University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Papers and Publications in Animal Science   Animal Science Department

2017

GENOMICS SYMPOSIUM: Using genomic approaches to uncover sources of variation in age at puberty and reproductive longevity in sows

Hiruni R. Wijesena
University of Nebraska-Lincoln, hwijesena2@unl.edu

Clay A. Lents
USDA, ARS, USMARC, clay.lents@ars.usda.gov

Jean-Jack Riethoven
University of Nebraska-Lincoln, jeanjack@unl.edu

Melanie D. Trenhaile-Grannemann
University of Nebraska-Lincoln, mtrenhaile2@unl.edu

Jennifer F. Thorson
USDA-ARS, MARC, jennifer.thorson2@montana.edu

See next page for additional authors

Follow this and additional works at: http://digitalcommons.unl.edu/animalscifacpub

Part of the Genetics and Genomics Commons, and the Meat Science Commons

Authors

Hiruni R. Wijesena, Clay A. Lents, Jean-Jack Riethoven, Melanie D. Trenhaile-Grannemann, Jennifer F. Thorson, Brittney N. Keel, Phillip S. Miller, Matt Spangler, Stephen D. Kachman, and Daniel C. Ciobanu

GENOMICS SYMPOSIUM: Using genomic approaches to uncover sources of variation in age at puberty and reproductive longevity in sows[1,2]

H. R. Wijesena,* 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*[3]

*Department of Animal Science, †Center for Biotechnology, and ‡Department of Statistics, University of Nebraska, Lincoln 68583; and §USDA, ARS, U.S. Meat Animal Research Center, Clay Center, NE, 68933[4]

ABSTRACT: Genetic variants associated with traits such as age at puberty and litter size could provide insight into the underlying genetic sources of variation impacting sow reproductive longevity and productivity. Genomewide characterization and gene expression profiling were used using gilts from the University of Nebraska–Lincoln swine resource population (n = 1,644) to identify genetic variants associated with age at puberty and litter size traits. From all reproductive traits studied, the largest fraction of phenotypic variation explained by the Porcine SNP60 BeadArray was for age at puberty (27.3%). In an evaluation data set, the predictive ability of all SNP from high-ranked 1-Mb windows (1 to 50%), based on genetic variance explained in training, was greater (12.3 to 36.8%) compared with the most informative SNP from these windows (6.5 to 23.7%). In the integrated data set (n = 1,644), the top 1% of the 1-Mb windows explained 6.7% of the genetic variation of age at puberty. One of the high-ranked windows detected (SSC2, 12–12.9 Mb) showed pleiotropic features, affecting both age at puberty and litter size traits. The RNA sequencing of the hypothalami arcuate nucleus uncovered 17 differentially expressed genes (adjusted P < 0.05) between gilts that became pubertal early (<155 d of age) and late (>180 d of age). Twelve of the differentially expressed genes are upregulated in the late pubertal gilts. One of these genes is involved in energy homeostasis (*FFAR2*), a function in which the arcuate nucleus plays an important contribution, linking nutrition with reproductive development. Energy restriction during the gilt development period delayed age at puberty by 7 d but increased the probability of a sow to produce up to 3 parities (P < 0.05). Identification of pleiotropic functional polymorphisms may improve accuracy of genomic prediction while facilitating a reduction in sow replacement rates and addressing welfare concerns.

Key words: hypothalamus, litter size, pleiotropic, puberty, reproductive longevity, swine

INTRODUCTION

Sow reproductive longevity, or number of litters produced by a sow during her lifetime, plays an important economic role in the swine industry. Sows that express puberty early in life, conceive, and farrow more than 3 litters during their lifetime are more likely to recover the development and maintenance costs (Stalder et al., 2003; Tart et al., 2013). Selection for reproductive longevity is challenging due to its expression late in life and low heritability (Tart et al., 2013). Age at puberty was shown to be the earliest indicator of reproductive longevity. Specifically, early onset of puberty was associated with a greater probability of sows to produce multiple parities during their 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.

Age at puberty is characterized by a moderate to high heritability (mean $h^2 = 0.37$ from 16 studies; reviewed by Bidanel, 2011) compared with other reproductive traits such as litter size (mean $h^2 = 0.11$ from 118 studies; Bidanel, 2011) or reproductive longevity ($h^2 = 0.04$; 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 currently used in breeding programs with a panel of pleiotropic DNA markers. Such a panel could be used early in life and assist the selection decision of superior females without recording age at puberty. In the current study, various genomic, nutritional, and physiological approaches were integrated to determine the genetics and energy restriction effects on age at puberty and reproductive longevity in sows.

