

Article

Identification of New Candidate Genes Related to Semen Traits in Duroc Pigs through Weighted Single-Step GWAS

Xiaoke Zhang ¹, Qing Lin ¹, Weili Liao ¹, Wenjing Zhang ¹, Tingting Li ¹, Jiaqi Li ¹, Zhe Zhang ¹,
Xiang Huang ^{2,*} and Hao Zhang ^{1,*}

¹ National Engineering Research Center for Breeding Swine Industry, Guangdong Provincial Key Lab of Agro-Animal Genomics and Molecular Breeding, College of Animal Science, South China Agricultural University, Guangzhou 510642, China

² Guangdong Guyue Technology Co., Ltd., Guangzhou 510980, China

* Correspondence: huangxiangxms@163.com (X.H.); zhanghao@scau.edu.cn (H.Z.)

Simple Summary: Due to the complexity of sperm cell reproduction and maturation, the genetic structure of semen traits remains largely unknown. In our study, we used weighted single-step GWAS to detect genetic regions and further candidate genes related to semen traits in Duroc boars. This study provides in-depth understanding of the genetic structure of semen traits and the biological information provided by gene networks, and can be applied to speed up the genetic process of semen traits in boars. The candidate genes *CATSPER1*, *STRA8*, *ZSWIM7*, *TEKT3*, *UBB*, *PTBP2*, *EIF2B2*, *MLH3*, and *CCDC70* were associated with semen traits in Duroc pigs.

Abstract: Semen traits play a key role in the pig industry because boar semen is widely used in purebred and crossbred pigs. The production of high-quality semen is crucial to ensuring a good result in artificial insemination. With the wide application of artificial insemination in the pig industry, more and more attention has been paid to the improvement of semen traits by genetic selection. The purpose of this study was to identify the genetic regions and candidate genes associated with semen traits of Duroc boars. We used weighted single-step GWAS to identify candidate genes associated with sperm motility, sperm progressive motility, sperm abnormality rate and total sperm count in Duroc pigs. In Duroc pigs, the three most important windows for sperm motility—sperm progressive motility, sperm abnormality rate, and total sperm count—explained 12.45%, 9.77%, 15.80%, and 12.15% of the genetic variance, respectively. Some genes that are reported to be associated with spermatogenesis, testicular function and male fertility in mammals have been detected previously. The candidate genes *CATSPER1*, *STRA8*, *ZSWIM7*, *TEKT3*, *UBB*, *PTBP2*, *EIF2B2*, *MLH3*, and *CCDC70* were associated with semen traits in Duroc pigs. We found a common candidate gene, *STRA8*, in sperm motility and sperm progressive motility, and common candidate genes *ZSWIM7*, *TEKT3* and *UBB* in sperm motility and sperm abnormality rate, which confirms the hypothesis of gene pleiotropy. Gene network enrichment analysis showed that *STRA8*, *UBB* and *CATSPER1* were enriched in the common biological process and participated in male meiosis and spermatogenesis. The SNPs of candidate genes can be given more weight in genome selection to improve the ability of genome prediction. This study provides further insight into the understanding the genetic structure of semen traits in Duroc boars.

Keywords: pig; semen traits; weighted single-step GWAS; candidate genes; gene network



Citation: Zhang, X.; Lin, Q.; Liao, W.; Zhang, W.; Li, T.; Li, J.; Zhang, Z.; Huang, X.; Zhang, H. Identification of New Candidate Genes Related to Semen Traits in Duroc Pigs through Weighted Single-Step GWAS. *Animals* **2023**, *13*, 365. <https://doi.org/10.3390/ani13030365>

Academic Editor: Krzysztof Flisikowski

Received: 2 December 2022

Revised: 16 January 2023

Accepted: 17 January 2023

Published: 20 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Semen traits play a key role in the wide application of artificial insemination in the pig industry. The economic profit of artificial pig breeding stations is highly dependent on the quantity and quality of semen [1], and the decline of semen quality is one of the main reasons for shortening the life span of boars [2]. Understanding the genetic background

and detecting the genetic markers related to semen traits are helpful to improve the genetic selection of semen traits and speed up the genetic process.

In recent years, with the rapid development of high-throughput genotyping and molecular technology, researchers can accurately identify quantitative trait loci by looking for the association between genetic markers and phenotypic records, which is called genome-wide association study (GWAS) [3]. GWAS has been successfully applied to QTL mapping of important economic traits in animal and plant breeding and detection of genetic risk factors of human diseases [4]. Candidate genes related to semen traits were also identified by GWAS in several previous studies [5–8]. However, due to the complexity of sperm cell reproduction and maturation, the genetic structure of semen characteristics is still unknown to a large extent [5]. The candidate genes found may be different for the same trait in the same breed of different populations. Gene mutations and impaired expression that control the whole process of spermatogenesis and sperm maturation will lead to problems in semen quality and fertility. Therefore, mining more QTL regions and candidate genes related to semen traits can broaden our understanding of the genetic structure of porcine semen traits.

Weighted single-step GWAS (WssGWAS), proposed by Wang et al. [9], is a method to estimate SNP effect based on the single-step best linear unbiased prediction (ssGBLUP) [10] of genome breeding values (GEBV) of all phenotypic, genotypic and pedigree-related animals. In addition, it allows the variance of SNP to vary, thus improving the accuracy of SNP effect estimation. Therefore, when the number of animals with both phenotype and genotype is small and the traits are controlled by QTL with large effects, among the established GWAS methods, weighted single-step GWAS is more suitable for the association study of domestic animals. This method has been applied to the growth, carcass and reproductive traits of livestock [11–16]. In addition, the candidate genes related to the QTL region identified in GWAS can be analyzed by gene networks. Gene networks are used to study the pathways and biological processes shared by these genes [17].

In this study, we used WssGWAS to detect genetic regions and further candidate genes related to semen traits in Duroc boars. This study will have an in-depth understanding of the genetic structure of semen traits and the biological information provided by gene networks and can be applied to speed up the genetic progress of semen traits in boars.

2. Materials and Methods

2.1. Population and Phenotype Data

Animals used in this study were from the same artificial insemination station (Guangzhou, China). A total of 24,983 semen records were collected from 583 Duroc pigs during 2020–2022. All boars had complete pedigree records for four generations. Four semen traits were measured. Sperm motility (SPMOT), sperm progressive motility (SPPMOT) and sperm abnormality rate (SPABR) were measured using Magapor Gesipor3.0 CASA system (Magapor S.L. Parque Científico Agroalimentario Valdeferrín-AulaDei, Ejea de los Caballeros, Zaragoza, Spain). Total sperm count (SPCOUNT) was calculated by multiplying semen volume (mL) by semen concentration (10^6 /mL, measured by the CASA system).

