Characterization of biological pathways associated with semen traits in the Thai multibreed dairy population

Mattaneeya Sarakul[a], Mauricio A. Elzo[b], Skorn Koonawootrittriron[a,*], Thanathip Suwanasopee[a], Danai Jattawa[a], Thawee Laodim[a]

[a]Department of Animal Science, Kasetsart University, Bangkok 10900, Thailand

[b]Department of Animal Sciences, University of Florida, Gainesville, FL 32611-0910, USA

[*]Corresponding author: Department of Animal Science, Faculty of Agriculture, Kasetsart University, Bangkok 10900, Thailand; Email: agrskk@ku.ac.th; Tel: +66-2-5791120; Fax: +66-2-5791120.

ABSTRACT

The objective of this research was to characterize biological pathways associated with semen volume (VOL), number of sperm (NS), and sperm motility (MOT) of dairy bulls in the Thai multibreed dairy population. Phenotypes for VOL (n = 13,535), NS (n = 12,773), and MOT (n = 12,660) came from 131 bulls of the Dairy Farming Promotion Organization of Thailand. Genotypic data consisted of 76,519 imputed and actual single nucleotide polymorphisms (SNP) from 72 animals. The SNP variances for VOL, NS, and MOT were estimated using a three-trait genomic-polygenic repeatability model. Fixed effects were contemporary group, ejaculate order, age of bull, ambient temperature, and heterosis. Random effects were animal additive genetic, permanent environmental, and residual. Individual SNP explaining at least 0.001% of the total genetic variance for each trait were selected to identify associated genes in the NCBI database (UMD *Bos taurus* 3.1 assembly) using the R package Map2NCBI. A set of 1,999 NCBI genes associated with all three semen traits was utilized for the pathway analysis conducted with the ClueGO plugin of Cytoscape using information from the Kyoto Encyclopedia of Genes and Genomes database. The pathway analysis revealed seven significant biological pathways involving 127 genes that explained 1.04% of the genetic variance for VOL, NS, and MOT. These genes were known to affect cell structure, motility, migration, proliferation, differentiation, survival, apoptosis, signal transduction, oxytocin release, calcium channel, neural development, and immune system functions related to sperm morphology and physiology during spermatogenesis.

Keywords: Cattle; Dairy; Genomic; Multibreed; SNP; Tropical

1. Introduction

Semen traits are important for cattle genetic improvement programs relying on artificial insemination. Bull quantity and quality of semen are influenced by genetic, physiological, management, and environmental factors (Mathevon et al., 1998; Druet et al., 2009; Karoui et al., 2011). Estimates of heritabilities for semen traits (volume, number of sperm, sperm motility), however, range from low (0.04) to moderate (0.27; Kapš et al., 2000; Druet et al., 2009; Nishimura et al., 2010; Sarakul et al., 2018) making the identification of superior animals in a population inaccurate. Recently, genome-wide association studies (GWAS) identified single nucleotide polymorphisms (SNP) related to semen traits in dairy populations (Hering et al., 2014a; Kaminski et al., 2016; Puglisi et al., 2016; Qin et al., 2016). Large numbers of SNP associated with semen volume, number of sperm, and sperm motility were identified in Holstein in temperate regions (Druet et al., 2009; Hering et al., 2014b; Suchocki and Szyda, 2015). Similarly, regions of SNP associated with semen volume and sperm motility were identified in all autosomes and the X chromosome in beef cattle in tropical regions (Fortes et al., 2012, 2013).

Unfortunately, individual SNP usually account for only a small proportion of trait genetic variances and cannot explain the biological function of these traits. A more reliable alternative would be to consider sets of SNP associated with genes in biological pathways that account for sizeable fractions of the genetic variance for semen traits. From research with the USA Holstein (Peñagaricano et al., 2013), there were sets of genes that were found to be associated with conception rate that are involved in multiple pathways (small GTPase mediated signal transduction, neurogenesis, calcium ion binding, and cytoskeleton); however, percentages of genetic variation accounted for by these pathways were not computed. Nevertheless, there is currently no information on SNP markers, genes, and biological pathways associated with semen traits in the Thai multibreed dairy population. Sets of SNP associated with genes that when expressed result in modulating biological pathways affecting semen traits could be included in customized genotyping chips to

increase the accuracy of genomic selection for semen traits in Thailand. Thus, the objective of the present research was to characterize biological pathways associated with semen volume, number of sperms, and sperm motility of dairy bulls in the Thai multibreed dairy population when bulls were maintained in tropical environmental conditions.

2. Materials and methods

2.1. Animals, management and traits

This study included 131 bulls with records for semen volume (n =13,535), number of sperm (n = 12,773), and sperm motility (n = 12,660) collected by the Semen Production and Dairy Genetic Evaluation Center of the Dairy Farming Promotion Organization of Thailand (DPO) from 2001 to 2017. These bulls were the offspring of 62 sires and 112 dams. Ninety-five percent of the bulls were Holstein crossbreds and 5% were Holstein purebreds. Bulls were, on the average, 92% Holstein (minimum = 62%; maximum = 100%). The 95% Holstein crossbred bulls had percentage composition of one or more of seven breeds (Brown Swiss, Jersey, Red Dane, Brahman, Red Sindhi, Sahiwal, and Thai Native; Koonawootrittriron et al., 2009).

Bulls were maintained in open-barn stalls, except at the times of semen collection. Bulls were fed concentrate (4 to 6 kg/d; 16% of CP, 2% fat, 14% fiber, and 13% moisture; Charoen Pokphand Foods, Bangkok, Thailand), fresh roughage (*ad libitum*), and had free access to water and a mineral supplement. Fresh roughage consisted of Guinea (*Panicum maximum*), Ruzi (*Brachiaria ruziziensis)*, Napier (*Pennisetum purpureum*), and Para (*Brachiaria mutica*) grasses cut and transported to bull stalls. Guinea and Ruzi grass hay and silage were provided during periods when there was limited availability of fresh grass (November to June).

Traits considered in this study were semen volume (VOL, ml), number of sperm (NS, million), and sperm motility (MOT, %). The three semen traits were measured by a single

trained technician throughout the years of the study. Semen volume was the amount of semen per ejaculate measured in a scaled tube. Number of sperm per ejaculate was equal to the product of semen volume (ml) by sperm concentration per ml (millions of sperm per ml). Sperm concentration was determined using a hemocytometer. A sample of 0.1 ml of semen was diluted with 9.9 ml of phosphate buffer saline (dilution factor 1:100). Subsamples of 0.1 µl of the diluted sample were placed in each of five counting areas of the hemocytometer. Sperm concentration (millions of sperm per ml) was computed as the average number of sperm per counting area multiplied first by 10,000 (number of sperm per 0.1 ml) and then by 100 (dilution factor) to yield number of sperm per ml. Sperm motility was estimated using five microcells with an optical microscope (magnification = 400x). The technician computed the average percentage of active motile sperm (spermatozoa moving forward progressively) in the five microcells twice. Sperm motility was defined as the mean of these two repeated measurements. Data recorded at each semen collection included bull identification, date and time, ejaculation number, ambient temperature (ºC), and collector's name.

2.2. Genotypic data

This research utilized genotypic data from 61 bulls and from 11 dams of sires obtained in a previous research project with the Thai multibreed dairy population (Jattawa et al., 2016). The DNA extraction and genotyping process was described in Jattawa et al. (2016). Briefly, genomic DNA was extracted from frozen semen using a GenElute™ Mammalian Genomic DNA Miniprep kit (Sigma®, USA), and from blood samples using a MasterPure™ DNA Purification kit (Epicentre® Biotechnologies, USA). DNA quality was assessed using a NanoDrop™ 2000 spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, USA). Samples prepared for genotyping had a minimum DNA concentration of 15 ng/µl and an

absorbance ratio of about 1.8 at 260/280 nm. The DNA samples were genotyped with GeneSeek Genomic Profiler (GGP) chips (GeneSeek Inc., Lincoln, NE, USA).

