Journal of Animal Science, 2021, Vol. 99, No. 7, 1–9

https://doi.org/10.1093/jas/skab188
Advance Access publication June 10, 2021
Received: 19 February 2021 and **Accepted:** 8 June 2021
Animal Genetics and Genomics

ANIMAL GENETICS AND GENOMICS

Identification of new semen trait-related candidate genes in Duroc boars through genome-wide association and weighted gene co-expression network analyses

Quanshun Mei,[†] Chuanke Fu,[†] Goutam Sahana,[‡] Yilong Chen,[†,‖] Lilin Yin,[†] Yuanxin Miao,[†,$] Shuhong Zhao,[†] and Tao Xiang[†,1]

[†]Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education & Key Laboratory of Swine Genetics and Breeding of Ministry of Agriculture, Huazhong Agricultural University, Wuhan 430070, China, [‡]Center for Quantitative Genetics and Genomics, Aarhus University, Tjele 8830, Denmark, [‖]Center of Breeding Production, Guangxi Yangxiang Agriculture and Husbandry Co., LTD, Guigang 537100, China, [$]School of Biological Engineering, Jingchu University of Technology, Jingmen 448000, China

[1]Corresponding author: Tao.Xiang@mail.hzau.edu.cn

Abstract

Semen traits are crucial in commercial pig production since semen from boars is widely used in artificial insemination for both purebred and crossbred pig production. Revealing the genetic architecture of semen traits potentially promotes the efficiencies of improving semen traits through artificial selection. This study is aimed to identify candidate genes related to the semen traits in Duroc boars. First, we identified the genes that were significantly associated with three semen traits, including sperm motility (**MO**), sperm concentration (**CON**), and semen volume (**VOL**) in a Duroc boar population through a genome-wide association study (**GWAS**). Second, we performed a weighted gene co-expression network analysis (**WGCNA**). A total of 2, 3, and 20 single-nucleotide polymorphisms were found to be significantly associated with MO, CON, and VOL, respectively. Based on the haplotype block analysis, we identified one genetic region associated with MO, which explained 6.15% of the genetic trait variance. *ENSSSCG00000018823* located within this region was considered as the candidate gene for regulating MO. Another genetic region explaining 1.95% of CON genetic variance was identified, and, in this region, *B9D2*, *PAFAH1B3*, *TMEM145*, and *CIC* were detected as the CON-related candidate genes. Two genetic regions that accounted for 2.23% and 2.48% of VOL genetic variance were identified, and, in these two regions, *WWC2*, *CDKN2AIP*, *ING2*, *TRAPPC11*, *STOX2*, and *PELO* were identified as VOL-related candidate genes. WGCNA analysis showed that, among these candidate genes, *B9D2*, *TMEM145*, *WWC2*, *CDKN2AIP*, *TRAPPC11*, and *PELO* were located within the most significant module eigengenes, confirming these candidate genes' role in regulating semen traits in Duroc boars. The identification of these candidate genes can help to better understand the genetic architecture of semen traits in boars. Our findings can be applied for semen traits improvement in Duroc boars.

Key words: Duroc pigs, semen traits

Abbreviations

AI	artificial insemination
CON	sperm concentration
DEBV	de-regressed estimated breeding values
DFI	DNA fragmentation index
GEBV	genomic estimated breeding values
GWAS	genome-wide association study
ING	inhibitor of growth
ME	module eigengenes
MLM	mixed linear model
MO	sperm motility
QTL	quantitative trait loci
REML	restricted maximum likelihood
snRNA	small nuclear RNA
SSGBLUP	single-step genomic best linear unbiased prediction
VOL	semen volume
WGCNA	weighted gene co-expression network analysis

Introduction

Semen traits play a key role in the pig industry since semen from boars is widely used for both purebred and crossbred pig production. With the widely use of artificial insemination (**AI**) in the pig production industry, there is an increasing interest in improving semen traits through genetic selection. One prerequisite for successful AI lies in the high quality of semen. Furthermore, the economic profit of AI boar station is highly dependent on the quantity and quality of semen (Robinson and Buhr, 2005), and the decreased quality of semen is one of the main reasons for shortening the longevity of boars (Koketsu and Sasaki, 2009). Involuntary replacement of boars results in an increased cost in pig production system (D'Allaire et al., 1992; Robinson and Buhr, 2005). In addition, the quantity and quality of semen from service boars also affect the reproductive performance of sows (McPherson et al., 2014; Myromslien et al., 2019).

With the fast development of high-throughput genotyping techniques, breeders utilize genomic selection to accelerate the genetic improvement of the target traits (Meuwissen et al., 2001). Revealing the genetic architecture of the target traits potentially enhances the efficiency of genomic selection of target trait. In the recent few years, many quantitative trait loci (**QTLs**) and candidate genes associated with the target traits have been identified via genome-wide association study (**GWAS**) in previous studies (Goddard and Hayes, 2009; Visscher et al., 2017). Candidate genes related to semen traits have also been identified in several previous studies through GWAS (Diniz et al., 2014; Marques et al., 2017, 2018; Gao et al., 2019; Zhao et al., 2020). However, the genetic architecture of semen traits remains largely unknown due to the complexity of the reproduction and maturation of sperm cells (Marques et al., 2018). Furthermore, previous studies have shown that mixed linear model (**MLM**) in GWAS is less powerful in detecting significantly associated markers (Lippert et al., 2011; Listgarten et al., 2012). The weak detection power might be due to the double fittings of candidate markers in the MLM with one fitting as a fixed effect for association test and the other one as a part of genomic relationship matrix. In order to overcome the above-mentioned detection weakness, in this study, we employed the MLM leaving-one-chromosome-out (MLMA-loco) algorithm (Yang et al., 2014) to perform GWAS, which excluded the effects of single-nucleotide polymorphisms (SNPs) on the chromosome where the candidate SNPs were located. This method has been confirmed to be more powerful than the standard MLM method (Yang et al., 2014).

In addition, the weighted gene co-expression network analysis (**WGCNA**) algorithm has been reported to play a key role in detecting the target trait-related candidate genes (Langfelder and Horvath, 2008). WGCNA algorithm can cluster the genes with similar functions into the same module based on the similarities of gene co-expression across different samples. To our knowledge, there have been few studies applying WGCNA in identifying genes significantly associated with semen traits.

In this study, we integrated GWAS and WGCNA to identify semen trait-related candidate genes. This study will provide an insight into the genetic architecture for semen traits, and our findings can be applied for accelerating the genetic progress of semen traits in boars.

Materials and Methods

Data

Phenotypic data were provided by a national pig nucleus herd in Guangxi Province, China. The data of three traits including sperm motility (**MO**), sperm concentration (**CON**), and semen volume (**VOL**) were collected from 2015 to 2018. MO and CON were evaluated using IMV TECHNOLOGIES IVOS II CASA system, and VOL was measured by electronic balance weighing. A total of 105,201 trait data were collected from 2,022 Duroc boars. Pedigrees could be traced back for about 10 generations, and the number of individuals in pedigree was 37,636. Descriptive statistics of the phenotypes are presented in Table 1. The average number of ejaculates per boar was 51.97, 52.10, and 52.10 for MO, CON, and VOL, respectively.