The phenotypes for four semen traits of Duroc pigs are shown in Table 1. According to the research of Marques et al. [18] and Wang et al. [19] and combined with the characteristics of our data, the quality control of phenotypes was: (1) frequency of semen collection <5 were excluded; (2) eliminating the semen records with an ejaculation volume ≤ 50 mL; (3) semen records with sperm motility <10% were excluded; (4) eliminating the semen records with adjacent semen collection interval >60 days and semen collection interval of 0 days.

Table 1. Descriptive statistics for the semen traits of Duroc pigs.

Trait	Number of Boars	Number of Phenotypes	Mean ± SD	CV(%) ^a	Minimum	Maximum
SPMOT/%	583	24,833	82.35 ± 12.08	14.67	10.00	100.00
SPPMOT/%	583	24,730	28.84 ± 18.44	63.92	1.00	100.00
SPABR/%	583	24,917	28.81 ± 12.49	43.35	4.00	100.00
SPCOUNT/10 ⁸	583	24,492	364.67 ± 190.59	52.26	50.00	2529.09

^a CV: Coefficient of variation.

2.2. Genotypic Data

Genomics DNA of pig semen samples were extracted and purified from 571 Duroc pigs. These boars were genotyped by using the KPS 50k SNP array (KPS Porcine Breeding Chip v1, Beijing, China) containing 57,566 SNPs. SNPs that unmapped the reference genome (*Sus scrofa* 11.1), and located in sexual chromosomes and with missing position information were removed, after which 50,897 SNPs were retained. After that, individuals with call rate <0.9, SNPs with call rate <0.9, minor allele frequency <0.01, and Hardy–Weinberg equilibrium <10⁶ were removed using plink v1.90 [20]. Finally, 38,054 SNPs were retained. Missing genotypes were finally imputed using Beagle software (version 4.1) [21].

2.3. Statistical Analyses

Variance components and heritability of SPMOT, SPPMOT, SPABR, and SPCOUNT traits were estimated with two methods using the average information restricted maximum likelihood (AIREML) [22] of the AIREMLF90 procedure by BLUPF90 software [23]. The two methods are pedigree-based best linear unbiased prediction (BLUP) and ssGBLUP, and calculated genetic and phenotype correlations of semen traits using “OPTION se_covar_function.”

The following multiple traits repeatability model was used to estimate variance components:

$$Y = X\beta + Za + Wp + \text{Age} + \text{Intv} + \varepsilon, \quad (1)$$

where y is the vector of phenotypic observed value, β is the vector of fixed effects (year-season of semen collection and birth parities of boars), year is 2020–2022, March–May is spring, June–August is summer, September–November is autumn, December–February is winter, the birth parity of boars is 1–6, $a \sim N(0, U\sigma_a^2)$ is a vector of additive genetic effects, U is pedigree-derived relationship matrix (A matrix) or H matrix, BLUP using A matrix and ssGBLUP using H matrix estimated variance components, $p \sim N(0, I\sigma_p^2)$ is a vector of random permanent environmental effects, covariates Age and Intv denote the month age of the boars when semen collection and the interval between two subsequent semen collections in days, respectively, $\varepsilon \sim N(0, I\sigma_\varepsilon^2)$ is the vector of random residuals, and X , Z , and W are the incidence matrices of β , a , and p , respectively. The σ_a^2 , σ_p^2 and σ_ε^2 components are the additive genetic, permanent environmental and residual variances, respectively. I is the identity matrix. H is the matrix that combines pedigree and genomic information [10,24], and was calculated as:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G_w^{-1} - A_{22}^{-1} \end{bmatrix} \quad (2)$$

where A_{22} is a submatrix of A for the genotyped individuals: $G_w = 0.9G + 0.1A_{22}$. These weights were used to be compatible with genomic and phenotypic information and to control bias. $G = \frac{ZDZ'}{\sum_{i=1}^n 2p_i(1-p_i)}$ is the genomic relationship matrix [25], where Z is a genotype matrix adjusted for allele frequencies (with 0, 1, and 2 representing genotypes AA, Aa, and aa, respectively). D is a diagonal matrix containing the SNP weight, n is the number of SNPs, and p_i is the minor allele frequency of the i^{th} SNP.

The weighted single-step GWAS was conducted using the BLUPF90 software family in an iterative way adapted for genomic analyses [11]. Briefly, for phenotype, pedigree and genomic file preprocessing, we made use of RENUMF90 and variance components were

estimated using AIREMLF90, which were then used in BLUPF90 to predict GEBV. SNP effects were then calculated using the postGSF90 [26] procedure. The association study was used for the iteration procedure according to Wang et al. [9] with the following steps:

Step 1: Procedure initialization, let $t = 1$, $D_{(t)} = I$, $G_{(t)} = \lambda ZD_{(t)}Z'$ and $\lambda = \frac{1}{\sum_{i=1}^n 2p_i(1-p_i)}$;

Step 2: Calculated GEBV of the entire dataset, via ssGBLUP method with $H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & (0.9G_{(t)} + 0.1A_{22})^{-1} - A_{22}^{-1} \end{bmatrix}$;

Step 3: Calculated marker effects, via $\hat{g}_t = \lambda D_{(t)}Z'G_{(t)}^{-1}\hat{a}$, where \hat{a} is the GEBV of the genotyped individuals;

Step 4: Calculated SNP weights for the next iteration, via $d_{i(t+1)} = \hat{g}_{i(t)}^2 2p_i(1-p_i)$;

Step 5: Normalized SNP weights, readjust the SNP weights to stabilize the total genetic variance via $D_{(t+1)} = \frac{\text{tr}(D_{(1)})}{\text{tr}(D_{(t+1)})} D_{(t+1)}$;

Step 6: Calculated G for the next iteration, via $G_{(t+1)} = \lambda ZD_{(t+1)}Z'$;

Step 7: Let $t = t + 1$ and iterate from step 2.

