12. doi: 10.3168/jds.S0022-0302(02)74367-1. PubMed PMID: 12487447.
124. Kisseleva T, Bhattacharya S, Braunstein J, Schindler CW. Signaling through the JAK/STAT pathway, recent advances and future challenges. Gene. 2002;285(1-2):1-24. doi: 10.1016/s0378-1119(02)00398-0. PubMed PMID: WOS:000175066900001.
125. Teglund S, McKay C, Schuetz E, van Deursen JM, Stravopodis D, Wang DM, et al. Stat5a and Stat5b proteins have essential and nonessential, or redundant, roles in cytokine responses. Cell. 1998;93(5):841-50. doi: 10.1016/s0092-8674(00)81444-0. PubMed PMID: WOS:000073956700019.
126. Tscherner A, Brown AC, Stalker L, Kao J, Dufort I, Sirard MA, et al. STAT3 signaling stimulates miR-21 expression in bovine cumulus cells during in vitro oocyte maturation. Sci Rep. 2018;8(1):11527. Epub 20180801. doi: 10.1038/s41598-018-29874-w. PubMed PMID: 30068990; PubMed Central PMCID: PMCPMC6070548.
127. Campbell K, Swann K. Ca2+ oscillations stimulate an ATP increase during fertilization of mouse eggs. Dev Biol. 2006;298(1):225-33. doi: 10.1016/j.ydbio.2006.06.032. PubMed PMID: WOS:000241071100020.
128. Dumollard R, Campbell K, Halet G, Carroll J, Swann K. Regulation of cytosolic and mitochondrial ATP levels in mouse eggs and zygotes. Dev Biol. 2008;316(2):431-40. doi: 10.1016/j.ydbio.2008.02.004. PubMed PMID: WOS:000254870800021.
129. Dumollard R, Maranngos P, Fitzharris G, Swann K, Duchen M, Carroll J. Sperm-triggered [Ca2+] oscillations and Ca2+ homeostasis in the mouse egg have an absolute requirement for mitochondrial ATP production. Development. 2004;131(13):3057-67. Epub 2004 May 26. doi: 10.1242/dev.01181.
130. Ben-Meir A, Burstein E, Borrego-Alvarez A, Chong J, Wong E, Yavorska T, et al. Coenzyme Q10 restores oocyte mitochondrial function and fertility during reproductive aging. Aging Cell. 2015;14(5):887-95. Epub 20150626. doi: 10.1111/acel.12368. PubMed PMID: 26111777; PubMed Central PMCID: PMCPMC4568976.
131. Tran UC, Clarke CF. Endogenous synthesis of coenzyme Q in eukaryotes. Mitochondrion. 2007;7 Suppl(Suppl):S62-71. Epub 20070330. doi: 10.1016/j.mito.2007.03.007. PubMed PMID: 17482885; PubMed Central PMCID: PMCPMC1974887.
132. Ortega MS, Wohlgemuth S, Tribulo P, Siqueira LG, Cole JB, Hansen PJ. A single nucleotide polymorphism in COQ9 affects mitochondrial and ovarian function and fertility in Holstein cows. Biol Reprod. 2017;96(3):652-63. doi: 10.1093/biolre/iox004. PubMed PMID: 28339599.

133. May-Panloup P, Chretien MF, Jacques C, Vasseur C, Malthiery Y, Reynier P. Low oocyte mitochondrial DNA content in ovarian insufficiency. Hum Reprod. 2005;20(3):593-7. Epub 20041217. doi: 10.1093/humrep/deh667. PubMed PMID: 15608038.
134. Reynier P, May-Panloup P, Chretien MF, Morgan CJ, Jean M, Savagner F, et al. Mitochondrial DNA content affects the fertilizability of human oocytes. Mol Hum Reprod. 2001;7(5):425-9. doi: 10.1093/molehr/7.5.425. PubMed PMID: 11331664.
135. Tsai T, St John JC. The role of mitochondrial DNA copy number, variants, and haplotypes in farm animal developmental outcome. Domest Anim Endocrinol. 2016;56 Suppl:S133-46. Epub 20160331. doi: 10.1016/j.domaniend.2016.03.005. PubMed PMID: 27345311.
136. Richter K, Haslbeck M, Buchner J. The heat shock response: life on the verge of death. Mol Cell. 2010;40(2):253-66. doi: 10.1016/j.molcel.2010.10.006. PubMed PMID: 20965420.
137. Lindquist S, Craig EA. The heat-shock proteins. Annu Rev Genet. 1988;22:631-77. doi: 10.1146/annurev.ge.22.120188.003215. PubMed PMID: WOS:A1988R470400022.
138. Adamowicz T, Pers E, Lechniak D. A new SNP in the 3'-UTR of the hsp 70-1 gene in Bos taurus and Bos indicus. Biochem Genet. 2005;43(11-12):623-7. doi: 10.1007/s10528-005-9119-2. PubMed PMID: 16382367.
139. Basirico L, Morera P, Primi V, Lacetera N, Nardone A, Bernabucci U. Cellular thermotolerance is associated with heat shock protein 70.1 genetic polymorphisms in Holstein lactating cows. Cell Stress & Chaperones. 2011;16(4):441-8. doi: 10.1007/s12192-011-0257-7. PubMed PMID: WOS:000291883900008.
140. Ortega MS, Rocha-Frigoni NAS, Mingoti GZ, Roth Z, Hansen PJ. Modification of embryonic resistance to heat shock in cattle by melatonin and genetic variation in HSPA1L. J Dairy Sci. 2016;99(11):9152-64. Epub 20160907. doi: 10.3168/jds.2016-11501. PubMed PMID: 27614828.
141. Rosenkrans C, Jr., Banks A, Reiter S, Looper M. Calving traits of crossbred Brahman cows are associated with Heat Shock Protein 70 genetic polymorphisms. Anim Reprod Sci. 2010;119(3-4):178-82. Epub 20100212. doi: 10.1016/j.anireprosci.2010.02.005. PubMed PMID: 20227203.
142. Qiu XB, Shao YM, Miao S, Wang L. The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. Cell Mol Life Sci. 2006;63(22):2560-70. doi: 10.1007/s00018-006-6192-6. PubMed PMID: 16952052.
143. Gotoh T, Terada K, Oyadomari S, Mori M. hsp70-DnaJ chaperone pair prevents nitric oxide- and CHOP-induced apoptosis by inhibiting translocation of Bax to mitochondria. Cell Death Differ. 2004;11(4):390-402. doi: 10.1038/sj.cdd.4401369. PubMed PMID: 14752510.
144. Zhang B, Penagaricano F, Driver A, Chen H, Khatib H. Differential expression of heat shock protein genes and their splice variants in bovine preimplantation embryos. J Dairy Sci. 2011;94(8):4174-82. doi: 10.3168/jds.2010-4137. PubMed PMID: WOS:000293011700044.
145. Crews DH. The relationship between beef sire carcass EPD and progeny phenotype. Can J Anim Sci. 2002;82(4):503-6. doi: 10.4141/a02-037. PubMed PMID: WOS:000180940800003.
146. Crews DH, Jr., Pollak EJ, Quaas RL. Evaluation of Simmental carcass EPD estimated using live and carcass data. J Anim Sci. 2004;82(3):661-7. doi: 10.2527/2004.823661x. PubMed PMID: 15032422.
147. Diaz C, Notter DR, Beal WE. Relationship between milk expected progeny differences of polled Hereford sires and actual milk-production of their crossbred daughters. J Anim Sci. 1992;70(2):396-402. doi: 10.2527/1992.702396x. PubMed PMID: WOS:A1992HC23000011.
148. Marston TT, Simms DD, Schalles RR, Zoellner KO, Martin LC, Fink GM. Relationship of milk production, milk expected progeny difference, and calf weaning weight in angus and

simmental cow-calf pairs. J Anim Sci. 1992;70(11):3304-10. doi: 10.2527/1992.70113304x. PubMed PMID: 1459890.
149. Minick JA, Buchanan DS, Rupert SD. Milk production of crossbred daughters of high- and low-milk EPD Angus and Hereford bulls. J Anim Sci. 2001;79(6):1386-93. doi: 10.2527/2001.7961386x. PubMed PMID: 11424673.
150. Nobre PR, Misztal I, Tsuruta S, Bertrand JK, Silva LO, Lopes PS. Genetic evaluation of growth in nellore cattle by multiple-trait and random regression models. J Anim Sci. 2003;81(4):927-32. doi: 10.2527/2003.814927x. PubMed PMID: 12723081.
151. Evans JL, Golden BL, Bourdon RM, Long KL. Additive genetic relationships between heifer pregnancy and scrotal circumference in Hereford cattle. J Anim Sci. 1999;77(10):2621-8. PubMed PMID: WOS:000084446300006.
152. Smith BA, Brinks JS, Richardson GV. Relationships of sire scrotal circumference to offspring reproduction and growth. J Anim Sci. 1989;67(11):2881-5. PubMed PMID: WOS:A1989CC21800007.
153. Van Melis MH, Eler JP, Rosa GJM, Ferraz JBS, Figueiredo LGG, Mattos EC, et al. Additive genetic relationships between scrotal circumference, heifer pregnancy, and stayability in Nellore cattle. J Anim Sci. 2010;88(12):3809-13. doi: 10.2527/jas.2009-2127. PubMed PMID: WOS:000284473700004.
154. Alward KJ, Cockrum RR, Ealy AD. Associations of antral follicle count with fertility in cattle: A review. JDS Commun. 2023;4(2):132-7. Epub 20230131. doi: 10.3168/jdsc.2022-0283. PubMed PMID: 36974207; PubMed Central PMCID: PMCPMC10039241.
155. Burns DS, Jimenez-Krassel F, Ireland JLH, Knight PG, Ireland JJ. Numbers of antral follicles during follicular waves in cattle: Evidence for high variation among animals, very high repeatability in individuals, and an inverse association with serum follicle-stimulating hormone concentrations. Biol Reprod. 2005;73(1):54-62. doi: 10.1095/biolreprod.104.036277. PubMed PMID: WOS:000229978700007.
156. Ireland J, Ward F, Jimenez-Krassel F, Ireland JLH, Smith GW, Lonergan P, et al. Follicle numbers are highly repeatable within individual animals but are inversely correlated with FSH concentrations and the proportion of good-quality embryos after ovarian stimulation in cattle. Hum Reprod. 2007;22(6):1687-95. doi: 10.1093/humrep/dem071. PubMed PMID: WOS:000247470700029.
157. Ireland JJ, Zielak-Steciwko AE, Jimenez-Krassel F, Folger J, Bettegowda A, Scheetz D, et al. Variation in the ovarian reserve is linked to alterations in intrafollicular estradiol production and ovarian biomarkers of follicular differentiation and oocyte quality in cattle. Biol Reprod. 2009;80(5):954-64. doi: 10.1095/biolreprod.108.073791. PubMed PMID: WOS:000265581800013.
158. Morotti F, Barreiros TRR, Machado FZ, Gonzalez SM, Marinho LSR, Seneda MM. Is the number of antral follicles an interesting selection criterium for fertility in cattle? Anim Reprod. 2015;12(3):479-86.
159. Cushman RA, Allan MF, Kuehn LA, Snelling WM, Cupp AS, Freetly HC. Evaluation of antral follicle count and ovarian morphology in crossbred beef cows: investigation of influence of stage of the estrous cycle, age, and birth weight. J Anim Sci. 2009;87(6):1971-80. Epub 20090313. doi: 10.2527/jas.2008-1728. PubMed PMID: 19286826.
160. McNeel AK, Cushman RA. Influence of puberty and antral follicle count on calving day in crossbred beef heifers. Theriogenology. 2015;84(7):1061-6. Epub 20150621. doi: 10.1016/j.theriogenology.2015.06.010. PubMed PMID: 26197954.

161. Ireland JL, Scheetz D, Jimenez-Krassel F, Themmen AP, Ward F, Lonergan P, et al. Antral follicle count reliably predicts number of morphologically healthy oocytes and follicles in ovaries of young adult cattle. Biol Reprod. 2008;79(6):1219-25. Epub 20080903. doi: 10.1095/biolreprod.108.071670. PubMed PMID: 18768912.
162. Jimenez-Krassel F, Folger JK, Ireland JLH, Smith GW, Hou X, Davis JS, et al. Evidence that high variation in ovarian reserves of healthy young adults has a negative impact on the corpus luteum and endometrium during estrous cycles in cattle. Biol Reprod. 2009;80(6):1272-81. doi: 10.1095/biolreprod.108.075093. PubMed PMID: WOS:000266247300023.
163. Martinez MF, Sanderson N, Quirke LD, Lawrence SB, Juengel JL. Association between antral follicle count and reproductive measures in New Zealand lactating dairy cows maintained in a pasture-based production system. Theriogenology. 2016;85(3):466-75. doi: 10.1016/j.theriogenology.2015.09.026. PubMed PMID: WOS:000368217900014.
164. Diskin MG, Morris DG. Embryonic and early foetal losses in cattle and other ruminants. Reprod Domest Anim. 2008;43 Suppl 2:260-7. doi: 10.1111/j.1439-0531.2008.01171.x. PubMed PMID: 18638133.
165. Inskeep EK. Preovulatory, postovulatory, and postmaternal recognition effects of concentrations of progesterone on embryonic survival in the cow. J Anim Sci. 2004;82 E-Suppl:E24-39. doi: 10.2527/2004.8213_supplE24x. PubMed PMID: 15471804.
166. Mossa F, Jimenez-Krassel F, Walsh S, Berry DP, Butler ST, Folger J, et al. Inherent capacity of the pituitary gland to produce gonadotropins is not influenced by the number of ovarian follicles >= 3 mm in diameter in cattle. Reprod Fertil Dev. 2010;22(3):550-7. doi: 10.1071/rd09100. PubMed PMID: WOS:000274934300007.
167. Haughian JM, Ginther OJ, Kot K, Wiltbank MC. Relationships between FSH patterns and follicular dynamics and the temporal associations among hormones in natural and GnRH-induced gonadotropin surges in heifers. Reproduction. 2004;127(1):23-33. doi: 10.1530/rep.1.00030. PubMed PMID: WOS:000188876600004.
168. Singh J, Dominguez M, Jaiswal R, Adams GP. A simple ultrasound test to predict the superstimulatory response in cattle. Theriogenology. 2004;62(1-2):227-43. doi: 10.1016/j.theriogenology.2003.09.020. PubMed PMID: 15159116.
169. Batista EOS, Macedo GG, Sala RV, Ortolan M, Sa MF, Del Valle TA, et al. Plasma antimullerian hormone as a predictor of ovarian antral follicular population in Bos indicus (Nelore) and Bos taurus (Holstein) heifers. Reprod Domest Anim. 2014;49(3):448-52. doi: 10.1111/rda.12304. PubMed PMID: WOS:000335228900014.
170. Cruz RS, Cushman RA, Vinoles C. Antral follicular count is a tool that may allow the selection of more precocious Bradford heifers at weaning. Theriogenology. 2018;119:35-42. doi: 10.1016/j.theriogenology.2018.06.010. PubMed PMID: WOS:000443530200005.
171. Morotti F, Zangirolamo AF, da Silva NC, da Silva CB, Rosa CO, Seneda MM. Antral follicle count in cattle: advantages, challenges, and controversy. Anim Reprod. 2017;14(3):514-20. doi: 10.21451/1984-3143-ar994. PubMed PMID: WOS:000407654600007.
172. Pence M, BreDahl R, editors. Clinical use of reproductive tract scoring to predict pregnancy outcome. American Association of Bovine Practitioners Conference; 1998; Spokane, WA.
173. Pence M, Ensley D, Berghaus R, Rossi J, Wilson T, Cannon PT, editors. Improving reproductive efficiency through use of reproductive tract scoring in a group of beef replacement heifers. American Association of Bovine Practitioners Conference; 2007: The Bovine Practitioner.

