5-2020

Translational Genomics For Improving Sow Fertility

Hiruni R. Wijesena
University of Nebraska - Lincoln, hrw@huskers.unl.edu

TRANSLATIONAL GENOMICS FOR IMPROVING SOW FERTILITY

by

Hiruni R. Wijesena

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Animal Science

(Animal Breeding and Genetics)

Under the Supervision of Professor Daniel Ciobanu

Lincoln, Nebraska

May, 2020

TRANSLATIONAL GENOMICS FOR IMPROVING SOW FERTILITY

Hiruni R. Wijesena, Ph.D.

University of Nebraska, 2020

Advisor: Daniel Ciobanu

Sow fertility traits, such as litter size and number of lifetime parities produced (reproductive longevity), are economically important. Selection for these traits is difficult because they are lowly heritable, polygenic, sex-limited, and express late in life. Age at puberty is an early indicator of reproductive longevity. Gilts that achieve puberty at an early age have a greater probability to produce more parities over their lifetime. However, measuring age at puberty is time consuming and tedious. Identifying pleiotropic polymorphisms that affect age at puberty and other fertility traits, including reproductive longevity, could help to improve the accuracy of genomic prediction for sow fertility traits. We developed a custom Affymetrix SNP array (*SowPro90*) including SNPs located in major QTL regions for age at puberty, other fertility and disease related traits, and potential loss of function SNPs. Genetic variants were identified using deep transcriptomic and genomic sequencing, gene network analysis, and genome-wide association (GWAS) carried out at University of Nebraska-Lincoln (UNL) and US Meat Animal Research Center (USMARC).

This novel SNP array was used to fine map the genetic sources associated with fertility traits. Using a Bayesian haplotype approach (BayesIM), *SowPro90* haplotypes were inferred and assigned to the entire UNL population and were used in an association analysis for age at puberty and other fertility traits. Five major QTL regions located on four chromosomes (SSC2, SSC7, SSC14, SSC18) were discovered for age at puberty. As

expected, a negative correlation ($r = -0.96$ to -0.10; $P<0.0001$) was observed between genomic estimated breeding values for age at puberty and reproductive longevity at these QTL. Some of the SNPs discovered in the major QTL regions for age at puberty were located in candidate genes for fertility traits (e.g. *P2RX3*, *OAS1*, *NR2F2*, *PTPN11*). These SNPs showed significant or suggestive effects on age at puberty, reproductive longevity, and litter size traits in the UNL population and litter size traits in the commercial sows. It will be beneficial to further characterize these SNPs and candidate genes to understand their impact on protein sequence and function, gene expression, splicing process, and how these changes affect phenotypic variation of fertility traits.

ACKNOWLEDGMENT

I would first like to thank my advisor, Dr. Daniel Ciobanu, for this tremendous opportunity and for his continuous guidance and support, helping me to become both a good researcher and a better person. He saw my potential and believed in me, even through the times I doubted myself. His door was always open when I needed help, but he also gave me space to figure things out on my own. His encouraging words like "great job" or "I'm proud of you" always put me in a better mood at the end of a long day and his constant encouragement for questions and discussions, along with his exciting stories and jokes, made working with him enjoyable and stress free. I would also like to thank my other committee members, Dr. Stephen Kachman, Dr. Matt Spangler, and Dr. Gary Rohrer for critiquing my work and sharing their knowledge through out my Ph.D journey. Their guidance and input helped me to improve my critical thinking and familiarize me with many analysis methods that helped to strengthen my research.

I take this opportunity to thank Dr. Jessica Petersen and Dr. Ron Lewis for their guidance and help during my graduate journey and adopting me as their own student when Dr. Ciobanu was not around. I am also thankful to them for giving me the opportunity to assist them with teaching and for their input and advice on how to improve my teaching skills. I'm also grateful to Dr. Clay Lents, Dr. Dan Nonneman, and Dr. Brittney Keel at the US Meat Animal Research Center for their assistance with tissue collection, bioinformatics applications, and development of *SowPro90* SNP array. I would also like to thank my fellow graduate students, Melanie Trenhaile Grannemann,

Lianna Walker, Kylee Sutton, and undergraduate students J'Nan Wittler for all their support and friendship.

I am especially thankful to the Department of Animal Science at University of Nebraska-Lincoln (UNL) and National Institute for Food and Agriculture for providing financial assistance for my research. Additionally, I would like to thank the farm staff at the UNL Swine Research Unit for helping with routine data collection and our industry partners for providing us with tissue samples and fertility data and helping us with designing and manufacturing *SowPro90*.

I am always grateful to my mother Hemamala, my father Uditha, and my brother, Naveen for always believing in me and reminding me that I am capable of more than I think I am. I would not be presenting this dissertation if it were not for their continued support and strength given to me, even being thousands of miles away. Last but not least I would like to thank my husband, Mark for standing by my side giving me support and courage I needed to make my graduate journey a success. He took interest in my research, helped with proofreading, and patiently listened to all my complaints and frustrations. He, along with everyone, knew exactly what to say or do to make my bad days better.

Hiruni R. Wijesena

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 4

CHAPTER 1: LITERATURE REVIEW

REPRODUCTIVE BIOLOGY IN FEMALE PIGS

The female reproductive cycle is initiated with the induction and controlled interaction of hormones in the hypothalamus-pituitary axis and the ovaries (Soede et al., 2011). Female reproduction events begin with the onset of puberty. Age at puberty in gilts is usually defined as the age at which a gilt first ovulates or express estrus (Bidanel, 2011). At the time of puberty, pulsatile secretion of gonadotropin releasing hormone (GnRH) from the hypothalamus stimulates the secretion of gonadotropic hormones such as luteinizing hormone (LH) and follicle-stimulating hormone (FSH) from the anterior pituitary. The LH and FSH act on the ovaries to stimulate gamete formation and secretion of gonadal steroid hormones (e.g. estrogen). In addition, gonadal hormones form feedback loops to regulate GnRH, LH, and FSH release. Once the estrogen levels exceed a certain threshold, it stimulates kisspeptin neurons in the hypothalamus to induce a preovulatory GnRH/LH surge leading to ovulation (d'Anglemont de Tassigny and Colledge, 2010; Duittoz et al., 2016). It is not clearly understood how environmental and nutritional factors trigger these initial key events during puberty (Duittoz et al., 2016). The timing of puberty varies among pig breeds. Chinese breeds attain puberty around three to four months of age while Western breeds take up to six to seven months to reach puberty (Bidanel, 2011; Soede et al., 2011).

Following the first ovulation, gilts start cycles of estrus until they get pregnant. Each estrus cycle spans approximately 18 to 24 days and consists of a follicular (5 to7

days) and a luteal phase (13 to 15 days). Gilts are usually bred on their second or third estrus. In the follicular phase, the number of small and medium follicles in the ovaries starts to decline leaving smaller number of follicles to be recruited for maturation (Knox, 2005). Approximately 6 days prior to ovulation, GnRH from the hypothalamus induces the release of FSH and LH from the anterior pituitary. FSH plays a major role in increasing the number of medium and large follicles needed for recruitment while LH is important for the development and maturation of the recruited follicles for ovulation (Guthrie et al., 1990; Knox, 2005). The recruited follicles then start to produce inhibin and 17β estradiol (Noguchi et al., 2010). Inhibin inhibits FSH secretion, thereby facilitating atresia of smaller follicles. Increased levels of estradiol induce a pre-ovulatory LH surge to prepare mature follicles for ovulation. Following the peak LH surge, ovulation occurs around 30 hours in pigs (Soede et al., 2011).

The period of ovulation is known as estrus and during this interval the females are receptive to males (i.e. standing response). Female receptivity to males can be observed by the immobile arched back and the cocked ears as well as redness and swelling of the vulva. In general, estrus last for 40 to 60 hours in the presence of a boar (Soede et al., 2011). The ovulation occurs approximately two thirds of this period and lasts for one to three hours (Soede et al., 1994; Soede et al., 1998). In pigs, 15 to 30 oocytes are released in each estrus cycle (Knox, 2005) and the ovulation rate increases until fourth or fifth parity (Bidanel, 2011). The follicular recruitment, development, and the rate of ovulation are affected by extrinsic factors such as nutrition (e.g. negative energy balance during lactation) and stress (Hazeleger et al., 2005; Turner and Tilbrook, 2006).

After ovulation, the luteal phase begins. At the time of ovulation the inhibin is decreased, therefore, the negative feedback on FSH is removed. As a result, one to two days following ovulation, FSH facilitates the initiation of a new follicular wave. Once ovulated, the ruptured follicle is formed into a corpora lutea and starts producing progesterone (Noguchi et al., 2010). Progesterone inhibits the release of FSH and LH suppressing the follicular development. In the absence of pregnancy, prostaglandin F2-alpha (PGF) is secreted from the uterus causing luteolysis around 15 days after ovulation (Soede et al., 2011). With the lysis of corpora lutea, the progesterone levels decrease and remove the negative feedback on gonadotropins to begin a new estrus cycle.

If the fertilization is successful, the pig embryo moves from the oviduct to the uterus and implantation begins around day 16 where the conceptus attaches to the luminal uterine epithelium. This is a non-invasive central type implantation (Bazer and Johnson, 2014). A proper implantation facilitates the establishment of a functional placenta for exchange of gasses, nutrients, and other molecules. Pigs have a fertilization rate close to 100% (Bidanel, 2011), but failure to elongate and implant result in 30 to 40% of coceptus death (Bazer and Johnson, 2014). The maternal recognition of pregnancy in pigs is estrogen. It is secreted from the pig blastocyst around 10 to 13 days of pregnancy. Estrogen prevents the secretion of PGF from the uterine epithelium and prevents the regression of corpora lutea. Progesterone secreted from corpora lutea is necessary to maintain the pregnancy to term (Bazer and Johnson, 2014).

Near to farrowing, the growing fetus starts secreting cortisol from the hypothalamic-pituitary-adrenal axis. Cortisol induces breakdown of progesterone and production of PGF. Oxytocin is produced from the maternal posterior pituitary during

uterine contractions. Oxytocin together with PGF regress the production of progesterone from the corpora lutea. During farrowing, PGF facilitates rupture of placental membranes, dilation of the cervix, contractions of the myometrium, placental expulsion, and uterine involution. Farrowing usually last for 3 to 5 hours (Bidanel, 2011). Hypoxia is a critical issue during parturition leading to approximately 10% of piglets being stillborn (Bidanel, 2011). A successful pregnancy lasts for around 114 to 116 days followed by 16 to 40 days of lactation (Soede et al., 2011). Prolactin secreted from the anterior pituitary is essential for lactation and maternal behavior (Bazer and Johnson, 2014). Colostrum is produced during the first 24 hours after parturition, and the peak milk production is observed around 21 days of lactation. During lactation sows are anestrus where they do not exhibit estrus and do not ovulate. Following weaning follicular development is initiated and ovulation occurs within four to 10 days.

ECONOMIC IMPORTANCE OF REPRODUCTIVE TRAITS

The length of sow reproductive life plays a major role in the profitability of swine industry. Improved sow reproductive performance and reproductive longevity enables farmers to keep mature females in the herd while reducing culling rates and replacement costs. Using a net present value analysis, Stalder et al. (2003) showed that in a breed-to-wean operation a gilt must remain in the breeding herd for at least three parities to reach a positive net present value covering all the replacement and maintenance costs. In most swine breeding herds, 40 to 50% of the sows are usually replaced by third or fourth parity when they have just begun to recover their replacement costs (D'Alleire et al., 1987).

Rodriguez-Zas et al. (2003) investigated sow reproductive longevity in 32 commercial herds representing the best and worst genetic lines (n = 8 lines) developed for reproductive longevity over a period of seven years (1995 to 2001). Between the best and the worst lines there was approximately one parity difference in herd life, 0.64 in litter size at weaning, and $52.39 in net income per litter per sow. This emphasized that a greater economic advantage is achieved by selecting for sows with greater reproductive longevity. They also found that at a net income of $50 per litter, a sow should produce at least two parities to cover the initial replacement cost.

Optimizing the herd artificial insemination program has a potential to maximize the farrowing rate and litter size while minimizing unnecessary semen and labor costs. Lamberson and Safranski (2000) developed a model to compare different artificial insemination schedules to identify the most profitable schedule with the highest conception rate. The schedule with four inseminations at 0, 12, 24, and 36 hours after estrus detection performed best in terms of farrowing rate and litter size. When the sows received three inseminations at 12, 24, and 36 hours combined with estrus detection twice per day yielded the highest economic return ($14.90) per bred sow. Four inseminations with estrus detection only once per day yielded second highest economic return ($13.75) per bred sow. Schedules with two inseminations had poorer farrowing rates and litter sizes, and very low positive/negative economic returns due to smaller number of piglets produced regardless of reduced cost of semen and labor.

Faust et al. (1993) evaluated the impact of different gilt replacement rates and genetic change for reproductive traits on economic return in nucleus, multiplier, and commercial level sow production. A simulation was generated where culling was

practiced after a maximum of one, five and 10 parities in the multiplier and commercial levels and after one and five parities in the nucleus. In all three levels, the system that practiced culling after parity one had the largest total cost. However, the same system had the highest genetic change due to continuous introduction of new genetics. At different combinations of maximum parity for culling in the three tiers showed that the return per pig increased annually from $0.85 (culling after parity one in all three levels) to $1.01 (culling after parity one in nucleus and multiplier, and after parity 10 in commercial level). Culling systems with lower replacement rates at commercial level had the lowest total costs, thus, were more profitable. The authors concluded that to maintain the balance between replacement rate and genetic superiority of the herd, a breeding system should always try to minimize the herd replacement rate and purchase genetically superior gilts when replenishing the breeding stock.

INDICATORS OF SOW REPRODUCTIVE LONGEVITY

Based on breeding goals and phenotypic information available, the definition for sow reproductive longevity varies among different studies. Some of these definitions include number of lifetime parities produced by a sow before culling (lifespan), probability of a sow reaching at least four parities (stayability), number of live piglets produced during the lifetime (lifetime prolificacy), duration from first farrowing to culling (herd life), and number of pigs produced per day of life (Hoge and Bates, 2011). Reproductive longevity depends on the ability of females to initiate and maintain ovarian cyclicity, rebreed, and farrow following successive parities. Therefore, it is a composite

trait influenced by many fertility traits such as age at puberty, ovulation rate, age at first service, conception rate, litter size traits, and wean to service interval (Tart et al., 2013).

From a range of pre-breeding gilt phenotypes, age at puberty was found to be the earliest indicator of reproductive longevity regardless of definition (Tart et al., 2013). Gilts that achieved puberty at an early age had mated and farrowed early in life, produced more lifetime number of parities, and remained in the breeding herd longer (Serenius and Stalder, 2006; Engblom et al., 2008; Patterson et al., 2010; Knauer et al., 2010; Tart et al., 2013). Tart et al. (2013) also observed a moderate negative relationship between the genomic estimated breeding values (GEBVs) for age at puberty and reproductive longevity (r = -0.45).

Gilts that attain puberty late were bred and farrowed at a later age compared to early pubertal gilts. Sows with delayed age at first farrowing were at a greater risk of being culled due to low reproductive performance or longevity (Serenius et al., 2006; Hoge and Bates; 2011). Knauer et al. (2010a) found that across six genetic lines, age at first farrowing had the largest effect on the ability to be retained in the herd explaining 6% of the phenotypic variation of reproductive longevity. Once age at first farrowing was removed from the model, age at puberty had the largest effect on reproductive longevity suggesting that age at puberty and age at first farrowing (in the absence of puberty data) are good early indicators of sow reproductive longevity.

Litter size traits such as more piglets born alive, fewer stillborn fetuses, and larger adjusted 21-day litter weight at first farrowing were significantly associated with longer reproductive life in sows (Knauer et al., 2010a; Hoge and Bates, 2011). Serenius et al. (2006) reported that sows with intermediate litter sizes had the highest chance of

remaining in the herd whereas sows with extreme litter sizes (< 9 and > 13) were at a greater risk of being culled.

Increased feed intake during lactation increased the probability of sows to produce up to four parities (Serenius et al., 2006; Knauer et al., 2010a). This effect was more profound in primiparous sows compared to multiparous sows. Sows with greater loss of back fat during lactation had a shorter reproductive lifespan indicating the importance of optimal back fat levels for rebreeding and maintenance of successive pregnancy to term (Serenius et al., 2006). Managing an optimum-feeding regime during lactation should be a main emphasis of swine producers to improve stayability of the breeding herd. Lactation length also had a significant association with stayability where sows with shorter lactation length (< 11 days) had a greater risk of being culled (Xue et al., 1997; Knauer et al., 2010a). Females that had > 30 days weaning to estrus interval after first parity were at 1.7 times greater risk of being culled compared to sows that returned to estrus less than four days (Tantasuparuk et al., 2001).

GENETIC VARIATION IN SOW REPRODUCTIVE TRAITS

Heritability of Reproductive Traits

Sow reproductive traits are usually lowly heritable and considerably influenced by environmental variation (Bidanel, 2011). Few traits that largely depend on the genotype of the female, such as age at puberty, age at first farrowing, ovulation rate, weaning to estrus interval, and estrus symptoms have moderate heritability (Bidanel, 2011). On the other hand, fertility rate and prolificacy traits including conception and farrowing rate,

litter size, litter weight, and piglet survival rate are lowly heritable since they are highly influenced by environment, management, and also by sow, boar, and piglet genotypes (Bidanel, 2011).

A wide range of heritability estimates were reported for reproductive traits in different studies depending on different trait definitions and populations studied. Heritability of age at puberty was estimated at 0.29 (Bidanel et al., 1996; Knauer et al., 2010b), 0.38 (Tart et al., 2013), 0.41 (Schneider et al., 2011), and 0.57 (Hsu, 2011). In a comprehensive review by Rothschild and Bidanel (1998), reported heritabilities for age at puberty ranging from 0 to 0.73 with a mean of 0.37. For other moderately heritable traits, the estimates varied from 0.1 to 0.59 for ovulation rate (Rothschild and Bidanel, 1998; Johnson et al., 1999; Hsu, 2011), 0.38 for age at first service (Holm et al., 2004), 0.16 to 0.22 for age at first farrowing (Serenius et al., 2008; Knauer et al., 2011), and 0.07 to 0.36 for weaning to estrus interval (Rothschild and Bidanel, 1998; Serenius and Stalder, 2006). Heritability estimates reported by Knauer et al. (2010b) for estrus symptoms (e.g. length of estrus, strength of the standing reflex with or without a boar, and vulva redness, swelling, and width) ranged from 0.13 to 0.58. These estimates were higher than the range (0.09 to 0.29) reported in the Rothschild and Bidanel (1998) review.

The heritability estimates for litter size traits such as total number of piglets born (TNB), number of piglets born alive (NBA), and number of piglets weaned per litter were similar and ranged from 0.02 to 0.18 (Johnson et al., 1999; Chen et al., 2003; Holm et al., 2004; Knauer et al., 2011; Tomiyama et al., 2011; Schneider et al., 2012a; Abell et al., 2013; Tart et al., 2013). The heritability of farrowing rate and farrowing interval were estimated between 0.03 and 0.1 (Rothschild and Bidanel, 1998;

Serenius and Stalder, 2006; Tomiyama et al., 2011; Abell et al., 2013). Relatively low heritability estimates (0.02 to 0.25) were reported for reproductive longevity related traits (Serenius and Stalder, 2006; Serenius and Stalder, 2007; Serenius et al., 2008; Mészáros et al., 2010; Abell et al., 2013; Tart et al., 2013).

Even though the heritability estimates for reproductive traits varied between studies due to different phenotypes targeted, size of the data sets, and statistical approaches, in most populations sufficient genetic variation exists for traits such as age at puberty, estrus symptoms, and litter size traits to be improved through selection (Serenius and Stalder, 2007; Serenius et al., 2006; Knauer et al., 2010b).

Genetic Correlation Between Reproductive Traits

Reproductive traits can be genetically correlated with one another if a single gene/allele are pleiotropic. In situations where the trait of interest is expressed late in life, difficult or expensive to measure, it is more feasible to select for an easily available and early expressed correlated trait.

Age at which gilts attain puberty is an important determinant of their future reproductive performance. A negative correlation (r = -0.27) was reported between age at puberty and ability of gilts to farrow their first litter (Knauer et al., 2011). There was a high positive genetic correlation (r = 0.76) between age at puberty and age at first farrowing suggesting gilts that attain puberty early can be bred early and will farrow at an earlier age (Knauer et al., 2011). Age at puberty onset was also positively correlated (r = 0.45) with weaning to estrus interval (Sterning et al., 1998). A moderate negative correlation (r = -0.45) was reported by Tart et al. (2013) between genomic estimated

breeding values (GEBVs) for age at puberty and reproductive longevity (measured as number of parities produced by a sow during lifetime). Negative correlations were also observed between GEBVs for age at puberty and TNB (r = -0.25) and NBA (r = -0.28). The earlier gilts attain puberty, the longer they stay in the herd and produce more litters and piglets during their lifetime suggesting that age at puberty is a good early indicator of sow reproductive longevity (Tart et al., 2013).

Several studies analyzed the effect of ovulation rate as an indicator trait to select for litter size (Johnson et al., 1984; Ruiz-Flores and Johnson, 2001). Ovulation rate showed a moderate, negative correlation with prenatal survival (r = -0.36) probably due to limited uterine capacity, farrowing survival (r = -0.27), and birth to weaning survival (r = -0.38) (Bidanel, 2011). Therefore, selecting for higher ovulation rate would not necessarily be beneficial in terms of improving litter size at birth and weaning. There was a moderate positive genetic correlation (r = 0.63) between shorter farrowing interval (gestation + lactation + days to next breeding) and number of litters per sow per year (Abell et al., 2013). A higher genetic correlation was also observed between litter size at birth and at weaning ($r \geq 0.73$) (Bidanel, 2011).

Sow maternal behavior, especially during early piglet life is important for pre-weaning piglet survival. A moderate, negative genetic correlation (r = -0.24) was observed between maternal behavior measured as response to piglet screaming and piglet mortality rate during the first days of lactation suggesting that selecting for sows with stronger maternal response could improve the survival rate of nursing piglets (Grandinson, 2005).

Between Breed Variation for Reproductive Traits

In addition to within line selection, swine breeding industry relies on heterosis for improved productivity and revenue, a phenomenon where hybrids are superior to the average of the purebred parental lines. To capture the effect of hetorosis, Chinese and Western (local) breeds were widely used for crossbreeding, since large variation for maternal reproductive traits was observed between the breeds (Bidanel, 2011; Soede et al., 2011). Chinese Meishan is a well-known breed for their superiority for maternal reproductive traits (Bidanel, 2011). In the early 1990s Meishan sows produced three to five more piglets per litter compared to Large White (Bidanel, 2011). Following the emphasis on selection for litter size in Western maternal breeds (e.g. Large White and Landrace), in early 2000s this difference was reduced to 1.1 piglets per litter (Canario et al., 2006). Meishans express puberty earlier, have higher conception and prenatal survival rates, and shorter weaning to estrus interval compared to Western breeds (Canario et al., 2006; Bidanel, 2011).

Differences in fertility and prolificacy traits were also observed between many Western breeds. Bidanel et al. (1996) analyzed the genetic variation of reproductive traits between important maternal breeds (Large White, Landrace, and their crossbred gilts). Large White gilts reached puberty 17 days later, had a higher ovulation rate (+1.3), but a lower embryo survival rate (-7.1%) compared to Landrace. Crossbred gilts reached puberty earlier and had higher embryo survival rates than both purebreds. Mészáros et al. (2010) found that Landrace sows stayed in the herd longer (+92 days) and produced more parities (+0.56) compared to Large White sows. Large White sows had more live piglets

born (0.6) compared to Landrace. Both Large White and Landrace females had higher prolificacy, conception rate, and lower pre-weaning mortality compared to pig breeds that were primarily selected for growth and carcass characteristics such as Duroc and Pietrain (Mészáros et al., 2010). Differences in sow fertility traits were even larger when the maternal breeds were compared to paternal breeds. The TNB ranged from approximately 10.0 piglets in paternal breeds (Duroc and Pietrain) to 14.5 piglets in maternal breeds (Large white and Landrace). The number of piglets weaned ranged from 7.8 (Pietrain) to approximately 11.0 (Large White and Landrace) (Bidanel, 2011).

GENOMIC APPROACHES TO UNDERSTAND AND IMPROVE SOW REPRODUCTIVE TRAITS

Genome-wide Association Studies

The direct use of DNA information in commercial pig breeding became popular in early 1990s with the discovery of a DNA polymorphism in *RYR1* gene (Hal-1843 marker) responsible for malignant hyperthermia syndrome (Fujii et al., 1991). A DNA test and marker-assisted-selection protocols were developed and used globally to control and reduce the rate of this syndrome. The test was very effective, since it was not labor intensive, did not involve progeny testing, and the syndrome was controlled by a single gene. However, most economically important livestock traits are quantitative in nature and controlled by many genes (quantitative trait loci or QTL) and genetic variants with smaller effects, therefore difficult to improve through just marker-assisted selection.

The first QTL associated with prolificacy traits in pigs was identified for litter size and mapped next to *estrogen receptor* (*ESR1*) gene (Rothschild et al., 1996).

In the late 1990s, methods were developed to include marker genotypes in conventional best linear unbiased prediction (BLUP) analysis to generate marker enhanced estimated breeding values (Fernando and Grossman, 1989; Knol et al., 2016). At the time, genetic evaluation of production traits included about 30 markers, most of them identified using candidate gene approach (Knol et al., 2016). In early 2000s, genome-wide association studies (GWAS) became popular in pig research with the development of DNA marker panels including large number of SNPs located across the genome, high throughput genotyping and sequencing techniques, and Bayesian statistical approaches to estimate effects for large number of SNPs (Samorè and Fontanesi, 2016). The first commercially available high-density marker panel for pigs was *Porcine SNP60 BeadArray* developed by Illumina (Illumina, San Diego, CA) with the input from swine genetic community. The majority of SNPs included in these panels were functionally neutral in nature, polymorphic across Western commercial breeds, and uniformly mapped across the genome. In GWAS, the genotypes of each SNP are tested for association with the targeted trait assuming that significant associations occur when the SNPs are in linkage disequilibrium (LD) with a causal mutation directly affecting the trait of interest (Hayes and Goddard, 2010).

Several Bayesian statistical models (e.g. BayesA, BayesB, BayesC, and BayesCπ) were developed to simultaneously assess the effects of thousands of markers to the target trait (Meuwissen et al., 2001). BayesA assumes each marker has a unique variance and they all have an effect on the trait. Therefore, the π value (percentage of markers that do

not have an effect) is set to zero (Meuwissen et al., 2001). BayesB assumes that each SNP has its own variance and only a fewer number of loci have an effect on the trait, while many do not. Therefore, the π value could range from zero to one (Meuwissen et al., 2001). BayesA is a special case of BayesB when π = zero. BayesC differs from BayesB assuming that all the markers have a common variance. The shrinkage of SNP effects due to LD is affected by the π value, therefore, in BayesCπ it assumes that π is unknown and all the SNPs have the same genetic variance. In BayesCπ different π values are fitted until convergence (Habier et al., 2011). In addition to these models, Kachman (2015) introduced a new Bayesian model called Bayes interval mapping (BayesIM) that fits haplotypes rather than individual SNPs, as is the case of Bayesian methods discussed earlier. BayesIM uses a hidden Markov model to generate haplotype clusters. The haplotype clusters are formed across the chromosomes while their effects are tested by evenly spacing the putative QTL along the genome.

Many GWAS have been carried out for sow reproductive traits, mainly for prolificacy traits, to map QTL and identify positional candidate genes. Several studies have reported QTL for different reproductive traits mapped to same chromosomal regions suggesting the pleiotropic effects of genes underlying these QTL. However, there is very limited progress in identification of functional variants/genes in these QTL.

Teat Number

Verardo et al. (2016) compared Poisson and Gaussian distributions for discrete reproductive traits and found that Gaussian models worked best for teat number. Sixty-five significant SNPs overlapping 57 positional candidate genes (e.g. *YLPM1*,

SYNDIG1L, *TGFB3*, and *VRTN*) were identified. These genes were used to generate gene-transcription factor networks. The most significant transcription factors identified for teat number (*SOX9* and *ELF5*) were involved in mammary gland development. In addition, *TINAGL1* (SSC7, 30 Mb) and *ICK* (SSC7, 134 Mb) genes were identified as candidate genes for teat number involved in regulation of cell cycle and apoptosis during mammary development (Verardo et al., 2015).

Ovulation Rate

A GWAS for ovulation rate in a composite pig population including females that produced zero to two parities mapped 22 QTL ($P \leq 0.001$) accounting for 71.10% of the total genetic variance (Schneider et al., 2014). A QTL on SSC1 (16 Mb) overlapping *ESR1* gene explained 3.61% of the genetic variance for ovulation rate. Another QTL on SSC17 (64 Mb) explained 23.78% of the genetic variance. The most compelling candidate gene in the region was *BMP7* involved in the regulation of ovarian functions. Two QTL on SSC2 (137.3 Mb and 139.9 Mb) explained 26.88% of the genetic variance (Schneider et al., 2014). *ADAMTS19* gene located on 137 Mb region is involved in premature ovarian failure in humans (Knauff et al., 2009) and *GDF9* gene located on 139 Mb region is known to control ovulation rate (Crawford and McNatty, 2012).

Litter Size Traits

An early study found a variation in *ESR1* (SSC1) being associated with litter size in Meishan synthetic lines (Rothschild et al., 1996) and Large White based commercial lines (Short et al., 1997). The additive effect estimates for favorable alleles varied from

0.42 (TNB) and 0.39 (NBA) pigs per litter in commercial Large White (Short et al., 1997) to 1.2 (TNB and NBA) pigs per litter in Meishan crosses (Rothschild et al., 1996). The same SSC1 QTL was also reported for ovulation rate (Schneider et al., 2014).

The proportion of phenotypic variance explained by *Porcine SNP60 BeadArray* SNPs for litter size traits was fairly low (0.001 to 0.4) (Onteru et al., 2012; Tart et al., 2013). Schneider et al. (2012b) reported number of overlapping QTL and candidate genes for TNB and NBA in first parity sows (*FEM1B* [SSC1, 173.5 Mb], *CRH* [SSC4, 71 Mb], *SNX7* [SSC4, 126 Mb], *HFM1* [SSC4, 132 Mb], and *ACSL3* [SSC15, 130 Mb]). A QTL identified by Schneider et al. (2012b) on SSC17 for NBA (64 to 65 Mb) also overlapped with a QTL for ovulation rate (Schneider et al., 2014).

In a GWAS for litter size traits by Onteru et al. (2012), approximately 50% of the genes located in the QTL regions were predicted to be involved in placental functions. Among them, *MEF2C* (SSC2), *PLSCR4* and *PLSCR5* (SSC13) were identified as candidates for both TNB and NBA in the first and second parities. Onteru et al. (2012) and Schneider et al. (2012b) reported QTL located on the same regions of SSC17 (30 Mb) and SSC1 (86 Mb) for TNB and NBA, respectively.

Two statistical models, 1) linear mixed model with a single SNP regression, 2) Bayesian mixture model including effects of all SNPs simultaneously (assuming a portion of the markers had small effects and a portion of the markers had large effects) were tested for TNB, day five litter size, and mortality in a large Landrace and Yorkshire population (Guo et al., 2016). Both models resulted in quite similar associations; however, the association signals were more clear and distinct in the Bayesian model. In Landrace sows, 1-Mb QTL regions on SSC2 (139 Mb, 141 Mb) and SSC3 (6 Mb)

explained 0.75 to 1% of the additive genetic variance for TNB. The SSC3 region was also associated with TNB in Yorkshire sows explaining 0.46% of the variance. In other studies, the same SSC2 region was associated with TNB, number mummified (Onteru et al., 2012), and ovulation rate (Schneider et al., 2014). For both litter size and mortality at day five, a 1-Mb QTL on SSC7 (34 Mb) was identified in Landrace explaining 0.38% and 1.99% of the variance, respectively (Guo et al., 2016). The QTL had a positive effect on litter size and a negative effect on mortality (Guo et al., 2016).

Bergfelder-Drüing et al. (2015) used Large White and Landrace pigs from different commercial breeding companies from Germany, Austria, and Switzerland for a GWAS for NBA. Distinct stratification was observed between and within breeds suggesting different breeding and selection strategies were used by different breeding companies. The study did not observe any overlapping QTL across breeds. In Large White, a QTL for NBA on SSC13 (27.9 Mb) overlapped with a QTL reported by Onteru et al. (2012). In Landrace, a QTL for NBA located on SSC9 (14.8 Mb) overlapped with a QTL for TNB by Onteru et al. (2012). Another QTL on SSC9 (139 Mb) was adjacent to *PTGS2*, a gene involved in placental attachment and embryo survival in pigs (Bergfelder-Drüing et al., 2015). *PTGS2* was one of the highly differentially expressed genes in the uterine endometrium between pigs with low and large litter sizes (Cordoba et al., (2015).

Lifetime Productivity Traits

Onteru et al. (2011) studied lifetime productivity data recorded in a commercial maternal pig line over a maximum of nine parities and identified overlapping candidate genes and QTL between lifetime total number born and lifetime number born alive

(*FUT9* [SSC1], *SLC22A18* [SSC2], and *P2RY6* [SSC9]). The QTL on SSC2 also overlapped with the number of mummified pigs (Holl et al., 2004). The *SLC22A18* gene in this region is known to be associated with reduced fetal intrauterine growth (Onteru et al., 2011).

Transcriptome Sequencing and Expression Profiling

The development of high throughput RNA sequencing techniques have enabled whole transcriptome analysis of specific tissues related to diseases, reproduction, and production traits in humans and farm animals. RNA sequencing is superior to Sanger sequencing and microarray-based expression profiling by providing greater coverage and higher resolution of the transcriptome (Kukurba and Montgomery, 2015). RNA sequencing enables identification of differentially expressed genes (DEGs), alternatively spliced genes, novel transcripts, complete gene structures, and gene coding polymorphisms. With many advances in sequencing technology, now it can be applied not only to mRNA, but also to different populations of RNA such as total RNA, pre-mRNA, and noncoding RNA (e.g. micro RNA (miRNA) and long non-coding RNA) (Ozsolak and Milos, 2011; Kukurba and Montgomery, 2015).

Among many reproductive traits, RNA sequencing was mainly applied for prolificacy traits in pigs. Kwon et al. (2016) analyzed the placenta transcriptome in Berkshire pigs that had larger (mean > 12) and smaller (mean < 6.5) litter sizes. Approximately 22% of the DEGs were associated with gene ontology terms related to reproduction (e.g. fecundity, prolificacy, and litter size). Two DEGs (*WNT9B* and *IL-6*) were highly expressed in the larger litter size group and were identified as upstream

regulators (i.e. transcription factors) of two other DEGs (*EGR2* and *LIPG*). Authors suggested that upregulation of *WNT9B* and *LIPG* could facilitate increased nutrition supply to the growing fetuses via the placenta to maintain a larger litter. In humans, the *LIPG* expression was very low in placentas from intrauterine growth-restricted pregnancies compared to placentas from normal pregnancies (Gauster et al., 2007).

In the same Berkshire population (Kwon et al., 2016), Hwang et al. (2017) studied genome-wide differences in DNA methylation using bisulfate sequencing and gene expression using RNA sequencing in the placental tissue in large and small litter size groups. DNA methylation is an important epigenetic modification that regulates gene expression. It involves addition of a methyl group to the fifth carbon of cytosine in CpG dinucleotides (Smith and Meissner, 2013). Genome-wide methylation profiling provided insights into the molecular processes involved in reproduction (Messerschmidt et al., 2014). The rate of methylation was higher in coding sequences and introns, but lower in promoter regions. The rate of methylation of CpG dinucleotides, especially in introns were lower in larger litter size group. Nine DEGs overlapped with differentially methylated regions. A positive relationship was observed between the differential expression and methylation, except for one gene.

Mammalian ovaries are important reproductive and endocrine organs where ovulation and secretion of many reproductive hormones take place (Soede et al., 2011). Integrating RNA sequencing, gene expression profiling, and gene pathway analysis of ovarian tissues in Yorkshire pigs with high and low litter sizes identified 21 DEGs that were involved in steroid biosynthesis and ovarian steroidogenesis pathways (Zhang et al., 2015). Nineteen of these genes were upregulated in pigs with higher litter size,

suggesting their involvement in ovarian functions. Among them, *HSD3B* and *STAR* genes regulate pre-ovulatory follicular maturation in chicken (Sechman et al., 2014) and *HSD17B2* and *EGR4* are key regulators of steroid hormone metabolism (Zhang et al., 2015). Several DEGs (*CYP11A1*, *RBP4*, *SLC5A10*, *SLC7A11*, and *SLC6A20B*) overlapped with previously identified QTL for litter size in pigs (Zhang et al., 2015).

Micro RNAs are a class of small RNAs that play a major role in regulating gene expression. Huang et al. (2016) looked at the expression of miRNAs in ovarian tissues in the same Yorkshire pig population used by Zhang et al. (2015). There were 37 differentially expressed miRNAs where, 21 were upregulated and 16 were downregulated in the high litter size sows. Among them, MiR-224 targets *Ptx3* gene in mice affecting ovulation and embryo development (Yao et al., 2014), MiR-99a is involved in cattle oocyte maturation (Tesfaye et al., 2009), and Let-7c regulates FSH secretion from follicle cells in mice (Yao et al., 2009).

Pigs usually have higher fertilization rates (Bidanel, 2011). However, large early embryonic loss (20 to 30%) at around 12 to 30 days of gestation, due to decreased placental efficiency and uterine capacity, affect the litter size (Lin et al., 2015). Changes in endometrial environment during early pregnancy play an important role in embryonic survival and successful pregnancy. Lin et al. (2015) evaluated gene expression profiles of the endometrium at the time of maternal recognition of pregnancy (day 12), conceptus attachment (day 18), and embryo implantation (day 25). Comparative gene expression was performed at day 12, 18, and 25 of pregnancy. The largest number of DEGs was observed between day 12 and 25 (n = 8,951). There were 188 common DEGs among the three stages of pregnancy. Several DEGs were associated with different aspects of early

pregnancy such as fibroblast growth (*FGF9*), immune response (*IRF1*, *S100A9*), adhesion (*OSTN*), prostaglandin synthesis (*PTGES*), and implantation marker (*STC1*). The immune system plays an active role during early pregnancy in mammals (Engelhardt et al., 1997). Lin et al. (2015) observed an over expression of immune response related genes (i.e. *S100A9*) during days 12 and 18 of pregnancy compared to day 25. One limitation of this study was the lack of expression comparison between non-pregnant and pregnant endometrium, which would have provided insight into DEGs between early pregnant and non-pregnant animals.