Bulls (n = 61) were genotyped with GGP80K, and dams (n = 11) were genotyped with GGP9K. Dams genotyped with GGP9K were imputed to GGP80K using version 2.2 of program FImpute (Sargolzaei et al., 2014). Imputation was conducted with the complete genotype dataset of the Thai multibreed population (n = 2,661; GGP9K = 1,412 cows, GGP20K = 570 cows, GGP26K = 540 cows, GGP80K = 89 sires and 50 cows; Jattawa et al., 2016). The four input genotypic files for FImpute only included SNP from autosomes (n = 29) and the X chromosome (SNP from the Y chromosome are not allowed by this program). Quality control consisted of removing SNP markers with call rates lower than 0.9 or minor allele frequencies lower than 0.04. The edited genotype file contained 76,519 SNP markers per animal.

2.3. SNP variances for semen traits

The variances explained by each of the SNP for each semen trait were computed with program POSTGSF90 (Misztal et al., 2018) using a two-step procedure (Wang et al., 2012), where the SNP most precise predictions are computed in the first step, and variances of SNP are computed iteratively in the second step, using the SNP with the most precise predictions from the first step as initial values.

Program POSTGSF90 requires predicted animal genomic values for VOL, NS, and MOT as inputs for the computation of the SNP most precise predictions. Predicted animal genomic values as well as variance components were obtained with a three-trait mixed repeatability model using a single-step genomic BLUP procedure (Aguilar et al., 2010; Wang et al., 2012) with the program AIREMLF90 (Tsuruta, 2014). Fixed effects were contemporary group (year-month of collection), ejaculate order (first or second), age of bull (mo), ambient temperature (^{o}C), and heterosis as a function of heterozygosity ([sire Holstein fraction × dam

O fraction] + [sire O fraction × dam Holstein fraction], O = breeds other than Holstein such as Brown Swiss, Red Danish, Jersey, Red Sindhi, Sahiwal, Brahman, and Thai Native). Random effects were animal additive genetic, permanent environmental, and residual. For a detailed description of the model and its assumptions refer to Sarakul et al. (2018).

2.4. Genes associated with semen traits

Percentages of genetic variances explained by each SNP for VOL, NS, and MOT were computed as ratios of individual SNP variances to total genetic variances multiplied by 100. The SNP that explained at least 0.001% of the total genetic variance for each one of the three traits (VOL, NS and MOT) were selected to identify associated genes. Genes associated with these SNP markers were identified using the UMD *Bos taurus* 3.1 assembly of the bovine genome at the National Center for Biotechnology Information (NCBI) using the R package Map2NCBI (Hanna and Riley, 2014). The Map2NCBI identified genes by the distance in base pairs (bp) between the SNP markers and the genes in the NCBI database (inside genes, within 2,500 bp, between 2,500 bp and 5,000 bp, between 5,000 bp and 25,000 bp, and over 25,000 bp).

2.5. Pathway analysis

The subset of genes identified by SNP accounting for at least 0.001% of the genetic variance for each semen trait that was either inside genes or within 2,500 bp from these genes was used to determine significantly enriched/depleted pathways in the Thai multibreed dairy population. Enriched/depleted pathways were detected by inputting the list of genes associated with all three semen traits into the ClueGO plugin of Cytoscape (Bindea et al., 2009) and relating these to *Bos taurus* genes in the Kyoto Encyclopedia of Genes and Genomes database (KEGG). The significance of enriched/depleted biological pathways was determined using

multiple two-sided hypergeometric tests corrected with a Bonferroni step down procedure (Holm, 1979). Pathways with a *P*-value < 0.05 were assumed to be significant.

3. Results and discussion

3.1. Number of SNP, SNP variance, and number of genes associated with VOL, NS, and MOT

Table 1 shows the numbers of SNP accounting for at least 0.001% of the genetic variance for VOL, NS and MOT in the Thai dairy cattle population classified by distance between SNP and genes in the NCBI database. The number of SNP located inside genes and within 2,500 bp of genes explaining at least 0.001% of the genetic variance for VOL, NS, and MOT represented 15% of SNP influencing these traits, and accounted for approximately 33% of the total genetic variance.

Data in Table 2 includes numbers of SNP and percentage of genetic variance accounted for by each chromosome for VOL, NS, and MOT. Large numbers of SNP for each trait were found in all 29 autosomes and the X chromosome for the three semen traits. Numbers of SNP per chromosome for all semen traits ranged from 138 for VOL in chromosome 27 to 577 for MOT in chromosome 19. The chromosomes explaining the greatest percentage of the genetic variance per trait were chromosome 19 for VOL (3.44%), and chromosome X for NS (3.71%) and MOT (3.47%). Conversely, chromosome 27 explained the least percentage of the genetic variance for all semen traits (0.43% for VOL, 0.60% for NS, and 0.54% for MOT). The wide distribution of SNP on all chromosomes and the low percentage of the genetic variance explained by each SNP clearly indicates the quantitative nature of these three semen traits.

There were large numbers of genes associated with VOL, NS, and MOT per chromosome for single traits, pairs of traits, and all three traits are included in Table 3. The number of genes associated with a single trait ranged from 56 (chromosome 27; VOL) to 338

(chromosome 5; VOL). The number of genes associated with pairs of traits fluctuated between 24 (chromosome 27; VOL and MOT) and 228 (chromosome 19; NS and MOT). Lastly, the number of genes associated with all three semen traits ranged from 16 (chromosome 27) to 141 (chromosome 19). The total numbers of genes associated with a single trait (5,576 genes for VOL, 5,667 genes for NS, and 5,694 genes for MOT) were greater than the corresponding total numbers of genes associated with two traits (3,215 genes between VOL and NS, 3,006 genes between VOL and MOT, and 3,771 genes between NS and MOT), and three traits (1,999 genes). On the average, 59% of the genes associated with VOL, NS, and MOT affected pairs of traits, and 35% affected all three traits. Genes influencing two and three semen traits provide a genomic basis for the genetic correlations previously estimated between these traits in this population by Sarakul et al. (2018).

3.2. Biological pathways associated with VOL, NS, and MOT

The set of genes that was associated with all three semen traits (n = 1,999) was used as input for the pathway analysis conducted with the ClueGO plugin of Cytoscape (Bindea et al., 2009) using information from the KEGG database. The ClueGO pathway analysis indicated there were seven significant biological pathways involving 127 genes affecting VOL, NS, and MOT (Table 4; Fig. 1; Supplemental Table 1). The information for total number of genes in Table 4 (n = 197) included the representation of the 127 genes in one pathway (n = 84) or multiple pathways (n = 43; 2 to 6 pathways). These seven biological pathways were classified into three categories: cellular processes, organismal systems, and environmental information processing. The 127 genes included in these three categories explained an average of 1.04% of the genetic variance for VOL, NS, and MOT (Table 4). The organismal systems category was composed of a greater number of genes (n = 72) than the cellular processes (n = 68) and environmental information processing (n = 57) categories (Table 4). The average genetic

variance for VOL, NS, and MOT explained by cellular processes pathways was larger (0.42%) than that explained by organismal systems (0.39%) and environmental information processing (0.22%) pathways (Table 4).

3.2.1. Cellular process pathways

Cellular processes were composed of two pathways (focal adhesion and actin cytoskeleton regulation; Table 4; Fig. 1; Supplemental Table 1). The focal adhesion pathway was composed of a greater number of genes (n = 36) than the regulation of actin cytoskeleton pathway (n = 32), and each pathway accounted for approximately 0.20% of the genetic variance for VOL, NS, and MOT (Table 4). There were 73% of the genes in the cellular processes category that were primarily involved in cell structure, shape, motility, migration, adhesion, proliferation, differentiation, survival, and apoptosis, and 27% were involved in other functions (angiogenesis, homeostasis, immune system, and signal transduction) contributing to physiological and morphological changes during spermatogenesis (Table 4; Supplemental Table 1).

3.2.1.1. Focal adhesion pathway

Genes *ACTN4, PTK2, AKT3,* and *PIK3CA* in the focal adhesion pathway affect sperm motility, transport, capacitation, morphology, and viability in mice, and humans (Du Plessis et al., 2004; Gallardo Bolaños et al., 2014; Gungor-Ordueri et al., 2014; Qian et al., 2014). The *ACTN4* gene (chromosome 18; codes for actinin alpha 4) was a member of both cellular processes pathways (focal adhesion and regulation of actin cytoskeleton).