This procedure was run for three iterations based on the predicted accuracies of GEBV according to Legarra et al. [27] and Zhang et al. [28], and was used by Wang et al. [9] and Marques et al. [5]. The SNP weights, G matrices, GEBV, and marker effects were updated at each iteration. Marker effects obtained from the third iteration were used to calculate the proportion of genetic variance explained by subsets of consecutive SNPs. The SNP window consisted of a region of consecutive SNPs located within 0.4 Mb, which is the average haplotype block size in commercial pig lines' mated SNP effects [29,30] of Duroc pigs in this study. The genetic variance explained by the i^{th} set of consecutive SNPs (i^{th} SNP window) was calculated via:

$$\frac{\text{Var}(a_i)}{\sigma_a^2} \times 100\% = \frac{\text{Var}\left(\sum_{j=1}^m Z_j g_j\right)}{\sigma_a^2} \times 100\% \quad (3)$$

where a_i is the genetic variance of the i^{th} SNP window, σ_a^2 is the total additive genetic variance, Z_j is the vector of the j^{th} SNP for all individuals and g_j is the effect of the j^{th} SNP within the i^{th} window. Manhattan plots of these windows were shown using the R software.

2.4. Candidate Gene Detection and Functional Enrichment Analysis

QTL regions were selected according to the genetic variance of chromosome windows. Windows explaining more than 1% genetic variance were selected as candidate QTL regions, within which candidate genes were searched. The threshold of 1% was chosen based on the literature [14,31,32] and the expected contribution of SNP windows. The expected proportion of average genetic variance explained by each window was 0.02% for the pigs (100/5039). The first three windows for single semen traits that explained the largest number of genetic variances were further extended to 0.4 Mb flanking regions of the midpoints both upstream and downstream.

Genome annotations were based on the gene database *Sus scrofa* 11.1 (<http://www.ensembl.org>, accessed on 11 October 2022). For all the candidate genes, we manually searched they National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov>, accessed on 11 October 2022) to see if they had a previously identified relationship with the traits under study. Gene Ontology (GO) [33] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [34] were used for functional enrichment analysis of candidate genes.

3. Results

3.1. Descriptive Statistics and Genetic Parameters for the Semen Traits

Descriptive statistics of phenotypes for all semen traits are given in Table 1. The coefficients of variation (CV) for SPMOT, SPPMOT, SPABR, and SPCOUNT traits were

15.61%, 68.01%, 43.72%, and 53.36%, respectively. Among the semen traits, the coefficient of variation in sperm motility was the smallest. Semen records per animal min were 5, maximum 103, and mean \pm SD 43.9 ± 26.2 . A distribution histogram with ejaculation times is shown in Figure S1.

In order to better understand the genetic structure of semen traits, we used two methods (BLUP and ssGBLUP) to estimate variance components, the latter to calculate genetic and phenotypic correlation. The variance components of semen traits are shown in Table 2, and the genetic and phenotypic correlations are shown in Figure S2. Among all semen traits, the heritability estimated by ssGBLUP was lower than that estimated by BLUP, and the repeatability of the two methods almost the same. Among them, the heritability and repeatability of SPABR were the highest. There was a very strong negative genetic correlation between SPMOT and SPABR traits: -0.7351 .

Table 2. Genetic parameters for the semen traits of Duroc pigs.

Trait	Models	σ_a^2 (SE)	σ_p^2 (SE)	σ_e^2 (SE)	h^2 (SE)	r_e (SE)
SPMOT	BLUP	50.472 \pm 11.526	15.672 \pm 7.563	110.890 \pm 1.026	0.285 \pm 0.058	0.374 \pm 0.020
	ssGBLUP	29.428 \pm 6.584	34.956 \pm 4.471	110.860 \pm 1.025	0.168 \pm 0.034	0.367 \pm 0.018
SPPMOT	BLUP	67.082 \pm 14.463	7.460 \pm 9.143	258.340 \pm 2.394	0.202 \pm 0.040	0.224 \pm 0.016
	ssGBLUP	39.267 \pm 8.315	31.831 \pm 5.095	258.310 \pm 2.394	0.119 \pm 0.023	0.216 \pm 0.014
SPABR	BLUP	65.056 \pm 17.395	47.887 \pm 12.085	71.246 \pm 0.658	0.353 \pm 0.082	0.613 \pm 0.019
	ssGBLUP	44.741 \pm 11.239	67.419 \pm 7.871	71.246 \pm 0.658	0.244 \pm 0.054	0.612 \pm 0.017
SPCOUNT	BLUP	9436.900 \pm 2092.000	2706.900 \pm 1358.500	22884.000 \pm 213.240	0.269 \pm 0.053	0.347 \pm 0.020
	ssGBLUP	6116.500 \pm 1247.900	5610.700 \pm 782.780	22881.000 \pm 213.200	0.177 \pm 0.032	0.339 \pm 0.018

σ_a^2 , Genetic variance; σ_p^2 , variance of environmental effect; σ_e^2 , residual variance; h^2 , heritability; r_e , Repeatability; SE, standard error.

3.2. WssGWAS Results of Semen Traits

In this study, we show the proportion of variances explained by each 0.4 Mb window for semen traits of Duroc pigs (Figure 1). The three most important QTL regions and the candidate genes are shown in Table 3. The three most important windows of SPMOT, SPPMOT, SPABR and SPCOUNT explained 12.45%, 9.77%, 15.80%, and 12.15% of the genetic variance of each trait, respectively (Figure 1 and Table 3).

Table 3. Three most important QTL regions and candidate genes for semen traits of Duroc pigs.

Traits ^a	Chr ^b	Position (Mb)	gVar (%) ^c	Nsnp	Candidate Genes
SOMOT	18	14.23–14.60	6.18	13	STRA8
	12	58.83–59.21	3.31	10	ZSWIM7, TEKT3, UBB
	2	6.19–6.58	3.87	6	CATSPER1
SPPMOT	4	121.17–121.57	3.84	10	PTBP2
	18	14.23–14.60	3.10	13	STRA8
SPABR	12	62.16–62.44	2.84	6	-
	12	58.83–59.21	8.07	10	ZSWIM7, TEKT3, UBB
	7	97.93–98.26	3.88	10	EIF2B2, MLH3
SPCOUNT	12	46.05–46.45	3.84	7	-
	2	17.69–18.09	6.21	8	-
	11	15.98–16.37	3.86	10	CCDC70
	2	15.98–16.09	2.08	3	-

^a SPMOT: sperm motility; SPPMOT: sperm progressive motility; SPABR: sperm abnormality rate; SPCOUNT: total sperm count. Within each trait, genomic regions were decreasingly sorted based on the proportion of genetic variance explained; ^b Chr: chromosome; ^c gVar (%): proportion of genetic variance explained by 0.4 Mb.