Cordoba et al. (2015) also studied gene expression in uterine endometrium at day 30 to 32 gestation in an Iberian × Meishan F2 population between sows with extreme EBVs for prolificacy. The tissues were collected at the fifth parity when the litter size has reached the maximum in this population. RNA sequencing identified 141 DEGs between the two groups. Gene ontology analysis proposed three main reproductive pathways (female pregnancy, maternal placenta development, and decidualization) to which DEGs belong. Twenty-five of the DEGs overlapped with QTL identified for litter size traits in the Pig QTL Database (Hu et al., 2013). Based on gene expression, localization with known QTL, and gene ontology analysis, five candidate genes (*HPGD*, *MMP8*, *PTGS2*, *PTHLH*, and *SCNN1G*) for litter size were identified. All the five genes were over expressed in high litter size group. The *PTGS2* gene is involved in early events of implantation (Kennedy et al., 2007) and the gene expression is substantially increased during early pregnancy suggesting that lower expression of the gene in low litter size group could lead to failure in implantation (Cordoba et al., 2015).

Genomic Selection

Traditionally genetic progress in the pig industry was mainly achieved through selection and crossbreeding strategies. In the late 1980s more progress was achieved with the introduction of BLUP animal models to evaluate males and females in the nucleus populations. In early 2000s, genomic selection was first introduced to predict the genetic merit of selection candidates early in life using their genomic information with limited need to phenotype (Meuwissen et al., 2001). A GEBV is calculated for each selection candidate by summing up allelic substitution effects for thousands of genetic markers (SNPs) across the pig genome (Samorè and Fontanesi, 2016).

Genomic selection is performed first by estimating the SNP effects (as a regression of the phenotype on the genotype) in animals with both phenotypic and genotypic data available (training population). The estimated SNP effects are then used to predict the GEBVs for animals with only genomic data available (evaluation population). Application of genomics is more advantageous for traits that are difficult or expensive to measure, expressed late in life, sex-limited, or lowly heritable. Genomic selection has the ability to increase the genetic gain by improving the prediction accuracy and decreasing the generation interval in traditional breeders equation.

In genomic selection, the prediction accuracy of GEBVs largely depends on the size and the breed composition of the training population, relatedness of the training and evaluation populations, marker density, heritability of the trait, and the statistical methods (Meuwissen, 2009; Cleveland et al., 2012). For example, Meuwissen (2009) reported that to obtain an accuracy of 0.3 for a lowly heritable trait, the training population should at

least contain 2,000 animals. The GEBV accuracy could also be improved by incorporating whole genome sequence data into evaluations since this information may capture causal variants and help the prediction equations to be more stable over time and across populations (Samorè and Fontanesi, 2016). One major limitation in application of genomic selection is the cost of genotyping. Cleveland and Hickey (2013) suggested strategies to implement cost effective low-density marker panels and evaluated the prediction accuracy with genotype imputation.

Genomic selection is widely applied to maternal and performance traits in the swine industry. Knol et al. (2016) illustrated how application of genomic selection improved not only a moderately heritable maternal trait (teat number, $h^2 = 0.4$), but also a binary, lowly heritable trait (post weaning piglet mortality, $h^2 = 0.05$). For teat number, in a conventional pedigree-based relationship matrix, addition of the most significant SNP increased the EBV accuracy by 7% and addition of four most significant SNPs increased the accuracy by 27%. Using a genomic relationship matrix instead of pedigree-based matrix increased the EBV accuracy by 50% (Knol et al., 2016). In early 2000s genetic progress for post-weaning mortality was mainly achieved through marker-assisted selection using five to 20 markers. With the development of high-density marker panels and implementation of genomic selection in 2010, the accuracy of EBV for this trait increased by 50% (0.14 to 0.22).

Lillehammer et al. (2011) evaluated the effect of genomic selection on maternal traits in a simulated Norwegian Landrace pig population. Genomic selection was applied to boars for a trait with heritability of 0.1 and measured (expressed) in females after the first parity. Different breeding schemes were simulated where one, two, and three males

from each litter and zero, 50%, and 100% females were genotyped. In a conventional breeding scheme, the boars are selected based on progeny testing. Information from older relatives leads to increased generation interval and lower accuracy resulting in a slower rate of genetic gain. Compared to conventional breeding scheme, genomic selection increased genetic gain from 23% (one male per litter and zero females genotyped) to 91% (three males per litter and 100% females genotyped). In all breeding schemes, the accuracy of selection was higher for females. In an additional simulation, Lillehammer et al. (2013) showed that genotyping 2,400 females each year doubled the maternal trait contribution to the total genetic gain in the herd, mainly via increasing the training population size. By comparing the genetic gain from different breeding schemes authors concluded that genotyping more females is important to reach a higher prediction accuracy and genetic gain, especially for maternal traits.

Cleveland et al. (2012) utilized a data set generated by the Pig Improvement Company (PIC) to test and validate different genomic prediction methods. The data set included individuals from a single nucleus genetic line (n = 3,534) genotyped with high-density genotypes (*Porcine SNP60 BeadArray*), phenotypes for five traits (h^2 ranging from 0.07 to 0.62), polygenic EBVs (PEBV; no genomic information), and complete pedigrees. The GEBVs were estimated using a BayesB approach using both phenotypes and de-regressed EBVs (a higher accuracy phenotype calculated using progeny and multigenerational pedigree information) and a single-step genomic BLUP approach that combines information from genotyped and un-genotyped animals. Accuracies for all traits improved with BayesB de-regressed EBVs and single-step approaches compared to BayesB phenotype-based approach. The prediction accuracy generally increased as the

trait heritability and the relatedness between training and evaluation sets increased. Increasing the relatedness between populations did not change the ranking of the statistical methods suggesting that prediction accuracy primarily depends on the information content used in different methods. This emphasizes the importance of additional genotyping, especially to improve lowly heritable traits using genomics.

The effectiveness of SNP panels used for genomic selection mainly depends on its ability to capture functional effects and predict cumulative additive genetic merit for animals in the evaluation population. This has been proven to work well within a population when the animals used in the training set are closely related to those used for evaluation. However, due to difficulty in measuring certain traits in commercial settings (e.g. age at puberty), it is necessary to transfer genomic information from experimental populations to potentially disjoint industry populations. Lucot et al. (2015) illustrated that for age at puberty, transferring SNP effects from training to evaluation populations resulted in low correlations between GEBVs and adjusted phenotypes. When all the SNPs from the top ranked 1, 5, 10, 20, and 50% 1-Mb windows identified in a training set (n = 820) were used in an evaluation set (n = 412) consisting of subsequent generations of similar genetics, the phenotypic variation that was explained ranged from 12.3 (top 1% 1-Mb windows) to 36.8% (top 20%). When only the highest ranked SNP from these subsets of 1-Mb windows (e.g., 1 SNP per 1-Mb window) were evaluated, the phenotypic variation captured was less and varied from 6.5 (top 1%) to 23.7% (top 50%). This is probably due to the fact that SNPs identified as highest ranked in the training set are not functional variants, and the LD between these SNPs and the functional polymorphisms is redefined in the evaluation set. However, Lucot et al. (2015)

emphasized that the knowledge of important regions can be captured using all SNPs in the region identified in the training set and re-estimating their effects in the evaluation population. Specifically, the correlations between GEBVs based on SNP effects estimated in the training set and the phenotypes of evaluation set was marginal (r = -0.01 to 0.17) compared with their effects retrained in the evaluation set for all (r = 0.46 to 0.81) or most informative SNPs (r = 0.30 to 0.65) from the high-ranked 1-Mb windows.

The swine breeding scheme follows a pyramidal structure with nucleus at the top, followed by multiplication, and commercial levels. The genomic evaluation and selection are largely performed within the purebreds at the nucleus level. The final product in swine industry is a crossbred animal and the goal is to improve crossbred performance in the commercial herd (Samorè and Fontanesi, 2016). It has been shown that genetic differences between purebred and crossbred animals as well as environmental differences between the two levels result in poor prediction of crossbred performance using purebred GEBVs (Dekkers, 2007). In addition, if breed specific effects of marker alleles exist, crossbred GEBVs calculated using only data from one purebred nucleus line might not be accurate. To investigate the breed specific effects on GEBV accuracy in crossbreds, Lopes et al. (2017) compared a traditional genomic selection model and a model that includes breed specific effects, trained on purebred (Large White and Landrace, n = 924 for each breed) or crossbred data (n = 924) for litter size and gestation length in pigs. A higher genetic correlation (r > 0.88) was observed between purebred and crossbred performance for both traits. The GEBV prediction accuracies of crossbred sows were highest when the training was done on crossbred data. Both models resulted in similar prediction accuracies for litter size (~ 0.23) and gestation length (~ 0.52) suggesting that

accounting for breed specific effects did not necessarily impact the prediction accuracy in this study. Authors suggested that to assess the benefit of adding breed specific relationships in genomic prediction models, a larger crossbred population of more distant purebreds must be used with traits that have a lower genetic correlation between purebred and crossbred performance.

Estimating marker effects on crossbred animals using commercial data to evaluate purebred nucleus animals was proposed as a method to incorporate commercial data into genomic evaluations (Dekkers, 2007). Toosi et al. (2010) reported that in a simulated data set when the training was performed in crossbreds using different marker densities to predict GEBVs in a purebred evaluation set, the prediction accuracy was slightly less (0.66 to 0.74) compared to when training was done in the same purebred population as the evaluation set (0.79 to 0.85). Ibánẽz-Escriche et al. (2009) also reported that accuracies based on crossbred data were slightly lower than the accuracies based on purebred data. But this difference was negligible when the parental breeds were closely related.

GENETIC ANALYSIS OF AGE AT PUBERTY IN MAMMALS

QTL and Candidate Genes for Age at Puberty in Pigs

Reducing age at which gilts attain puberty and farrow is economically important for swine producers since it can reduce the time and money spent on the replacement gilts before they produce a litter. Early age at puberty is associated with more litters per sow during her lifetime (Tart et al., 2013). However, measuring age at puberty in a

commercial setting is difficult and labor intensive. Pleiotropic genetic markers that can predict the propensity of individuals to attain puberty early and produce more parities could be more effective in selecting for superior females to retain in the breeding herd. Quantitative trait loci mapping and GWAS are widely used to identify candidate genes and polymorphisms associated with age at puberty in gilts. Many QTL were reported by different studies in different populations suggesting that age at puberty in gilts is a classical quantitative trait affected by many genes with smaller effects.

QTL Mapping in Western × Chinese F_2 Populations

Several QTL mapping studies reported a locus on SSC7 (54 to 58 cM) near the Swine Leucocyte Antigen Complex II (SLAII) region associated with age at puberty in F2 crossbred populations (Bidanel et al., 2008; Yang et al., 2008). In a Large White × Meishan population, the SSC7 QTL explained 2.9% of the phenotypic variance of age at puberty and negative effect of Large White alleles was observed (Bidanel et al., 2008). This was expected since many Western maternal breeds mature late compared to Chinese breeds (Bidanel, 2011). In a White Duroc × Erhualian population, the same QTL explained 8% of the phenotypic variance of age at puberty (Yang et al., 2008). The Duroc allele was associated with early age at puberty. Duroc is a late maturing breed compared to many Chinese breeds, thus, it was interesting to observe the favorable allele at this locus in Duroc. Several other QTL were reported for both reproductive and growth traits at the same SSC7 SLAII locus (Sanchez et al., 2006). The SLAII region is polygenic, very polymorphic, and characterized by high LD, therefore, whether this QTL identified for different traits is represented by a single gene due to pleiotropic effects or many

closely linked genes involved in the variation of different traits still needs to be determined.

Rohrer et al. (1999) reported a QTL on SSC10 (125 cM) in a White composite × Meishan population with an additive effect of 27.6 days for age at puberty. The Meishan allele at this locus was favorable for earlier age at puberty. Another QTL on SSC1 (105 cM) had an additive effect of 9.35 days (Rohrer et al., 1999). The same QTL was reported by Kuehn et al. (2009) with a relatively smaller additive effect (3.2 days) in a population of Yorkshire × Landrace dams mated with Duroc or Landrace boars. An intronic polymorphism in the *PAX5* gene overlapping SSC1 QTL had the largest effect on age at puberty ($P = 0.036$). However, authors did not find a direct functional relationship of this gene with age at puberty or other reproductive traits.

In a White Duroc × Erhualian crossbred population, Yang et al. (2008) reported a QTL on SSC1 (114 cM) explaining 3.9% of the phenotypic variance of age at puberty. Two more QTL on SSC8 (77 cM) and SSC17 (88 cM) explained 2% and 2.4% of the phenotypic variance for age at puberty, respectively (Yang et al., 2008). The Erhualian allele was associated with earlier age at puberty in all three QTL regions.

Genome-wide Association Studies in Commercial Crossbred Populations

Tart et al. (2013) conducted a GWAS for age at puberty in a population of Large White × Landrace crossbreds and Nebraska Index Line (NIL) dams mated with Landrace boars from two unrelated commercial lines. The NIL was originated by crossing a high indexing line selected for increased ovulation and embryonic survival for eight generations with a randomly selected control line (Ruız-Flores and Johnson, 2001).

Subsequently, the NIL was selected for increased litter size for 29 generations, while the last 12 generations also included within litter selection for increased growth and reduced backfat (Miller et al., 2011; Hsu and Johnson, 2014). In the GWAS by Tart et al. (2013), the top 1% of 1-Mb windows explained 11% of the genetic variation of age at puberty. Three SNPs located on SSC5 (27 to 28Mb), SSC8 (36 to 37 Mb), and SSC12 (1.2 to 2 Mb) exhibited pleiotropic additive effects with age at puberty and the number of parities produced during lifetime. Individuals with five favorable alleles across the three loci reached puberty 7 days earlier and produced 1.36 more parities compared to individuals that did not carry any favorable alleles. *AVPR1A* was a candidate gene in the SSC5 pleiotropic QTL region. The gene encodes a G-protein-coupled receptor involved in social and reproductive behavior (Caldwell et al., 2008; Walum et al., 2008; Gobrogge et al., 2009). There were three non-synonymous SNPs identified in this gene (*G31E*, *G256D*, and *K377Q*). Homozygozity for the favorable *31G* allele of *AVPR1A G31E* SNP (*BGIS0007637*) was associated with 5.8 days early expression of puberty and 0.53 more lifetime number of parities compared to homozygozity for the *31E* allele (Tart et al., 2013).

In a composite gilt population developed using maternal and terminal Landrace, Duroc, and Yorkshire lines, Nonneman et al. (2016) reported two QTL on SSC12 (15 Mb) and SSC7 (75 Mb) that explained 9.7% and 7.1% of the total genetic variance of age at puberty, respectively. The most interesting candidate gene on SSC12 was *GH1*. A reduction of serum growth hormone levels was observed towards puberty in gilts (Klindt and Stone, 1984). Insufficiency of growth hormone caused ovarian dysfunction leading to problems in sexual maturation and irregular menstrual cycle in women (Spiliotis et al.,

2003). The candidate gene reported on the SSC7 QTL region was *PRKD1*. Although a direct relationship of this gene with fertility was not reported, it is involved in body mass index in humans. Wang et al. (2006) reported that as body mass index increased, the age at puberty decreased in humans.

Six common QTL regions for age at puberty were reported by Tart et al. (2013) and Nonneman et al. (2016). Candidate genes in these regions (*IQCHI* and *RORA*, SSC1; *CRTC1*, SSC2; *AQP8* and *GPRC5B*, SSC3; *NEGR1*, SSC6) were also identified as candidates for age at menarche in humans (Elk et al., 2010; Demerath et al., 2013). *CRTC1* on SSC2 is involved in the Leptine-Kisspeptin-GnRH signalling pathway. Leptin increases the expression of *CRTC1* in the hypothalamus and *CRTC1* increases the expression of kisspeptin, which then activates GnRH to initiate puberty (Li et al., 2008). A deletion in the mouse *RORA* gene was associated with delayed puberty and lower number of mature oocytes compared to wild type mice (Guastavino et al., 1992).

A potential QTL reported by Tart et al. (2013) on SSC10 (70 to 70.9 Mb) was also reported by Rohrer et al. (1999) and Nonneman et al. (2016). *AKR1C2* is a candidate gene located in this region. *AKR1C* gene family members are involved in regulation of steroid hormones during puberty (Griffin and Mellon, 2001) and teat number (Hirooka et al., 2001). A non-synonymous SNP in *AKR1C2* gene was associated with age at puberty in a Meishan, Landrace, and Large White composite population (Nonneman et al., 2006).

Nonneman et al. (2014) reported another QTL on SSC14 (114.6 Mb) overlapping *NHLH2* gene. Mice deficient in *NHLH2* had decreased number of GnRH neurons in the hypothalamus. Targeted deletions in the *NHLH2* gene were associated with increased age at puberty and shorter reproductive lifespan (Johnson et al., 2004; Cogliati et al., 2007).

QTL and Candidate Genes for Age at Puberty in Other Livestock Species

Cattle

Bos indicus (e.g. Brahman) and Tropical Composite breeds are typically older at puberty (22 to 40 months with an average of 25 months) (Nogueira, 2004; Abeygunawardena and Dematawewa, 2004; Fortes et al., 2012) compared to *Bos taurus* breeds (12 to 15 months) (Thallman et al., 1999; Day and Nogueira, 2013). Selecting for early pubertal females, especially in late maturing tropical breeds is important to improve their herd reproductive performance.

A moderate heritability was reported for age at puberty in Australian Brahman ($h^2 = 0.56$) and Tropical Composite ($h^2 = 0.52$) heifers. Age at puberty was defined as the age at which the first corpus luteum was detected after frequent ovarian ultrasound scans (Fortes et al., 2012; Hawken et al., 2012). In a GWAS for age at puberty in Brahman, a major QTL was identified on BTA14 (22 to 28 Mb) (Fortes et al., 2012; Hawken et al., 2012). The top marker (*BTB-02056709*) located at 25 Mb explained 5.1% of the genetic variation for age at puberty. Another marker (*Hapmap46986-BTA-34282*) in the same region explained 4.7% and 2.3% of genetic variance in both Brahman and Tropical Composites, respectively (Hawken et al., 2012). The top QTL in Tropical Composites was mapped on BTA5 (96 Mb) and the top marker explained 4% of the genetic variation for age at puberty (Hawken et al., 2012).

The QTL on BTA14 was reported for various other bovine phenotypes (e.g. weight, stature, height, and prolificacy traits) suggesting pleiotropic effects associated with this region (Fortes et al., 2016). Two genes, *PENK* and *PLAG1* were potential

candidates for age at puberty in this region. *PENK* regulates the secretion of GnRH in mammals (Taylor et al., 2007). In sheep peripubertal progesterone surges during puberty transition increased the expression of *PENK* (Taylor et al., 2007). Canovas et al. (2014) also observed increased expression of *PENK* in the hypothalamus of heifers reaching puberty, suggesting a role of *PENK* in puberty onset mainly via regulating *GnRH* expression in the hypothalamus.

PLAG1 gene is associated with fertility, litter size, pre-and post-natal growth traits in mice (Hensen et al., 2004) and reduced average pause length in egg laying in chicken (Chen et al., 2007). Karim et al. (2011) identified a *G* to *C* nucleotide substitution (*rs109231213*) located in the 3' UTR of *PLAG1* gene associated with bovine stature. The SNP was also responsible for differential expression of *PLAG1*. Fortes et al. (2013a) evaluated the effect of this SNP on fertility, carcass, and feed efficiency traits in *Bos taurus* (no age at puberty data available), *Bos indicus*, and Tropical Composite cattle. The *C* allele was associated with late age at puberty in *Bos indicus* (+38 days) and Tropical composites (+25 days), but also with increased hip height, and weight. The *C* allele frequency was intermediate in *Bos indicus* (0.52) and Tropical Composites (0.68), but near fixation in *Bos taurus* (0.96). Less heterozygosity observed in this region especially in *Bos indicus* indicates a recent, strong selection for this mutation. Due to its favorable effects on growth traits, it can be speculated that heifers with a larger frame size were selected as replacements, thus, increased the frequency of the *C* allele, despite its unfavorable effects on age at puberty.

Canovas et al. (2014) characterized the transcriptome of five bovine tissues related to reproduction (hypothalamus, pituitary gland, ovary, uterus, and endometrium)

in pre- (n = 4) and post- (n = 4) pubertal Brangus (5/8 Angus × 3/8 Brahman) heifers. The largest number of upregulated genes (204 out of 275 genes) was observed in the post puberty hypothalamus. The most significant DEGs in the hypothalamus were *AVP* and *OXT*, encoding hormones involved in complex sexual and maternal behavior. The most significant DEGs in uterus and endometrium were *TDGF1* and *PENK*, both involved in GnRH regulation, estrus cycle, embryonic development, and early pregnancy. These genes were also upregulated in hypothalamus and pituitary of post pubertal heifers. *SIX6* and *PROP1* were differentially expressed in the endometrium and hypothalamus and were involved in regulating GnRH secretion, pituitary development, and expression of reproductive hormones (LH, FSH, prolactin etc.) (Canovas et al., 2014). *SIX6* is part of the *SIX* family, an important group of transcription factors involved in developmental processors and tissue differentiation. Differential expression was also observed for other *SIX* family members (e.g. *SIX3* and *SIX5*) in Brahman, Brangus, or both breeds (Fortes et al., 2016). Integrating transcriptomic data with GWAS uncovered seven DEGs (*MGC157266*, *C10H11ORF46*, *PENK*, *ELF5*, *FAM19A4*, *CPNE5*, and *MMD2*) overlapping QTL for age at puberty (Canovas et al., 2014). Candidate polymorphisms potentially associated with the observed effects were identified within the majority of these genes.

The IGF1 pathway plays a major role in initiating puberty onset and other reproductive functions, mainly through regulating GnRH neurons in the hypothalamus (Fortes et al., 2013b). In a candidate gene approach, SNPs were identified in *IGF* pathway genes associated with age at puberty in both Brahman and Tropical Composite breeds (Fortes et al., 2013b). A SNP in the *IGF1R* gene had the strongest effect ($P <$

0.00009) explaining 2% of the genetic variation of age at puberty with an estimated effect of 49 days in Brahmans. Two other SNPs in *IGFIR* were found in Brahmans and Tropical Composites associated with age at puberty.

Sheep

Haldar et al. (2013) used whole genome sequencing and candidate gene approach to identify genes and polymorphisms associated with age at puberty in New Zealand Davisdale sheep. Ewes usually reach puberty between five to 12 months depending on the breed and season of birth (Haldar et al., 2013). Four rams representing extremes for their daughter's age at puberty were selected for whole genome sequencing. Two non-conservative missense SNPs (*R62C* and *P1019S*) in *LEPR* gene were identified affecting age at puberty in ewes ($P < 0.001$). The *R62C* SNP was located near two cysteine residues that were highly conserved across species. The *P1019S* was also located at a relatively conserved region. Approximately 93% of the ewes homozygous for the wild type allele for both SNPs (*R62* and *P1019*) attained puberty before 1 year of age where as only 70% of the ewes homozygous for the mutant alleles (*C62* and *S1019*) attained puberty before 1 year of age. Ewes that were homozygous for the *C62* and *S1019* alleles, but attained puberty before 1 year of age were 17 days older at puberty compared to their wild type counterparts. Leptin is synthesized in the white adipocyte tissues (Garcia et al., 2002). Once the females reach an optimum body fat mass threshold, leptin signals the hypothalamus to initiate puberty. As a result, increased level of leptin has been observed during puberty in livestock species (Garcia et al., 2002; Rosales Nieto et al., 2013).

QTL and Candidate Genes for Age at Menarche in Humans

Female reproductive cycle in humans starts with the first menstrual bleeding known as menarche. In a comprehensive study including data collected from 67 countries, age at menarche ranged from 12 to 16.2 years with a mean of 13.5 years (Thomas et al., 2001). Over the decades, the mean age at menarche has shown a steady decline from 13.3 years in the early 1900s to 12.4 years in 1980s in different ethnic groups in the United States (McDowell et al., 2007). A similar pattern was reported from other parts of the word including the Netherlands (Fredriks et al., 2000), Spain (Cabanes et al., 2009), Korea (Ahn et al., 2013), and Thailand (Jaruratanasirikul et al., 2014).

Age at menarche is an important indicator of subsequent reproductive events such as age at first conception (Sandler et al., 1984) and probability to become pregnant (Zhang et al., 2017). Women that reach menarche later than 14 years were less likely to become pregnant compared to women that reached menarche at 13 to 14 years (Zhang et al., 2017). Early age at menarche (< 12 years) is associated with several health risks including breast cancer (Peeters et al., 1995), high blood pressure and glucose intolerance (Remsberg et al., 2005), and cardiovascular diseases (Lakshman et al., 2009).

Candidate Gene Approach

Heritability estimates for age at menarche based on twin and familial studies were moderate to high (h^2 = 0.44 to 0.95) (Kaprio et al., 1995; Towne et al., 2005; Anderson et al., 2007) suggesting that genetic factors play a major role in the variation of this trait. Before the era of high throughput genotyping and deep DNA and RNA sequencing,

candidate gene approach was utilized to identify genetic sources associated with variation in menarche onset. Since hormones involved in the hypothalamus-pituitary-gonadal axis play a major role in the timing of menarche (Barbier, 2014), they were studied extensively to identify polymorphisms associated with menarche (Table 1.1).

Onset of menarche is initiated with the increased exposure of reproductive tissues to estrogen. This exposure is facilitated via two main estrogen receptors (ER), ERα and ERβ. Several ER gene polymorphisms were associated with age at menarche and other reproductive disorders in humans (Table 1.1, Stavrou et al., 2002). *CYP* gene family members play a major role in biosynthesis and metabolism of ovarian hormones (Kadlubar et al., 2001). Polymorphisms in two key genes (*CYP17* and *CYP19*) involved in biosynthesis of estrogen in the lipid precursor cells were associated with age at menarche (Table 1.1, Guo et al., 2006). *IGF1* is considered an ideal candidate gene for onset of menarche since it is involved in metabolic processes in the hypothalamic-pituitary-ovarian axis in mammals via regulation of reproductive hormones (Table 1.1). IGF1 is known to stimulate GnRH activity (Belgorosky and Rivarola, 1998), FSH-mediated production of estradiol and progesterone (Zhao et al., 2007), and GnRH and LH release in rats (Hiney et al., 1991). Vitamin D receptor (*VDR*) also plays a role in reproductive organ development (Yoshizawa et al., 1997). In mice, a disrupted *VDR* gene failed to form fully functional ovaries. Polymorphisms in *VDR* gene were known to affect age at menarche (Table 1.1; Kitagawa et al., 1998).

Genome-wide Association Studies

Application of high throughput genotyping and GWAS for age at menarche began

with four large-scale (n = 15,000 to 18,000) studies carried out in women of European ancestry (He et al., 2009; Ong et al., 2009; Perry et al., 2009; Sulem et al., 2009). All studies reported genetic variants located on HSA 6q21, in or near *LIN28B*, a gene associated with age at menarche. The SNP *rs314276* located within intron 2 of *LIN28B* explained 0.2% of the genetic variance and each *C* allele was associated with 0.12 years earlier expression of menarche ($P = 2.8 \times 10^{-10}$; Ong et al., 2009). Two *C*/*T* nucleotide substitutions (*rs314280* and *rs7759938*) located near *LIN28B* were also associated with age at menarche ($P = 1.8 \times 10^{-14}$ and $P = 7.0 \times 10^{-9}$, respectively) (Sulem et al., 2009; Perry et al., 2009). These variants were also associated with development of secondary sex characteristics (e.g. breast development and pubertal hair growth), height, and body mass index (Ong et al., 2009; Sulem et al., 2009).

The LIN28B is a small (<30kDa) RNA binding protein mainly expressed in embryonic stem cells to maintain their pluripotent state (Thornton and Gregory, 2012). The LIN28 members bind to *let-7* miRNA and regulate *Let-7* expression during embryonic development and stem cell differentiation. A transgenic mouse model overexpressing *LIN28A* exhibited increased body size and delayed expression of puberty suggesting the involvement of the *LIN28* family members on age at menarche.

He et al. (2009) and Perry et al. (2009) reported another intergenic region with a major signal for age at menarche located near HSA 9q31.2. The largest effect in this region was observed for *rs2090409*, the *A* allele having associated with approximately 0.1 years earlier onset of menarche. The closest gene to this SNP is *TMEM38B*. In mice this gene is highly expressed in brain and loss of function mutations have resulted in neonatal lethality (Perry et al., 2009). In humans, mutations in this gene cause bone

malfunctions such as osteogenesis imperfecta (Volodarsky et al., 2013; Rubinato et al., 2014).

Following the initial GWAS for age at menarche, Elk et al. (2010) carried out a meta-analysis of 32 association studies using women with European descent (n = 87,802). This study confirmed the previously reported two loci at HSA 6q21 and 9q31.2. In addition, 30 novel loci were identified at genome-wise significance level. Among them, a SNP (*rs1079866*) located approximately 250 Kb downstream of *INHBA* explained the largest association with age at menarche. The protein coded by *INHBA* is involved in the biosynthesis of Inhibin A hormone. During expression of puberty, ovarian granulosa cells increase the production of inhibin A to act on anterior pituitary and hypothalamus to inhibit the secretion of FSH and GnRH, respectively (Burger, 1993). Four additional loci (in or near *FTO*, *TRA2B*, *TMEM18*, and *SEC16B* genes) overlapped with previously reported loci for body mass index. Wang et al. (2006) reported a strong negative genetic correlation between age at menarche and obesity related traits (e.g. fat mass, body mass index). Elk et al. (2010) observed that for all the four common loci, alleles that increase body mass index were associated with earlier age at menarche confirming the direction of the correlation observed by Wang et al. (2006). Pathway analyses identified coenzyme A and fatty acid biosynthesis as biological processes related to timing of menarche (Elk et al., 2010).

Table 1.1: Significant SNPs located in candidate genes for age at menarche in humans

Gene	SNP/restriction site	Exon/intron	Genotype method	Effect on age at menarche	*P* value	Population	Sources
ERα	*XbaI (X/x*)	Intron 1	PCR-RFLP	*XX* genotype delayed menarche by 0.61 years compared to *xx* genotype	0.017	North-Western Greece	Stavrou et al., (2002)
	PvuII (P/p)	Intron 1		*PP* genotype delayed menarche by 0.24 years compared to *pp* genotype	0.24	North-Western Greece	Stavrou et al., (2002)
ERβ	*AluI (A/G*)	3' UTR	PCR-RFLP	*AA* genotype delayed menarche by 0.57 years compared to *AG* genotype	0.005	North-Western Greece	Stavrou et al., (2006)
ERα and *ERβ*	*XbaI (X/x*) *PvuII (P/p*) *AluI (A/G*)			*XX/PP* genotype at *ERα* and *AA* genotype at *ERβ* delayed menarche by 0.89 years compared to alternate homozygotes	0.042	North-Western Greece	Stavrou et al., (2006)

Gene	**SNP/restriction site**	**Exon/intron**	**Genotype method**	**Effect on age at menarche**	***P* value**	**Population**	**Sources**
CYP17	*MspAI* *(A1/A2)*	5' UTR	PCR-RFLP	*A1A1* genotype delayed menarche by 0.5 years compared to *A1A2* genotype	0.027	Japanese	Gorai et al., (2003)
CYP19	*rs2445761* (*A/G*)	Intron 1	*Illumina BeadArray* SNP genotyping	*GG* genotype delayed menarche by 1 year compared to *AA* genotype	1.2×10^{-6}	Caucasian	Guo et al., (2006)
IGF1	*rs6214* (***A***/*G*) *GA* haplotype	Exon 4	*Illumina BeadArray* SNP genotyping	*GA* haplotype carrying *A* allele at rs6214 delayed menarche by 0.3 years than non-carriers	0.024	Caucasian	Zhao et al., (2007)
VDR	*ApaI* *(A/a)*		PCR-RFLP	*aa* genotype delayed menarche by 0.4 years compared to *Aa* genotype	< 0.05	Japanese	Kitagawa et al., (1998)
ERα and *VDR*	*XbaI* *PvuII* *ApaI*			*PX* haplotype at *ERα* and *aa* genotype at *VDR* delayed menarche by 0.5 years	0.01	Chinese	Xu et al., (2005)

LITERATURE CITED

Abell, C. E., J. W. Mabry, J. C. M. Dekkers, and K. J. Stalder. 2013. Relationship between litters per sow per year sire breeding values and sire progeny means for farrowing rate, removal parity and lifetime born alive. J. Anim. Breed. Genet. 130:64-71. doi:10.1111/j.1439-0388.2012.01012.x

Abeygunawardena, H., and C. M. B. Dematawewa. 2004. Pre-pubertal and postpartum anestrus in tropical Zebu cattle. Anim. Rreprod. Sci. 82-83:373-387. doi:10.1016/j.anireprosci.2004.05.006

Ahn, J. H., S. W. Lim, B. S. Song, J. Seo, J. A. Lee, D. H. Kim, and J. S. Lim. 2013. Age at menarche in the Korean female: secular trends and relationship to adulthood body mass index. Ann. Pediatr. Endocrinol. Metab. 18:60-64. doi:10.6065/apem.2013.18.2.60

Anderson, C. A., D. L. Duffy, N. G. Martin, and P. M. Visscher. 2007. Estimation of variance components for age at menarche in twin families. Behav. Genet. 37:668-677. doi:10.1007/s10519-007-9163-2

Barbieri, R. L. 2014. The endocrinology of the menstrual cycle. Methods. Mol. Biol. 1154:145-169. doi:10.1007/978-1-4939-0659-8_7

Bazer, F. W., and G. A. Johnson. 2014. Pig blastocyst-uterine interactions. Differentiation. 87:52-65. doi:10.1016/j.diff.2013.11.005

Belgorosky, A., and M. A. Rivarola. 1998. Irreversible increase of serum IGF-1 and IGFBP-3 levels in GnRH-dependent precocious puberty of different etiologies: implications for the onset of puberty. Horm. Res. 49:226-232. doi:10.1159/000023176

Bergfelder-Drüing, S., C. Grosse-Brinkhaus, B. Lind, M. Erbe, K. Schellander, H. Simianer, and E. Tholen. 2015. A Genome-Wide Association Study in Large White and Landrace Pig Populations for Number Piglets Born Alive. PLoS ONE 10:e0117468. doi:10.1371/journal.pone.0117468

Bidanel, J.P. 2011. Biology and genetics of reproduction. In The Genetics of the Pig. 2nd Edition. Edited by Rothschild, M.F. and Ruvinsky, A. Wallingford, UK: CAB International; 218-241.