The *ACTN4* gene encodes for an actin binding protein from the spectrin gene superfamily that helps regulate the organization of actin microfilaments at the Sertoli-spermatid interface, which is important for the passage of spermatids across the seminiferous epithelium

during spermatogenesis (Qian et al., 2014). Gene *PTK2* (chromosome 14; codes for protein tyrosine kinase 2; also known focal adhesion kinase, FAK) was present in three pathways (focal adhesion, actin cytoskeleton regulation and axon guidance). The encoded protein has an essential role in cell migration (Ritt et al., 2013), regulation of actin polymerization, and remodeling of the actin cytoskeleton during acrosome reaction (Roa-Espitia et al., 2016). In addition, *PTK2* facilitates the movement of spermatids across the epithelium and preleptotene spermatocytes across the blood-testis barrier during spermatogenesis (Gungor-Ordueri et al., 2014) and is involved in the regulation of intracellular calcium ion during sperm capacitation and acrosome reaction (Breitbart and Naor, 1999). The *AKT3* gene (chromosome 16; codes for AKT serine/threonine kinase 3; also known as protein kinase B) was involved in four pathways (focal adhesion, B cell receptor signaling, rap1 signaling, and sphingolipid signaling) with its encoded protein influencing sperm motility and viability by maintaining spermatozoa membrane integrity (Gallardo Bolaños et al., 2014). The *PIK3CA* gene (chromosome 1; codes for phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha) was involved in six pathways (focal adhesion, actin cytoskeleton regulation, axon guidance, B cell receptor signaling, rap1 signaling, and sphingolipid signaling). The *PIK3CA* protein belongs to the phosphoinositide-3-kinase (PI3K) family of enzymes that has a major role in cell motility, transport, growth, survival, proliferation, differentiation, and cytoskeletal organization (Fruman et al., 1998; Kordi-Tamandani and Mir, 2015). More specifically, PI3K influences sperm capacitation and acrosome reaction as well as sperm motility and sperm-oocyte binding (Fisher et al., 1998; Du Plessis et al., 2004).

3.2.1.2. Actin cytoskeleton regulation pathway

Genes in the actin cytoskeleton regulation pathway are important for cell shape, motility, migration, and maintenance of the cytoskeletal structure (Table 4; Fig. 1;

Supplemental Table 1). Cellular events regulated by genes in the actin cytoskeleton pathway (e.g., *APC, SPATA13, ARHGEF4, ARHGEF6, FGD1, PIP4K2A, PIP4K2B*) occurring during spermatogenesis are responsible for extensive changes in cell shape, cell size, and germ cell movement (Lie et al., 2010). The *APC* gene (chromosome 10; codes for adenomatosis polyposis coli, an antagonist to the Wnt signaling pathway) is involved in cell migration, adhesion, and organization of the actin and microtubule networks (Bienz and Hamada, 2004; Aoki and Taketo, 2007; GeneCards, 2018). Proteins encoded by genes *ARHGEF4* and *ARHGEF6* (chromosomes 2 and X; members of the rho guanosine triphosphate family), participate in numerous cellular processes mediated by G protein coupled receptors that participate in the regulation of the actin cytoskeleton (Chi et al., 2013; GeneCards, 2018). The *SPATA13* gene (chromosome 12; codes for spermatogenesis-associated protein 13) regulates cell migration and adhesion through the involvement of the Rho family of GTPases (Bristow et al., 2009; GeneCards, 2018). The encoded protein of the *FGD1* gene (chromosome X; codes for faciogenital dysplasia 1 protein) contributes to the regulation of the actin cytoskeleton and cell shape (Hou et al., 2003; GeneCards, 2018). The proteins encoded by the *PIP4K2A* (chromosome 13; codes for phosphatidylinositol-5-phosphate 4-kinase type 2 alpha) and *PIP4K2B* genes (chromosome 19; codes for phosphatidylinositol-5-phosphate 4-kinase type 2 beta) help regulate cell proliferation, differentiation, and motility (GeneCards, 2018).

3.2.2. Organismal system pathways

The organismal systems category included three pathways (oxytocin signaling, axon guidance, B cell receptor signaling; Table 4; Fig. 1; Supplemental Table 1). The three organismal system pathways included 72 genes that explained an average of 0.39% of the genetic variance for VOL, NS, and MOT (Table 4). The axon guidance pathway included more genes (n = 30) than the oxytocin signaling (n = 26), and the B cell receptor signaling pathways

281 (n = 16). Accordingly, the axon guidance pathway explained a larger percentage of the genetic
282 variance for VOL, NS, and MOT (0.19%) than the percentage explained by the oxytocin
283 signaling pathway (0.12%) and the B cell receptor signaling pathway (0.07%). Among the 72
284 genes included in the organismal system category, 52% of them were involved in regulating
285 oxytocin release, calcium channel regulation, neural development, immune system, and the
286 remaining 48% were associated with angiogenesis, cellular processes, and signal transduction
287 functions influencing VOL, NS, and MOT (Table 4; Supplemental Table 1).
288
289 *3.2.2.1. Oxytocin signaling pathway*
290 Oxytocin signaling pathway genes *CD38, CACNA1C, CACNA2D3, CACNB2,*
291 *CACNB4, CACNG2, CACNG4, CACNG5, ITPR1, RYR1, RYR2, RYR3, KCNJ12,* and *EEF2K*
292 were reported to regulate oxytocin release and calcium channel activity that influence sperm
293 motility and sperm capacitation in humans and mice (Chiarella et al., 2004; Catterall, 2011;
294 Kiss et al., 2011; Mannowetz et al., 2017; Table 4; Fig. 1; Supplemental Table 1). In addition,
295 the oxytocin signaling pathway has a role in the contraction of the reproductive tract during
296 sperm release (Thackare et al., 2006). Proteins coded by genes *CACNA1C, CACNA2D3,*
297 *CACNB2, CACNB4, CACNG2, CACNG4, CACNG5, KCNJ12,* and *EEF2K* (chromosome 5,
298 22, 13, 2, 5, 19, 19, 19, 25; members of voltage-gated calcium channel families) are essential
299 for cellular signal transduction needed to initiate a variety of physiological actions including
300 the regulation of sperm ion channel activity and calcium ion binding that is associated with
301 sperm motility and fertility (Catterall, 2011; Mannowetz et al., 2017; GeneCards, 2018). Gene
302 *ITPR1* (chromosome 22; codes for intracellular inositol 1,4,5-trisphosphate receptor type 1)
303 and genes *RYR1, RYR2,* and *RYR3* (chromosomes 18, 28, and 10; code for ryanodine receptors
304 1, 2, and 3) mediate the release of Ca^{++} from intracellular stores including the endoplasmic
305 reticulum (Lanner et al., 2010; GeneCards, 2018). One of these genes (*RYR1*) is known to be

306 expressed in spermatogonia, pachytene spermatocytes, and round spermatids (Chiarella et al.,
307 2004). Furthermore, the *CD38* gene (chromosome 6) codes for adenosine monophosphate
308 ribosyl cyclase 1 which participates in the synthesis of the Ca^{++} second messengers and in the
309 regulation of oxytocin secretion (Kiss et al., 2011; Rah and Kim, 2015).
310
311 *3.2.2.2. Axon guidance pathway*
312 Axon guidance pathway genes *NTN1, NTN4, NTNG1, DPYSL2, ROBO2, SEMA3C,*
313 *SEMA3E, SEMA5B, SRGAP1,* and *SRGAP3* are involved in axon growth and migration
314 relevant to male and female fertility (Sun et al., 2011; Newquist et al., 2013; Martins-de-Souza
315 et al., 2015; GeneCards, 2018; Table 4; Fig. 1; Supplemental Table 1). The encoded protein
316 from genes *NTN1, NTN4, and NTNG1* (chromosome 19, 5, and 3; code for netrin 1, 4 and G1
317 proteins) contributes to axon migration during neural development (Sun et al., 2011;
318 GeneCards, 2018). In particular, Netrin 1 stimulates the growth of GnRH axons and dendrites
319 through intracellular and extracellular Ca^{++} (Low et al., 2012; GeneCards, 2018). Similarly,
320 proteins from genes *DPYSL2* (chromosome 8; codes for dihydropyrimidinase-related protein
321 2), *ROBO2* (chromosome 1; codes for roundabout guidance receptor 2), and *SEMA3C,*
322 *SEMA3E,* and *SEMA5B* (Chromosome 4, 4, and 1; code for semaphorin 3C, 3E, and 5B)
323 regulate axon growth and cell migration during the development of the nervous system
324 (GeneCards, 2018). Conversely, proteins from genes *SRGAP1* (chromosome 5; codes for rho
325 GTPase-activating protein 13) and S*RGAP3* (chromosome 12; codes for rho GTPase-activating
326 protein 14) function as activators of GTPases that downregulate neuronal migration
327 (GeneCards, 2018).
328
329 *3.2.2.3. B cell receptor signaling pathway*

