Genome-wide association study on milk production and somatic cell score for Thai dairy cattle using weighted single-step approach with random regression test-day model

S. Buaban,1 K. Lengnudum,2 W. Boonkum,3 and P. Phakdeedindan4,5*
1Bureau of Animal Husbandry and Genetic Improvement, Department of Livestock Development, Pathum Thani 12000, Thailand
2Bureau of Biotechnology in Livestock Production, Department of Livestock Development, Pathum Thani 12000, Thailand
3Department of Animal Science, Faculty of Agriculture, Khon Kaen University, Khon Kaen 40002, Thailand
4Department of Animal Husbandry, Faculty of Veterinary Science, Chulalongkorn University, Bangkok 10330, Thailand
5Genomics and Precision Dentistry Research Unit, Department of Physiology, Faculty of Dentistry, Chulalongkorn University, Bangkok 10330, Thailand

ABSTRACTS

Genome-wide association studies are a powerful tool to identify genomic regions and variants associated with phenotypes. However, only limited mutual confirmation from different studies is available. The objectives of this study were to identify genomic regions as well as genes and pathways associated with the first-lactation milk, fat, protein, and total solid yields; fat, protein, and total solid percentage; and somatic cell score (SCS) in a Thai dairy cattle population. Effects of SNPs were estimated by a weighted single-step GWAS, which back-solved the genomic breeding values predicted using single-step genomic BLUP (ssGBLUP) fitting a single-trait random regression test-day model. Genomic regions that explained at least 0.5% of the total genetic variance were selected for further analyses of candidate genes. Despite the small number of genotyped animals, genomic predictions led to an improvement in the accuracy over the traditional BLUP. Genomic predictions using weighted ssGBLUP were slightly better than the ssGBLUP. The genomic regions associated with milk production traits contained 210 candidate genes on 19 chromosomes [Bos taurus autosome (BTA) 1 to 7, 9, 11 to 16, 20 to 21, 26 to 27 and 29], whereas 21 candidate genes on 3 chromosomes (BTA 11, 16, and 21) were associated with SCS. Many genomic regions explained a small fraction of the genetic variance, indicating polygenic inheritance of the studied traits. Several candidate genes coincided with previous reports for milk production traits in Holstein cattle, especially a large region of genes on BTA14. We identified 141 and 5 novel genes related to milk production and SCS, respectively. These novel genes were also found to be functionally related to heat tolerance (e.g., SLC45A2, IRAG1, and LOC101902172), longevity (e.g., SYT10 and LOC101903327), and fertility (e.g., PAG1). These findings may be attributed to indirect selection in our population. Identified biological networks including intracellular cell transportation and protein catabolism implicate milk production, whereas the immunological pathways such as lymphocyte activation are closely related to SCS. Further studies are required to validate our findings before exploiting them in genomic selection.

Key words: candidate gene, longitudinal trait, pathway analysis, Thai dairy cattle, weighted single-step genome-wide association study

INTRODUCTION

Milk production and udder health are economically important traits affecting dairy farming profitability. Improvement of milk production traits will directly bring greater benefits to dairy operations, and improved mastitis resistance will reduce cost of mastitis treatments. These polygenic traits are affected by a variety of factors (e.g., management practices, environmental conditions, animal physiological stage), and controlled by many genes and variants with small effects on the observed phenotype (Snelling et al., 2013). Improvement of management and nutrition along with intense genetic selection can lead to increased milk production and reduce the prevalence of mastitis (Rupp and Boichard, 2003).

In the last decade, the use of genomic information to infer the genomic estimated breeding values (GEBV) of an individual, known as genomic selection (Meuwissen et al., 2001), has become a routine practice in several livestock species facilitated by rapid advances in high-throughput SNP genotyping technologies. Genome-wide association studies are becoming practical tools
to identify SNP associated with QTL or genes with major effects on phenotypes (Zhang et al., 2016). The GWAS are advantageous compared with the traditional QTL mapping strategy due to a greater power to detect causal variants with modest effects and ability to define narrower genomic regions harboring causal variants (Hirschhorn and Daly, 2005). Therefore, GWAS have been widely regarded as a primary tool for identifying QTL and genomic regions associated with phenotypes.

The weighted single-step GWAS (WssGWAS; Wang et al., 2012) are a method that allows the estimation of SNP effects using GEBV predicted by single-step genomic BLUP (ssGBLUP, Aguilar et al., 2010) based on all phenotypes, genotypes, and pedigree information. This approach allows complex models such as random regression and multiple traits to be efficiently implemented for genomic prediction of longitudinal traits (Kang et al., 2017). In addition, WssGWAS allow unequal variances for SNPs, which result in improved precision of the estimation of SNP effects (Wang et al., 2012). The WssGWAS may perform better than classical GWAS methods when the number of animals with both phenotypes and genotypes is small, and the traits are controlled by QTL with large effects (Marques et al., 2018). The WssGWAS have been implemented in a variety of traits in livestock including broiler chickens (Fragomeni et al., 2014; Wang et al., 2014), dairy cattle (Tiezzi et al., 2015; Zhou et al., 2019), pigs (Howard et al., 2015; Marques et al., 2018; Wu et al., 2018), and beef cattle (Irano et al., 2016; Lemos et al., 2016; Valente et al., 2016; Silva et al., 2017; de Melo et al., 2017; Silva et al., 2019).

In Thailand, crossbred cattle constitute the majority of the total dairy cattle population. In 2018, the crossbred population reached more than 283,000 and was mainly raised by smallholders (Buaban et al., 2020). The majority of crossbred cows had at least 75% genetic fraction of Holstein Friesian. Other breed fractions were mainly from Bos indicus (Sahiwal, Red Sindhi, Brahman, and Thai native), which are robust to the tropical environment (hot and humid) in Thailand (Buaban, 2005). A single-trait random regression test-day model (RR-TDM) is currently used for the official genetic evaluation for milk production and milk component traits in dairy cattle because it can improve the accuracy of genetic evaluation and provide better modeling (Buaban et al., 2020; Department of Livestock Development, 2020). In addition, RR-TDM maximizes the amount of information gathered from each animal. Considering the advantages of RR-TDM and ssGBLUP in genetic prediction, it is expected that their combination should result in more accurate predictions, less bias, (Koivula et al., 2015; Baba et al., 2017; Kang et al., 2018; Oliveira et al., 2019b,d) and a better performance of GWAS (Oliveira et al., 2019a,c; Freitas et al., 2020).

Information about regions of the genome involved in milk production traits and SCS is currently lacking in the Thai dairy cattle population. Mastitis and high SCC can lead to changes in milk composition as a result of a decrease in synthesis due to epithelial cell damage (Malek dos Reis et al., 2013). This finding is directly related to mastitis, which is a major problem in the Thai dairy population (Jarassaeng et al., 2012; Hinthong et al., 2017; Horpiencharoen et al., 2019; Pumipuntu et al., 2019). The objectives of this study were to (1) identify genomic regions associated with milk production traits (milk yield and milk components) and SCS in Thai dairy cattle using WssGWAS, and (2) identify genes and pathways influencing the studied traits.

**MATERIALS AND METHODS**

**Phenotypic, Pedigree, and Genomic Data**

Monthly test-day records of the first-lactation cows calving between November 1993 and March 2017 were extracted from the centralized database of the Bureau of Biotechnology in Livestock Production, Department of Livestock Development of Thailand (Pathum Thani, Thailand). After the extraction, data were edited to retain only cows with known sire, age at first calving between 18 and 48 mo, DIM between 5 and 305 d, the first test date between 5 and 60 d, and daily milk yield between 1 and 45 kg. In addition, a minimum of 5 records were defined for each herd test-date contemporary group (Table 1). Each test-day record comprised at most 8 traits: yields for milk (MY, kg), fat (FY, kg), protein (PY, kg), and total solids (TSY, kg); percentages for fat (FP), protein (PP), and total solids (TSP); and SCC (1,000 cells/mL). Included in TSP are all constituents such as protein, fat, lactose, mineral materials, trace elements, and vitamins of the milk, except water. As the SCC has a highly skewed distribution, log-transformation was applied as log2(SCC/100) + 3 (Ali and Shook, 1980). Because some traits were not available on a test day, we classified the traits into 3 groups by missing pattern: MY, milk component traits (FY, PY, TSY, FP, PP, TSP), and SCS. We split the full data into 3 files by trait groups. In each data set, the pedigree was traced back to 3 generations from the recorded cows.

A total of 892 crossbred Thai dairy animals (154 bulls and 738 cows) were genotyped using the Illumina BovineSNP50 BeadChip (Illumina Inc.) version 2 (containing 54,609 SNPs) and version 3 (containing 53,218 SNPs), and only 50,908 common SNPs were considered. Quality control (QC) retained SNPs with call
Prediction of GEBV

First, (co)variance components for each trait were estimated based on the integration of the RR-TDM into single-step GBLUP procedure (SS-RR-TDM) using the same RR-TDM currently used in the Thai official genetic evaluations (Department of Livestock Development, 2020). The AIREMLF90 and BLUPF90 programs (Misztal et al., 2018) were used to estimate the (co)variance components and solve the mixed model equations. In matrix notation, the model equation for the SS-RR-TDM was

\[
y = X_1b_1 + X_2b_2 + Vh + Za + Wp + e, \quad [1]
\]

where \( y \) is the vector of test-day milk records, \( b_1 \) is the vector of systematic effects consisting of AI unit-year-season of calving effect (for MY and component traits), breed-age at calving, or year-season of calving effect (for SCS), \( b_2 \) is the vector of fixed regression coefficients on Legendre polynomials (LP) nested within breed-age at calving, \( h \) is the vector of random herd-year month of testing effect, \( a \) and \( p \) are the vectors of random regression coefficients on LP nested within the additive genetic effect and permanent environmental effect, respectively, and \( e \) is the vector of the residuals.