Bidanel, J. P., J. Gruand, and C. Legault. 1996. Genetic variability of age and weight at puberty, ovulation rate and embryo survival in gilts and relations with production traits. Genet. Sel. Evol. 28:103-115. doi:10.1186/1297-9686-28-1-103

Bidanel, J. P., A. Rosendo, N. Iannuccelli, J. Riquet, H. Gilbert, J. C. Caritez, Y. Billon, Y. Amigues, A. Prunier, and D. Milan. 2008. Detection of quantitative trait loci

for teat number and female reproductive traits in Meishan × Large White F2 pigs. Animal. 2:813-820. doi:10.1017/S1751731108002097

Burger, H. G. 1993. Evidence for a negative feedback role of inhibin in follicle stimulating hormone regulation in women. Hum. Reprod. 8 Suppl 2:129-132. doi: 10.1093/humrep/8.suppl_2.129

Cabanes, A., N. Ascunce, E. Vidal, M. Ederra, A. Barcos, N. Erdozain, V. Lope, and M. Pollán. 2009. Decline in age at menarche among Spanish women born from 1925 to 1962. BMC. Public. Health. 9:449. doi:10.1186/1471-2458-9-449

Caldwell, H. K., H.-J. Lee, A. H. Macbeth, and W. S. Young. 2008. Vasopressin: Behavioral Roles of an "Original" Neuropeptide. Prog. Neurobiol. 84:1-24. doi: 10.1016/j.pneurobio.2007.10.007

Canario, L., E. Cantoni, E. Le Bihan, J. C. Caritez, Y. Billon, J. P. Bidanel, and J. L. Foulley. 2006. Between-breed variability of stillbirth and its relationship with sow and piglet characteristics. J. Anim. Sci. 84:3185-3196. doi:10.2527/jas.2005-775

Cánovas, A., A. Reverter, K. L. DeAtley, R. L. Ashley, M. L. Colgrave, M. R. S. Fortes, A. Islas-Trejo, S. Lehnert, L. Porto-Neto, G. Rincón, G. A. Silver, W. M. Snelling, J. F. Medrano, and M. G. Thomas. 2014. Multi-Tissue Omics Analyses Reveal Molecular Regulatory Networks for Puberty in Composite Beef Cattle. PLoS ONE. 9:e102551. doi:10.1371/journal.pone.0102551

Chen, C. F., Y. L. Shiue, C. J. Yen, P. C. Tang, H. C. Chang, and Y. P. Lee. 2007. Laying traits and underlying transcripts, expressed in the hypothalamus and pituitary gland, that were associated with egg production variability in chickens. Theriogenology. 68:1305-1315. doi:10.1016/j.theriogenology.2007.08.032

Chen, P., T. J. Baas, J. W. Mabry, and K. J. Koehler. 2003. Genetic correlations between lean growth and litter traits in U.S. Yorkshire, Duroc, Hampshire, and Landrace pigs. J. Anim. Sci. 81:1700-1705. doi:10.2527/2003.8171700x

Cleveland, M. A., and J. M. Hickey. 2013. Practical implementation of cost-effective genomic selection in commercial pig breeding using imputation1. J. Anim. Sci. 91:3583-3592. doi:10.2527/jas.2013-6270

Cleveland, M. A., J. M. Hickey, and S. Forni. 2012. A common dataset for genomic analysis of livestock populations. G3 (Bethesda). 2:429-435. doi:10.1534/g3.111.001453

Cogliati, T., P. Delgado-Romero, E. R. Norwitz, J. Guduric-Fuchs, U. B. Kaiser, S. Wray, and I. R. Kirsch. 2007. Pubertal impairment in Nhlh2 null mice is associated with hypothalamic and pituitary deficiencies. Mol. Endocrinol. 21:3013-3027. doi:10.1210/me.2005-0337

Córdoba, S., I. Balcells, A. Castelló, C. Ovilo, J. L. Noguera, O. Timoneda, and A. Sánchez. 2015. Endometrial gene expression profile of pregnant sows with extreme phenotypes for reproductive efficiency. Sci. Rep. 5:14416. doi:10.1038/srep14416

Crawford, J. L., and K. P. McNatty. 2012. The ratio of growth differentiation factor 9: bone morphogenetic protein 15 mRNA expression is tightly co-regulated and differs between species over a wide range of ovulation rates. Mol. Cell. Endocrinol. 348:339-343. doi:10.1016/j.mce.2011.09.033

D'Allaire, S., T. E. Stein, and A. D. Leman. 1987. Culling patterns in selected Minnesota swine breeding herds. Can. J. Vet. Res. 51:506-512.

d'Anglemont de Tassigny, X., and W. H. Colledge. 2010. The Role of Kisspeptin Signaling in Reproduction. Physiology. 25:207-217. doi:10.1152/physiol.00009.2010

Day, M. L., and G. P. Nogueira. 2013. Management of age at puberty in beef heifers to optimize efficiency of beef production. Animal Frontiers. 3:6-11. doi:10.2527/af.2013-0027

Dekkers, J. C. 2007. Marker-assisted selection for commercial crossbred performance. J. Anim. Sci. 85:2104-2114. doi:10.2527/jas.2006-683

Demerath, E. W., C. T. Liu, N. Franceschini, G. Chen, J. R. Palmer, E. N. Smith, C. T. Chen, C. B. Ambrosone, A. M. Arnold, E. V. Bandera, G. S. Berenson, L. Bernstein, A. Britton, A. R. Cappola, C. S. Carlson, S. J. Chanock, W. Chen, Z. Chen, S. L. Deming, C. E. Elks, M. K. Evans, Z. Gajdos, B. E. Henderson, J. J. Hu, S. Ingles, E. M. John, K. F. Kerr, L. N. Kolonel, L. Le Marchand, X. Lu, R. C. Millikan, S. K. Musani, N. L. Nock, K. North, S. Nyante, M. F. Press, J. L. Rodriquez-Gil, E. A. Ruiz-Narvaez, N. J. Schork, S. R. Srinivasan, N. F. Woods, W. Zheng, R. G. Ziegler, A. Zonderman, G. Heiss, B. Gwen Windham, M. Wellons, S. S. Murray, M. Nalls, T. Pastinen, A. Rajkovic, J. Hirschhorn, L. Adrienne Cupples, C. Kooperberg, J. M. Murabito, and C. A. Haiman. 2013. Genome-wide association study of age at menarche in African-American women. Hum. Mol. Genet. 22:3329-3346. doi:10.1093/hmg/ddt181

Duittoz, A.H., Y. Tillet, D. Le Bourhis, and L. Schibler. 2016. The timing of puberty (oocyte quality and management). Anim. Reprod. 13:313-333. doi:10.21451/1984-3143-AR874

Elks, C. E., J. R. Perry, P. Sulem, D. I. Chasman, N. Franceschini, C. He, K. L. Lunetta, J. A. Visser, E. M. Byrne, D. L. Cousminer, D. F. Gudbjartsson, T. Esko, B. Feenstra, J. J. Hottenga, D. L. Koller, Z. Kutalik, P. Lin, M. Mangino, M. Marongiu, P. F. McArdle, A. V. Smith, L. Stolk, S. H. van Wingerden, J. H. Zhao, E. Albrecht, T. Corre, E. Ingelsson, C. Hayward, P. K. Magnusson, E. N. Smith, S. Ulivi, N. M. Warrington, L. Zgaga, H. Alavere, N. Amin, T. Aspelund, S. Bandinelli, I. Barroso, G. S. Berenson, S. Bergmann, H. Blackburn, E. Boerwinkle, J. E. Buring, F. Busonero, H. Campbell, S. J. Chanock, W. Chen, M. C. Cornelis, D. Couper, A. D. Coviello, P. d'Adamo, U. de Faire, E. J. de Geus, P. Deloukas, A. Doring, G. D. Smith, D. F. Easton, G. Eiriksdottir, V. Emilsson, J. Eriksson, L. Ferrucci, A. R. Folsom, T. Foroud, M. Garcia, P. Gasparini, F. Geller, C. Gieger, V. Gudnason, P. Hall, S. E. Hankinson, L. Ferreli, A. C. Heath, D. G. Hernandez, A. Hofman, F. B. Hu, T. Illig, M. R. Jarvelin, A. D. Johnson, D.

Karasik, K. T. Khaw, D. P. Kiel, T. O. Kilpelainen, I. Kolcic, P. Kraft, L. J. Launer, J. S. Laven, S. Li, J. Liu, D. Levy, N. G. Martin, W. L. McArdle, M. Melbye, V. Mooser, J. C. Murray, S. S. Murray, M. A. Nalls, P. Navarro, M. Nelis, A. R. Ness, K. Northstone, B. A. Oostra, M. Peacock, L. J. Palmer, A. Palotie, G. Pare, A. N. Parker, N. L. Pedersen, L. Peltonen, C. E. Pennell, P. Pharoah, O. Polasek, A. S. Plump, A. Pouta, E. Porcu, T. Rafnar, J. P. Rice, S. M. Ring, F. Rivadeneira, I. Rudan, C. Sala, V. Salomaa, S. Sanna, D. Schlessinger, N. J. Schork, A. Scuteri, A. V. Segre, A. R. Shuldiner, N. Soranzo, U. Sovio, S. R. Srinivasan, D. P. Strachan, M. L. Tammesoo, E. Tikkanen, D. Toniolo, K. Tsui, L. Tryggvadottir, J. Tyrer, M. Uda, R. M. van Dam, J. B. van Meurs, P. Vollenweider, G. Waeber, N. J. Wareham, D. M. Waterworth, M. N. Weedon, H. E. Wichmann, G. Willemsen, J. F. Wilson, A. F. Wright, L. Young, G. Zhai, W. V. Zhuang, L. J. Bierut, D. I. Boomsma, H. A. Boyd, L. Crisponi, E. W. Demerath, C. M. van Duijn, M. J. Econs, T. B. Harris, D. J. Hunter, R. J. Loos, A. Metspalu, G. W. Montgomery, P. M. Ridker, T. D. Spector, E. A. Streeten, K. Stefansson, U. Thorsteinsdottir, A. G. Uitterlinden, E. Widen, J. M. Murabito, K. K. Ong, and A. Murray. 2010. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. Nat. Genet. 42:1077-1085. doi:10.1038/ng.714

Engblom, L., N. Lundeheim, E. Strandberg, P. Schneider Mdel, A. M. Dalin, and K. Andersson. 2008. Factors affecting length of productive life in Swedish commercial sows. J. Anim. Sci. 86:432-441. doi:10.2527/jas.2007-0310

Engelhardt, H., B. A. Croy, and G. J. King. 1997. Role of uterine immune cells in early pregnancy in pigs. J. Reprod. Fertil. Suppl. 52:115-131.

Faust, M. A., O. W. Robison, and M. W. Tess. 1993. Genetic and economic analyses of sow replacement rates in the commercial tier of a hierarchical swine breeding structure. J. Anim. Sci. 71:1400-1406. doi:10.2527/1993.7161400x

Fernando, R.L., and M. Grossman. 1989. Marker assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 21:467–477. doi:10.1186/1297-9686-21-4-467

Fortes, M. R. S., K. Kemper, S. Sasazaki, A. Reverter, J. E. Pryce, W. Barendse, R. Bunch, R. McCulloch, B. Harrison, S. Bolormaa, Y. D. Zhang, R. J. Hawken, M. E. Goddard, and S. A. Lehnert. 2013a. Evidence for pleiotropism and recent selection in the PLAG1 region in Australian Beef cattle. Anim. Genet. 44:636-647. doi:10.1111/age.12075

Fortes, M. R. S., S. A. Lehnert, S. Bolormaa, C. Reich, G. Fordyce, N. J. Corbet, V. Whan, R. J. Hawken, and A. Reverter. 2012. Finding genes for economically important traits: Brahman cattle puberty. Anim. Prod. Sci. 52:143-150. doi:10.1071/AN11165

Fortes, M. R. S., Y. Li, E. Collis, Y. Zhang, and R. J. Hawken. 2013b. The IGF1 pathway genes and their association with age of puberty in cattle. Anim. Genet. 44:91-95. doi:10.1111/j.1365-2052.2012.02367.x

Fortes, M. R. S., L. T. Nguyen, L. R. Porto Neto, A. Reverter, S. S. Moore, S. A. Lehnert, and M. G. Thomas. 2016. Polymorphisms and genes associated with puberty in heifers. Theriogenology. 86:333-339. doi:10.1016/j.theriogenology.2016.04.046

Fredriks, A. M., S. van Buuren, R. J. F. Burgmeijer, J. F. Meulmeester, R. J. Beuker, E. Brugman, M. J. Roede, S. P. Verloove-Vanhorick, and J.-M. Wit. 2000. Continuing Positive Secular Growth Change in the Netherlands 1955–1997. Pediatr. Res. 47:316-323. doi:10.1203/00006450-200003000-00006

Fujii, J., K. Otsu, F. Zorzato, S. de Leon, V. K. Khanna, J. E. Weiler, P. J. O'Brien, and D. H. MacLennan. 1991. Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. Science. 253:448-451. doi:10.1126/science.1862346

Garcia, M. R., M. Amstalden, S. W. Williams, R. L. Stanko, C. D. Morrison, D. H. Keisler, S. E. Nizielski, and G. L. Williams. 2002. Serum leptin and its adipose gene expression during pubertal development, the estrous cycle, and different seasons in cattle. J. Anim. Sci. 80:2158-2167.

Gauster, M., U. Hiden, A. Blaschitz, S. Frank, U. Lang, G. Alvino, I. Cetin, G. Desoye, and C. Wadsack. 2007. Dysregulation of placental endothelial lipase and lipoprotein lipase in intrauterine growth-restricted pregnancies. J. Clin. Endocrinol. Metab. 92:2256-2263. doi:10.1210/jc.2006-2403

Gobrogge, K. L., Y. Liu, L. J. Young, and Z. Wang. 2009. Anterior hypothalamic vasopressin regulates pair-bonding and drug-induced aggression in a monogamous rodent. Proc. Natl. Acad. Sci. U. S. A. 106:19144-19149. doi:10.1073/pnas.0908620106

Grandinson, K. 2005. Genetic background of maternal behaviour and its relation to offspring survival. Livest. Prod. Sci. 93:43-50. doi:10.1016/j.livprodsci.2004.11.005

Griffin, L. D., and S. H. Mellon. 2001. Biosynthesis of the neurosteroid 3 alpha-hydroxy-4-pregnen-20-one (3 alpha hp), a specific inhibitor of FSH release. Endocrinology. 142:4617-4622. doi:10.1210/endo.142.11.8477

Guastavino, J. M., and K. Larsson. 1992. The staggerer gene curtails the reproductive life span of females. Behav. Genet. 22:101-112. doi:10.1007/bf01066795

Guo, X., G. Su, O. F. Christensen, L. Janss, and M. S. Lund. 2016. Genome-wide association analyses using a Bayesian approach for litter size and piglet mortality in Danish Landrace and Yorkshire pigs. BMC Genomics. 17:468. doi:10.1186/s12864-016-2806-z

Guo, Y., D. H. Xiong, T. L. Yang, Y. F. Guo, R. R. Recker, and H. W. Deng. 2006. Polymorphisms of estrogen-biosynthesis genes CYP17 and CYP19 may influence age at menarche: a genetic association study in Caucasian females. Hum. Mol. Genet. 15:2401-2408. doi:10.1093/hmg/ddl155

Guthrie, H. D., D. J. Bolt, and B. S. Cooper. 1990. Effects of gonadotropin treatment on ovarian follicle growth and granulosal cell aromatase activity in prepuberal gilts. J. Anim. Sci. 68:3719-3726. doi:10.2527/1990.68113719x

Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick. 2011. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics. 12:186. doi:10.1186/1471-2105-12-186

Haldar, A., M. C. French, R. Brauning, S. J. Edwards, A. R. O'Connell, P. A. Farquhar, G. H. Davis, P. D. Johnstone, and J. L. Juengel. 2014. Single-Nucleotide Polymorphisms in the LEPR Gene Are Associated with Divergent Phenotypes for Age at Onset of Puberty in Davisdale Ewes1. Biol. Reprod. 90:33 doi:10.1095/biolreprod.113.115923

Hawken, R. J., Y. D. Zhang, M. R. S. Fortes, E. Collis, W. C. Barris, N. J. Corbet, P. J. Williams, G. Fordyce, R. G. Holroyd, J. R. W. Walkley, W. Barendse, D. J. Johnston, K. C. Prayaga, B. Tier, A. Reverter, and S. A. Lehnert. 2012. Genome-wide association studies of female reproduction in tropically adapted beef cattle1. J. Anim. Sci. 90:1398-1410. doi:10.2527/jas.2011-4410

Hayes, B., and M. Goddard. 2010. Genome-wide association and genomic selection in animal breedingThis article is one of a selection of papers from the conference "Exploiting Genome-wide Association in Oilseed Brassicas: a model for genetic improvement of major OECD crops for sustainable farming". Genome. 53:876-883. doi:10.1139/G10-076

Hazeleger, W., N. M. Soede, and B. Kemp. 2005. The effect of feeding strategy during the pre-follicular phase on subsequent follicular development in the pig. Domest. Anim. Endocrin. 29:362-370. doi:10.1016/j.domaniend.2005.03.007

He, C., P. Kraft, C. Chen, J. E. Buring, G. Pare, S. E. Hankinson, S. J. Chanock, P. M. Ridker, D. J. Hunter, and D. I. Chasman. 2009. Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. Nat. Genet. 41:724-728. doi:10.1038/ng.385

Hensen, K., C. Braem, J. Declercq, F. Van Dyck, M. Dewerchin, L. Fiette, C. Denef, and W. J. Van de Ven. 2004. Targeted disruption of the murine Plag1 proto-oncogene causes growth retardation and reduced fertility. Dev. Growth. Differ. 46:459-470. doi:10.1111/j.1440-169x.2004.00762.x

Hiney, J. K., S. R. Ojeda, and W. L. Dees. 1991. Insulin-like growth factor I: a possible metabolic signal involved in the regulation of female puberty. Neuroendocrinology. 54:420-423. doi:10.1159/000125924

Hirooka, H., D. J. de Koning, B. Harlizius, J. A. van Arendonk, A. P. Rattink, M. A. Groenen, E. W. Brascamp, and H. Bovenhuis. 2001. A whole-genome scan for quantitative trait loci affecting teat number in pigs. J. Anim. Sci 79:2320-2326. doi:10.2527/2001.7992320x

Hoge, M. D., and R. O. Bates. 2011. Developmental factors that influence sow longevity. J. Anim. Sci. 89:1238-1245. doi:10.2527/jas.2010-3175

Holl, J. W., J. P. Cassady, D. Pomp, and R. K. Johnson. 2004. A genome scan for quantitative trait loci and imprinted regions affecting reproduction in pigs. J. Anim. Sci. 82:3421-3429. doi:10.2527/2004.82123421x

Holm, B., M. Bakken, G. Klemetsdal, and O. Vangen. 2004. Genetic correlations between reproduction and production traits in swine. J. Anim. Sci. 82:3458-3464. doi:10.2527/2004.82123458x

Hsu, W.L. 2011. Analysis of long-term selection (28 generations) for reproduction, growth, and carcass traits in swine. PhD dissertation. University of Nebraska – Lincoln, Animal Science Department.

Hsu, W. L., and R. K. Johnson. 2014. Analysis of 28 generations of selection for reproduction, growth, and carcass traits in swine. J. Anim. Sci. 92:4806-4822. doi:10.2527/jas.2014-8125

Hu, Z.-L., C. A. Park, X.-L. Wu, and J. M. Reecy. 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic. Acids. Res. 41:D871-D879. doi:10.1093/nar/gks1150

Huang, L., Z. J. Yin, Y. F. Feng, X. D. Zhang, T. Wu, Y. Y. Ding, P. F. Ye, K. Fu, and M. Q. Zhang. 2016. Identification and differential expression of microRNAs in the ovaries of pigs (Sus scrofa) with high and low litter sizes. Anim. Genet. 47:543-551. doi:10.1111/age.12452

Hwang, J. H., S. M. An, S. Kwon, D. H. Park, T. W. Kim, D. G. Kang, G. E. Yu, I.-S. Kim, H. C. Park, J. Ha, and C. W. Kim. 2017. DNA methylation patterns and gene expression associated with litter size in Berkshire pig placenta. PLOS ONE 12:e0184539. doi:10.1371/journal.pone.0184539

Ibánẽz-Escriche, N., R. L. Fernando, A. Toosi, and J. C. M. Dekkers. 2009. Genomic selection of purebreds for crossbred performance. Genet. Sel. Evol. 41:12. doi:10.1186/1297-9686-41-12

Jaruratanasirikul, S., A. Chanpong, N. Tassanakijpanich, and H. Sriplung. 2014. Declining age of puberty of school girls in southern Thailand. World. J. Pediatr. 10:256-261. doi:10.1007/s12519-014-0472-2

Johnson, R. K., M. K. Nielsen, and D. S. Casey. 1999. Responses in ovulation rate, embryonal survival, and litter traits in swine to 14 generations of selection to increase litter size. J. Anim. Sci. 77:541-557. doi:10.2527/1999.773541x

Johnson, R. K., D. R. Zimmerman, and R. J. Kittok. 1984. Selection for components of reproduction in swine. Livest. Prod. Sci. 11:541-558. doi:10.1016/0301-6226(84)90070-8

Johnson, S. A., C. L. Marin-Bivens, M. Miele, C. A. Coyle, R. Fissore, and D. J. Good. 2004. The Nhlh2 transcription factor is required for female sexual behavior and

reproductive longevity. Horm. Behav. 46:420-427. doi:10.1016/j.yhbeh.2004.03.006

Kachman, S. D. 2015. Genomic prediction model based on haplotype clusters. In: Joint Stat. Meet., August 11, 2015, Seattle, WA. https://ww2.amstat.org/meetings/jsm/2015/onlinepro-gram/AbstractDetails.cfm?abstractid=314884 (Accessed 25 February 2016.)

Kadlubar, F., G. Bcrkowitz, R. R. Delongchamp, B. L. Green, C. Wang, and M. S. Wolff. 2001. The putative high activity variant *CYP3A4*1B* predicts the onset of puberty in young girls. Phys. Chem. Chem. Phys.. 159:216.

Kaprio, J., A. Rimpelä, T. Winter, R. J. Viken, M. Rimpelä, and R. J. Rose. 1995. Common genetic influences on BMI and age at menarche. Hum. Biol. 67:739-753.

Karim, L., H. Takeda, L. Lin, T. Druet, J. A. C. Arias, D. Baurain, N. Cambisano, S. R. Davis, F. Farnir, B. Grisart, B. L. Harris, M. D. Keehan, M. D. Littlejohn, R. J. Spelman, M. Georges, and W. Coppieters. 2011. Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. Nat. Genet. 43:405-413. doi:10.1038/ng.814

Kennedy, T. G., C. Gillio-Meina, and S. H. Phang. 2007. Prostaglandins and the initiation of blastocyst implantation and decidualization. Reproduction. 134:635-643. doi:10.1530/rep-07-0328

Kitagawa, I., Y. Kitagawa, Y. Kawase, T. Nagaya, and S. Tokudome. 1998. Advanced onset of menarche and higher bone mineral density depending on vitamin D receptor gene polymorphism. Eur. J. Endocrinol. 139:522-527.

Klindt, J., and R. T. Stone. 1984. Porcine growth hormone and prolactin: concentrations in the fetus and secretory patterns in the growing pig. Growth. 48:1-15.

Knauer, M., K. J. Stalder, T. Serenius, T. J. Baas, P. J. Berger, L. Karriker, R. N. Goodwin, R. K. Johnson, J. W. Mabry, R. K. Miller, O. W. Robison, and M. D. Tokach. 2010a. Factors associated with sow stayability in 6 genotypes. J. Anim. Sci. 88:3486-3492. doi:10.2527/jas.2009-2319

Knauer, M. T., J. P. Cassady, D. W. Newcom, and M. T. See. 2010b. Estimates of variance components for genetic correlations among swine estrus traits. J. Anim. Sci. 88:2913-2919. doi:10.2527/jas.2009-2639

Knauer, M. T., J. P. Cassady, D. W. Newcom, and M. T. See. 2011. Phenotypic and genetic correlations between gilt estrus, puberty, growth, composition, and structural conformation traits with first-litter reproductive measures J. Anim. Sci. 89:935-942. doi:10.2527/jas.2009-2673

Knauff, E. A., L. Franke, M. A. van Es, L. H. van den Berg, Y. T. van der Schouw, J. S. Laven, C. B. Lambalk, A. Hoek, A. J. Goverde, S. Christin-Maitre, A. J. Hsueh, C. Wijmenga, and B. C. Fauser. 2009. Genome-wide association study in

premature ovarian failure patients suggests ADAMTS19 as a possible candidate gene. Hum. Reprod. 24:2372-2378. doi:10.1093/humrep/dep197

Knol, E. F., B. Nielsen, and P. W. Knap. 2016. Genomic selection in commercial pig breeding. Animal Frontiers. 6:15-22. doi:10.2527/af.2016-0003

Knox, R. V. 2005. Recruitment and selection of ovarian follicles for determination of ovulation rate in the pig. Domest. Anim. Endocrin. 29:385-397. doi:10.1016/j.domaniend.2005.02.025

Kuehn, L. A., D. J. Nonneman, J. M. Klindt, and T. H. Wise. 2009. Genetic relationships of body composition, serum leptin, and age at puberty in gilts12. J. Anim. Sci. 87:477-483. doi:10.2527/jas.2008-0936

Kukurba, K. R., and S. B. Montgomery. 2015. RNA Sequencing and Analysis. Cold. Spring. Harb. Protoc. 2015:pdb.top084970. doi:10.1101/pdb.top084970

Kwon, S. G., J. H. Hwang, D. H. Park, T. W. Kim, D. G. Kang, K. H. Kang, I.-S. Kim, H. C. Park, C.-S. Na, J. Ha, and C. W. Kim. 2016. Identification of Differentially Expressed Genes Associated with Litter Size in Berkshire Pig Placenta. PLoS ONE. 11:e0153311. doi:10.1371/journal.pone.0153311

Lakshman, R., N. G. Forouhi, S. J. Sharp, R. Luben, S. A. Bingham, K. T. Khaw, N. J. Wareham, and K. K. Ong. 2009. Early age at menarche associated with cardiovascular disease and mortality. J. Clin. Endocrinol. Metab. 94:4953-4960. doi:10.1210/jc.2009-1789

Lamberson, W. R., and T. J. Safranski. 2000. A model for economic comparison of swine insemination programs. Theriogenology. 54:799-807. doi:10.1016/s0093-691x(00)00391-5

Li, S., J. Ren, G. Yang, Y. Guo, and L. Huang. 2008. Characterization of the porcine Kisspeptins receptor gene and evaluation as candidate for timing of puberty in sows. J. Anim. Breed. Genet. 125:219-227. doi:10.1111/j.1439-0388.2008.00732.x

Lillehammer, M., T. H. E. Meuwissen, and A. K. Sonesson. 2011. Genomic selection for maternal traits in pigs. J. Anim. Sci. 89:3908-3916. doi:10.2527/jas.2011-4044

Lillehammer, M., T. H. E. Meuwissen, and A. K. Sonesson. 2013. Genomic selection for two traits in a maternal pig breeding scheme1. J. Anim. Sci. 91:3079-3087. doi:10.2527/jas.2012-5113

Lin, H., H. Wang, Y. Wang, C. Liu, C. Wang, and J. Guo. 2015. Transcriptomic Analysis of the Porcine Endometrium during Embryo Implantation. Genes. 6:1330-1346. doi:10.3390/genes6041330

Lopes, M. S., H. Bovenhuis, A. M. Hidalgo, J. A. M. van Arendonk, E. F. Knol, and J. W. M. Bastiaansen. 2017. Genomic selection for crossbred performance accounting for breed-specific effects. Genet. Sel. Evol. 49:51-51. doi:10.1186/s12711-017-0328-z

Lucot, K. L., M. L. Spangler, M. D. Trenhaile, S. D. Kachman, and D. C. Ciobanu. 2015. Evaluation of reduced subsets of single nucleotide polymorphisms for the prediction of age at puberty in sows. Anim. Genet. 46:403-409. doi:10.1111/age.12310

McDowell, M. A., D. J. Brody, and J. P. Hughes. 2007. Has Age at Menarche Changed? Results from the National Health and Nutrition Examination Survey (NHANES) 1999–2004. J. Adolescent. Health. 40:227-231. doi:10.1016/j.jadohealth.2006.10.002

Messerschmidt, D. M., B. B. Knowles, and D. Solter. 2014. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Gene. Dev. 28:812-828. doi:10.1101/gad.234294.113

Meszaros, G., J. Palos, V. Ducrocq, and J. Solkner. 2010. Heritability of longevity in Large White and Landrace sows using continuous time and grouped data models. Genet. Sel. Evol. 42:13. doi:10.1186/1297-9686-42-13

Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157:1819-1829.

Meuwissen, T. H. E. 2009. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. Genet. Sel. Evol. 41:35. doi:10.1186/1297-9686-41-35

Miller, P. S., R. Moreno, and R. K. Johnson. 2011. Effects of restricting energy during the gilt developmental period on growth and reproduction of lines differing in lean growth rate: responses in feed intake, growth, and age at puberty. J. Anim. Sci. 89:342-354. doi:10.2527/jas.2010-3111

Noguchi, M., K. Yoshioka, S. Itoh, C. Suzuki, S. Arai, Y. Wada, Y. Hasegawa, and H. Kaneko. 2010. Peripheral concentrations of inhibin A, ovarian steroids, and gonadotropins associated with follicular development throughout the estrous cycle of the sow. Reproduction. 139:153-161. doi:10.1530/rep-09-0018

Nogueira, G. P. 2004. Puberty in South American Bos indicus (Zebu) cattle. Anim. Rreprod. Sci. 82-83:361-372. doi:10.1016/j.anireprosci.2004.04.007

Nonneman, D., C. Lents, G. Rohrer, L. Rempel, and J. Vallet. 2014. Genome-wide association with delayed puberty in swine. Anim. Genet. 45:130-132. doi:10.1111/age.12087

Nonneman, D. J., J. F. Schneider, C. A. Lents, R. T. Wiedmann, J. L. Vallet, and G. A. Rohrer. 2016. Genome-wide association and identification of candidate genes for age at puberty in swine. BMC. Genet. 17:50. doi:10.1186/s12863-016-0352-y

Nonneman, D. J., T. H. Wise, J. J. Ford, L. A. Kuehn, and G. A. Rohrer. 2006. Characterization of the aldo-keto reductase 1C gene cluster on pig chromosome 10: possible associations with reproductive traits. BMC. Vet. Res. 2:28. doi:10.1186/1746-6148-2-28

Ong, K. K., C. E. Elks, S. Li, J. H. Zhao, J. Luan, L. B. Andersen, S. A. Bingham, S. Brage, G. D. Smith, U. Ekelund, C. J. Gillson, B. Glaser, J. Golding, R. Hardy, K. T. Khaw, D. Kuh, R. Luben, M. Marcus, M. A. McGeehin, A. R. Ness, K. Northstone, S. M. Ring, C. Rubin, M. A. Sims, K. Song, D. P. Strachan, P. Vollenweider, G. Waeber, D. M. Waterworth, A. Wong, P. Deloukas, I. Barroso, V. Mooser, R. J. Loos, and N. J. Wareham. 2009. Genetic variation in LIN28B is associated with the timing of puberty. Nat. Genet. 41:729-733. doi:10.1038/ng.382

Onteru, S. K., B. Fan, Z.-Q. Du, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2012. A whole-genome association study for pig reproductive traits. Anim. Genet. 43:18-26. doi:10.1111/j.1365-2052.2011.02213.x

Onteru, S. K., B. Fan, M. T. Nikkila, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2011. Whole-genome association analyses for lifetime reproductive traits in the pig. J. Anim. Sci. 89:988-995. doi:10.2527/jas.2010-3236

Ozsolak, F., and P. M. Milos. 2011. RNA sequencing: advances, challenges and opportunities. Nat. Rev. Genet. 12:87-98. doi:10.1038/nrg2934

Patterson, J. L., E. Beltranena, and G. R. Foxcroft. 2010. The effect of gilt age at first estrus and breeding on third estrus on sow body weight changes and long-term reproductive performance. J. Anim. Sci. 88:2500-2513. doi:10.2527/jas.2008-1756

Peeters, P. H., A. L. Verbeek, A. Krol, M. M. Matthyssen, and F. de Waard. 1995. Age at menarche and breast cancer risk in nulliparous women. Breast. Cancer. Res. Tr. 33:55-61.

Perry, J. R., L. Stolk, N. Franceschini, K. L. Lunetta, G. Zhai, P. F. McArdle, A. V. Smith, T. Aspelund, S. Bandinelli, E. Boerwinkle, L. Cherkas, G. Eiriksdottir, K. Estrada, L. Ferrucci, A. R. Folsom, M. Garcia, V. Gudnason, A. Hofman, D. Karasik, D. P. Kiel, L. J. Launer, J. van Meurs, M. A. Nalls, F. Rivadeneira, A. R. Shuldiner, A. Singleton, N. Soranzo, T. Tanaka, J. A. Visser, M. N. Weedon, S. G. Wilson, V. Zhuang, E. A. Streeten, T. B. Harris, A. Murray, T. D. Spector, E. W. Demerath, A. G. Uitterlinden, and J. M. Murabito. 2009. Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. Nat. Genet. 41:648-650. doi:10.1038/ng.386

Remsberg, K. E., E. W. Demerath, C. M. Schubert, W. C. Chumlea, S. S. Sun, and R. M. Siervogel. 2005. Early menarche and the development of cardiovascular disease risk factors in adolescent girls: the Fels Longitudinal Study. J. Clin. Endocrinol. Metab. 90:2718-2724. doi:10.1210/jc.2004-1991

Rodriguez-Zas, S. L., B. R. Southey, R. V. Knox, J. F. Connor, J. F. Lowe, and B. J. Roskamp. 2003. Bioeconomic evaluation of sow longevity and profitability. J. Anim. Sci. 81:2915-2922. doi:10.2527/2003.81122915x

Rohrer, G. A., J. J. Ford, T. H. Wise, J. L. Vallet, and R. K. Christenson. 1999. Identification of quantitative trait loci affecting female reproductive traits in a

multigeneration Meishan-White composite swine population1. J. Anim. Sci. 77:1385-1391. doi:10.2527/1999.7761385x

Rosales Nieto, C. A., M. B. Ferguson, C. A. Macleay, J. R. Briegel, G. B. Martin, and A. N. Thompson. 2013. Selection for superior growth advances the onset of puberty and increases reproductive performance in ewe lambs. Animal. 7:990-997. doi:10.1017/s1751731113000074

Rothschild, M. F. and J. P. Bidanel. 1998. Biology and genetics of reproduction. In: Rothschild, M. F. and A. Ruvinsky (eds). The genetics of the pig. CAB International, Wallingford, UK. 313-343.

Rothschild, M., C. Jacobson, D. Vaske, C. Tuggle, L. Wang, T. Short, G. Eckardt, S. Sasaki, A. Vincent, D. McLaren, O. Southwood, H. van der Steen, A. Mileham, and G. Plastow. 1996. The estrogen receptor locus is associated with a major gene influencing litter size in pigs. Proc. Natl. Acad. Sci. U. S. A. 93:201-205. doi:10.1073/pnas.93.1.201

Rubinato, E., A. Morgan, A. D'Eustacchio, V. Pecile, G. Gortani, P. Gasparini, and F. Faletra. 2014. A novel deletion mutation involving TMEM38B in a patient with autosomal recessive osteogenesis imperfecta. Gene. 545:290-292. doi:10.1016/j.gene.2014.05.028

Ruiz-Flores, A., and R. K. Johnson. 2001. Direct and correlated responses to two-stage selection for ovulation rate and number of fully formed pigs at birth in swine. J. Anim. Sci. 79:2286-2297. doi:10.2527/2001.7992286x

Rybak, A., H. Fuchs, L. Smirnova, C. Brandt, E. E. Pohl, R. Nitsch, and F. G. Wulczyn. 2008. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. Nat. Cell. Biol. 10:987-993. doi:10.1038/ncb1759

Samorè, A. B., and L. Fontanesi. 2016. Genomic selection in pigs: state of the art and perspectives. Ital. J. Anim. Sci. 15:211-232. doi:10.1080/1828051X.2016.1172034

Sanchez, M. P., J. Riquet, N. Iannuccelli, J. Gogue, Y. Billon, O. Demeure, J. C. Caritez, G. Burgaud, K. Feve, M. Bonnet, C. Pery, H. Lagant, P. Le Roy, J. P. Bidanel, and D. Milan. 2006. Effects of quantitative trait loci on chromosomes 1, 2, 4, and 7 on growth, carcass, and meat quality traits in backcross Meishan x Large White pigs. J. Anim. Sci. 84:526-537. doi:10.2527/2006.843526x

Sandler, D. P., A. J. Wilcox, and L. F. Horney. 1984. Age at menarche and subsequent reproductive events. Am. J. Epidemiol. 119:765-774. doi:10.1093/oxfordjournals.aje.a113797

Schneider, J. F., D. J. Nonneman, R. T. Wiedmann, J. L. Vallet, and G. A. Rohrer. 2014. Genomewide association and identification of candidate genes for ovulation rate in swine12. J. Anim. Sci. 92:3792-3803. doi:10.2527/jas.2014-7788

Schneider, J. F., L. A. Rempel, and G. A. Rohrer. 2012a. Genome-wide association study of swine farrowing traits. Part I: Genetic and genomic parameter estimates. J. Anim. Sci. 90:3353-3359. doi: 10.2527/jas.2011-4729

Schneider, J. F., L. A. Rempel, G. A. Rohrer, and T. M. Brown-Brandl. 2011. Genetic parameter estimates among scale activity score and farrowing disposition with reproductive traits in swine. J. Anim. Sci. 89:3514-3521. doi:10.2527/jas.2010-3510

Schneider, J. F., L. A. Rempel, W. M. Snelling, R. T. Wiedmann, D. J. Nonneman, and G. A. Rohrer. 2012b. Genome-wide association study of swine farrowing traits. Part II: Bayesian analysis of marker data. J. Anim. Sci. 90:3360-3367. doi:10.2527/jas.2011-4759

Sechman, A., P. Antos, D. Katarzynska, A. Grzegorzewska, D. Wojtysiak, and A. Hrabia. 2014. Effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin on secretion of steroids and STAR, HSD3B and CYP19A1 mRNA expression in chicken ovarian follicles. Toxicol. Lett. 225:264-274. doi:10.1016/j.toxlet.2013.12.021

Serenius, T., and K. J. Stalder. 2006. Selection for sow longevity. J. Anim. Sci. 84 Suppl:E166-171.

Serenius, T., and K. J. Stalder. 2007. Length of productive life of crossbred sows is affected by farm management, leg conformation, sow's own prolificacy, sow's origin parity and genetics. Animal. 1:745-750. doi:10.1017/s175173110769185x

Serenius, T., K. J. Stalder, T. J. Baas, J. W. Mabry, R. N. Goodwin, R. K. Johnson, O. W. Robison, M. Tokach, and R. K. Miller. 2006. National Pork Producers Council Maternal Line National Genetic Evaluation Program: a comparison of sow longevity and trait associations with sow longevity. J. Anim. Sci. 84:2590-2595. doi:10.2527/jas.2005-499

Serenius, T., K. J. Stalder, and R. L. Fernando. 2008. Genetic associations of sow longevity with age at first farrowing, number of piglets weaned, and wean to insemination interval in the Finnish Landrace swine population. J. Anim. Sci. 86:3324-3329. doi:10.2527/jas.2007-0484

Short, T. H., M. F. Rothschild, O. I. Southwood, D. G. McLaren, A. de Vries, H. van der Steen, G. R. Eckardt, C. K. Tuggle, J. Helm, D. A. Vaske, A. J. Mileham, and G. S. Plastow. 1997. Effect of the estrogen receptor locus on reproduction and production traits in four commercial pig lines. J. Anim. Sci. 75:3138-3142. doi:10.2527/1997.75123138x

Smith, Z. D., and A. Meissner. 2013. DNA methylation: roles in mammalian development. Nat. Rev. Genet. 14:204-220. doi:10.1038/nrg3354

Soede, N., W. Hazeleger, and B. Kemp. 1998. Follicle Size and the Process of Ovulation in Sows as Studied with Ultrasound. Reprod. Domest. Anim. 33:239-244. doi:10.1111/j.1439-0531.1998.tb01350.x

Soede, N. M., F. A. Helmond, and B. Kemp. 1994. Periovulatory profiles of oestradiol, LH and progesterone in relation to oestrus and embryo mortality in multiparous sows using transrectal ultrasonography to detect ovulation. J. Reprod. Fertil. 101:633-641.

Soede, N. M., P. Langendijk, and B. Kemp. 2011. Reproductive cycles in pigs. Anim. Rreprod. Sci. 124:251-258. doi:10.1016/j.anireprosci.2011.02.025

Spiliotis, B. E. 2003. Growth hormone insufficiency and its impact on ovarian function. Ann. N. Y. Acad. Sci. 997:77-84. doi:10.1196/annals.1290.009

Stalder, K.J., C. Lacy, C.L. Cross, and G.E. Conatser. 2003. Financial impact of average parity of culled females in a breed-to-wean swine operation using replacement gilt net present value analysis. J. Swine. Health. Prod. 11:69-74.

Stavrou, I., C. Zois, J. P. Ioannidis, and A. Tsatsoulis. 2002. Association of polymorphisms of the oestrogen receptor alpha gene with the age of menarche. Hum. Reprod.17:1101-1105.