The B cell receptor signaling pathway genes *MALT1*, *NFATC2*, *FCGR2B*, and *CR2* are important for the immune response (Rao et al., 1997; GeneCards, 2018). Gene MALT1 (chromosome 24; codes for mucosa-associated lymphoid tissue lymphoma translocation protein 1 paracaspase) and has a role in the activation of transcription factor NF-KappaB, which affects the expression of inflammatory genes such as cytokines and chemokines (Lawrence, 2009). Gene *NFATC2* (chromosome 13; codes for the nuclear factor of activated T cells 2) belongs to a group of NFAT, which stimulates transcription of cytokine genes during immune response (Rao et al., 1997; GeneCards, 2018). The protein encoded by gene *FCGR2B* (chromosome 3; codes for the Fc fragment of the IgG receptor IIb) is involved in phagocytosis of immune complexes and lowering the production of B-cell antibodies, whereas gene *CR2* (chromosome 26) encodes complement C3d Receptor 2 that participates in B lymphocyte activation and maturation (GeneCards, 2018).

3.2.3. Environmental information processing pathways

The environmental information processing category included two pathways (Rap1 signaling and sphingolipid signaling; Table 4; Fig. 1; Supplemental Table 1). These two pathways include 57 genes and explain an average of 0.23% of the genetic variance for VOL, NS, and MOT. The Rap1 signaling pathway contained a greater number of genes (n = 33) and explained a larger percentage of the genetic variance for the three semen traits (0.15%) than the sphingolipid signaling pathway (24 genes and 0.07%). There were 70% of the genes in the environmental information processing category that were involved in cell signaling, cell differentiation, cell survival, spermatogenesis, sperm maturation, activation, and capacitation, and 30% of these genes were involved in other functions (embryonic development, immune system, and angiogenesis; Table 4; Fig. 1; Supplemental Table 1).

355 *3.2.3.1. Rap1 signaling pathway*

356 Genes *RAC1, PRKCI, PRKCZ,* and *PRKCB* in the Rap1 signaling pathway are involved
357 in cell signalling, motility, differentiation, survival, and apoptosis during spermatogenesis and
358 fertilization (Garbi et al., 2000; Takashima et al., 2001; Lui et al., 2003; Ickowicz et al., 2012;
359 Hering et al., 2014a; GeneCards, 2018; Table 4; Fig. 1; Supplemental Table 1). Importantly,
360 the *RAC1* gene (chromosome 25; codes for Rac family small GTPase 1) was involved in six
361 pathways (focal adhesion, actin cytoskeleton regulation, axon guidance, B cell receptor
362 signaling, rap1 signaling, and sphingolipid signaling). The Rac GTPase 1 is essential for the
363 regulation of transmigration and proliferation of spermatogonial stem cells, cell movement,
364 and cell adhesion during spermatogenesis (Takashima et al., 2001; Lui et al., 2003). Similarly,
365 genes *PRKC*I (chromosome 1; codes for protein kinase C lota), *PRKCZ* (chromosome 16;
366 codes for protein kinase C zeta), and *PRKCB* (chromosome 25; codes for protein kinase C beta)
367 were also involved in six pathways (focal adhesion, oxytocin signaling, axon guidance, B cell
368 receptor signaling, Rap1 signaling, and sphingolipid signaling). These genes are members of
369 the protein kinase C family and have essential roles in signal transduction mechanisms involved
370 in multiple cellular processes that are essential for sperm capacitation and sperm motility
371 (Garbi et al., 2000; Ickowicz et al., 2012; Hering et al., 2014a; GeneCards, 2018).

372

373 *3.2.3.2. Sphingolipid signaling pathway*

374 Sphingolipid signaling genes *CERS3, CERS5, PLD1*, *PLD2, PRKCE, PPP2R5A,*
375 *SGMS1, NFKB1, PLCB1, FYN, MAP3K5* and *MAPK10* contribute to multiple cellular
376 processes including cytoskeletal organization, cell cycle, cell transcription, apoptosis, division,
377 growth, regulation of polyunsaturated fatty acids, immunity, inflammation, and heat shock
378 response (Liscovitch et al., 2000; Pentikainen et al., 2002; Coaxum et al., 2003; Mizutani et
379 al., 2006; Wathes et al., 2007; Li et al., 2009; Luo et al., 2012; Wittmann et al., 2016;

GeneCards, 2018; Table 4; Fig. 1; Supplemental Table 1). Briefly, gene *CERS3* (chromosome 21; codes for ceramide synthase 3) modifies lipid structures required for spermatogenesis (Mizutani et al., 2006; GeneCards, 2018). Genes *CERS5* (chromosome 5; codes for ceramide synthase 5) and *PRKCE* (chromosome 11; codes for protein kinase C epsilon) are involved in cellular signaling pathways regulating cell cycle, differentiation, heat shock response, senescence, and apoptosis (Ruvolo, 2001; GeneCards, 2018). Genes *PLD1* and *PLD2* (chromosomes 1 and 19) code for phospholipase D1 and D2 that participate in signal transduction, cytoskeletal dynamics, and membrane vesicle trafficking (Liscovitch et al., 2000; GeneCards, 2018), whereas gene *PPP2R5A* (chromosome 16) codes for protein phosphatase 2 regulatory subunit B alpha that downregulates cell division and cell growth (GeneCards, 2018). Gene *SGMS1* (chromosome 26) encodes sphingomyelin synthase 1, an enzyme essential for the regulation and maintenance of polyunsaturated fatty acids (PUFA) in the testes, which have a major effect on male fertility (Wittmann et al., 2016). A high PUFA content in the plasma membrane of spermatozoa is important for sperm development and flow during fertilization (Wathes et al., 2007). The *NFKB1* gene (chromosome 6; codes for nuclear factor kappa b subunit 1) was involved in two pathways (B cell receptor signaling and sphingolipid signaling). The *NFKB1* protein is essential for signal transduction associated with cell differentiation, growth, and apoptosis in male germ cells (Pentikainen et al., 2002; GeneCards, 2018). The *PLCB1* gene (chromosome 13; codes for phospholipase c beta 1) was present in three pathways (oxytocin signaling, rap1 signaling, and sphingolipid signaling). The *PLCB1* protein participates in intracellular transduction of numerous extracellular signals including neurotransmitters and hormones regulating functions of the central nervous system (Caricasole et al., 2000; GeneCards, 2018). The *FYN* gene (chromosome 9; codes for tyrosine-protein kinase Fyn) contributed to three pathways (focal adhesion, axon guidance, and sphingolipid signaling). Tyrosine-protein kinase Fyn participates in the regulation of cell growth, sperm

head shaping, and acrosome reaction (Luo et al., 2012). Genes *MAPK10* (chromosome 6) and *MAP3K5* (chromosome 9) are members of the mitogen-activated protein kinase family (MAPK) and were part of two pathways (focal adhesion and sphingolipid signaling). The MAPK proteins regulate cell adhesion, proliferation, differentiation, survival, and death, all of which are important for spermatogenesis, sperm capacitation, and acrosome reaction during fertilization (Wong and Cheng, 2005; Almog and Naor, 2008; Li et al., 2009).