Genomic data were obtained from 1,411 boars by using GeneSeek Porcine 50K SNP chip. These boar tissues were collected since late 2016 after a genomic selection program was started. Quality controls of genomic data were as follows: individuals with call rate less than 90% were first excluded; SNPs with call rate less than 90% were removed as well; SNPs with minor allele frequency less than 0.05 were filtered; SNPs deviated strongly from Hardy–Weinberg equilibrium within breed ($P < 10^{-7}$) were also ruled out. After quality control, 1,411 individuals and 38,888 SNPs were retained. Missing SNP markers were imputed by using software Beagle 5.0 (Browning et al., 2018).

Statistical model

A multi-trait single-step genomic best linear unbiased prediction (**SSGBLUP**) method was used to estimate variance components and genomic estimated breeding values (**GEBV**) (Legarra et al., 2009; Christensen and Lund, 2010). The model was as follows :

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{Wpe} + \boldsymbol{e} \quad (1)$$

where $\boldsymbol{y}$ is the vector of phenotypic recordings; $\boldsymbol{b}$ is the vector of fixed effects including effects of section (physical

Table 1. Descriptive statistics of semen traits

Trait	Number of records	Mean	SD	Max	Min
MO, %	99,569	88.62	7.11	100.00	10.00
CON, 10^8/mL	99,878	4.71	2.20	45.24	0.01
VOL, mL	99,924	163.11	63.51	757.00	1.89

unit for animals during a certain period), year-month when the semen was collected; covariates of intervals between two successive semen collections (expressed as day); and ages of boar at the time of semen collection (expressed as month); $\boldsymbol{u}$ represents the random additive genetic effects of animals; $\boldsymbol{pe}$ is the vector of permanent environmental effects, and $\boldsymbol{e}$ is the vector of the random residual error. It was assumed that random effects followed normal distributions of $\boldsymbol{u} \sim N(0, \sum_u \otimes \boldsymbol{H})$, $\boldsymbol{pe} \sim N(0, \sum_{pe} \otimes \boldsymbol{I})$, and $\boldsymbol{e} \sim N(0, \sum_e \otimes \boldsymbol{I})$, where $\boldsymbol{H}$ is the combined pedigree-based and marker-based relationship matrix that was constructed as previously reported (Legarra et al., 2009; Christensen and Lund, 2010); $\boldsymbol{I}$ is the corresponding identity matrix; $\sum_u$ is the genetic (co)variance matrix; $\sum_{pe}$ is the permanent environmental (co)variance matrix; $\sum_e$ is the residual (co) variance matrix; and $\otimes$ denotes the Kronecker product. The (co)variance matrices were as follows:

$$\sum_u = \begin{bmatrix} \sigma^2_{u_1} & \sigma_{u_1u_2} & \sigma_{u_1u_3} \\ \sigma_{u_2u_1} & \sigma^2_{u_2} & \sigma_{u_2u_3} \\ \sigma_{u_3u_1} & \sigma_{u_3u_2} & \sigma^2_{u_3} \end{bmatrix}; \sum_{pe} = \begin{bmatrix} \sigma^2_{pe_1} & \sigma_{pe_1pe_2} & \sigma_{pe_1pe_3} \\ \sigma_{pe_2pe_1} & \sigma^2_{pe_2} & \sigma_{pe_2pe_3} \\ \sigma_{pe_3pe_1} & \sigma_{pe_3pe_2} & \sigma^2_{pe_3} \end{bmatrix};$$

$$\sum_e = \begin{bmatrix} \sigma^2_{e_1} & \sigma_{e_1e_2} & \sigma_{e_1e_3} \\ \sigma_{e_2e_1} & \sigma^2_{e_2} & \sigma_{e_2e_3} \\ \sigma_{e_3e_1} & \sigma_{e_3e_2} & \sigma^2_{e_3} \end{bmatrix}$$

where subscript 1 denotes MO; subscript 2 indicates CON; and subscript 3 represents VOL.

The above calculation was carried out using the restricted maximum likelihood (**REML**) algorithm in the software DMU (Madsen and Jensen, 2013). In order to remove the contribution of information from relatives, de-regressed estimated breeding values (**DEBVs**) were used as the response variable in GWAS analysis (Ostersen et al., 2011). DEBVs of genotyped boars were calculated by weighting EBVs (Garrick et al., 2009). The weighting factor (w_i) for individual i was calculated with the following formula:

$$w_i = \frac{1 - h^2}{(c + [(1 - r_i^2)/(r_i^2)])h^2}$$

where h^2 is the heritability of the trait; r_i^2 indicates the reliability of EBV of the individual i, and c refers to the proportion of genetic variation that could not be explained by the genetic information. In this study, c was assumed to be constant 0.5 (Sevillano et al., 2015). In terms of the reliability of EBVs, the SSGBLUP method outperformed traditional BLUP for non-genotyped animals and GBLUP for genotyped animals (Christensen et al., 2012).

After obtaining DEBVs of 1,411 genotyped boars, genome-wide association analysis of each semen trait was performed by using the software GCTA (Yang et al., 2011). The statistical model was as follows:

$$\boldsymbol{y} = \boldsymbol{Pf} + \boldsymbol{Xb} + \boldsymbol{Wg} + \boldsymbol{e}, \quad (2)$$

where $\boldsymbol{y}$ is the vector of DEBVs for semen trait in the genotyped Duroc pigs; $\boldsymbol{P}$ is a matrix containing top three principal components of genomic relationships, and these three principal components explained >80% variance across individuals; $\boldsymbol{f}$ is the vector of corresponding regression coefficients; $\boldsymbol{X}$ is a matrix of the tested SNP genotypes with entries 0, 1, and 2, indicating genotypes AA, AB, and BB, respectively; b is the fixed additive genetic effect of analyzed SNP; $\boldsymbol{g}$ is a vector of random polygenic effects; $\boldsymbol{W}$ is the incidence matrix relating the DEBVs to the corresponding random polygenic effects. It was assumed that $\boldsymbol{g}$ followed a normal distribution with mean of 0 and variance of $\boldsymbol{G}\sigma_g^2$, where $\boldsymbol{G}$ was the genomic realized relationship matrix. The relationship matrix $\boldsymbol{G}$ was constructed as follows (VanRaden, 2008):

$$\boldsymbol{G} = \frac{\boldsymbol{ZZ'}}{\sum_{i=1}^{m} 2p_iq_i}$$

where m is the number of SNPs; p_i is the frequency of the second allele at marker i with $q_i = 1 - p_i$; $\boldsymbol{Z}$ is the incidence matrix with elements of $2 - 2p_i$, $1 - 2p_i$, and $-2p_i$ for genotypes AA, Aa, and aa, respectively; $\boldsymbol{e}$ is a vector of residual effects, following a normal distribution of $\boldsymbol{e} \sim N(0, \boldsymbol{D}\sigma_e^2)$, where $\boldsymbol{D}$ is a diagonal matrix with elements $d_{ii} = (1 - r^2_{DEBV})/r^2_{DEBV}$ and r^2_{DEBV} referred to the reliabilities for DEBVs. Significant test of SNP effects was implemented by a two-tailed t-test. Bonferroni corrections were set as the genome-wide significant threshold [$-\log_{10}$ (0.05/number of SNPs)] = 5.89). Manhattan plots for these semen traits were drawn via CMplot function in rMVP package (Yin et al., 2021).