The matrices \( X_1, X_2, V, Z, \) and \( W \) are the correspondent incidence matrices. The order of \( L \) was calculated based on DIM as defined by Gengler et al. (1999). The LP of third (constant, linear, quadratic, and cubic) and second orders (constant, linear, and quadratic) were used for MY and content traits, and for SCS, respectively. The same order of LP was used for both fixed and random regression effects. The residual variable was assumed to be homogeneous throughout lactation to decrease the model complexity. Random effects were assumed to be normally distributed with zero means, and the covariance structure for models was defined as

\[
\begin{bmatrix}
\mathbf{h} \\
\mathbf{a} \\
\mathbf{p} \\
\mathbf{e}
\end{bmatrix}
\begin{bmatrix}
\mathbf{I} \sigma^2_{htm} & 0 & 0 & 0 \\
0 & \mathbf{H} \otimes \mathbf{G}_0 & 0 & 0 \\
0 & 0 & \mathbf{I} \otimes \mathbf{P}_0 & 0 \\
0 & 0 & 0 & \mathbf{I} \sigma^2_e
\end{bmatrix}, \quad [2]
\]

where \( \mathbf{G}_0 \) and \( \mathbf{P}_0 \) are 4×4 (for milk yield and component traits) or 3×3 (for SCS) (co)variance matrices of the random regression coefficients for the animal and permanent environment effects respectively, is a modified genetic relationship matrix that combines pedigree-based (\( \mathbf{A} \)) and genomic-based (\( \mathbf{G} \)) relationship matrices as shown by Legarra et al. (2009), \( \mathbf{I} \) is an identity matrix, \( \otimes \) is the Kronecker product operator, \( \sigma^2_{htm} \) is the variance of the random herd-year month of testing effects, and \( \sigma^2_e \) is the residual variance. The inverse of \( \mathbf{H} \) needed for mixed model equations is given by

\[
\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix}
0 & 0 \\
0 & \tau (\alpha \mathbf{G} + \beta \mathbf{A}_{22})^{-1} - \omega \mathbf{A}_{22}^{-1}
\end{bmatrix}, \quad [3]
\]

where \( \mathbf{A} \) is the numerator relationship matrix based on pedigree for all animals; \( \mathbf{A}_{22} \) is the numerator relationship matrix for genotyped animals; and \( \alpha, \beta, \omega, \) and \( \tau \) are weighting factors. To avoid singularity problems, \( \alpha \) and \( \beta \) were selected to be 0.95 and 0.05, respectively (VanRaden, 2008). The weights \( \tau \) and \( \omega \) \((\tau = 1.00, \omega = 1.00)\) were selected to be 0.95 and 0.05, respectively. The matrices \( X_1, X_2, V, Z, \) and \( W \) are the correspondent incidence matrices. The order of \( L \) was calculated based on DIM as defined by Gengler et al. (1999). The LP of third (constant, linear, quadratic, and cubic) and second orders (constant, linear, and quadratic) were used for MY and content traits, and for SCS, respectively. The same order of LP was used for both fixed and random regression effects. The residual variable was assumed to be normally distributed with zero means, and the covariance structure for models was defined as

\[
\begin{bmatrix}
\mathbf{h} \\
\mathbf{a} \\
\mathbf{p} \\
\mathbf{e}
\end{bmatrix}
\begin{bmatrix}
\mathbf{I} \sigma^2_{htm} & 0 & 0 & 0 \\
0 & \mathbf{H} \otimes \mathbf{G}_0 & 0 & 0 \\
0 & 0 & \mathbf{I} \otimes \mathbf{P}_0 & 0 \\
0 & 0 & 0 & \mathbf{I} \sigma^2_e
\end{bmatrix}, \quad [2]
\]
Weights were chosen such that they provided the least bias in the preliminary validation study. The value of $\omega$ was lower than the value (0.7) usually used in beef and dairy cattle ssGBLUP evaluations (Lourenco et al., 2020) and this could be in part due to some degrees of pedigree incompleteness in the studied population. As a result of the potential incomplete pedigree, inbreeding was ignored in $A$ matrix although it could have some effects on the results. The $G$ matrix was created according to the second method of VanRaden (2008) as follows:

$$
G = \frac{1}{2} \sum_{i=1}^{m} p_i (1 - p_i)
$$

where $Z$ is a matrix of gene content adjusted for allele frequencies; $D$ is a diagonal matrix of weights for SNP variances (initially $D = I$); $m$ is the number of SNP markers; and $p_i$ is the frequency of the reference allele of the $i$th SNP.

The individual genomic breeding value was presented as a 305-d production basis. The 305-d genomic breeding value was defined as the sum of the breeding values between DIM 5 to 305 for yield traits as described by Jamrozik et al. (1997). For milk component traits and SCS, the sum was replaced with the average of breeding values as a 305-d production basis. The 305-d genomic breeding value ($GEBV$) for a 305-d basis, and all other terms are as defined earlier. Using the breeding value for 305-d basis, $GEBV$ for a 305-d basis, and all other terms are as described for variance component estimation were applied in the WssGBLUP. The additive genomic breeding value ($a_g$; the GEBV for 305-d basis) was back-solved to SNP effects accounting for their shared genomic variance ($\sigma^2_g$), using the following formula:

$$
u = DZG^{-1}a_g,
$$

where $u$ is the vector of SNP effects, $a_g$ is the vector of GEBV for a 305-d basis, and all other terms are as defined earlier. Using the breeding value for 305-d basis, thus all predictions of random regression coefficients associated with LP for each individual were already condensed in the SNP effects.

In the first round of the iterative process, the variance absorbed by each marker was obtained as $2p_i (1 - p_i)$ and $D = I$. In successive iterations, to highlight regions with high impact on the genetic variation of the trait, a weighted $G$ matrix was created, where expected marker contributions were replaced with realized variances. Thus, elements of $D$ were

$$
D_{ii} = \frac{2p_i (1 - p_i)}{m} u^2_i,
$$

where $u_i$ is the effect of the $i$th SNP estimated in the previous iteration. New marker effects were obtained by considering the weighted $G$ matrix in Equation [5]. The detailed description of the iterative reweighting procedure can be seen in Scenario 1 in Wang et al. (2012).

In the first implementation of WssGBLUP, Wang et al. (2012) suggested using a linear weight following the formula for genetic variance due to an additive locus (Falconer and Mackay, 1996) as $d^{*}_{j(t+1)} = \frac{u^2_j}{2p_j (1 - p_j)}$. However, Lourenco et al. (2017) showed that this method did not reach convergence under a more polygenic scenario because of extreme weights. Therefore, nonlinear $A$ weights were used in this study as described by VanRaden (2008):

$$
d^{*}_{j(t+1)} = \frac{u_j}{CT \cdot sd(\hat{u})^2},
$$

where $CT$ is a constant that determines the departure from normality; $u_j$ is the absolute estimated SNP effect for marker $j$, and $sd(\hat{u})$ is the standard deviation of the estimated SNP effects. Nonlinear $A$ weights had good convergence properties and avoided extreme values (Garcia et al., 2018). This is because the maximum change in weights is limited by the minimum between 5 and the exponent of $CT$. In our study, $CT$ was 1.25 following Legarra et al. (2018) and VanRaden (2008).

Detection of informative regions or loci based on single SNPs may result in noise or underestimation due to the high ratio between the number of SNPs and the number of genotyped animals (Wang et al., 2014). Moreover, adjacent SNPs may be in high LD with the same QTL in high-density SNP panels leading to the QTL effect spreading over all SNPs in high LD (Fan et al., 2011). For this reason, the proportion of genetic variance was calculated for each nonoverlapping 1-Mb windows to account for LD, which is more appropriate than using single SNP.

Performance of genetic and genomic predictions for young bulls was evaluated by comparing individual theoretical accuracy of 305-d GEBV (or EBV). Thus, pedigree-based RR-TDM was also used to obtain EBV in addition to SS-RR-TDM. Genomic predictions with WssGBLUP were also used to compare genomic evaluation methods. The young bulls were progeny-tested.
bulls in the last 6 yr from the data set with more than 75% reliability of EBV. The theoretical accuracy of 305-d GEBV (or EBV) was calculated based on the prediction error variance (Misztal and Wiggans, 1988). The prediction error variance was presented using the 305-d GEBV (or EBV) basis (e.g., Mrode, 2014). The number of iterations for weighting procedure in WssGWAS was determined as the iteration number that maximized the accuracy of predictions. The proportion of genetic variance was then calculated using a nonoverlapping 20-SNP moving windows (approximately 1 Mb).

The percentage of genetic variance explained by the \( i \)th region was calculated as follows:

\[
\frac{\text{var} \left( a_i \right)}{\sigma^2_a} \times 100\% = \frac{\text{var} \left( \sum_{j=1}^{20} z_j \hat{u}_j \right)}{\sigma^2_u} \times 100\%, \quad [7]
\]

where \( a_i \) is the genetic value of the \( i \)th region that consists of 20 adjacent SNPs, \( \sigma^2_a \) is the total additive genetic variance, \( z_i \) is the vector of gene content of the \( j \)th SNP for all individuals, and \( \hat{u}_j \) is the marker effect of the \( j \)th SNP within the \( i \)th region (Zhang et al., 2010).

To identify important genomic regions related to the analyzed traits, moving windows of 20 adjacent SNPs explaining at least 0.5% of the total genetic variance were selected for further analyses. This threshold has been previously investigated in cattle for complex traits (Fragomeni et al., 2014; Oliveira et al., 2017; Silva et al., 2017; Lee et al., 2019; Zhou et al., 2019) and in the expected contribution of SNP windows (Sollero et al., 2017). Manhattan plots of these windows were generated using gnuplot 5.2 (Williams and Kelley, 2019).

Candidate Genes and Functional Analyses

To identify possible candidate genes associated with milk production traits and SCS, genes that were located within the most important nonoverlapping genomic windows (i.e., between the start and end of genomic coordinates of the selected windows) were further investigated. We identified genes using the National Center for Biotechnology Information (NCBI) Map Viewer tool for the bovine genome with the UMD 3.1 assembly as the reference map (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/055/GCF_000003055.6_Bos_taurus_UMD_3.1.1/).

For all identified genes, literature and database searches (NCBI: https://www.ncbi.nlm.nih.gov/; Genecards: https://www.genecards.org/; and UniProt: https://www.uniprot.org/) were performed to investigate the metabolic function of the genes identified. A list of genes located in proximity to the windows was used for performing a gene network analysis using the online resource GeneMania (Warde-Farley et al., 2010). Moreover, gene ontology analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (Huang et al., 2009a,b).

RESULTS AND DISCUSSION

Variance Component and Heritability Estimations

Descriptive statistics for test-day milk production traits and SCS in the data set used for genetic evaluation and 305-d variance components and heritability estimates \( h^2_{305d} \) for additive, herd test month, permanent environmental and residual variances are presented in Table 2. Estimated heritabilities were 0.49, 0.43, 0.54, 0.52, 0.31, 0.50, 0.46, and 0.32 for 305-d MY, 305-d FY, 305-d PY, 305-d TSY, average FP, average PP, average TSP, and average SCS, respectively.