3.3. Final remarks

The 127 genes found to be relevant for VOL, NS, MOT based on percentage of genetic variance explained for each trait (minimum 0.001%) and proximity of SNP to a gene in the NCBI database (inside or within 2,500 bp of genes) and the proteins encoded by these genes are known to affect cell structure, motility, migration, proliferation, differentiation, survival, apoptosis, signal transduction, calcium channel permeability, oxytocin release, neural development, and immune system performance. These proteins encoded by these genes have also been found to influence sperm morphology, sperm motility, transport, capacitation, and viability. The morphological and physiological effects of the proteins encoded by these 127 genes during spermatogenesis translated into measurable contributions to the genetic variances for VOL, NS, and MOT in the Thai dairy multibreed population. Although the percent of the genetic variance accounted for by the 127 genes was small (1.04%), these genes will be useful for the construction of genotyping chips that contain biologically relevant SNP markers. The small fraction of the genetic variance explained by this set of genes confirmed the quantitative nature of VOL, NS, and MOT, indicating that perhaps over 20,000 SNP markers representing biologically relevant genes will be needed to fully account for the genetic variance of these three traits in the Thai dairy multibreed population. This would likely require genotyping animals with a substantially denser genotyping chip such as the Illumina BovineHD 777K chip

if the restrictions of a 0.001% minimum genetic variance explained by each SNP marker and a distance of 2,500 bp or less from an NCBI gene were kept. A less expensive option would be to relax these two restrictions and utilize a somewhat denser chip than the one used here, such as GeneSeek GGPHD 150K or 250K chip.

It should be emphasized that the genes identified in the present study were determined with the specific information on phenotypes, pedigree, genotypes used in this research. Thus, the set of identified genes may change with additional information from new sires, genotyping chips of greater density, and restrictions utilized in future research in this population. Furthermore, the genetic variance of biologically relevant genes for VOL, NS, and MOT may also vary depending on the breed composition and linkage disequilibrium in the population as well as environmental conditions, thus studies such as the present one need to be conducted in each dairy population to determine similarities and differences across populations. Results from these studies will be particularly applicable to Holstein sires used in multiple countries with tropical and temperate environmental conditions.

4. Conclusions

Biological pathways analysis indicated that VOL, NS, and MOT were affected by nearly 200 genes involved in cellular process, organismal system, and environmental information processing pathways. These genes were known to have a role in cell structure, motility, migration, proliferation, differentiation, survival, apoptosis, signal transduction, oxytocin release, calcium channel, neural development, and immune system functions affecting sperm morphology and physiology during spermatogenesis. The percentage of the genetic variance explained by these genes for each trait was low, confirming the quantitative nature of VOL, NS, and MOT. Consequently, future research would need to be conducted to either utilize greater-density genotyping chips, relaxing the restrictions on SNP markers used to

identify genes in the NCBI database, or both. Genes identified as a result of the present research will be useful for constructing genotyping chips with biologically relevant SNP markers for genomic selection tailored to the Thai dairy cattle that are maintained in tropical environmental conditions.

Conflict of interests

The authors declare that they have no conflicts of interest.

Acknowledgements

The authors thank the Royal Golden Jubilee Ph.D. Program of the Thailand Research Fund for awarding the scholarship to first author (Grant No. PHD/0090/2559), the University of Florida for supporting the training of the first author as research scholar, Kasetsart University [Project No. SK(KS)1.58] and the Dairy Farming Promotion Organization of Thailand for supporting tissue sampling and genomic SNP genotype acquisition, and the Thai dairy farmers for their participation and support of this investigation.

References

Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S., Lawlor, T.J., 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93, 743-752.

Almog, T., Naor, Z., 2008. Mitogen activated protein kinases (MAPKs) as regulators of spermatogenesis and spermatozoa functions. Mol. Cell. Endocrinol. 282, 39-44.

Aoki, K., Taketo, M.M., 2007. Adenomatous polyposis coli (APC): a multi-functional tumor suppressor gene. J. Cell Sci. 120, 3327-3335.

479 Bienz, M., Hamada, F., 2004. Adenomatous polyposis coli proteins and cell adhesion. Curr.
480 Opin. Cell Biol. 16, 528-535.
481 Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman,
482 W.H., Pages, F., Trajanoski, Z., Galon, J., 2009. ClueGO: a Cytoscape plug-in to
483 decipher functionally grouped gene ontology and pathway annotation networks.
484 Bioinformatics 25, 1091-1093.
485 Breitbart, H., Naor, Z., 1999. Protein kinases in mammalian sperm capacitation and the
486 acrosome reaction. Rev. Reprod. 4, 151-159.
487 Bristow, J.M., Sellers, M.H., Majumdar, D., Anderson, B., Hu, L., Webb, D.J., 2009. The
488 rho-family GEF Asef2 activates Rac to modulate adhesion and actin dynamics and
489 thereby regulate cell migration. J. Cell Sci. 122, 4535-4546.
490 Caricasole, A., Sala, C., Roncarati, R., Formenti, E., Terstappen, G.C., 2000. Cloning and
491 characterization of the human phosphoinositide-specific phospholipase C-beta 1 (PLC
492 beta 1). Biochim. Biophys. Acta. 1517, 63-72.
493 Catterall, W.A., 2011. Voltage-gated calcium channels. Cold Spring Harb. Perspect. Biol. 3,
494 a003947.
495 Chi, X., Wang, S., Huang, Y., Stamnes, M., Chen, J.L., 2013. Roles of rho GTPases in
496 intracellular transport and cellular transformation. Int. J. Mol. Sci. 14, 7089-7108.
497 Chiarella, P., Puglisi, R., Sorrentino, V., Boitani, C., Stefanini, M., 2004. Ryanodine
498 receptors are expressed and functionally active in mouse spermatogenic cells and their
499 inhibition interferes with spermatogonial differentiation. J. Cell Sci. 117, 4127-4134.
500 Coaxum, S.D., Martin, J.L., Mestril, R., 2003. Overexpression of heat shock proteins
501 differentially modulates protein kinase C expression in rat neonatal cardiomyocytes.
502 Cell Stress Chaperones 8, 297-302.

503 Druet, T., Fritz, S., Sellem, E., Basso, B., Gérard, O., Salas-Cortes, L., Humblot, P., Druart,
504 X., Eggen, A., 2009. Estimation of genetic parameters and genome scan for 15 semen
505 characteristics traits of Holstein bulls. J. Anim. Breed. Genet. 126, 269-277.
506 Du Plessis, S.S., Franken, D.R., Baldi, E., Luconi, M., 2004. Phosphatidylinositol 3-kinase
507 inhibition enhances human sperm motility and sperm-zona pellucida binding. Int. J.
508 Androl. 27, 19-26.
509 Fisher, H.M., Brewis, I.A., Barratt, C.L.R, Cooke, I.D., Moore, H.D.M., 1998.
510 Phosphoinositide 3-kinase is involved in the induction of the human sperm acrosome
511 reaction downstream of tyrosine phosphorylation. Mol. Hum. Reprod. 4, 849-855.
512 Fortes, M. R., Reverter, A., Hawken, R.J., Bolormaa, S., Lehnert, S.A., 2012. Candidate
513 genes associated with testicular development, sperm quality, and hormone levels of
514 inhibin, luteinizing hormone, and insulin-like growth factor 1 in Brahman bulls. Biol.
515 Reprod. 58, 1-8.
516 Fortes, M. R., Reverter, A., Kelly, M., McCulloch, R., Lehert, A.S., 2013. Genome-wide
517 association study for inhibin, luteinizing hormone, insulin-like growth factor 1,
518 testicular size and semen traits in bovine species. Andrology 1, 644-650.
519 Fruman, D.A., Meyers, R.E., Cantley, L.C., 1998. Phosphoinositide kinases. Annu. Rev.
520 Biochem. 67, 481-507.
521 Gallardo Bolaños, J.M., Balao da Silva, C.M., Martin Muñoz, P., Morillo Rodríguez, A.,
522 Plaza Dávila, M., Rodríguez-Martínez, H., Aparicio, I.M., Tapia, J.A., Ortega
523 Ferrusola, C., Peña, F.J., 2014. Phosphorylated AKT preserves stallion sperm viability
524 and motility by inhibiting caspases 3 and 7. Reproduction 148, 221-235.
525 Garbi, M., Rubinstein, S., Lax, Y., Breitbart, H., 2000. Activation of protein kinase Cα in the
526 lysophosphatidic acid-induced bovine sperm acrosome reaction and phospholipase D1
527 regulation. Biol. Reprod. 63, 1271-1277.