In addition, genetic variances explained by SNPs located within candidate region were calculated as follows:

$$\boldsymbol{y} = \boldsymbol{Pf} + \boldsymbol{Z}_1\boldsymbol{g}_1 + \boldsymbol{Z}_2\boldsymbol{g}_2 + \boldsymbol{e} \quad (3)$$

where $\boldsymbol{y}$, $\boldsymbol{P}$, $\boldsymbol{f}$, and $\boldsymbol{e}$ are the same as in the equation 2; $\boldsymbol{g}_1$ is the vector of genetic effects, following a normal distribution of $\boldsymbol{g}_1 \sim N(0, \boldsymbol{G}_1\sigma^2_{g_1})$, where $\boldsymbol{G}_1$ is the genomic realized relationship matrix that was constructed according to SNPs located within the candidate region, and $\sigma^2_{g_1}$ is the genetic variance captured by SNPs located within the candidate region; $\boldsymbol{g}_2$ is the vector of polygenetic effects, following a normal distribution of $\boldsymbol{g}_2 \sim N(0, \boldsymbol{G}_2\sigma^2_{g_2})$ where $\boldsymbol{G}_2$ is the genomic realized relationship matrix that constructed by the remaining SNPs, and $\sigma^2_{g_2}$ is the genetic variance captured by SNPs located outside the candidate region; $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are the incidence matrices relating the DEBVs to the corresponding random effects. Then, genetic variance explained by SNPs located within candidate region was calculated as $\frac{\sigma^2_{g_1}}{\sigma^2_{g_1}+\sigma^2_{g_2}}$. This calculation was performed by the genome-based restricted maximum likelihood (GREML) method in software DMU (Madsen and Jensen, 2013).

Haplotype block analysis

In order to detect the candidate QTL regions significantly associated with target traits, we employed software Haploview (Barrett et al., 2005) to perform haplotype block analysis. Haplotype block detection was performed on the whole chromosome containing the most significant SNP. Haplotype blocks were defined as the upper and lower confidence intervals of the estimates of pairwise D′ falling within certain threshold values (Gabriel et al., 2002). A haplotype block containing the most significant SNP was designated as the candidate QTL region.

Co-expression network analysis

The public dataset GSE74934 was obtained from the GEO database (http://www.ncbi.nlm.nih.gov/geo/) (Van Son et al., 2017), and this dataset contained 11 testis tissue samples from 11 Duroc boars. Of these 11 samples, 5 samples had low sperm DNA fragmentation index (**DFI**), and 6 samples had high sperm DFI. The sperm DFI has been reported to be related to three seme

traits (MO, CON, and VOL) in previous studies (Stahl et al., 2015; Lu et al., 2018). Thus, this study chose DFI as an indicator trait.

WGCNA in R package was used to perform the WGCNA (Langfelder and Horvath, 2008). As suggested by Langfelder and Horvath (2008), an appropriate soft power should be chosen when scale-free R^2 reached 0.8, and mean connectivity was at a low level. Considering this, we selected soft power of 10 to construct networks (Supplementary Figure 1). Dendrograms and heatmaps were generated to visualize the co-expression of genes. We merged modules with similar co-expression patterns (correlation > 85%; Supplementary Figure 2). Then, the merged module genes were used to calculate the module eigengenes (**MEs**) for further identifying modules that were significantly associated with sperm DFI. As described in Langfelder and Horvath (2008), gene significance was defined as the correlation of gene expression profiles with an external trait y, and module significance was designated as the average absolute gene significance of all genes in a certain module. Statistical significance between the ME E and the trait y could be obtained from a univariate regression model between E and y. Modules with high trait significance referred to the modules significantly associated with the sample trait. Therefore, the most significant MEs were regarded as the candidate genes.

Identification of candidate genes

Ensemble database (www.ensembl.org) was used to identify genes located within the candidate QTL region. The genes located within the most significant ME (identified by WGCNA) were considered as the important candidate genes.

Results

Variance components

Estimated variance components of MO, CON, and VOL are presented in Table 2. In this study, the heritabilities of MO, CON, and VOL were 0.17, 0.20, and 0.22, respectively. VOL was negatively correlated with MO (–0.34) and CON (–0.56), and MO was positively correlated with CON (0.54).

GWAS results for three semen traits

In total, 25 SNPs were found to be significantly associated with three semen traits in Duroc boars, and the most significant SNPs associated with these three semen traits are presented in Table 3. The remaining SNPs are available in Supplementary Table 1. As shown in Figure 1, for MO, the most significantly associated SNP (ALGA0034850) was located on chromosome 6. For CON, the most significantly associated SNP (WU_10.2_6_45143458) was also located on chromosome 6. For VOL, two most significantly associated SNPs were located on chromosome 15 (H3GA0044276) and chromosome 16 (ALGA0090092).

Through haplotype block analysis, we identified four candidate regions containing the most significantly associated SNPs on each chromosome (Figure 2). For MO, we detected one haplotype block ranging from 23,477,206 to 23,812,216 bp on SSC6, which contained SNP ALGA0034850 (Figure 2A). For CON, we detected one haplotype block spanning 467 kb (from 49,350,594 to 49,817,645 bp) on SSC6, which contained SNP WU_10.2_6_45143458 (Figure 2B). For VOL, we detected two haplotype blocks: one block ranged from 44,707,245 to 45,048,041 bp on SSC15, which contained SNP H3GA0044276 (Figure 2C), and the other one ranged from 31,722,381 to 32,220,183 bp on SSC16, which contained SNP ALGA0090092 (Figure 2D). The genes located within these blocks were regarded as candidate genes (Table 3). MO-related candidate gene was identified as *ENSSSCG00000018823*. CON-related candidate genes were identified as B9 domain containing 2 (*B9D2*), platelet-activating factor acetylhydrolase 1b catalytic subunit 3 (*PAFAH1B3*), transmembrane protein 145 (*TMEM145*), and capicua transcriptional repressor (*CIC*). VOL-related candidate genes were identified as WW and C2 domain containing 2 (*WWC2*), CDKN2A interacting protein (*CDKN2AIP*), inhibitor of growth family member 2 (*ING2*), trafficking protein particle complex 11 (*TRAPPC11*), storkhead box 2 (*STOX2*), and pelota mRNA surveillance and ribosome rescue factor (*PELO*).