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. Currently, a total of 1,644 females have been produced in 14 batches (**B**) that have been extensively phenotyped and genotyped (Fig. 1). A detailed description of the resource population was previously reported by Miller et al. (2011). Briefly, the dams of the experimental females were Large White × Landrace (**LR**) crossbreeds (B1
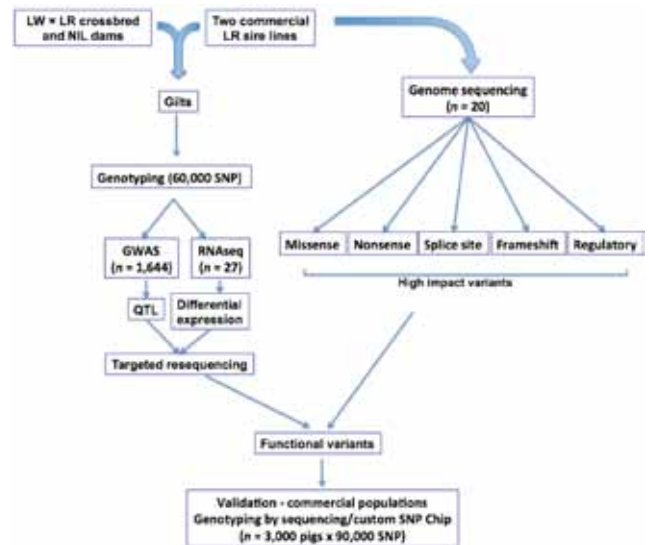
Figure 1. Schematic representation of the approaches used in the study. The dams of the experimental gilts were Large White (LW) × Landrace (LR) crossbreeds and Nebraska Index Line (NIL; University of Nebraska-Lincoln,Lincoln,NE). The dams were bred with LR boars from 2 unrelated commercial lines. The genetic approaches used included genomewide association study (GWAS) to identify QTL associated with age at puberty and genome and RNA sequencing (RNAseq) to identify functional variants. The candidate SNP will be incorporated into a custom SNP chip and validated in 3,000 commercial pigs.

to B4) and Nebraska Index Line (**NIL**; University of Nebraska-Lincoln,Lincoln,NE; B1 to B14). The NIL was developed based on commercial crossbreeds (Large White × LR) and was selected for increased litter size for 18 generations (Hsu and Johnson, 2014). The dams were bred with LR boars from 2 unrelated commercial lines. The first batches (B1 to B4) were sired by boars from the LR1 line and the remaining (B5 to B14) were sired by boars from the LR2 line. Each batch is considered a separate generation of dams and sires. The number of sires per batch varied from 5 (B13) to 12 (B3, B5, and B14), whereas the number of litters 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.

Experimental Diets

All gilts were fed a common diet from birth to 123 d of age. During the development period (123 to 240 d), until they were moved to the breeding barn, gilts were allocated to an ad libitum standard corn–soybean–based diet (diet A), an energy-restricted diet with approximately 20% less ME (diet B), or an energy- and lysine-restricted diet (diet C) as described in detail in Trenhaile et al. (2015). In the latest batch (B14), diet C (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 restricted diet. After being moved to the breeding

barn, all the animals were fed a common diet. All diets met or exceeded nutritional requirements (NRC, 2012).

Reproductive Phenotypes

Detection of age at puberty in experimental gilts began at approximately 130 d of age and continued until all the gilts expressed estrus at least twice within a development pen or until they reached 240 d of age. Detection of estrus was achieved by moving gilts once per day to an adjacent pen where they were exposed in the same pen to a mature intact boar for 15 min. Age at puberty was defined as the age at which a gilt first expressed estrus (Miller et al., 2011). The experimental females were maintained through 4 parities unless they died or were culled. Culling usually occurred due to failure to express estrus before 240 d 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 mummified and stillborn, and lifetime number of parities produced (**LT-NP**), were recorded for as many as 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 by Tart et al. (2013).

Genotyping

Tail snips or ear notches were collected from gilts shortly after birth, and DNA was isolated ($n = 1,644$) using the DNeasy or Puregene tissue kit (Qiagen Inc., Valencia, CA). The quality of DNA was assessed using a NanoDrop Spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA). 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 sample and SNP call rate less than 80% were removed, leaving 53,529 SNP for further analysis.

Genomewide Association Analyses

The proportion of genetic variance explained by high-density SNP genotypes for age at puberty and litter size traits in experimental gilts was estimated using a BayesB model implemented by GenSel software (Fernando and Garrick, 2008) and Bayes interval mapping (**BayesIM**) recently introduced by Kachman (2015) that fits haplotypes rather than individual SNP, as is the case of BayesB. The SNP were mapped to the Sscrofa 10.2 reference genome assembly (http://sup-port.illumina.com/sequencing/sequencing_software/igenome.html [accessed 16 March 2016]). The BayesB analysis was performed setting the pi value to 0.99, assuming that 0.01 of the SNP have a nonzero effect on the analyzed phenotype. Previous reports showed that a faster convergence is reached when a small set of SNP is expected to have a nonzero effect in each sampling (Onteru et al., 2012). Batch, diet, litter, 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 prediction value (**GPV**) 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, 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 the first 1,000 iterations discarded as burn-in. Random effects included sire, litter, and developmental pen, and fixed effects included batch and diet.