A total of 112 candidate genes were detected in the QTL regions of SPMOT, SPPMOT, SPABR and SPCOUNT traits (Table S1), of which 9 genes were reported to be associated with mammal spermiogenesis, testes functioning, and male fertility (Table 3). Furthermore,

16, 17, 12, and 13 QTL regions (windows that explained more than 1% of total genetic variances) were found for SPMOT, SPPMOT, SPABR and SPCOUNT (Figure S3).

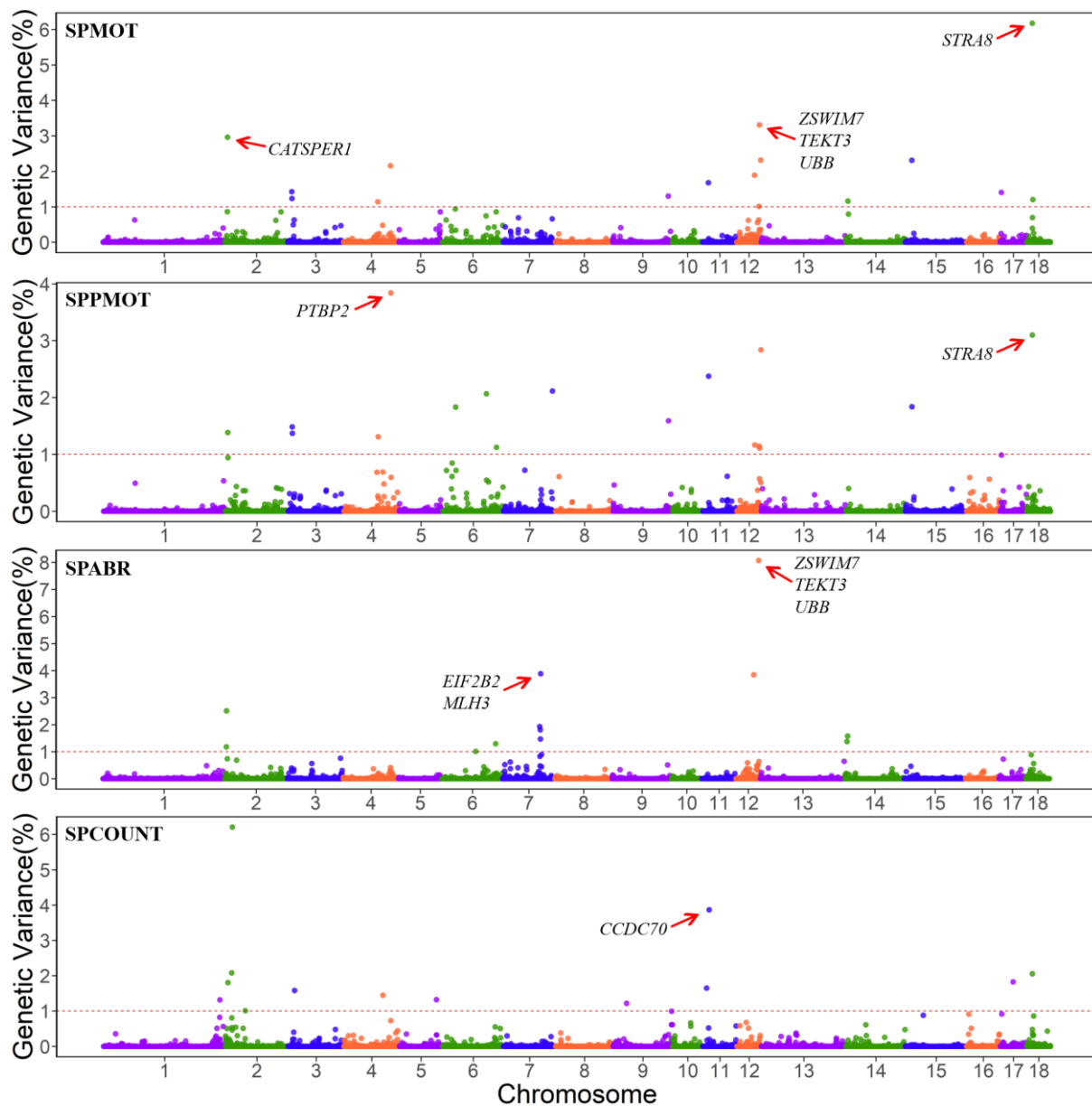


Figure 1. Manhattan plot for GWAS results of semen traits in Duroc pigs.

3.3. GO Terms and KEGG Pathway Enrichment Analysis

The GO terms used to enrich identified genes, with a total of four GO terms related to semen traits, are shown in Table S2. The reproductive process, reproduction, meiotic cell cycle, and microtubule-based movement are important processes of the male reproductive process (Figure 2).

3.4. Association Network Diagram between GO Terms

The association network diagram of the GO terms of genes' biological processes can be seen in Figure 3. The biological processes of the reproductive process cilium assembly, reproduction, and meiotic cell cycle work together to affect male meiotic nuclear division. In the mammal, the product of meiotic division of the male germ cell is the spermatozoon, and the union between the male and female germ cells in the process called fertilization results in the formation of a new organism.