GREML analysis

GREML analysis revealed that the above-mentioned QTL regions explained 1.95% to 6.15% of the total genetic variances for different traits. For MO, the candidate region explained 6.15% of the total genetic variance; for CON, the candidate region explained 1.95% of the total genetic variance; for VOL, two candidate regions explained 2.23% and 2.48% of the total genetic variance, respectively.

WGCNA results

To further verify the candidate genes identified by GWAS, we performed WGCNA. The WGCNA results showed that a total of 11 co-expression modules were detected. Module ME-turquoise was the most significantly associated module with the highest correlation coefficients between module and sperm DFI (Figure 3). Of the 11 detected semen trait-related candidate genes, 6 genes were located within module ME-turquoise, including *B9D2*, *TMEM145*, *WWC2*, *CDKN2AIP*, *TRAPPC11*, and *PELO*.

Table 2. Variance components of semen traits

Trait	σ_a^2	σ_p^2	σ_e^2	h_a^2 (se)[1]
MO	6.89e-04	4.63e-04	2.98e-03	0.17(0.02)
CON	0.95	0.86	2.83	0.20(0.02)
VOL	822.29	655.92	2,225.86	0.22(0.02)

[1]se, standard error of heritability.

Table 3. The most significant SNPs detected in the GWAS for semen quality traits

Trait	SNP	Chromosome	Position	P-value	Effect	MAF[1]	Candidate gene
MO	ALGA0034850	6	23477206	1.19e-07	–0.01	0.31	*ENSSSCG00000018823*
CON	WU_10.2_6_45143458	6	49350594	1.05e-06	–0.28	0.38	*B9D2*, *PAFAH1B3*, *TMEM145*, *and CIC*
VOL	H3GA0044276	15	45048041	4.22e-07	–9.91	0.47	*WWC2*, *CDKN2AIP*, *ING2*, *TRAPPC11*, *and STOX2*
	ALGA0090092	16	31722381	2.24e-07	9.72	0.36	*PELO*

[1]MAF, minor allele frequency.

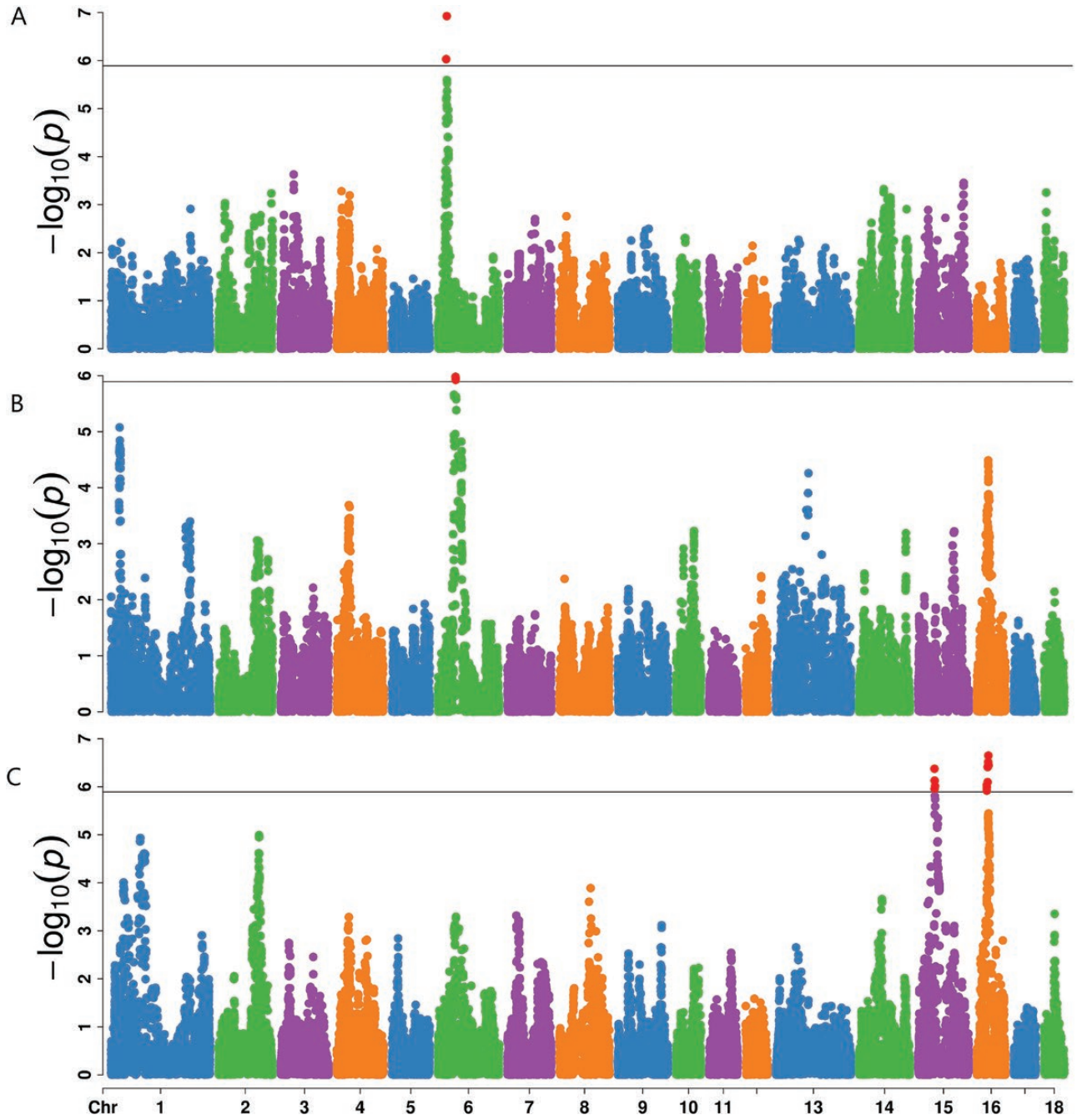

Figure 1. Manhattan plots of three semen quality traits. (A) MO, (B) CON, and (C) VOL. The black lines divide SNP with *P*-values < 2.24e-07. The signals of genome-wide significant SNPs are colored in red.

Discussion

In GWAS and haplotype block analysis, we identified one candidate gene (*ENSSSCG00000018823*) associated with MO, four candidate genes (*B9D2*, *PAFAH1B3*, *TMEM145*, and *CIC*) associated with CON, and six candidate genes (*WWC2*, *CDKN2AIP*, *ING2*, *TRAPPC11*, *STOX2*, and *PELO*) associated with VOL. We further employed the WGCNA algorithm to detect the candidate genes located within the most significant ME. Integrating the results of GWAS and WGCNA, we found thatsix genes (*B9D2*, *TMEM145*, *WWC2*, *CDKN2AIP*, *TRAPPC11*, and *PELO*) were located within the significant ME, and these six genes were the candidate genes regulating semen traits. It should be noted that not all candidate genes identified by GWAS could be further verified by WGCNA. Different traits exhibited different associated genomic regions.