174. LeFever DG, Odde KG. Predicting reproductive performance in beef heifers by reproductive tract evaluation before breeding. CSU Beef Program Report: Colorado State University, 1986.
175. Burris MJ, Priode BM. Effect of calving date on subsequent calving performance. J Anim Sci. 1958;17(3):527-33. doi: 10.2527/jas1958.173527x. PubMed PMID: WOS:A1958WM02300005.
176. Lesmeist.Jl, Burfenin.Pj, Blackwel.Rl. Date of first calving in beef cows and subsequent calf production. J Anim Sci. 1973;36(1):1-6. doi: 10.2527/jas1973.3611. PubMed PMID: WOS:A1973O625700001.
177. Holm DE, Thompson PN, Irons PC. The value of reproductive tract scoring as a predictor of fertility and production outcomes in beef heifers. J Anim Sci. 2009;87(6):1934-40. doi: 10.2527/jas.2008-1579. PubMed PMID: WOS:000266108600012.
178. Rosenkrans KS, Hardin DK. Repeatability and accuracy of reproductive tract scoring to determine pubertal status in beef heifers. Theriogenology. 2003;59(5-6):1087-92. doi: 10.1016/s0093-691x(02)01171-8. PubMed PMID: 12527058.
179. Dorsey BR, Kasimanickam R, Whittier WD, Nebel RL, Wahlberg ML, Hall JB. Effect of time from estrus to AI on pregnancy rates in estrous synchronized beef heifers. Anim Reprod Sci. 2011;127(1-2):1-6. Epub 20110730. doi: 10.1016/j.anireprosci.2011.07.014. PubMed PMID: 21911186.
180. Larson LL, Ball PJ. Regulation of estrous cycles in dairy cattle: A review. Theriogenology. 1992;38(2):255-67. doi: 10.1016/0093-691x(92)90234-i. PubMed PMID: 16727134.
181. Gonzalez-Padilla E, Ruiz R, LeFever D, Denham A, Wiltbank JN. Puberty in beef heifers. III. Induction of fertile estrus. J Anim Sci. 1975;40(6):1110-8. doi: 10.2527/jas1975.4061110x. PubMed PMID: 1141064.
182. Patterson DJ, Corah LR, Brethour JR. Response of prepubertal Bos-taurus and Bos-indicus X Bos-taurus heifers to melengestrol acetate with or without gonadotropin-releasing hormone. Theriogenology. 1990;33(3):661-8. doi: 10.1016/0093-691x(90)90543-3. PubMed PMID: WOS:A1990CT25600010.
183. Mallory DA, Wilson DJ, Busch DC, Ellersieck MR, Smith MF, Patterson DJ. Comparison of long-term progestin-based estrus synchronization protocols in beef heifers. J Anim Sci. 2010;88(11):3568-78. Epub 2010/07/27. doi: 10.2527/jas.2010-3084. PubMed PMID: 20656979.
184. Patterson DJ, Thomas JM, Martin NT, Nash JM, Smith MF. Control of estrus and ovulation in beef heifers. Vet Clin North Am Food Anim Pract. 2013;29(3):591-617. Epub 20130905. doi: 10.1016/j.cvfa.2013.07.009. PubMed PMID: 24182437.
185. Kaim M, Rosenberg M, Folman Y. Management of reproduction in dairy heifers based on the synchronization of estrous cycles. Theriogenology. 1990;34(3):537-47. doi: 10.1016/0093-691x(90)90010-q. PubMed PMID: 16726859.
186. Lamb GC, Larson JE, Geary TW, Stevenson JS, Johnson SK, Day ML, et al. Synchronization of estrus and artificial insemination in replacement beef heifers using gonadotropin-releasing hormone, prostaglandin F2alpha, and progesterone. J Anim Sci. 2006;84(11):3000-9. doi: 10.2527/jas.2006-220. PubMed PMID: 17032794.
187. Karakaya-Bilen E, Ribeiro ES, Bisinotto RS, Gumen A, Santos JEP. Effect of presynchronization with prostaglandin F-2 alpha before the 5-d timed AI protocol on ovarian responses and pregnancy in dairy heifers. Theriogenology. 2019;132:138-43. doi: 10.1016/j.theriogenology.2019.03.019. PubMed PMID: WOS:000469159900017.

188. Oosthuizen N, Fontes P, Lamb C. Presynchronization with Prostaglandin F-2 alpha and prolonged exposure to exogenous progesterone impacts estrus expression and alters fertility in beef heifers. J Anim Sci. 2020;98:41-. doi: 10.1093/jas/skz397.094. PubMed PMID: WOS:000605982900095.
189. Colazo MG, Kastelic JP, Mainar-Jaime RC, Gavaga QA, Whittaker PR, Small JA, et al. Resynchronization of previously timed-inseminated beef heifers with progestins. Theriogenology. 2006;65(3):557-72. Epub 20050721. doi: 10.1016/j.theriogenology.2005.06.001. PubMed PMID: 16039702.
190. Sa Filho MF, Marques MO, Girotto R, Santos FA, Sala RV, Barbuio JP, et al. Resynchronization with unknown pregnancy status using progestin-based timed artificial insemination protocol in beef cattle. Theriogenology. 2014;81(2):284-90. Epub 20130928. doi: 10.1016/j.theriogenology.2013.09.027. PubMed PMID: 24139935.
191. Stevenson JS, Johnson SK, Medina-Britos MA, Richardson-Adams AM, Lamb GC. Resynchronization of estrus in cattle of unknown pregnancy status using estrogen, progesterone, or both. Journal of Animal Science. 2003;81(7):1681-92. PubMed PMID: WOS:000186718100003.
192. Minten MA, Bilby TR, Bruno RG, Allen CC, Madsen CA, Wang Z, et al. Effects of fertility on gene expression and function of the bovine endometrium. PLoS One. 2013;8(8):e69444. Epub 20130805. doi: 10.1371/journal.pone.0069444. PubMed PMID: 23940519; PubMed Central PMCID: PMCPMC3734181.
193. Bazer FW, First NL. Pregnancy and parturition. J Anim Sci. 1983;57 Suppl 2:425-60. PubMed PMID: 6352591.
194. Killeen AP, Morris DG, Kenny DA, Mullen MP, Diskin MG, Waters SM. Global gene expression in endometrium of high and low fertility heifers during the mid-luteal phase of the estrous cycle. BMC Genomics. 2014;15. doi: 10.1186/1471-2164-15-234. PubMed PMID: WOS:000334951000001.
195. Moraes JGN, Behura SK, Geary TW, Hansen PJ, Neibergs HL, Spencer TE. Uterine influences on conceptus development in fertility-classified animals. P Natl Acad Sci USA. 2018;115(8):E1749-E58. doi: 10.1073/pnas.1721191115. PubMed PMID: WOS:000425495000010.
196. Geary TW, Burns GW, Moraes JGN, Moss JI, Denicol AC, Dobbs KB, et al. Identification of beef heifers with superior uterine capacity for pregnancy. Biol Reprod. 2016;95(2). doi: 10.1095/biolreprod.116.141390. PubMed PMID: WOS:000385942800013.
197. Fair T. The contribution of the maternal immune system to the establishment of pregnancy in cattle. Front Immunol. 2015;6. doi: 10.3389/fimmu.2015.00007. PubMed PMID: WOS:000354603500001.
198. Dickinson SE, Griffin BA, Elmore MF, Kriese-Anderson L, Elmore JB, Dyce PW, et al. Transcriptome profiles in peripheral white blood cells at the time of artificial insemination discriminate beef heifers with different fertility potential. BMC Genomics. 2018;19. doi: 10.1186/s12864-018-4505-4. PubMed PMID: WOS:000424781200001.
199. Moorey SE, Walker BN, Elmore MF, Elmore JB, Rodning SP, Biase FH. Rewiring of gene expression in circulating white blood cells is associated with pregnancy outcome in heifers (Bos taurus). Sci Rep. 2020;10(1):16786. Epub 2020/10/10. doi: 10.1038/s41598-020-73694-w. PubMed PMID: 33033295; PubMed Central PMCID: PMCPMC7544915.

200. Banerjee P, Diniz WJS, Hollingsworth R, Rodning SP, Dyce PW. mRNA signatures in peripheral white blood cells predict reproductive potential in beef heifers at weaning. Genes. 2023;14(2). doi: 10.3390/genes14020498. PubMed PMID: WOS:000945185700001.
201. Ambros V. The functions of animal microRNAs. Nature. 2004;431(7006):350-5. doi: 10.1038/nature02871. PubMed PMID: 15372042.
202. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004;116(2):281-97. doi: 10.1016/s0092-8674(04)00045-5. PubMed PMID: 14744438.
203. Chen L, Zhang J, Feng Y, Li R, Sun X, Du W, et al. MiR-410 regulates MET to influence the proliferation and invasion of glioma. Int J Biochem Cell Biol. 2012;44(11):1711-7. Epub 20120627. doi: 10.1016/j.biocel.2012.06.027. PubMed PMID: 22750473.
204. Gilad S, Meiri E, Yogev Y, Benjamin S, Lebanony D, Yerushalmi N, et al. Serum microRNAs are promising novel biomarkers. PLoS One. 2008;3(9):e3148. Epub 20080905. doi: 10.1371/journal.pone.0003148. PubMed PMID: 18773077; PubMed Central PMCID: PMCPMC2519789.
205. Li J, Smyth P, Flavin R, Cahill S, Denning K, Aherne S, et al. Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. BMC Biotechnol. 2007;7:36. Epub 20070629. doi: 10.1186/1472-6750-7-36. PubMed PMID: 17603869; PubMed Central PMCID: PMCPMC1914054.
206. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature. 2005;435(7043):834-8. doi: 10.1038/nature03702. PubMed PMID: 15944708.
207. Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. P Natl Acad Sci USA. 2006;103(7):2257-61. Epub 20060203. doi: 10.1073/pnas.0510565103. PubMed PMID: 16461460; PubMed Central PMCID: PMCPMC1413718.
208. Jairajpuri DS, Malalla ZH, Mahmood N, Khan F, Almawi WY. Differentially expressed circulating microRNAs associated with idiopathic recurrent pregnancy loss. Gene. 2021;768:145334. Epub 20201202. doi: 10.1016/j.gene.2020.145334. PubMed PMID: 33278550.
209. Luo H, Han Y, Liu J, Zhang Y. Identification of microRNAs in granulosa cells from patients with different levels of ovarian reserve function and the potential regulatory function of miR-23a in granulosa cell apoptosis. Gene. 2019;686:250-60. Epub 20181117. doi: 10.1016/j.gene.2018.11.025. PubMed PMID: 30453069.
210. Qasemi M, Amidi F. Extracellular microRNA profiling in human follicular fluid: new biomarkers in female reproductive potential. J Assist Reprod Genet. 2020;37(8):1769-80. Epub 20200709. doi: 10.1007/s10815-020-01860-0. PubMed PMID: 32642870; PubMed Central PMCID: PMCPMC7468023.
211. Sorensen AE, Udesen PB, Wissing ML, Englund ALM, Dalgaard LT. MicroRNAs related to androgen metabolism and polycystic ovary syndrome. Chem Biol Interact. 2016;259(Pt A):8-16. Epub 20160605. doi: 10.1016/j.cbi.2016.06.008. PubMed PMID: 27270454.
212. Hu SJ, Ren G, Liu JL, Zhao ZA, Yu YS, Su RW, et al. MicroRNA expression and regulation in mouse uterus during embryo implantation. J Biol Chem. 2008;283(34):23473-84. Epub 20080613. doi: 10.1074/jbc.M800406200. PubMed PMID: 18556655.
213. Tang F, Kaneda M, O'Carroll D, Hajkova P, Barton SC, Sun YA, et al. Maternal microRNAs are essential for mouse zygotic development. Genes Dev. 2007;21(6):644-8. doi: 10.1101/gad.418707. PubMed PMID: 17369397; PubMed Central PMCID: PMCPMC1820938.

214. Yao GD, Liang M, Liang N, Yin MM, Lu MR, Lian J, et al. MicroRNA-224 is involved in the regulation of mouse cumulus expansion by targeting Ptx3. Mol Cell Endocrinol. 2014;382(1):244-53. doi: 10.1016/j.mce.2013.10.014. PubMed PMID: WOS:000330421600026.
215. Song Y, An X, Zhang L, Fu M, Peng J, Han P, et al. Identification and profiling of microRNAs in goat endometrium during embryo implantation. PLoS One. 2015;10(4):e0122202. Epub 20150417. doi: 10.1371/journal.pone.0122202. PubMed PMID: 25886011; PubMed Central PMCID: PMCPMC4401794.
216. Zang X, Zhou C, Wang W, Gan J, Li Y, Liu D, et al. Differential microRNA expression involved in endometrial receptivity of goats. Biomolecules. 2021;11(3). Epub 20210322. doi: 10.3390/biom11030472. PubMed PMID: 33810054; PubMed Central PMCID: PMCPMC8004627.
217. Zhang XD, Zhang YH, Ling YH, Liu Y, Cao HG, Yin ZJ, et al. Characterization and differential expression of microRNAs in the ovaries of pregnant and non-pregnant goats (Capra hircus). BMC Genomics. 2013;14:157. Epub 20130307. doi: 10.1186/1471-2164-14-157. PubMed PMID: 23497306; PubMed Central PMCID: PMCPMC3599660.
218. Kaczmarek MM, Reliszko ZP, Szuszkiewicz J, Nitkiewicz A, Guzewska MM, Myszczynski K, et al. Profiling circulating microRNAs in the serum of pregnant and non-pregnant pigs reveals a plethora of reproductive status-dependent microRNAs. Animal. 2021;15(4):100182. Epub 20210224. doi: 10.1016/j.animal.2021.100182. PubMed PMID: 33640292.
219. Krawczynski K, Najmula J, Bauersachs S, Kaczmarek MM. MicroRNAome of porcine conceptuses and trophoblasts: Expression profile of microRNAs and their potential to regulate genes crucial for establishment of pregnancy. Biol Reprod. 2015;92(1). doi: ARTN 2110.1095/biolreprod.114.123588. PubMed PMID: WOS:000350806200006.