GeneCards. 2018. The human gene database. (Accessed 01 March 2018). https://www.genecards.org/.

Gungor-Ordueri, N.E., Mruk, D.D., Wan, H.T., Wong, E.W.P., Celik-Ozenci, C., Lie, P.P.Y., Cheng, C.Y., 2014. New insights into FAK function and regulation during spermatogenesis. Histol. Histopathol. 29, 977-989.

Hanna, L.L.H., Riley, D.G., 2014. Mapping genomic markers to closets feature using the R package Map2NCBI. Livest. Sci. 162, 59-65.

Hering, D.M., Olenski, K., Kaminski, S., 2014a. Genome-wide association study for poor sperm motility in Holstein-Friesian bulls. Anim. Reprod. Sci. 146, 89-97.

Hering, D.M., Olenski, K., Rusc, A., Kaminski, S., 2014b. Genome-wide association study for semen volume and total number of sperm in Holstein-Friesian bulls. Anim. Reprod. Sci. 151, 126-130.

Holm, S., 1979. A simple sequential rejective multiple test procedure. Scand. J. Stat. 6, 65-70.

Hou, P., Estrada, L., Kinley, A.W., Parsons, J.T., Vojtek, A.B., Gorski, J.L., 2003. Fgd1, the Cdc42 GEF responsible for faciogenital dysplasia, directly interacts with cortactin and mAbp1 to modulate cell shape. Hum. Mol. Genet. 12, 1981-1993.

Ickowicz, D., Finkelstein, M., Breitbart, H., 2012. Mechanism of sperm capacitation and the acrosome reaction: role of protein kinases. Asian J. Androl. 14, 816-821.

Jattawa, D., Elzo, M.A., Koonawootrittriron, S., Suwanasopee, T., 2016. Imputation accuracy from low to moderate density single nucleotide polymorphism chips in a Thai multibreed dairy cattle population. Asian-Australasian J. Anim. Sci. 29, 464-470.

Kaminski, S., Hering, D.M., Olenski, K., Lecewicz, M., Kordan, W., 2016. Genome-wide association study for sperm membrane integrity in frozen-thawed semen of Holstein-Friesian bulls. Anim. Reprod. Sci. 170, 135-140.

553 Kapš, M., Posavi, M., Stipic, N., Mikulic, B., 2000. Genetic evaluation of semen and growth
554 traits of young Simmental bulls in performance test. Agric. Conspec. Scientif. 65, 15-
555 20.
556 Karoui, S., Díaz, C., Serrano, M., Cue, R., Celorrio, I., Carabaño, M.J., 2011. Time trends,
557 environmental factors and genetic basis of semen traits collected in Holstein bulls
558 under commercial conditions. Anim. Reprod. Sci. 124, 28-38.
559 Kiss, I., Levy-Gigi, E., Keri, S., 2011. CD 38 expression, attachment style and habituation of
560 arousal in relation to trust-related oxytocin release. Biol. Psychol. 88, 223-226.
561 Koonawootrittriron, S., Elzo, M.A., Thongprapi, T., 2009. Genetic trends in a Holstein x
562 other breeds multibreed dairy population in Central Thailand. Livest. Sci. 122, 186-
563 192.
564 Kordi-Tamandani, D.M., Mir, A., 2015. Relationship between phosphoinositide-3-kinase
565 genetic polymorphism and schizophrenia. Nord. J. Psychiatry 70, 272-275.
566 Lanner, J.T., Georgiou, D.K., Joshi, A.D., Hamilton, S.L., 2010. Ryanodine receptors:
567 structure, expression, molecular details, and function in calcium release. Cold Spring
568 Harb. Perspect. Biol. 2, a003996.
569 Lawrence, T., 2009. The nuclear factor NF-kB pathway inflammation. Cold Spring Harb.
570 Perspect. Biol. 1, a001651.
571 Li, M.W.W., Mruk, D.D., Cheng, C.Y., 2009. Mitogen-activated protein kinases in male
572 reproductive function. Trends Mol. Med. 15,159-168.
573 Lie, P. P.Y., Mruk, D. D., Lee, W. M., Cheng, C.Y., 2010. Cytoskeletal dynamics and
574 spermatogenesis. Phil. Trans. R. Soc. B. 365, 1581-1592.
575 Liscovitch, M., Czarny, M., Fiucci, G., Tang, X., 2000. Phospholipase D: molecular and cell
576 biology of a novel gene family. Biochem. J. 345, 401-415.

577 Low, V.F., Fiorini, Z., Fisher, L., Jasoni, C.L., 2012. Netrin-1 stimulates developing GnRH
578 neurons to extend neuritis to the median eminence in a calcium-dependent manner
579 PLoS One 7, e46999.
580 Lui, W.Y., Lee, W.M., Cheng, C.Y., 2003. Rho GTPases and spermatogenesis. Biochim.
581 Biophys. Acta 1593, 121-129.
582 Luo, J., Gupta, V., Kern, B., Tash, J.S., Sanchez, G., Blanco, G., Kinsey, W.H., 2012. Role of
583 FYN kinase in spermatogenesis: defects characteristic of Fyn-null sperm in mice.
584 Biol. Reprod. 86, 1-8.
585 Mannowetz, N., Miller, M.R., Lishko, P.V., 2017. Regulation of the sperm calcium channel
586 CatSper by endogenous steroids and plant triterpenoids. Proc. Natl. Acad. Sci. U.S.A.
587 114, 5743-5748.
588 Martins-de-Souza, D., Cassoli, J.S., Nascimento, J.M., Hensley, K., Guest, P.C. Pinzon-
589 Velasco, A.M., Turck, C.W., 2015. The protein interactome of collapsin response
590 mediator protein-2 (CRMP2/DPYSL2) reveals novel partner proteins in brain tissue.
591 Proteomics Clin. Appl. 9, 817-831.
592 Mathevon, M., Buhr, M.M., Dekkers, J.C.M., 1998. Environmental, management, and
593 genetic factors affecting semen production in Holstein bulls. J. Dairy Sci. 81, 3321-
594 3330.
595 Misztal, I., Tsuruta, S., Lourenco, D., Masuda, Y., Aguilar, I., Legarra, A., Vitezica, Z., 2018.
596 Manual for BLUPF90 family of programs. Univ. Georgia, Athens, GA, p 1-142.
597 http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all6.pdf.
598 Mizutani, Y., Kihara, A., Igarashi, Y., 2006. LASS3 (longevity assurance homologue 3) is a
599 mainly testis-specific (dihydro) ceramide synthase with relatively broad substrate
600 specificity. Biochem. J. 398, 531-538.