Comparisons Between Genetic and Genomic Prediction Methods

Table 3 presents the average of individual theoretical accuracies of EBV and GEBV obtained from ssGBLUP and WssGBLUP (at the second iteration) for 305-d milk yield, milk component, and SCS traits. Across all

<table>
<thead>
<tr>
<th>Trait</th>
<th>Mean</th>
<th>SD</th>
<th>( \sigma^2_a )</th>
<th>( \sigma^2_{\text{lin}} )</th>
<th>( \sigma^2_{\text{p}} )</th>
<th>( \sigma^2_{\text{e}} )</th>
<th>( h^2_{305d} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>305-d MY, kg</td>
<td>13.89</td>
<td>4.53</td>
<td>353,825.40</td>
<td>169.34</td>
<td>368,823.60</td>
<td>565.34</td>
<td>0.49</td>
</tr>
<tr>
<td>305-d FY, kg</td>
<td>0.49</td>
<td>0.20</td>
<td>517.05</td>
<td>0.95</td>
<td>671.55</td>
<td>2.97</td>
<td>0.43</td>
</tr>
<tr>
<td>305-d PY, kg</td>
<td>0.44</td>
<td>0.14</td>
<td>393.36</td>
<td>0.30</td>
<td>332.59</td>
<td>0.93</td>
<td>0.54</td>
</tr>
<tr>
<td>305-d TSY, kg</td>
<td>1.69</td>
<td>0.55</td>
<td>5,675.20</td>
<td>3.39</td>
<td>5,112.85</td>
<td>11.25</td>
<td>0.52</td>
</tr>
<tr>
<td>Average FP, %</td>
<td>3.56</td>
<td>0.96</td>
<td>6,500.22</td>
<td>40.11</td>
<td>1,4427.73</td>
<td>40.11</td>
<td>0.54</td>
</tr>
<tr>
<td>Average PP, %</td>
<td>3.13</td>
<td>0.39</td>
<td>1,238.12</td>
<td>6.12</td>
<td>1,190.99</td>
<td>18.35</td>
<td>0.50</td>
</tr>
<tr>
<td>Average TSP, %</td>
<td>3.56</td>
<td>0.96</td>
<td>6,500.22</td>
<td>40.11</td>
<td>1,4427.73</td>
<td>40.11</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Table 2. Mean and SD of test-day milk production traits and SCS in the data set used for genetic evaluation; 305-d variance components \( \sigma^2_a, \sigma^2_{\text{lin}}, \sigma^2_{\text{p}} \), and \( \sigma^2_{\text{e}} \) and heritability estimates \( h^2_{305d} \) are also presented.
traits, WssGBLUP provided the greatest accuracies, followed by ssGBLUP and BLUP. For WssGBLUP, the accuracies were maximized by the second iteration for all traits, then remained constant from the third to fifth iterations (results not shown). Thus, 2 iterations of weighting were used in WssGWAS. The gain in accuracy over BLUP for young bulls was 0.015 and 0.017 for ssGBLUP and WssGBLUP, respectively. The increase in accuracy in genomic predictions over EBV is consistent with previous studies in dairy cattle (VanRaden et al., 2009; Ding et al., 2013). Notably, small improvements of accuracies obtained using WssGBLUP compared with ssGBLUP (accuracy gain of 0.002) could be the fact that the studied traits are (highly) polygenic where small or no gain in accuracy are expected with SNP weighting (Hayes et al., 2009; Lourenco et al., 2017). In this study, the optimal weight with 2 iterations is sufficient to maximize genomic accuracy which is in accordance with those reported in previous studies (Wang et al., 2012; Zhang et al., 2016). In addition, this iteration number was consistent with other literatures, providing the greatest predictive ability and least inflation (Legarra et al., 2008; Garcia et al., 2018), as well as high stability of the marker effect estimates (Wang et al., 2014). In spite of the limited number of genotype animals, fitting a ssGBLUP increases the accuracy of breeding values in Thai dairy cattle, eventually improving overall genetic gain.

The resolution of Manhattan plots was slightly improved when SNP weights were added. The second iteration of weights SNP weighting resulted in the least noisy percentage of additive variance explained by windows of 20 adjacent SNPs (Figure 1). Lourenco et al. (2017) showed that WssGBLUP reduced noise which resulted in an increased precision of estimating QTL effects and positions in simulation study. Small improvement of the resolution observed in this study could be attributed to a polygenic nature of the studied traits. In addition, a sample size can critically affect statistical power in GWAS study (Hong and Park, 2012). The small number of genotyped animals used in this study may lead to a low power to detect SNP related to QTL which resulted in a large number of SNPs with small effects across all traits analyzed (all SNPs explained less than 1% of the genetic variance of a trait).

Although the small number of genotyped animals could limit the improved accuracy of genomic prediction (Kosgey and Okeyo, 2007; Mrode et al., 2019), in the present study a positive effect on the accuracy of young bulls was observed. This could be related to the ongoing breeding program (e.g., the existence of a progeny testing scheme and extensive mating within a population), which enhanced relationships between reference population and selection candidates (Habier et al., 2010; Clark et al., 2012; Mrode et al., 2019). Some degree of improvement in accuracy of genomic prediction with the small number of genotyped animals were previously reported in developing countries (Neves et al., 2014; Silva et al., 2016; Boison et al., 2017). Nevertheless, more data on genotypes and records likely facilitate the improvement of genomic prediction accuracy and the reliability (and resolution) of GWAS results for the Thai dairy cattle population. Therefore, it is necessary to enlarge the reference database with phenotypes and genotypes in both cows and bulls.

In this study, the models were assumed to have constant residual variance along the lactation to decrease model complexity. Different assumptions of residual variance (i.e., homogeneity and heterogeneity) have been previously used in ssGWAS procedures with random regression models (Howard et al., 2015; Oliveira et al., 2019b). Howard et al. (2015) modeled homogeneous residual variance due to its simplicity and similar results being observed with heterogeneity assumption in

### Table 3. Average theoretical accuracy of EBV in the pedigree-based model and genetic EBV with ssGBLUP and WssGBLUP1 for genotyped young bulls in 305-d production yield, milk component traits, and SCS

<table>
<thead>
<tr>
<th>Trait</th>
<th>Young bulls</th>
<th>BLUP</th>
<th>ssGBLUP</th>
<th>WssGBLUP_1</th>
<th>WssGBLUP_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>305-d MY</td>
<td>28</td>
<td>0.9149</td>
<td>0.9260</td>
<td>0.9262</td>
<td>0.9263</td>
</tr>
<tr>
<td>305-d FY</td>
<td>28</td>
<td>0.8972</td>
<td>0.9124</td>
<td>0.9126</td>
<td>0.9126</td>
</tr>
<tr>
<td>305-d PY</td>
<td>28</td>
<td>0.9208</td>
<td>0.9305</td>
<td>0.9307</td>
<td>0.9307</td>
</tr>
<tr>
<td>305-d TSY</td>
<td>28</td>
<td>0.9196</td>
<td>0.9296</td>
<td>0.9298</td>
<td>0.9298</td>
</tr>
<tr>
<td>Average FP</td>
<td>12</td>
<td>0.8773</td>
<td>0.8970</td>
<td>0.8974</td>
<td>0.8974</td>
</tr>
<tr>
<td>Average PP</td>
<td>25</td>
<td>0.8943</td>
<td>0.9104</td>
<td>0.9108</td>
<td>0.9108</td>
</tr>
<tr>
<td>Average TSP</td>
<td>27</td>
<td>0.8942</td>
<td>0.9107</td>
<td>0.9110</td>
<td>0.9110</td>
</tr>
<tr>
<td>Average SCS</td>
<td>13</td>
<td>0.8787</td>
<td>0.8971</td>
<td>0.8975</td>
<td>0.8975</td>
</tr>
</tbody>
</table>

1BLUP = pedigree-based model; ssGBLUP = single-step GBLUP; WssGBLUP_1 = first iteration weighted single-step GBLUP; WssGBLUP_2 = second iteration weighted single-step GBLUP.

2305-d MY = 305-d milk yield; 305-d FY = 305-d fat yield; 305-d PY = 305-d protein yield; 305-d TSY = 305-d total solids yield; average FP = average fat percentage; average PP = average protein percentage; average TSP = average total solids percentage.
Figure 1. Manhattan plots of the additive genetic variance explained by windows of 20 adjacent SNPs for 305-d milk yield: (A) no weight (ssGWAS), (B) the first iteration of SNP weighting (WssGWAS1), and (C) the second iteration of SNP weighting (WssGWAS2).
Figure 2. Manhattan plots of the additive genetic variance explained by windows of 20 adjacent SNPs for milk production traits and SCS in Thai dairy cattle: (A) milk yield, (B) fat yield, (C) protein yield, (D) TS yield, (E) fat percentage, (F) protein percentage, (G) TS percentage, and (H) SCS.
their preliminary study. It might be that if the accuracy of models was affected by the assumption of residual variance, the accuracy of estimated SNP effects could also be affected, impacting GWAS results. Further investigations might be necessary to compare model accuracies and GWAS results using different assumptions of residual variance.

Identification of Genomic Region and Candidate Genes

In the present study, a WssGBLUP method was used to investigate genomic regions and identified candidate genes responsible for milk yield, milk components, and SCS using medium density SNP panels. The Manhattan plots of percentage additive genetic variance explained by 20-SNP moving windows are reported in Figure 2. The largest genetic variance explained by a window was 0.7% for milk production traits and 0.6% for SCS, respectively. However, most windows explained less than 0.5%, and these low-contributing regions spread across the genome for all traits analyzed.

This indicates that both milk production traits and SCS are moderately to highly polygenic, in which many regions across the genome contribute to the genetic variation of the traits (Figure 2). A summary of the 36 informative windows that explained the largest proportion of additive genetic variance (at least 0.5%) and the important candidate genes for all traits studied are presented in Table 4. The regions associated with milk production traits were located on chromosomes BTA 1, 2, 3, 4, 5, 6, 7, 9, 11, 12, 13, 14, 15, 16, 20, 21, 26, 27, and 29, whereas regions associated with SCS were found on chromosomes BTA 11, 16, and 21. We compared the results with databases (NCBI, Genecards, and Uniprot) and found 210 and 21 reported QTL related to milk production traits and SCS, respectively. In addition, we found several genes which were not previously reported for the studied traits, but they were related to the analyzed traits. For example, TMEM181, a fat-related QTL in marbling score (Mokry et al., 2013), was identified as novel candidate genes for fat yield in the present study. Relatively concentrated areas related to milk production traits and SCS were noted on chromosomes BTA 1, 5, 7, 12, 14, 16, 20, and 21.