Sterning, M., L. Rydhmer, and L. Eliasson-Selling. 1998. Relationships between age at puberty and interval from weaning to estrus and between estrus signs at puberty and after the first weaning in pigs. J. Anim. Sci. 76:353-359. doi:10.2527/1998.762353x

Sulem, P., D. F. Gudbjartsson, T. Rafnar, H. Holm, E. J. Olafsdottir, G. H. Olafsdottir, T. Jonsson, P. Alexandersen, B. Feenstra, H. A. Boyd, K. K. Aben, A. L. M. Verbeek, N. Roeleveld, A. Jonasdottir, U. Styrkarsdottir, V. Steinthorsdottir, A. Karason, S. N. Stacey, J. Gudmundsson, M. Jakobsdottir, G. Thorleifsson, G. Hardarson, J. Gulcher, A. Kong, L. A. Kiemeney, M. Melbye, C. Christiansen, L. Tryggvadottir, U. Thorsteinsdottir, and K. Stefansson. 2009. Genome-wide association study identifies sequence variants on 6q21 associated with age at menarche. Nat. Genet. 41:734. doi:10.1038/ng.383

Tantasuparuk, W., N. Lundeheim, A. M. Dalin, A. Kunavongkrit, and S. Einarsson. 2001. Weaning-to-service interval in primiparous sows and its relationship with longevity and piglet production. Livest. Prod. Sci. 69:155-162. doi:10.1016/S0301-6226(00)00256-6

Tart, J. K., R. K. Johnson, J. W. Bundy, N. N. Ferdinand, A. M. McKnite, J. R. Wood, P. S. Miller, M. F. Rothschild, M. L. Spangler, D. J. Garrick, S. D. Kachman, and D. C. Ciobanu. 2013. Genome-wide prediction of age at puberty and reproductive longevity in sows. Anim. Genet. 44:387-397. doi:10.1111/age.12028

Taylor, J. A., M.-L. Goubillon, K. D. Broad, and J. E. Robinson. 2007. Steroid Control of Gonadotropin-Releasing Hormone Secretion: Associated Changes in Pro-Opiomelanocortin and Preproenkephalin Messenger RNA Expression in the Ovine Hypothalamus1. Biol. Reprod. 76:524-531. doi:10.1095/biolreprod.106.055533

Tesfaye, D., D. Worku, F. Rings, C. Phatsara, E. Tholen, K. Schellander, and M. Hoelker. 2009. Identification and expression profiling of microRNAs during bovine oocyte maturation using heterologous approach. Mol. Reprod. Dev. 76:665-677. doi:10.1002/mrd.21005

Thallman, R. M., L. V. Cundiff, K. E. Gregory, and R. M. Koch. 1999. Germplasm evaluation in beef cattle--Cycle IV: postweaning growth and puberty of heifers. J. Anim. Sci. 77:2651-2659. doi:10.2527/1999.77102651x

Thomas, F., F. Renaud, E. Benefice, T. de Meeus, and J. F. Guegan. 2001. International variability of ages at menarche and menopause: patterns and main determinants. Hum. Biol. 73:271-290.

Thornton, J. E., and R. I. Gregory. 2012. How does Lin28 let-7 control development and disease? Trends. Cell. Biol. 22:474-482. doi:10.1016/j.tcb.2012.06.001

TOMIYAMA, M., S. KUBO, T. TAKAGI, and K. SUZUKI. 2011. Evaluation of genetic trends and determination of the optimal number of cumulative records of parity required in reproductive traits in a Large White pig population. Anim. Sci. J. 82:621-626. doi:10.1111/j.1740-0929.2011.00889.x

Toosi, A., R. L. Fernando, and J. C. Dekkers. 2010. Genomic selection in admixed and crossbred populations. J. Anim. Sci. 88:32-46. doi:10.2527/jas.2009-1975

Towne, B., S. A. Czerwinski, E. W. Demerath, J. Blangero, A. F. Roche, and R. M. Siervogel. 2005. Heritability of age at menarche in girls from the Fels Longitudinal Study. Am. J. Phys. Anthropol. 128:210-219. doi:10.1002/ajpa.20106

Turner, A. I., and A. J. Tilbrook. 2006. Stress, cortisol and reproduction in female pigs. Soc. Reprod. Fertil. Suppl. 62:191-203.

Verardo, L. L., M. S. Lopes, S. Wijga, O. Madsen, F. F. Silva, M. A. M. Groenen, E. F. Knol, P. S. Lopes, and S. E. F. Guimarães. 2016. After genome-wide association studies: Gene networks elucidating candidate genes divergences for number of teats across two pig populations1. J. Anim. Sci. 94:1446-1458. doi:10.2527/jas.2015-9917

Verardo, L. L., F. F. Silva, L. Varona, M. D. V. Resende, J. W. M. Bastiaansen, P. S. Lopes, and S. E. F. Guimarães. 2015. Bayesian GWAS and network analysis revealed new candidate genes for number of teats in pigs. J. Appl. Genet. 56:123-132. doi:10.1007/s13353-014-0240-y

Volodarsky, M., B. Markus, I. Cohen, O. Staretz-Chacham, H. Flusser, D. Landau, I. Shelef, Y. Langer, and O. S. Birk. 2013. A deletion mutation in TMEM38B associated with autosomal recessive osteogenesis imperfecta. Hum. Mutat. 34:582-586. doi:10.1002/humu.22274

Walum, H., L. Westberg, S. Henningsson, J. M. Neiderhiser, D. Reiss, W. Igl, J. M. Ganiban, E. L. Spotts, N. L. Pedersen, E. Eriksson, and P. Lichtenstein. 2008. Genetic variation in the vasopressin receptor 1a gene (AVPR1A) associates with

pair-bonding behavior in humans. Proc. Natl. Acad. Sci. U. S. A. 105:14153-14156. doi:10.1073/pnas.0803081105

Wang, W., L. J. Zhao, Y. Z. Liu, R. R. Recker, and H. W. Deng. 2006. Genetic and environmental correlations between obesity phenotypes and age at menarche. Int. J. Obesity. 30:1595-1600. doi:10.1038/sj.ijo.0803322

Xue, J., G. D. Dial, W. E. Marsh, and T. Lucia. 1997. Association between lactation length and sow reproductive performance and longevity. J. Am. Vet. Med. Assoc. 210:935-938.

Yang, G., J. Ren, S. Li, H. Mao, Y. Guo, Z. Zou, D. Ren, J. Ma, and L. Huang. 2008. Genome-wide identification of QTL for age at puberty in gilts using a large intercross F2 population between White Duroc × Erhualian. Genet. Sel. Evol. 40:529. doi:10.1186/1297-9686-40-5-529

Yao, G., M. Liang, N. Liang, M. Yin, M. Lu, J. Lian, Y. Wang, and F. Sun. 2014. MicroRNA-224 is involved in the regulation of mouse cumulus expansion by targeting Ptx3. Mol. Cell. Endocrinol. 382:244-253. doi:10.1016/j.mce.2013.10.014

Yao, N., C. L. Lu, J. J. Zhao, H. F. Xia, D. G. Sun, X. Q. Shi, C. Wang, D. Li, Y. Cui, and X. Ma. 2009. A network of miRNAs expressed in the ovary are regulated by FSH. Front. Biosci. 14:3239-3245. doi:10.2741/3447

Yoshizawa, T., Y. Handa, Y. Uematsu, S. Takeda, K. Sekine, Y. Yoshihara, T. Kawakami, K. Arioka, H. Sato, Y. Uchiyama, S. Masushige, A. Fukamizu, T. Matsumoto, and S. Kato. 1997. Mice lacking the vitamin D receptor exhibit impaired bone formation, uterine hypoplasia and growth retardation after weaning. Nat. Genet. 16:391-396. doi:10.1038/ng0897-391

Zhang, Q., Y. Y. Wang, Y. Zhang, H. G. Zhang, Y. Yang, Y. He, J. H. Xu, J. Zhao, Z. Q. Peng, and X. Ma. 2017. The influence of age at menarche, menstrual cycle length and bleeding duration on time to pregnancy: a large prospective cohort study among rural Chinese women. BJOG. 124:1654-1662. doi:10.1111/1471-0528.14469

Zhang, X., L. Huang, T. Wu, Y. Feng, Y. Ding, P. Ye, and Z. Yin. 2015. Transcriptomic Analysis of Ovaries from Pigs with High And Low Litter Size. PLOS ONE. 10:e0139514. doi:10.1371/journal.pone.0139514

Zhao, J., D. H. Xiong, Y. Guo, T. L. Yang, R. R. Recker, and H. W. Deng. 2007. Polymorphism in the insulin-like growth factor 1 gene is associated with age at menarche in caucasian females. Hum. Reprod. 22:1789-1794. doi:10.1093/humrep/dem052

CHAPTER 2: INTEGRATION OF GENOMIC APPROACHES TO IDENTIFY SOURCES OF VARIATION IN AGE AT PUBERTY AND REPRODUCTIVE LONGEVITY IN SOWS*

* Parts of this chapter were published in Journal of Animal Science (2017)

INTRODUCTION

Sow reproductive longevity, or the number of litters produced by sows during their lifetime, plays an important economic role in the swine industry. Specifically, reproductive failure accounts for approximately 35% of culling rates in breeding females, causing economic and welfare barriers for swine producers (Mote, 2008). The sows that express puberty early in life, conceive, and farrow more than three litters during their lifetime are more likely to recover the development and maintenance costs (Stalder et al., 2003; Tart et al., 2013). Currently, approximately 43% of females fail to meet this requirement and are removed from the breeding herd at a young age (Mote, 2008). Thus, identifying and selecting females with greater reproductive longevity prior to retaining them in the herd would increase the number of parities produced by a sow during lifetime, improving sustainability of the swine industry.

Selection for reproductive longevity is challenging due to its complex nature, expression late in life, and low heritability (Tart et al., 2013). Reproductive longevity is a composite phenotype that includes multiple fertility traits. There is substantial interest in identifying early indicators of reproductive longevity. Age at puberty was shown to be an

early indicator of reproductive longevity. Specifically, early onset of puberty was associated with a greater probability of sows to produce multiple parities during lifetime (Serenius and Stalder, 2006; Tart et al., 2013). However, determining the age at which a gilt expresses first estrus in commercial settings is impractical, because it is tedious and time consuming, thus, not used as a selection criterion for breeding programs.

Age at puberty is characterized by a moderate to high heritability (mean h^2 = 0.37, 16 studies, reviewed by Bidanel, 2011) compared to other reproductive traits such as litter size (mean h^2 = 0.11, 118 studies, Bidanel, 2011) or reproductive longevity measured as lifetime number of parities produced (h^2 = 0.04, Tart et al., 2013). We hypothesize that major genetic variants associated with differences in puberty onset will explain a portion of the variation in reproductive longevity. A possible solution for selecting superior breeding females would be to complement traditional fertility related phenotypes used currently in breeding programs with a panel of pleiotropic DNA markers associated with age at puberty and reproductive longevity. Such a panel could be used early in life and assist selection decision of females without having to record age at puberty.

In the current study, various genomic approaches such as genome-wide association, RNA and whole genome sequencing, and gene networks and pathway analysis were used to determine candidate genes and pleiotropic sources of genetic variation associated with age at puberty and reproductive longevity in sows. These genetic variants will be integrated into a novel genotyping array to improve accuracy of genomic prediction of sow fertility while facilitating a reduction in sow replacement rates and addressing welfare concerns.

MATERIALS AND METHODS

This study was approved by the University of Nebraska-Lincoln (**UNL**) Institutional Animal Care and Use Committee.

The Resource Population

The UNL swine resource population was developed to study the roles of genetics and nutrition on reproductive development and longevity of sows. A total of 1,644 females produced in 14 batches (B) that have been extensively phenotyped and genotyped were used (Figure 2.1). Detailed description of the resource population was previously reported in Miller et al. (2011). Briefly, the dams of the experimental females were Large White (LW) × Landrace (LR) crossbreds (B1 to B4) and Nebraska Index Line (NIL) (B1 to B14). The NIL originated from commercial crossbreds (LW × LR) and was selected for increased litter size for 29 generations, while the last 12 generations also included within litter selection for increased lean growth (Hsu and Johnson, 2014). The dams were bred with Landrace (LR) boars from two unrelated commercial lines. The first batches (B1 to B4) were sired by boars from LR1 line and the remaining batches (B5 to B14) were sired by boars from LR2 line. Each batch is considered a separate generation of gilts produced by a distinct group of dams and sires. The number of sires per batch ranged from 5 (B13) to 12 (B3, B5, and B14). The number of dams in each batch varied from 21 (B8) to 65 (B2). The size of the batch varied from 91 to 153 gilts. Due to farrowing space limitations (96 pens), not more than 110 randomly selected gilts were bred per batch.

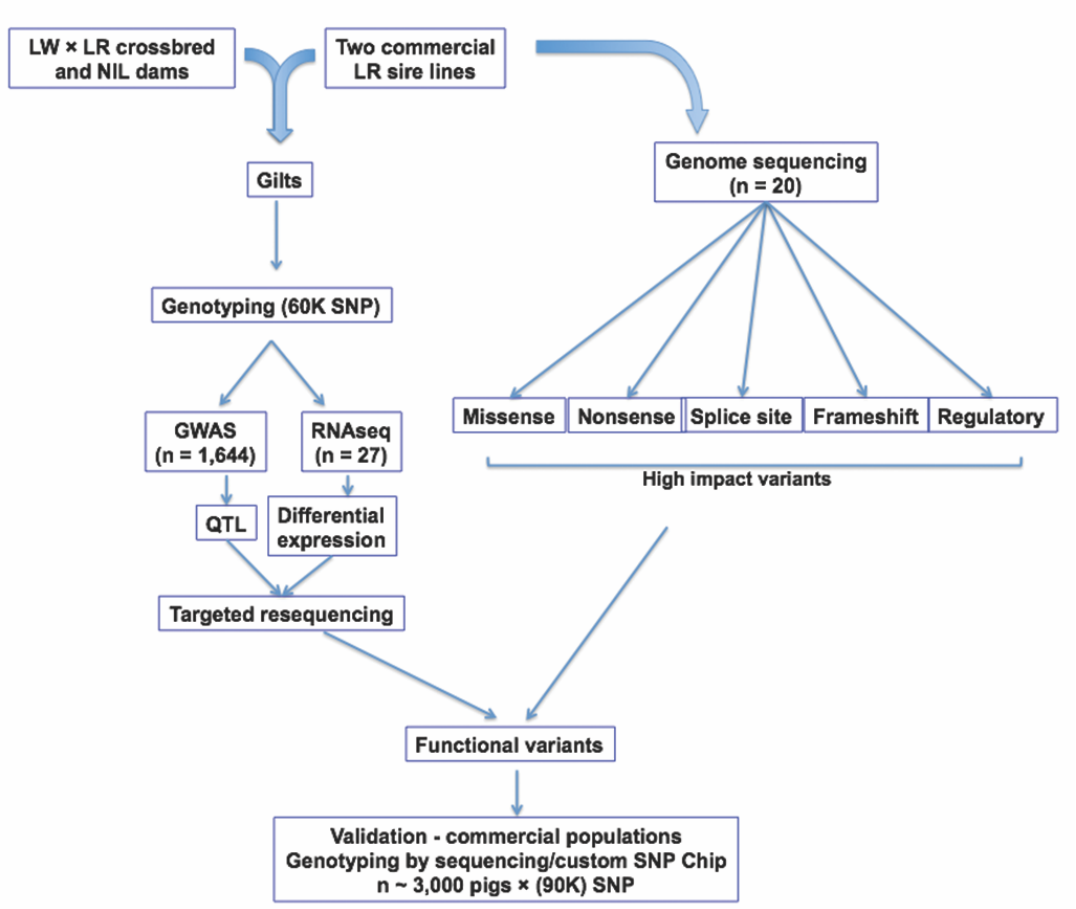

Figure 2.1: Schematic representation of the genomic approaches used in the study. The dams of the experimental gilts were Large White × Landrace (LW × LR) crossbreds and Nebraska Index Line (NIL) sows. The dams were bred with LR boars from 2 unrelated commercial lines. The genetic approaches used included genome-wide association (GWAS) to identify quantitative trait loci (QTL) associated with age at puberty, and genome and RNA sequencing (RNAseq) to identify functional variants. The candidate SNPs will be incorporated into a custom SNP chip and validated in ~3,000 commercial pigs.

Experimental Diets

All gilts were fed a common diet from birth to 123 days of age. During the development period (pre-breeding, 123 to 240 days), until they were moved to the

breeding barn, gilts were allocated either to an *ad libitum* standard corn-soybean based diet (diet A), an energy-restricted diet with approximately 20% less metabolizable energy (ME; diet B) or an energy and lysine restricted diet (diet C) as described in detail in Trenhaile (2015). In B14, diet C (previously used in B11 to B13) was replaced with a high-lysine diet containing the same ME as in the standard diet and ME:lysine ratio as in the energy restricted diet. After being moved to the breeding barn, all the animals were fed a common standard diet. All diets met or exceeded nutritional requirements (NRC, 2012).

Reproductive Phenotypes

Detection of age at puberty in experimental gilts began at approximately 130 days of age and continued until all the gilts within a development pen expressed estrus at least twice or until they reached 240 days of age. Detection of estrus was achieved by moving all gilts from a pen once a day to an adjacent pen where they were exposed to a mature intact boar for 15 min. Age at puberty was defined as the age at which a gilt first expressed estrus. The experimental females were maintained through 4 parities unless they died or were culled. Culling occurred due to failure to express estrus before 240 days of age, failure to conceive or farrow, or for major feet and leg problems. Litter size traits, including total number of piglets born (TNB), number of piglets born alive (NBA), number of mummified and stillborn piglets, and reproductive longevity measured as lifetime number of parities produced (LTNP), were recorded for up to 4 parities. Reproductive longevity was also analyzed as the probability of the sows to produce successive parities. The effect of age at puberty (as a covariate) on these probabilities was

tested using generalized linear mixed models as described in Tart et al. (2013).

Genotyping

Tail snips or ear notches were collected from gilts shortly after birth. DNA was isolated from these tissues (n = 1,644) using the DNeasy or Puregene tissue kits (Qiagen, Valencia, CA). The quality of DNA was assessed using a NanoDrop (Thermo Fisher Scientific Inc., Waltham, MA) or Epoch (BioTek Inc., Winooski, VT) spectrophotometers. All gilts used in the study were genotyped with the *Porcine SNP60 BeadArray* (Illumina Inc., San Diego, CA). Genotypes with an Illumina quality score less than 0.4 and samples and SNPs with a call rate less than 80% were removed leaving 53,529 SNPs for further analysis.

Genome-wide Association Analyses

The proportion of genetic variance for age at puberty and LTNP in experimental gilts explained by high-density SNP genotypes was estimated using a BayesB model implemented by GenSel software (Fernando and Garrick, 2008). Bayes Interval Mapping (BayesIM), a model recently introduced by Kachman (2015), was also used to estimate the proportion of genetic variance for age at puberty explained by high-density SNP genotypes. BayesIM fits haplotypes into association analysis rather than individual SNPs as is the case of BayesB. The SNPs were mapped to the Sscrofa10.2 reference genome assembly (http:// support.illumina.com/sequencing/sequencing_ software/ igenome.html – [accessed March 7, 2016]). The BayesB analysis was carried out by setting the π value to 0.99, assuming that 0.01 of the SNPs have a nonzero effect on the analyzed phenotype.

Contemporary group (batch and diet), genetic line, litter/dam, sire, and developmental pen were included as fixed effects. The Markov Chain Monte Carlo chain included 41,000 samples with the first 1,000 being discarded as burn in. The posterior mean of the genetic and phenotypic variances explained by each 1-Mb window was calculated using effects generated from each 40th sample (Tart et al., 2013). Genomic estimated breeding value (GEBV) was calculated for all gilts using high-density genotypes and the mean posterior SNP effects. The BayesIM was performed setting the pi value to 0.96, quantitative trait loci (QTL) frequency to 200 Kb, number of haplotype states to 16, average haplotype length to 500 Kb, and number of iterations to estimate haplotype parameters to 25. There were 82,000 iterations included in the analysis with first 1,000 iterations discarded as burn in. Fixed effects included contemporary group (batch and diet) and genetic line where as random effects included litter/dam and developmental pen.

Gene Ontology Analysis

The non-overlapping 1-Mb windows across the genome were ranked based on the genetic variance explained for age at puberty and LTNP. The top 1% of 1-Mb windows associated with largest proportion of genetic variance were extended by 0.5 Mb in both directions. Gene annotation of positional candidate genes was obtained using the Sscrofa 10.2 genome build and gene ontology terms were obtained using BIOMART tool in the Ensembl database (version 86; https://may2017.archive.ensembl.org/biomart/martview/3107c5e5934984add6a0f70708ee5537 [accessed 16 March 2016]). Human orthologs of swine positional candidate genes

were obtained from Ensembl Genes 86 Database for pathway and functional analysis for DAVID (https://david.ncifcrf.gov) and Ingenuity Pathway Analysis (IPA).

Genome Sequencing

A subset of 20 sires, representing both ends of the distribution for the average GEBV of daughters age at puberty, were selected for whole genome sequencing. Single-end sequencing was carried out using Ion Proton sequencing as described in the manufacturer protocol (Thermo Fisher Scientific Inc., www.thermofisher.com). Sequence reads were filtered with the prinseq-lite software (Schmieder and Edwards, 2011) by, 1) trimming bases on both read ends when the mean quality in a sliding window of 2 bases dropped below 20, 2) removing duplicates if they occur more than 6 times, 3) removing any read with a non-called base, and 4) requiring all reads to be at least 30 nucleotides long. Filtered and trimmed sequence reads were aligned to the Sscrofa 10.2 genome assembly downloaded from Ensembl using the bowtie2 package. Only the high-quality alignments (Phred score $\geq$ 30) were retained for downstream analysis (Langmead et al., 2009; Langmead and Salzberg, 2012). In order to improve SNP detection, realignment around indels was performed using GATK software tools, RealignerTargetCreator and IndelRealigner (DePristo et al., 2011) leveraging the data available in the dbSNP database (March 2012; ftp://ftp.ncbi.nih.gov/snp/organisms/pig_9823) followed by GATK's BaseCalibrator to reduce the effects of sequence artifacts. Genetic variants were uncovered using the multiallelic and rare-variant options of BCFtools using default settings (Narasimhan et al., 2016).

RNA Sequencing

Collection of the hypothalamus from pre-pubertal gilts was performed before boar exposure, approximately 2 weeks before the gilts were 140 days of age (n = 12 gilts from 12 litters). The pubertal status was confirmed by examining the ovaries at slaughter. The post-pubertal group was composed of gilts fed the three experimental diets, A (n = 10), B (n = 8), and C (n = 7). Age at puberty in the UNL population ranged from 128 to 256 days with an average at 166 days.

The hypothalamus was dissected from the brain by making the following cuts: rostral to the optic chiasm, caudal to the mammillary body, lateral to the hypothalamic sulci, and dorsal to the anterior commissure. Hypothalami from pre-pubertal (n = 12) and post-pubertal gilts (n = 25) representing the same litters were collected and frozen in liquid nitrogen vapor before being placed on dry ice and stored at -80 ^{o}C until the isolation of arcuate nucleus (ARC) from the hypothalamus.

The ARC was isolated using a micropunch procedure. Frozen coronal sections (250 μm) were cut using a CM1950 cryostat (Leica Biosystems, Inc., Buffalo Grove, IL) and mounted onto charged Premier microscope slides (Life Science Products Inc., Manassas, VA). Sections containing the ARC were identified based on anatomical references (Kineman et al., 1988; Kineman et al., 1989; Amstalden et al., 2010). A 2-mm biopsy punch (Miltex Inc., York, PA) was used to bilaterally microisolate the ARC from each section. Micropunches were immediately transferred to a frozen microtube and placed on dry ice and stored at -80 ^{o}C until isolation of RNA. Total RNA was isolated from micropunches by extraction with Trizol (Thermo Fisher Scientific Inc.) followed by

precipitation with isopropanol. The pellet was resuspended in RNAse-free water and RNA purified on RNEasy Mini Columns (Qiagen) according to the manufacturer's protocol for on-column digestion with DNase. The quantity and quality of RNA were determined by NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific Inc.) and microfluidic analysis with an automated electrophoresis system using Agilent 2100 Bioanalyzer (Agilent Technologies, Foster City, CA).

The RNA sequencing was performed using Ion Proton sequencing as described in the manufacturer protocol (Thermo Fisher Scientific Inc., www.thermofisher.com). The RNA sequencing reads were aligned to the Sscrofa10.2 reference genome as explained in the two-step alignment approach used for Ion Proton transcriptome data (Blair et al., 2014). Briefly, the adaptors attached to the RNAseq reads were removed using Cutadapt (version 1.4; Martin, 2011). The quality of raw reads including basic statistics, sequence quality, and content were examined using FastQC (version 0.11; Andrews, 2010). The sequence reads were trimmed and filtered using Trim galore (version 0.4; Krueger, 2015). Phred33 score was used for quality trimming. Low-quality bases in the 5' end were removed and nucleotides with quality base calls less than 22 were trimmed off from the 3' end. The filtered reads were first aligned to the build 10.2 Sscrofa reference genome using Tophat (version 2.1; Trapnell et al., 2012; Blair et al., 2014). The unmapped reads from Tophat were then realigned to the reference genome using the local option of the Bowtie package (version 2.2; Langmead and Salzberg, 2012; Blair et al., 2014). This option aligns long reads to the genome by trimming the ends of reads to achieve the greatest possible alignment score. The alignment outputs from Tophat and local Bowtie were merged with Picard (version 2.1.1; Wysoker et al., 2013). The number

of reads mapped to each gene in the reference annotation was obtained using HTSeq (version 0.6.1p1; Anders, 2014).

Gene Expression Profiling and Pathway Analysis

Differentially expressed genes (DEGs) comparing pre and post pubertal gilts as well as gilts that exhibited puberty early (< 155 days of age) and late (> 180 days of age) were determined using DESeq2 package (Love et al., 2014). The DESeq2 uses a statistical approach based on a generalized linear model and a negative binomial distribution to obtain gene read counts and identify DEGs. The analysis was performed using the default parameters considering a gene being differentially expressed at adjusted *P* (*Padj*) < 0.1.

Ingenuity Pathway Analysis used DEGs as input to identify specific pathways and regulatory networks. This software utilizes comprehensive record of literature available on Ingenuity Pathway Knowledge Base to transform a set of genes into number of relevant networks representing the relationship between genes, their upstream regulators, and downstream gene products. The Regulator Effects tool was used to identify the potential transcriptional regulators of DEGs and their predicted effect on gene expression. Currently, IPA does not support swine, therefore, human orthologs of swine DEGs were obtained from Ensembl Genes 86 Database. Fifty-eight swine DEGs with human orthologs were available for IPA.

RESULTS AND DISCUSSION

Age at Puberty is an Indicator of Sow Reproductive Longevity

Previously we reported that from a range of pre-breeding gilt phenotypes (birth weight, weaning weight, age at puberty, 230-day body weight, backfat thickness, and longissimus muscle area), age at puberty was the only phenotype that affected probability of a gilt to produce the first litter, regardless of their genetic line and development diet (n = 852; Tart et al., 2013). In an updated analysis including a larger data set (n = 1,428) we found that age at puberty affects the probability of the sows to produce up to three parities ($P < 0.001$). Consistent with our initial analysis (Tart et al., 2013), the likelihood of a female generating a parity increased as age at puberty decreased (Figure 2.2), confirming the observed effect of age at puberty on multiple parities.

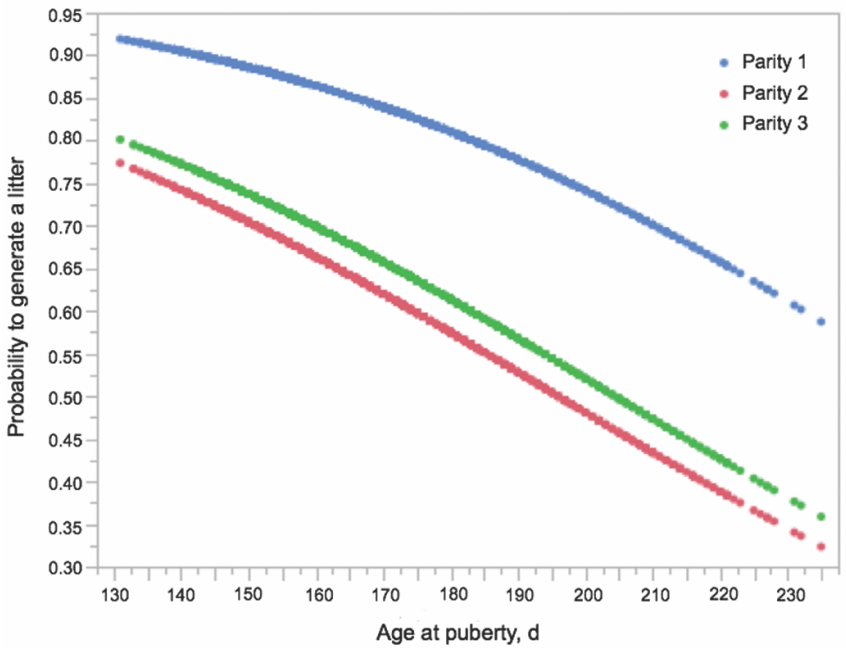

Figure 2.2. The effect of gilt age at puberty on the probability to produce up to 3 parities.

Age at puberty had the largest estimate of heritability from all the reproductive traits measured in our resource population (h^2 = 0.42; Lucot et al., 2015). In comparison, the heritability of NBA and TNB for parity 1 was 0.16 and 0.12, respectively (Trenhaile et al., 2016). The contribution of combined SNP effects to the phenotypic variation was the largest for age at puberty (27.3%) and limited for litter size (< 10%; Table 2.1). The genetic variation of age at puberty is affected by many loci with relatively small effects and the probabilities of the major 1-Mb non-overlapping windows to have effects larger than the average windows are less than 0.30 (Figure 2.3).

Table 2.1: Posterior means of variance components of age at puberty and litter size traits based on SNP effects estimated by BayesB model.

Trait[1]	*n*	Genetic variance	Residual variance	Total variance	Genomic h^2, %
AP	1,644	93.09	195.31	268.50	27.2
NBA-P1	903	1.04	12.63	13.67	7.6
NBA-P2	903	0.30	11.60	11.90	2.5
TNB-P1	903	0.36	9.29	9.65	3.7
TNB-P2	903	0.38	12.04	12.42	3.1

[1]AP = age at puberty (days); NBA-P1 = number born alive in parity 1 (piglets/litter); NBA-P2 = number born alive in parity 2 (piglets/litter); TNB-P1 = total number born in parity 1 (piglets/litter); TNB-P2 = total number born in parity 2 (piglets/litter).

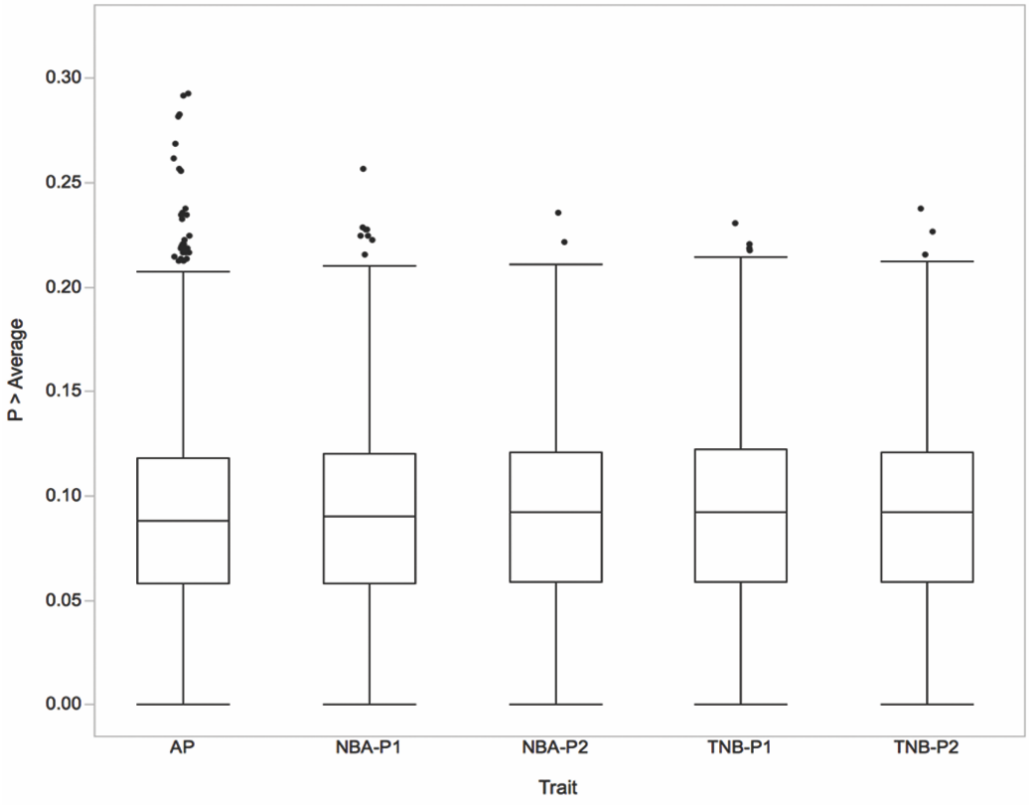

Figure 2.3. Box plot of the probability of 1-Mb windows having effects greater than average (including quartiles and outliers) on fertility traits. Age at puberty (AP), number of piglets born alive at parity 1 (NBA-P1), and 2 (NBA-P2), total number of piglets born at parity 1 (TNB-P1), and 2 (TNB-P2). Age at puberty is a typical quantitative trait influenced by large number of genes with no evidence of major loci explaining substantial phenotypic variation. The probability of the major 1-Mb windows to have an effect greater than the average is less than 0.30.

Identification of Genomic Regions and Candidate Genes Associated with Phenotypic Variation of the Targeted Traits

Age at Puberty

In order to uncover sources of variation that affect age at puberty, we employed

two Bayesian mixture models. Genome-wide association analysis carried out using a BayesB model uncovered major 1-Mb windows associated with age at puberty located on SSC2 (12 to 12.9 Mb), SSC9 (22 to 22.9, 82 to 82.9, and 106.2 to 106.9 Mb), and SSC13 (211 to 211.9 Mb) that explained 0.32 to 0.61% of the genetic variation for age at puberty (Figure 2.4a). The top 1% of major 1-Mb windows (n = 26) explained 6.9% of the genetic variation of age at puberty. The top SNP located on SSC2 QTL (*DIAS0004771*, 12.9 Mb) was shown to have significant additive effect on age at puberty (n = 1,614; $P < 0.01$). As the number of favorable alleles increased age at puberty decreased by 2.4 days. The same top region on SSC2 (12.8 Mb) explained the largest fraction of the total genetic variance for age at puberty using BayesIM model (Figure 2.4b). A high pairwise correlation ($r = 0.8$) was observed between GEBVs obtained from SNP- (BayesB) and haplotype-based (BayesIM) models suggesting that both models captured common loci responsible for genetic variation (Figure 2.4c). Some of the major regions associated with age at puberty identified in the UNL population (top 1% 1-Mb windows) such as the regions on SSC5 (4 Mb) and SSC12 (57 Mb) were also reported from a different population of crossbred gilts (Nonneman et al., 2016).

The top 1% of major 1-Mb windows mapped by both BayesB and BayesIM models uncovered positional candidate genes that have known post-pubertal reproductive functions such as fertilization (*CLIC4*, SSC6, 76 Mb; *NR2F2*, SSC7, 88.9 Mb), placental development (*NR2F2*, SSC7, 88.9 Mb), progesterone secretion and luteinisation (*FZD4*, SSC9, 22.9 Mb), and female pregnancy and embryo implantation (*LIF*, SSC14, 50.2 Mb).

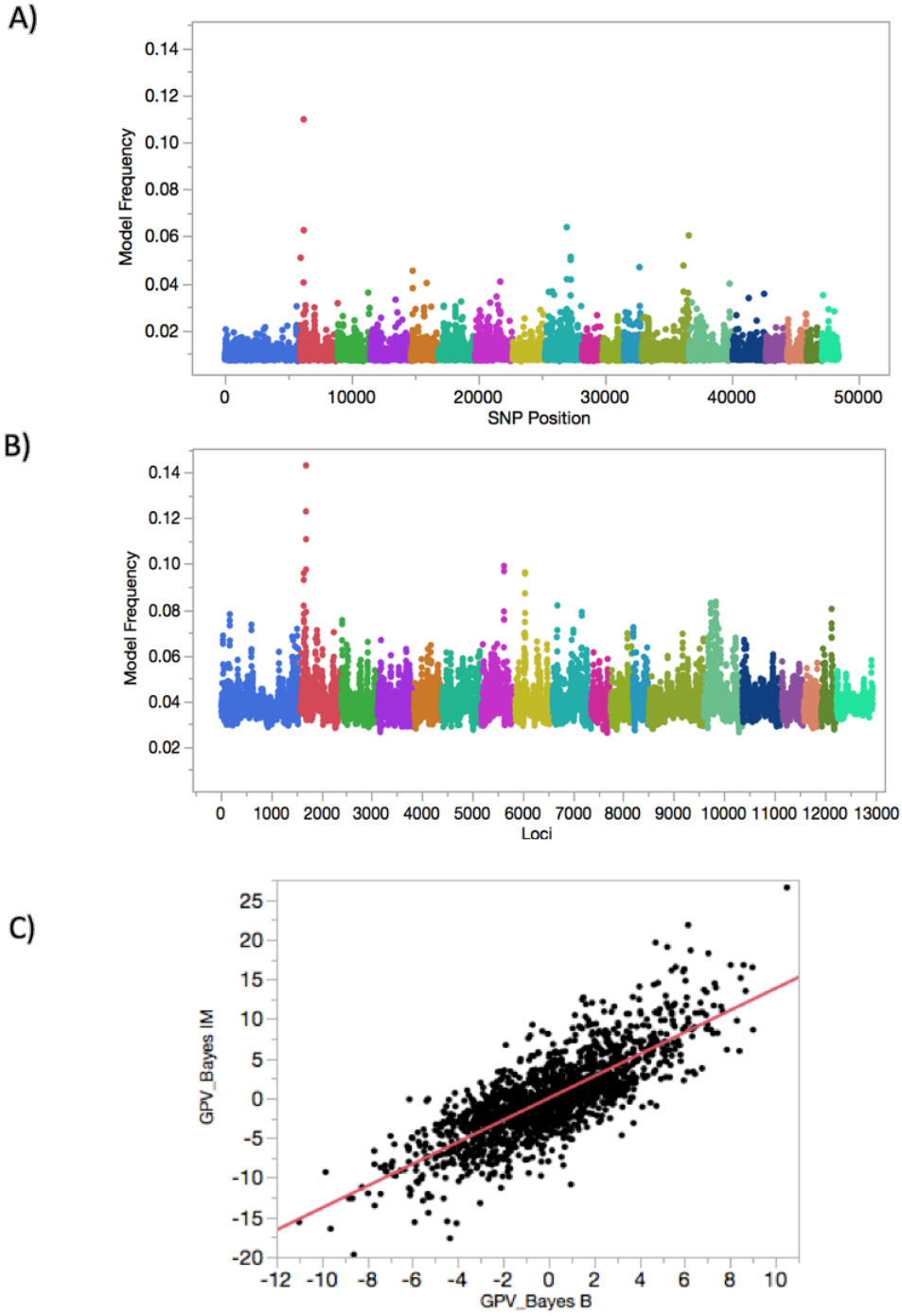

Figure 2.4. Genome-wide association analysis for age at puberty. The autosomes from SSC1 to 18 and chromosome X are represented by different colors. A) BayesB model. Each dot represents a SNP. B) Bayes interval mapping (BayesIM) model. Each dot represents a 200 Kb haplotype. C) Correlation of the GEBV between BayesB and BayesIM models ($r = 0.8$).