601 Newquist, G., Hogan, J., Walker, K., Lamanuzzi, M., Bowser, M., Kidd, T., 2013. Control of
602 male and female fertility by the netrin axon guidance genes. PLoS One 8, e72524.
603 Nishimura, K., Honda, T., Oyama, K., 2010. Genetic variability of semen characteristics in
604 Japanese Black bulls. Anim. Genet. 38, 73-76.
605 Pentikainen, V., Suomalainen, L., Erkkila, K., Martelin, E., Parvinen, M., Pentikainen, M.O.,
606 Dunkel, L., 2002. Nuclear factor-kappa B activation in human testicular apoptosis.
607 Am. J. Pathol. 160, 205-218.
608 Peñagaricano, F., Weigel, K.A., Rosa, G.J.M., Khatib, H., 2013. Inferring quantitative trait pathway
609 associated with bull fertility from a genome-wide association study. Front Genet. 3, 307.
610 Puglisi, R., Gaspa, G., Balduzzi, D., Severgnini, A., Vanni, R., Macciotta, N.P.P., Galli, A.,
611 2016. Genome wide analysis of bull sperm quality and fertility traits. Reprod. Dom.
612 Anim. 51, 840-843.
613 Qian, X., Mruk, D.D., Cheng, Y.H., Tang, E.I., Han, D., Lee, W.M., Wong, E.W.P., Cheng,
614 C.Y., 2014. Actin binding proteins, spermatid transport and spermiation. Semin. Cell
615 Dev. Biol. 30, 75-85.
616 Qin, C., Yin, H., Zhang, X., Sun, D., Zhang, Q., Liu, J., Ding, X., Zhang, Y., Zhang, S.,
617 2016. Genome-wide association study for semen traits of the bulls in Chinese
618 Holstein. Anim. Genet. 48, 80-84.
619 Rah, S.Y., Kim, U.H., 2015. CD38-mediated Ca^{2+} signaling contributes to glucagon-induced
620 hepatic gluconeogenesis. Sci. Rep. 5, 10741.
621 Rao, A., Luo, C., Hogan, P.G., 1997. Transcription factors of the NFAT family: Regulation
622 and function. Ann. Rev. Immunol. 15, 707-747.
623 Ritt, M., Guan, J.L., Sivaramakrishnan, S., 2013. Visualizing and manipulating focal
624 adhesion kinase regulation in live cells. J. Biol. Chem. 288, 8875–8886.

625 Roa-Espitia, A.L., Hernández-Rendón, E.R., Baltiérrez-Hoyos, R., Muñoz-Gotera, R.J., Cote-
626 Vélez, A., Jiménez, I., González-Márquez, H., Hernández-González, E.O., 2016.
627 Focal adhesion kinase is required for actin polymerization and remodeling of the
628 cytoskeleton during sperm capacitation. Biol. Open 5, 1189-1199.
629 Ruvolo, P.P., 2001. Ceramide regulates cellular homeostasis via diverse stress signaling
630 pathways. Leukemia 15, 1153-1160.
631 Sarakul, M., Elzo, M.A., Koonawootrittriron, S., Suwanasopee, T., Jattawa, D., 2018. Genetic
632 parameters, predictions, and rankings for semen production traits in a Thailand multi-
633 breed dairy population using genomic-polygenic and polygenic models. Anim.
634 Reprod. Sci. 195, 71-79.
635 Sargolzaei, M., Chesnais, J.P., Schenkel, F.S., 2014. A new approach for efficient genotype
636 imputation using information from relatives. BMC Genomics 15, 478.
637 Suchocki, T., Szyda, J., 2015. Genome-wide association study for semen production traits in
638 Holstein-Friesian bulls. J. Dairy Sci. 98, 5774-5780.
639 Sun, K.L.W., Correia, J.P., Kennedy, T. E., 2011. Netrins: versatile extracellular cues with
640 diverse functions. Development 138, 2153-2169.
641 Takashima, S., Kanatsu-Shinohara, M., Tanaka, T., Takehashi, M., Morimoto, H., Shinohara,
642 T., 2011. Rac mediates mouse spermatogonial stem cell homing to germline niches by
643 regulating transmigration through the blood-testis barrier. Cell Stem Cell 9, 463-475.
644 Thackare, H., Nicholson, H.D., Whittington, K., 2006. Oxytocin-its role in male reproduction
645 and new potential therapeutic uses. Hum. Reprod. Update. 12, 437-448.
646 Tsuruta, S., 2014. Average Information REML with several options including EM-REML
647 and heterogeneous residual variances. (Accessed 1 November 2016).
648 http://nce.ads.uga.edu/wiki/doku.php?id=application_programs.

649 Wang, H., Misztal, I., Aguilar, I., Legarra, A., Muir, W.M., 2012. Genome-wide association
650 mapping including phenotypes from relatives without genotypes. Genet. Res. Camb.
651 94, 73-83.
652 Wathes, D.C., Abayasekara, D.R.E, Aitken, R.J., 2007. Polyunsaturated fatty acids in male
653 and female reproduction. Biol. Reprod. 77, 190-201.
654 Wittmann, A., Grimm, M.O.W., Scherthan, H., Horsch, M., Beckers, J., Fuchs, H., Gailus-
655 Durner, V., Hrabě de Angelis, M., Ford, S.J., Burton, N.C., Razansky, D., Trümbach,
656 D., Aichler, M., Walch, A.K., Calzada-Wack, J., Neff, F., Wurst, W., Hartmann, T.,
657 Floss, T., 2016. Sphingomyelin synthase 1 is essential for male fertility in mice. PLoS
658 One 11, e0164298.
659 Wong, C.H., Cheng, C.Y., 2005. Mitogen-activated protein kinases, adherens junction
660 dynamics, and spermatogenesis: a review of recent data. Dev. Biol. 286, 1-15.
661

662 **Table 1**

663 Number of SNP for semen volume (VOL), number of sperm (NS), and sperm motility (MOT)
664 explaining at least 0.001% of the genetic variance classified by distance between SNP and
665 genes in the NCBI database

666

Distance between SNP and gene	VOL	NS	MOT
Inside gene	9,645	9,808	9,830
≤ 2,500 bp	1,307	1,277	1,317
Total	10,952	11,085	11,147

667

668 **Table 2**

669 Number of SNP and genetic variance (%) for semen volume (VOL), number of sperm (NS),
670 and sperm motility (MOT) accounting for at least 0.001% of the genetic variance located
671 inside genes or within 2500 bp of genes in the NCBI database

672

Chromosome	Number of SNP markers			Genetic variance (%)		
	VOL	NS	MOT	VOL	NS	MOT
1	513	559	524	1.59	1.85	1.78
2	417	472	476	1.17	1.40	1.40
3	498	530	537	1.58	1.57	1.64
4	529	523	528	1.55	1.58	1.53
5	572	554	548	1.80	1.81	1.59
6	404	452	484	1.18	1.56	1.60
7	500	398	432	1.53	1.15	1.19
8	452	442	388	1.48	1.34	1.07
9	387	369	385	1.30	1.24	1.23
10	529	503	479	1.67	1.68	1.50
11	470	427	469	1.46	1.21	1.53
12	348	303	318	1.29	1.01	1.04
13	399	383	412	1.35	1.09	1.33
14	339	304	338	1.08	0.89	1.14
15	430	412	418	1.19	1.22	1.28
16	284	431	406	0.84	1.35	1.20
17	307	321	371	1.03	0.96	1.20
18	368	370	329	1.23	1.05	1.01
19	555	513	577	3.44	3.08	3.25
20	239	253	256	0.79	0.85	0.87
21	235	240	234	0.72	0.71	0.70
22	196	232	242	0.58	0.81	0.76
23	218	227	215	0.59	0.66	0.63
24	200	205	193	0.60	0.72	0.67
25	226	214	217	0.61	0.73	0.65
26	239	245	235	0.81	0.93	0.76
27	138	178	169	0.43	0.60	0.54
28	264	283	275	0.82	0.96	0.90
29	289	316	292	0.86	1.07	0.95
X	407	426	400	2.95	3.71	3.47
Total	10,952	11,085	11,147	37.53	38.77	38.41

673
674

675 **Table 3**

676 Number of genes associated with semen volume (VOL), number of sperm (NS), and sperm
677 motility (MOT) per chromosome identified by SNP located inside genes or within 2500 bp of
678 genes in the NCBI database

679

Chromosome	One trait[1]			Two traits[2]			Three traits[3]
	VOL	NS	MOT	VOL-NS	VOL-MOT	NS-MOT	VOL-NS-MOT
1	266	289	261	161	136	190	96
2	223	237	251	135	123	162	87
3	284	302	295	160	153	195	101
4	201	203	217	108	111	142	72
5	338	315	310	197	171	192	109
6	182	208	210	107	114	151	77
7	274	228	257	134	133	140	72
8	216	212	199	130	117	138	85
9	160	176	188	95	97	129	68
10	258	238	239	146	140	156	87
11	246	227	245	122	126	154	75
12	115	114	111	61	69	73	43
13	215	201	221	116	121	123	65
14	138	134	146	71	78	94	51
15	240	231	234	128	113	144	74
16	155	219	199	106	85	150	70
17	158	169	202	86	98	122	58
18	249	272	216	145	101	145	60
19	330	315	320	209	204	228	141
20	96	89	94	58	50	55	33
21	119	135	138	60	60	90	35
22	102	110	119	57	54	79	37
23	126	130	119	74	64	84	45
24	95	105	98	61	48	70	38
25	144	141	141	79	74	92	45
26	124	120	121	79	68	83	51
27	56	71	73	31	24	47	16
28	104	107	110	61	59	84	46
29	149	152	142	79	66	93	41
X	213	217	218	159	149	166	121
Total	5,576	5,667	5,694	3,215	3,006	3,771	1,999