Candidate Genes for MY

Genomic regions associated with 305-d MY were located on chromosomes 1 (from 103,434,094 to 104,344,681 bp), 29 (from 37,949,466 to 40,507,958 bp), and 3 (from 49,431,468 to 50,835,532 bp). Within these windows, there were 42 known genes and 50 uncharacterized genes (Figure 2A and Table 4).

Two genes were previously reported to directly regulate milk yield (Benyamin et al., 2014; Zabolewicz et al., 2014; Do et al., 2017b; Yodklaew et al., 2017). The gene DR1 plays a role in lactoferrin expression in milk, which is related to calcium metabolism, skeletal homeostasis, and immune system (Zabolewicz et al., 2014) and influences mRNA network regulation of milk yield (Do et al., 2017b). LOC100847667 has been reported as a marker involving peak yield in Thai multibreed dairy cattle population (Yodklaew et al., 2017). This agreement could be due to the similar genetic background of the dairy cattle breeds in Thailand where Holstein and Bos indicus cattle are commonly crossed.

Other genes were demonstrated to have mechanisms indirectly related to milk yield (Ibeagha-Awemu et al., 2016; Yang et al., 2016). The gene VPS37C, a subunit of the protein complex required for sorting ubiquitinated transmembrane proteins into internal vesicles in cell transportation, is related to the milk fatty acid and butyric acid (Ibeagha-Awemu et al., 2016). The gene TMED5 is a protein in a secretory pathway that is required for the maintenance of the Golgi apparatus and vesicular protein trafficking. This protein can be found in a milk fat globule membrane of colostrum and milk (Yang et al., 2016). Moreover, several genes in the significant windows were reported to regulate mammary physiology (Farke et al., 2008; Oh et al., 2013; Benyamin et al., 2014; Chen et al., 2015; Han et al., 2018). A member of ABC transporters, ABCA4, is expressed in the bovine mammary gland and may be implicated in lipid and cholesterol transport into milk (Farke et al., 2008). The gene GCLM acts as an antioxidant in mammary tissues (Han et al., 2018), whereas DNTTIP2 is a candidate marker to identify mastitis in cows (Chen et al., 2015). Both BCAR3 and FNB1 are responsible for cell proliferation, migration, and cytoskeleton reorganization (Oh et al., 2013; Benyamin et al., 2014).

Furthermore, many proteins were found to be related to pregnancy, such as members of pregnancy-associated glycoprotein (PAG; PAG1, PAG2, PAG3, PAG4, PAG5, PAG6, PAG7, PAG8, PAG9, PAG10, PAG11, PAG12, PAG14, PAG15, PAG16, PAG17, PAG18, PAG19, PAG20, and PAG21), PAG1-like (LOC784241, LOC783673, LOC615840, LOC785405, LOC785376, LOC785335, LOC784923, and LOC101903717), and MGC157408 (Green et al., 2000; Szafranska et al., 2006). The gene GCLM is related indirectly to milk yield (Ibeagha-Awemu et al., 2016; Yang et al., 2016). The gene VPS37C is a protein in a secretory pathway that is related to the milk fatty acid and butyric acid (Ibeagha-Awemu et al., 2016). The gene TMED5 is a protein in a secretory pathway that is required for the maintenance of the Golgi apparatus and vesicular protein trafficking. This protein can be found in a milk fat globule membrane of colostrum and milk (Yang et al., 2016). Moreover, several genes in the significant windows were reported to regulate mammary physiology (Farke et al., 2008; Oh et al., 2013; Benyamin et al., 2014; Chen et al., 2015; Han et al., 2018). A member of ABC transporters, ABCA4, is expressed in the bovine mammary gland and may be implicated in lipid and cholesterol transport into milk (Farke et al., 2008). The gene GCLM acts as an antioxidant in mammary tissues (Han et al., 2018), whereas DNTTIP2 is a candidate marker to identify mastitis in cows (Chen et al., 2015). Both BCAR3 and FNB1 are responsible for cell proliferation, migration, and cytoskeleton reorganization (Oh et al., 2013; Benyamin et al., 2014).

Furthermore, many proteins were found to be related to pregnancy, such as members of pregnancy-associated glycoprotein (PAG; PAG1, PAG2, PAG3, PAG4, PAG5, PAG6, PAG7, PAG8, PAG9, PAG10, PAG11, PAG12, PAG14, PAG15, PAG16, PAG17, PAG18, PAG19, PAG20, and PAG21), PAG1-like (LOC784241, LOC783673, LOC615840, LOC785405, LOC785376, LOC785335, LOC784923, and LOC101903717), and MGC157408 (Green et al., 2000; Szafranska et al., 2006). The gene GCLM is related indirectly to milk yield (Ibeagha-Awemu et al., 2016; Yang et al., 2016). The gene VPS37C is a protein in a secretory pathway that is related to the milk fatty acid and butyric acid (Ibeagha-Awemu et al., 2016). The gene TMED5 is a protein in a secretory pathway that is required for the maintenance of the Golgi apparatus and vesicular protein trafficking. This protein can be found in a milk fat globule membrane of colostrum and milk (Yang et al., 2016). Moreover, several genes in the significant windows were reported to regulate mammary physiology (Farke et al., 2008; Oh et al., 2013; Benyamin et al., 2014; Chen et al., 2015; Han et al., 2018). A member of ABC transporters, ABCA4, is expressed in the bovine mammary gland and may be implicated in lipid and cholesterol transport into milk (Farke et al., 2008). The gene GCLM acts as an antioxidant in mammary tissues (Han et al., 2018), whereas DNTTIP2 is a candidate marker to identify mastitis in cows (Chen et al., 2015). Both BCAR3 and FNB1 are responsible for cell proliferation, migration, and cytoskeleton reorganization (Oh et al., 2013; Benyamin et al., 2014).
<table>
<thead>
<tr>
<th>Trait</th>
<th>Window regions (nonoverlapping)</th>
<th>Var. (%)</th>
<th>Chr</th>
<th>Start (bp)</th>
<th>Stop (bp)</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>MY</td>
<td>1,724-1,743</td>
<td>0.5</td>
<td>1</td>
<td>103,434,094</td>
<td>104,344,681</td>
<td>LOC104970979, LOC781603</td>
</tr>
<tr>
<td>FY</td>
<td>39,309-39,328</td>
<td>0.6</td>
<td>26</td>
<td>37,336,603</td>
<td>38,233,337</td>
<td>LOC101905185, FNBP1L, DR1, LOC101905706, CCDC18, LOC107132311, TMED5, MFP2, FAM69A, RPL3</td>
</tr>
<tr>
<td>TSY</td>
<td>1,724-1,743</td>
<td>0.5</td>
<td>1</td>
<td>103,434,094</td>
<td>104,344,681</td>
<td>LOC101905185, FNBP1L, DR1, LOC101905706, CCDC18, LOC107132311, TMED5, MFP2, FAM69A, RPL3</td>
</tr>
<tr>
<td>FP</td>
<td>13,004-13,023</td>
<td>0.7</td>
<td>14</td>
<td>5,774,815</td>
<td>6,413,319</td>
<td>LOC104970979, LOC781603</td>
</tr>
<tr>
<td></td>
<td>22,477-22,496</td>
<td>0.6</td>
<td>12</td>
<td>40,009,119</td>
<td>42,606,604</td>
<td>LOC101905185, FNBP1L, DR1, LOC101905706, CCDC18, LOC107132311, TMED5, MFP2, FAM69A, RPL3</td>
</tr>
<tr>
<td></td>
<td>10,105-10,124</td>
<td>0.6</td>
<td>5</td>
<td>76,339,294</td>
<td>77,095,345</td>
<td>LOC101905185, FNBP1L, DR1, LOC101905706, CCDC18, LOC107132311, TMED5, MFP2, FAM69A, RPL3</td>
</tr>
<tr>
<td></td>
<td>1,740-1,759</td>
<td>0.6</td>
<td>12</td>
<td>104,220,701</td>
<td>105,570,832</td>
<td>LOC104970979, LOC781603</td>
</tr>
<tr>
<td></td>
<td>4,823-4,842</td>
<td>0.5</td>
<td>2</td>
<td>130,141,723</td>
<td>131,135,560</td>
<td>LOC101905185, FNBP1L, DR1, LOC101905706, CCDC18, LOC107132311, TMED5, MFP2, FAM69A, RPL3</td>
</tr>
<tr>
<td></td>
<td>1,972-1,991</td>
<td>0.5</td>
<td>16</td>
<td>118,532,290</td>
<td>119,421,373</td>
<td>LOC101905185, FNBP1L, DR1, LOC101905706, CCDC18, LOC107132311, TMED5, MFP2, FAM69A, RPL3</td>
</tr>
<tr>
<td></td>
<td>10,669-10,688</td>
<td>0.5</td>
<td>5</td>
<td>113,468,743</td>
<td>114,504,935</td>
<td>LOC101905185, FNBP1L, DR1, LOC101905706, CCDC18, LOC107132311, TMED5, MFP2, FAM69A, RPL3</td>
</tr>
</tbody>
</table>