Lifetime Number of Parities

The combined *Porcine SNP60 BeadArray* SNP effects explained only a limited proportion (4%) of the phenotypic variation of LTNP. Genome-wide association using BayesB model identified top 1-Mb windows associated with LTNP on SSC1 (117 to 117.9 Mb) and SSC13 (158.0 to 158.9 Mb) explaining 0.41% and 0.24% of the genetic variation, respectively (Figure 2.5).

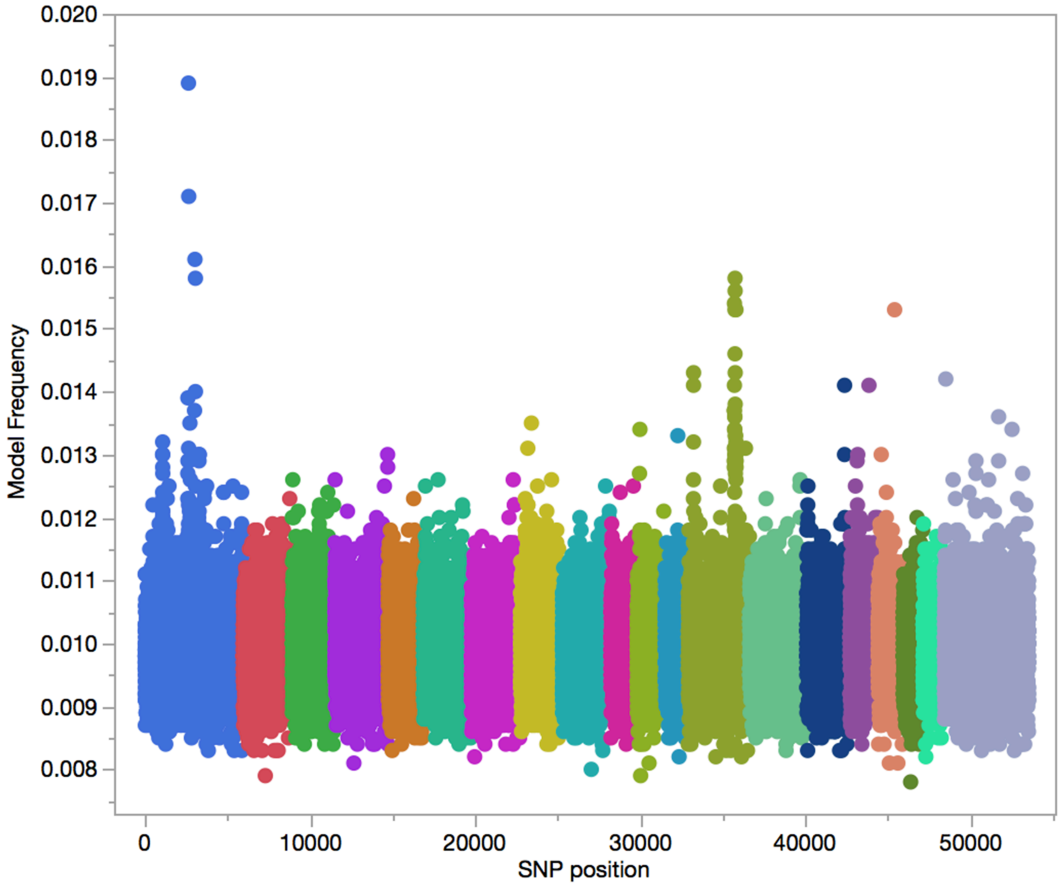

Figure 2.5. Genome-wide association analysis for lifetime number of parities (LTNP). The autosomes from SSC1 to 18 and chromosome X are represented by different colors. Each dot represents a SNP and there were 2 major 1-Mb windows located on SSC1 (117 to 117.9 Mb) and SSC13 (158.0 to 158.9 Mb) associated with LTNP.

The top 1% of the 1-Mb windows (n = 26) explained 3.6% of the genetic variation of LTNP. For the top SNP located on SSC1 QTL (*ALGA0005365*, 117 Mb), as the number of favorable alleles increased LTNP increased by 0.19 litters (n = 1,214; $P < 0.05$). Positional candidate genes overlapping the top 1-Mb windows were identified implicated in reproductive functions such as estrogen receptor binding (*TRIP4*, SSC1, 118.4 Mb), oocyte maturation and maintaining ovulation cycles (*TYRO3*, SSC1, 144.8 Mb; *WASH1*, SSC5, 69.6 Mb), fertilization (*NECTIN3*, SSC13, 158.5 Mb), female pregnancy and embryo implantation (*HSF1*, SSC4, 0.6 Mb; *FKBP4*, SSC5, 69.5 Mb; *LIF*, SSC14, 50.2 Mb), fetal placenta generation (*RSPO3*, SSC1, 39.6 Mb; *HEY2*, SSC1, 40.9 Mb), maternal placenta development (*PRDX3*, SSC14, 140.6 Mb), and in utero embryonic development (*MAN2A1*, SSC2, 119.1 Mb; *HSF1*, SSC4, 0.6 Mb; *TEAD4*, SSC5, 69.4 Mb; *SRSF1*, SSC12, 35.2 Mb).

Pleiotropic Regions for Age at Puberty and Lifetime Number of Parities

There was a weak negative correlation ($r = -0.07$, $P = 0.006$) between the GEBVs for age at puberty and LTNP based on SNP effects estimated by BayesB. Two common 1-Mb regions (SSC2, 12 to 12. 9 Mb and SSC14, 50 to 50.9 Mb) were identified within the top 1% 1-Mb windows for both age at puberty and LTNP. The SSC2 QTL was located near a 1-Mb window identified in the same population as a potential selection sweep region for litter size traits (Trenhaile et al., 2016). Gene ontology analysis of *P2RX3*, a candidate gene in this area, indicated its involvement in embryo implantation and maintenance of pregnancy. Alleles fixed in NIL but polymorphic in lines not subjected to selection for sow fertility and litter size indicated *P2RX3* as a potential

source of the large litter size in NIL (Trenhaile et al., 2016). The SSC2 QTL region also harbors many olfactory receptor genes (e.g. *OR6Q1*, *OR9Q1*, *OR9Q2*, *OR10Q1*, *OR5B2*, *OR5B3*, *OR5B12, OR5B17*, *OR5B21*) involved in detection of chemical stimulus and odorant binding. The olfactory cues from the boar are necessary to stimulate puberty in gilts (Pearce and Hughes, 1987). The olfactory receptor neurons provide sensory inputs to the olfactory bulb. Olfactory bulbectomy in gilts have delayed the puberty onset in the presence of a boar, compared to intact gilts, highlighting the importance of olfactory system in attaining puberty in gilts (Kirkwood et al., 1981). A positional candidate gene (*LIF*) located on SSC14 QTL region was associated with maternal processes involved in female pregnancy such as embryo implantation and decidualization (geneontology.org). The gene is highly expressed in the pig placenta (biogps.org). Point mutations in the coding region of *LIF* gene were associated with blastocyst implantation failure in humans (Stewart et al., 1992; Giess et al., 1999).

Genome-wide association based on BayesB revealed that some of the SNPs without a physical location in the Sscrofa 10.2 genome build also represented an important group of the top 0.1% of SNPs associated with age at puberty (11 of the 53 top SNPs). Two of the unmapped SNPs, *ASGA0092359* and *ASGA0008471,* are within the top three SNPs for their effect on age at puberty. The *ASGA0008471* was shown to have additive pleiotropic effects ($P < 0.05$). The favorable homozygote genotype was associated with 4.3 days earlier expression of age at puberty (n = 1,614; $P = 0.01$) and 0.29 more litters (n = 1,214; $P < 0.10$) compared to the alternate homozygote. Linkage disequilibrium (LD) estimates with multiple mapped SNPs from *Porcine SNP60 BeadArray* revealed that *ASGA0008471* is most likely located on SSC2 (2.5 to 2.9 Mb,

$r^2 > 0.25$) and *ASGA0092359* is located on SSC5 (65 to 68 Mb, $r^2 > 0.12$). Once the updated genome build, Sscrofa 11.1 (https://support.illumina.com/sequencing/sequencing_software/igenome.html [Accessed 20 February 2018]) was released, the *Porcine SNP60 BeadArray* content including the unmapped SNPs (n = 5,121 SNPs) was mapped to the new reference genome using blat command line tool available from UCSC genome browser (http://hgdownload.soe.ucsc.edu/admin/ [Accessed 25 February 2018]). Ninety percent of the SNPs without a physical position in Sscrofa10.2 genome build were mapped to Sscrofa 11.1 reference genome confirming the chromosomal position of *ASGA0008471* (SSC2, 4.1 Mb) and *ASGA0092359* (SSC5, 64.7 Mb) SNPs obtained from LD mapping.

Whole Genome Sequencing Uncovered Potential Sources of Genetic Variation

In order to identify genetic variants outside the limited capability of *Porcine SNP60 BeadArray*, we performed next generation genome sequencing on 20 sires that represents both ends of the distribution for average daughter's GEBV for age at puberty. The average number of gilts with available GEBV per sequenced sire was 21.8. Individual genomic coverage varied from 16.2 to 26.7X with an average of 22.2X coverage per boar. The average length of the sequencing reads after filtering was 164.9 bp. Following filtering with a Phred quality score (≥ 20) and pooled reads depth (≥ 20), 11,896,069 SNPs and 1,074,512 indels were uncovered among the 20 samples. The majority of the discovered SNPs were intergenic (65.7%). Intronic SNPs were the most prevalent (96.3%) from all polymorphisms located in genes, followed by SNPs located in the 5' and 3' untranslated region (2.1%) and the coding region (1.6%). Some of these

polymorphisms, especially those located in the extended areas of the major 1-Mb windows associated with phenotypic differences for the targeted traits could be potential sources of genetic variation.

RNA Sequencing of Arcuate Nucleus Provides Expression Profiling of Gilts with Different Puberty Status

High throughput RNA sequencing reads were obtained from ARC from gilts representing pre- and post-pubertal time points, and early and late puberty gilts. On average, 55.3 million raw, single-end Ion Proton reads with an average length of 150 bp were obtained per gilt. After trimming the reads based on quality, 90% of the raw reads per gilt (~50 million) were available for transcriptome analysis. Using a 2-step alignment process (Blair et al., 2013) 94.4% of the trimmed reads were mapped to the genome and of those 45.1% of the reads were mapped to annotated genes.

As expected, a large number of genes (n = 5.8K; *Padj* <0.1) were found differentially expressed between pre-pubertal and post-pubertal (early or late) gilts. Differential expression between early and late pubertal gilts was observed for 70 genes (*Padj* < 0.1), including genes involved in age at menarche in humans (*LIN28B*, SSC1, 80.4 Mb) and energy homeostasis (*FFAR2*, SSC6, 40.2 Mb). The LIN28 proteins regulate micro RNAs involved in embryonic development and stem cell differentiation (Thornton and Gregory, 2012). A transgenic mouse model overexpressing *LIN28A* exhibited increased body size and delayed expression of puberty suggesting the involvement of LIN28 proteins in age at menarche (Thornton and Gregory, 2012). Arcuate nucleus is one of the major sites in the hypothalamus involved in integrating central and peripheral

signals that regulate energy homeostasis (Sahu, 2004; Hausman et al., 2012) and links nutrition with reproductive development in gilts (Barb et al., 2006), thus, we expected to see expression differences in genes involved in energy homeostasis between the two groups. However, *FFAR2* and *LIN28B* were not located in the top 1% of QTL regions associated with age at puberty in our population.

Fifty-five of the DEGs were up regulated in gilts exhibiting puberty at later ages compared to gilts with early age at puberty. Three of the DEGs (*FAM111B;* SSC2, 11.8 Mb; *CDADC1,* SSC11, 18.8 Mb; and *HERPUD2*, SSC18, 42.1 Mb) overlapped with major (top 1%) QTL regions associated with age at puberty. *FAM111B* is associated with delayed puberty in humans (genecards.org). Genetic variants located upstream of the transcription start site (-1 to -500 bp) affecting potential cis-binding motifs were identified as possible sources of differential expression of these three genes and variation in onset of puberty. For example, SNP-affected motifs for two transcription factors (*MAX* and *SP2*) known to regulate the expression of both *CDADC1* and *FAM111B* were identified in the proximal promoter of these genes. Using Ingenuity Pathway Analysis (IPA) Regulator Effects tool, 363 upstream regulators of the 70 DEGs were identified. Thirty-eight upstream regulators of six DEGs (*CDKN1A*, *DPP4*, *FFAR2*, *LCN2*, *PGK1* and *SAMHD1*) overlapped with major QTL regions for age at puberty identified in the UNL population. Missense SNPs were identified in four upstream regulators (*RAD9A*, SSC2, 3.8 Mb*; APC* SSC2, 121.5 Mb; *IL17B*, SSC2, 157.3 Mb; and *CDCA2*, SSC14, 10.6 Mb), which could be potential trans-modulators and could influence variation in age at puberty via downstream differentially expressed genes.

Ingenuity Pathway Analysis was also used to predict the downstream effects of

DEGs and to identify pathways and regulatory networks to which DEGs belong. There were five networks significantly enriched in the data set of 58 DEGs used in the analysis. One of the networks included candidate DEGs for age at puberty such as *FAM111B* and *LIN28B* as well as other genes and sex steroids (i.e. *KISS1*, *KISS1R*, *ESR1*, *AR*, and estradiol) involved in puberty onset (Figure 2.6). Estrogen receptor alpha activates release of kisspeptin from kisspeptin neurons in hypothalamic nuclei during puberty. KISS1 receptors are located on GnRH neurons and upon ligand binding stimulate pulsatile release of GnRH to initiate puberty (Mayer et al., 2010). In humans, loss of function mutations in *KISS1* and their receptor (*KISS1R*) caused deficiency in GnRH secretion and delay in puberty emphasizing the importance of kisspeptin signaling in GnRH regulation and its involvement in puberty onset (Semple et al., 2005; Silveira et al., 2010). This specific network emphasizes the potential involvement of these DEGs in regulation of reproductive functions in mammals.

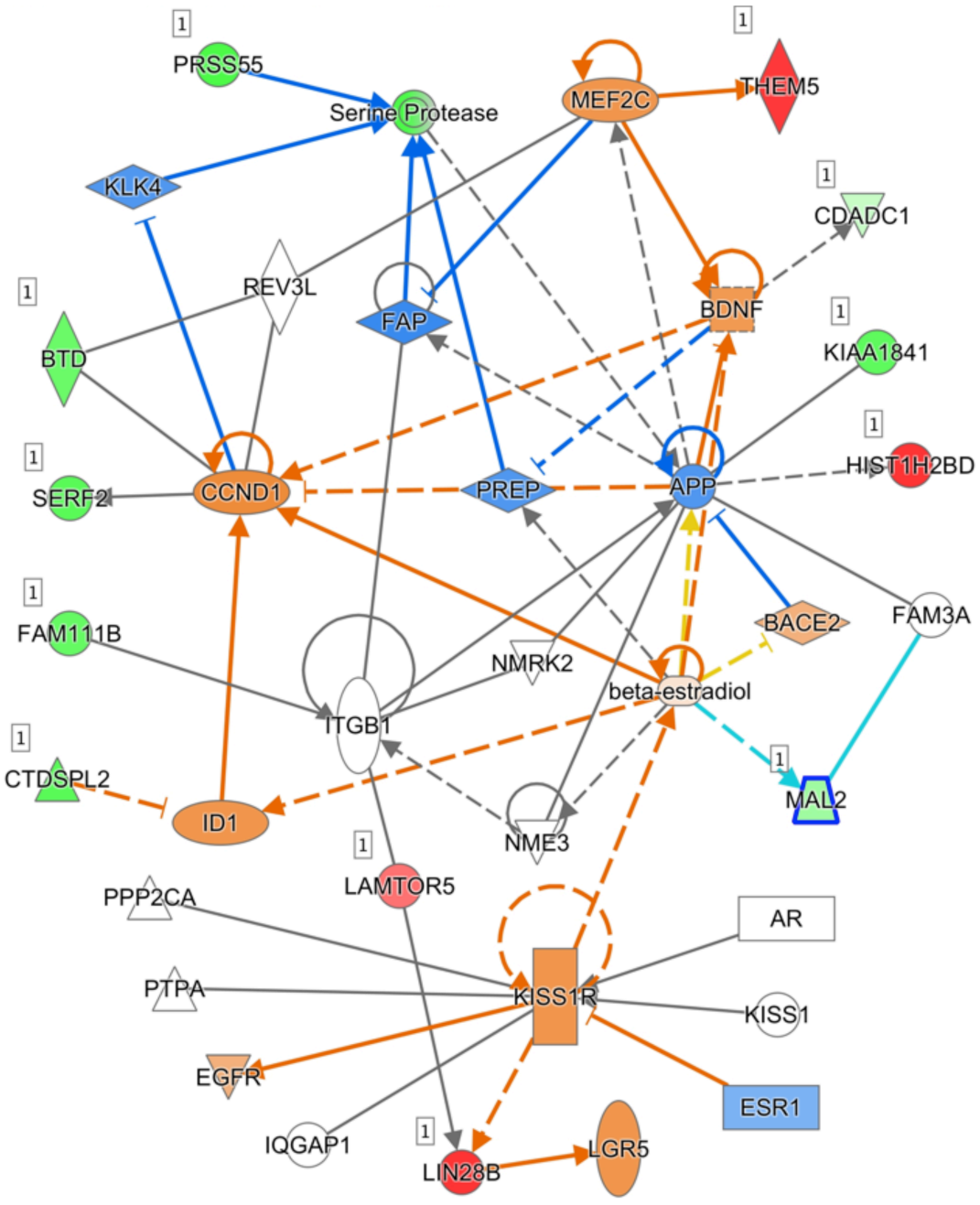

Figure 2.6. The gene network identified using Ingenuity Pathway Analysis. This pathway included DEGs such as *FAM111B* and *LIN28B* and other genes and sex steroids involved in puberty onset such as *KiSS1*, *KiSS1R*, *ESR1*, *AR*, and βestradiol.

CONCLUSIONS

The ultimate goal of this study was to identify genes and functional polymorphisms associated with early onset of puberty and reproductive longevity. To achieve this purpose, we combined the results obtained from genome-wide associations, genome and transcriptome sequencing, and gene expression profiling. A customized SNP array will be constructed incorporating the potential functional variants located in the regions with the largest effect on the targeted traits. The array will be applied in several commercial and research populations for evaluation. The SNP enriched regions with large effects on fertility traits will increase the ability of genomic information to be transferred between populations. Applying this approach when selecting replacement gilts will benefit the swine industry by lowering the production costs due to improved reproductive efficiency, reduction in sow culling and gilt replacement rates, and improving animal welfare.

LITERATURE CITED

Amstalden, M., L. M. Coolen, A. M. Hemmerle, H. J. Billings, J. M. Connors, R. L. Goodman, and M. N. Lehman. 2010. Neurokinin 3 receptor immunoreactivity in the septal region, preoptic area, and hypothalamus of the female sheep: Colocalization in neurokinin B cells of the arcuate nucleus but not in gonadotrophin-releasing hormone neurons. J. Neuroendocrinol. 22:1-12. doi:10.1111/j.1365-2826.2009.01930.x

Anders, S., P. T. Pyl, and W. Huber. 2014. HTSeq – a Python framework to work with high-throughput sequencing data. Bioinformatics. 31:166-169. doi:10.1093/bioinformatics/btu638

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (Accessed 14 December 2016.)

Barb, C. R., R. R. Kraeling, and G. B. Rampacek. 2002. Metabolic regulation of the neuroendocrine axis in pigs. Reprod. Suppl. 59:203-217.

Barb, C. R., G. J. Hausman, and R. Rekaya. 2006. Gene expression in the brain-pituitary adipose tissue axis and luteinising hormone secretion during pubertal development in the gilt. Soc. Reprod. Fertil. Suppl. 62:33-44.

Bidanel, J. P. Biology and genetics of reproduction. In M. F. Rothschild and A. Ruvinsky (eds.) 2011. The genetics of the pig. 228-232. C.A.B. International, Oxfordshire, UK.

Blair, K., S. Ivkovic, V. Mladenovic, F. Torri, and P. Zecevic. 2013. Comparative analysis of software tools for aligning RNA-Seq reads from the Ion Proton platform. ION World Conference, October 21-22, 2013, Boston, MA.

Caldwell, H. K., H. J. Lee, A. H. Macbeth, and W. S. Young. 2008. Vasopressin: behavioral roles of an “original” neuropeptide. Prog. Neurobiol. 84:1–24. doi:10.1016/j.pneurobio.2007.10.007

Chen, X., J. Fu, and A. Wang. 2016. Expression of genes involved in progesterone receptor paracrine signaling and their effect on litter size in pigs. J. Anim. Sci. Biotechnol. 7:31. doi:10.1186/s40104-016-0090-z

DePristo, M. A. E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, An. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, A. B. Gabriel, D. Altshuler, and M. J. Daly. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491-498. doi:10.1038/ng.806

Fernando, R. L., and D. J. Garrick. 2008. GenSel-User manual for a portfolio of genomic selection related analyses. Animal Breeding and Genetics, Iowa State University, Ames, IA.

Fischer, D., A. Laiho, A. Gyenesei, and A. Sironen. 2015. Identification of reproduction-related gene polymorphisms using whole transcriptome sequencing in the Large White pig population. G3(Bethesda) 5:1351-1360. doi:10.1534/g3.115.018382

Giess, R., I. Tanasescu, T. Steck, and M. Sendtner. 1999. Leukaemia inhibitory factor gene mutations in infertile women. Mol. Hum. Reprod. 5:581-586. doi: 10.1093/molehr/5.6.581

Gobrogge, K. L., Y. Liu, L. J. Young, and Z. Wang. 2009. Anterior hypothalamic vasopressin regulates pair-bonding and drug- induced aggression in a monogamous rodent. Proc. Natl. Acad. Sci. USA. 106:19144–19149. doi:10.1073/pnas.0908620106

Hausman, G. J., C. R. Barb, and C. A. Lents. 2012. Leptin and reproductive function. Biochimie. 94:2075-2081. doi:10.1016/j.biochi.2012.02.022

Hsu, W. L. and R. K. Johnson. 2014. Analysis of 28 generations of selection for reproduction, growth, and carcass traits in swine. J. Anim. Sci. 92:4806-4822. doi:10.2527/jas.2014-8125

Kachman, S.D. 2015. Genomic prediction model based on haplotype clusters. Joint Statistical Meetings, August 11, 2015, Seattle, WA.

Kineman, R. D., L. S. Leshin, J. W. Crim, G. B. Rampacek, and R. R. Kraeling. 1988. Localization of luteinizing hormone-releasing hormone in the forebrain of the pig. Biol. Reprod. 39:665-672. doi:10.1095/biolreprod39.3.665

Kineman, R. D., R. R. Kraeling, J. W. Crim, L. S. Leshin, C. R. Barb, and G. B Rampecek. 1989. Localization of proopiomelanocortin (POMC) immunoreactive neurons in the forebrain of the pig. Biol. Reprod. 40:1119-1126. doi:10.1095/biolreprod40.5.1119

Kirkwood, R. N., J. M. Forbes, and P. E. Hughes. 1981. Influence of boar contact on attainment of puberty in gilts after removal of the olfactory bulbs. J. Reprod. Fertil. 61:193-196. doi: 10.1530/jrf.0.0610193

Krueger, F. 2015. Trim Galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Available online at http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (Accessed 14 December 2016.)

Kwon, S. G., J. H. Hwang, da H. Park, T. W. Kim, D. G. Kang, K. H. Kang, I. S. Kim, H. C. Park, C. S. Na, J. Ha, and C. W. Kim. 2016. Identification of differentially expressed genes associated with litter size in Berkshire pig placenta. PLoS ONE

11:e0153311. doi:10.1371/journal.pone.0153311

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome. Biol. 10:R25. doi:10.1186/gb-2009-10-3-r25

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9:357-359. doi:10.1038/nmeth.1923

Lehman, M. N., L. M. Coolen, and R. L. Goodman. 2010. Minireview: Kisspeptin/neurokinin B/dynorphin (KNDy) cells of the arcuate nucleus: A central node in the control of gonadotropin-releasing hormone secretion. Endocrinology. 151:3479-3489. doi:10.1210/en.2010-0022

Love, M. I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome. Biol. 15:550. doi:10.1186/s13059-014-0550-8

Lucot, K. L., M. L. Spangler, M. D. Trenhaile, S. D. Kachman, and D. C. Ciobanu. 2015. Evaluation of reduced subsets of single nucleotide polymorphisms for the prediction of age at puberty in sows. Anim. Genet. 46:403-409. doi:10.1111/age.12310

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 1:10-12, doi:10.14806/ej.17.1.200

Mayer, C., M. Acosta-Martinez, S. L. Dubois, A. Wolfe, S. Radovick, U. Boehm, and J. E. Levine. 2010. Timing and completion of puberty in female mice depend on estrogen receptor alpha-signaling in kisspeptin neurons. Proc. Natl. Acad. Sci. U. S. A. 107:22693-22698. doi: 10.1073/pnas.1012406108

Miller, P.S., R. Moreno, and K. Johnson. 2011. Effects of restricting energy during the gilt developmental period on growth and reproduction of lines differing in lean growth rate: Responses in feed intake, growth, and age at puberty. J. Anim. Sci. 89: 342-354. doi:10.2527/jas.2010-3111

Mote, B. E. 2008. The use of genetic markers to improve sow productive life and genetic abnormalities.. PhD Thesis, Iowa State University, Ames, Iowa

Narasimhan, V., P. Danecek, A. Scally, Y. Xue, C. Tyler-Smith, and R. Durbin. 2016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics. 32:1749-1751. doi:10.1093/bioinformatics/btw044

Nonneman D. J., J. F. Schneider, C. A. Lents, R. T. Wiedmann, J. L. Vallet, and G. A. Rohrer. 2016. Genome-wide association and identification of candidate genes for age at puberty in swine. BMC. Genet. 17:50. doi:10.1186/s12863-016-0352-y

NRC. 2012. Nutrient Requirements of Swine. 11th revised edition. National Academies

Press- National Research Council, Washington, DC. doi:10.17226/13298

Pearce, G. P., and P. E. Hughes. 1987. The influence of male contact on plasma cortisol concentrations in the prepubertal gilt. J. Reprod. Fertil. 80:417-424. doi: 10.1530/jrf.0.0800417

Sahu, A. 2004. Minireview: A hypothalamic role in energy balance with special emphasis on leptin. Endocrinology. 145:2613–2620. doi:10.1210/en.2004-0032

Schmieder, R., and R. Edwards. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 27:863-864. doi: 10.1093/bioinformatics/btr026

Semple, R. K., J. C. Achermann, J. Ellery, I. S. Farooqi, F. E. Karet, R. G. Stanhope, S. O'Rahilly, and S. A. Aparicio. 2005. Two novel missense mutations in g protein-coupled receptor 54 in a patient with hypogonadotropic hypogonadism. J. Clin. Endocrinol. Metab. 90:1849-1855. doi: 10.1210/jc.2004-1418

Serenius, T., and K. J. Stalder. 2006. Selection for sow longevity. J. Anim. Sci. 84 (E. Suppl.): E166-E171. doi:10.2527/2006.8413_supplE166x

Silveira, L. G., S. D. Noel, A. P. Silveira-Neto, A. P. Abreu, V. N. Brito, M. G. Santos, S. D. Bianco, W. Kuohung, S. Xu, M. Gryngarten, M. E. Escobar, I. J. Arnhold, B. B. Mendonca, U. B. Kaiser, and A. C. Latronico. 2010. Mutations of the KISS1 gene in disorders of puberty. J. Clin. Endocrinol. Metab. 95:2276-2280. doi: 10.1210/jc.2009-2421

Stalder, K. J., R. C. Lacy, T. L. Cross, and G. E. Conatser. 2003. Financial impact of average parity of culled females in a breed-to-wean swine operation using replacement gilt net present value analysis. J. Swine. Health. Prod. 11:69-74.

Stewart, C. L., P. Kaspar, L. J. Brunet, H. Bhatt, I. Gadi, F. Köntgen, and S. J. Abbondanzo. 1992. Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. Nature. 359:76-79. doi: 10.1038/359076a0

Tart, J. K., R. K. Johnson, J. W. Bundy, N. N. Ferdinand, A. M. McKnite, J. R. Wood, P. S. Miller, M. F. Rothschild, M. L. Spangler, D. J. Garrick, S. D. Kachman, and D. C. Ciobanu. 2013. Genome-wide prediction of age at puberty and reproductive longevity in sows. Anim. Genet. 44:387-397. doi:10.1111/age.12028

Thornton, J. E., and R. I. Gregory. 2012. How does Lin28 let-7 control development and disease? Trends. Cell. Biol. 22:474-482. doi: 10.1016/j.tcb.2012.06.001

Trapnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, and L. Pachter. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7:562-578. doi:10.1038/nprot.2012.016

Trenhaile, M. D. 2015. Genomic analysis of sow reproductive traits: Identification of selective sweeps, major genes, and genotype by diet interactions. MS Thesis,

University of Nebraska, Lincoln. Nebraska

Trenhaile, M. D., J. L. Petersen, S. D. Kachman, R. K. Johnson, and D. C. Ciobanu. 2016. Long-term selection for litter size in swine results in shifts in allelic frequency in regions involved in reproductive processes. Anim. Genet. 47:534-542. doi:10.1111/age.12448

Walum, H., L. Westberg, S. Henningsson, J. M. Neiderhiser, D. Reiss, W. Igl, J. M. Ganiban, E. L. Spotts, N. L. Pedersen, E. Eriksson, and P. Lichtenstein. 2008. Genetic variation in the vasopressin receptor 1a gene (AVPR1A) associates with pair-bonding behavior in humans. Proc. Natl. Acad. Sci. USA. 105:14153–14156. doi:10.1073/pnas.0803081105

Wysoker, A., K. Tibbetts, and T. Fennel. 2013. Picards tools version 1.90. http://picard.sourceforge.net (Accessed 14 December 2016.)

Yeo, Z. X., M. Chan, Y. S. Yap, P. Ang, S. Rozen, and A. S. G. Lee. 2012. Improving indel detection specificity of the Ion Torrent PGM benchtop sequencer. PloS ONE 7:e45798. doi:10.1371/journal.pone.0045798

Zhang, X., L. Huang, T. Wu, Y. Feng, Y. Ding, P. Ye, and Z. Yin. 2015. Transcriptomic analysis of ovaries from pigs with high and low litter size. PLoS ONE 10:e0139514. doi:10.1371/journal.pone.0139514

CHAPTER 3: DEVELOPMENT AND GENOTYPE QUALITY EVALUATION OF "*SOWPRO90*", A NEW GENOTYPING ARRAY FOR SWINE*

* Parts of this chapter were published in Journal of Animal Science (2019)

INTRODUCTION

Sow fertility is one of the most important factors that impact productivity of swine operations (Serenius and Stalder, 2006). Sow reproductive traits are generally lowly heritable (Tart et al., 2013; Trenhaile et al., 2016) and expressed late in life making early selection for these traits difficult. Age at puberty is one of the earliest indicators of reproductive longevity (Tart et al., 2013). Late onset of puberty was associated with a reduction in service rate (Graves, 2015) and a decreased probability to generate multiple parities (Tart et al., 2013; Lucot et al., 2015). As a result, identification of pleiotropic sources that influence phenotypic variation of age at puberty and fertility traits expressed late in life such as sow reproductive longevity, would have an important impact on any modern genetic program.

Disease susceptibility and exposure to other environmental stressors are critical factors that cause major economic losses for swine producers (Rowland et al., 2012; Engle et al., 2014). The innate and adaptive immunity play a major role in providing first and second line defense against pathogen entrance and clearance. Therefore, identifying genetic variants associated with higher disease resistance or genetic fitness are important

goals of swine industry. For example, recently, a group of North American swine genetic companies formed a consortium (PigGen Canada) to study the role of host genetics in susceptibility to important viral pathogens.

Multiple genomic approaches carried out at University of Nebraska-Lincoln (UNL) and US Meat Animal Research Center (USMARC) including transcriptomic and genomic sequencing, genome-wide association (GWAS), and gene network and pathway analysis were integrated to identify candidate genes and potential genetic variants influencing fertility (Tart et al., 2013; Trenhaile et al., 2016; Nonneman et al., 2016b; Wijesena et al., 2017), innate and adaptive immune response (Engle et al., 2014; Kreikemeier et al., 2015; Walker et al., 2018), and SNPs with predicted loss of function (Keel et al., 2017). These variants were integrated into "*SowPro90,*" a custom Affymetrix Axiom myDesign SNP array. This application has the potential to improve genomic prediction for both fertility and disease related traits.

Diverse genotyping platforms with varying SNP densities are used in genomic evaluation in different livestock species (Mullen et al., 2013; Berry et al., 2016; Biochard et al., 2018). Previous reports showed that quality metrics and distribution of quality data differ across genotyping platforms in human (Hong et al., 2012) and livestock (Berry et al., 2016). This study evaluated data obtained from two genotyping approaches, *SowPro90* and *Porcine SNP60 BeadArray*, and established optimal quality control parameters across platforms. The findings and strategy for quality control used here could be helpful in identifying consistent, high quality genotypes for genomic evaluations, especially when integrating genotype data from different platforms.

MATERIALS AND METHODS

This study was approved by the University of Nebraska-Lincoln Institutional Animal Care and Use Committee (Project ID: 1677).

Animal Populations

Tissue and DNA samples were available from 1,644 experimental sows from the UNL resource population. The sows were developed to investigate the effect of genetics and diet on their reproductive potential. The genetic makeup of dams of the experimental sows was comprised of Nebraska Index Line and commercial Landrace × Large White crossbred lines while the sires were from two unrelated commercial Landrace lines. During developmental (pre-breeding) period (123 to 240 days of age), gilts were subjected to different dietary treatments and starting at approximately 130 days of age they were subjected to daily estrus detection. The age at which a gilt expressed first estrus was defined as the age at puberty. In addition, litter size traits (i.e. total number of piglets born, number of piglets born alive, number of stillborn and mummified piglets, and life-time number of parities produced) were also recorded for experimental animals up to four parities. A detailed description of the resource population and the phenotypic data collected was previously reported (Miller et al., 2011; Wijesena et al., 2017). In addition, tissue and DNA samples as well as fertility data were available from 2,309 animals from two commercial populations with different genetics including Landrace and Yorkshire pigs as well as maternal Landrace × Large White crossbred sows.

Genotypic Data Collection

The DNA was isolated from tail tissue samples collected from the 1,664 sows in the UNL population generated in 14 batches as described in Wijesena et al. (2017). Genotyping was completed with *Porcine SNP60 BeadArray* versions 1 and 2 (Illumina Inc. San Diego, CA), and SNPs with a GenCall score ≥ 0.4 and SNPs and samples with a call rate ≥ 80% were retained for downstream analysis (n = 53,529; Wijesena et al., 2017). In addition, 277 sows in the UNL population representing both extremes of the distribution for their genomic estimated breeding values (GEBVs) for age at puberty (approximately 10% of the gilts with GEBVs representing early age at puberty and approximately 8% of the gilts representing late age at puberty) were also genotyped with *SowPro90* (Thermo Fisher Scientific Inc. Waltham, MA). Moreover, 2,309 animals from the two commercial populations including Landrace, Yorkshire, and crossbred animals were genotyped with *SowPro90*.

Genomic Approaches for Novel Genetic Variants Identification

RNA Sequencing

The RNA sequencing data was obtained from multiple swine populations and tissues. These include the hypothalamic arcuate nucleus from pre pubertal (n = 12) and post pubertal gilts (n = 25) originating from the UNL population (Wijesena et al., 2017) and peripheral blood from commercial maternal crossbred (Large white × Landrace) pigs that expressed high and low levels of viremia following an experimental infection with Porcine circovirus 2b (PCV2b, n = 8, Walker et al., 2018).

The SNP detection was carried out using Genome Analysis Toolkit (GATK, version 3.1, DePristo et al., 2011) and Picard tools (version 2.1.1, Wysoker et al., 2013). Briefly, a sequence dictionary was created for the Sscrofa 10.2 reference genome (http://support.illumina.com/sequencing/sequencing_software/ igenome.html [accessed 7 March 2016]) using CreateSequenceDictonary tool in Picard (version 2.1.1; Wysoker et al., 2013). Aligned RNA sequence BAM files were processed using Picard tools - AddOrReplaceReadGroups, MarkDuplicates, and ReorderSam. The sequence reads were split into exons and any leftover intronic regions were hard clipped using GATK SplitNCigarReads tool. The variants were called using the HaplotypeCaller tool and filtered using VariantFiltration tool in GATK (FisherStrand > 30.0 and QualitybyDepth < 2.0; Van der Auwera et al., 2013). The individual VCF files generated for each sample containing high quality variant calls were then merged within each data set using GATK CombineVariant tool.

Genome Sequencing

Commercial Landrace sires (n = 20) representing both ends of the distribution for average genomic prediction values for their daughters' age at puberty were selected for whole-genome sequencing (Wijesena et al., 2017). Eleven of the sequenced sires represented early age at puberty and nine sires represented late age at puberty. The sequence reads were mapped to Sscrofa 10.2 reference genome and DNA variants were detected using default settings in the multiallelic and rare-variant option of BCFtools (Wijesena et al., 2017). Seventy-two founders in a USMARC experimental swine herd (12 Duroc and 12 Landrace boars and 48 Yorkshire × Landrace composite sows) were

also sequenced to identify putative functional variants across the swine genome such as loss of function, non-synonymous, and regulatory SNPs (Keel et al., 2017). Variant calling and filtering were performed as described in Keel et al. (2017).

Design of the *SowPro90* SNP Array

The *SowPro90* SNP array was designed and manufactured based on Affymetrix Axiom myDesign technology (Thermo Fisher Scientific Inc. Waltham, MA) and included 103,476 SNPs. The SNPs were obtained from sources mentioned above originated from transcriptomic and genome sequencing. In addition, SNPs from available commercial genotyping platforms, mainly *Porcine SNP60 BeadArray* were added to provide the necessary scaffold for imputation.