GWAS showed that *ENSSSCG00000018823* was MO-related candidate gene, and this gene was an U2 spliceosomal RNA, which was a branch of small nuclear RNA (**snRNA**) molecules found in virtually all eukaryotic organisms. In mouse mature spermatozoa, U2 snRNA was found to be located in the sperm nucleus; this indicated that U2 snRNA might regulate the development of spermatozoon (Concha et al., 1993). In human sperm, U2 snRNA might be involved in the process of sperm development via miR-1246, which regulated sperm development by either suppressing or enhancing mRNA translation (Ørom et al., 2008; Pantano et al., 2015).

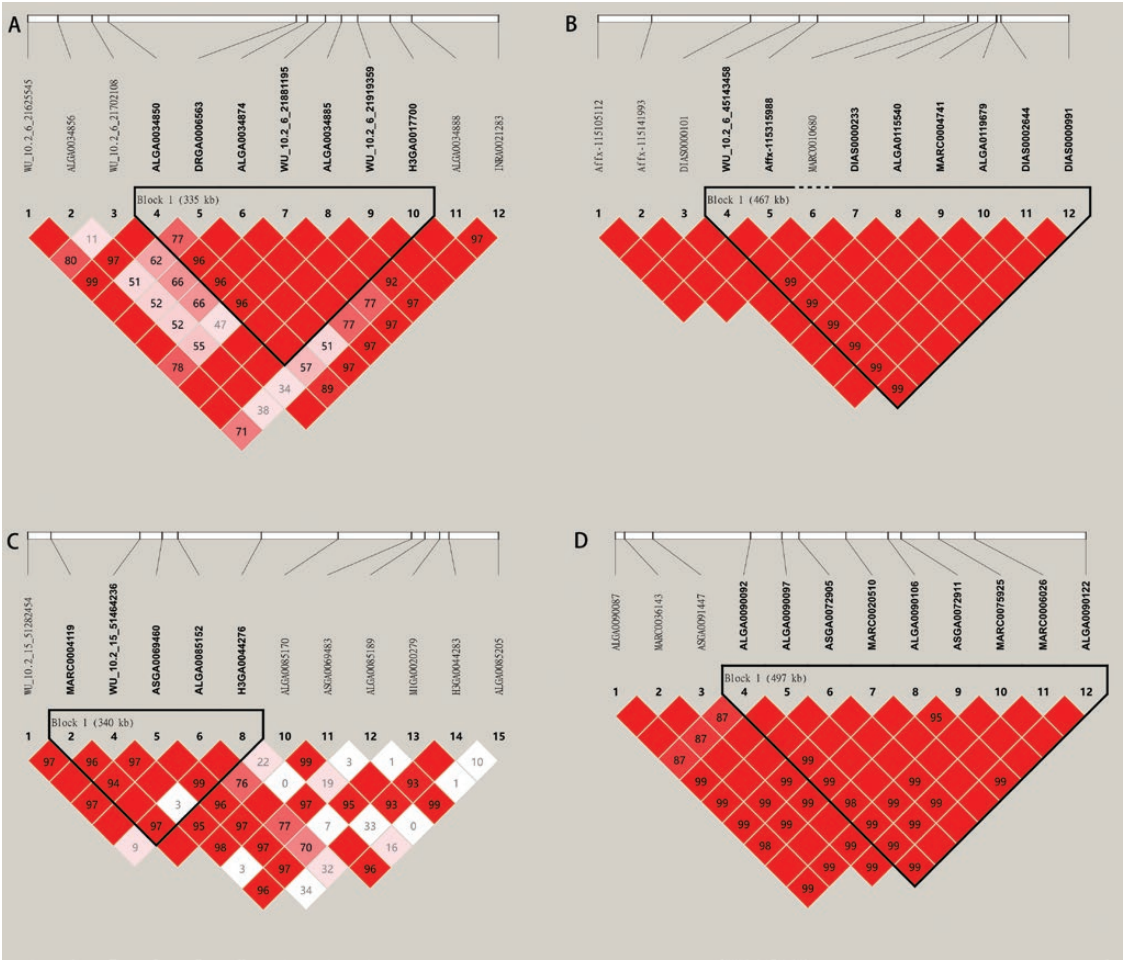

Figure 2. Haplotype block analysis for significant associated genomic regions for three semen traits: MO, CON, and VOL. Markers in blocks are shown in bold. (A) Linkage disequilibrium block detected in the regions from 23,477,206 to 23,812,216 bp on SSC6 associated with MO; (B) Linkage disequilibrium block detected in the regions from 49,350,594 to 49,817,645 bp on SSC6 associated with CON; (C) Linkage disequilibrium block detected in the regions from 44,707,245 to 45,048,041 bp on SSC15 associated with VOL; (D) Linkage disequilibrium block detected in the regions from 31,722,381 to 32,220,183 bp on SSC16 associated with VOL.

GWAS results showed that *B9D2*, *PAFAH1B3*, *TMEM145*, and *CIC* were the CON-related candidate genes. *B9D2* was upregulated during mucociliary differentiation, and the *B9D2*-encoded protein was localized to basal bodies and cilia. Recently, *B9D2* has been reported to migrate away from the centriole to expose the sperm axoneme to the cytoplasm in *Drosophila spermatids* (Basiri et al., 2014). Additionally, in the absence of *B9D2*, sperm flagella showed significant ultrastructural defects in *Drosophila* male germ cells (Vieillard et al., 2016). *PAFAH1B3* encoded an acetylhydrolase catalyzing the removal of an acetyl group from the glycerol backbone of the platelet-activating factor. *PAFAH1B3* was found to be correlated with LIS1 protein function, which was important for restoring spermatogenesis and male fertility in mouse model (Yan et al., 2003). *TMEM145* encoded a transmembrane protein involved in G protein-coupled receptor signaling pathway, and, in human prostate cancer cell lines, putative homozygous deletion was found in *TMEM145* (Seim et al., 2017). *CIC* is a member of the high-mobility-group box superfamily of transcriptional repressors. *CIC* inhibition reduced sperm hyperactivated motility and acrosome reaction, and the report that citrate induced these key sperm activities confirmed a role of *CIC* in sperm physiological process (Cappello et al., 2012).