Continued...
<table>
<thead>
<tr>
<th>Trait</th>
<th>Window regions (nonoverlapping)</th>
<th>Var. (%)</th>
<th>Chr</th>
<th>Start (bp)</th>
<th>Stop (bp)</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP</td>
<td>39,708–39,727</td>
<td>0.6</td>
<td>27</td>
<td>10,697,257</td>
<td>11,822,624</td>
<td>LOC100818735, TRNAC-GCA, TRNAW-CCA, LOC101908395</td>
</tr>
<tr>
<td></td>
<td>13,147–13,166</td>
<td>0.6</td>
<td>7</td>
<td>17,871,862</td>
<td>19,200,954</td>
<td>PRR36, LRRCNH, MAP2K7, TGFB3I2, SNAPC2, CTXNL1, TMM14, MIR2455, ELAVL1, CCL25, FBN3, CERSA, CD200, NDUFA7, RPS8, KANK2, ANGPTL4, RAB11B, MARC2H, HNRNPM, PRAM1, ZNF144, MYO1F, MIR339B, ADAMTS10, LOC101904824, ACTL9, LOC101729019, LOC101903704, LOC101905537, LOC101712915, LOC101829356, LOC101829364, LOC101829371, TRIP10, MIR6535, LOC100818893, C3, LOC101732612, LOC101905865, MIR3432B, TNFSF14, CD70, TNFSF9, LOC101905992, TUBB4A, DENNDIC</td>
</tr>
<tr>
<td></td>
<td>32,999–33,018</td>
<td>0.5</td>
<td>20</td>
<td>31,933,394</td>
<td>33,433,160</td>
<td>GHR, LOC101793269, LOC101793266, BFQ01, OXCT1, TRNAY-GUA, PLCX3, C6</td>
</tr>
<tr>
<td></td>
<td>24,736–24,755</td>
<td>0.5</td>
<td>14</td>
<td>1,463,676</td>
<td>2,553,525</td>
<td>C14H8orf33, ZNF53, RPL8, ZNF7, COMMD5, ARHGAQ9, MIR3208, C14H8orf82, LRRCC4, LRRCA4, RECC4Q, MFSQ3, GPT, PPP1R16A, FOXH1, KIFC2, CYHRI, TNSL1, VPS28, SLC39A4, CPSF1, ADCK5, SLC52A2, FBXL6, TMEM249, SCRT1, DGAT1, HSF1, BOP1, SCX, MROH1, MIR1839, LOC520328, HCH1, LOC101793958, MAF1, SHARPIN, CYC1, GPA1, LOC101733095, EXOSC4, OPLAH, SMPD5, SPATC1, TRNAG-UCC, LOC101908059, GRINA, PARP10, LOC101793966, MIR3390, EPPK1, NRRBP2, PUF60, SCIR1, LOC506831, LOC101907863, FMAM3B, MAPK15, CDC66, ZNF8323, MAMC2, Pycr1, TIGD5, EEF1D, NAPRT1, MROH6, GSMDM, ZC2H9, MFAA, RHPNI, TOP1MT, ZNF686, GLI1, ZFPL1, GIPHIHP1, PPIMK1, ABCG2, LOC781421, PKD2, PPP1, MEPE, IBSP, LOC101972726, TRNA-CGC, LAP3, MED28, FAM18B, NCAPC, DCAF16, LOC1R1, USHPB1, NR2F6, OCRL1, LOC101969050, USE1, MYO9B, HUA88, LOC101969051, LOC101972787, CPAMDS, NIP50, FZL3, SIN3B, NWDL, LOC101969054, TMEM338A, SMMIM7, MED26, SLC32A1, CHERP, LOC535121, SPT5, WBP2N1, NAGA, FAM109B, SMDT1, NDUFA6, LOC785804, MGC127055, CYP2D4, TCF20, MIR2424, LOC101793976, LOC101973595, NEAM1, SERH2L, LOC101903498, TRNAG-CCC, RRPIA, POLDII3, CYB5R3, LOC101972579, AGXABT, LOC10175259, ARFGAP3, LOC100298383, PACSIN2, LOC101904708, TTT1, LOC101847895, BIK, MCAT</td>
</tr>
<tr>
<td>TSP</td>
<td>11,439–11,458</td>
<td>0.8</td>
<td>6</td>
<td>37,868,743</td>
<td>38,892,842</td>
<td>33,831–33,850, 21</td>
</tr>
<tr>
<td></td>
<td>13,004–13,023</td>
<td>0.6</td>
<td>7</td>
<td>5,774,810</td>
<td>6,413,319</td>
<td>4,987–4,996, 21</td>
</tr>
<tr>
<td></td>
<td>13,147–13,166</td>
<td>0.6</td>
<td>14</td>
<td>7,718,808</td>
<td>8,545,690</td>
<td>4,472–4,478, 21</td>
</tr>
<tr>
<td>SCS</td>
<td>22,477–22,496</td>
<td>0.5</td>
<td>12</td>
<td>40,009,119</td>
<td>42,606,004</td>
<td>27,928–27,947, 16</td>
</tr>
<tr>
<td></td>
<td>20,925–20,944</td>
<td>0.6</td>
<td>11</td>
<td>45,736,640</td>
<td>46,794,995</td>
<td>20,925–20,944, 11</td>
</tr>
<tr>
<td></td>
<td>34,820–34,839</td>
<td>0.5</td>
<td>21</td>
<td>68,751,694</td>
<td>70,050,042</td>
<td>34,820–34,839, 21</td>
</tr>
</tbody>
</table>

1. MY = milk yield, FY = fat yield, PY = protein yield, TSY = total solids yield, FP = fat percentage, PP = protein percentage, TSP = total solid percentage.
2. Genomic variances absorbed by 20-SNP moving windows obtained using single-step genomic BLUP.
3. Genes with start and stop positions within the window were considered. Genes linked to production traits and SCS are in bold and italic font, whereas novel candidate genes are bold, italic, and underlined.
Candidate Genes for FY

For 305-d FY, there were 5 chromosomes that contain associated genomic regions including chromosomes 26 (from 37,336,603 to 38,233,337 bp), 13 (from 46,150,079 to 46,841,024 bp), 1 (from 103,434,094 to 104,344,681 bp), 20 (from 39,625,124 to 40,437,346 bp), and 9 (from 96,298,532 to 97,397,264 bp). Within these regions, 32 known genes and 24 uncharacterized genes were found to be associated with 305-d fat yield (Figure 2B and Table 4).

Three genes were previously reported as candidate genes for milk fat, namely, ADAMTS12 (Oliveira et al., 2018), EZR (Venturini et al., 2014), and IDI1 (Connor et al., 2008; Liu et al., 2009; Khayatzadeh et al., 2016). For IDI1, it encodes an enzyme in cholesterol synthesis which is related to nutrient transport for milk secretion in the mammary gland (Connor et al., 2008; Khayatzadeh et al., 2016). Moreover, IDI1 also plays role in follicle development (Liu et al., 2009).

Additionally, other genes were demonstrated to have mechanisms related to fat yield. The gene RXFP3 can inhibit oxytocin and vasopressin release (Kania et al., 2017), which affects milk flow and milk fat concentration (Högberg et al., 2014). The gene AMACR is an enzyme involved in fatty acids' metabolism (Wright et al., 2011); FNDC1 downregulates osteogenic differentiation pathway of the adipose stem cells (Zhao et al., 2018). The G protein-coupled receptor, TMEM181, is a fat-related QTL in marbling score (Mokry et al., 2013). Additionally, KCNK18 regulates many cellular processes, such as muscle contraction, hormonal regulation, osmoregulation, maintenance of the action potential and ion flow, and this gene has been previously reported to be associated with MY and PY (Andres-Enguix et al., 2012; Maher et al., 2013; Cai et al., 2020).

Furthermore, HSPA12A is a member of the heat shock protein family, other members of which, such as Hsp90, are related to thermotolerance and milk yield (Sailo et al., 2015). Both TRNAT-UGU and LOC104975279 were reported as candidates for milk yield in buffalo (da Costa Barros et al., 2018). The gene PNLIPRP2 was reported as a milk protein candidate gene in Holstein cattle (Li et al., 2010a). Both EMX2 and VAX1 are candidate genes for milk casein (Fomichev et al., 2012); SLC18A2 is an ATP-dependent transporter of monoamines and it is also a candidate gene for SCS in Holstein (Kolbehdari et al., 2009). The gene C1QTNF3 is associated with SCS by playing a role in metabolism, inflammation, and cell survival (Tiezzi et al., 2015; Li et al., 2017). A transporter protein in melanin synthesis, SLC45A2, is a determinant of the skin color (Branicki et al., 2008; Wang et al., 2016; Le et al., 2020). It was reported that skin color is correlated with heat tolerance in sheep (McManus et al., 2011).

Candidate Genes for PY

Genomic regions with a high impact on 305-d PY were located in chromosomes 4 (from 61,234,542 to 62,019,561 bp) and 1 (from 103,434,094 to 104,344,681 bp). A total of 6 known genes and 4 uncharacterized genes associated with protein yield were identified in these regions (Figure 2C and Table 4). From these regions, genomic regions located in chromosome 1 (from 103,434,094 to 104,344,681 bp) were found in both PY and FY. The region contains 2 genes of unknown function, LOC104970979 and LOC781603. In addition, these 2 uncharacterized genes were identified for MY and TSY, and needed further investigation. However, a few genes have been reported to have mechanism indirectly related to protein yield, and HERPUD2 plays a role in metabolic stress (Rabhi et al., 2016). Metabolic stress occurs in lactating cows when homeostasis is disrupted, causing significant production losses (Putman et al., 2018). The actin-binding protein ANLN has functions in cell growth and migration and can influence levels of β-CN and α-LA (a major protein component) that are produced by the epithelial cells of the mammary gland (Rezaei et al., 2016; Layman et al., 2018). The gene EEPD1 was identified, and its function is to maintain genomic stability, which is essential to an animal's health and survival (Wu et al., 2015; Feng and Riddle, 2020).

Candidate Genes for TSY

Genomic regions with a high impact on 305-d TSY were located on 2 chromosomes including chromosomes 1 (from 103,434,094 to 104,344,681 bp) and 15 (from 42,657,355 to 43,538,866 bp). A total of 9 known genes and 5 uncharacterized genes associated with TSY were identified in these windows (Figure 2D and Table 4). However, only a few genes have previously been reported to have mechanism indirectly related to total solid. The gene IRAG1 is found in the membrane of the endoplasmic reticulum, functioning in intracellular calcium modulation, body growth, and adaptation to the arid environment (Widmann et al., 2013; Oberheim Bush and Nedergaard, 2017; Goshu et al., 2018). The gene ADM is found to be involved in several pathways including vasodilation, hormone regulation, angiogenesis, and antimicrobial activity (Schönauer et al., 2017).
Finally, SBF2, a member of the myotubularin-related protein family, is a candidate for fertility traits (Wu et al., 2014).

**Candidate Genes for FP**

Four chromosomes were found to carry genomic regions associated with PP, including chromosomes 27 (from 10,697,257 to 11,822,624 bp), 7 (from 17,871,862 to 19,220,954 bp), 20 (from 31,933,394 to 33,433,160 bp) and 14 (from 1,463,676 to 2,553,525 bp). A total of 116 known genes and 29 uncharacterized genes associated with PP were identified (Figure 2F and Table 4).