Briefly, transcriptomic and genomic sequence data were used to identify SNPs located in genes and their proximal promoters (± 2 Kb region flanking the transcription start site [TSS]) that overlapped the top 1% of QTL for age at puberty discovered by GWAS in the UNL (Wijesena et al., 2017) and USMARC (Nonneman et al., 2016b) populations. In the UNL population the genes were identified in major 1-Mb windows extended by 500 Kb in both directions (n = 42 windows). These major windows explained the largest proportion of genetic variance for age at puberty (Wijesena et al., 2017). Similarly, the genes overlapping major QTL for age at puberty in the USMARC population were identified in the 5-SNP QTL windows extended by 300 Kb in both directions (n = 222 windows) (Nonneman et al., 2016b). Another major portion of the array included SNPs located in genes with ontologies associated with innate and adaptive immunity, and also SNPs known to affect viral disease susceptibility (Walker et al., 2018;

Walker et al., 2019). The immunity related gene ontology terms were obtained from Ensembl BioMart (https://may2017.archive.ensembl.org/biomart/martview/[accessed 2 May 2017]). Additionally, SNPs in the proximal promoter (-1 to -500 bp of TSS) of differentially expressed genes between gilts that expressed puberty at different ages, their upstream regulatory genes (e.g. transcription factors, Wijesena et al., 2017), genes overlapping selection sweep regions for litter size traits (Trenhaile et al., 2016), and genes associated with structural soundness (Trenhaile, personal communication) were included in *SowPro90*. A large majority of SNPs incorporated in the array were positional candidate gene-based, located in their coding (e.g. non-synonymous, synonymous, splice region, stop gained, and stop lost) and untranslated regions (5' and 3'). The position of the genes was identified based on Sscrofa 10.2 reference genome annotation. The potential functional SNP consequence was obtained using Ensembl Variant Effect Predictor (https://may2017.archive.ensembl.org/info/docs/tools/vep/index.html [accessed 25 May 2017]). The *SowPro90* array was also consisted of potential loss of function SNPs (Keel et al., 2017) as well as genetic markers for age at puberty identified in the USMARC studies (Nonneman et al., 2016a).

The majority of the scaffold SNPs incorporated in the SNP array was obtained from the *Porcine SNP60 BeadArray* having a minor allele frequency > 0.05 in the UNL and commercial maternal crossbred data sets (Wijesena et al., 2017; Walker et al., 2018). The array was also supplemented with SNPs originated from the *Neogen Porcine GGPHD Array* (Neogen Genomics, Lincoln, NE) if they overlapped the top 1% of the QTL regions for age at puberty (Wijesena et al., 2017) and SNPs from the Affymetrix

Axiom Pig High Density (PigHD) Array (Groenen, 2015; Thermo Fisher Scientific Inc. Waltham, MA) located in the Swine Leukocyte Antigen complex I and II (Table 3.1). The *SowPro90* array is commercially available and the array content can be found in Wijesena et al. (2019).

The *SowPro90* Genotype Quality Evaluation

The genotype quality of *SowPro90* was evaluated by assessing the genotype concordance defined as proportion of identical genotypes for common SNPs between *SowPro90* and *Porcine SNP60 BeadArray* using 277 UNL animals genotyped with both platforms. There were 49,710 *Porcine SNP60 BeadArray* SNPs included in the *SowPro90* design.

For *SowPro90*, the CEL files from all genotyped samples (n = 2,586) were imported into SNPolisher tool in Axiom Analysis Suite (AxAS; Thermo Fisher Scientific Inc. Waltham, MA) together with library files and diploid threshold parameter settings while the rest of the parameters were set at default levels. The optimum threshold for SNP call rate was obtained by analyzing the distribution of the data quality and genotype concordance across platforms at 2% SNP call rate increments from 80% to 100%. The sample call rate threshold was obtained by analyzing the data quality and distribution at different sample call rates (80%, 90%, 93%, and 97%). Finally, the *SowPro90* data (n =103,476) was re-analyzed using all genotyped animals (n = 2,586) and the newly established optimum SNP and sample call rates in order to generate the genotypes for future downstream analysis. To assess the genome-wide distribution of SNPs across platforms, the *SowPro90* and *Porcine SNP60 BeadArray* SNPs were mapped to the

Sscrofa 11.1 reference genome assembly (https://support.illumina.com/sequencing/sequencing_software/igenome.html [Accessed 10 August 2018]).

RESULTS AND DISCUSSION

Development of the *SowPro90* SNP Array

Reproductive longevity is a composite trait influenced by many fertility phenotypes, has a low heritability ($h^2 = 0.04$), and expressed late in life (Tart et al. 2013). Previous research found that age at puberty, a trait with moderate heritability ($h^2 = 0.42$), is one of the earliest indicators of reproductive longevity (Tart et al. 2013). Understanding the pleiotropic sources influencing phenotypic variation of age at puberty and other fertility traits could help in the development of a reliable approach to improve genomic prediction for sow reproductive longevity and other fertility traits. Genetic variants (SNPs) overlapping QTL for age at puberty and fertility traits as well as other economically important traits such as susceptibility to viral diseases were integrated into *SowPro90*, a custom Axiom myDesign SNP array (Thermo Fisher Scientific Inc. Waltham, MA.). The *SowPro90* included 103,476 SNPs overlapping 4,171 transcribed genes (Table 3.1). Similar custom SNP panels targeting economically important traits have been developed in cattle industry to aid in genomic selection. For example, Mullen et al. (2013) and Boichard et al. (2018) developed custom SNP panels for dairy and beef cattle to uncover genes associated with quantitative traits, lethal recessive, and congenital disorders. These panels included SNPs from low density *Illumina BovineLD BeadChip*

and causative variants such as loss of function and non-synonymous polymorphisms.

Approximately 50% of the *SowPro90* SNPs (n = 51,463) were identified using transcriptomic and genomic sequencing as described in Wijesena et al. (2017). Of these, 32,964 SNPs were located in 2,288 genes overlapping major QTL regions for age at puberty discovered in prior studies (Nonneman et al., 2016b; Wijesena et al., 2017) (Table 3.1). The SNP array was also supplemented with 16,271 SNPs located in 1,015 genes involved in innate and adaptive immunity including genes overlapping the Swine Leukocyte Antigen complex I and II and other genes influencing viral disease susceptibility (Kreikemeier et al., 2015; Walker et al. 2018) (Table 3.1). The rest of the SNPs identified were located in differentially expressed genes (including their proximal promoters) in the hypothalamic arcuate nucleus of gilts that expressed puberty early versus late as well as their upstream regulatory genes or *trans* modulators (Wijesena et al., 2017), SNPs in genes overlapping 11 selective sweep regions for litter size traits (Trenhaile et al., 2016), and SNPs in genes associated with structural soundness (Fan et al., 2009; Fan et al., 2011). Additionally, 565 SNPs located in 504 genes characterized by potential loss of function (Keel et al., 2017) as well as previously reported DNA markers for age at puberty (Nonneman et al., 2016a) were included. The remaining approximately 50% of the array content (n = 51,316) was comprised of scaffold SNPs obtained from *Porcine SNP60 BeadArray* and SNPs originating from other commercially available platforms (e.g. *Neogen Porcine GGPHD Array* and Affymetrix *PigHD array*) overlapping the top 1% QTL for targeted traits (Table 3.1).

Table 3.1: Number of SNPs and overlapping genes included in *SowPro90*

SNP category	Number of SNPs	Number of genes
SNPs in genes and regulatory regions (RNA and genome sequencing)		
42 QTL for age at puberty (UNL)	11,474	788
222 QTL for age at puberty (USMARC)	21,490	1,500
Adaptive and immunity genes	16,271	1,015
Differentially expressed genes in hypothalamic arcuate nucleus	107	17
Upstream regulatory genes of differentially expressed genes	308	31
11 selection sweep regions for litter size	1,286	220
Structural soundness genes	607	224
Predicted loss of function SNPs	617	376
SNPs from commercial genotyping platforms		
Illumina *Porcine SNP60 BeadArray*	49,710	
Neogen *Porcine GGPHD Array*	1,012	
Affymetrix *Axiom PigHD Array*	594	
Total	103,476	4,171

Initial Quality Evaluation of the *SowPro90*

The Affymetrix AxAS software classified SNPs into six quality classes, including, 1) polymorphic SNPs with three genotype clusters that passed all the quality control parameters (poly high resolution), 2) SNPs that were monomorphic (mono high

resolution), 3) SNPs with only one homozygote and heterozygote genotype clusters (no minor homozygotes), 4) SNPs with more than one heterozygote cluster or the average signal for heterozygote cluster much lower than for the homozygote clusters (off target SNPs), 5) SNPs with genotype call rate below the threshold (e.g. < 80%) and, 6) SNPs that failed one or more quality control parameters (other) (Figure 3.1). Polymorphic and monomorphic high-resolution SNPs and SNPs lacking minor homozygotes were recommended for downstream analysis.

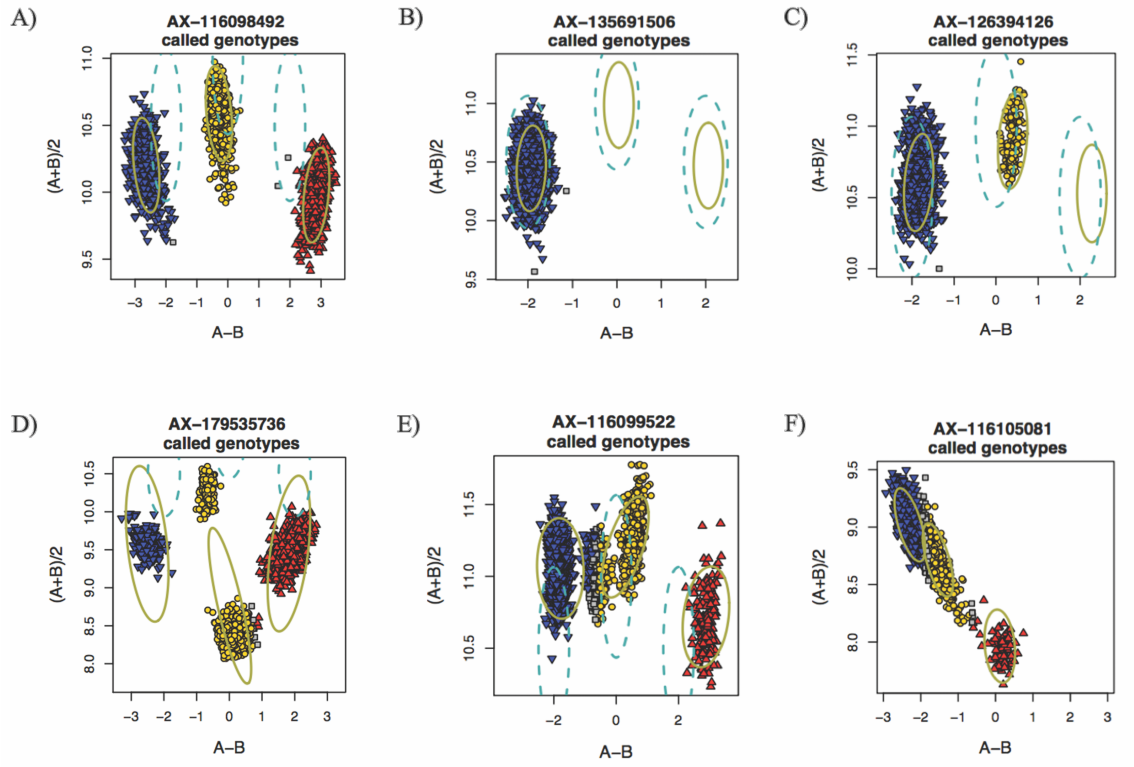

Figure 3.1: Example of SNP classification based on cluster properties by Axiom Analysis Suite. Red – AA genotype, Yellow – AB genotype, Blue – BB genotype, and Grey – No genotype call. SNPs with three genotype clusters (Poly high resolution), monomorphic SNPs (Mono high resolution), and SNPs lacking minor homozygotes (No minor homozygotes) were recommended for downstream analysis.

The genotype quality of *SowPro90* was evaluated by merging genotype data from three populations (UNL and two commercial populations, n = 2,586), rather than evaluating each population or plate (96-well) separately in order to achieve an optimum genotype clustering and reliable genotype calls. For example, the number of recommended SNPs at default call rates (97% SNP and 94% sample call rates) ranged from 62,145 to 94,428 when genotype quality was evaluated in 16 separate plates and only 36,897 (36%) SNPs were consistently recommended across all plates. We hypothesize that the variation in data is a result of limited genetic diversity present in single plates. Plates usually do not include randomly assigned samples from different genetic backgrounds, but rather batches of samples of similar genetics. Some of the non-recommended SNPs could therefore result from a lack of genetic diversity within a plate (or within a genetic line). For example, SNPs that appeared to lack heterozygote genotypes with both homozygote genotypes (AA and BB) present were not recommended. The proportion of this group of SNPs varied from 3 to 5% when samples from three genetic lines were allocated to three separate plates and analyzed individually. When the genotype data from the three populations were merged (n = 2,586), 20% to 86% of these SNPs were recommended and "rescued" in each population since heterozygote calls were previously miscalled as one of the homozygote calls (AA or BB) (Figure 3.2). The absence or rarity of one of the homozygote genotype in samples from the same genetic line (or single plates) could limit the power to distinguish homozygote from heterozygote clusters. This problem does not exist when a large and diverse set of samples is analyzed together, and all three genotypes are expected to be present.

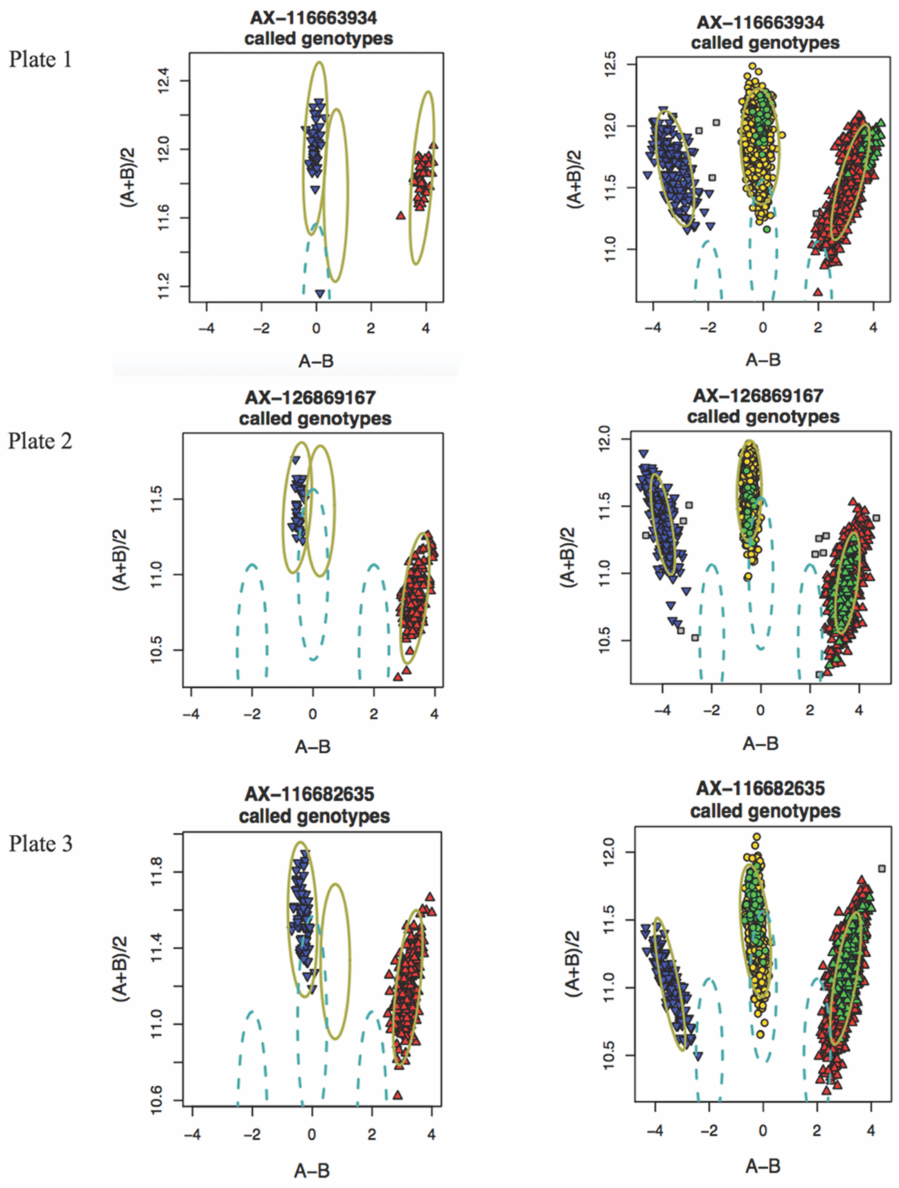

Figure 3.2: Example of genotype calling in three individual plates, each representing a genetic line. These plates (left) generated non-recommended SNPs that appeared to lack heterozygote genotypes but both homozygote genotypes being present (AA and BB). When the three populations were merged (right) these SNPs were recommended since one of the homozygote calls (AA or BB) were miscalled in individual plates when they were actually heterozygotes. Each data point represents the genotype of an animal. Red – AA genotype, Yellow – AB genotype, Blue – BB genotype, and Gray – No genotype call. The green dots represent the animals in each individual plate.

The Genotype Concordance Rate per SNP Between *SowPro90* and *Porcine SNP60 BeadArray*

The evaluation of the genotype concordance was performed using animals (n = 277) genotyped with both *SowPro90* and *Porcine SNP60 BeadArray*. There were 49,710 common SNPs in both platforms. Of those, 44,708 SNPs with ≥ 80% call rate in both platforms were selected for evaluation of the genotype concordance. This call rate is generally considered acceptable in high-density genotyping (Tart et al., 2013).

The mean genotype concordance rate per SNP across the 44,708 SNPs was 98.4%. A small proportion (0.65%) of the homozygote genotypes in one platform were called as alternate homozygote genotypes in the other platform (Table 3.2a). A subset of SNPs with < 90% genotype concordance rate (n = 2,418) exhibited higher incidences of calling heterozygote variants as homozygotes and calling homozygote genotypes in one platform as alternate homozygote genotypes in the other platform. Of those, approximately 13% of the heterozygous *Porcine SNP60 BeadArray* SNPs were called homozygotes in *SowPro90* and approximately 11% of homozygote *Porcine SNP60 BeadArray* SNPs were called alternate homozygotes in *SowPro90* (Table 3.2b). Due to higher incidences of genotype discordance, only SNPs with genotype concordance ≥ 90% were recommended for downstream analysis. Selecting for SNPs with ≥ 90% genotype concordance rate (n = 42,290) increased the overall mean genotype concordance to 99.5% (Table 3.2c).

Table 3.2: Genotype occurrence (%) in *Porcine SNP60 BeadArray* and *SowPro90* using common genotyped animals (n = 277).

A) All SNPs (n = 44,708), B) SNPs with < 90% genotype concordance (n = 2,418), and C) SNPs with ≥ 90% genotype concordance (42,290)

A.

Porcine SNP60 BeadArray	***SowPro90***		
	AA	**AB**	**BB**
AA	98.39	0.96	0.66
AB	0.90	98.22	0.88
BB	0.64	0.70	98.66

B.

Porcine SNP60 BeadArray	***SowPro90***		
	AA	**AB**	**BB**
AA	81.69	6.90	11.41
AB	13.39	73.34	13.27
BB	11.00	5.24	83.75

C.

Porcine SNP60 BeadArray	***SowPro90***		
	AA	**AB**	**BB**
AA	99.32	0.65	0.03
AB	0.20	99.61	0.19
BB	0.03	0.46	99.51

Similar mean genotype concordance rates were reported by other studies in livestock and humans. An evaluation of 49,859 SNPs in sheep samples (n = 84) genotyped by Illumina and Affymetrix platforms reported a 98.1% mean genotype concordance rate per SNP (Berry et al., 2016). In this study a small proportion (0.3%) of homozygous genotypes in one platform were called as alternate homozygous in the other platform. In humans, a comparison between six technical replicates genotyped with both Illumina and Affymetrix platforms reported a mean genotype concordance of 98.8% (Hong et al., 2012). In a simulation study, Hong et al. (2012) reported that using genotypes with lower concordance in GWAS could lead to spurious odds ratios of genetic markers. Jiang et al., (2013) evaluated the within sample genotype concordance between Illumina and Affymetrix platforms in humans for 146,885 SNPs and reported a mean genotype concordance of 99.9%.

A potential source of limited genotype concordance across SNPs could be represented by minor allele frequency (MAF). Across the 44,708 SNPs used for genotype concordance evaluation, the number of SNPs in different *SowPro90* MAF categories ranged from 369 (0.83%) monomorphic SNPs to 5,608 (12.5%) having a MAF between 0.45 and 0.50 (Table 3.3). The lowest mean genotype concordance (61.6%) was observed for monomorphic SNPs while the highest (98.89%) was observed for SNPs with MAF > 0.05 to ≤ 0.10 (Table 3.3). The study of Berry et al. (2016) observed similar results.

Table 3.3: Number and genotype concordance of SNPs in each minor allele frequency (MAF) category

MAF category (*SowPro90*)	**Common SNPs used for genotype concordance evaluation (44,708 SNPs)**	**Mean genotype concordance (44,708 SNPs)**	***SowPro90* at ≥ 97% SNP and ≥ 93% sample call rate (89,040 SNPs)**
0	369	61.63	9,293
> 0 to ≤ 0.05	2593	98.06	9,269
> 0.05 to ≤ 0.1	3338	98.89	7,643
> 0.1 to ≤ 0.15	3713	98.76	7,149
> 0.15 to ≤ 0.2	4111	98.83	7,298
> 0.2 to ≤ 0.25	4439	98.68	7,835
> 0.25 to ≤ 0.3	4910	98.73	7,856
> 0.3 to ≤ 0.35	5116	98.78	8,134
> 0.35 to ≤ 0.4	5055	98.73	8,015
> 0.4 to ≤ 0.45	5456	98.76	8,241
> 0.45 to ≤ 0.5	5608	98.73	8,307

Evaluation of Optimal SNP Call Rate

Concordance between the genotypes across platforms was evaluated within 2% SNP call rate ranges starting from 80% up to 100% (Figure 3.3). For *SowPro90*, the SNP call rates of the majority of SNPs (n = 40,939, 91.6%,) were distributed between 98% to 100% (Figure 3.3a). In this range, there were 40,155 SNPs (98.1%) with ≥ 90% genotype concordance between the platforms and 35,767 of these SNPs (89.1%) had ≥ 99% genotype concordance. Based on the distribution of SNP genotype call rates and the

number of SNPs with ≥ 90% genotype concordance, a SNP call rate of ≥ 97% was considered to be the optimal threshold for *SowPro90* quality evaluation which allowed retention of a maximum number of SNPs (n = 42,151, 94.3%). The mean genotype concordance of the SNPs with ≥ 97% SNP call rate was 98.7%.

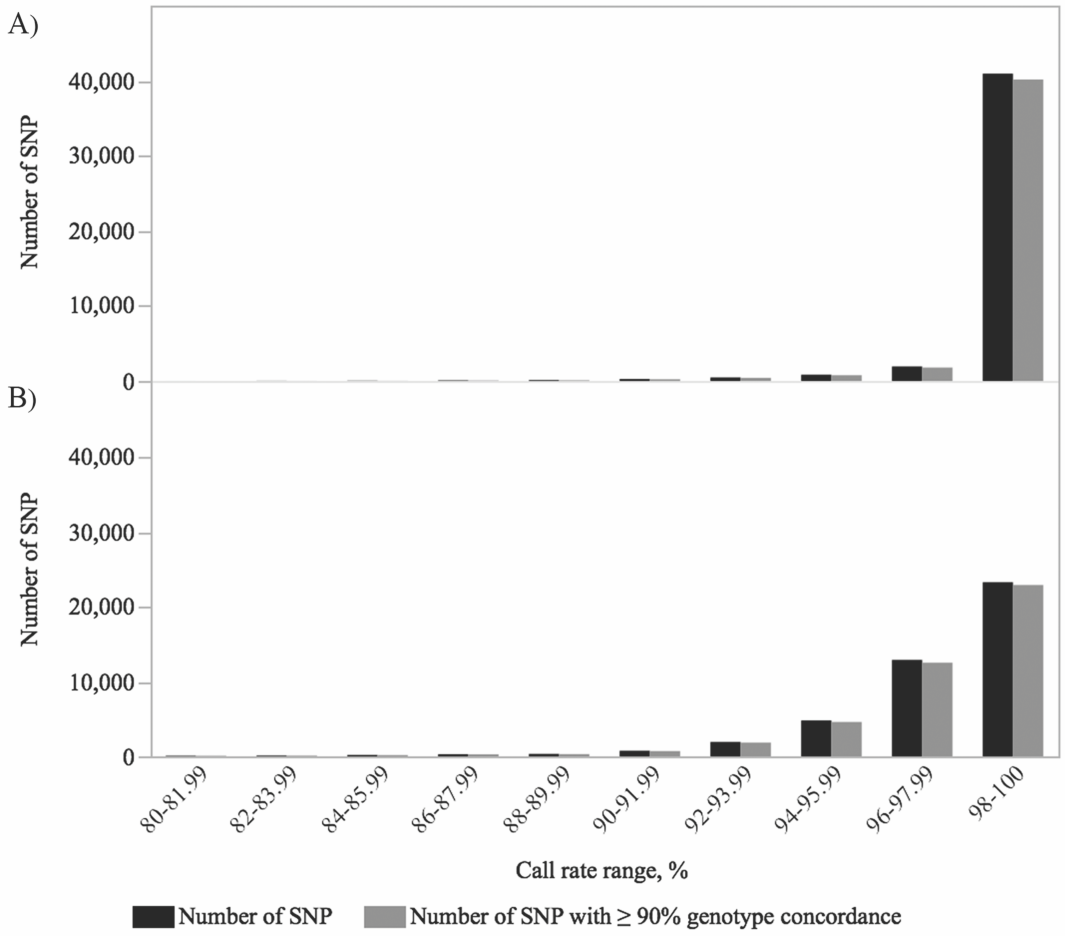

Figure 3.3: Distribution of SNP call rates at 80% SNP and 80% sample call rates. A) *SowPro90* SNPs and B) *Porcine SNP60 BeadArray* SNPs. Black - Total number of SNPs and Gray – Number of SNPs with ≥ 90% concordance.

The SNP call rates for most of the SNPs (n = 41,043, 91.8%) on the *Porcine SNP60 BeadArray* were distributed between 94% to 100% and in this range there were

40,085 SNPs (97.7%) with ≥ 90% genotype concordance between platforms (Figure 3.3b). Based on call rate distribution and concordance data, a SNP call rate of ≥ 94% was considered to be the optimal threshold for *Porcine SNP60 BeadArray* (Figure 3.3b). The mean genotype concordance of the SNPs with ≥ 94% SNP call rate was 98.6%.

Evaluation of Optimal Sample Call Rate

To identify the optimal sample call rate for *SowPro90*, the SNP array (n = 103,476) was re-analyzed at a SNP call rate ≥ 97% and different sample call rates (e.g. 80%, 90%, 93%, and 97%) using all the genotyped animals (n = 2,586). The largest number of genotyped animals (n = 2,571) was retained at ≥ 80% sample call rate (Figure 3.4a). An increase in sample call rate (and removing low quality samples) led to an improvement in genotype clustering and a larger number of SNPs that passed the filtering criteria (Figure 3.5). For example, at ≥ 97% sample call rate there were 932 additional SNPs (+1%) retained compared to ≥ 80% sample call rate (Figure 3.4b) while 72 animals (-2.9%) failed the threshold parameters (Figure 3.4a). Therefore, to retain the maximum number of animals with highest quality genotypes, a less stringent ≥ 93% sample call rate was considered to be the optimal threshold for *SowPro90*. In this case there were 308 additional SNPs retained with only 11 animals failing this threshold parameter compared to ≥ 80% sample call rate.

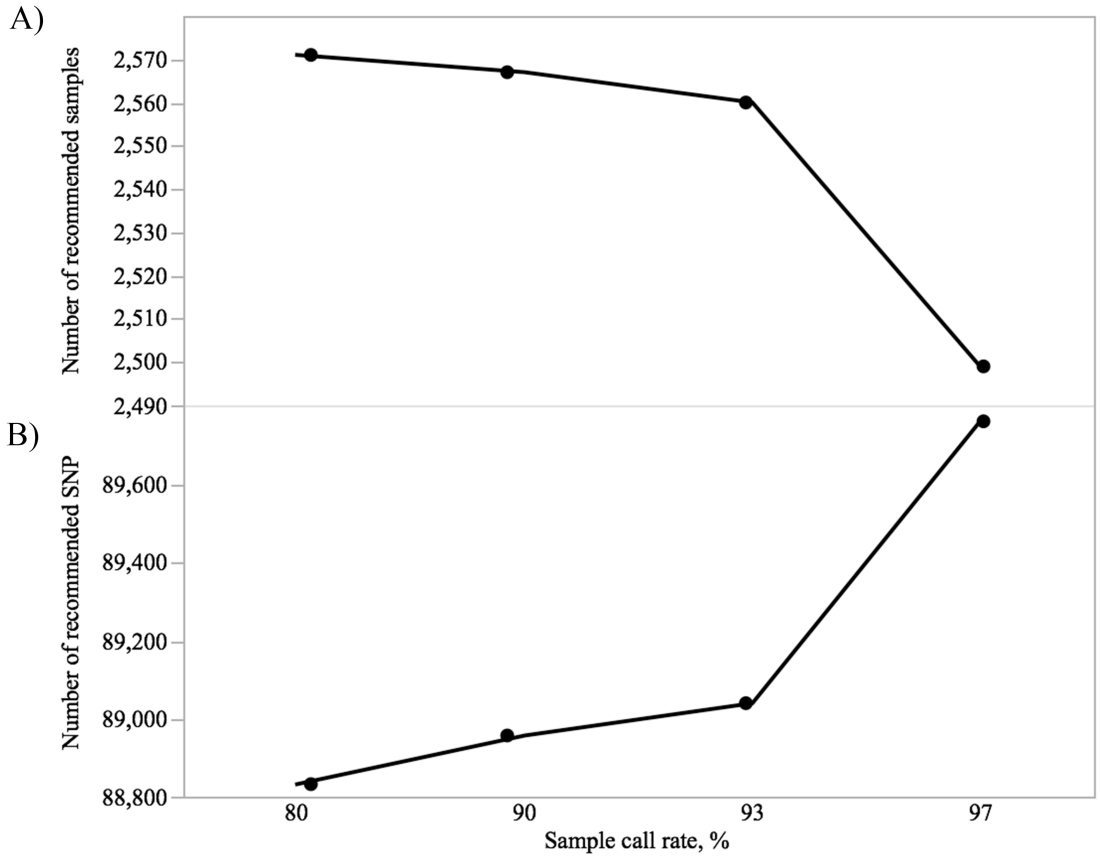

Figure 3.4: A) Number of recommended samples and B) number of recommended SNPs at 97% SNP call rate and different sample call rates for *SowPro90*.

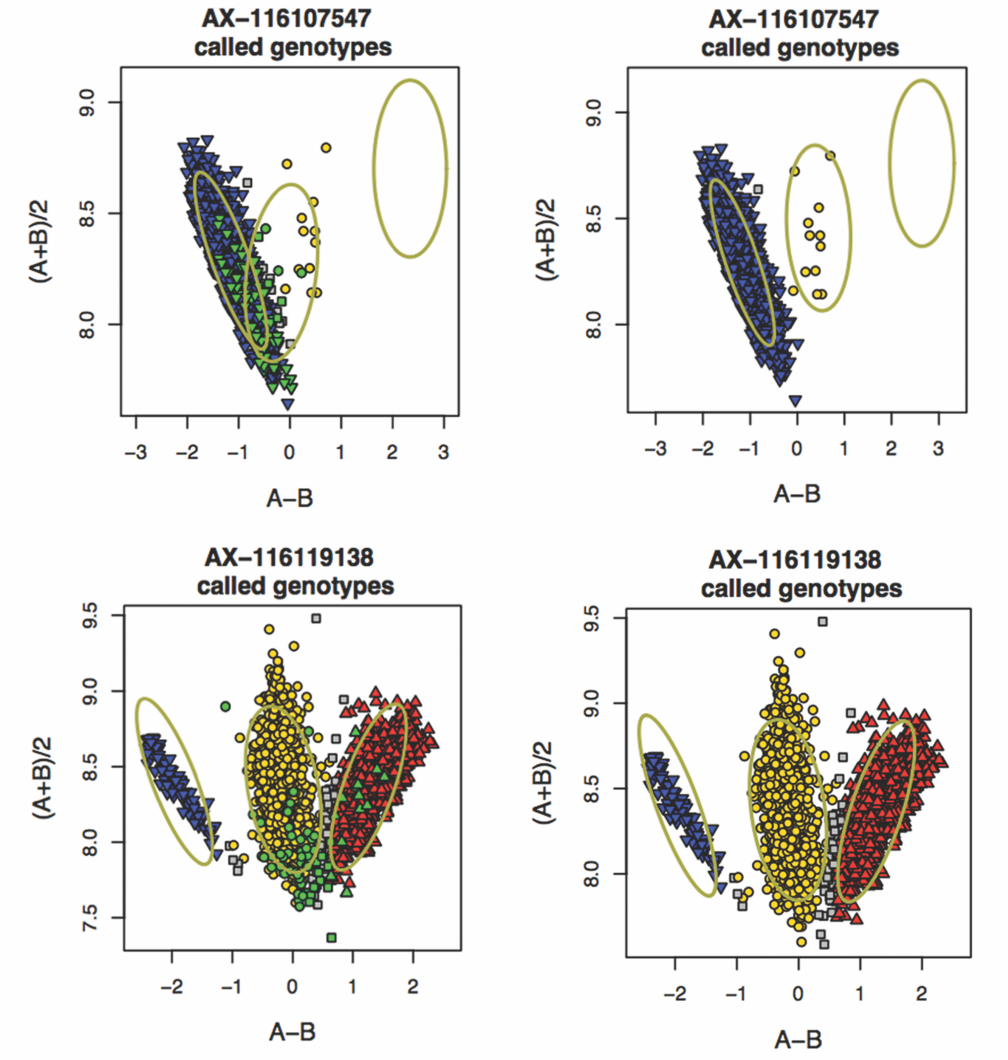

Figure 3.5: An increase in sample call rate improved the overall genotype clustering as illustrated by two examples. The genotype clustering was compared at 97% SNP and 80% (left) and 97% (right) sample call rates. Each data point represents the genotype of an animal. Red – AA genotype, Yellow – AB genotype, Blue – BB genotype, and Gray – No genotype call. The green dots represent the 72 animals that were removed when the sample call rate increased from 80% to 97%.

Similar to above, the *Porcine SNP60 BeadArray* (n = 61,565) data was re-analyzed at ≥ 94% SNP call rate and different sample call rates (e.g. 80%, 85%, 90%, and 93%) using 1,836 genotyped animals. The largest number of SNPs was retained at 93% sample call rate with 57% of the animals failing this filtering criteria (Figure 3.6). In order to retain the maximum number of animals with highest quality genotypes, 85% was determined to be the optimal sample call rate for *Porcine SNP60 BeadArray* data, retaining 53,668 SNPs and 1,668 (91%) animals (Figure 3.6). This sample call rate was also suggested as the optimum by Purfield et al. (2016).

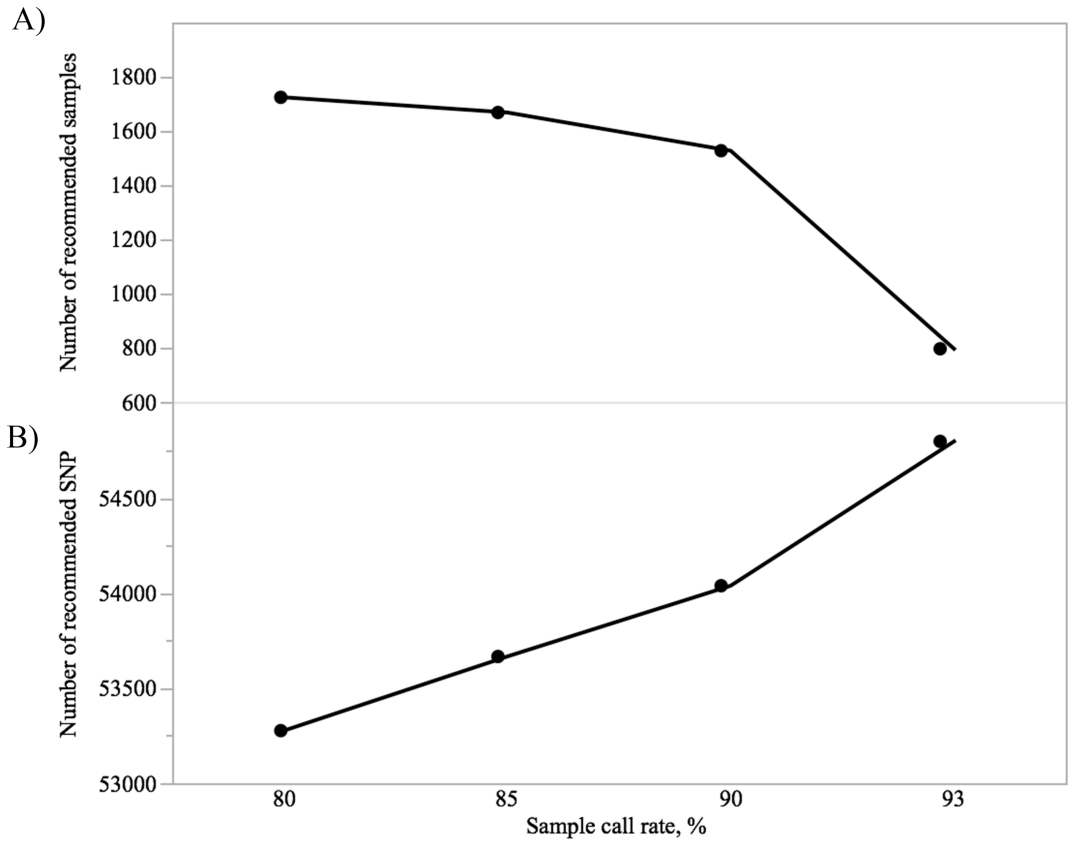

Figure 3.6: A) Number of recommended samples and B) number of recommended SNPs at 94% SNP call rate and different sample call rates for *Porcine SNP60 BeadArray*.