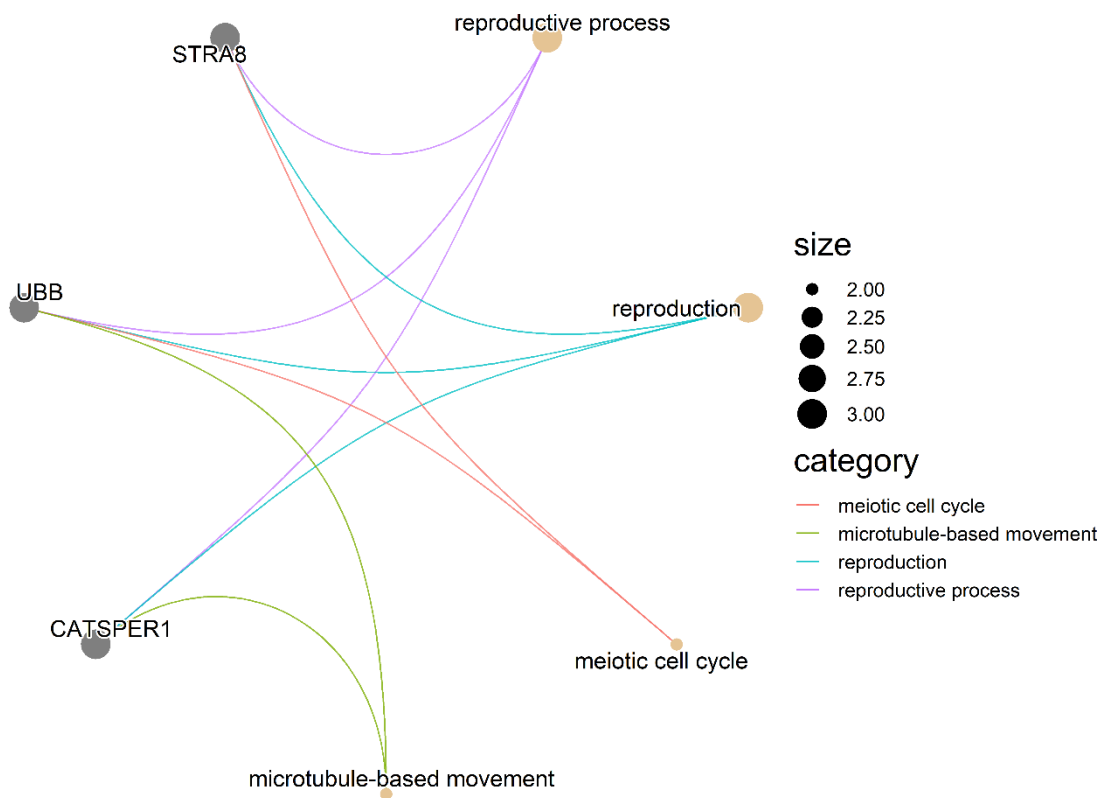


Figure 2. Gene network of biological processes for semen traits of candidate genes. Gray nodes represent candidate genes, yellow nodes represent biological processes, and different biological processes and candidate genes are connected by different colors.

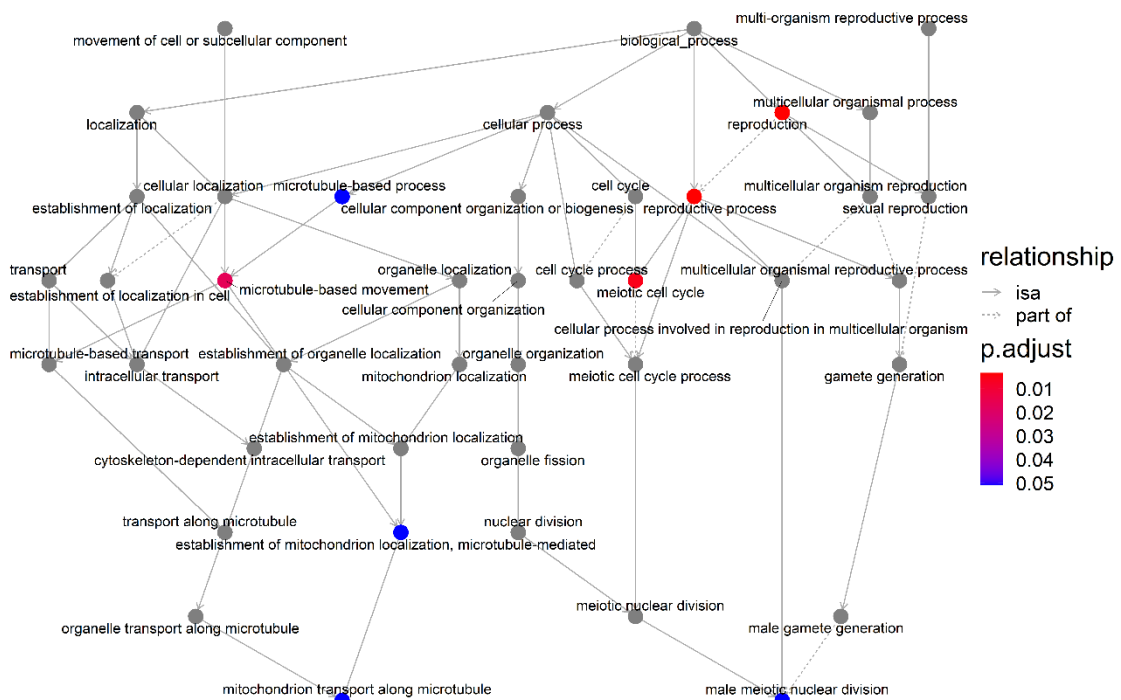


Figure 3. Association network diagram between GO terms of enrichment genes. The solid line represents the complete correlation of the two biological processes, and the dotted line represents the partial correlation. The color represents the size of the adjusted *p* value.

4. Discussion

In this study, we chose the WssGWAS method because of the following advantages. (1) It can integrate all phenotypes, genotypes and pedigree data at the same time, thus avoiding the calculation of the pseudo-phenotypes of genotyped animals to integrate all phenotypic information. It uses information from those without genotypes to improve the statistical power of QTL detection. (2) It allows different weights to be used according to the importance of SNPs, which deviates from the unrealistic assumptions of the GBLUP infinitesimal model and improves the accuracy of SNP effect estimation [9]. (3) It offers the possibility of utilizing SNP windows, the percentage of genetic variance explained by a series of successive SNPs. Continuous SNP windows in GWAS may be more successful in finding QTL regions than single SNP analyses, due to linkage disequilibrium (LD).

The results of variance components showed that the repeatability of ssGBLUP was similar to that of BLUP and the additive variance of ssGBLUP lower than BLUP. Previous studies have shown that the additive variances and heritabilities estimated by the pedigree-based BLUP method may be too high. Compared with BLUP, ssGBLUP has a lower standard deviation [35]. In this study, we also found that the standard error estimated by ssGBLUP was smaller. That the ssGBLUP method uses both pedigree and genotype information to estimate the genetic parameters renders it more accurate in theory. The results of the two methods show that semen traits belong to medium heritability traits and great genetic progress can be obtained through selection. The heritabilities of SPMOT, SPPMOT, SPABR and SPCOUNT traits estimated by ssGBLUP were 0.168, 0.119, 0.244 and 0.177, respectively. This is basically consistent with the results of Gao et al. [6]: 0.160, 0.161, 0.261 and 0.183.

In this study, we identified several QTL regions related to semen traits in Duroc pigs. The search regions of candidate genes are not only limited to the SNP window but also include upstream and downstream flanking regions. It is important to use larger genomic regions to identify genes, because the SNPs in the window may be in the high LD and the QTL in the surrounding regions. We found the same QTL region (chr18:14.23–14.60) in SPMOT and SPPMOT, and determined that the *STRA8* gene is a candidate gene for these two traits. The same QTL region (chr12:58.83–59.21) was found in SPMOT and SPABR, and the candidate genes of *ZSWIM7*, *TEKT3* and *UBB* were identified, which may be related to the high genetic correlation between SPMOT and SPABR (Figure S2). These results confirm the hypothesis of gene pleiotropy.