The interaction between high-density SNP genotypes and diet on age at puberty and LT-NP was conducted using a BayesB model, including line, batch, and diet as fixed effects (GenSel package; Fernando and Garrick, 2008). The genotypes were coded to differentiate the main effect and the interactions. Forty-one thousand iterations were performed, with the first 1,000 discarded as burn-in. The pi value was set to 0.99 for main effects and 0.995 for interaction effects.

Diet-dependent SNP effects discovered via BayesB were validated using single marker association analyses. Additive and dominance general linear mixed models included batch, diet, SNP, and SNP × diet interaction as fixed effects and litter as a random effect. Individual SNP effects were also tested to characterize the effect of developmental energy intake and age at puberty on sow probability to generate parities 1 to 3. The model included age at puberty as a covariate; diet, batch, and SNP as fixed effects; and litter as a random effect.

The nonoverlapped 1-Mb windows across the genome were ranked based on the genetic variance explained for age at puberty. The top 10% of 1-Mb windows associated with largest proportion of genetic variance were extended by 0.5 Mb in both directions for gene annotation characterization of positional candidate genes using the Sscrofa 10.2 genome build and the Ensembl gene annotation for *Sus scrofa* (version 86; http://www.ensembl.org/sus_scrofa/info/index.html [accessed 16 March 2016]).

Genome Sequencing

A subset of 16 sires was selected for whole-genome sequencing, representing both ends of the distribution for the average GPV for age at puberty. Single-end sequencing was performed using Ion Proton sequencing, as described in the manufacturer's protocol (Thermo Fisher Scientific Inc.). 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 exact duplicates if they occurred 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 Bowtie 2 package, and only high-quality alignments (Phred score $\geq$ 30) were retained for downstream analysis (Langmead et al., 2009; Langmead and Salzberg, 2012). To improve SNP detection, realignment around indels was performed using GATK software tools, RealignerTargetCreator (DePristo et al., 2011) and IndelRealigner (DePristo et al., 2011), along with the dbSNP database (ftp://ftp.ncbi.nih.gov/snp/organisms/pig_9823 [accessed 1 March 2012]) followed by GATK's BaseCalibrator (DePristo et al. 2011) to reduce the effects of sequence artifacts. Genetic variants were uncovered using the multiallelic and rare-variant option of BCFtools (Narasimhan et al., 2016) using default settings.

Ribonucleic Acid Sequencing

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 different prepubertal ($n = 12$) and postpubertal 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°C until arcuate nucleus (**ARC**) isolation from the hypothalamus. Collection of the hypothalamus from prepubertal gilts was performed before boar exposure, approximately 2 wk before the gilts were 140 d of age ($n = 12$ gilts from 12 litters). The prepubertal status was confirmed by examining the ovaries at slaughter. The postpubertal group was composed of gilts fed the 3 experimental diets, A ($n = 10$), B ($n = 8$), and C ($n = 7$). Age at puberty ranged from 132 to 215 d. The ARC was isolated using a micropunch procedure. Frozen coronal sections (250 μm) were cut using a cryostat (CM1950; Leica Biosystems, Inc., Buffalo Grove, IL) and mounted onto charged microscope slides (Premier; Life Science Products Inc., Manassas, VA). Sections containing the ARC were identified based on anatomical references (Kineman et al., 1988, 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 microcentrifuge tube, placed on dry ice, and then stored at −80°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 ribonuclease-free water and RNA purified on RNEasy Mini Columns (Qiagen Inc.) according to the manufacturer's protocol for on-column digestion with deoxyribonuclease. The quantity and quality of RNA were determined by spectrophotometry (NanoDrop 8000; Thermo Fisher Scientific Inc.) and microfluidic analysis with an Agilent automated electrophoresis system (Agilent 2100 Bioanalyzer; Agilent Technologies, Foster City, CA).

The RNA sequencing was performed using Ion Proton sequencing as described in the manufacturer's protocol (Thermo Fisher Scientific Inc.). The RNA sequencing reads were aligned to the Sscrofa 10.2 reference genome, as explained in the 2-step alignment approach used for Ion Proton transcriptome data (Sun et al., 2013). Briefly, the adaptors attached to the RNA sequencing 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). The Phred33 score was used for quality trimming. Low-quality bases in the 5′ end were removed, and nucleotides with base calls less than 22 were trimmed off from the 3′ end. The filtered reads were first aligned to the Sscrofa 10.2 reference genome using TopHat (version 2.1; Trapnell et al., 2012; Sun et al., 2013). The unmapped reads from TopHat were then aligned to the reference genome using the local option of the Bowtie package (version 2.2; Langmead and Salzberg, 2012; Sun et al., 2013). This option aligns long reads to the genome by trimming the ends of the 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 et al., 2014).