680 [1]Number of genes associated with a single trait

681 [2]Number of genes associated with two traits

682 [3]Number of genes associated with three traits

683 **Table 4**

684 Biological pathways involving genes associated with semen volume (VOL), number of sperm (NS) and sperm motility (MOT) in the Thai

685 multibreed dairy population

686

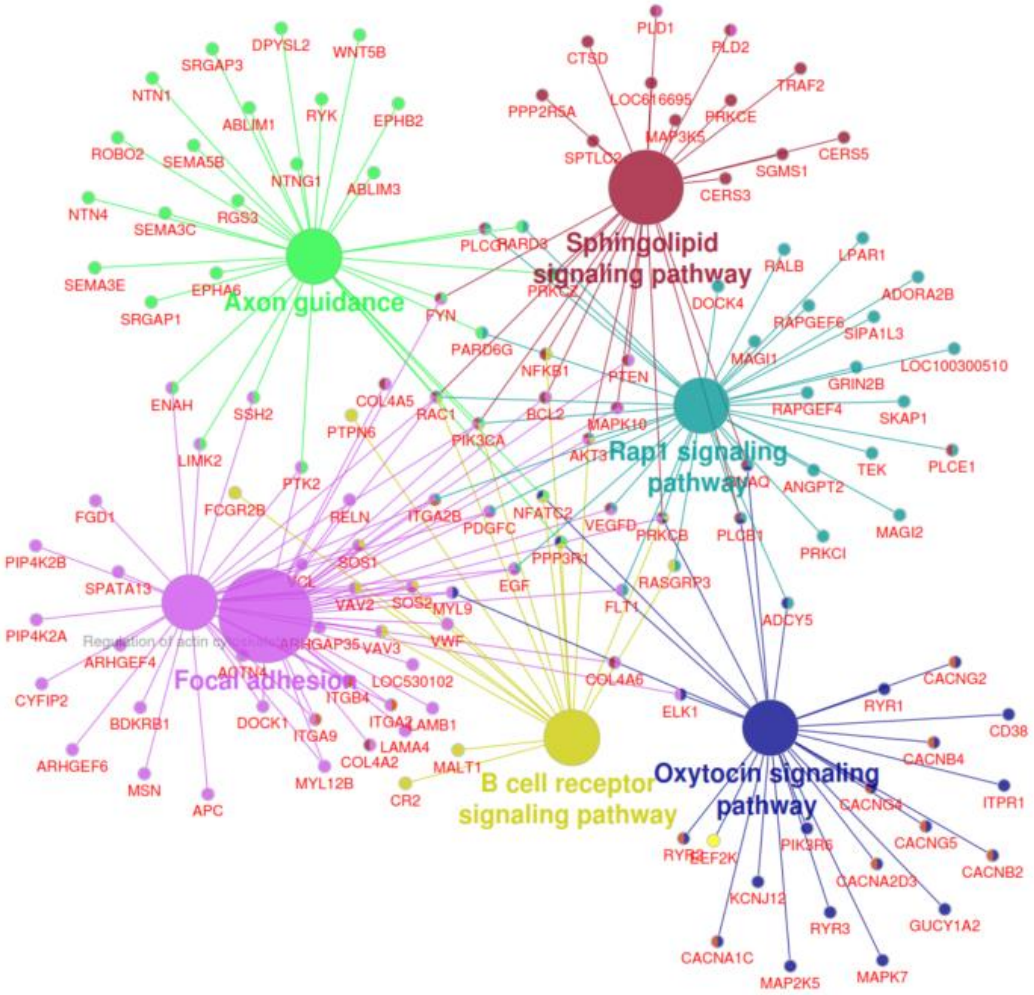
Category	Pathway[1]	*P*-Value[2]	Percent genetic variance[3]	No. genes	Gene symbol
Cellular processes			0.42	68	
	Focal adhesion	0.0002	0.21	36	*ACTN4, AKT3, ARHGAP35, BCL2, COL4A2, COL4A5, COL4A6, DOCK1, EGF, ELK1, FLT1, FYN, ITGA2, ITGA2B, ITGA9, ITGB4, LAMA4, LAMB1, LOC530102, MAPK10, MYL12B, MYL9, PDGFC, PIK3CA, PRKCB, PTEN, PTK2, RAC1, RELN, SOS1, SOS2, VAV2, VAV3, VCL, VEGFD, VWF*
	Actin cytoskeleton regulation	0.0310	0.20	32	*ACTN4, APC, ARHGAP35, ARHGEF4, ARHGEF6, BDKRB1, CYFIP2, DOCK1, EGF, ENAH, FGD1, ITGA2, ITGA2B, ITGA9, ITGB4, LIMK2, MSN, MYL12B, MYL9, PDGFC, PIK3CA, PIP4K2A, PIP4K2B, PTK2, RAC1, SOS1, SOS2, SPATA13, SSH2, VAV2, VAV3, VCL*
Organismal systems			0.39	72	
	Oxytocin signaling	0.0183	0.12	26	*ADCY5, CACNA1C, CACNA2D3, CACNB2, CACNB4, CACNG2, CACNG4, CACNG5, CD38, EEF2K, ELK1, GNAQ, GUCY1A2, ITPR1, KCNJ12, MAP2K5, MAPK7, MYL9, NFATC2, PIK3R6, PLCB1, PPP3R1, PRKCB, RYR1, RYR2, RYR3*

	Axon guidance	0.0067	0.20	30	*ABLIM1, ABLIM3, DPYSL2, ENAH, EPHA6, EPHB2, FYN, LIMK2, NFATC2, NTN1, NTN4, NTNG1, PARD3, PARD6G, PIK3CA, PLCG1, PPP3R1, PRKCZ, PTK2, RAC1, RGS3, ROBO2, RYK, SEMA3C, SEMA3E, SEMA5B, SRGAP1, SRGAP3, SSH2, WNT5B*
	B cell receptor signaling	0.0474	0.08	16	*AKT3, CR2, FCGR2B, MALT1, NFATC2, NFKB1, PIK3CA, PPP3R1, PRKCB, PTPN6, RAC1, RASGRP3, SOS1, SOS2, VAV2, VAV3*
Environmental Information Processing			0.23	57	
	Rap1 signaling	0.0137	0.15	33	*ADCY5, ADORA2B, AKT3, ANGPT2, DOCK4, EGF, FLT1, GNAQ, GRIN2B, ITGA2B, LOC100300510, LPAR1, MAGI1, MAGI2, PARD3, PARD6G, PDGFC, PIK3CA, PLCB1, PLCE1, PLCG1, PRKCB, PRKCI, PRKCZ, RAC1, RALB, RAPGEF4, RAPGEF6, RASGRP3, SIPA1L3, SKAP1, TEK, VEGFD*
	Sphingolipid signaling	0.0022	0.07	24	*AKT3, BCL2, CERS3, CERS5, CTSD, FYN, GNAQ, LOC616695, MAP3K5, MAPK10, NFKB1, PIK3CA, PLCB1, PLD1, PLD2, PPP2R5A, PRKCB, PRKCE, PRKCZ, PTEN, RAC1, SGMS1, SPTLC2, TRAF2*
All			**1.04**	**197**	

[1]Kyoto Encyclopedia of Genes and Genomes (KEGG) database

[2]Bonferroni step down

[3]Mean percent of genetic variance explained by category or pathway across the three semen traits (VOL, NS, and MOT)



690

691 **Fig. 1.** Biological pathways influencing semen traits identified by ClueGO using *Bos taurus*

692 genes in the KEGG database

693

694