The *STRA8*, *UBB* and *CATSPER1* genes are enriched into a common biological process and participate in male meiotic spermatogenesis. In order to further confirm the identified candidate genes related to semen traits, further molecular experiments are needed in future research.

For sperm motility, *STRA8*, *ZSWIM7*, *TEKT3*, *UBB*, and *CATSPER1* are significant candidate genes. *STRA8* (stimulated by retinoic acid 8) controls the initiation of meiosis in male germ cells by activating the expression of meiotic genes [36]. The spermatozoa of *STRA8* knockout mice can eventually form sperm cells that cannot complete meiosis [37], which will increase germ cell apoptosis. This is known as the goalkeeper of male meiosis [38]. It has been confirmed that it is an important candidate gene affecting spermatogenesis in mice [39–41] and Atlantic salmon [42]. Recurrent *ZSWIM7* (zinc finger SWIM-type containing 7) mutations lead to human male infertility [43]. The homozygous variation of *ZSWIM7* leads to azoospermia in men and primary ovarian insufficiency in women [44]. *TEKT3* (tektin 3) is a filamentous protein associated with microtubules in cilia, flagella, basal bodies and centrioles. *TEKT3* is necessary for progressive sperm motility in mice [45]. The translocation of *TEKT3* in bull spermatozoa may be related to capacitation or overactivation [46]. The expression of the polyubiquitin gene *UBB* plays an important role in maintaining RNA binding regulatory factors and piRNA-metabolic proteins in testis to complete mouse spermatogenesis [47]. The targeted destruction of the polyubiquitin gene *UBB* leads to male and female infertility in mice, and germ cells are blocked during meiotic prophase I [48]. The *CATSPER1* (cation channel, sperm associated 1) protein is localized in sperm tail

and expressed in human testicular tissue in the form of meiosis and postmeiosis [49]. There is high expression of *CATSPER1* in human azoospermic semen [50]. The downregulation of *CATSPER1* channel in epididymal sperm contributes to the pathogenesis of asthenospermia in rats [51–53].

For sperm progressive motility, *PTBP2* and *STRA8* are significant candidate genes. The splicing regulation of *PTBP2* (polypyrimidine tract binding protein 2) is very important for the communication between germ cells and Sertoli cells (multifunctional somatic cells necessary for spermatogenesis) [54]. The RNA binding protein *PTBP2* plays an important role in the development of male germ cells and spermatogenesis in mice [55,56].

For sperm abnormality rate, *EIF2B2*, *MLH3*, *ZSWIM7*, *TEKT3*, and *UBB* are significant candidate genes. In human male asthenospermia, *EIF2B2* (eukaryotic translation initiation factor 2B subunit beta) is important for sperm motility [57–59]. The loss of function of the DNA mismatch repair gene *MLH3* can lead to male infertility with azoospermia or severe oligozoospermia.

For total sperm count, *CCDC70* is a significant candidate gene. *CCDC70* (coiled-coil domain containing 70) gene is highly expressed in mouse testis, mainly expressed in sperm cells, round sperm group and epididymal epithelial cells, and participates in the regulation of spermatogenesis and epididymal sperm maturation.

5. Conclusions

We used weighted single-step GWAS to identify candidate genes related to sperm motility, sperm progressive motility, sperm abnormality rate, and total sperm count in Duroc boars. These results can be used as genetic markers to improve semen production and quality. This study provides information for further understanding the genetic structure of semen traits in Duroc boars.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/ani13030365/s1>. Figure S1: Distribution histogram with ejaculation times of Duroc pigs. Figure S2: Genetic and phenotypic correlation for semen traits of Duroc pigs; Figure S3: Distribution of the four classes in four semen traits of Duroc pigs; Table S1: All candidate genes for semen traits of Duroc pigs; Table S2: GO terms where the candidate genes were significantly ($p < 0.05$) enriched.

Author Contributions: Conceptualization, X.Z., Z.Z. and H.Z.; methodology, X.Z.; software, X.Z.; validation, W.Z. and T.L.; formal analysis, X.Z. and W.Z.; investigation, T.L.; resources, X.H.; data curation, Q.L.; writing—original draft preparation, X.Z. and W.L.; writing—review and editing, X.Z., Q.L. and H.Z.; visualization, X.Z.; supervision, J.L., X.H. and H.Z.; project administration, Z.Z. and H.Z.; funding acquisition, H.Z. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the earmarked fund for the China Agriculture Research System (CARS-35), and the Guangdong Provincial Key R&D Program (2022B0202090002).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated and/or analyzed during the current study are not publicly available, since the studied population consists of the nucleus herd of Guangdong Guyue Technology Co., Ltd., China, but are available from the corresponding author on reasonable request.

Acknowledgments: We thank the National Supercomputer Center in Guangzhou for its computing support. Additionally, we thank Chao Yue for his help that greatly improved our manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

(1) Sperm motility (SPMOT): the proportion of motile sperm cells to total sperm count when the semen temperature is about 37 °C.

- (2) Sperm progressive motility (SPPMOT): the proportion of sperm cells that move in a straight line.
- (3) Sperm abnormality rate (SPABR): the proportion of sperm cells with abnormal morphology.
- (4) Total sperm count (SPCOUNT): the total number of sperm cells in semen.