Final Genotype Evaluation of the *SowPro90*

At ≥ 97% SNP and ≥ 93% sample call rates, there were 89,040 (86%) recommended SNPs and 2,560 (98.7%) samples that passed the quality thresholds for *SowPro90*. The recommended SNPs included 74,661 poly high-resolution, 9,293 mono high-resolution, and 5,086 SNPs without homozygotes for the minor allele. The monomorphic SNPs were presumably sequencing artifacts as the majority of these SNPs (94%) originated from transcriptome (73%) and genome (21%) sequencing. The average observed heterozygosity of polymorphic SNPs was 0.35 and the average MAF of the recommended array content was 0.25. The number of recommended SNPs in different MAF categories (excluding monomorphic SNPs) ranged from 7,149 (> 0.1 to 0.15) to 9,269 (0 to 0.05) (Table 3.3).

In the *Porcine SNP60 BeadArray* there were 1,812 SNPs overlapping 42 QTL windows for age at puberty identified in the UNL population. These regions were enriched with 13,511 SNPs in *SowPro90*. The average distance between SNPs in the enriched QTL regions was 5,150 bp for *SowPro90* compared to 38,753 bp for *Porcine SNP60 BeadArray.* In the updated swine genome assembly (Sscrofa 11.1) the SNPs included in *SowPro90* were distributed across the 18 autosomes and the X chromosome ranging from 1,669 (SSCX) to 8,413 SNP (SSC7) per chromosome. At the genome-wide level there were an average of 36 SNPs per 1-Mb window for *SowPro90* compared to 21 SNPs for the *Porcine SNP60 BeadArray*, an improvement of 71%.

CONCLUSIONS

A custom Affymetrix Axiom myDesign SNP array (*SowPro90*) was developed including potential functional genetic variants associated with fertility and disease related traits. This SNP array can provide benefits to swine industry by improving the herd reproductive efficiency, disease resistance, and thereby decreasing production cost and improving animal welfare.

Evaluating genotype quality and concordance across genotyping platforms, we observed that distribution of genotype quality across various platforms tends to differ, likely due to different chemistries and allelic detection approaches used for genotyping. For example, the majority of *SowPro90* SNPs (91.6%) had a SNP call rate ≥ 98% while for *Porcine SNP60 BeadArray* the majority of SNPs (91.8%) had a SNP call rate ≥ 94% (Figure 3.3) suggesting that these platforms used different stringency levels when calling genotypes. For these specific ranges, a high genotype concordance rate (≥ 98.5%) between platforms was observed. Based on these observations it is not ideal to use the same threshold parameters for quality evaluations across different genotyping platforms. The approach used in this study, assessing genotype concordance between two genotyping platforms at different SNP and sample call rates, allowed identification of specific quality thresholds necessary to retain the maximum number of SNPs and samples with high quality. This strategy will be helpful when integrating data from various genotyping sources for different applications such as genomic evaluations and genome-wide association.

LITERATURE CITED

Berry, D. P., A. O'Brien, E. Wall, K. McDermott, S. Randles, P. Flynn, S. Park, J. Grose, R. Weld, and N. McHugh. 2016. Inter- and intra-reproducibility of genotypes from sheep technical replicates on Illumina and Affymetrix platforms. Genet. Sel. Evol. 48:86. doi:10.1186/s12711-016-0267-0

Boichard, D., M. Boussaha, A. Capitan, D. Rocha, C. Hoze, M.-P. Sanchez, T. Tribout, R. Letaief, P. Croiseau, C. Grohs, W. Li, C. Harland, C. Charlier, M. S. Lund, G. Sahana, M. Georges, S. Barbier, W. Coppieters, S. Fritz, and B. Guldbrandtsen. 2018. Experience from large scale use of the EuroGenomics custom SNP chip in cattle Proceedings of the World Congress on Genetics Applied to Livestock Production No. Molecular Genetics 4. p 675.

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, An. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, A. B. Gabriel, D. Altshuler, and M. J. Daly. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491-498. doi:10.1038/ng.806

Engle, T. B., E. E. Jobman, T. W. Moural, A. M. McKnite, J. W. Bundy, S. Y. Barnes, E. H. Davis, J. A. Galeota, T. E. Burkey, G. S. Plastow, S. D. Kachman, and D. C. Ciobanu. 2014. Variation in time and magnitude of immune response and viremia in experimental challenges with Porcine circovirus 2b. BMC. Vet. Res. 10:286. doi:10.1186/s12917-014-0286-4

Fan, B., S. K. Onteru, Z. Q. Du, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2011. Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. PLoS One. 6: e14726. doi:10.1371/journal.pone.0014726

Fan, B., S. K. Onteru, B. E. Mote, T. Serenius, K. J. Stalder, and M. F. Rothschild. 2009. Large-scale association study for structural soundness and leg locomotion traits in the pig. Genet. Sel. Evol. 41:14. doi:10.1186/1297-9686-41-14

Graves, K. L. 2015. Factors associated with puberty onset and reproductive performance of gilts. Graduate Theses and Dissertations. 14580. https://lib.dr.iastate.edu/etd/14580

Groenen, M. A. M. 2015. Development of a high-density Axiom® porcine genotyping array to meet research and commercial needs. Plant and Animal Genome Conference, January 10-14, 2015, San Diego, CA

Hong, H., L. Xu, J. Liu, W. D. Jones, Z. Su, B. Ning, R. Perkins, W. Ge, K. Miclaus, L. Zhang, K. Park, B. Green, T. Han, H. Fang, C. G. Lambert, S. C. Vega, S. M. Lin,

N. Jafari, W. Czika, R. D. Wolfinger, F. Goodsaid, W. Tong, and L. Shi. 2012. Technical Reproducibility of Genotyping SNP Arrays Used in Genome-Wide Association Studies. PLoS One. 7:e44483. doi:10.1371/journal.pone.0044483

Jiang, L., D. Willner, P. Danoy, H. Xu, and M. A. Brown. 2013. Comparison of the performance of two commercial genome-wide association study genotyping platforms in Han Chinese samples. G3 (Bethesda) 3:23-29. doi:10.1534/g3.112.004069

Keel, B. N., D. J. Nonneman, and G. A. Rohrer. 2017. A survey of single nucleotide polymorphisms identified from whole-genome sequencing and their functional effect in the porcine genome. Anim. Genet. 48:404-411. doi:10.1111/age.12557

Kreikemeier, C. A., T. B. Engle, K. L. Lucot, S. D. Kachman, T. E. Burkey, and D. C. Ciobanu. 2015. Genome-wide analysis of TNF-alpha response in pigs challenged with porcine circovirus 2b. Anim. Genet. 46:205-208. doi:10.1111/age.12262

Lucot, K. L., M. L. Spangler, M. D. Trenhaile, S. D. Kachman, and D. C. Ciobanu. 2015. Evaluation of reduced subsets of single nucleotide polymorphisms for the prediction of age at puberty in sows. Anim. Genet. 46:403-409. doi:10.1111/age.12310

Miller, P. S., R. Moreno, and R. K. Johnson. 2011. Effects of restricting energy during the gilt developmental period on growth and reproduction of lines differing in lean growth rate: responses in feed intake, growth, and age at puberty. J. Anim. Sci. 89:342-354. doi:10.2527/jas.2010-3111

Mullen, M., M. McClure, J. F. Kearney, S. Waters, R. Weld, P. Flynn, C. Creevey, A. R. Cromie, and D. Berry. 2013. Development of a custom SNP chip for dairy and beef cattle breeding, parentage, and research. INTERBULL Nantes. 47.

Nonneman, D.J., C.A. Lents, T.S. Kalbfleisch, J.L.Vallet, and G.A. Rohrer. 2016a. Potential functional variants associated with age at puberty in a validation population of swine. In: Plant and Animal Genome Conference, January 9-13, 2016, San Diego, CA

Nonneman, D. J., J. F. Schneider, C. A. Lents, R. T. Wiedmann, J. L. Vallet, and G. A. Rohrer. 2016b. Genome-wide association and identification of candidate genes for age at puberty in swine. BMC. Genetics. 17:50. doi:10.1186/s12863-016-0352-y

Purfield, D. C., M. McClure, and D. P. Berry. 2016. Justification for setting the individual animal genotype call rate threshold at eighty-five percent. J. Anim. Sci. 94:4558-4569. doi:10.2527/jas.2016-0802

Rowland, R., J. Lunney, and J. Dekkers. 2012. Control of porcine reproductive and respiratory syndrome (PRRS) through genetic improvements in disease resistance and tolerance. Front. Genet. 3(260)(Review) doi:10.3389/fgene.2012.00260

Serenius, T., and K. J. Stalder. 2006. Selection for sow longevity. J. Anim. Sci. 84 Suppl:E166-171.

Tart, J. K., R. K. Johnson, J. W. Bundy, N. N. Ferdinand, A. M. McKnite, J. R. Wood, P. S. Miller, M. F. Rothschild, M. L. Spangler, D. J. Garrick, S. D. Kachman, and D. C. Ciobanu. 2013. Genome-wide prediction of age at puberty and reproductive longevity in sows. Anim. Genet. 44:387-397. doi:10.1111/age.12028

Trenhaile, M. D., J. L. Petersen, S. D. Kachman, R. K. Johnson, and D. C. Ciobanu. 2016. Long-term selection for litter size in swine results in shifts in allelic frequency in regions involved in reproductive processes. Anim. Genet. 47:534-542. doi:10.1111/age.12448

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, and M. A. DePristo. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 43:11.10.11-33. doi:10.1002/0471250953.bi1110s43

Walker, L. R., T. B. Engle, H. Vu, E. R. Tosky, D. J. Nonneman, T. P. L. Smith, T. Borza, T. E. Burkey, G. S. Plastow, S. D. Kachman, and D. C. Ciobanu. 2018. Synaptogyrin-2 influences replication of Porcine circovirus 2. PLoS. Genet. 14: e1007750. doi:10.1371/journal.pgen.1007750

Walker, L. R., E. E. Jobman, K. M. Sutton, J. Wittler, R. K. Johnson, and D. C. Ciobanu. 2019. Genome-wide association analysis for porcine reproductive and respiratory syndrome virus susceptibility traits in two genetic populations of pigs1. J. Anim. Sci 97:3253-3261. doi: 10.1093/jas/skz184

Wijesena, H. R., C. A. Lents, J. J. Riethoven, M. D. Trenhaile-Grannemann, J. F. Thorson, B. N. Keel, P. S. Miller, M. L. Spangler, S. D. Kachman, and D. C. Ciobanu. 2017. GENOMICS SYMPOSIUM: Using genomic approaches to uncover sources of variation in age at puberty and reproductive longevity in sows. J. Anim. Sci. 95:4196-4205. doi:10.2527/jas2016.1334

Wijesena, H. R., G. A. Rohrer, D. J. Nonneman, B. N. Keel, J. L. Petersen, S. D. Kachman, and D. C. Ciobanu. 2019. Evaluation of genotype quality parameters for SowPro90, a new genotyping array for swine. J. Anim. Sci. 97:3262-3273. doi: 10.1093/jas/skz185

Wysoker, A., K. Tibbetts, and T. Fennel. 2013. Picard tools version 1.90. Available online at http://picard.sourceforge.net (Accessed 14 December 2016)

CHAPTER 4: FINE MAPPING GENETIC VARIANTS ASSOCIATED WITH AGE AT PUBERTY AND SOW FERTILITY USING *SOWPRO90* GENOTYPING ARRAY

INTRODUCTION

Improving reproductive traits (e.g. litter size, reproductive longevity) via traditional breeding and quantitative genetic approaches faces challenges given these traits are lowly heritable, sex-limited, express late in life, and are polygenic, with each gene variant having a relatively small effect. Age at puberty is an early and indirect indicator of sow fertility (Serenius and Stalder, 2006; Tart et al., 2013). Gilts that reach puberty earlier have a greater probability of producing more litters during their lifetime. Identification of common genetic sources that influence variation of age at puberty and fertility traits has the potential to facilitate genomic prediction for reproductive longevity.

Age at puberty is a moderately heritable trait (mean h^2 = 0.32, 16 studies reviewed by Bidanel, 2011), but traditional selection for this trait is not widely practiced in commercial settings due to extensive time and labor requirements for daily estrus detection via boar exposure. With the aim of improving accuracy of genomic prediction for sow fertility traits, we developed a custom SNP panel (*SowPro90*) including candidate genetic variants associated with age at puberty and fertility traits identified in resource populations developed at University of Nebraska-Lincoln (UNL) and US Meat Animal Research Center (USMARC) (Wijesena et al., 2019).

A Bayesian mixture model that fits haplotypes into association analysis (Bayes

interval mapping, BayesIM) (Kachman, 2015) instead of individual SNPs (e.g. BayesB, BayesC) was used to infer *SowPro90* haplotypes to the entire UNL population, previously genotyped with *Porcine SNP60 BeadArray* to fine map major genetic variants associated with age at puberty. This study was also intended to evaluate the effect of major QTL regions identified for age at puberty using *SowPro90* on other fertility traits such as litter size and reproductive longevity in the UNL and commercial maternal crossbred populations.

MATERIALS AND METHODS

This study was approved by the University of Nebraska-Lincoln (**UNL**) Institutional Animal Care and Use Committee.

The Resource Population

The UNL swine resource population was extensively genotyped and phenotyped for several fertility and growth traits to study the role of genetics and nutrition on reproductive development and longevity of sows. The females used in this study were developed in 18 non-overlapping batches (B) (n = 2,054). The dams of the experimental gilts were either Large White × Landrace crossbreds or Nebraska Index Line (NIL) while the sires were from two unrelated commercial Landrace lines. During the 117 day development (pre-breeding) period (123 to 240 days of age), gilts were subjected to four main experimental diets (with or without phase feeding) over the 18 batches. The diets included, 1) *ad libitum* standard corn-soybean based diet, 2) energy-restricted diet with

approximately 20% less metabolizable energy (ME), 3) energy and lysine restricted diet, and 4) a diet containing high protein and lysine with same ME as in the standard diet and ME:lysine ratio as in the restricted diet. A detailed description of the resource population and dietary treatments can be found in Miller et al. (2011) and Trenhaile (2015). In addition, 2,309 animals from two maternal commercial populations were used representing 1) Landrace, Yorkshire (paternal genetic lines of UNL resource population), and 2) Landrace × Large White crossbred lines (Wijesena et al., 2019).

Phenotypes

Estrous detection in the UNL population began at approximately 130 days of age and the age at which gilts expressed first estrus was defined as the age at puberty. Estrus detection was carried out daily by exposing the gilts to a mature intact boar for 15 min in an adjacent pen until all the gilts expressed estrus twice within a development pen or until they reached 240 days of age. The females were maintained up to four parities and reproductive data were collected unless they died or culled for reproductive or structural failure.

The commercial Landrace × Large White crossbred females entered five different farms as gilts during 2013. Reproductive data were available from up to eight parities unless the females were culled due to reproductive issues, old age or other structural and health reasons. It should also be noted that age at puberty data were not available for commercial females.

The reproductive data available in both UNL and commercial data sets included litter size traits such as total number of piglets born in the first two parities (TNB-P1,

TNB-P2), number born alive (NBA-P1, NBA-P2), and reproductive longevity defined as lifetime number of parities (LTNP) produced (up to four parities in UNL and up to three or more parities in commercial sows). For LTNP, the number of parities produced was recorded up until reproductive failure and females that were culled due to non-reproductive reasons were edited out of the data.

Genotyping

Tail snips or ear notches were collected from UNL gilts (n = 2,054) shortly after birth. DNA was isolated from these tissues using the DNeasy or Puregene tissue kits (Qiagen, Valencia, CA). The quality of DNA was assessed using a NanoDrop (Thermo Fisher Scientific Inc., Waltham, MA), Epoch (BioTek Inc., Winooski, VT) spectrophotometers or Qubit fluorometer (Thermo Fisher Scientific Inc., Waltham, MA). Gilts from B1 to B14 (n = 1,556) were genotyped with *Porcine SNP60 BeadArray* (Illumina Inc., San Diego, CA). Gilts from B15 to B17 (n = 375), commercial crossbred females (n = 1,972), and commercial maternal parental lines (n = 314) were genotyped with *SowPro90* custom SNP array (Thermo Fisher Scientific Inc., Waltham, MA) (Wijesena et al., 2019). A subset of gilts from B1 to B14 (UNL extreme gilts; n = 270) representing the tails of the distribution for *Porcine SNP60 BeadArray* (53,529 SNPs) derived genomic estimated breeding values (GEBVs) for age at puberty (early, n=147; late, n=123) were also genotyped with *SowPro90*. After initial quality evaluation of *SowPro90*, the array content was updated (*SowPro91*) excluding monomorphic and non-recommended SNPs (Wijesena et al., 2019) and including novel SNPs located in genes across the genome and Swine Leukocyte Antigen complex. The last two UNL batches

(B18 and B19, n = 256) were genotyped with *SowPro91*. *Porcine SNP60 BeadArray* genotypes were filtered at Illumina quality score ≥ 0.4 and sample and SNP call rate ≥ 80% leaving 53,529 SNPs for further analysis. *SowPro90* was filtered at ≥ 93% sample and ≥ 97% SNP call rate leaving 86,452 SNPs for downstream analysis.

Infer *SowPro90* Haplotypes and Genome-wide Association Analyses Using BayesIM

A reference population including a subset of UNL sows (n = 768) and sows and boars from the parental genetic lines (n = 314) genotyped with *SowPro90* and *SowPro91* were used to infer *SowPro90* haplotypes to the entire UNL population previously genotyped with *Porcine SNP60 BeadArray* (n = 1,286) using BayesIM (Kachman, 2015). Only SNPs that were common between *SowPro90* and *SowPro91* (n = 77,695 SNPs) were used for B18 females genotyped with *SowPro91* (n = 123).

A haplotype-based genome-wide association (GWAS) using BayesIM was performed to estimate the proportion of genetic variance explained by high-density SNP genotypes for age at puberty (B1 to B18, n = 2,054) and LTNP (B1 to B17, n = 1,853) in the UNL population. The SNPs were mapped to Sscrofa11.1 reference genome assembly (http://igenomes.illumina.com.s3-website-us-east-1.amazonaws.com/Sus_scrofa/Ensembl/Sscrofa11.1/Sus_scrofa_Ensembl_Sscrofa11.1.tar.gz – [accessed November 7, 2016]). The BayesIM was performed setting the π value to 0.99, assuming that the probability of a nonzero haplotype effect at a given locus is 0.01 (1– π), QTL frequency to 50 Kb (n = 47,756 haplotype windows), number of haplotype states to eight, average haplotype length to 250 Kb, and number of iterations to estimate haplotype parameters to 25. The Markov Chain Monte Carlo chain included 82,000

samples with the first 1,000 being discarded as burn in (Wilson-Wells and Kachman, 2016; Schweer et al., 2018). The fixed effects included contemporary group (batch and diet) and genetic line. Random effects included litter/dam and developmental pen. Genetic variance and haplotype effects were estimated for each 50 Kb QTL region.

Correlation Between Genomic Estimated Breeding Value for Age at Puberty and LTNP

The haplotype windows (50 Kb) were ranked based on the genetic variance explained for age at puberty and the top five distinct QTL regions located on four chromosomes were identified. These regions were extended in both directions to obtain a 1-Mb QTL interval. Genomic estimated breeding values for age at puberty were calculated for all the gilts (n = 2,054) for five major 1-Mb windows using PullRegions option in BayesIM. Similarly, GEBVs were calculated for LTNP (n = 1,853) for major 1-Mb QTL regions for age at puberty and 16 1-Mb sliding windows by 250 Kb from the top window location in both directions. The pairwise correlation between GEBV for age at puberty and LTNP for major QTL regions and adjacent sliding windows was calculated using JMP (version Pro 14.1.0, SAS Institute Inc., Cary, NC). Haplotype effects were calculated for age at puberty and LTNP for the 5-Mb region (region between the 16 sliding 1-Mb windows) flanking the top QTL regions for age at puberty.

Single Marker Association and Linkage Disequilibrium Analyses

A general linear mixed model was used to test the additive effect of single SNPs located in the major 1-Mb windows (plus 0.2 Mb flanking region in both directions) on

age at puberty in a subset of gilts (B1 to B14, n = 270) genotyped with *SowPro90*. These gilts represented extremes of the distribution for *Porcine SNP60 BeadArray* derived GEBVs for age at puberty (UNL extreme gilts). Association of these SNPs with other targeted fertility traits (e.g. LTNP, TNB P1, P2, and NBA P1, P2) was also tested in the same subset of UNL sows. In addition, SNP effects were tested in the subsequent batches of UNL sows (B15 to B17, n = 375) and also in commercial Landrace × Large White crossbred sows (n = 1,972) genotyped with *SowPro90* for age at puberty (only UNL data), LTNP, and litter size traits (UNL and commercial data). In the UNL population, the model included SNP genotype as a covariate, contemporary group (batch and diet) as a fixed effect and sire, litter/dam, and development pen as random effects. In the commercial crossbred population, SNP genotype was fitted as a covariate, contemporary group (farm and entry month), birth farm, and farm entry age were fitted as fixed effects and sire and dam/litter were fitted as random effects. In order to obtain least square means for each genotype the SNP was treated as a fixed effect. Pairwise comparison of least square means for each genotype was based on Tukey test. Linkage Disequilibrium (LD) was assessed in the top QTL regions (e.g. SSC7 and SSC14 region 1) using *SowPro90* genotypes and HAPLOVIEW (Barrett et al., 2005).

Gene Ontology and Variant Effect Predictor Analyses

The top QTL regions (plus 0.2 Mb flanking region on both directions) were characterized for positional candidate genes using the Sscrofa build 11.1. The candidate genes and their gene ontology terms were obtained using BIOMART tool in the Ensembl database (version 99;

http://uswest.ensembl.org/biomart/martview/e0267c5280d49b8815f710187ba39839 [accessed 10 February 2020]). The SIFT scores and consequences of potential functional SNPs were obtained using Variant Effect Predictor (VEP) in the Ensembl database (https://uswest.ensembl.org/Tools/VEP [accessed 4 March 2020]).

RESULTS AND DISCUSSION

Development of *SowPro90*

To improve genomic prediction accuracy for lowly heritable traits that express late in life such as reproductive longevity, this research was aimed to identify pleiotropic genetic variants associated with age at puberty and other fertility traits. A novel custom SNP panel, *SowPro90* was developed by saturating with SNPs in genes located in the major QTL regions for age at puberty identified in UNL and USMARC populations (using *Porcine SNP60 BeadArray*) (Wijesena et al., 2019). *SowPro90* included 7X more SNPs (n = 13,511 SNPs) overlapping QTL regions for age at puberty in the UNL data set compared to *Porcine SNP60 BeadArray* (n = 1,812 SNPs).

Fine Mapping Genomic Regions Associated with Age at Puberty and Reproductive Longevity Using *SowPro90*

The majority of the UNL population (B1 to B14) was genotyped with *Porcine SNP60 BeadArray*, while the last five batches were genotyped with *SowPro90* (B15 to B17) and *SowPro91* (B18 and B19). In order to fine map the genomic regions and genetic variants affecting age at puberty, a novel BayesIM approach was implemented to infer

SowPro90 haplotypes in all animals previously genotyped with *Porcine SNP60 BeadArray* (Figure 4.1). As a reference data set to infer haplotypes, we used UNL samples genotyped with *SowPro90* and *SowPro91* and commercial parental lines genotyped with *SowPro90*. The sires of the UNL resource population originated from these commercial lines.

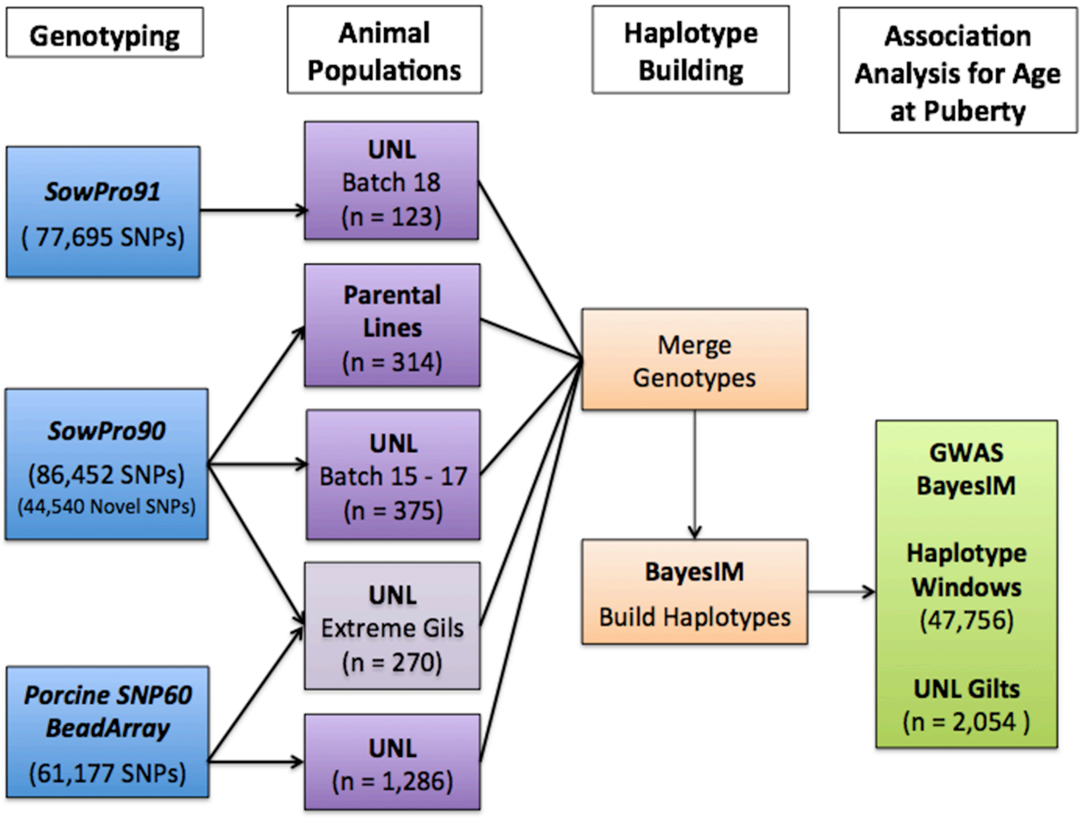

Figure 4.1: Schematic representation of the approach used for haplotype-based association analysis using BayesIM.

This approach allowed us to assign haplotypes to the entire UNL population and to utilize them in a haplotype-based association analysis (n = 47,756 haplotype windows) for age at puberty (n = 2,054) and LTNP (n = 1,853). For age at puberty, five major 1-Mb

QTL regions were identified on SSC2 (13.6 to 14.6 Mb), SSC7 (83.5 to 84.5 Mb), SSC14 (38.7 to 39.7 Mb, region 1; 45.05 to 46.05 Mb, region 2), and SSC18 (37.35 to 38.35 Mb) (Figure 4.2). The QTL on SSC2, SSC7, and SSC14 also overlapped the top 1% major 1-Mb windows for age at puberty identified in the previous GWAS in the UNL population (n = 1,644) using *Porcine SNP60 BeadArray* (Wijesena et al., 2017). Four of these top QTL regions were enriched with SNPs in *SowPro90* and *SowPro91* (Wijesena et al., 2019).

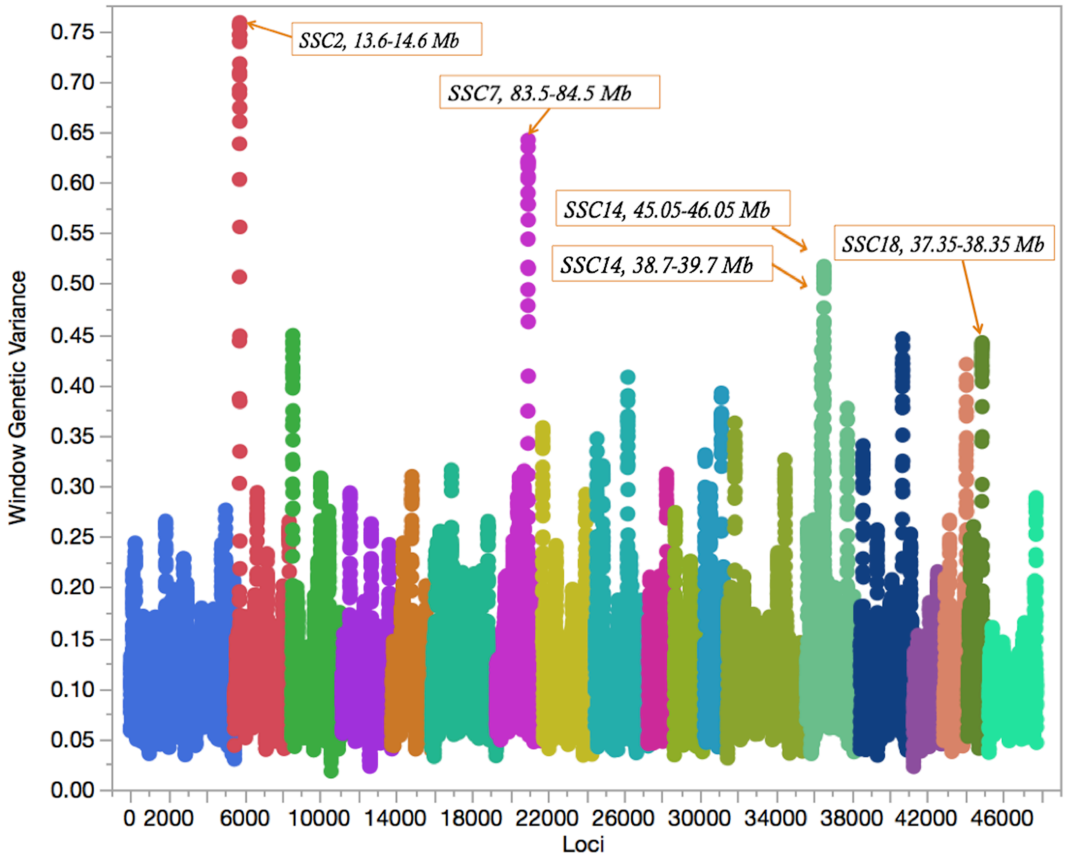

Figure 4.2: Genome-wide association analysis for age at puberty using *SowPro90*. The autosomes, from SSC1 to SSC18, followed by chromosome X are represented by different colors. Each dot represents a 50 Kb haplotype window. The boxed labels indicate the top five QTL regions for age at puberty.

Haplotype Effects and Correlation Between Genomic Estimated Breeding Values for Age at Puberty and LTNP

We expected that some of the top regions identified for age at puberty would have an effect on other fertility phenotypes including reproductive longevity. While the major QTL for age at puberty were distinct, due to nature of the phenotype, the QTL for LTNP were less defined. To evaluate the effect of these major age at puberty QTL on reproductive longevity, the correlation between GEBVs for age at puberty and LTNP (n = 1,845) was calculated for the top five regions. Additionally, the GEBVs were calculated for LTNP in 16 1-Mb sliding windows by 250 Kb from the top windows for age at puberty in both directions (2.5 Mb distant from the center of the top windows on each direction). We used this approach to observe the extent and the decay of the correlation between age at puberty and reproductive longevity (Figure 4.3). As expected, negative correlations between age at puberty and LTNP were observed across top age at puberty QTL regions (r = -0.96 to -0.10) ($P < 0.0001$). As the sliding windows for LTNP shifted away from the major QTL locations for age at puberty, the correlation diminished reaching zero (Figure 4.3).

To further investigate the effect of these QTL regions on both traits, the haplotype effects (n = 8 haplotypes) were calculated for each region for age at puberty and LTNP. For all the QTL regions, the haplotypes with largest effects on age at puberty had an opposite effect on LTNP. As expected, the effects for LTNP were not as prominent as in age at puberty. For example, different directional effects for age at puberty and LTNP were observed for haplotype 3 located on SSC7 QTL region (Figure 4.4). None of these

top QTL regions associated with age at puberty were among the top QTL identified for LTNP. We consider LTNP is a complex, composite phenotype, largely affected by environment, making it difficult to fine map the underlying genetic sources driving the phenotypic variation.

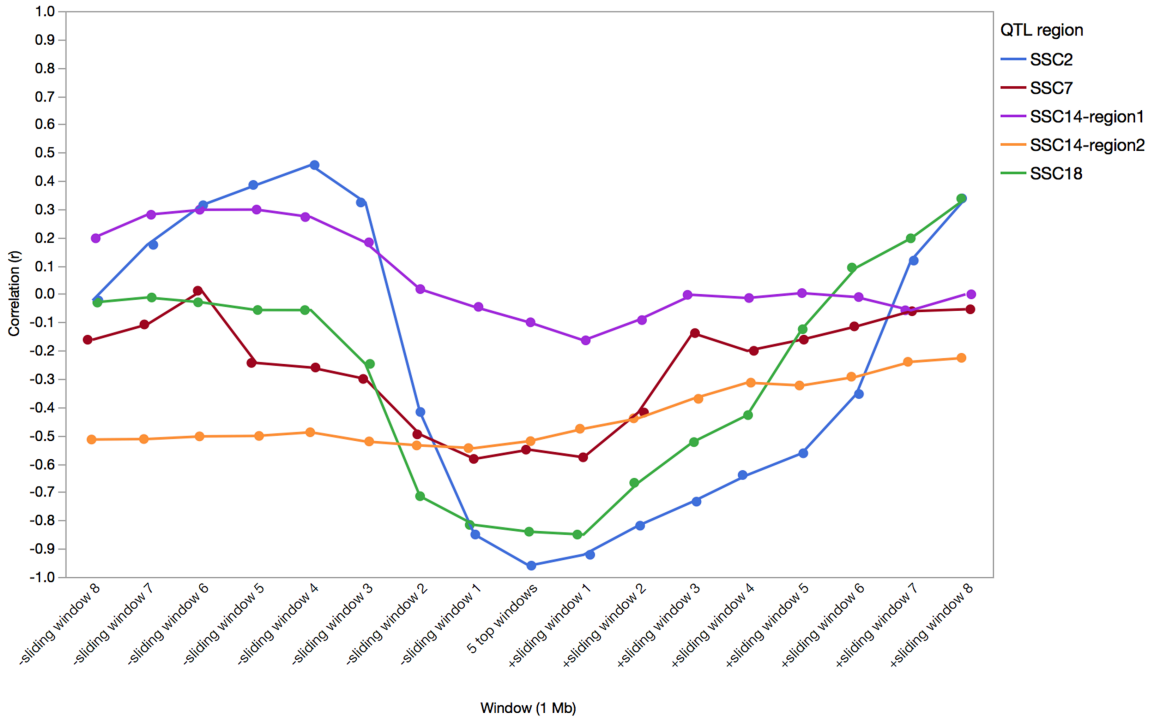

Figure 4.3: Correlation of genomic estimated breeding values for age at puberty (top windows) and LTNP (top windows for age at puberty and adjacent sliding windows). Adjacent windows for LTNP are sliding by 250 Kb. The distance between the center of the top QTL and the last sliding window is 2.5 Mb. All the regions show negative correlations at the top QTL window and the correlation reaches zero as moving away from the major QTL regions.

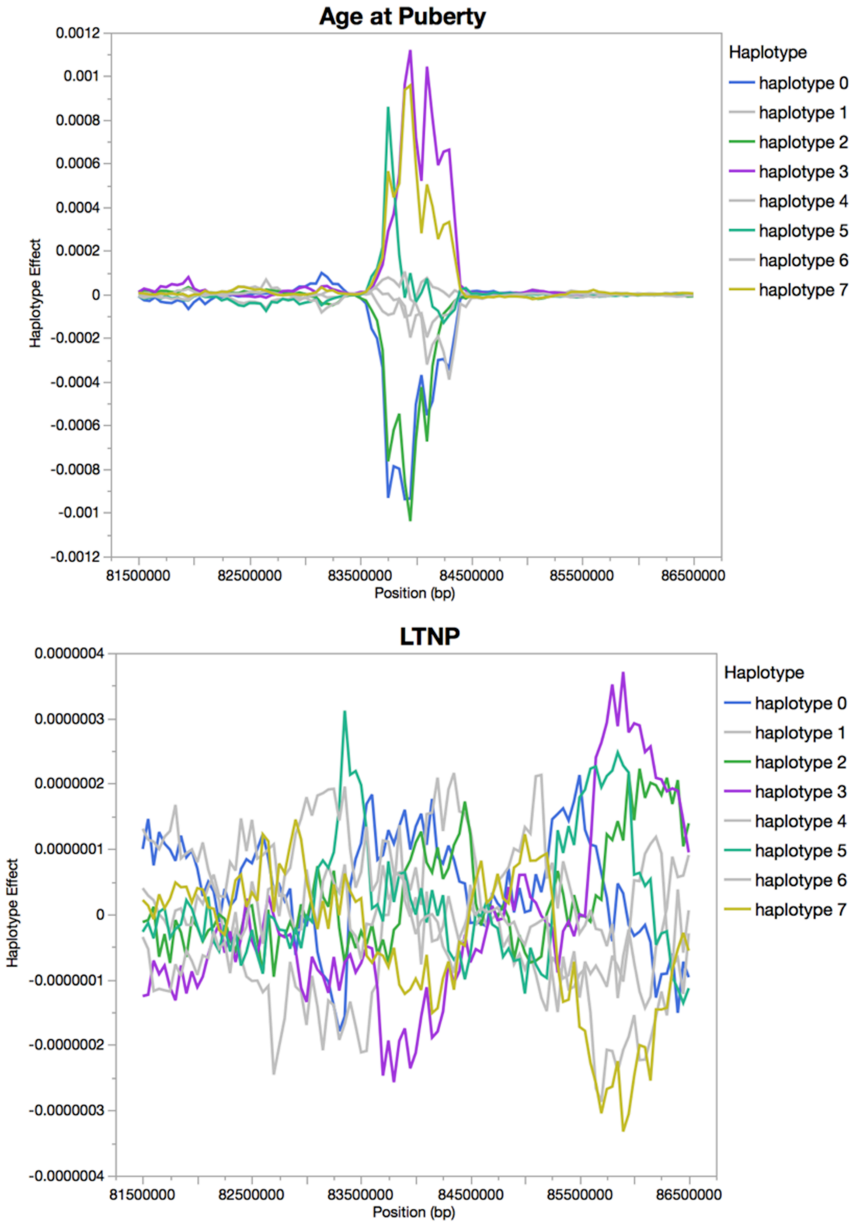

Figure 4.4: Haplotype effects for SSC7 QTL (5-Mb region including the sliding windows). Haplotypes with largest effect on age at puberty were shown in different colors while the haplotypes with smaller effects were shown in gray for both traits. An opposite direction of the haplotype effects was observed between age at puberty and LTNP. For example, Haplotype 3 (purple) was associated with increasing age at puberty and decreasing LTNP.