220. Su LJ, Liu RZ, Cheng W, Zhu MJ, Li XP, Zhao SH, et al. Expression patterns of microRNAs in porcine endometrium and their potential roles in embryo implantation and placentation. Plos One. 2014;9(2). doi: ARTN e8786710.1371/journal.pone.0087867. PubMed PMID: WOS:000330829200079.

221. Gad A, Sanchez JM, Browne JA, Nemcova L, Laurincik J, Prochazka R, et al. Plasma extracellular vesicle miRNAs as potential biomarkers of superstimulatory response in cattle. Sci Rep. 2020;10(1). doi: ARTN 1913010.1038/s41598-020-76152-9. PubMed PMID: WOS:000587666100026.

222. Ioannidis J, Donadeu FX. Circulating miRNA signatures of early pregnancy in cattle. BMC Genomics. 2016;17:184. Epub 20160303. doi: 10.1186/s12864-016-2529-1. PubMed PMID: 26939708; PubMed Central PMCID: PMCPMC4778341.
223. Pohler KG, Green JA, Moley LA, Gunewardena S, Hung WT, Payton RR, et al. Circulating microRNA as candidates for early embryonic viability in cattle. Mol Reprod Dev. 2017;84(8):731-43. doi: 10.1002/mrd.22856. PubMed PMID: WOS:000409342000008.
224. Sinha PB, Tesfaye D, Rings F, Hossien M, Hoelker M, Held E, et al. MicroRNA-130b is involved in bovine granulosa and cumulus cells function, oocyte maturation and blastocyst formation. J Ovarian Res. 2017;10. doi: ARTN 3710.1186/s13048-017-0336-1. PubMed PMID: WOS:000403589900001.

225. Massirer KB, Pasquinelli AE. The evolving role of microRNAs in animal gene expression. Bioessays. 2006;28(5):449-52. doi: 10.1002/bies.20406. PubMed PMID: 16615087.

226. Butler AE, Cunningham TK, Ramachandran V, Diboun I, Halama A, Sathyapalan T, et al. Association of microRNAs with embryo development and fertilization in women undergoing subfertility treatments: A pilot study. Front Reprod Health. 2021;3:719326. Epub 20210923. doi: 10.3389/frph.2021.719326. PubMed PMID: 36303988; PubMed Central PMCID: PMCPMC9580729.
227. Butler AE, Ramachandran V, Hayat S, Dargham SR, Cunningham TK, Benurwar M, et al. Expression of microRNA in follicular fluid in women with and without PCOS. Sci Rep. 2019;9. doi: 10.1038/s41598-019-52856-5. PubMed PMID: WOS:000495370900006.
228. Machtinger R, Rodosthenous RS, Adir M, Mansour A, Racowsky C, Baccarelli AA, et al. Extracellular microRNAs in follicular fluid and their potential association with oocyte fertilization and embryo quality: an exploratory study. J Assist Reprod Gen. 2017;34(4):525-33. doi: 10.1007/s10815-017-0876-8. PubMed PMID: WOS:000399824500015.
229. Martinez RM, Liang L, Racowsky C, Dioni L, Mansur A, Adir M, et al. Extracellular microRNAs profile in human follicular fluid and IVF outcomes. Sci Rep. 2018;8(1):17036. Epub 20181119. doi: 10.1038/s41598-018-35379-3. PubMed PMID: 30451969; PubMed Central PMCID: PMCPMC6242846.
230. Feng RZ, Sang Q, Zhu Y, Fu W, Liu M, Xu Y, et al. MiRNA-320 in the human follicular fluid is associated with embryo quality in vivo and affects mouse embryonic development in vitro. Sci Rep. 2015;5. doi: 10.1038/srep08689. PubMed PMID: WOS:000350350500010.
231. Lin XY, Beckers E, Mc Cafferty S, Gansemans Y, Szymanska KJ, Pavani KC, et al. Bovine embryo-secreted microRNA-30c is a potential non-invasive biomarker for hampered preimplantation developmental competence. Fronti Genet. 2019;10. doi: ARTN 31510.3389/fgene.2019.00315. PubMed PMID: WOS:000463623400002.

232. Melo-Baez B, Wong YS, Aguilera CJ, Cabezas J, Mancanares ACF, Riadi G, et al. MicroRNAs from extracellular vesicles secreted by bovine embryos as early biomarkers of developmental competence. Int J Mol Sci. 2020;21(23). doi: ARTN 888810.3390/ijms21238888. PubMed PMID: WOS:000597921200001.

233. Pasquariello R, Manzoni EFM, Fiandanese N, Viglino A, Pocar P, Brevini TAL, et al. Implications of miRNA expression pattern in bovine oocytes and follicular fluids for developmental competence. Theriogenology. 2020;145:77-85. doi: 10.1016/j.theriogenology.2020.01.027. PubMed PMID: WOS:000516888400010.
234. Coticchio G, Pennetta F, Rizzo R, Tarozzi N, Nadalini M, Orlando G, et al. Embryo morphokinetic score is associated with biomarkers of developmental competence and implantation. J Assist Reprod Gen. 2021;38(7):1737-43. doi: 10.1007/s10815-021-02162-9. PubMed PMID: WOS:000636970900001.
235. Fang F, Li ZL, Yu JY, Long YT, Zhao Q, Ding XF, et al. MicroRNAs secreted by human embryos could be potential biomarkers for clinical outcomes of assisted reproductive technology. J Adv Res. 2021;31:25-34. doi: 10.1016/j.jare.2021.01.003. PubMed PMID: WOS:000664399500003.
236. Scalici E, Traver S, Mullet T, Molinari N, Ferrieres A, Brunet C, et al. Circulating microRNAs in follicular fluid, powerful tools to explore in vitro fertilization process. Sci Rep. 2016;6:24976. Epub 20160422. doi: 10.1038/srep24976. PubMed PMID: 27102646; PubMed Central PMCID: PMCPMC4840336.
237. Banerjee P, Diniz WJS, Rodning SP, Dyce PW. miRNA expression profiles of peripheral white blood cells from beef heifers with varying reproductive potential. Front Genet.

2023;14:1174145. Epub 20230510. doi: 10.3389/fgene.2023.1174145. PubMed PMID: 37234872; PubMed Central PMCID: PMCPMC10206245.
238. Otsuka M, Zheng M, Hayashi M, Lee JD, Yoshino O, Lin S, et al. Impaired microRNA processing causes corpus luteum insufficiency and infertility in mice. J Clin Invest. 2008;118(5):1944-54. doi: 10.1172/JCI33680. PubMed PMID: 18398510; PubMed Central PMCID: PMCPMC2289794.
239. Ponsuksili S, Tesfaye D, Schellander K, Hoelker M, Hadlich F, Schwerin M, et al. Differential expression of miRNAs and their target mRNAs in endometria prior to maternal recognition of pregnancy associates with endometrial receptivity for in vivo- and in vitro-produced bovine embryos. Biol Reprod. 2014;91(6). doi: ARTN 13510.1095/biolreprod.114.121392. PubMed PMID: WOS:000346164800006.

240. Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances in human liver. Electrophoresis. 1997;18(3-4):533-7. doi: 10.1002/elps.1150180333. PubMed PMID: 9150937.
241. Graves PR, Haystead TA. Molecular biologist's guide to proteomics. Microbiol Mol Biol Rev. 2002;66(1):39-63; table of contents. doi: 10.1128/MMBR.66.1.39-63.2002. PubMed PMID: 11875127; PubMed Central PMCID: PMCPMC120780.
242. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol. 1999;19(3):1720-30. doi: 10.1128/mcb.19.3.1720. PubMed PMID: WOS:000078697900010.
243. Baldassini WA, Bonilha SFM, Branco RH, Vieira JCS, Padilha PM, Lanna DPD. Proteomic investigation of liver from beef cattle (Bos indicus) divergently ranked on residual feed intake. Mol Biol Rep. 2018;45(6):2765-73. Epub 20180903. doi: 10.1007/s11033-018-4341-2. PubMed PMID: 30178216.
244. Bathla S, Sindhu A, Kumar S, Dubey SK, Pattnaik S, Rawat P, et al. Tandem Mass Tag (TMT)-based quantitative proteomics reveals potential targets associated with onset of Sub-clinical Mastitis in cows. Sci Rep. 2020;10(1):9321. Epub 20200609. doi: 10.1038/s41598-020-66211-6. PubMed PMID: 32518370; PubMed Central PMCID: PMCPMC7283279.
245. Boudon S, Ounaissi D, Viala D, Monteils V, Picard B, Cassar-Malek I. Label free shotgun proteomics for the identification of protein biomarkers for beef tenderness in muscle and plasma of heifers. J Proteomics. 2020;217:103685. Epub 20200210. doi: 10.1016/j.jprot.2020.103685. PubMed PMID: 32058039.
246. Gagaoua M, Bonnet M, Picard B. Protein array-based approach to evaluate biomarkers of beef tenderness and marbling in cows: Understanding of the underlying mechanisms and prediction. Foods. 2020;9(9). Epub 20200826. doi: 10.3390/foods9091180. PubMed PMID: 32858893; PubMed Central PMCID: PMCPMC7554754.
247. Marrella MA, Biase FH. A multi-omics analysis identifies molecular features associated with heifer fertility in a case-control design including Angus and Holstein cattle. 2022.
248. Zhao X, Yang Y, Sun BF, Zhao YL, Yang YG. FTO and obesity: mechanisms of association. Curr Diab Rep. 2014;14(5):486. Epub 2014/03/15. doi: 10.1007/s11892-014-0486-0. PubMed PMID: 24627050.
249. Gegenfurtner K, Frohlich T, Flenkenthaler F, Kosters M, Fritz S, Desnoes O, et al. Genetic merit for fertility alters the bovine uterine luminal fluid proteomedagger. Biol Reprod. 2020;102(3):730-9. doi: 10.1093/biolre/ioz216. PubMed PMID: 31786596.
250. Moraes JGN, Behura SK, Bishop JV, Hansen TR, Geary TW, Spencer TE. Analysis of the uterine lumen in fertility-classified heifers: II. Proteins and metabolites. Biol Reprod.

2020;102(3):571-87. doi: 10.1093/biolre/ioz197. PubMed PMID: 31616912; PubMed Central PMCID: PMCPMC7331878.
251. D'Occhio MJ, Baruselli PS, Campanile G. Metabolic health, the metabolome and reproduction in female cattle: a review. Ital J Anim Sci. 2019;18(1):858-67. doi: 10.1080/1828051x.2019.1600385. PubMed PMID: WOS:000466726100001.
252. Fontanesi L. Metabolomics and livestock genomics: Insights into a phenotyping frontier and its applications in animal breeding. Animal Frontiers. 2016;6(1):73-9. doi: 10.2527/af.2016-0011. PubMed PMID: WOS:000457256200011.
253. Pantophlet AJ, Roelofsen H, de Vries MP, Gerrits WJJ, van den Borne J, Vonk RJ. The use of metabolic profiling to identify insulin resistance in veal calves. PLoS One. 2017;12(6):e0179612. Epub 20170615. doi: 10.1371/journal.pone.0179612. PubMed PMID: 28617863; PubMed Central PMCID: PMCPMC5472311.
254. Hessock EA, Edwards JL, Schrick FN, Payton RR, Campagna SR, Pollock AB, et al. Metabolite abundance in bovine preovulatory follicular fluid is influenced by follicle developmental progression post estrous onset in cattle. Front Cell Dev Biol. 2023;11. doi: 10.3389/fcell.2023.1156060. PubMed PMID: WOS:001000410600001.
255. Orsi NM, Gopichandran N, Leese HJ, Picton HM, Harris SE. Fluctuations in bovine ovarian follicular fluid composition throughout the oestrous cycle. Reproduction. 2005;129(2):219-28. doi: 10.1530/rep.1.00460. PubMed PMID: 15695616.
256. Bender K, Walsh S, Evans ACO, Fair T, Brennan L. Metabolite concentrations in follicular fluid may explain differences in fertility between heifers and lactating cows. Reproduction. 2010;139(6):1047-55. doi: 10.1530/rep-10-0068. PubMed PMID: WOS:000277781400012.
257. Leroy J, Rizos D, Sturmey R, Bossaert P, Gutierrez-Adan A, Van Hoeck V, et al. Intrafollicular conditions as a major link between maternal metabolism and oocyte quality: a focus on dairy cow fertility. Reprod Fertil Dev. 2012;24(1):1-12. doi: 10.1071/rd11901. PubMed PMID: WOS:000297647800002.
258. Revelli A, Delle Piane L, Casano S, Molinari E, Massobrio M, Rinaudo P. Follicular fluid content and oocyte quality: From single biochemical markers to metabolomics. Reprod Biol Endocrinol. 2009;7:40. Epub 20090504. doi: 10.1186/1477-7827-7-40. PubMed PMID: 19413899; PubMed Central PMCID: PMCPMC2685803.
259. Sinclair KD, Lunn LA, Kwong WY, Wonnacott K, Linforth RST, Craigon J. Amino acid and fatty acid composition of follicular fluid as predictors of in-vitro embryo development. Reprod Biomed Online. 2008;16(6):859-68. doi: 10.1016/s1472-6483(10)60153-8. PubMed PMID: WOS:000256541400015.
260. Matoba S, Bender K, Fahey AG, Mamo S, Brennan L, Lonergan P, et al. Predictive value of bovine follicular components as markers of oocyte developmental potential. Reprod Fertil Dev. 2014;26(2):337-45. doi: 10.1071/rd13007. PubMed PMID: WOS:000329311900010.
261. De Wit AA, Cesar ML, Kruip TA. Effect of urea during in vitro maturation on nuclear maturation and embryo development of bovine cumulus-oocyte-complexes. J Dairy Sci. 2001;84(8):1800-4. doi: 10.3168/jds.S0022-0302(01)74618-8. PubMed PMID: 11518303.
262. Phillips KM, Read CC, Kriese-Anderson LA, Rodning SP, Brandebourg TD, Biase FH, et al. Plasma metabolomic profiles differ at the time of artificial insemination based on pregnancy outcome, in Bos taurus beef heifers. Sci Rep. 2018;8. doi: 10.1038/s41598-018-31605-0. PubMed PMID: WOS:000443543700006.