Differentially expressed genes for pre- and postpubertal gilts that included gilts that exhibited puberty early (<155 d of age) or late (>180 d of age) were determined using the DESeq2 package (Love et al., 2014). DESeq2 uses a statistical approach based on a generalized linear model and a negative binomial distribution to model gene read counts and identify differentially expressed genes. The analysis was performed using the
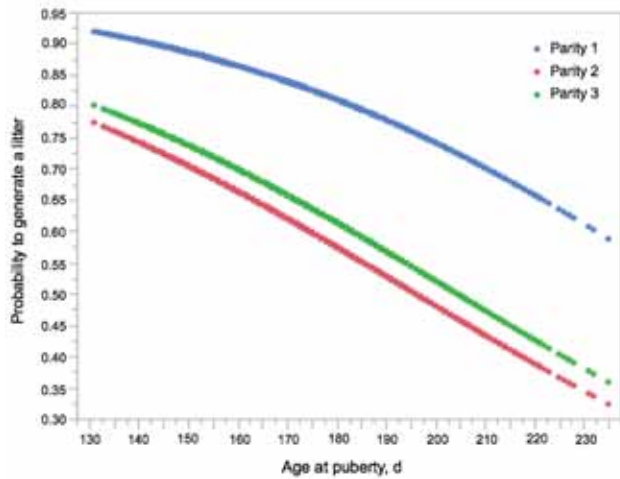
Figure 2. The effect of gilt age at puberty on the probability to produce up to 3 parities. The likelihood of a female generating a parity increased as age at puberty decreased ($P < 0.0001$).

Table 1. Posterior means of variance components of age at puberty and litter size traits based on SNP effects estimated by BayesB statistical model

Trait[1]	No.	Genetic variance	Residual variance	Total variance	Phenotypic variance explained by SNP, %
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; NBA-P1 = number of piglets born alive in parity 1; NBA-P2 = number of piglets born alive in parity 2; TNB-P1 = total number of piglets born in parity 1; TNB-P2 = total number of piglets born in parity 2.

default parameters, and a gene was considered differentially expressed at adjusted $P < 0.05$.

RESULTS AND DISCUSSION

Age at Puberty is an Indicator of Sow Reproductive Longevity

We previously reported that from a range of prebreeding gilt phenotypes (i.e., birth weight, weaning weight, age at puberty, 230-d BW, backfat thickness, and LM 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 developmental 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 sows producing up to 3 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 (Fig. 2), confirming the observed effect of age at puberty on multiple parities.

Evaluation of the Ability to Transfer Genomic Predictions Based on Major Loci Across Populations

Age at puberty had the greater 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 greatest for age at puberty (27.2%) and limited for litter size (<10%; Table 1). The genetic variation of age at puberty is affected by many loci with relatively small effects, and the probabilities of the ma-

jor 1-Mb nonoverlapping windows to have effects larger than the average windows are less than 0.30 (Fig. 3).

The effectiveness of a marker panel mainly depends on its ability to capture functional effects and predict cumulative additive genetic merit for animals not contained in the training 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 prediction. However, the expense related to measuring age at puberty necessitates that genomic information be transferred from experimental populations to potentially disjoint industry populations. Lucot et al. (2015) clearly illustrated that transferring SNP effects from a training population to various evaluation sets resulted in low correlations between GPV and adjusted phenotypes. When all the SNP from the top ranked 1, 5, 10, 20, and 50% 1-Mb windows identified in a training set (B1–B7; $n = 820$) were used in an evaluation set (B8–B11; $n = 412$) 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%). The addition of more than the top 20% of high-ranked 1-Mb windows resulted in a decline in the proportion of phenotypic variation that was explained. 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 finding is probably due to the fact that SNP identified as highest ranked in the training set are not functional variants, and the linkage disequilibrium between these SNP and the actual 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 SNP in the region identified in the training set and re-estimating their effects in the target population. Specifically, the correlations between GPV based on SNP effects estimated in the training set and the phenotype was marginal (between −0.01 to 0.17) compared with their effects retrained in the evaluation
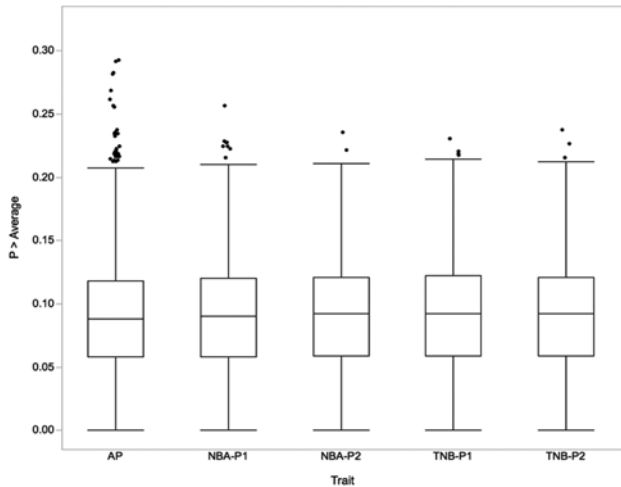
Figure 3. Box plot of the probability of 1-Mb windows having effects greater than average (including quartiles and outliers) on age at puberty (AP), number of piglets born alive in parity 1 (NBA-P1), number of piglets born alive in parity 2 (NBA-P2), total number of piglets born in parity 1 (TNB-P1), and total number of piglets born in parity 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.