Importantly, a large region on chromosome 14 was reported for milk production and compositions in previous studies such as C14H8orf532, ZNF34, RPL8, ZNF7, COMMMD5, ARHGAP39, MIR2308, C14H8orf82, LRRC24, LRRC14, RECQ4L, MFSD3, GPT, PPP1R16A, FOXH1, KIFC2, CYHR1, TONSL, VPS28, SLC39A4, CPSF1, ADCK5, SLC52A2, FBXL6, TMEM1249, SCRT1, DGAT1, HSF1, BOP1, SCX, MROH1, HGH1, MAF1, SHARPIN, CYC1, GPA1, EXOSC4, OPLAH, SMPD5, SPC1, GRNA, PARP10, FAM83H, MAPK15, ZNF623 and GPHHP1 (Salem, 2012; Buitenhuis et al., 2014; Jiang et al., 2014; Shaohua et al., 2015; Nayeri et al., 2016; Palombo et al., 2018; Cai et al., 2020). The gene GPT modulates glutamine, which is a major amino acid in milk protein (Meijer et al., 1993), and RPL8 influences the expression of genes encoding fat-related enzymes (ACACA, FASN, ACS5, FABP3, SREBP1, DGAT1, GPAM, PLIN2, PLIN5, and CIDEA) in bovine mammary epithelial cells (Zheng et al., 2019). For DGAT1, an enzyme in triglyceride synthesis, it has been previously demonstrated that its mutation is associated with milk yield, fat yield and protein yield (Strzalkowska et al., 2005; Näslund et al., 2008; Sun et al., 2009; Lu et al., 2015). Another 3 genes were also reported to directly regulate milk protein: GHR is a causative locus underlying milk yield and milk protein in Chinese Holstein cows (Sun et al., 2009), and transfer RNA (TRNAC-GCA and TRNAW-CCA) are fat yield, milk yield and protein yield-related QTL in buffalo (Venturini et al., 2014). Both RPS28 and microRNA MIR2455 are related to milk whey exosomes in cattle (Izumi et al., 2015; Bhat et al., 2019).

Other genes have been reported to have mechanisms for protein percentage. The gene MAP2K7 is involved in the signal transduction mediating the cell responses to environmental stresses and is a candidate region for milk yield in buffalo (Du et al., 2019); LRRCS6 forms volume-regulated anion channels, a key structure that regulates osmotic volume regulation (König and Stauber, 2019); and OXCT1 plays a role in ketone body catabolism in mammary glands during lactation, and it is related to fat percentage (Bionaz and Loor, 2008). The QTL regions associated with milk protein in this study have been previously found to be related to SCS (SNAPC2, CTXN1, ELAVL1, CCL25, FB3N, CERS4, CD320, NDUFA7, KANK3, ANGPT4, RAB11B, MYO1F, ADAMTS10, ACTL9, GFUS, PYCR3, TIGD5, EEF1D, NAPRT, MROH6, GSDMD, ZC3H3, MAFA, and RHPN1) and immune system (TRIP10, C3, TNFSF14, CD70, and TNFSF9; Griesbeck-Zilch et al., 2009; Li et al., 2010b; Chen et al., 2015; Vrentas et al., 2018; Wu et al., 2019; Rabe et al., 2020). In addition, MicroRNA MIR339B plays a regulatory role in hematopoiesis, mastitis, and breast cancer (Bruchova et al., 2007; Wu et al., 2010; Li et al., 2014).

**Candidate Genes for FP**

Genomic regions associated with FP were located on chromosomes 7 (from 5,774,810 to 6,413,319 bp), 5 (from 76,339,294 to 77,095,345 bp), 1 (from 104,220,701 to 105,570,832 bp), 14 (from 7,718,808 to 8,454,690 bp), 2 (from 130,141,723 to 31,135,560 bp), 1 (from 118,532,290 to 119,421,373 bp), 5 (from 113,468,743 to 114,594,935 bp), and 16 (from 45,049,521 to 46,937,172 bp). These regions harbored 84 known and 39 uncharacterized genes identified as being associated with FP (Figure 2E and Table 4).

Five genes were previously reported to directly regulate milk fat. Candidate gene LOC535121 is responsible for fat and protein percentages in Holstein cows (Cole et al., 2011). The gene MCAT is related to fatty acid synthesis in the lactating period (Sun et al., 2019), whereas TRNAS-GGA and TRNAC-GCA7 are known to be involved in FY, MY, and PY in buffalo milk (Venturini et al., 2014). In addition, a component of intracellular transportation, ARFGER3, has been reported as a candidate gene for milk production traits in Danish Jersey (Kartberg et al., 2010; Mai et al., 2010).

Other genes, such as MIR30D, Ndufa6, and SMDT1, could also indirectly relate to fat percentage. The gene MIR30D regulates insulin, which influences lipid metabolism via SLC27A1 and SLC27A4 (Czech, 2002; Tang et al., 2009), and Ndufa6 and SMDT1 play a role in mitochondria in which their metabolism affects MY and milk fat (Stemberger et al., 1984; Brown et al., 1988; Angerer et al., 2014; Kamer and Mootha, 2015). Other candidate genes, WWTR1, KDM1A, and NAGA, were also linked to lipid metabolism, adipogenesis, or adipocyte differentiation (Sambat et al., 2016; Wang et al., 2016). In addition, PCDH9, C1QA, C1QB, C1QC, EPHA8, LOC781650, LOC782622, TRNAS-GGA, and TRNAG-CCC are candidate genes previously reported to be related to fat deposition and backfat thickness.
(Júnior et al., 2016; Seong et al., 2016; de Lemos et al., 2018; Zhang et al., 2019). Whereas TRNAG-CCC6-1 is a candidate gene for milk lactose and SCS (de Camargo et al., 2015; Costa et al., 2019), miR2284U could be involved in milk yield (Raven et al., 2016). Both TRNAV-CAC and TRNAC-ACA were reported to be linked to protein percentage in milk (de Camargo et al., 2015; da Costa Barros et al., 2018). Calcium homeostasis could be related to osmotic status in milk, and 3 genes, TMEM38A, CHERP, and SMDT1, were identified as being associated with such mechanism (Yazawa et al., 2007; Lin-Moshier et al., 2013; Xie et al., 2019). Additionally, ZBTB40 may be related to calcium metabolism because it influences bone mineral density (Nan et al., 2020). Several genes were also identified as candidate genes for fat percentage; however, they are also linked to reproductive system and embryo development. These genes are USHBPI, MYO9B, HAUS8, SIN3B, NWD1, SLC35E1, USP18, WBP2NL, COMMD2, MIR34A, SLC45A1, ERRFI1, PARK7, TNFRSF9, UTS2, and CAMTA1 (Tong et al., 2006; Malcuit et al., 2009; Forde et al., 2011; Miyamoto et al., 2011; Labrecque et al., 2013; Santonocito et al., 2013; Waters et al., 2014; Crisà et al., 2016; Li et al., 2016; Zielak-Steciwko and Evans, 2016; Biase, 2017; Mallam and Marcotte, 2017; Marchak et al., 2017; Yamada and Sakakibara, 2018; Sánchez et al., 2019). Although the aforementioned genes could be indirectly linked to fat percentage, other candidate genes were also identified within regions associated with this trait. However, functions of such genes are not obviously close to milk fat mechanism. Previously, SYT10 and LOC101903327 were reported to be candidate genes for longevity traits in cattle (Mészáros et al., 2014; Jiang et al., 2019b), whereas LOC101903328 was reported to be a candidate for heat resistance (Jiang et al., 2019a). In addition, MIR30B, ZFAT, SPSB1, and BIK were reported to influence cell death and immunological mechanisms and may affect SCS (Germain et al., 2005; Fujimoto et al., 2009; Tang et al., 2012; Mastrangelo et al., 2018).

Candidate Genes for TSP

Genomic regions that were identified to be associated with TSP were located in chromosomes 6 (from 37,868,743 to 38,892,842 bp), 7 (from 5,774,810 to 6,413,319 bp), 5 (from 113,468,743 to 114,594,935 bp, 21 from 8,031,396 to 8,955,497 bp), 14 (from 7,718,808 to 8,454,690 bp) and 12 (from 40,009,119 to 42,606,004 bp). A total of 61 known and 25 uncharacterized genes associated with TSP were found in these regions (Figure 2G and Table 4).

Several candidate genes for TSP coincide with previous literature such as PPM1K, ARF GAP3, LOC535121, MCAT, and ABCG2-LAP3 region. The gene PPM1K, which regulates mitochondrial permeability, has been previously detected in a region of associated SNPs for lactation persistence and feed efficiency in beef cattle (Do et al., 2017a; Abo-Ismail et al., 2018). ARF GAP3 is a candidate gene for milk production traits in Danish Jersey (Mai et al., 2010). Uncharacterized LOC535121 was identified to be a candidate locus for fat and protein percentages in Holstein breed (Cole et al., 2011). The gene MCAT encodes capable enzyme for fatty acid synthesis during lactation (Sun et al., 2019). Previous significant QTL on BTA 6 containing 6 genes (ABCG2, PKD2, SPP1, MEPE, IBSP, and LAP3) was identified to be a candidate locus for fat production traits (Cohen-Zinder et al., 2005; Olsen et al., 2005, 2007) were also identified in this study. In this region, ABCG2 could be a causative gene affecting milk yield and milk composition in Holstein and Norwegian Red cattle breeds (Cohen-Zinder et al., 2005; Olsen et al., 2007; Alim et al., 2013). The gene SPP1 plays a crucial role in mammary gland development, lactation, and the synthesis of β-CN and whey protein (Nemir et al., 2000), whereas LAP3 is a regulator of hormone, protein maturation, protein inactivation and protein degradation (Zheng et al., 2011). Genes PKD2, CHERP, and TMEM38A are linked to calcium homeostasis and osmotic status in blood and milk (Olsen et al., 2007; Yazawa et al., 2007; Liu-Moshier et al., 2013). The gene TRNAG-ACA is a candidate for protein percentage (de Camargo et al., 2015), whereas NCAPG was found to be related to milk production (Weikard et al., 2012). Additionally, TRNAG-CCC6-1, which functions in decoding mRNA into protein, is a candidate gene for milk lactose in Fleckvieh cattle and for SCS in dairy buffaloes (de Camargo et al., 2015; Costa et al., 2019), and TRNAC-GCA7-1 is a candidate for milk production and composition (Venturini et al., 2014).