263. Davis TC, White RR. Breeding animals to feed people: The many roles of animal reproduction in ensuring global food security. Theriogenology. 2020;150:27-33. Epub 2020/02/24. doi: 10.1016/j.theriogenology.2020.01.041. PubMed PMID: 32088028.
264. Cushman RA, Kill LK, Funston RN, Mousel EM, Perry GA. Heifer calving date positively influences calf weaning weights through six parturitions. J Anim Sci. 2013;91(9):4486-91. doi: 10.2527/jas.2013-6465. PubMed PMID: WOS:000323602200046.
265. Heinrichs AJ. Raising dairy replacements to meet the needs of the 21st century. J Dairy Sci. 1993;76(10):3179-87. doi: 10.3168/jds.S0022-0302(93)77656-0. PubMed PMID: 8227639.
266. Hoffman PC. Optimum body size of Holstein replacement heifers. J Anim Sci. 1997;75(3):836-45. PubMed PMID: WOS:A1997WP59300033.
267. Boulton AC, Rushton J, Wathes DC. A study of dairy heifer rearing practices from birth to weaning and their associated costs on UK dairy farms. Open J Anim Sci. 2015;5:185-97.
268. Svensson C, Hultgren J. Associations between housing, management, and morbidity during rearing and subsequent first-lactation milk production of dairy cows in southwest Sweden. J Dairy Sci. 2008;91(4):1510-8. doi: 10.3168/jds.2007-0235. PubMed PMID: WOS:000254123800024.
269. Hultgren J, Svensson C. Calving interval in dairy cows in relation to heifer rearing conditions in Southwest Sweden. Reprod Domest Anim. 2010;45(1):136-41. doi: 10.1111/j.1439-0531.2008.01273.x. PubMed PMID: WOS:000273823300022.
270. Correa MT, Curtis CR, Erb HN, White ME. Effect of calfhood morbidity on age at first calving in New York Holstein herds. Prev Vet Med. 1988;6(4):253-62. doi: https://doi.org/10.1016/0167-5877(88)90037-2.
271. Koltes JE, Cole JB, Clemmens R, Dilger RN, Kramer LM, Lunney JK, et al. A vision for development and utilization of high-throughput phenotyping and big data analytics in livestock. Front Genet. 2019;10:1197. Epub 20191217. doi: 10.3389/fgene.2019.01197. PubMed PMID: 31921279; PubMed Central PMCID: PMCPMC6934059.
272. Iso-Touru T, Sahana G, Guldbrandtsen B, Lund MS, Vilkki J. Genome-wide association analysis of milk yield traits in Nordic Red Cattle using imputed whole genome sequence variants. BMC Genet. 2016;17. doi: ARTN 5510.1186/s12863-016-0363-8. PubMed PMID: WOS:000372737400001.

273. Jiang L, Liu J, Sun D, Ma P, Ding X, Yu Y, et al. Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One. 2010;5(10):e13661. Epub 20101027. doi: 10.1371/journal.pone.0013661. PubMed PMID: 21048968; PubMed Central PMCID: PMCPMC2965099.
274. Raven LA, Cocks BG, Hayes BJ. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics. 2014;15. doi: Artn 6210.1186/1471-2164-15-62. PubMed PMID: WOS:000331115000001.

275. Chen SY, Schenkel FS, Melo ALP, Oliveira HR, Pedrosa VB, Araujo AC, et al. Identifying pleiotropic variants and candidate genes for fertility and reproduction traits in Holstein cattle via association studies based on imputed whole-genome sequence genotypes. BMC Genomics. 2022;23(1):331. Epub 2022/04/29. doi: 10.1186/s12864-022-08555-z. PubMed PMID: 35484513; PubMed Central PMCID: PMCPMC9052698.
276. Mao X, Sahana G, Johansson AM, Liu A, Ismael A, Lovendahl P, et al. Genome-wide association mapping for dominance effects in female fertility using real and simulated data from

Danish Holstein cattle. Sci Rep. 2020;10(1):2953. Epub 20200219. doi: 10.1038/s41598-020-59788-5. PubMed PMID: 32076041; PubMed Central PMCID: PMCPMC7031268.
277. Fang LZ, Sahana G, Su GS, Yu Y, Zhang SL, Lund MS, et al. Integrating sequence-based GWAS and RNA-seq provides novel insights into the genetic basis of mastitis and milk production in dairy cattle. Sci Rep-Uk. 2017;7. doi: ARTN 4556010.1038/srep45560. PubMed PMID: WOS:000398414800001.

278. Freebern E, Santos DJA, Fang L, Jiang J, Parker Gaddis KL, Liu GE, et al. GWAS and fine-mapping of livability and six disease traits in Holstein cattle. BMC Genomics. 2020;21(1):41. Epub 20200113. doi: 10.1186/s12864-020-6461-z. PubMed PMID: 31931710; PubMed Central PMCID: PMCPMC6958677.
279. Neibergs HL, Seabury CM, Wojtowicz AJ, Wang Z, Scraggs E, Kiser JN, et al. Susceptibility loci revealed for bovine respiratory disease complex in pre-weaned holstein calves. BMC Genomics. 2014;15(1):1164. Epub 20141222. doi: 10.1186/1471-2164-15-1164. PubMed PMID: 25534905; PubMed Central PMCID: PMCPMC4445561.
280. Banerjee P, Diniz WJS, Hollingsworth R, Rodning SP, Dyce PW. mRNA signatures in peripheral white blood cells predict reproductive potential in beef heiferes at weaning. Genes. 2023;14(2). doi: https://doi.org/10.3390/genes14020498.
281. Sekimoto T. Extracellular signal-dependent nuclear import of Stat1 is mediated by nuclear pore-targeting complex formation with NPI-1, but not Rch1. The EMBO Journal. 1997;16(23):7067-77. doi: 10.1093/emboj/16.23.7067.
282. Ushijima R, Sakaguchi N, Kano A, Maruyama A, Miyamoto Y, Sekimoto T, et al. Extracellular signal-dependent nuclear import of STAT3 is mediated by various importin αs. Biochem Bioph Res Co. 2005;330(3):880-6. doi: DOI 10.1016/j.bbrc.2005.03.063. PubMed PMID: WOS:000228427200036.
283. Du Y, Bi J, Liu J, Liu X, Wu X, Jiang P, et al. 3Cpro of foot-and-mouth disease virus antagonizes the interferon signaling pathway by blocking STAT1/STAT2 nuclear translocation. J Virol. 2014;88(9):4908-20. Epub 20140219. doi: 10.1128/JVI.03668-13. PubMed PMID: 24554650; PubMed Central PMCID: PMCPMC3993825.
284. Reid SP, Leung LW, Hartman AL, Martinez O, Shaw ML, Carbonnelle C, et al. Ebola virus VP24 binds karyopherin alpha1 and blocks STAT1 nuclear accumulation. J Virol. 2006;80(11):5156-67. doi: 10.1128/JVI.02349-05. PubMed PMID: 16698996; PubMed Central PMCID: PMCPMC1472181.
285. Wang CY, Sun MH, Yuan XH, Ji LF, Jin Y, Cardona CJ, et al. Enterovirus 71 suppresses interferon responses by blocking Janus kinase (JAK)/signal transducer and activator of transcription (STAT) signaling through inducing karyopherin-α1 degradation. J Biol Chem. 2017;292(24):10262-74. doi: 10.1074/jbc.M116.745729. PubMed PMID: WOS:000403580600034.
286. Wang R, Nan Y, Yu Y, Zhang YJ. Porcine reproductive and respiratory syndrome virus Nsp1beta inhibits interferon-activated JAK/STAT signal transduction by inducing karyopherin-alpha1 degradation. J Virol. 2013;87(9):5219-28. Epub 20130228. doi: 10.1128/JVI.02643-12. PubMed PMID: 23449802; PubMed Central PMCID: PMCPMC3624296.
287. Hasankhani A, Bahrami A, Sheybani N, Fatehi F, Abadeh R, Ghaem Maghami Farahani H, et al. Integrated network analysis to identify key modules and potential hub genes involved in Bovine Respiratory Disease: A systems biology approach. Front Genet. 2021;12:753839. Epub 20211018. doi: 10.3389/fgene.2021.753839. PubMed PMID: 34733317; PubMed Central PMCID: PMCPMC8559434.

288. Sun HZ, Srithayakumar V, Jiminez J, Jin W, Hosseini A, Raszek M, et al. Longitudinal blood transcriptomic analysis to identify molecular regulatory patterns of bovine respiratory disease in beef cattle. Genomics. 2020;112(6):3968-77. Epub 20200708. doi: 10.1016/j.ygeno.2020.07.014. PubMed PMID: 32650099.
289. Wang J, Zhou XM, Pan B, Wang HN, Shi FS, Gan WQ, et al. Expression pattern of interferon-inducible transcriptional genes in neutrophils during Bovine Tuberculosis infection. DNA Cell Biol. 2013;32(8):480-6. doi: 10.1089/dna.2012.1941. PubMed PMID: WOS:000322315800008.
290. Bianchi E, Sun Y, Almansa-Ordonez A, Woods M, Goulding D, Martinez-Martin N, et al. Control of oviductal fluid flow by the G-protein coupled receptor Adgrd1 is essential for murine embryo transit. Nat Commun. 2021;12(1). doi: ARTN 125110.1038/s41467-021-21512-w. PubMed PMID: WOS:000623781900025.

291. Tubau-Juni N, Bassaganya-Riera J, Leber A, Zoccoli-Rodriguez V, Kronsteiner B, Viladomiu M, et al. Identification of new regulatory genes through expression pattern analysis of a global RNA-seq dataset from a co-culture system. Sci Rep-Uk. 2020;10(1). doi: ARTN 1150610.1038/s41598-020-68439-8. PubMed PMID: WOS:000550580900023.

292. Basak T, Ain R. Molecular regulation of trophoblast stem cell self-renewal and giant cell differentiation by the Hippo components YAP and LATS1. Stem Cell Res Ther. 2022;13(1):189. Epub 20220507. doi: 10.1186/s13287-022-02844-w. PubMed PMID: 35526072; PubMed Central PMCID: PMCPMC9080189.
293. Duan X, Zhang HL, Wu LL, Liu MY, Pan MH, Ou XH, et al. Involvement of LIMK1/2 in actin assembly during mouse embryo development. Cell Cycle. 2018;17(11):1381-9. Epub 20180725. doi: 10.1080/15384101.2018.1482138. PubMed PMID: 29943641; PubMed Central PMCID: PMCPMC6110583.
294. Duan X, Zhang Y, Chen KL, Zhang HL, Wu LL, Liu HL, et al. The small GTPase RhoA regulates the LIMK1/2-cofilin pathway to modulate cytoskeletal dynamics in oocyte meiosis. J Cell Physiol. 2018;233(8):6088-97. doi: 10.1002/jcp.26450. PubMed PMID: WOS:000430797600052.
295. Kwon J, Seong MJ, Piao X, Jo YJ, Kim NH. LIMK1/2 are required for actin filament and cell junction assembly in porcine embryos developing in vitro. Asian-Australas J Anim Sci. 2020;33(10):1579-89. Epub 20200113. doi: 10.5713/ajas.19.0744. PubMed PMID: 32054159; PubMed Central PMCID: PMCPMC7463081.
296. Chen C, Lodish HF. Global analysis of induced transcription factors and cofactors identifies Tfdp2 as an essential coregulator during terminal erythropoiesis. Exp Hematol. 2014;42(6):464-76. doi: 10.1016/j.exphem.2014.03.001. PubMed PMID: WOS:000338482100006.
297. Hitchens MR, Robbins PD. The role of the transcription factor DP in apoptosis. Apoptosis. 2003;8(5):461-8. doi: 10.1023/a:1025586207239. PubMed PMID: 12975577.
298. Zhu M, Li X, Sun R, Shi P, Cao A, Zhang L, et al. The C/EBPbeta-dependent induction of TFDP2 facilitates Porcine Reproductive and Respiratory Syndrome Virus proliferation. Virol Sin. 2021;36(6):1341-51. Epub 20210617. doi: 10.1007/s12250-021-00403-w. PubMed PMID: 34138404; PubMed Central PMCID: PMCPMC8209777.
299. Bandara LR, Buck VM, Zamanian M, Johnston LH, Lathangue NB. Functional synergy between Dp-1 and E2f-1 in the cell cycle-regulating transcription factor Drtf1/E2f. Embo J.

1993;12(11):4317-24. doi: DOI 10.1002/j.1460-2075.1993.tb06116.x. PubMed PMID: WOS:A1993MC97400029.
300. Helin K, Wu CL, Fattaey AR, Lees JA, Dynlacht BD, Ngwu C, et al. Heterodimerization of the transcription factors E2F-1 and DP-1 leads to cooperative trans-activation. Genes Dev. 1993;7(10):1850-61. doi: 10.1101/gad.7.10.1850. PubMed PMID: 8405995.
301. Krek W, Livingston DM, Shirodkar S. Binding to DNA and the retinoblastoma gene-product promoted by complex-formation of different E2f family members. Science. 1993;262(5139):1557-60. doi: DOI 10.1126/science.8248803. PubMed PMID: WOS:A1993MK32900036.
302. Magae J, Wu CL, Illenye S, Harlow E, Heintz NH. Nuclear localization of DP and E2F transcription factors by heterodimeric partners and retinoblastoma protein family members. J Cell Sci. 1996;109 ( Pt 7):1717-26. doi: 10.1242/jcs.109.7.1717. PubMed PMID: 8832394.
303. Wu CL, Zukerberg LR, Ngwu C, Harlow E, Lees JA. In vivo association of E2F and DP family proteins. Mol Cell Biol. 1995;15(5):2536-46. doi: 10.1128/MCB.15.5.2536. PubMed PMID: 7739537; PubMed Central PMCID: PMCPMC230484.
304. Danielian PS, Bender Kim CF, Caron AM, Vasile E, Bronson RT, Lees JA. E2f4 is required for normal development of the airway epithelium. Dev Biol. 2007;305(2):564-76. doi: 10.1016/j.ydbio.2007.02.037.
305. Geiser V, Jones C. Stimulation of bovine herpesvirus-1 productive infection by the adenovirus E1A gene and a cell cycle regulatory gene, E2F-4. J Gen Virol. 2003;84(Pt 4):929-38. doi: 10.1099/vir.0.18915-0. PubMed PMID: 12655094.
306. Horcajo P, Coronado M, Pastor-Fernández I, Collantes-Fernández E, Román LRS, Reyes-Palomares A, et al. Whole-transcriptome analysis reveals virulence-specific pathogen-host interactions at the placenta in bovine neosporosis. Front Immunol. 2023;14. doi: 10.3389/fimmu.2023.1198609. PubMed PMID: WOS:001038200700001.
307. Humbert PO, Rogers C, Ganiatsas S, Landsberg RL, Trimarchi JM, Dandapani S, et al. E2F4 is essential for normal erythrocyte maturation and neonatal viability. Molecular Cell. 2000;6(2):281-91. doi: 10.1016/s1097-2765(00)00029-0.
308. Jiang WJ, Sun MH, Li XH, Lee SH, Heo G, Zhou D, et al. E2F4 regulates cell cycle to mediate embryonic development in pigs. Theriogenology. 2023;196:227-35. Epub 20221103. doi: 10.1016/j.theriogenology.2022.10.040. PubMed PMID: 36427391.
309. Kim J, Yim GW, Lee DW, Kim YT, Lee YJ, Rhee YJ. Knockdown of E2F4 suppresses the growth of ovarian cancer cells through the cell cycle pathway. Int J Clin Exp Pathol. 2021;14(8):866-74. PubMed PMID: WOS:000728384600002.
310. Lee EY, Yuan TL, Danielian PS, West JC, Lees JA. E2F4 cooperates with pRB in the development of extra-embryonic tissues. Dev Biol. 2009;332(1):104-15. Epub 20090509. doi: 10.1016/j.ydbio.2009.05.541. PubMed PMID: 19433082; PubMed Central PMCID: PMCPMC2832217.
311. Wan Z, Zhi N, Wong S, Keyvanfar K, Liu D, Raghavachari N, et al. Human parvovirus B19 causes cell cycle arrest of human erythroid progenitors via deregulation of the E2F family of transcription factors. J Clin Invest. 2010;120(10):3530-44. Epub 20100920. doi: 10.1172/JCI41805. PubMed PMID: 20890043; PubMed Central PMCID: PMCPMC2947219.
312. da Silva RB, Yang MY, Caixeta ES, Castilho AC, Amorim RL, Price CA, et al. Fibroblast growth factor 18 regulates steroidogenesis in fetal bovine ovarian tissue in vitro. Mol Reprod Dev. 2019;86(2):166-74. Epub 20190110. doi: 10.1002/mrd.23091. PubMed PMID: 30625262.