set for all (0.46 to 0.81) or most informative SNP (0.30 to 0.65) from the high-ranked 1-Mb windows (Lucot et al., 2015). To further refine the transfer of information from training to target populations, elucidating causal mutations will be necessary.

Identification of Genomic Regions and Candidate Genes

To uncover pleotropic sources of variation that affect both age at puberty and reproductive longevity, we used 2 Bayesian mixture models. Genomewide association analysis performed 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 from 0.32 to 0.61% of the genetic variation for age at puberty (Fig. 4a). The top 1% major 1-Mb windows ($n = 26$) explained 6.7% of genetic variation of age at puberty. The position of the SSC2 window was located near a 1-Mb window we identified for litter size traits in the same population and also in the proximity of a potential selection sweep region (Trenhaile et al., 2016). The same region (12.8 Mb) explained the greatest fraction of the total genetic variance for age at puberty using a BayesIM model (Fig. 4b). A candidate gene in this area, *P2X3R*, is involved in placental attachment and maintenance of pregnancy. Alleles fixed in NIL but polymorphic in lines not subjected to selection indicated *P2X3R* as a potential source of the large litter size in NIL (Trenhaile et al., 2016).

In general, we observed that both SNP- (BayesB) and haplotype-based (BayesIM) models captured a portion of the same major regions associated with age at puberty, such as the one located on SSC2 (12 to 12.9 Mb; Fig. 4a and 4b). This finding is also based on a high pairwise correlation between GPV obtained from BayesB and BayesIM ($r = 0.8$), indicating that both models capture common loci responsible for genetic variation (Fig. 4c). Major regions associated with age at puberty (top 1% 1-Mb windows), such as the regions on SSC5 (4 Mb) and SSC12 (57 Mb), were 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 uncovered genes that have known postpubertal involvement, such as genes involved in fertilization (*CLIC4* [SSC6, 76 Mb] and *NR2F2* [SSC7, 88.9 Mb]), placental development (*NR2F2* [SSC7, 88.9 Mb]), progesterone secretion and luteinization (*FZD4* [SSC9, 22.9 Mb]), and pregnancy and placental attachment (*LIF* [SSC14, 67.2 Mb]; http://geneontology.org [accessed 17 March 2016]). In our previous study, Tart et al. (2013) identified *AVPR1A* as a candidate gene in a pleotropic region (SSC5, 27 to 28 Mb) associated with both age at puberty and lifetime number of parities. This gene is 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 3 nonsynonymous SNP identified in this gene (*G31E*, *G256D*, and *K377Q*). Homozygosity for the favorable *G* allele of *G31E* was associated with 5.8 d early expression of puberty and 0.53 more LT-NP (Tart et al., 2013) compared with *AA* genotype (Tart et al., 2013). Lucot et al. (2015) validated this finding in a larger data set. In the current study, the frequency of the favorable *G* allele across the resource population (B1 to B12) continuously increased from 0.42 in females unable to produce a single litter to 0.50 in females that produced 3 parities ($P < 0.05$; 0.46 in parity 1 and 0.47 in parity 2). These results illustrate how this type of polymorphism has potential to be useful in improving reproductive longevity in sow populations.