GWAS results showed that *WWC2*, *CDKN2AIP*, *ING2*, *TRAPPC11*, *STOX2*, and *PELO* were VOL-related candidate genes. WWC2 encodes a member of the WW- and C2-domain-containing family proteins. In prostate cancer, *WWC2* was found to be differentially expressed in prostate cancer (Sun et al., 2016). *CDKN2AIP* encodes the proteins regulating the DNA damage response through several different signaling pathways. *CDKN2AIP* played an important role in spermatogonial self-renewal and proliferation in spermatogonial stem cells in *CDKN2AIP*-knockout mouse (Cui et al., 2020). *ING2* is a member of the inhibitor of growth (**ING**) family. *ING2* was defined as a novel regulator of spermatogenesis through both p53- and chromatin-mediated mechanisms (Saito et al., 2010). *TRAPPC11* encoded a subunit of TRAPP complex, which was crucial for male fertility (Riedel et al., 2018). *STOX2* encoded a Storkhead-box_winged helix domain-containing protein; the expression of *STOX2* was significantly increased in the testes of ΔXmiR mice, in which a few abnormal seminiferous tubules were observed (Ota et al., 2019). This indicated that *STOX2* might be involved in spermatogenesis. *STOX2* was reported to be upregulated when mrhl RNA was downregulated in mouse spermatogonial cells, and the Western blot result showed that the downregulation of mrhl RNA was associated with spermatogonia differentiation

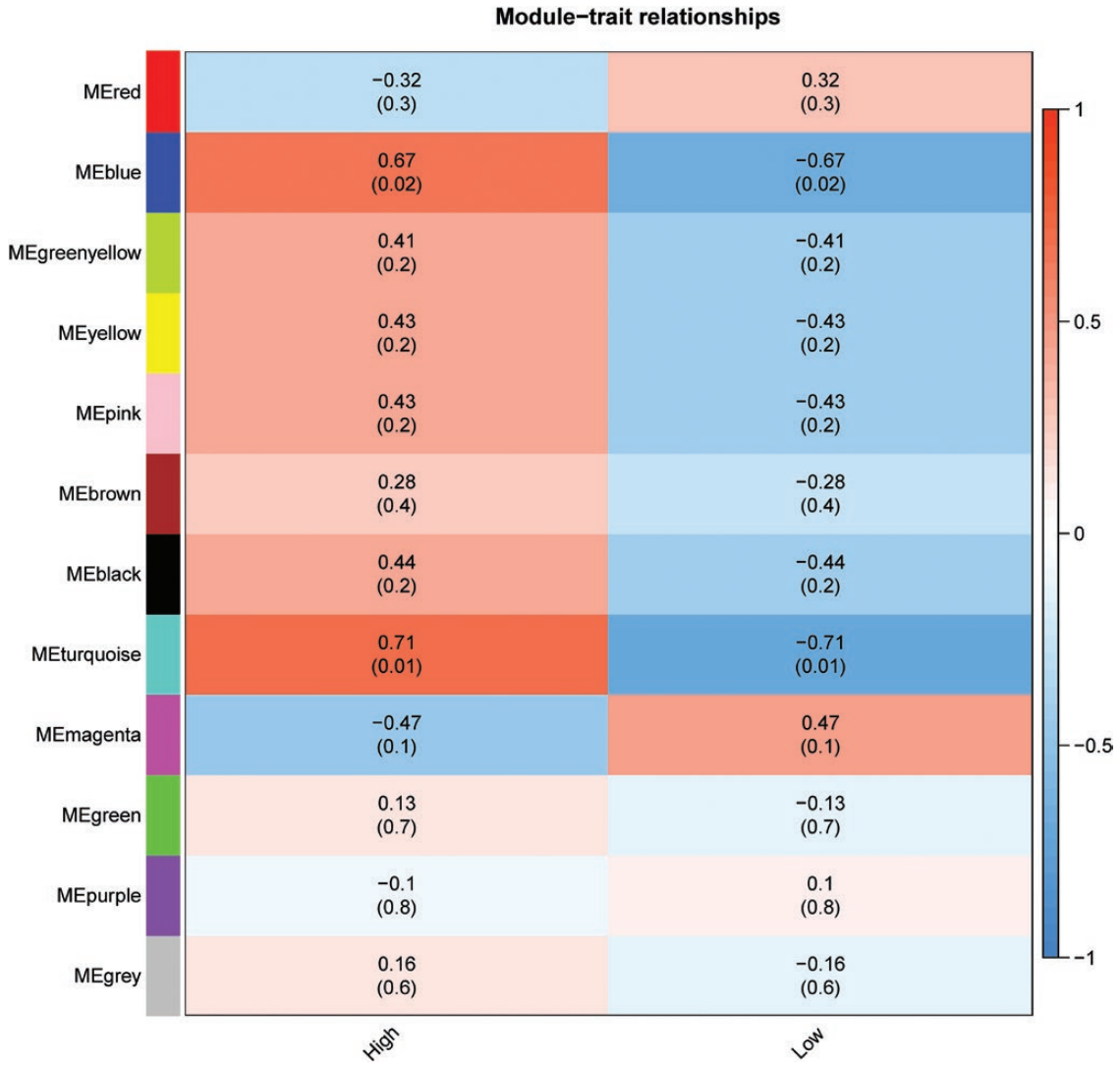

Figure 3. Heatmap of the correlation between MEs and sperm DFI (high vs. low). Each row in the table corresponds to a consensus module. Numbers in the table report the correlations of the corresponding MEs and traits, with the *P*-values printed below the correlations in parentheses. The table is color coded by correlation according to the color legend.

in 7- and 21-d-old mice (Akhade et al., 2014). *PELO* encoded a protein containing a conserved nuclear localization signal, and *PELO* deficiency in adult mice resulted in the depletion of all germ cells (Raju et al., 2015).

WGCNA results indicated that *B9D2*, *TMEM145*, *WWC2*, *CDKN2AIP*, *TRAPPC11*, and *PELO* were located within the most significant ME. GWAS and WGCNA have been integrated and applied in the previous studies of other species (Farber, 2013; Deng et al., 2019). In this study, WGCNA was used as the supplementation of GWAS to further validate candidate genes identified by GWAS. Recently, system biology approach has been used for detecting candidate genes potentially affecting sperm quality (Gòdia et al., 2020). To our knowledge, the integration of GWAS and WGCNA is still seldom reported for identifying semen trait-related candidate genes in pig. It should be noted that to further confirm the identified candidate SNPs and genes, further laboratory experiments are still needed in the future study.

After identification of semen trait-related candidate genes, one possible approach to utilize this information is to assign more weights to SNPs located within candidate genes in genomic selection. This method is usually termed as weighted BLUP, which has been successfully applied for improving the genomic prediction (Zhang et al., 2016).

Some other semen trait-related candidate genes in pigs have also been reported in previous studies. For example, *MTFMT* was associated with MO in Landrace pigs (Diniz et al., 2014); *HPGDS*, *SPATA7*, and *METTL3* were associated with MO in Landrace pigs (Marques et al., 2018); *SCN8A*, *DNAI2*, *IQCG*, and *LOC102167830* were associated with MO in Yorkshire pigs; *PPP2R2B*, *NEK2*, *NDRG*, *ADAM7*, *SKP2*, and *RNASET2* were associated with MO in Duroc pigs (Gao et al., 2019).

Overall, the semen trait-related candidate genes might be breed specific and even population specific, suggesting that semen traits are complex traits. The semen trait regulation mechanism may vary with breeds and populations. Hence, it is necessary to further investigate the genetic mechanism underlying semen traits. Our findings provide an insight into the genetic architecture of semen traits, and they could also be used in breeding scheme to accelerate the genetic progress of semen traits in Duroc boars.

Supplementary Data

Supplementary data are available at *Journal of Animal Science* online.