313. Portela VM, Machado M, Buratini J, Jr., Zamberlam G, Amorim RL, Goncalves P, et al. Expression and function of fibroblast growth factor 18 in the ovarian follicle in cattle. Biol Reprod. 2010;83(3):339-46. Epub 20100519. doi: 10.1095/biolreprod.110.084277. PubMed PMID: 20484739.
314. Jolly PD, Tisdall DJ, Heath DA, Lun S, Mcnatty KP. Apoptosis in bovine granulosa-cells in relation to steroid-synthesis, cyclic adenosine-3',5'-monophosphate response to follicle-stimulating-hormone and luteinizing-hormone, and follicular atresia. Biol Reprod. 1994;51(5):934-44. doi: DOI 10.1095/biolreprod51.5.934. PubMed PMID: WOS:A1994PL95500015.
315. Portela VM, Dirandeh E, Guerrero-Netro HM, Zamberlam G, Barreta MH, Goetten AF, et al. The role of fibroblast growth factor-18 in follicular atresia in cattle. Biol Reprod. 2015;92(1):14. Epub 20141119. doi: 10.1095/biolreprod.114.121376. PubMed PMID: 25411391.
316. Yang MY, Rajamahendran R. Involvement of apoptosis in the atresia of nonovulatory dominant follicle during the bovine estrous cycle. Biol Reprod. 2000;63(5):1313-21. doi: 10.1095/biolreprod63.5.1313. PubMed PMID: 11058534.
317. Guillemot F, Zimmer C. From Cradle to Grave: The multiple roles of fibroblast growth factors in neural development. Neuron. 2011;71(4):574-88. doi: 10.1016/j.neuron.2011.08.002. PubMed PMID: WOS:000294521600005.
318. Hung IH, Schoenwolf GC, Lewandoski M, Ornitz DM. A combined series of Fgf9 and Fgf18 mutant alleles identifies unique and redundant roles in skeletal development. Dev Biol. 2016;411(1):72-84. Epub 20160116. doi: 10.1016/j.ydbio.2016.01.008. PubMed PMID: 26794256; PubMed Central PMCID: PMCPMC4801039.
319. Mok GF, Cardenas R, Anderton H, Campbell KHS, Sweetman D. Interactions between FGF18 and retinoic acid regulate differentiation of chick embryo limb myoblasts. Dev Biol. 2014;396(2):214-23. doi: 10.1016/j.ydbio.2014.10.004. PubMed PMID: WOS:000346228100005.
320. Ohbayashi N, Shibayama M, Kurotaki Y, Imanishi M, Fujimori T, Itoh N, et al. FGF18 is required for normal cell proliferation and differentiation during osteogenesis and chondrogenesis. Gene Dev. 2002;16(7):870-9. doi: 10.1101/gad.965702. PubMed PMID: WOS:000174971800011.
321. Usui H, Shibayama M, Ohbayashi N, Konishi M, Takada S, Itoh N. Fgf18 is required for embryonic lung alveolar development. Biochem Bioph Res Co. 2004;322(3):887-92. doi: 10.1016/j.bbrc.2004.07.198. PubMed PMID: WOS:000223710500026.
322. Zhao XL, Brade T, Cunningham TJ, Duester G. Retinoic acid controls expression of tissue remodeling genes and at the digit-interdigit junction. Dev Dynam. 2010;239(2):665-71. doi: 10.1002/dvdy.22188. PubMed PMID: WOS:000274819300026.

323. Ihanka R, Gentleman R. R: A language for data analysis and graphics. J Comput Graph Stat. 1996;5(3):299-314. doi: https://doi.org/10.2307/1390807.
324. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75. Epub 2007/08/19. doi: 10.1086/519795. PubMed PMID: 17701901; PubMed Central PMCID: PMCPMC1950838.
325. Wickham H. ggplot2: Elegant graphics for data analysis. Use R. 2009:1-212. doi: 10.1007/978-0-387-98141-3. PubMed PMID: WOS:000269437100014.
326. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 2006;34(Database

issue):D590-8. Epub 2005/12/31. doi: 10.1093/nar/gkj144. PubMed PMID: 16381938; PubMed Central PMCID: PMCPMC1347506.
327. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661-78. Epub 2007/06/08. doi: 10.1038/nature05911. PubMed PMID: 17554300; PubMed Central PMCID: PMCPMC2719288.
328. Neupane M, Geary TW, Kiser JN, Burns GW, Hansen PJ, Spencer TE, et al. Loci and pathways associated with uterine capacity for pregnancy and fertility in beef cattle. PLoS One. 2017;12(12):e0188997. Epub 2017/12/12. doi: 10.1371/journal.pone.0188997. PubMed PMID: 29228019; PubMed Central PMCID: PMCPMC5724891.
329. Stegemiller MR, Murdoch GK, Rowan TN, Davenport KM, Becker GM, Hall JB, et al. Genome-wide association analyses of fertility traits in beef heifers. Genes (Basel). 2021;12(2). Epub 2021/02/06. doi: 10.3390/genes12020217. PubMed PMID: 33540904; PubMed Central PMCID: PMCPMC7913221.
330. Rajagopal VM. ggman: R package for Manhattan plots 2017. Available from: https://github.com/drveera/ggman.
331. Shi H, Kichaev G, Pasaniuc B. Contrasting the genetic architecture of 30 complex traits from summary association data. Am J Hum Genet. 2016;99(1):139-53. Epub 2016/06/28. doi: 10.1016/j.ajhg.2016.05.013. PubMed PMID: 27346688; PubMed Central PMCID: PMCPMC5005444.
332. Shi H, Burch KS, Johnson R, Freund MK, Kichaev G, Mancuso N, et al. Localizing components of shared transethnic genetic architecture of complex traits from GWAS summary data. Am J Hum Genet. 2020;106(6):805-17. Epub 2020/05/23. doi: 10.1016/j.ajhg.2020.04.012. PubMed PMID: 32442408; PubMed Central PMCID: PMCPMC7273527.
333. Goddard ME, Kemper KE, MacLeod IM, Chamberlain AJ, Hayes BJ. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. Proc Biol Sci. 2016;283(1835). Epub 2016/07/22. doi: 10.1098/rspb.2016.0569. PubMed PMID: 27440663; PubMed Central PMCID: PMCPMC4971198.
334. Eyre-Walker A. Evolution in health and medicine Sackler colloquium: Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. P Natl Acad Sci USA. 2010;107 Suppl 1(Suppl 1):1752-6. Epub 2010/02/27. doi: 10.1073/pnas.0906182107. PubMed PMID: 20133822; PubMed Central PMCID: PMCPMC2868283.
335. Watanabe K, Stringer S, Frei O, Umicevic Mirkov M, de Leeuw C, Polderman TJC, et al. A global overview of pleiotropy and genetic architecture in complex traits. Nat Genet. 2019;51(9):1339-48. Epub 2019/08/21. doi: 10.1038/s41588-019-0481-0. PubMed PMID: 31427789.
336. Williams RB, Chan EK, Cowley MJ, Little PF. The influence of genetic variation on gene expression. Genome res. 2007;17(12):1707-16. Epub 2007/12/08. doi: 10.1101/gr.6981507. PubMed PMID: 18063559.
337. Kreitmaier P, Katsoula G, Zeggini E. Insights from multi-omics integration in complex disease primary tissues. Trends Genet. 2023;39(1):46-58. Epub 2022/09/23. doi: 10.1016/j.tig.2022.08.005. PubMed PMID: 36137835.
338. Consortium GT. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369(6509):1318-30. Epub 2020/09/12. doi: 10.1126/science.aaz1776. PubMed PMID: 32913098; PubMed Central PMCID: PMCPMC7737656.

339. Kim-Hellmuth S, Aguet F, Oliva M, Munoz-Aguirre M, Kasela S, Wucher V, et al. Cell type-specific genetic regulation of gene expression across human tissues. Science. 2020;369(6509). Epub 2020/09/12. doi: 10.1126/science.aaz8528. PubMed PMID: 32913075; PubMed Central PMCID: PMCPMC8051643.
340. Liu S, Gao Y, Canela-Xandri O, Wang S, Yu Y, Cai W, et al. A multi-tissue atlas of regulatory variants in cattle. Nat Genet. 2022;54(9):1438-47. Epub 2022/08/12. doi: 10.1038/s41588-022-01153-5. PubMed PMID: 35953587.
341. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348(6235):648-60. Epub 2015/05/09. doi: 10.1126/science.1262110. PubMed PMID: 25954001; PubMed Central PMCID: PMCPMC4547484.
342. Gregersen PK. Closing the gap between genotype and phenotype. Nat Genet. 2009;41(9):958-9. Epub 2009/08/28. doi: 10.1038/ng0909-958. PubMed PMID: 19710714.
343. Dendrou CA, Plagnol V, Fung E, Yang JH, Downes K, Cooper JD, et al. Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable human bioresource. Nat Genet. 2009;41(9):1011-5. Epub 2009/08/25. doi: 10.1038/ng.434. PubMed PMID: 19701192; PubMed Central PMCID: PMCPMC2749506.
344. Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. Philos Trans R Soc Lond B Biol Sci. 2013;368(1620):20120362. Epub 2013/05/08. doi: 10.1098/rstb.2012.0362. PubMed PMID: 23650636; PubMed Central PMCID: PMCPMC3682727.
345. Kendziorski CM, Chen M, Yuan M, Lan H, Attie AD. Statistical methods for expression quantitative trait loci (eQTL) mapping. Biometrics. 2006;62(1):19-27. Epub 2006/03/18. doi: 10.1111/j.1541-0420.2005.00437.x. PubMed PMID: 16542225.
346. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11:R25. doi: 10.1186/gb-2010-11-3-r25. PubMed PMID: 20196867.
347. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5:621-8. doi: 10.1038/nmeth.1226. PubMed PMID: 18516045.
348. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323. Epub 2011/08/06. doi: 10.1186/1471-2105-12-323. PubMed PMID: 21816040; PubMed Central PMCID: PMCPMC3163565.
349. Yang J, Wang D, Yang Y, Yang W, Jin W, Niu X, et al. A systematic comparison of normalization methods for eQTL analysis. Brief Bioinform. 2021;22(6). Epub 2021/05/21. doi: 10.1093/bib/bbab193. PubMed PMID: 34015824.
350. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. doi: 10.1186/PREACCEPT-8897612761307401. PubMed PMID: 25516281.
351. Mason VC, Schaefer RJ, McCue ME, Leeb T, Gerber V. eQTL discovery and their association with severe equine asthma in European Warmblood horses. BMC Genomics. 2018;19(1):581. Epub 2018/08/04. doi: 10.1186/s12864-018-4938-9. PubMed PMID: 30071827; PubMed Central PMCID: PMCPMC6090848.
352. Zeng B, Lloyd-Jones LR, Montgomery GW, Metspalu A, Esko T, Franke L, et al. Comprehensive multiple eQTL detection and its application to GWAS interpretation. Genetics.

2019;212(3):905-18. Epub 2019/05/28. doi: 10.1534/genetics.119.302091. PubMed PMID: 31123039; PubMed Central PMCID: PMCPMC6614888.
353. Strunz T, Grassmann F, Gayan J, Nahkuri S, Souza-Costa D, Maugeais C, et al. A mega-analysis of expression quantitative trait loci (eQTL) provides insight into the regulatory architecture of gene expression variation in liver. Sci Rep. 2018;8(1):5865. Epub 2018/04/14. doi: 10.1038/s41598-018-24219-z. PubMed PMID: 29650998; PubMed Central PMCID: PMCPMC5897392.
354. Albert FW, Bloom JS, Siegel J, Day L, Kruglyak L. Genetics of trans-regulatory variation in gene expression. Elife. 2018;7. Epub 2018/07/18. doi: 10.7554/eLife.35471. PubMed PMID: 30014850; PubMed Central PMCID: PMCPMC6072440.
355. Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, Kolberg L, et al. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. Nat Genet. 2021;53(9):1290-9. Epub 2021/09/09. doi: 10.1038/s41588-021-00924-w. PubMed PMID: 34493866; PubMed Central PMCID: PMCPMC8423625.
356. Beasley TM, Erickson S, Allison DB. Rank-based inverse normal transformations are increasingly used, but are they merited? Behav Genet. 2009;39(5):580-95. Epub 2009/06/16. doi: 10.1007/s10519-009-9281-0. PubMed PMID: 19526352; PubMed Central PMCID: PMCPMC2921808.
357. McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139-40. doi: 10.1093/bioinformatics/btp616. PubMed PMID: 19910308.
358. Hardcastle TJ, Kelly KA. baySeq: empirical bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics. 2010;11:422. Epub 2010/08/12. doi: 10.1186/1471-2105-11-422. PubMed PMID: 20698981; PubMed Central PMCID: PMCPMC2928208.
359. Robinson MD, Smyth GK. Moderated statistical tests for assessing differences in tag abundance. Bioinformatics. 2007;23(21):2881-7. Epub 2007/09/21. doi: 10.1093/bioinformatics/btm453. PubMed PMID: 17881408.
360. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106. Epub 2010/10/29. doi: 10.1186/gb-2010-11-10-r106. PubMed PMID: 20979621; PubMed Central PMCID: PMCPMC3218662.
361. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5:R80. doi: 10.1186/gb-2004-5-10-r80. PubMed PMID: 15461798.
362. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012;40:4288-97. doi: 10.1093/nar/gks042. PubMed PMID: 22287627.
363. Dickinson SE, Biase FH. Transcriptome data of peripheral white blood cells from beef heifers collected at the time of artificial insemination. Data Brief. 2018;18:706-9. Epub 2018/06/15. doi: 10.1016/j.dib.2018.03.062. PubMed PMID: 29900224; PubMed Central PMCID: PMCPMC5996294.
364. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20. Epub 2014/04/04. doi: 10.1093/bioinformatics/btu170. PubMed PMID: 24695404; PubMed Central PMCID: PMCPMC4103590.

365. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907-15. Epub 2019/08/04. doi: 10.1038/s41587-019-0201-4. PubMed PMID: 31375807.
366. Elsik CG, Tellam RL, Worley KC, Gibbs RA, Muzny DM, Weinstock GM, et al. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science. 2009;324(5926):522-8. doi: 10.1126/science.1169588.
367. Rosen BD, Bickhart DM, Schnabel RD, Koren S, Elsik CG, Tseng E, et al. De novo assembly of the cattle reference genome with single-molecule sequencing. GigaScience. 2020;9(3). doi: https://doi.org/10.1093/gigascience/giaa021.
368. Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2014. 2014;42(Database issue:D749-D755).
369. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimatin from sequencing data. Bioinformatics. 2011;27(21):2987-93.
370. Tischler G, Leonard S. biobambam: tools for read pair collation based algorithms on BAM files. Source Code Biol Med. 2014;9(1). doi: Artn 1310.1186/1751-0473-9-13. PubMed PMID: WOS:000215862300013.

371. Auwera GAVd, O'Connor BD. Genomics in the cloud: Using Docker, GATK, and WDL in Terra (1st Edition): O'Reilly Media; 2020.
372. Ihaka R, Gentleman R. R: A language for data analysis and graphics. J Comput Graph Stat. 1996;5(3):299-314. doi: {ISBN} 3-900051-07-0.
373. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17(1):122. doi: 10.1186/s13059-016-0974-4.
374. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. 2012;131(4):281-5. Epub 2012/08/09. doi: 10.1007/s12064-012-0162-3. PubMed PMID: 22872506.
375. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904-9. Epub 2006/07/25. doi: 10.1038/ng1847. PubMed PMID: 16862161.
376. Huang QQ, Ritchie SC, Brozynska M, Inouye M. Power, false discovery rate and Winner's Curse in eQTL studies. Nucleic Acids Res. 2018;46(22):e133. Epub 2018/09/07. doi: 10.1093/nar/gky780. PubMed PMID: 30189032; PubMed Central PMCID: PMCPMC6294523.
377. Graffelman J. Exploring diallelic genetic markers: The HardyWeinberg package. J Stat Softw. 2015;64(3):1-23. PubMed PMID: WOS:000352911000001.
378. Shabalin AA. Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28(10).
379. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 2014;46(11):1173-86. Epub 2014/10/06. doi: 10.1038/ng.3097. PubMed PMID: 25282103; PubMed Central PMCID: PMCPMC4250049.
380. Vicente CT, Revez JA, Ferreira MAR. Lessons from ten years of genome-wide association studies of asthma. Clin Transl Immunology. 2017;6(12):e165. Epub 2018/01/16. doi: 10.1038/cti.2017.54. PubMed PMID: 29333270; PubMed Central PMCID: PMCPMC5750453.
381. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. Genetic Epidemiology. 2008;32(3):227-34. doi: 10.1002/gepi.20297. PubMed PMID: WOS:000254690500004.

382. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol. 2008;32(4):381-5. Epub 2008/03/19. doi: 10.1002/gepi.20303. PubMed PMID: 18348202.
383. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P, et al. A haplotype map of the human genome. Nature. 2005;437(7063):1299-320. doi: 10.1038/nature04226. PubMed PMID: WOS:000232829100044.
384. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. J Roy Stat Soc B Met. 1995;57(1):289-300. PubMed PMID: WOS:A1995QE45300017.
385. Lund SP, Nettleton D, McCarthy DJ, Smyth GK. Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. Stat Appl Genet Mol Biol. 2012;11(5). doi: Artn 810.1515/1544-6115.1826. PubMed PMID: WOS:000310129100009.
386. Lun ATL, Chen YS, Smyth GK. It's DE-licious: A recipe for differential expression analyses of RNA-seq experiments using quasi-likelihood methods in edgeR. Methods Mol Biol. 2016;1418:391-416. doi: 10.1007/978-1-4939-3578-9_19. PubMed PMID: WOS:000376532300020.
387. Horita N, Kaneko T. Genetic model selection for a case-control study and a meta-analysis. Meta Gene. 2015;5:1-8. Epub 2015/06/05. doi: 10.1016/j.mgene.2015.04.003. PubMed PMID: 26042205; PubMed Central PMCID: PMCPMC4443430.
388. Wilke CO. cowplot: Streamlined plot theme and plot annotations for 'ggplot2 2020. Available from: https://wilkelab.org/cowplot/.
389. Sievert C. Interactive web-based data visualization with R, plotly, and shiny: Chapman and Hall/CRC; 2020.
390. Wickham H. ggplot2: Elegant graphics for data analysis. New York: Springer-Vergag; 2009.
391. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome res. 2003;13(11):2498-504. doi: 10.1101/gr.1239303. PubMed PMID: 14597658; PubMed Central PMCID: PMC403769.
392. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11.
393. Holm S. A simple sequentially rejective multiple test procedure. Scand Stat Theory Appl. 1979;6(2):65–70.
394. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. Genome Biol. 2016;17(1):122. Epub 2016/06/09. doi: 10.1186/s13059-016-0974-4. PubMed PMID: 27268795; PubMed Central PMCID: PMCPMC4893825.
395. Hunt SE, Moore B, Amode RM, Armean IM, Lemos D, Mushtaq A, et al. Annotating and prioritizing genomic variants using the Ensembl Variant Effect Predictor-A tutorial. Hum Mutat. 2021. Epub 2021/11/25. doi: 10.1002/humu.24298. PubMed PMID: 34816521.
396. Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome res. 1999;9(8):677-9. Epub 1999/08/14. PubMed PMID: 10447503.
397. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28(10):1353-8. Epub 2012/04/12. doi: 10.1093/bioinformatics/bts163. PubMed PMID: 22492648; PubMed Central PMCID: PMCPMC3348564.

398. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25-9. Epub 2000/05/10. doi: 10.1038/75556. PubMed PMID: 10802651; PubMed Central PMCID: PMCPMC3037419.
399. Noel-MacDonnell JR, Usset J, Goode EL, Fridley BL. Assessment of data transformations for model-based clustering of RNA-Seq data. PLoS One. 2018;13(2):e0191758. Epub 2018/02/28. doi: 10.1371/journal.pone.0191758. PubMed PMID: 29485993; PubMed Central PMCID: PMCPMC5828440.
400. Zwiener I, Frisch B, Binder H. Transforming RNA-Seq data to improve the performance of prognostic gene signatures. PLoS One. 2014;9(1):e85150. Epub 2014/01/15. doi: 10.1371/journal.pone.0085150. PubMed PMID: 24416353; PubMed Central PMCID: PMCPMC3885686.
401. Jain D, Hodonsky CJ, Schick UM, Morrison JV, Minnerath S, Brown L, et al. Genome-wide association of white blood cell counts in Hispanic/Latino Americans: the Hispanic Community Health Study/Study of Latinos. Hum Mol Genet. 2017;26(6):1193-204. Epub 2017/02/06. doi: 10.1093/hmg/ddx024. PubMed PMID: 28158719; PubMed Central PMCID: PMCPMC5968624.
402. Jehl F, Degalez F, Bernard M, Lecerf F, Lagoutte L, Desert C, et al. RNA-seq data for reliable SNP detection and genotype calling: Interest for coding variant characterization and cis-regulation analysis by allele-specific expression in livestock species. Front Genet. 2021;12:655707. Epub 2021/07/16. doi: 10.3389/fgene.2021.655707. PubMed PMID: 34262593; PubMed Central PMCID: PMCPMC8273700.
403. Lam S, Zeidan J, Miglior F, Suarez-Vega A, Gomez-Redondo I, Fonseca PAS, et al. Development and comparison of RNA-sequencing pipelines for more accurate SNP identification: practical example of functional SNP detection associated with feed efficiency in Nellore beef cattle. BMC Genomics. 2020;21(1):703. Epub 2020/10/10. doi: 10.1186/s12864-020-07107-7. PubMed PMID: 33032519; PubMed Central PMCID: PMCPMC7545862.
404. Lam S, Miglior F, Fonseca PAS, Gomez-Redondo I, Zeidan J, Suarez-Vega A, et al. Identification of functional candidate variants and genes for feed efficiency in Holstein and Jersey cattle breeds using RNA-sequencing. J Dairy Sci. 2021;104(2):1928-50. Epub 2020/12/29. doi: 10.3168/jds.2020-18241. PubMed PMID: 33358171.
405. Bakhtiarizadeh MR, Alamouti AA. RNA-Seq based genetic variant discovery provides new insights into controlling fat deposition in the tail of sheep. Sci Rep. 2020;10(1):13525. Epub 2020/08/13. doi: 10.1038/s41598-020-70527-8. PubMed PMID: 32782325; PubMed Central PMCID: PMCPMC7419499.
406. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491-8. Epub 2011/04/12. doi: 10.1038/ng.806. PubMed PMID: 21478889; PubMed Central PMCID: PMCPMC3083463.
407. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2). Epub 2021/02/17. doi: 10.1093/gigascience/giab008. PubMed PMID: 33590861; PubMed Central PMCID: PMCPMC7931819.
408. Brouard JS, Schenkel F, Marete A, Bissonnette N. The GATK joint genotyping workflow is appropriate for calling variants in RNA-seq experiments. J Anim Sci Biotechnol. 2019;10:44.

Epub 2019/06/30. doi: 10.1186/s40104-019-0359-0. PubMed PMID: 31249686; PubMed Central PMCID: PMCPMC6587293.
409. Nodzak C. Introductory methods for eQTL analyses. Methods Mol Biol. 2020;2082:3-14. Epub 2019/12/19. doi: 10.1007/978-1-0716-0026-9_1. PubMed PMID: 31849004.
410. van den Berg I, Hayes BJ, Chamberlain AJ, Goddard ME. Overlap between eQTL and QTL associated with production traits and fertility in dairy cattle. BMC Genomics. 2019;20(1):291. Epub 2019/04/17. doi: 10.1186/s12864-019-5656-7. PubMed PMID: 30987590; PubMed Central PMCID: PMCPMC6466667.
411. Lee YL, Takeda H, Costa Monteiro Moreira G, Karim L, Mullaart E, Coppieters W, et al. A 12 kb multi-allelic copy number variation encompassing a GC gene enhancer is associated with mastitis resistance in dairy cattle. Plos Genet. 2021;17(7):e1009331. Epub 2021/07/22. doi: 10.1371/journal.pgen.1009331. PubMed PMID: 34288907; PubMed Central PMCID: PMCPMC8328317.
412. Fang L, Cai W, Liu S, Canela-Xandri O, Gao Y, Jiang J, et al. Comprehensive analyses of 723 transcriptomes enhance genetic and biological interpretations for complex traits in cattle. Genome res. 2020;30(5):790-801. Epub 2020/05/20. doi: 10.1101/gr.250704.119. PubMed PMID: 32424068; PubMed Central PMCID: PMCPMC7263193.
413. Canive M, Fernandez-Jimenez N, Casais R, Vazquez P, Lavin JL, Bilbao JR, et al. Identification of loci associated with susceptibility to bovine paratuberculosis and with the dysregulation of the MECOM, eEF1A2, and U1 spliceosomal RNA expression. Sci Rep. 2021;11(1):313. Epub 2021/01/13. doi: 10.1038/s41598-020-79619-x. PubMed PMID: 33432064; PubMed Central PMCID: PMCPMC7801378.
414. Gjuvsland AB, Plahte E, Adnoy T, Omholt SW. Allele interaction--single locus genetics meets regulatory biology. PLoS One. 2010;5(2):e9379. Epub 2010/02/27. doi: 10.1371/journal.pone.0009379. PubMed PMID: 20186347; PubMed Central PMCID: PMCPMC2826424.
415. Elston RC, Satagopan JM, Sun S. Genetic terminology. Methods Mol Biol. 2012;850:1-9. Epub 2012/02/07. doi: 10.1007/978-1-61779-555-8_1. PubMed PMID: 22307690; PubMed Central PMCID: PMCPMC4450815.
416. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, et al. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc. 2013;8:1765-86. doi: 10.1038/nprot.2013.099. PubMed PMID: 23975260.
417. Hanson C, Cairns J, Wang L, Sinha S. Principled multi-omic analysis reveals gene regulatory mechanisms of phenotype variation. Genome research. 2018;28(8):1207-16. Epub 2018/06/15. doi: 10.1101/gr.227066.117. PubMed PMID: 29898900; PubMed Central PMCID: PMCPMC6071639.
418. Doss S, Schadt EE, Drake TA, Lusis AJ. Cis-acting expression quantitative trait loci in mice. Genome research. 2005;15(5):681-91. Epub 2005/04/20. doi: 10.1101/gr.3216905. PubMed PMID: 15837804; PubMed Central PMCID: PMCPMC1088296.
419. Loguercio S, Overall RW, Michaelson JJ, Wiltshire T, Pletcher MT, Miller BH, et al. Integrative analysis of low- and high-resolution eQTL. PLoS One. 2010;5(11):e13920. Epub 2010/11/19. doi: 10.1371/journal.pone.0013920. PubMed PMID: 21085707; PubMed Central PMCID: PMCPMC2978079.
420. Goring HH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole SA, et al. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nat Genet. 2007;39(10):1208-16. Epub 2007/09/18. doi: 10.1038/ng2119. PubMed PMID: 17873875.