A major shortcoming for some of the high-density commercially available SNP panels is that a large proportion of the SNP do not have an assigned position in the reference genome. For example, 5,121 SNP from the Porcine SNP60 BeadArray do not have a physical location. This is particularly problematic if the desire is to use a haplotype-based statistical model (e.g., BayesIM). Genomewide association revealed that unmapped SNP represented an important group of the top 0.1% SNP associated with age at puberty (11 of the 53 SNP). For example, 2 unmapped SNP, *ASGA0092359* and *ASGA0008471*, are within the top 3 SNP for their effect on age at puberty. Linkage disequilibrium analysis across Porcine SNP60 BeadArray SNP revealed that
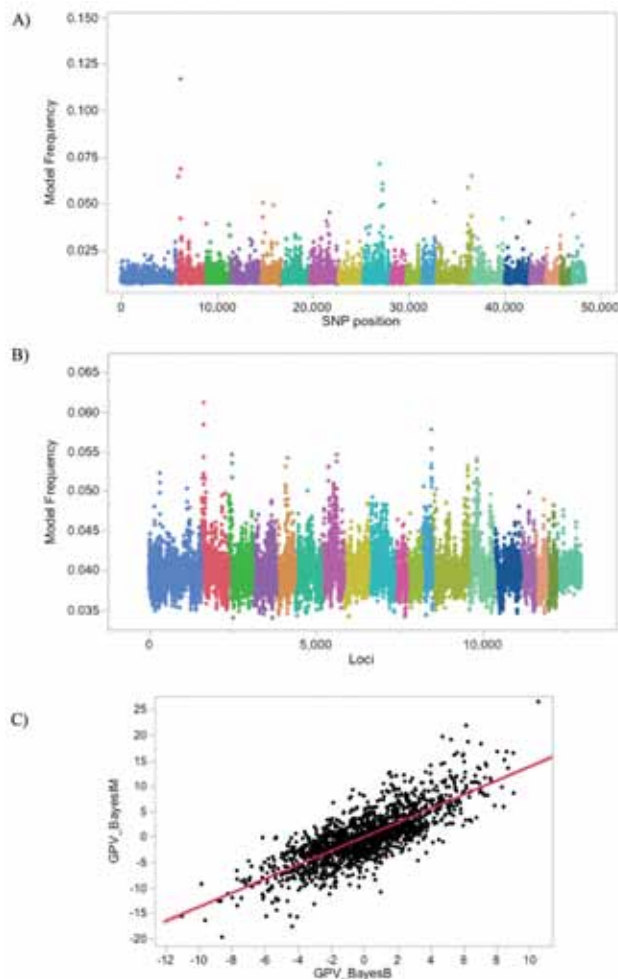
Figure 4. Genomewide association analysis for age at puberty. The autosomes, from SSC1 to 18, followed by chromosome X are represented by different colors. A) BayesB model. Each dot represents a SNP and there were 5 high-ranked 1-Mb windows including SSC2 (12 to 12.9 Mb), SSC9 (22 to 22.9, 82 to 82.9, and 106.2 to 106.9 Mb), and SSC13 (211 to 211.9 Mb) that explained from 0.32 to 0.61% of the genetic variation for age at puberty. B) Bayes interval mapping (BayesIM) model. Each dot represents a 200 kb haplotype. Both SNP- (BayesB) and haplotype-based (BayesIM) models captured some of the same high-ranked regions associated with age at puberty (SSC2, 12 to 12.9 Mb). C) Correlation of the genomic prediction values (GPV) between BayesB and BayesIM models ($r = 0.8$) indicates that both models capture common loci responsible for genetic variation.

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; $r^2 > 0.12$) based on linkage disequilibrium estimates with multiple mapped SNP from the Porcine SNP60 BeadArray. We encountered similar results in a QTL analysis for susceptibility to porcine circovirus 2 (Tosky et al., 2016). In this study, we used a chromosomal scaffold based on long sequencing reads and a combination of RNA sequencing alignments and gene prediction software to fully annotate and map the phenotypic variation. A similar approach or an updated reference genome will be used in the future to fully characterize major QTL regions for age at puberty. This process is critical because a SNP such as *ASGA0008471*

was shown to have additive pleiotropic effects ($P < 0.05$). The favorable homozygote genotype was associated with 4.3 d earlier expression of age at puberty ($n = 1,614$; $P = 0.01$) and 0.29 more LT-NP ($n = 1,214$; $P < 0.10$) compared with the alternate homozygote.

Whole-Genome Sequencing Uncovered Potential Sources of Genetic Variation

To identify genetic variants outside the limited capability of the Porcine SNP60 BeadArray, we performed next generation genome sequencing on 16 sires that represent both ends of the distribution for average daughter's GPV for age at puberty. The average number of gilts with available GPV per sequenced sire was 21.8. Individual genomic coverage varied from 16.2- to 26.7-fold) with an average of 22.2-fold coverage. The average length of the sequencing reads after filtering was 164.9 bp. The uncovered genetic variants filtered for Phred quality score ($\geq$20), pooled reads depth ($\geq$10), and B-Allele Frequency score ($\geq$0.5 for indels only; Yeo et al., 2012) included 11,201,995 SNP and 1,007,486 indels among samples. The majority of the discovered SNP were intergenic (60.0%). Intronic SNP were the most prevalent (96.3%) from all polymorphisms located in genes followed by 5′ and 3′ untranslated region (1.9%) and the coding region (1.4%). The effect of the depth of pooled sequencing reads on polymorphisms discovery was evaluated; no major changes in the distribution of SNP localization were found. A number of these polymorphisms could be potential sources of genetic variation if they are located in the extended areas of the major 1-Mb windows associated with phenotypic differences for the targeted traits.