Other genes could have mechanism indirectly related to total solid percentage: NDUFA6 and SMDT1 regulate mitochondria, which influence milk yield and milk fat (Stemberger et al., 1984; Brown et al., 1988; Angerer et al., 2014; Kamer and Mootha, 2015). The gene IGF1R regulates mammary gland development and milk synthesis (Szwczuk, 2015), and ARRD4 is involved in cell metabolism and related to TXNIP, which regulates glucose metabolism (Parikh et al., 2007; Zhou and Chng, 2013; Feuer et al., 2016). A member of the nucleotide sugar transporter SLC35 family, SLC35E1, was found to be presented in colostrum milk (Crisà et al., 2016). The gene MED28 is expressed in the mammary gland during lactation and is related to breast cancer (Cohen-Zinder et al., 2005; Huang et al., 2012), whereas NAGA was reported to be involved in lipid metabolism (Kimura et al., 2005). Adipocytes play a vital role in lipid-related energy homeostasis and adipocytes...
Candidate Genes for SCS

Genomic regions associated with SCS were located in 3 chromosomes, 16 (from 26,671,488 to 27,692,042 bp), 11 (from 45,736,640 to 46,794,995 bp), and 21 (from 68,751,694 to 70,050,042 bp) where 52 known and 19 uncharacterized genes were found to be associated with the trait (Figure 2H and Table 4).

Milk somatic cells consist of milk-secreting cells and immune cells (Alhussien and Dang, 2018). Importantly, genes known to be related to immunity, inflammation, or cell proliferation were identified. The gene MIA3 regulates extracellular matrix composition that is expressed in mammary tissue after bacterial infection (Chen et al., 2015; Bauersachs et al., 2017). A member of cytokines in an interleukin superfamily (IL1A, IL1B, IL1F10, IL36A, IL36B, IL36G, and IL37) and 2 IL1 receptors (IL36RN and IL1RN) were identified on BTA 11. They play prominent roles in modulating innate and adaptive immune responses and in the process of inflammation (Benveniste, 2014; Queen et al., 2019).

In addition, RCOR1 is implicated in B cell differentiation and inflammatory responses (Yao et al., 2015; Xiong et al., 2020). Also identified was TRAF3, plays crucial roles in mechanisms of immunity in bovine mammary epithelial cells (Song et al., 2017). Genes CDC42BPB, EXOC3L4 and TNAIP2 are implicated in intracellular transportation of vesicles, and have been previously detected in a candidate region for innate immunity in Canadian Holstein cows (Porat-Shliom et al., 2013; de Klerk et al., 2018). The gene PPP1R13B is an activator of p53, a regulator in immunity and apoptotic pathway (Mitchell et al., 2002; Liu et al., 2005), whereas TAF1A regulates cell proliferation and is implicated in breast cancer (Bergstrahl et al., 2007; Rossetti et al., 2016). Other genes were also identified to be linked to SCS in this population; however, they have been also found to be related to milk yield and composition in other studies. These genes are TRNAT-UGU, TRNAC-GCA7-1, HHIPL2 and NCK2 (Venturini et al., 2014; da Costa Barros et al., 2018; Johnston et al., 2018; Oliveira et al., 2018).

In this study, 210 and 21 candidate genes were identified as influence milk production and SCS, respectively (Table 4). These candidate genes largely describe polygenic inheritance of the traits. Several candidate genes were coincided with other previous reports for milk production in Holstein cattle, especially a large region of genes on BTA14 (Atashi et al., 2020; Cai et al., 2020; Silva et al., 2020; Wang et al., 2020). This could be attributed to the predominance of Holstein blood in Thai dairy cattle population by upgrading the local population with Holstein and later, mating within population to keep the proportion of Holstein blood between 87.5 and 93.75% as being considered the most suitable for tropical condition in Thailand (Department of Livestock Development, 2020). On the other hand, some identified genes are different from other breeds; for instance, LRRC24 was a candidate gene for milk composition in our population but Jersey nor Ayrshire (Oliveira et al., 2019a). Moreover, we also identified 141 novel genes for milk production and compositions and 5 prospective genes for SCS (Table 4). Most of these genes have mechanism indirectly related to SCS, and they have not been shown to influence the trait in previous studies.

The genes previously reported in literature to be involved in fertility, longevity, and environmental adaptation such as heat resistance were included in several traits. In Thailand’s context, such genes could likely influence milk production and composition because low milk production and fertility, as well as mastitis problems, are the main causes of involuntary culling. In addition, dairy cattle population in Thailand was naturally selected by high temperature and humidity in a tropical environment. That environmental condition is not optimal for Holstein populations compared with that of other countries (Boonkum et al., 2011; Collier et al., 2011). Some genomic regions were associated with 2 or more traits. These regions were located on chromosomes 7 (from 5,774,810 to 6,413,319 bp), 5 (from 113,468,743 to 114,594,935 bp), 14 (from 7,718,808 to 8,454,690 bp) and 12 (from 40,009,119 to 42,606,004 bp) were identified to be associated with both TSP and FP. In addition, 2 uncharacterized genes, LOC104970979 and LOC781603, were identified for MY, FY, PY, and
The results suggest these genes may be biological or spurious pleiotropic effects (Solovieff et al., 2013). Further investigation is needed to be conducted.

For SCS, a large member of interleukin superfamily (IL1A, IL1B, IL1F10, IL36A, IL36B, IL36G, and IL37) as well as other genes related to immunity significantly influencing SCS were identified. One of the limitations of this study is that the small number of genotyped animals used in this study could result in inaccurate estimation of SNP effects using ssGBLUP because the inverse of G may be noninformative. Furthermore, we annotated those SNPs on Bovine UMD 3.1, which was sequenced from a beef breed (Hereford). It should be also noted that UMD 3.1 provided genomic difference from Holstein cow genome such as variations in insertion/deletion accounting for 48,537,190 bp of the entire genome (Køks et al., 2013). This suggests critical concerns for annotation and linked gene identification (Eilbeck et al., 2009; Florea et al., 2011; Wu et al., 2013). Annotation of these SNPs on their own genome map should be more appropriate.

**Network and Pathway Analysis**

**Milk Production and Composition.** Gene network analysis of the gene list in milk production demonstrated using GeneMania revealed the dense co-expression network (74.02%; Figure 3). The network included 170 genes with 1,137 interactions. Among those genes, PKD2, TONSL, TNFSF14, WWTR1, TNFSF14, CHERP, PRAP10 and GPIHBP1 are involved in several regulations of intracellular cell transportations. These findings are consistent with the results from functional analyses in gene ontology (GO, Table 5). The pathways involved in cellular process and cell transportation were identified, such as protein ubiquitination (GO:0016567), endoplasmic reticulum membrane (GO:0005789), actin cytoskeleton organization (GO:0030036), positive regulation of receptor recycling (GO:0001921), cell-cell adherens junction (GO:0005913), and extracellular exosome (GO:0070062). In addition, some pathways were found to be related to calcium release, such as the release of sequestered calcium ion into cytosol (GO:0051209). These cellular mechanisms may influence the transportation of milk constituents via mammary secretory cells. The GO term analysis also identified osmoregulation (GO:0006970), which influences fluid quantity in the body including milk. Furthermore, pathways related to protein, a major component in milk, were also identified, such as protein catabolic process (GO:0030163), aspartic-type endopeptidase activity (GO:0004190), and proteolysis (GO:0006508). Other pathways related to oxygen-containing radicals (reactive oxygen species) were also identified, such as age-dependent response to reactive oxygen species (GO:0001315), superoxide metabolic process (GO:0006801), superoxide dismutase activity (GO:0004784). Indeed, the uncontrolled free radicals can lead to the breakdown of vital biochemical compounds, such as lipids and protein, and they can affect milk composition (Khan et al., 2019). Immunoregulation and inflammation were also identified in GO of milk production and composition, such as complement activation (GO:0006958), immune response (GO:0006955), positive regulation of activation of membrane attack complex (GO:0001970), endocytosis (GO:0006897), and blood microparticle (GO:0072562). Results from this study conformed to gene expression patterns in human mammary epithelial cells and breast cancers in GeneMania (Perou et al., 1999; Bild et al., 2006; Warde-Farley et al., 2010).

**Udder Health.** Gene network analysis of gene list in udder health revealed the co-expression network (26.97%; Figure 4). The network included 29 genes with 141 interactions. These genes are mainly related to immunological pathways such as lymphocyte activation. Table 5 presents the pathways identified using genes annotated for SCS with the highest explained variance. Most pathways confirm the finding of genes involved in immunological mechanism, inflammation, and infection, such as cytokine-mediated signaling pathway (GO:0019221), positive regulation of IL-6 production (GO:0032755), fever generation (GO:0001660), exocytosis (GO:0006887), negative regulation of cytokine-mediated signaling pathway (GO:0001960), immune response (GO:0006955), intrinsic apoptotic signaling pathway (GO:0097193), toll-like receptor signaling pathway (GO:0002224), positive regulation of I-kappaB kinase/NF-kappaB signaling (GO:0043123), IL-1 receptor binding (GO:0005149), and cytokine activity (GO:0005125). Other pathways are related to cell division (positive regulation of cell division, GO:0051781), DNA translation (transcription from RNA polymerase I promoter, GO:0006360), and cellular process (extracellular space, GO:0005615; exocyst, GO:0000145; ARF guanyl-nucleotide exchange factor activity, GO:0005086).

Based on our findings, in spite of a small number of genotyped animals, genomic prediction seems to be useful for improving genetic gain in the Thai dairy cattle breeding program. The accuracies of predictions using WssGBLUP were greater than ssGBLUP and pedigree-based BLUP for all studied traits. Results from a weighted ssGWAS revealed many genomic regions and candidate genes associated with 7 milk production traits and SCS. However, further investigation is needed for some genomic regions with known or uncharacterized genes (e.g., LOC104970979).
and LOC781603). Furthermore, future studies should also consider the usefulness of the random regression model in estimating genomic merits and SNP effects at particular period within a lactation, which will help to elucidate the associated genomic regions and genes specific to time for milk-related traits. In addition, the heterogeneity assumption of residual variance should be considered to better account for measurement errors at different stages of lactation.