Acknowledgments

This study was financially supported by the National Natural Science Foundation of China (31802039), the National Key Research and Development Program of China (2019YFE0115400), Major Science and Technology Projects in Hubei Province (2020ABA016), the Fundamental Research Funds for the Central Universities (2662020DKPY005), the China Agriculture Research System of MOF and MARA (CARS-35), the key research and development program in Guangxi Province (912269910019), and the Natural Science Foundation of Jingmen City (2020YFYB045).

Conflict of interest statement

The authors declare no real or perceived conflicts of interest.

Literature Cited

Akhade, V. S., G. Arun, S. Donakonda, and M. R. Rao. 2014. Genome wide chromatin occupancy of mrhl RNA and its role in gene regulation in mouse spermatogonial cells. *RNA Biol.* **11**:1262–1279. doi:10.1080/15476286.2014.996070

Barrett, J. C., B. Fry, J. Maller, and M. J. Daly. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. **21**:263–265. doi:10.1093/bioinformatics/bth457

Basiri, M. L., A. Ha, A. Chadha, N. M. Clark, A. Polyanovsky, B. Cook, and T. Avidor-Reiss. 2014. A migrating ciliary gate compartmentalizes the site of axoneme assembly in *Drosophila spermatids*. *Curr. Biol.* **24**:2622–2631. doi:10.1016/j.cub.2014.09.047

Browning, B. L., Y. Zhou, and S. R. Browning. 2018. A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **103**:338–348. doi:10.1016/j.ajhg.2018.07.015

Cappello, A. R., C. Guido, A. Santoro, M. Santoro, L. Capobianco, D. Montanaro, M. Madeo, S. Andò, V. Dolce, and S. Aquila. 2012. The mitochondrial citrate carrier (CIC) is present and regulates insulin secretion by human male gamete. *Endocrinol.* **153**:1743–1754. doi:10.1210/en.2011-1562

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* **42**:2. doi:10.1186/1297-9686-42-2

Christensen, O. F., P. Madsen, B. Nielsen, T. Ostersen, and G. Su. 2021. Single-step methods for genomic evaluation in pigs. *Animal*. **6**:1565–1571. doi:10.1017/S1751731112000742

Concha, I. I., U. Urzua, A. Yañez, R. Schroeder, C. Pessot, and L. O. Burzio. 1993. U1 and U2 snRNA are localized in the sperm nucleus. *Exp. Cell Res.* **204**:378–381. doi:10.1006/excr.1993.1046

Cui, W., X. He, X. Zhai, H. Zhang, Y. Zhang, F. Jin, X. Song, D. Wu, Q. Shi, and L. Li. 2020. CARF promotes spermatogonial self-renewal and proliferation through Wnt signaling pathway. *Cell Discov.* **6**:85. doi:10.1038/s41421-020-00212-7

D´Allaire, S., A. D. Leman, and R. Drolet. 1992. Optimizing longevity in sows and boars. *Vet. Clin. North Am. Food Anim. Pract.* **8**:545–557. doi:10.1016/s0749-0720(15)30703-9

Deng, T., A. Liang, S. Liang, X. Ma, X. Lu, A. Duan, C. Pang, G. Hua, S. Liu, and G. Campanile. 2019. Integrative analysis of transcriptome and GWAS data to identify the hub genes associated with milk yield trait in buffalo. *Front Genet.* **10**:36. doi:10.3389/fgene.2019.00036

Diniz, D. B., M. S. Lopes, M. L. Broekhuijse, P. S. Lopes, B. Harlizius, S. E. Guimarães, N. Duijvesteijn, E. F. Knol, and F. F. Silva. 2014. A genome-wide association study reveals a novel candidate gene for sperm motility in pigs. *Anim. Reprod. Sci.* **151**:201–207. doi:10.1016/j.anireprosci.2014.10.014

Farber, C. R. 2013. Systems-level analysis of genome-wide association data. *G3 (Bethesda)* **3**:119–129. doi:10.1534/g3.112.004788

Gabriel, S. B., S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, *et al.* 2002. The structure of haplotype blocks in the human genome. *Science*. **296**:2225–2229. doi:10.1126/science.1069424

Gao, N., Y. Chen, X. Liu, Y. Zhao, L. Zhu, A. Liu, W. Jiang, X. Peng, C. Zhang, Z. Tang, *et al.* 2019. Weighted single-step GWAS identified candidate genes associated with semen traits in a Duroc boar population. *BMC Genomics*. **20**:797. doi:10.1186/s12864-019-6164-5

Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol.* **41**:55. doi:10.1186/1297-9686-41-55

Goddard, M. E., and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* **10**:381–391. doi:10.1038/nrg2575

Gòdia, M., A. Reverter, R. González-Prendes, Y. Ramayo-Caldas, A. Castelló, J. E. Rodríguez-Gil, A. Sánchez, and A. Clop. 2020. A systems biology framework integrating GWAS and RNA-seq to shed light on the molecular basis of sperm quality in swine. *Genet. Sel. Evol.* **52**:72. doi:10.1186/s12711-020-00592-0

Koketsu, Y., and Y. Sasaki. 2009. Boar culling and mortality in commercial swine breeding herds. *Theriogenology* **71**:1186–1191. doi:10.1016/j.theriogenology.2008.12.018

Langfelder, P., and S. Horvath. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. **9**:559. doi:10.1186/1471-2105-9-559

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* **92**:4656–4663. doi:10.3168/jds.2009-2061

Lippert, C., J. Listgarten, Y. Liu, C. M. Kadie, R. I. Davidson, and D. Heckerman. 2011. FaST linear mixed models for genome-wide association studies. *Nat. Methods* **8**:833–835. doi:10.1038/nmeth.1681

Listgarten, J., C. Lippert, C. M. Kadie, R. I. Davidson, E. Eskin, and D. Heckerman. 2012. Improved linear mixed models for genome-wide association studies. *Nat. Methods* **9**:525–526. doi:10.1038/nmeth.2037

Lu, J. C., J. Jing, L. Chen, Y. F. Ge, R. X. Feng, Y. J. Liang, and B. Yao. 2018. Analysis of human sperm DNA fragmentation index (DFI) related factors: a report of 1010 subfertile men in China. *Reprod. Biol. Endocrinol.* **16**:23. doi:10.1186/s12958-018-0345-y

Madsen, P., and J. Jensen. 2013. A user's guide to DMU. A package for analysing multivariate mixed models. Version 6, release 5.2. Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, University of Aarhus.