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

Recent studies have demonstrated that a considerable proportion of phenotypic variation observed in reproductive traits is a result of variations in gene expression due to polymorphisms in the regulatory regions (Chen et al., 2016). High-throughput RNA sequencing is an effective method to perform genomewide quantification of gene expression as well as to reveal new genes and splice site variants. For example, RNA sequencing has been used in identifying differentially expressed genes associated with litter size traits in the placenta in Berkshire pigs (Kwon et al., 2016) and the ovaries in Yorkshire pigs (Zhang et al., 2015). Fischer et al. (2015) analyzed the transcriptome of the testis and oviduct of Finnish Large White pigs and identified polymorphisms in differentially expressed genes associated with reproduction.

The hypothalamic ARC contains neurons that are involved in regulating pubertal onset and maintenance of estrous cycles through the control of gonadotropin secretion (Lehman et al., 2010; Redmond et al., 2011). High-throughput RNA sequencing reads were obtained from the ARC from gilts representing pre- and postpubertal time points and early and late pubertal gilts fed 3 different dietary treatments. 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 (Sun et al., 2013), we were able to map 94.4% of the trimmed reads to the genome. Of these mapped reads, 45.1% of the reads were mapped to actual genes.

As expected, a large number of genes were found differentially expressed between prepubertal and postpubertal gilts (early or late; $n > 5{,}800$; adjusted $P < 0.05$). Differential expression between early and late pubertal gilts was observed for 17 genes (adjusted $P < 0.05$), including a gene involved in energy homeostasis (*FFAR2* [SSC6, 40.2 Mb]). The ARC 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). Twelve of these genes are upregulated in gilts exhibiting puberty at later ages compared with gilts with early age at puberty. None of the differentially expressed genes were located in the QTL regions associated with age at puberty. We speculate that some differences in expression could be a result of *trans*-modulation. For example, there are 74 genes known in humans (37 genes have swine orthologs) to be upstream regulators of the differentially expressed genes identified. Two of these upstream genes are located in the vicinity of a QTL for age at puberty (SSC13, 217 Mb), and as a result, they could influence variation in age at puberty via downstream differentially expressed genes. In the future, gene expression and pathway analysis combined with genetic variants located in the QTL areas will be integrated to expand our understanding of the genetic role in puberty onset.

Genomewide SNP × Diet Interaction on Age at Puberty and Reproductive Longevity

Genetic background and caloric intake affect the neuroendocrine axis and, as a result, the age when females express first estrus (Barb et al., 2002). In a previous study, using the same UNL population, Miller et al. (2011) demonstrated that energy restriction reduced BW, backfat, and LM area and delayed reproductive development. Our research showed that early expression of
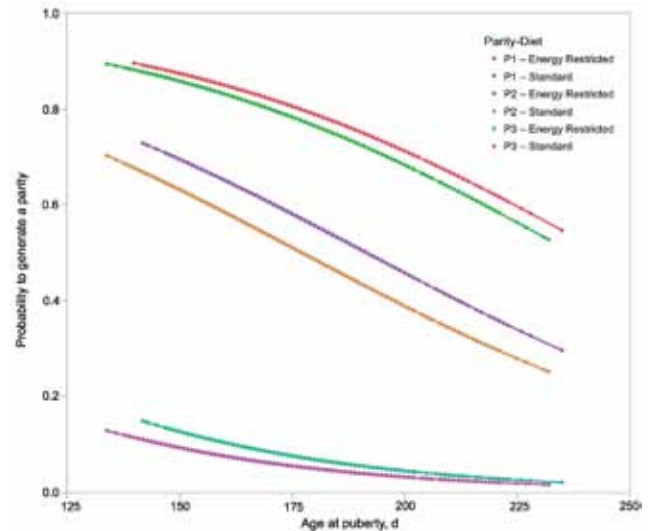

Figure 5. The effect of age at puberty and energy input during gilt development period on the probability to produce up to 3 parities. Energy restriction delayed age at puberty by approximately 7 d and increased the probability of the sows to produce parity 2 (P2) and parity 3 (P3; $P < 0.05$). Although not significant, parity 1 (P1) followed a similar trend ($P < 0.4$).