Single Marker Association and Linkage Disequilibrium Analyses

In order to identify major SNPs associated with fertility traits in UNL and commercial populations, all SNPs (n = 510 SNPs) representing the five major QTL regions for age at puberty were used in a single marker association analysis. Initially, the SNP effects on age at puberty were estimated in UNL extreme gilts (B1 to B14) using *SowPro90* genotypes. Of all the SNPs tested, 35.5% had significant ($P < 0.05$) additive effects for age at puberty. Of those significant SNPs, 71.3% were located in 29 genes. All the SNPs (n = 510 SNPs) were also evaluated in a data set consisting of gilts from subsequent UNL batches (B15 to B17). Nineteen percent of the SNPs showed significant additive effects on age at puberty, while 6.5% of the SNPs were significant in both UNL data sets. Age at puberty was not available in the commercial data set. Therefore, we were unable to validate these SNPs for age at puberty across populations.

Ten SNPs with the largest additive effects ($P < 0.05$) on age at puberty (in the UNL extreme gilts) were selected for each QTL region (n = 50 SNPs) to evaluate their impact on other fertility traits (LTNP, TNB, and NBA) in the UNL extreme and commercial data sets. In the UNL extreme data set, a subset of SNPs had significant effects on LTNP (16%) and TNB-P1 (18%). In the commercial data set, 2% of the SNPs had a suggestive effect ($P < 0.15$) on LTNP, however, the expected opposite trend between age at puberty (decrease) and LTNP (increase) was not observed. In the same data set a subset of SNPs were identified with significant or suggestive effects for TNB-P1 (14%), TNB-P2 (2%), NBA-P1 (26%), and NBA-P2 (6%).

SSC2 QTL Region (13.4 to 14.8 Mb)

The QTL located on this SSC2 region showed evidence of pleiotropy for age at puberty and litter size traits in the UNL population. The SNP with the largest additive effect on age at puberty (*AX-116162218*, 14.7 Mb, $P = 0.0001$) explained 24 days difference in age at puberty between homozygote genotypes ($P = 0.04$) in the UNL extreme gilts. As the number of favorable alleles increased, age at puberty decreased by 11.6 days ($P = 0.0001$) and both TNB-P1 and NBA-P1 increased by 0.54 ($P = 0.21$) and 0.88 ($P = 0.08$) piglets per litter, respectively. For TNB, a numerical increase continued in P2 with a suggestive additive effect of 1.2 piglets per litter ($P = 0.12$) (Table 4.1). The expected impact of this top SNP on LTNP was not observed in any of the data sets (Table 4.1).

The top SNP (*AX-116162218*) in this region was located in an intergenic region. The closest gene to this SNP was a small nuclear RNA (*U6*, ~8 Kb downstream of the SNP). Seven of the other top SNPs with the largest effects on age at puberty were located in three genes (*P2RX3*, *SSRP1*, and *PTPRJ*). The SNPs were located in either coding (synonymous) or 3' untranslated region (UTR). Three of these synonymous SNPs significantly associated with age at puberty were located in *P2RX3*, a gene involved in implantation and pregnancy (Slater et al., 2000). This QTL region was previously identified as a major QTL and a potential selection sweep for litter size traits in the UNL resource population (Trenhaile et al., 2016). Polymorphisms in the *P2RX3* gene were fixed in Nebraska Index Line (NIL), a line extensively selected for litter size traits for over 29 generations (Trenhaile et al., 2016). The UNL experimental gilts were originated

from NIL dams and that could explain the observed effect of this region on litter size traits in our population.

SSC7 QTL Region (83.3 to 84.7 Mb)

The SNP associated with the largest effect on age at puberty (*AX-116689678*, 83.3 Mb, $P = 0.01$) also explained some of the phenotypic variation in LTNP and NBA-P1. As the number of favorable alleles increased, age at puberty decreased by 6 days in UNL extreme gilts ($P = 0.01$) and by 4.1 days in subsequent UNL batches ($P = 0.02$), while LTNP and NBA-P1 increased by 0.37 litters ($P = 0.03$) and 0.53 ($P = 0.24$) piglets per litter, respectively in the UNL extreme sows (Table 4.1, Figure 4.5).

Since this SNP was significantly associated with LTNP, we expected to see a change in allele frequency across parities in the UNL population. The frequency of the favorable allele increased from 0.41 in sows that were unable to generate a parity to 0.48 in sows that produced four parities (Figure 4.6). Similarly, in the commercial data set (n = 904) the frequency of the same allele increased from 0.18 in sows unable to generate a parity to 0.19 in sows that produced three or more parities (Figure 4.6), even though the effect of this SNP on LTNP was not significant in this data set ($P = 0.72$).

There were no known annotated genes identified in this QTL region in the current swine reference genome. Therefore, this region was extended by 0.5 Mb in both directions. *NR2F2* (83.07 Mb), is a candidate gene located upstream (~280 Kb) of the top SNP, and is implicated in fertilization, progesterone receptor signaling during embryonic implantation, and litter size in pigs (Chen et al., 2016). Chen et al. (2016) reported a -204*C>A* SNP located in the promoter region of *NR2F2* gene that disrupts the binding

site for CREP transcription factor. The *CC* genotype was associated with higher litter size in Large White sows.

Four SNPs in *NR2F2* gene were included in *SowPro90*. The SNPs were located in the proximal promoter (-1,044 bp; *AX-134892687*), 5' UTR (*AX-135052973*), and intron 1 (*AX-123958897*, *AX-179489698*). The LD (r^2) between these SNPs with the top SNP in the SSC7 QTL ranged from 0.32 to 0.68. The effects of these SNPs on fertility traits were evaluated in a subsequent analysis (Table 4.1). These SNPs showed significant additive effects for age at puberty in the UNL B15 to B17 ($P < 0.05$) gilts but suggestive ($P < 0.15$) or non-significant effects in the UNL extreme gilts (Table 4.1). Three of the *NR2F2* SNPs (*AX-135052973*, *AX-123958897*, *AX-179489698*) showed expected (opposite) direction of the association for age at puberty ($P < 0.05$) (UNL B15 to B17) and LTNP in the commercial data set (although the effects were not significant, $P = 0.34$ to 0.56) (Table 4.1). The lack of missense polymorphisms and differential expression for this gene (Wijesena et al., 2017; Chapter 2), lead us to explore isoform diversity as a genetic source of phenotypic variation. Two *NR2F2* isoforms were identified in the UNL population using RNA sequencing data from the hypothalamic arcute nucleus (ARC) of gilts with different pubertal status. The same two transcripts were also predicted in public databases (e.g. ensembl.org). The isoforms were translated into two distinct proteins (281 and 414 amino acids). Further characterization of this gene could provide insights into the role of these SNPs on differential splicing and function of the different isoforms on phenotypic variation in fertility traits.

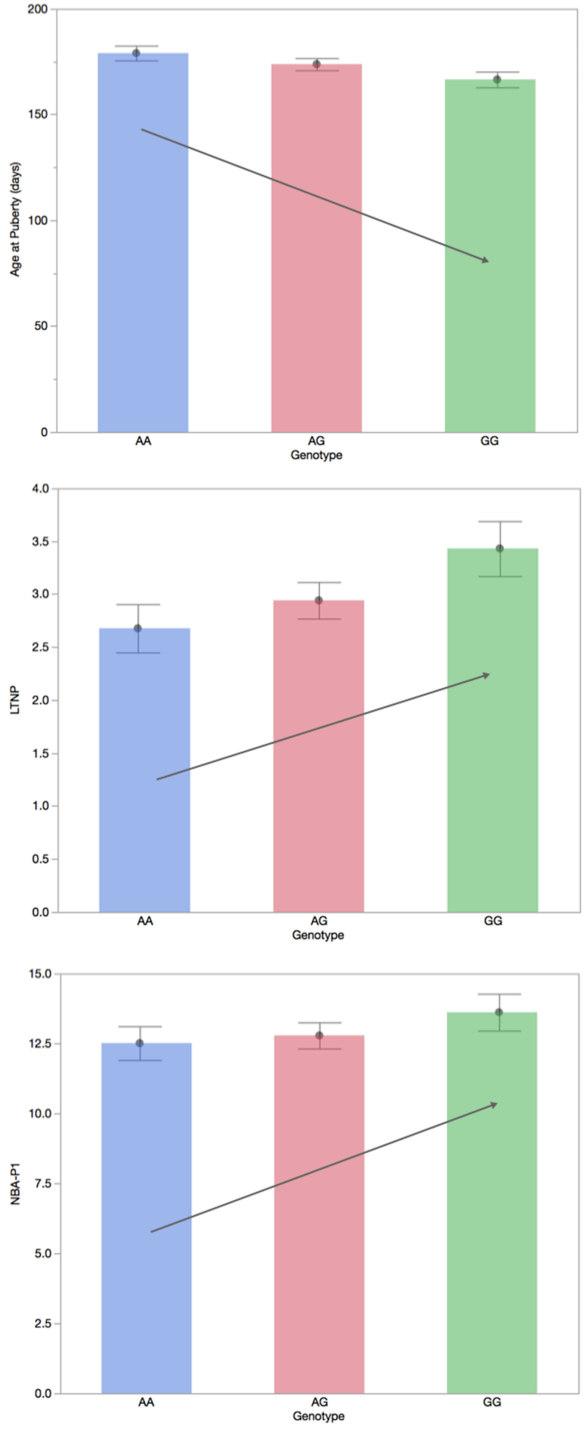

Figure 4.5: Age at puberty (*P* = 0.01), LTNP (*P* = 0.03), and NBA-P1 (*P* = 0.24) least square means and standard errors for *AX-116689678* SNP genotypes in the UNL extreme data set.

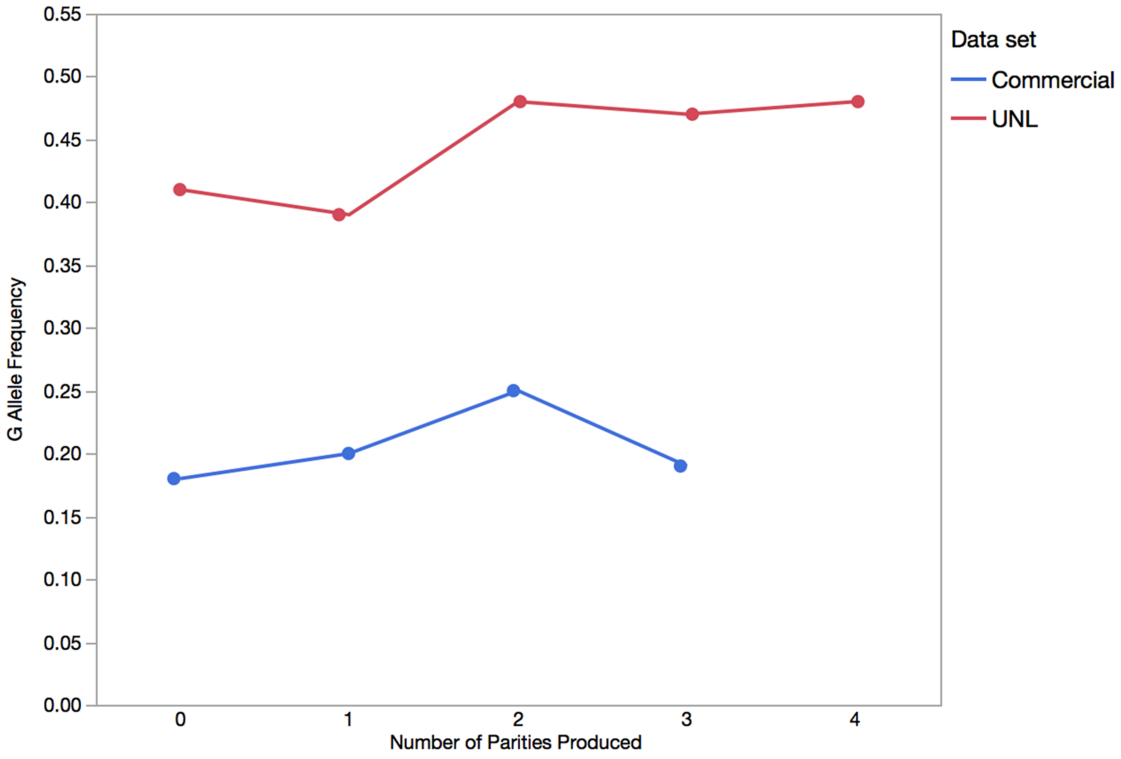

Figure 4.6: The frequency of favorable (*G*) *AX-116689678* allele across parities in the UNL and commercial populations. The allele frequency increased from 0.41 in sows that were unable to generate a parity to 0.48 in sows that produced four parities in the UNL population (n = 645). Similarly, an increase in the allele frequency was observed from 0.18 in sows unable to generate a parity to 0.19 in sows that produced three or more parities in the commercial population (n = 904).

SSC14 QTL Region 1 (38.5 to 39.9 Mb)

The QTL mapped on SSC14 (38.7 to 39.7 Mb) was characterized by high LD in both UNL and commercial populations. The SNP associated with the largest effect on age at puberty (*AX-141921242*, 39.2 Mb, $P < 0.0001$) was in high LD ($r^2 > 0.95$) with next nine SNPs with the largest effects for age at puberty in the UNL population. In UNL extreme gilts, as the number of favorable alleles for this SNP increased, age at puberty significantly decreased (13.1 days; $P < 0.0001$), while there was a suggestive increase in LTNP (0.29 litters; $P = 0.10$) (Table 4.1; Figure 4.7). Due to high LD in the region, similar effects were observed for other top nine SNPs ranging from 11.9 to 13.1 days for age at puberty ($P < 0.0001$) and 0.24 to 0.29 litters for LTNP ($P = 0.16$ to 0.19). The expected effect of these SNPs on age at puberty was not observed in the subsequent UNL batches (B15 to B17) as the frequency of favorable homozygous genotype was very low (0.53%) across these SNPs. Eight SNPs including the top SNP showed suggestive effects ($P < 0.15$) for NBA-P1 in the commercial data set (Table 4.1; Figure 4.7). The additive effects of these SNPs ranged from 0.16 to 0.23 piglets per litter.

These 10 SNPs with the largest effects on age at puberty were located in intronic and untranslated regions of seven genes (*OAS1*, *OAS2*, *RPH3A*, *PTPN11*, *HECTD4*, *TRAFD1*, and *NAA25*). Since there were no obvious missense mutations among these top SNPs that could affect the protein sequence and function, we expect that some of these SNPs would affect gene expression, RNA stability or splicing events. However, none of these genes were differentially expressed in the hypothalamic ARC of UNL gilts.

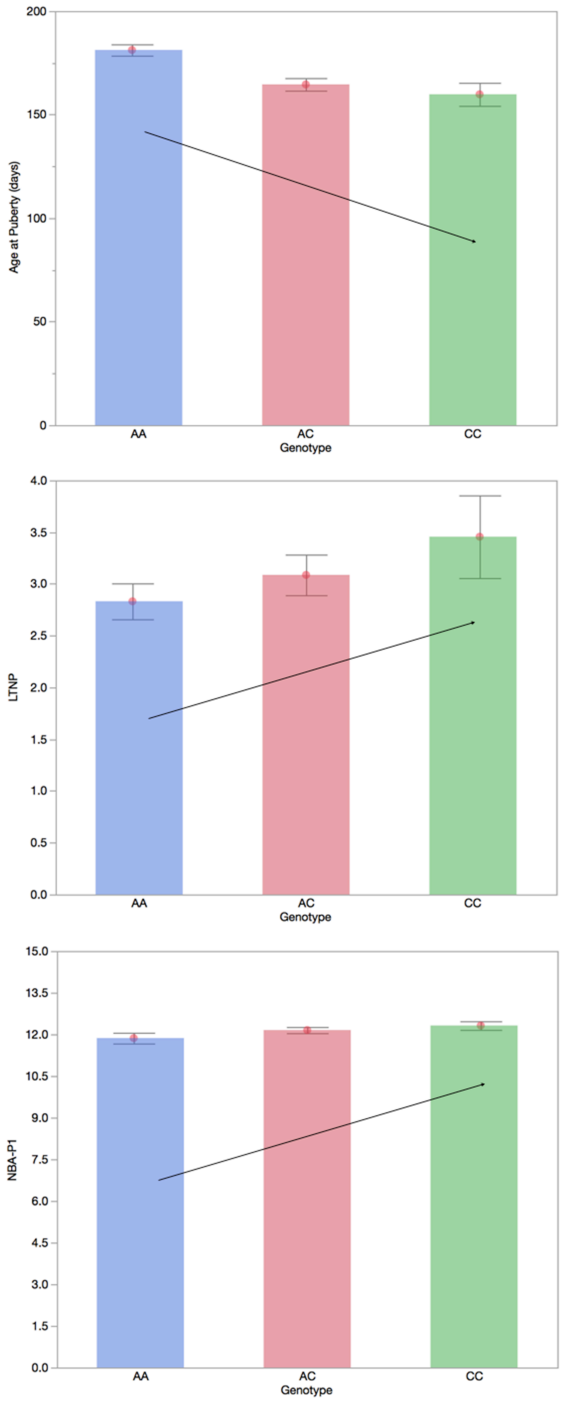

Figure 4.7: Age at puberty (UNL; $P < 0.0001$), LTNP (UNL; $P = 0.10$), and NBA-P1 (Commercial; $P = 0.06$) least square means and standard errors for *AX-141921242* SNP genotypes in the UNL extreme and commercial data sets.

The top SNP (*AX-141921242*) was located in the 3' UTR of *PTPN11* gene while two other SNPs (among the top 10 SNPs) were located in intron 1 of the same gene. Three alternatively spliced *PTPN11* isoforms were identified in the UNL population using RNA sequence data of the hypothalamic ARC. There were five predicted *PTPN11* isoforms and amino acid sequences available in the Ensembl database (ensembl.org). *PTPN11* is widely involved in male fertility mainly via regulating proliferation and survival of spermatogonial stem cells, cellular adhesion in Sertoli cells to protect germ cells, and steroid (e.g. testosterone) production in Leydig cells (Puri and Walker, 2016).

Outside the selected top 10 SNPs, there were two missense SNPs associated with age at puberty (*AX-179504119* and *AX-140952109*; $P < 0.0001$) and were located in *OAS1* and *OAS2* genes, respectively. Both SNPs had significant ($P < 0.05$) additive effects on NBA-P1 in the commercial data set. Based on the SIFT score (> 0.05), these amino acid substitutions did not have deleterious effects on the protein function. *OAS1* is a gene implicated in fertility traits in mammals. Promoter and exonic polymorphisms were associated with reproductive traits (e.g. age at first breeding and calving, pregnancy rate) in cattle of Indian origin (Alex et al., 2018a; Alex et al., 2018b). In mice, a novel transcript (*OAS1D*) was identified encoding an OAS1-like protein with 59% sequence identity with *OAS1*. It is exclusively expressed in growing oocytes and early embryos. Deficiency in *OASD1* caused defects in ovarian follicular development and decreased efficiency in ovulation leading to reduced fertility (Yan et al., 2005).

Table 4.1: Single marker association results of the SNPs that represent top QTL regions for age at puberty on fertility traits

SNP	Chr	Position (bp)	Genotype	Trait[1]	Data Set[2]	*N*	*P*[3]	Least Square Means ± Standard Error[4]		
								11	12	22
AX-116162218	2	14770110	*C/T*	AP	UNL-Ex	270	0.0001	155.69 ± 13.50	168.19 ± 2.56	179.70 ± 2.82
					UNL B15-17	375	0.7913	167.65 ± 4.79	162.54 ± 1.95	163.37 ± 2.20
				LTNP	UNL-Ex	259	0.6206	2.69 ± 0.98	2.94 ± 0.17	3.04 ± 0.19
					Commercial	904	0.6493	2.70 ± 0.15	2.95 ± 0.07	2.90 ± 0.06
				TNB-P1	UNL-Ex	189	0.2139	15.56 ± 2.02	14.64 ± 0.37	14.12 ± 0.42
					Commercial	1,675	0.0872	12.73 ± 0.28	12.94 ± 0.13	12.87 ± 0.12
				TNB-P2	UNL-Ex	119	0.1224	15.58 ± 4.30	13.80 ± 0.50	12.68 ± 0.59
					Commercial	1,359	0.9746	13.96 0.36	13.39 0.17	13.61 0.15
				NBA-P1	UNL-Ex	189	0.0843	14.85 ± 2.43	13.48 ± 0.41	12.63 ± 0.46
					Commercial	1,675	0.4527	12.15 ± 0.28	12.25 ± 0.13	12.08 ± 0.12
				NBA-P2	UNL-Ex	119	0.1255	12.54 ± 4.65	12.54 ± 0.51	11.28 ± 0.61
					Commercial	1,359	0.9798	13.39 ± 0.35	12.79 ± 0.17	13.01 ± 0.16
AX-116689678	7	83373091	*A/G*	AP	UNL-Ex	270	0.0105	179.06 ± 3.56	173.75 ± 2.87	166.52 ± 3.71
					UNL B15-17	375	0.0166	167.83 ± 2.42	161.59 ± 1.94	160.18 ± 2.74
				LTNP	UNL-Ex	259	0.0278	2.68 ± 0.23	2.94 ± 0.17	3.43 ± 0.26
					Commercial	904	0.7173	2.90 ± 0.06	2.90 ± 0.08	3.05 ± 0.23
				TNB-P1	UNL-Ex	189	0.8738	14.46 ± 0.51	14.45 ± 0.42	14.33 ± 0.51
					Commercial	1,675	0.1455	12.96 ± 0.11	12.73 ± 0.14	12.67 ± 0.38
				TNB-P2	UNL-Ex	119	NA	13.73 ± 0.72	13.35 ± 0.52	12.90 ± 0.70
					Commercial	1,359	0.8540	13.50 ± 0.14	13.68 ± 0.19	13.14 ± 0.48
				NBA-P1	UNL-Ex	189	0.2377	12.51 ± 0.60	12.78 ± 0.47	13.62 ± 0.66
					Commercial	1,675	0.8869	12.16 ± 0.11	12.13 ± 0.14	12.16 ± 0.39
				NBA-P2	UNL-Ex	119	0.8244	11.81 ± 0.84	12.28 ± 0.62	11.63 ± 0.80
					Commercial	1,359	0.4233	12.86 ± 0.15	13.11 ± 0.19	12.83 ± 0.47

SNP	Chr	Position (bp)	Genotype	Trait[1]	Data Set[2]	*N*	*P*[3]	Least Square Means ± Standard Error[4]		
								11	**12**	**22**
AX-134892687 (*NR2F2*)	7	83091580	*C/T*	AP	UNL-Ex	270	0.1305	170.23 ± 2.90	175.50 ± 2.95	177.01 ± 5.00
					UNL B15-17	375	0.0065	155.98 ± 3.52	163.12 ± 1.90	167.53 ± 2.41
Promoter				LTNP	UNL-Ex	259	0.0594	3.26 ± 0.19	2.79 ± 0.18	2.69 ± 0.34
					Commercial	904	0.4942	1.17 ± 0.84	2.98 ± 0.09	2.88 ± 0.06
				TNB-P1	UNL-Ex	189	0.7901	14.39 ± 0.43	14.39 ± 0.42	14.68 ± 0.76
					Commercial	1,675	0.6246	11.41 ± 2.09	12.96 ± 0.16	12.85 ± 0.10
				TNB-P2	UNL-Ex	119	NA	12.51 ± 0.51	13.82 ± 0.50	15.58 ± 1.10
					Commercial	1,359	0.1628	12.87 ± 3.39	13.78 ± 0.21	13.46 ± 0.14
				NBA-P1	UNL-Ex	189	0.7719	12.81 ± 0.48	12.95 ± 0.48	13.06 ± 0.91
					Commercial	1,675	0.2338	10.06 ± 2.10	12.33 ± 0.16	12.09 ± 0.11
				NBA-P2	UNL-Ex	119	0.1762	11.37 ± 0.57	12.20 ± 0.6	13.84 ± 1.23
					Commercial	1,359	0.1762	11.61 ± 3.37	13.18 ± 0.21	12.86 ± 0.14
AX-135052973 (*NR2F2*)	7	83090047	*C/T*	AP	UNL-Ex	270	0.2192	171.32 ± 2.86	174.51 ± 2.98	177.69 ± 5.26
					UNL B15-17	375	0.0042	160.46 ± 2.12	163.22 ± 1.98	170.10 ± 2.83
5' UTR				LTNP	UNL-Ex	259	0.1729	3.19 ± 0.19	2.81 ± 0.18	2.85 ± 0.36
					Commercial	904	0.3378	2.95 ± 0.14	2.94 ± 0.07	2.86 ± 0.07
				TNB-P1	UNL-Ex	189	0.8959	14.50 ± 0.42	14.25 ± 0.43	14.97 ± 0.82
					Commercial	1,675	0.4241	12.83 ± 0.24	12.81 ± 0.12	12.96 ± 0.12
				TNB-P2	UNL-Ex	119	NA	12.39 ± 0.50	14.02 ± 0.51	15.07 ± 1.14
					Commercial	1,359	0.5872	13.58 ± 0.32	13.60 ± 0.16	13.47 ± 0.16
				NBA-P1	UNL-Ex	189	0.6371	12.80 ± 0.47	12.94 ± 0.49	13.32 ± 0.97
					Commercial	1,675	0.6064	12.21 ± 0.25	12.18 ± 0.12	12.11 ± 0.13
				NBA-P2	UNL-Ex	119	0.0765	11.24 ± 0.56	12.49 ± 0.59	13.13 ± 1.28
					Commercial	1,359	0.5150	12.95 ± 0.32	13.03 ± 0.16	12.85 ± 0.17

SNP	Chr	Position (bp)	Genotype	Trait[1]	Data Set[2]	*n*	*P*[3]	Least Square Means ± Standard Error[4]		
								11	12	22
AX-179489698	7	83084320	*C/T*	AP	UNL-Ex	270	0.3096	171.47 ± 2.89	174.00 ± 2.96	177.07 ± 5.51
(*NR2F2*)					UNL B15-17	375	0.0289	161.83 ± 2.11	162.04 ± 1.99	171.00 ± 3.09
Intron 1				LTNP	UNL-Ex	259	0.1158	3.21 ± 0.19	2.80 ± 0.18	2.78 ± 0.38
					Commercial	904	0.4730	2.98 ± 0.17	2.92 ± 0.07	2.88 ± 0.07
				TNB-P1	UNL-Ex	189	0.6476	14.37 ± 0.44	14.32 ± 0.43	15.05 ± 0.87
					Commercial	1,675	0.3970	12.63 ± 0.29	12.86 ± 0.12	12.92 ± 0.12
				TNB-P2	UNL-Ex	119	0.0377	12.47 ± 0.56	14.04 ± 0.58	14.98 ± 1.35
					Commercial	1,359	0.3050	13.78 ± 0.37	13.61 ± 0.16	13.45 ± 0.16
				NBA-P1	UNL-Ex	189	0.9842	13.01 ± 0.47	12.85 ± 0.47	13.17 ± 1.00
					Commercial	1,675	0.5568	12.10 ± 0.29	12.22 ± 0.13	12.08 ± 0.12
				NBA-P2	UNL-Ex	119	0.1358	11.36 ± 0.58	12.48 ± 0.59	13.03 ± 1.37
					Commercial	1,359	0.2122	13.16 ± 0.37	13.04 ± 0.17	12.82 ± 0.16
AX-123958897	7	83088085	*A/G*	AP	UNL-Ex	270	0.0885	175.47 ± 3.71	175.91 ± 2.82	166.89 ± 3.59
(*NR2F2*)					UNL B15-17	375	0.0038	167.76 ± 2.38	162.17 ± 2.02	156.40 ± 3.71
Intron 1				LTNP	UNL-Ex	259	0.0146	2.57 ± 0.24	2.96 ± 0.17	3.40 ± 0.25
					Commercial	904	0.5561	2.90 ± 0.05	2.97 ± 0.12	NA
				TNB-P1	UNL-Ex	189	0.1907	14.87 ± 0.53	14.51 ± 0.40	13.89 ± 0.55
					Commercial	1,675	0.9868	12.89 ± 0.09	12.91 ± 0.22	10.01 ± 2.97
				TNB-P2	UNL-Ex	119	NA	14.67 ± 0.96	12.97 ± 0.50	13.00 ± 0.81
					Commercial	1,359	0.2663	13.49 ± 0.13	13.83 ± 0.29	12.93 ± 3.39
				NBA-P1	UNL-Ex	189	0.7808	13.16 ± 0.64	12.79 ± 0.46	12.95 ± 0.63
					Commercial	1,675	0.6732	12.14 ± 0.10	12.27 ± 0.22	8.99 ± 2.97
				NBA-P2	UNL-Ex	119	NA	13.24 ± 0.83	11.54 ± 0.49	11.80 ± 0.71
					Commercial	1,359	0.0987	12.87 ± 0.13	13.38 ± 0.29	11.68 ± 3.36

SNP	Chr	Position (bp)	Genotype	Trait[1]	Data Set[2]	*n*	*P*[3]	Least Square Means ± Standard Error[4]		
								11	12	22
AX-141921242	14	39203611	*A/C*	AP	UNL-Ex	270	<0.0001	181.24 ± 2.74	164.62 ± 3.02	159.83 ± 5.57
					UNL B15-17	375	0.0577	164.46 ± 1.77	160.10 ± 2.12	165.29 ± 10.68
				LTNP	UNL-Ex	259	0.1046	2.83 ± 0.17	3.09 ± 0.20	3.46 ± 0.40
					Commercial	904	0.4854	2.92 ± 0.11	2.93 ± 0.06	2.85 ± 0.08
				TNB-P1	UNL-Ex	189	0.4084	14.18 ± 0.4	15.02 ± 0.44	14.20 ± 0.79
					Commercial	1,675	0.1888	12.71 ± 0.19	12.87 ± 0.11	13.02 ± 0.15
				TNB-P2	UNL-Ex	119	NA	12.83 ± 0.60	13.76 ± 0.64	13.50 ± 1.22
					Commercial	1,359	0.7512	13.56 ± 0.25	13.50 ± 0.15	13.63 ± 0.20
				NBA-P1	UNL-Ex	189	0.6806	12.80 ± 0.46	13.69 ± 0.50	12.28 ± 0.94
					Commercial	1,675	0.0630	11.86 ± 0.19	12.15 ± 0.11	12.32 ± 0.16
				NBA-P2	UNL-Ex	119	NA	11.94 ± 0.61	11.91 ± 0.68	12.62 ± 1.19
					Commercial	1,359	0.2988	12.84 ± 0.25	12.89 ± 0.15	13.12 ± 0.20

[1]AP: Age at puberty (days). LTNP: lifetime number of parities (litters). TNB: total number born (piglets/litter). NBA: number born alive (piglets/litter)
[2]UNL-Ex: UNL extreme gilts (Batch 1-13). UNL B15-17: UNL gilts (Batch 15-17). Commercial: Large White × Landrace maternal crossbreds
[3]*P*-value for overall test of the effect of genotypes
[4]The alleles (1 and 2) are designated based on the alphabetical order of SNP variants (A, C, G and T)

CONCLUSIONS

The goal of this study was to fine map the genetic sources that explain variation in age at puberty and other fertility traits to enable accurate prediction of traits with low heritability and expressed late in life such as reproductive longevity. For this purpose, a custom SNP array (*SowPro90*) enriched in SNPs located in genes overlapping QTL regions for age at puberty was developed. Since only a subset of UNL gilts were genotyped with *SowPro90*, a novel BayesIM approach was utilized to infer and assign *SowPro90* haplotypes to the entire UNL population.

Five QTL regions on four chromosomes (SSC2, SSC7, SSC14, and SSC8) with the largest effects on age at puberty were identified in the UNL swine resource population using a haplotype based GWAS. As expected, a negative correlation (r = -0.96 to -0.10; $P < 0.0001$) was observed between GEBVs for age at puberty and LTNP at the major QTL sites for age at puberty. The haplotypes with largest effects on age at puberty showed an opposite effect on LTNP at each QTL region.

In this study we discovered QTL regions and genetic variants that explained variation in multiple fertility traits including age at puberty, LTNP, and litter size in the UNL population. Approximately 70% of the SNPs significantly associated with age at puberty were located in genes. Among them, *P2RX3*, *NR2F2*, *PTPN11*, and *OAS1* were candidate genes for fertility traits. The SNPs were located in coding (missense, synonymous), intronic, and untranslated regions of these genes. Some of the SNPs identified in our experimental population associated with age at puberty were evaluated

for their effects on LTNP and litter size traits in a commercial data set. Suggestive associations were observed for some of these SNPs with litter size traits but not with LTNP. A reason for this could be that reproductive longevity is a composite phenotype largely affected by environment. These candidate polymorphisms can be further characterized to understand their role in gene expression, splicing process, protein sequence and function, and how these changes affect the fertility phenotypes.

LITERATURE CITED

Alex, R., K. Ramesha, U. Singh, S. Kumar, R. Alyethodi, R. Deb, S. Rai, S. Sharma, G. Sengar, A. Kumar, and B. Prakash. 2018a. Association analysis of novel polymorphisms in 2', 5'-oligoadenylate synthetase gene with reproductive traits in indigenous and cross-bred cattle of Indian Origin. Reprod. Domest. Anim. 53:442-449. doi: 10.1111/rda.13129

Alex, R., K. P. Ramesha, U. Singh, S. Kumar, R. R. Alyethodi, R. Deb, S. Sharma, G. S. Sengar, A. Kumar, and B. Prakash. 2018b. Promoter variants of OAS1 gene are associated with reproductive performance and incidence of normal calving in cattle. Theriogenology. 108:255-261. doi: 10.1016/j.theriogenology.2017.12.002

Barrett J.C., B. Fry, J. Maller, and M. J. Daly. 2005. HAPLOVIEW: analysis and visualization of LD and haplotype maps. Bioinformatics. 21, 263–5. doi: 10.1093/bioinformatics/bth457

Bidanel, J.P. 2011. Biology and genetics of reproduction. In The Genetics of the Pig. 2nd Edition. Edited by Rothschild, M.F. and Ruvinsky, A. Wallingford, UK: CAB International; 218-241.

Chen, X., J. Fu, and A. Wang. 2016. Expression of genes involved in progesterone receptor paracrine signaling and their effect on litter size in pigs. J. Anim. Sci. Biotechno. 7:31. doi: 10.1186/s40104-016-0090-z

Kachman, S. D. 2015. Genomic prediction model based on haplotype clusters. In: Joint Stat. Meet., August 11, 2015, Seattle, WA. https://ww2.amstat.org/meetings/jsm/2015/onlinepro-gram/AbstractDetails.cfm?abstractid=314884 (Accessed 25 February 2016)

Miller, P.S., R. Moreno, and K. Johnson. 2011. Effects of restricting energy during the gilt developmental period on growth and reproduction of lines differing in lean growth rate: Responses in feed intake, growth, and age at puberty. J. Anim. Sci. 89: 342-354. doi: 10.2527/jas.2010-3111

Puri, P., and W. H. Walker. 2016. The regulation of male fertility by the PTPN11 tyrosine phosphatase. Semin. Cell. Dev. Biol. 59:27-34. doi: 10.1016/j.semcdb.2016.01.020

Schweer, K. R., S. D. Kachman, L. A. Kuehn, H. C. Freetly, J. E. Pollak, and M. L. Spangler. 2018. Genome-wide association study for feed efficiency traits using SNP and haplotype models1. J. Anim. Sci. 96:2086-2098. doi:10.1093/jas/sky119

Serenius, T., and K. J. Stalder. 2006. Selection for sow longevity. J. Anim. Sci. 84 Suppl:E166-171.

Tart, J. K., R. K. Johnson, J. W. Bundy, N. N. Ferdinand, A. M. McKnite, J. R. Wood, P. S. Miller, M. F. Rothschild, M. L. Spangler, D. J. Garrick, S. D. Kachman, and D. C. Ciobanu. 2013. Genome-wide prediction of age at puberty and reproductive longevity in sows. Anim. Genet. 44:387-397. doi: 10.1111/age.12028

Trenhaile, M. D. 2015. Genomic analysis of sow reproductive traits: Identification of selective sweeps, major genes, and genotype by diet interactions. MS Thesis, University of Nebraska, Lincoln.

Trenhaile, M. D., J. L. Petersen, S. D. Kachman, R. K. Johnson, and D. C. Ciobanu. 2016. Long-term selection for litter size in swine results in shifts in allelic frequency in regions involved in reproductive processes. Anim. Genet. 47:534-542. doi:10.1111/age.12448

Wijesena, H. R., C. A. Lents, J. J. Riethoven, M. D. Trenhaile-Grannemann, J. F. Thorson, B. N. Keel, P. S. Miller, M. L. Spangler, S. D. Kachman, and D. C. Ciobanu. 2017. GENOMICS SYMPOSIUM: Using genomic approaches to uncover sources of variation in age at puberty and reproductive longevity in sows. J. Anim. Sci. 95:4196-4205. doi: 10.2527/jas2016.1334

Wijesena, H. R., G. A. Rohrer, D. J. Nonneman, B. N. Keel, J. L. Petersen, S. D. Kachman, and D. C. Ciobanu. 2019. Evaluation of genotype quality parameters for SowPro90, a new genotyping array for swine1. J. Anim. Sci. 97:3262-3273. doi: 10.1093/jas/skz185

Wilson-Wells, D. F., and S. D. Kachman. 2016. A Bayesian GWAS method utilizing haplotype clusters for a composite breed population. 2016. Annual Conference on Applied Statistics in Agriculture; May 1-3, 2016; Manhattan, KS. https://newprairiepress.org/cgi/viewcontent.cgi?article=1469&context=agstatconference (Accessed 20 March, 2020)

Yan, W., L. Ma, P. Stein, S. A. Pangas, K. H. Burns, Y. Bai, R. M. Schultz, and M. M. Matzuk. 2005. Mice Deficient in Oocyte-Specific Oligoadenylate Synthetase-Like Protein OAS1D Display Reduced Fertility. Mol. Cell. Biol. 25:4615. doi: 10.1128/MCB.25.11.4615-4624.2005