References

- Robinson, J.A.; Buhr, M.M. Impact of genetic selection on management of boar replacement. *Theriogenology* **2005**, *63*, 668–678. [[CrossRef](#)] [[PubMed](#)]
- Koketsu, Y.; Sasaki, Y. Boar culling and mortality in commercial swine breeding herds. *Theriogenology* **2009**, *71*, 1186–1191. [[CrossRef](#)] [[PubMed](#)]
- Hirschhorn, J.N.; Daly, M.J. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **2005**, *6*, 95–108. [[CrossRef](#)]
- De, R.; Bush, W.S.; Moore, J.H. Bioinformatics challenges in genome-wide association studies (GWAS). *Methods Mol. Biol.* **2014**, *1168*, 63–81. [[CrossRef](#)]
- Marques, D.B.D.; Bastiaansen, J.W.M.; Broekhuijse, M.; Lopes, M.S.; Knol, E.F.; Harlizius, B.; Guimaraes, S.E.F.; Silva, F.F.; Lopes, P.S. Weighted single-step GWAS and gene network analysis reveal new candidate genes for semen traits in pigs. *Genet. Sel. Evol.* **2018**, *50*, 40. [[CrossRef](#)]
- Gao, N.; Chen, Y.; Liu, X.; Zhao, Y.; Zhu, L.; Liu, A.; Jiang, W.; Peng, X.; Zhang, C.; Tang, Z.; et al. Weighted single-step GWAS identified candidate genes associated with semen traits in a Duroc boar population. *BMC Genomics* **2019**, *20*, 797. [[CrossRef](#)] [[PubMed](#)]
- Zhao, Y.; Gao, N.; Li, X.; El-Ashram, S.; Wang, Z.; Zhu, L.; Jiang, W.; Peng, X.; Zhang, C.; Chen, Y.; et al. Identifying candidate genes associated with sperm morphology abnormalities using weighted single-step GWAS in a Duroc boar population. *Theriogenology* **2020**, *141*, 9–15. [[CrossRef](#)]
- Mei, Q.; Fu, C.; Sahana, G.; Chen, Y.; Yin, L.; Miao, Y.; Zhao, S.; Xiang, T. Identification of new semen trait-related candidate genes in Duroc boars through genome-wide association and weighted gene co-expression network analyses. *J. Anim. Sci.* **2021**, *99*, skab188. [[CrossRef](#)]
- Wang, H.; Misztal, I.; Aguilar, I.; Legarra, A.; Muir, W.M. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res.* **2012**, *94*, 73–83. [[CrossRef](#)]
- Aguilar, I.; Misztal, I.; Johnson, D.L.; Legarra, A.; Tsuruta, S.; Lawlor, T.J. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* **2010**, *93*, 743–752. [[CrossRef](#)]
- Wang, H.; Misztal, I.; Aguilar, I.; Legarra, A.; Fernando, R.L.; Vitezica, Z.; Okimoto, R.; Wing, T.; Hawken, R.; Muir, W.M. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. *Front. Genet.* **2014**, *5*, 134. [[CrossRef](#)] [[PubMed](#)]
- Howard, J.T.; Jiao, S.; Tiezzi, F.; Huang, Y.; Gray, K.A.; Maltecca, C. Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on Duroc Boars. *BMC Genet.* **2015**, *16*, 59. [[CrossRef](#)] [[PubMed](#)]
- Tiezzi, F.; Parker-Gaddis, K.L.; Cole, J.B.; Clay, J.S.; Maltecca, C. A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. *PLoS ONE* **2015**, *10*, e0114919. [[CrossRef](#)]
- Lemos, M.V.; Chiaia, H.L.; Berton, M.P.; Feitosa, F.L.; Aboujaoud, C.; Camargo, G.M.; Pereira, A.S.; Albuquerque, L.G.; Ferrinho, A.M.; Mueller, L.F.; et al. Genome-wide association between single nucleotide polymorphisms with beef fatty acid profile in Nellore cattle using the single step procedure. *BMC Genomics* **2016**, *17*, 213. [[CrossRef](#)] [[PubMed](#)]
- Valente, T.S.; Baldi, F.; Sant’Anna, A.C.; Albuquerque, L.G.; Paranhos da Costa, M.J. Genome-Wide Association Study between Single Nucleotide Polymorphisms and Flight Speed in Nellore Cattle. *PLoS ONE* **2016**, *11*. [[CrossRef](#)]
- Melo, T.P.; de Camargo, G.M.F.; de Albuquerque, L.G.; Carvalheiro, R. Genome-wide association study provides strong evidence of genes affecting the reproductive performance of Nellore beef cows. *PLoS ONE* **2017**, *12*, e0178551. [[CrossRef](#)] [[PubMed](#)]
- Verardo, L.L.; Silva, F.F.; Varona, L.; Resende, M.D.; Bastiaansen, J.W.; Lopes, P.S.; Guimaraes, S.E. Bayesian GWAS and network analysis revealed new candidate genes for number of teats in pigs. *J. Appl. Genet.* **2015**, *56*, 123–132. [[CrossRef](#)]
- Marques, D.B.D.; Lopes, M.S.; Broekhuijse, M.; Guimaraes, S.E.F.; Knol, E.F.; Bastiaansen, J.W.M.; Silva, F.F.; Lopes, P.S. Genetic parameters for semen quality and quantity traits in five pig lines. *J. Anim. Sci.* **2017**, *95*, 4251–4259. [[CrossRef](#)]
- Wang, C.; Li, J.; Wei, H.; Zhou, Y.; Tan, J.; Sun, H.; Jiang, S.; Peng, J. Effects of feeding regimen on weight gain, semen characteristics, libido, and lameness in 170- to 250-kilogram Duroc boars. *J. Anim. Sci.* **2016**, *94*, 4666–4676. [[CrossRef](#)]
- Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [[CrossRef](#)]
- Browning, B.L.; Browning, S.R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* **2009**, *84*, 210–223. [[CrossRef](#)] [[PubMed](#)]