age at puberty is associated with reproductive longevity. We have also demonstrated that energy restriction delayed age at puberty by approximately 7 d but had positive sow fertility effects by increasing the probability to achieve parity 2 and 3 ($P < 0.05$). Although not significant ($P < 0.4$), parity 1 followed a similar trend (Fig. 5). When analyzed at the genomewide level, the effect of SNP × diet interactions on age at puberty marginally contributed to the proportion of phenotypic variation explained by SNP main effects (0.25% of additional phenotypic variation explained; Trenhaile et al., 2015). All SNP identified to interact with diet and affecting age at puberty ($n = 8$) and LT-NP ($n = 4$) have significant genotype × diet effects ($P < 0.05$; Table 2). The SNP had significant additive effects in at least one diet, with opposite effects among the dietary treatments (Fig. 5). As expected, no significant effects were observed when the analysis was performed on the integrated data set. As described above, one gene associated with a pleiotropic effect was *AVPR1A* (Tart et al., 2013). Initial findings of the second parity success demonstrated that the *GG* genotype was favorable across developmental diets and different from *AG* and *AA* genotypes ($P < 0.05$). However, when fed an energy-restricted diet, sows with the *AG* genotype are just as likely to produce parity 2 as sows with the favorable *GG* genotype fed the standard diet (Lucot et al., 2015). This is an example where the environment (in this case, a change in diet) can overwrite the genetic role in phenotypic variation. However, when the cumulative effects of all SNP are considered, reranking of GPV among dietary treatments investigated herein is not expected. A greater understanding of the molecular mechanisms involved in the interaction between nutri-

Table 2. Posterior means of variance components of age at puberty based on SNP × diet interaction effects estimated by BayesB analysis

GWAS[1]	Genetic variance	Residual variance	Total variance	Phenotypic variance explained by SNP, %
SNP	90.3	229.3	319.6	28.26
SNP × diet	97.3	244.6	342.1	28.51

[1]GWAS = genomewide association study.

tion and sow genome could be useful in designing efficient strategies to develop highly prolific sows.

Summary and Conclusions

The ultimate goal of this project was to enable accurate genomic prediction by identifying genes and functional polymorphisms associated with early onset of puberty and reproductive longevity. To achieve our purpose we combined the results obtained from a genome-wide association, genomic and transcriptomic sequencing, and gene expression profiling of the experimental animals. A customized marker panel will be constructed incorporating the functional variants and variants with the greatest effect on the targeted traits, and it will be applied in several commercial populations for evaluation. Functional polymorphisms and SNP-enriched regions with large effects on fertility traits will increase the ability of genomic information to be transferred between populations. Applying these tools when selecting replacement gilts will benefit the swine industry by decreasing the production costs due to improved reproductive efficiency, reduction in sow culling and gilt replacement rates, and improving animal well-being.

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. 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. 2011. Biology and genetics of reproduction. In: M. F. Rothschild and A. Ruvinsky, editors, The genetics of the pig. CAB Int., Oxfordshire, UK. p. 218–241. doi:10.1079/9781845937560.0218

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, A. 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

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. In: Joint Stat. Meet., August 11, 2015, Seattle, WA. https://ww2.amstat.org/meetings/jsm/2015/onlineprogram/AbstractDetails.cfm?abstractid=314884 (Accessed 25 February 2016.)

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

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

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

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

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

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

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

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

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 rev. ed. Natl. Acad. Press, Washington, DC.

Onteru, S. K., B. Fan, Z. Q., Du, D. J., Garrick, K. J., Stalder, 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

Redmond, J. S., G. M. Baez-Sandoval, K. M. Spell, T. E. Spencer, C. A. Lents, G. L. Williams, and M. Amstalden. 2011. Developmental changes in hypothalamic Kiss1 expression during activation of the pulsatile release of luteinising hormone in maturing ewe lambs. J. Neuroendocrinol. 23:815–822. doi:10.1111/j.1365-2826.2011.02177.x

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

Serenius, T., and K. J. Stalder. 2006. Selection for sow longevity. J. Anim. Sci. 84(E. Suppl.):E166–E171.

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.

Sun, Y. A., M. Taylor, D. Dhingra, K. Giorda, G. Meredith, L. Qu, C. S. Jose, K. Bramlett, and S. Ghosh. 2013. Two-step alignments for Ion ProtonTM sequencer RNA-Seq analysis. In: ION World Conf., October 21–22, 2013, Boston, MA. http://find.lifetechnologies.com/sequencing/ionworld/ltfurl/posters-2626HD-5088T5.html (Accessed 2 April 2016.)

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

Tosky, E. R., T. P. L. Smith, T. B. Engle, E. H. Davis, T. W. Moural, D. J. Nonneman, J. Chin, T. E. Burkey, G. S. Plastow, S. D. Kachman, and D. C. Ciobanu. 2016. Investigation and validation of the role of a QTL on SSC12 in PCV2b viremia. In: Plant Anim. Genome Conf., January 9–13, 2016, San Diego, CA. p. 169. https://pag.confex.com/pag/xxiv/webprogram/Paper21181.html (Accessed 18 November 2016.)

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., S. Kachman, P. Miller, R. Johnson, and D. Ciobanu. 2015. Genomic analysis of the interaction between energy intake and SNP genotypes on age at puberty. In: Plant Anim. Genome Conf., January 10–14, 2015, San Diego, CA. p. 233. https://pag.confex.com/pag/xxiii/webprogram/Paper16525.html (Accessed 20 August 2016.).

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. Picard tools version 1.90. http://broadinstitute.github.io/picard. (Accessed 5 April 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