421. Wilkie AO. The molecular basis of genetic dominance. J Med Genet. 1994;31(2):89-98. Epub 1994/02/01. doi: 10.1136/jmg.31.2.89. PubMed PMID: 8182727; PubMed Central PMCID: PMCPMC1049666.
422. Kacser H, Burns JA. The molecular basis of dominance. Genetics. 1981;97(3-4):639-66. Epub 1981/03/01. doi: 10.1093/genetics/97.3-4.639. PubMed PMID: 7297851; PubMed Central PMCID: PMCPMC1214416.
423. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics. 2016;32(10):1479-85. Epub 2015/12/29. doi: 10.1093/bioinformatics/btv722. PubMed PMID: 26708335; PubMed Central PMCID: PMCPMC4866519.
424. Henchion M, Hayes M, Mullen AM, Fenelon M, Tiwari B. Future protein supply and demand: strategies and factors influencing a sustainable equilibrium. Foods. 2017;6(7). Epub 2017/07/21. doi: 10.3390/foods6070053. PubMed PMID: 28726744; PubMed Central PMCID: PMCPMC5532560.
425. Cushman RA, Kill LK, Funston RN, Mousel EM, Perry GA. Heifer calving date positively influences calf weaning weights through six parturitions. J Anim Sci. 2013;91(9):4486-91. doi: 10.2527/jas2013-6465. PubMed PMID: WOS:000323602200046.
426. Lesmeister JL, Burfening PJ, Blackwell RL. Date of first calving in beef cows and subsequent calf production. J Anim Sci. 1973;36(1):1-6. doi: 10.2527/jas1973.3611.
427. Damiran D, Larson KA, Pearce LT, Erickson NE, Lardner BHA. Effect of calving period on beef cow longevity and lifetime productivity in western Canada. Transl Anim Sci. 2018;2(suppl_1):S61-S5. doi: 10.1093/tas/txy020.
428. Hoffman PC. Optimum body size of Holstein replacement heifers. J Anim Sci. 1997;75(3):836-45. Epub 1997/03/01. doi: 10.2527/1997.753836x. PubMed PMID: 9078504.
429. Peters SO, Kizilkaya K, Garrick DJ, Fernando RL, Reecy JM, Weaber RL, et al. Bayesian genome-wide association analysis of growth and yearling ultrasound measures of carcass traits in Brangus heifers. J Anim Sci. 2012;90(10):3398-409. doi: 10.2527/jas2012-4507.
430. Fortes MRS, Snelling WM, Reverter A, Nagaraj SH, Lehnert SA, Hawken RJ, et al. Gene network analyses of first service conception in Brangus heifers: use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors. J Anim Sci. 2012;90:2894-906. doi: 10.2527/jas.2011-4601. PubMed PMID: 22739780.
431. McAllister CM, Speidel SE, Crews DH, Jr., Enns RM. Genetic parameters for intramuscular fat percentage, marbling score, scrotal circumference, and heifer pregnancy in Red Angus cattle. J Anim Sci. 2011;89(7):2068-72. Epub 2011/02/01. doi: 10.2527/jas.2010-3538. PubMed PMID: 21278121.
432. Boddhireddy P, Kelly MJ, Northcutt S, Prayaga KC, Rumph J, DeNise S. Genomic predictions in Angus cattle: Comparisons of sample size, response variables, and clustering methods for cross-validation. J Anim Sci. 2014;92(2):485-97. doi: 10.2527/jas.2013-6757.
433. Walsh SW, Williams EJ, Evans AC. A review of the causes of poor fertility in high milk producing dairy cows. Anim Reprod Sci. 2011;123(3-4):127-38. Epub 2011/01/25. doi: 10.1016/j.anireprosci.2010.12.001. PubMed PMID: 21255947; PubMed Central PMCID: PMCPMC7125520.
434. McDaneld TG, Kuehn LA, Thomas MG, Snelling WM, Smith TPL, Pollak EJ, et al. Genomewide association study of reproductive efficiency in female cattle. J Anim Sci. 2014;92:1945-57. doi: 10.2527/jas.2012-6807. PubMed PMID: 24782394.

435. McDaneld TG, Kuehn LA, Thomas MG, Snelling WM, Sonstegard TS, Matukumalli LK, et al. Y are you not pregnant: identification of Y chromosome segments in female cattle with decreased reproductive efficiency. J Anim Sci. 2012;90:2142-51. doi: 10.2527/jas.2011-4536. PubMed PMID: 22408089.
436. de Camargo GM, Costa RB, de Albuquerque LG, Regitano LC, Baldi F, Tonhati H. Association between JY-1 gene polymorphisms and reproductive traits in beef cattle. Gene. 2014;533(2):477-80. Epub 2013/10/23. doi: 10.1016/j.gene.2013.09.126. PubMed PMID: 24144840.
437. Dias MM, Souza FR, Takada L, Feitosa FL, Costa RB, Diaz ID, et al. Study of lipid metabolism-related genes as candidate genes of sexual precocity in Nellore cattle. Genet Mol Res. 2015;14(1):234-43. Epub 2015/03/03. doi: 10.4238/2015.January.16.7. PubMed PMID: 25729955.
438. Irano N, de Camargo GM, Costa RB, Terakado AP, Magalhaes AF, Silva RM, et al. Genome-wide association study for indicator traits of sexual precocity in Nellore cattle. PLoS One. 2016;11(8):e0159502. Epub 2016/08/06. doi: 10.1371/journal.pone.0159502. PubMed PMID: 27494397; PubMed Central PMCID: PMCPMC4975395.
439. Junior GAO, Perez BC, Cole JB, Santana MHA, Silveira J, Mazzoni G, et al. Genomic study and medical subject headings enrichment analysis of early pregnancy rate and antral follicle numbers in Nelore heifers. J Anim Sci. 2017;95(11):4796-812. Epub 2018/01/03. doi: 10.2527/jas2017.1752. PubMed PMID: 29293733; PubMed Central PMCID: PMCPMC6292327.
440. Frischknecht M, Bapst B, Seefried FR, Signer-Hasler H, Garrick D, Stricker C, et al. Genome-wide association studies of fertility and calving traits in Brown Swiss cattle using imputed whole-genome sequences. BMC Genomics. 2017;18(1):910. Epub 2017/11/28. doi: 10.1186/s12864-017-4308-z. PubMed PMID: 29178833; PubMed Central PMCID: PMCPMC5702100.
441. Jiang J, Ma L, Prakapenka D, VanRaden PM, Cole JB, Da Y. A large-scale genome-wide association study in U.S. Holstein cattle. Front Genet. 2019;10:412. Epub 2019/05/30. doi: 10.3389/fgene.2019.00412. PubMed PMID: 31139206; PubMed Central PMCID: PMCPMC6527781.
442. Nayeri S, Sargolzaei M, Abo-Ismail MK, May N, Miller SP, Schenkel F, et al. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. BMC Genet. 2016;17(1):75. Epub 2016/06/12. doi: 10.1186/s12863-016-0386-1. PubMed PMID: 27287773; PubMed Central PMCID: PMCPMC4901445.
443. Kiser JN, Keuter EM, Seabury CM, Neupane M, Moraes JGN, Dalton J, et al. Validation of 46 loci associated with female fertility traits in cattle. BMC Genomics. 2019;20(1):576. Epub 2019/07/14. doi: 10.1186/s12864-019-5935-3. PubMed PMID: 31299913; PubMed Central PMCID: PMCPMC6624949.
444. Kiser JN, Clancey E, Moraes JGN, Dalton J, Burns GW, Spencer TE, et al. Identification of loci associated with conception rate in primiparous Holstein cows. BMC Genomics. 2019;20(1):840. Epub 2019/11/14. doi: 10.1186/s12864-019-6203-2. PubMed PMID: 31718557; PubMed Central PMCID: PMCPMC6852976.
445. Phillips KM, Read CC, Kriese-Anderson LA, Rodning SP, Brandebourg TD, Biase FH, et al. Plasma metabolomic profiles differ at the time of artificial insemination based on pregnancy outcome, in Bos taurus beef heifers. Sci Rep. 2018;8(1):13196. Epub 2018/09/06. doi: 10.1038/s41598-018-31605-0. PubMed PMID: 30181662; PubMed Central PMCID: PMCPMC6123494.

446. Pryce JE, Royal MD, Garnsworthy PC, Mao IL. Fertility in the high-producing dairy cow. Livest Prod Sci. 2004;86(1-3):125-35. doi: 10.1016/S0301-6226(03)00145-3. PubMed PMID: WOS:000188949000012.
447. Wathes DC, Fenwick M, Cheng Z, Bourne N, Llewellyn S, Morris DG, et al. Influence of negative energy balance on cyclicity and fertility in the high producing dairy cow. Theriogenology. 2007;68 Suppl 1:S232-41. Epub 2007/05/04. doi: 10.1016/j.theriogenology.2007.04.006. PubMed PMID: 17475319.
448. Diskin MG, Kenny DA. Managing the reproductive performance of beef cows. Theriogenology. 2016;86(1):379-87. Epub 2016/05/18. doi: 10.1016/j.theriogenology.2016.04.052. PubMed PMID: 27180327.
449. Randel RD. Nutrition and postpartum rebreeding in cattle. J Anim Sci. 1990;68(3):853-62. Epub 1990/03/01. doi: 10.2527/1990.683853x. PubMed PMID: 2180880.
450. Breuel KF, Lewis PE, Schrick FN, Lishman AW, Inskeep EK, Butcher RL. Factors affecting fertility in the postpartum cow: role of the oocyte and follicle in conception rate. Biol Reprod. 1993;48(3):655-61. Epub 1993/03/01. doi: 10.1095/biolreprod48.3.655. PubMed PMID: 8452940.
451. Okano A, Tomizuka T. Ultrasonic observation of postpartum uterine involution in the cow. Theriogenology. 1987;27(2):369-76. Epub 1987/02/01. doi: 10.1016/0093-691x(87)90225-1. PubMed PMID: 16726242.
452. Sheldon IM, Williams EJ, Miller AN, Nash DM, Herath S. Uterine diseases in cattle after parturition. Vet J. 2008;176(1):115-21. Epub 2008/03/11. doi: 10.1016/j.tvjl.2007.12.031. PubMed PMID: 18329302; PubMed Central PMCID: PMCPMC2706386.
453. Sheldon IM, Molinari PCC, Ormsby TJR, Bromfield JJ. Preventing postpartum uterine disease in dairy cattle depends on avoiding, tolerating and resisting pathogenic bacteria. Theriogenology. 2020;150:158-65. Epub 2020/01/25. doi: 10.1016/j.theriogenology.2020.01.017. PubMed PMID: 31973964; PubMed Central PMCID: PMCPMC7234917.
454. Sheldon IM, Lewis GS, LeBlanc S, Gilbert RO. Defining postpartum uterine disease in cattle. Theriogenology. 2006;65(8):1516-30. Epub 2005/10/18. doi: 10.1016/j.theriogenology.2005.08.021. PubMed PMID: 16226305.
455. Crowley WF, Jr., Pitteloud N, Seminara S. New genes controlling human reproduction and how you find them. Trans Am Clin Climatol Assoc. 2008;119:29-37; discussion -8. Epub 2008/07/04. PubMed PMID: 18596868; PubMed Central PMCID: PMCPMC2394706.
456. Yatsenko SA, Rajkovic A. Genetics of human female infertility. Biol Reprod. 2019;101(3):549-66. Epub 2019/05/12. doi: 10.1093/biolre/ioz084. PubMed PMID: 31077289; PubMed Central PMCID: PMCPMC8127036.
457. Mann CJ. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. Emerg Med J. 2003;20(1):54-60. Epub 2003/01/21. doi: 10.1136/emj.20.1.54. PubMed PMID: 12533370; PubMed Central PMCID: PMCPMC1726024.
458. Lima FS, Ribeiro ES, Bisinotto RS, Greco LF, Martinez N, Amstalden M, et al. Hormonal manipulations in the 5-day timed artificial insemination protocol to optimize estrous cycle synchrony and fertility in dairy heifers. J Dairy Sci. 2013;96(11):7054-65. Epub 2013/09/10. doi: 10.3168/jds.2013-7093. PubMed PMID: 24011941.
459. Patterson D, Kojima F, Smith M. A review of methods to synchronize estrus in replacement beef heifers and postpartum cows. J Anim Sci. 2003;81(14_suppl_2):E166-E77.
460. Wilson C, Dias NW, Pancini S, Mercadante V, Biase FH. Delayed processing of blood samples impairs the accuracy of mRNA-based biomarkers. Sci Rep. 2022;12(1):8196. Epub

2022/05/18. doi: 10.1038/s41598-022-12178-5. PubMed PMID: 35581252; PubMed Central PMCID: PMCPMC9113984.
461. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc. 2010;5(9):1564-73. Epub 2010/11/19. doi: 10.1038/nprot.2010.116. PubMed PMID: 21085122; PubMed Central PMCID: PMCPMC3025522.
462. Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2014. Nucleic Acids Res. 2014;42(Database issue):D749-55. Epub 2013/12/10. doi: 10.1093/nar/gkt1196. PubMed PMID: 24316576; PubMed Central PMCID: PMCPMC3964975.
463. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016;11(9):1650-67. Epub 2016/08/26. doi: 10.1038/nprot.2016.095. PubMed PMID: 27560171; PubMed Central PMCID: PMCPMC5032908.
464. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357-60. Epub 2015/03/10. doi: 10.1038/nmeth.3317. PubMed PMID: 25751142; PubMed Central PMCID: PMCPMC4655817.
465. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. Epub 2009/06/10. doi: 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PubMed Central PMCID: PMCPMC2723002.
466. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-30. Epub 2013/11/15. doi: 10.1093/bioinformatics/btt656. PubMed PMID: 24227677.
467. Orsburn BC. Proteome Discoverer-a community enhanced data processing suite for protein informatics. Proteomes. 2021;9(1). Epub 2021/04/04. doi: 10.3390/proteomes9010015. PubMed PMID: 33806881; PubMed Central PMCID: PMCPMC8006021.
468. Che R, Jack JR, Motsinger-Reif AA, Brown CC. An adaptive permutation approach for genome-wide association study: evaluation and recommendations for use. BioData Min. 2014;7:9. Epub 2014/07/01. doi: 10.1186/1756-0381-7-9. PubMed PMID: 24976866; PubMed Central PMCID: PMCPMC4070098.
469. Storey JD, Tibshirani R. Statistical significance for genomewide studies. P Natl Acad Sci USA. 2003;100(16):9440-5. Epub 2003/07/29. doi: 10.1073/pnas.1530509100. PubMed PMID: 12883005; PubMed Central PMCID: PMCPMC170937.
470. Nakayasu ES, Gritsenko M, Piehowski PD, Gao Y, Orton DJ, Schepmoes AA, et al. Tutorial: best practices and considerations for mass-spectrometry-based protein biomarker discovery and validation. Nat Protoc. 2021;16(8):3737-60. Epub 2021/07/11. doi: 10.1038/s41596-021-00566-6. PubMed PMID: 34244696; PubMed Central PMCID: PMCPMC8830262.
471. Kalpić D, Hlupić N, Lovrić M. Student's t-Tests. In: Lovric M, editor. International Encyclopedia of Statistical Science. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 1559-63.
472. Smyth GK. Limma: linear models for microarray data. Bioinformatics and computational biology solutions using R and Bioconductor: Springer; 2005. p. 397-420.
473. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat appl genet mol biol. 2004;3:Article3. doi: 10.2202/1544-6115.1027. PubMed PMID: 16646809.

474. Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. Ann Appl Stat. 2016;10(2):946-63. Epub 2017/04/04. doi: 10.1214/16-AOAS920. PubMed PMID: 28367255; PubMed Central PMCID: PMCPMC5373812.
475. Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, et al. Multi-omics factor analysis-a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol. 2018;14(6):e8124. Epub 2018/06/22. doi: 10.15252/msb.20178124. PubMed PMID: 29925568; PubMed Central PMCID: PMCPMC6010767.
476. Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC, et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biol. 2020;21(1):111. Epub 2020/05/13. doi: 10.1186/s13059-020-02015-1. PubMed PMID: 32393329; PubMed Central PMCID: PMCPMC7212577.
477. Orr TJ, Garland T, Jr. Complex reproductive traits and whole-organism performance. Integr Comp Biol. 2017;57(2):407-22. Epub 2017/09/02. doi: 10.1093/icb/icx052. PubMed PMID: 28859419.
478. Hu ZL, Park CA, Reecy JM. Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. Nucleic Acids Res. 2022;50(D1):D956-D61. Epub 2021/12/02. doi: 10.1093/nar/gkab1116. PubMed PMID: 34850103; PubMed Central PMCID: PMCPMC8728226.
479. Yin H, Hou X, Zhang T, Shi L, Su YQ. Participation of EML6 in the regulation of oocyte meiotic progression in mice. J Biomed Res. 2019;34(1):44-53. Epub 2019/04/30. doi: 10.7555/JBR.33.20190014. PubMed PMID: 35081682; PubMed Central PMCID: PMCPMC7007726.
480. Yin H, Zhang T, Wang H, Hu X, Hou X, Fang X, et al. Echinoderm microtubule associated protein like 1 is indispensable for oocyte spindle assembly and meiotic progression in mice. Front Cell Dev Biol. 2021;9:687522. Epub 2021/06/15. doi: 10.3389/fcell.2021.687522. PubMed PMID: 34124073; PubMed Central PMCID: PMCPMC8194061.
481. Walker BN, Biase FH. The blueprint of RNA storages relative to oocyte developmental competence in cattle (Bos taurus). Biol Reprod. 2020;102(4):784-94. Epub 2020/01/27. doi: 10.1093/biolre/ioaa015. PubMed PMID: 31982908.
482. de la Iglesia R, Mansego ML, Sanchez-Muniz FJ, Zulet MA, Martinez JA. Arylesterase activity is associated with antioxidant intake and paraoxonase-1 (PON1) gene methylation in metabolic syndrome patients following an energy restricted diet. EXCLI J. 2014;13:416-26. Epub 2014/01/01. PubMed PMID: 26417268; PubMed Central PMCID: PMCPMC4464483.
483. Pessentheiner AR, Huber K, Pelzmann HJ, Prokesch A, Radner FPW, Wolinski H, et al. APMAP interacts with lysyl oxidase-like proteins, and disruption of Apmap leads to beneficial visceral adipose tissue expansion. FASEB J. 2017;31(9):4088-103. Epub 2017/06/01. doi: 10.1096/fj.201601337R. PubMed PMID: 28559441; PubMed Central PMCID: PMCPMC5566180.
484. Corton M, Botella-Carretero JI, Lopez JA, Camafeita E, San Millan JL, Escobar-Morreale HF, et al. Proteomic analysis of human omental adipose tissue in the polycystic ovary syndrome using two-dimensional difference gel electrophoresis and mass spectrometry. Hum Reprod. 2008;23(3):651-61. Epub 2007/12/25. doi: 10.1093/humrep/dem380. PubMed PMID: 18156650.
485. Braschi B, Omran H, Witman GB, Pazour GJ, Pfister KK, Bruford EA, et al. Consensus nomenclature for dyneins and associated assembly factors. J Cell Biol. 2022;221(2). Epub

2022/01/11. doi: 10.1083/jcb.202109014. PubMed PMID: 35006274; PubMed Central PMCID: PMCPMC8754002.
486. Blyth M, Wellesley D. Ectopic pregnancy in primary ciliary dyskinesia. J Obstet Gynaecol. 2008;28(3):358. Epub 2008/06/24. doi: 10.1080/01443610802058742. PubMed PMID: 18569496.
487. Manenti G, Galbiati F, Gianni-Barrera R, Pettinicchio A, Acevedo A, Dragani TA. Haplotype sharing suggests that a genomic segment containing six genes accounts for the pulmonary adenoma susceptibility 1 (Pas1) locus activity in mice. Oncogene. 2004;23(25):4495-504. Epub 2004/04/06. doi: 10.1038/sj.onc.1207584. PubMed PMID: 15064703.
488. Kabbout M, Garcia MM, Fujimoto J, Liu DD, Woods D, Chow CW, et al. ETS2 mediated tumor suppressive function and MET oncogene inhibition in human non-small cell lung cancer. Clin Cancer Res. 2013;19(13):3383-95. Epub 2013/05/11. doi: 10.1158/1078-0432.CCR-13-0341. PubMed PMID: 23659968; PubMed Central PMCID: PMCPMC3846434.
489. Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol. 2011;7(12):885-7. Epub 2011/10/18. doi: 10.1038/nchembio.687. PubMed PMID: 22002720; PubMed Central PMCID: PMCPMC3218240.
490. Mauer J, Luo X, Blanjoie A, Jiao X, Grozhik AV, Patil DP, et al. Reversible methylation of m(6)A(m) in the 5' cap controls mRNA stability. Nature. 2017;541(7637):371-5. Epub 2016/12/22. doi: 10.1038/nature21022. PubMed PMID: 28002401; PubMed Central PMCID: PMCPMC5513158.
491. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat Genet. 2009;41(1):25-34. Epub 2008/12/17. doi: 10.1038/ng.287. PubMed PMID: 19079261; PubMed Central PMCID: PMCPMC2695662.
492. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet. 2009;41(1):18-24. Epub 2008/12/17. doi: 10.1038/ng.274. PubMed PMID: 19079260.
493. Fischer J, Koch L, Emmerling C, Vierkotten J, Peters T, Bruning JC, et al. Inactivation of the Fto gene protects from obesity. Nature. 2009;458(7240):894-8. Epub 2009/02/24. doi: 10.1038/nature07848. PubMed PMID: 19234441.
494. Liu AL, Xie HJ, Xie HY, Liu J, Yin J, Hu JS, et al. Association between fat mass and obesity associated (FTO) gene rs9939609 A/T polymorphism and polycystic ovary syndrome: a systematic review and meta-analysis. BMC Med Genet. 2017;18(1):89. Epub 2017/08/23. doi: 10.1186/s12881-017-0452-1. PubMed PMID: 28826396; PubMed Central PMCID: PMCPMC5563909.
495. Cheung MK, Gulati P, O'Rahilly S, Yeo GS. FTO expression is regulated by availability of essential amino acids. Int J Obes (Lond). 2013;37(5):744-7. Epub 2012/05/23. doi: 10.1038/ijo.2012.77. PubMed PMID: 22614055.
496. Boissel S, Reish O, Proulx K, Kawagoe-Takaki H, Sedgwick B, Yeo GS, et al. Loss-of-function mutation in the dioxygenase-encoding FTO gene causes severe growth retardation and multiple malformations. Am J Hum Genet. 2009;85(1):106-11. Epub 2009/06/30. doi: 10.1016/j.ajhg.2009.06.002. PubMed PMID: 19559399; PubMed Central PMCID: PMCPMC2706958.
497. Michenet A, Saintilan R, Venot E, Phocas F. Insights into the genetic variation of maternal behavior and suckling performance of continental beef cows. Genet Sel Evol. 2016;48(1):45. Epub

2016/06/24. doi: 10.1186/s12711-016-0223-z. PubMed PMID: 27335091; PubMed Central PMCID: PMCPMC4918023.
498. Albarran-Portillo B, Pollott GE. The relationship between fertility and lactation characteristics in Holstein cows on United Kingdom commercial dairy farms. J Dairy Sci. 2013;96(1):635-46. Epub 2012/11/13. doi: 10.3168/jds.2012-5632. PubMed PMID: 23141835.
499. An N, Yu Z, Yang X. Expression differentiation is not helpful in identifying prognostic genes based on TCGA datasets. Mol Ther Nucleic Acids. 2018;11:292-9. Epub 2018/06/03. doi: 10.1016/j.omtn.2018.02.013. PubMed PMID: 29858064; PubMed Central PMCID: PMCPMC5992444.
500. Lo A, Chernoff H, Zheng T, Lo SH. Why significant variables aren't automatically good predictors. P Natl Acad Sci USA. 2015;112(45):13892-7. Epub 2015/10/28. doi: 10.1073/pnas.1518285112. PubMed PMID: 26504198; PubMed Central PMCID: PMCPMC4653162.
501. Hollywood K, Brison DR, Goodacre R. Metabolomics: current technologies and future trends. Proteomics. 2006;6(17):4716-23. doi: 10.1002/pmic.200600106. PubMed PMID: 16888765.
502. Liu X, Locasale JW. Metabolomics: A Primer. Trends Biochem Sci. 2017;42(4):274-84. Epub 20170211. doi: 10.1016/j.tibs.2017.01.004. PubMed PMID: 28196646; PubMed Central PMCID: PMCPMC5376220.
503. Zhang J, Shi H, Li S, Cao Z, Yang H, Wang Y. Integrative hepatic metabolomics and proteomics reveal insights into the mechanism of different feed efficiency with high or low dietary forage levels in Holstein heifers. J Proteomics. 2019;194:1-13. Epub 20181228. doi: 10.1016/j.jprot.2018.12.026. PubMed PMID: 30594576.
504. Leal LN, Doelman J, Keppler BR, Steele MA, Martin-Tereso J. Preweaning nutrient supply alters serum metabolomics profiles related to protein and energy metabolism and hepatic function in Holstein heifer calves. J Dairy Sci. 2021;104(7):7711-24. Epub 20210423. doi: 10.3168/jds.2020-19867. PubMed PMID: 33896629.
505. Jorge-Smeding E, Polakof S, Bonnet M, Durand S, Centeno D, Petera M, et al. Untargeted metabolomics confirms the association between plasma branched chain amino acids and residual feed intake in beef heifers. PLoS One. 2022;17(11):e0277458. Epub 20221129. doi: 10.1371/journal.pone.0277458. PubMed PMID: 36445891; PubMed Central PMCID: PMCPMC9707789.
506. Horn EJ, Read CC, Edwards JL, Schrick FN, Rhinehart JD, Payton RR, et al. Preovulatory follicular fluid and serum metabolome profiles in lactating beef cows with thin, moderate, and obese body condition. J Anim Sci. 2022;100(7). doi: 10.1093/jas/skac152. PubMed PMID: 35772755; PubMed Central PMCID: PMCPMC9246665.
507. Zhang H, Wu L, Xu C, Xia C, Sun L, Shu S. Plasma metabolomic profiling of dairy cows affected with ketosis using gas chromatography/mass spectrometry. BMC Vet Res. 2013;9:186. Epub 20130926. doi: 10.1186/1746-6148-9-186. PubMed PMID: 24070026; PubMed Central PMCID: PMCPMC3849279.
508. Li Y, Xu C, Xia C, Zhang H, Sun L, Gao Y. Plasma metabolic profiling of dairy cows affected with clinical ketosis using LC/MS technology. Vet Q. 2014;34(3):152-8. Epub 20141009. doi: 10.1080/01652176.2014.962116. PubMed PMID: 25299384.
509. Bender K, Walsh S, Evans AC, Fair T, Brennan L. Metabolite concentrations in follicular fluid may explain differences in fertility between heifers and lactating cows. Reproduction. 2010;139(6):1047-55. Epub 20100412. doi: 10.1530/REP-10-0068. PubMed PMID: 20385782.

510. Read CC, Edwards L, Schrick N, Rhinehart JD, Payton RR, Campagna SR, et al. Correlation between pre-ovulatory follicle diameter and follicular fluid metabolome profiles in lactating beef cows. Metabolites. 2021;11(9). doi: ARTN 62310.3390/metabo11090623. PubMed PMID: WOS:000701408900001.

511. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, et al. The human serum metabolome. PLoS One. 2011;6(2):e16957. Epub 20110216. doi: 10.1371/journal.pone.0016957. PubMed PMID: 21359215; PubMed Central PMCID: PMCPMC3040193.
512. Qiu S, Cai Y, Yao H, Lin C, Xie Y, Tang S, et al. Small molecule metabolites: discovery of biomarkers and therapeutic targets. Signal Transduct Target Ther. 2023;8(1):132. Epub 20230320. doi: 10.1038/s41392-023-01399-3. PubMed PMID: 36941259; PubMed Central PMCID: PMCPMC10026263.
513. Marrella MA, Biase FH. A multi-omics analysis identifies molecular features associated with fertility in heifers (Bos taurus). Sci Rep. 2023;13(1):12664. Epub 2023/08/05. doi: 10.1038/s41598-023-39858-0. PubMed PMID: 37542054.
514. Lu W, Clasquin MF, Melamud E, Amador-Noguez D, Caudy AA, Rabinowitz JD. Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer. Anal Chem. 2010;82(8):3212-21. doi: 10.1021/ac902837x. PubMed PMID: 20349993; PubMed Central PMCID: PMCPMC2863137.
515. Rabinowitz JD, Kimball E. Acidic acetonitrile for cellular metabolome extraction from Escherichia coli. Anal Chem. 2007;79(16):6167-73. Epub 20070714. doi: 10.1021/ac070470c. PubMed PMID: 17630720.
516. Bartel J, Krumsiek J, Theis FJ. Statistical methods for the analysis of high-throughput metabolomics data. Comput Struct Biotechnol J. 2013;4:e201301009. Epub 20130322. doi: 10.5936/csbj.201301009. PubMed PMID: 24688690; PubMed Central PMCID: PMCPMC3962125.
517. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. Epub 20150120. doi: 10.1093/nar/gkv007. PubMed PMID: 25605792; PubMed Central PMCID: PMCPMC4402510.
518. Perry GA, Smith MF, Lucy MC, Green JA, Parks TE, MacNeil MD, et al. Relationship between follicle size at insemination and pregnancy success. P Natl Acad Sci USA. 2005;102(14):5268-73. doi: 10.1073/pnas.0501700102. PubMed PMID: WOS:000228195800062.
519. Northrop EJ, Rich JJJ, Cushman RA, McNeel AK, Soares EM, Brooks K, et al. Effects of preovulatory estradiol on uterine environment and conceptus survival from fertilization to maternal recognition of pregnancy. Biol Reprod. 2018;99(3):629-38. doi: 10.1093/biolre/ioy086. PubMed PMID: 29672673.
520. Zheng Y, Zhu Y, Zhuge T, Li B, Gu C. Metabolomics analysis discovers estrogen altering cell proliferation via the pentose phosphate pathway in infertility patient endometria. Front Endocrinol (Lausanne). 2021;12:791174. Epub 20211115. doi: 10.3389/fendo.2021.791174. PubMed PMID: 34867831; PubMed Central PMCID: PMCPMC8636142.
521. Diener C, Dai CL, Wilmanski T, Baloni P, Smith B, Rappaport N, et al. Genome-microbiome interplay provides insight into the determinants of the human blood metabolome. Nat Metab. 2022;4(11):1560-72. Epub 20221110. doi: 10.1038/s42255-022-00670-1. PubMed PMID: 36357685; PubMed Central PMCID: PMCPMC9691620.