CONCLUSIONS

This study performed a weighted GWAS using single-step procedure and found many genomic regions and

Figure 3. Gene network analysis of milk production using GeneMania (Warde-Farley et al., 2010). Network includes candidate genes and related genes (170 genes, circles) with 1,137 interactions (edges).
<table>
<thead>
<tr>
<th>Pathway Description</th>
<th>Gene</th>
<th>P-value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>MY, FY, PY, and TSY</td>
<td>GO 003163: protein catabolic process</td>
<td>24</td>
<td>5.3E-43</td>
</tr>
<tr>
<td></td>
<td>GO 0006508: proteolysis</td>
<td>24</td>
<td>1.2E-26</td>
</tr>
<tr>
<td></td>
<td>GO 0001315: age-dependent response to reactive oxygen species</td>
<td>2</td>
<td>1.0E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0006801: superoxide metabolic process</td>
<td>2</td>
<td>7.5E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0004190: aspartic-endopeptidase activity</td>
<td>25</td>
<td>1.4E-46</td>
</tr>
<tr>
<td></td>
<td>GO 0004784: superoxide dismutase activity</td>
<td>2</td>
<td>2.5E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0006958: complement activation, classical pathway</td>
<td>4</td>
<td>5.5E-04</td>
</tr>
<tr>
<td></td>
<td>GO 0006955: immune response</td>
<td>8</td>
<td>7.0E-03</td>
</tr>
<tr>
<td></td>
<td>GO 0001970: positive regulation of activation of membrane attack complex</td>
<td>2</td>
<td>2.1E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0051209: release of sequestered calcium into cytosol</td>
<td>3</td>
<td>2.5E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0030036: actin cytoskeleton organization</td>
<td>4</td>
<td>4.9E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0006897: endocytosis</td>
<td>4</td>
<td>6.7E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0001921: positive regulation of receptor recycling</td>
<td>2</td>
<td>8.0E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0006970: response to osmotic stress</td>
<td>2</td>
<td>8.9E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0001657: protein ubiquitination</td>
<td>5</td>
<td>9.5E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0007262: blood microparticle</td>
<td>5</td>
<td>6.1E-03</td>
</tr>
<tr>
<td></td>
<td>GO 0005789: endoplasmic reticulum membrane</td>
<td>9</td>
<td>2.60E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0005038: recycling endosome membrane</td>
<td>3</td>
<td>3.20E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0005913: cell-cell adherens junction</td>
<td>4</td>
<td>6.0E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0007062: extracellular exosome</td>
<td>28</td>
<td>9.80E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0035925: mRNA 3′-UTR AU-rich region binding</td>
<td>2</td>
<td>4.80E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0016874: ligase activity</td>
<td>4</td>
<td>6.30E-02</td>
</tr>
<tr>
<td>SCS</td>
<td>GO 0019221: cytokine-mediated signaling pathway</td>
<td>9</td>
<td>5.30E-10</td>
</tr>
<tr>
<td></td>
<td>GO 0032755: positive regulation of interleukin-6 production</td>
<td>3</td>
<td>5.10E-03</td>
</tr>
<tr>
<td></td>
<td>GO 0035600: fever generation</td>
<td>2</td>
<td>5.30E-03</td>
</tr>
<tr>
<td></td>
<td>GO 0036887: exocytosis</td>
<td>3</td>
<td>6.70E-03</td>
</tr>
<tr>
<td></td>
<td>GO 0001960: negative regulation of cytokine-mediated signaling pathway</td>
<td>2</td>
<td>1.60E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0001601: exocytosis localization</td>
<td>2</td>
<td>1.90E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0006955: immune response</td>
<td>2</td>
<td>2.00E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0007190: intrinsic apoptotic signaling pathway</td>
<td>2</td>
<td>3.70E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0006360: transcription from RNA polymerase I promoter</td>
<td>2</td>
<td>3.70E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0002224: toll-like receptor signaling pathway</td>
<td>2</td>
<td>3.70E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0004121: positive regulation of I- kappaB kinase/ NF-kappaB signaling</td>
<td>3</td>
<td>4.50E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0001761: positive regulation of cell division</td>
<td>2</td>
<td>6.50E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0006515: extracellular space</td>
<td>11</td>
<td>9.1E-05</td>
</tr>
<tr>
<td></td>
<td>GO 0000145: exocyst</td>
<td>2</td>
<td>3.70E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0005149: interleukin-1 receptor binding</td>
<td>9</td>
<td>5.4E-19</td>
</tr>
<tr>
<td></td>
<td>GO 0005125: cytokine activity</td>
<td>9</td>
<td>1.2E-09</td>
</tr>
<tr>
<td></td>
<td>GO 0005086: ARF guanyl-nucleotide exchange factor activity</td>
<td>2</td>
<td>5.50E-02</td>
</tr>
<tr>
<td></td>
<td>GO 0000149: SNARe binding</td>
<td>2</td>
<td>9.4E-02</td>
</tr>
</tbody>
</table>

\(^{1}\)MY = milk yield, FY = fat yield, PY = protein yield, TSY = total solids yield, FP = fat percentage, PP = protein percentage, TSP = total solid percentage.

Table 5. Pathways identified by the Database for Annotation, Visualization and Integrated Discovery (DAVID version 6.7) in gene ontology (GO) term
candidate genes associated with 7 milk production traits and SCS, indicating a polygenic nature of the studied traits. Functional analysis revealed genes potentially influencing milk production and composition, including cellular transportations, protein catabolism, and osmoregulation. Moreover, the identified genes related to immunological pathway strongly support their effects on SCS. Novel candidate genes implicated in milk production traits and SCS were also identified. A group of candidate genes for milk production traits were found to be overlapped with other traits (SCS, fertility, longevity, and heat tolerance), likely being

![Gene network](image)

**Figure 4.** Gene network produced using GeneMania (Warde-Farley et al., 2010) for udder health. The network consists of candidate genes and related genes (29 genes, circles) connected by 141 interactions (edges).
attributed to involuntary selection in the population (e.g., culling of cows with fertility and mastitis problems). Knowledge about the genetic architecture of the tropical dairy cattle population is important and will provide valuable information for future studies. Further research with more information on animals, records, and genotypes is required to validate our findings.

ACKNOWLEDGMENTS

This study was a part of the Genomic Evaluation for Genetic Improvement of Tropical Holstein Dairy Cattle in Thailand project, originally established by the Bureau of Biotechnology in Livestock Production (BBLP), the Department of Livestock Development, and Khon Kaen University, with support from the Agricultural Research Development Agency. The authors also acknowledge BBLP for providing the genotype and test-day data. The authors have not stated any conflicts of interest.

REFERENCES


Forde, N., F. Carter, T. E. Spencer, F. W. Bazer, O. Sandra, N. Mau-
souri-Attia, L. A. Okumu, P. A. McGregor, J. P. Mehta, R. Mc-
Bride, P. O’Gaora, J. F. Roche, and P. Lonergan. 2011. Conceptus-
induced changes in the endometrial transcriptome: How soon does the cow
doi.org/10.1095/biolreprod.110.090019.
Fragomeni, B. O., I. Misztal, D. L. Lourenco, I. Aguilar, R. Okimoto,
and W. M. Muir. 2014. Changes in variance explained by top
SNP windows generations for three traits in broiler chicken. Front.
Freitas, P. H. F., H. R. Oliveira, F. F. Silva, A. Fleming, F. S. Schen-
kel, F. Miglior, and L. F. Brito. 2020. Short communication: Time-
dependent genetic parameters and single-step genomic associate-
ment analyses for predicted milk fatty acid composition in Ayr-
Fujimoto, T., K. Doi, M. Koyanagi, T. Tsumoda, Y. Takashima, Y.
Yoshida, T. Sasaki, and S. Shirasawa. 2009. ZFAT is an anti-
apoptotic molecule and critical for cell survival in MOLT-4 cells.
García, A. L. S., B. Bosworth, G. Waldbieser, I. Misztal, S. Tsuruta,
and W. M. Muir. 2014. Changes in variance explained by top
SNP windows generations for three traits in broiler chicken. Front.
Freitas, P. H. F., H. R. Oliveira, F. F. Silva, A. Fleming, F. S. Schen-
kel, F. Miglior, and L. F. Brito. 2020. Short communication: Time-
dependent genetic parameters and single-step genome-wide assoxi-
ation analyses for predicted milk fatty acid composition in Ayr-
Fujimoto, T., K. Doi, M. Koyanagi, T. Tsumoda, Y. Takashima, Y.
Yoshida, T. Sasaki, and S. Shirasawa. 2009. ZFAT is an anti-
apoptotic molecule and critical for cell survival in MOLT-4 cells.
García, A. L. S., B. Bosworth, G. Waldbieser, I. Misztal, S. Tsuruta,
and W. M. Muir. 2014. Changes in variance explained by top
SNP windows generations for three traits in broiler chicken. Front.
Freitas, P. H. F., H. R. Oliveira, F. F. Silva, A. Fleming, F. S. Schen-
kel, F. Miglior, and L. F. Brito. 2020. Short communication: Time-
dependent genetic parameters and single-step genome-wide assoxi-
ation analyses for predicted milk fatty acid composition in Ayr-
Fujimoto, T., K. Doi, M. Koyanagi, T. Tsumoda, Y. Takashima, Y.
Yoshida, T. Sasaki, and S. Shirasawa. 2009. ZFAT is an anti-
apoptotic molecule and critical for cell survival in MOLT-4 cells.
García, A. L. S., B. Bosworth, G. Waldbieser, I. Misztal, S. Tsuruta,
and W. M. Muir. 2014. Changes in variance explained by top
SNP windows generations for three traits in broiler chicken. Front.


Lotfan, M., S. Kumar, M. L. Yadav, S. Jaswal, S. Bathla, P. Rawat, R. Baithalu, T. K. Mohanty, A. K. Dang, and J. K. Kaushik. 2014. Pregnancy Associated Glycoprotein 7 (PAG7) is the most


Xie, K.-F., D.-D. Guo, and X.-J. Luo. 2019. SMDT1-driven c


ORCIDS

S. Buaban https://orcid.org/0000-0002-2239-2151
K. Lengnudum https://orcid.org/0000-0001-6544-8162
W. Boonkum https://orcid.org/0000-0003-2239-2300
P. Phakdeedinand https://orcid.org/0000-0002-5994-5215

Journal of Dairy Science Vol. 105 No. 1, 2022