Marques, D. B. D., J. W. M. Bastiaansen, M. Broekhuijse, M. S. Lopes, E. F. Knol, B. Harlizius, S. E. F. Guimarães, F. F. Silva, and P. S. Lopes. 2018. Weighted single-step GWAS and gene network analysis reveal new candidate genes for semen traits in pigs. *Genet. Sel. Evol.* **50**:40. doi:10.1186/s12711-018-0412-z

Marques, D. B. D., M. S. Lopes, M. Broekhuijse, S. E. F. Guimarães, E. F. Knol, J. W. M. Bastiaansen, F. F. Silva, and P. S. Lopes. 2017. Genetic parameters for semen quality and quantity traits in five pig lines. *J. Anim. Sci.* **95**:4251–4259. doi:10.2527/jas2017.1683

McPherson, F. J., S. G. Nielsen, and P. J. Chenoweth. 2014. Semen effects on insemination outcomes in sows. *Anim. Reprod. Sci.* **151**:28–33. doi:10.1016/j.anireprosci.2014.09.021

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

Myromslien, F. D., N. H. Tremoen, I. Andersen-Ranberg, R. Fransplass, E. B. Stenseth, T. T. Zeremichael, M. Van Son, E. Grindflek, and A. H. Gaustad. 2019. Sperm DNA integrity in Landrace and Duroc boar semen and its relationship to litter size. *Reprod. Domest. Anim.* **54**:160–166. doi:10.1111/rda.13322

Ørom, U. A., F. C. Nielsen, and A. H. Lund. 2008. MicroRNA-10a binds the 5′UTR of ribosomal protein mRNAs and enhances their translation. *Mol. Cell.* **30**:460–471. doi:10.1016/j.molcel.2008.05.001

Ostersen, T., O. F. Christensen, M. Henryon, B. Nielsen, G. Su, and P. Madsen. 2011. Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. *Genet. Sel. Evol.* **43**:38. doi:10.1186/1297-9686-43-38

Ota, H., Y. Ito-Matsuoka, and Y. Matsui. 2019. Identification of the X-linked germ cell specific miRNAs (XmiRs) and their functions. *PLoS One*. **14**:e0211739. doi:10.1371/journal.pone.0211739

Pantano, L., M. Jodar, M. Bak, J. L. Ballescà, N. Tommerup, R. Oliva, and T. Vavouri. 2015. The small RNA content of human sperm reveals pseudogene-derived piRNAs complementary to protein-coding genes. *RNA*. **21**:1085–1095. doi:10.1261/rna.046482.114

Raju, P., G. Nyamsuren, M. Elkenani, A. Kata, E. Tsagaan, W. Engel, and I. M. Adham. 2015. Pelota mediates gonocyte maturation and maintenance of spermatogonial stem cells in mouse testes. *Reproduction*. **149**:213–221. doi:10.1530/REP-14-0391

Riedel, F., A. Galindo, N. Muschalik, and S. Munro. 2018. The two TRAPP complexes of metazoans have distinct roles and act on different Rab GTPases. *J. Cell Biol.* **217**:601–617. doi:10.1083/jcb.201705068

Robinson, J. A., and M. M. Buhr. 2005. Impact of genetic selection on management of boar replacement. *Theriogenology*. **63**:668–678. doi:10.1016/j.theriogenology.2004.09.040

Saito, M., K. Kumamoto, A. I. Robles, I. Horikawa, B. Furusato, S. Okamura, A. Goto, T. Yamashita, M. Nagashima, T. L. Lee, *et al.* 2010. Targeted disruption of ING2 results in defective spermatogenesis and development of soft-tissue sarcomas. *PLoS One*. **5**:e15541. doi:10.1371/journal.pone.0015541

Seim, I., P. L. Jeffery, P. B. Thomas, C. C. Nelson, and L. K. Chopin. 2017. Whole-genome sequence of the metastatic PC3 and LNCaP human prostate cancer cell lines. *G3 (Bethesda)*. **7**:1731–1741. doi:10.1534/g3.117.039909

Sevillano, C. A., M. S. Lopes, B. Harlizius, E. H. Hanenberg, E. F. Knol, and J. W. Bastiaansen. 2015. Genome-wide association study using deregressed breeding values for cryptorchidism and scrotal/inguinal hernia in two pig lines. *Genet. Sel. Evol.* **47**:18. doi:10.1186/s12711-015-0096-6

Stahl, P. J., C. Cogan, A. Mehta, A. Bolyakov, D. A. Paduch, and M. Goldstein. 2015. Concordance among sperm deoxyribonucleic acid integrity assays and semen parameters. *Fertil. Steril.* **104**:56–61.e1. doi:10.1016/j.fertnstert.2015.04.023

Sun, Y., X. Jia, L. Hou, and X. Liu. 2016. Screening of differently expressed miRNA and mRNA in prostate cancer by integrated analysis of transcription data. *Urology*. **94**:313.e1–313.e6. doi:10.1016/j.urology.2016.04.041

Van Son, M., N. H. Tremoen, A. H. Gaustad, F. D. Myromslien, D. I. Våge, E.-B. Stenseth, T. T. Zeremichael, and E. Grindflek. 2017. RNA sequencing reveals candidate genes and polymorphisms related to sperm DNA integrity in testis tissue from boars. *BMC Vet. Res.* **13**:362. doi:10.1186/s12917-017-1279-x

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**:4414–4423. doi:10.3168/jds.2007-0980

Vieillard, J., M. Paschaki, J. L. Duteyrat, C. Augière, E. Cortier, J. A. Lapart, J. Thomas, and B. Durand. 2016. Transition zone assembly and its contribution to axoneme formation in *Drosophila* male germ cells. *J. Cell Biol.* **214**:875–889. doi:10.1083/jcb.201603086

Visscher, P. M., N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, M. A. Brown, and J. Yang. 2017. 10 Years of GWAS discovery: Biology, function, and translation. *Am. J. Hum. Genet.* **101**:5–22. doi:10.1016/j.ajhg.2017.06.005

Yan, W., A. H. Assadi, A. Wynshaw-Boris, G. Eichele, M. M. Matzuk, and G. D. Clark. 2003. Previously uncharacterized roles of platelet-activating factor acetylhydrolase 1b complex in mouse spermatogenesis. *Proc. Natl. Acad. Sci.* **100**:7189–7194. doi:10.1073/pnas.1236145100

Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher. 2011. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**:76–82. doi:10.1016/j.ajhg.2010.11.011

Yang, J., N. A. Zaitlen, M. E. Goddard, P. M. Visscher, and A. L. Price. 2014. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46**:100–106. doi:10.1038/ng.2876

Yin, L., H. Zhang, Z. Tang, J. Xu, D. Yin, Z. Zhang, X. Yuan, M. Zhu, S. Zhao, and X. Li. 2021. rmvp: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, Proteomics & Bioinformatics*. doi:10.1016/j.gpb.2020.10.007

Zhang, X., D. Lourenco, I. Aguilar, A. Legarra, and I. Misztal. 2016. Weighting strategies for single-step genomic BLUP: an iterative approach for accurate calculation of GEBV and GWAS. *Front Genet.* 7. doi:10.3389/fgene.2016.00151

Zhao, Y., N. Gao, X. Li, S. El-Ashram, Z. Wang, L. Zhu, W. Jiang, X. Peng, C. Zhang, Y. Chen, *et al.* 2020. Identifying candidate genes associated with sperm morphology abnormalities using weighted single-step GWAS in a Duroc boar population. *Theriogenology*. **141**:9–15. doi:10.1016/j.theriogenology.2019.08.031