22. Jensen, J.; Mantysaari, E.A.; Madsen, P.; Thompson, R. Residual maximum likelihood estimation of (Co)variance components in multivariate mixed linear models using average information. *J. Indian Soc. Agric. Stat.* **1997**, *49*, 215–236.
23. Misztal, I.; Tsuruta, S.; Strabel, T.; Auvray, B.; Druet, T.; Lee, D.H. BLUPF90 and related programs (BGF90). In Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, 19–23 August 2002; pp. 743–744.
24. Lund, O.F.C.M.S. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* **2010**, *42*, 2. [[CrossRef](#)]
25. VanRaden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [[CrossRef](#)] [[PubMed](#)]
26. Aguilar, I.; Misztal, I.; Legarra, A.; Tsuruta, S. Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. *J. Anim. Breed. Genet.* **2011**, *128*, 422–428. [[CrossRef](#)] [[PubMed](#)]
27. Legarra, A.; Robert-Granie, C.; Manfredi, E.; Elsen, J.M. Performance of genomic selection in mice. *Genetics* **2008**, *180*, 611–618. [[CrossRef](#)] [[PubMed](#)]
28. Zhang, X.; Lourenco, D.; Aguilar, I.; Legarra, A.; Misztal, I. Weighting strategies for single-step genomic BLUP: An iterative approach for accurate calculation of GEBV and GWAS. *Front. Genet.* **2016**, *7*, 151. [[CrossRef](#)]
29. Veroneze R, L.P.; Guimarães, S.E.; Silva, F.F.; Lopes, M.S.; Harlizius, B.; Knol, E.F. Linkage disequilibrium and haplotype block structure in six commercial pig lines. *J. Anim. Sci.* **2013**, *91*, 3493–3501. [[CrossRef](#)]
30. Amaral, A.J.; Megens, H.J.; Crooijmans, R.P.; Heuven, H.C.; Groenen, M.A. Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics* **2008**, *179*, 569–579. [[CrossRef](#)]
31. Gonzalez-Pena, D.; Gao, G.; Baranski, M.; Moen, T.; Cleveland, B.M.; Kenney, P.B.; Vallejo, R.L.; Palti, Y.; Leeds, T.D. Genome-wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in rainbow trout (*Oncorhynchus mykiss*). *Front. Genet.* **2016**, *7*, 203. [[CrossRef](#)]
32. Irano, N.; de Camargo, G.M.; Costa, R.B.; Terakado, A.P.; Magalhaes, A.F.; Silva, R.M.; Dias, M.M.; Bignardi, A.B.; Baldi, F.; Carvalheiro, R.; et al. Genome-wide association study for indicator traits of sexual precocity in Nelore cattle. *PLoS ONE* **2016**, *11*, e0159502. [[CrossRef](#)] [[PubMed](#)]
33. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)] [[PubMed](#)]
34. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic. Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)] [[PubMed](#)]
35. Odegard, J.; Meuwissen, T.H. Estimation of heritability from limited family data using genome-wide identity-by-descent sharing. *Genet. Sel. Evol.* **2012**, *44*, 16. [[CrossRef](#)]
36. Sun, S.; Jiang, Y.; Zhang, Q.; Pan, H.; Li, X.; Yang, L.; Huang, M.; Wei, W.; Wang, X.; Qiu, M.; et al. Znhit1 controls meiotic initiation in male germ cells by coordinating with Stra8 to activate meiotic gene expression. *Dev. Cell.* **2022**, *57*, 901–913. [[CrossRef](#)]
37. Sinha, N.; Whelan, E.C.; Tobias, J.W.; Avarbock, M.; Stefanovski, D.; Brinster, R.L. Roles of Stra8 and Tcerg1l in retinoic acid induced spermatogonial differentiation in mousedagger. *Biol. Reprod.* **2021**, *105*, 503–518. [[CrossRef](#)]
38. Niu, C.; Guo, J.; Shen, X.; Ma, S.; Xia, M.; Xia, J.; Zheng, Y. Meiotic gatekeeper STRA8 regulates cell cycle by interacting with SETD8 during spermatogenesis. *J. Cell. Mol. Med.* **2020**, *24*, 4194–4211. [[CrossRef](#)]
39. Shen, X.; Niu, C.; Guo, J.; Xia, M.; Xia, J.; Hu, Y.; Zheng, Y. Stra8 may inhibit apoptosis during mouse spermatogenesis via the AKT signaling pathway. *Int. J. Mol. Med.* **2018**, *42*, 2819–2830. [[CrossRef](#)]
40. Feng, C.W.; Burnet, G.; Spiller, C.M.; Cheung, F.K.M.; Chawengsaksophak, K.; Koopman, P.; Bowles, J. Identification of regulatory elements required for Stra8 expression in fetal ovarian germ cells of the mouse. *Development* **2021**, *148*, dev194977. [[CrossRef](#)]
41. Gewiss, R.L.; Shelden, E.A.; Griswold, M.D. STRA8 induces transcriptional changes in germ cells during spermatogonial development. *Mol. Reprod. Dev.* **2021**, *88*, 128–140. [[CrossRef](#)]
42. Skaftnesmo, K.O.; Crespo, D.; Kleppe, L.; Andersson, E.; Edvardsen, R.B.; Norberg, B.; Fjellidal, P.G.; Hansen, T.J.; Schulz, R.W.; Wargelius, A. Loss of stra8 Increases Germ Cell Apoptosis but Is Still Compatible With Sperm Production in Atlantic Salmon (*Salmo salar*). *Front. Cell. Dev. Biol.* **2021**, *9*, 657192. [[CrossRef](#)] [[PubMed](#)]
43. Li, Y.; Wu, Y.; Zhou, J.; Zhang, H.; Zhang, Y.; Ma, H.; Jiang, X.; Shi, Q. A recurrent ZSWIM7 mutation causes male infertility resulting from decreased meiotic recombination. *Hum. Reprod.* **2021**, *36*, 1436–1445. [[CrossRef](#)]
44. Hussain, S.; Nawaz, S.; Khan, I.; Khan, N.; Hussain, S.; Ullah, I.; Fakhro, K.A.; Ahmad, W. A novel homozygous variant in homologous recombination repair gene ZSWIM7 causes azoospermia in males and primary ovarian insufficiency in females. *Eur. J. Med. Genet.* **2022**, *65*, 104629. [[CrossRef](#)]
45. Roy, A.; Lin, Y.N.; Agno, J.E.; DeMayo, F.J.; Matzuk, M.M. Tektin 3 is required for progressive sperm motility in mice. *Mol. Reprod. Dev.* **2009**, *76*, 453–459. [[CrossRef](#)] [[PubMed](#)]
46. Tsukamoto, M.; Hiyama, E.; Hirotsu, K.; Gotoh, T.; Inai, T.; Iida, H. Translocation of Tektin 3 to the equatorial segment of heads in bull spermatozoa exposed to dibutyryl cAMP and calyculin A. *Mol. Reprod. Dev.* **2017**, *84*, 30–43. [[CrossRef](#)] [[PubMed](#)]
47. Han, B.; Jung, B.K.; Park, S.H.; Song, K.J.; Anwar, M.A.; Ryu, K.Y.; Kim, K.P. Polyubiquitin gene Ubb is required for upregulation of Piwi protein level during mouse testis development. *Cell. Death Discov.* **2021**, *7*, 194. [[CrossRef](#)] [[PubMed](#)]
48. Sinnar, S.A.; Small, C.L.; Evanoff, R.M.; Reinholdt, L.G.; Griswold, M.D.; Kopito, R.R.; Ryu, K.Y. Altered testicular gene expression patterns in mice lacking the polyubiquitin gene Ubb. *Mol. Reprod. Dev.* **2011**, *78*, 415–425. [[CrossRef](#)]
49. Li, H.G.; Liao, A.H.; Ding, X.F.; Zhou, H.; Xiong, C.L. The expression and significance of CATSPER1 in human testis and ejaculated spermatozoa. *Asian J. Androl.* **2006**, *8*, 301–306. [[CrossRef](#)]

50. Manfredola, F.; Ferraro, B.; Sellitto, C.; Rocco, D.; Fasano, S.; Pierantoni, R.; Chianese, R. CRISP2, CATSPER1 and PATE1 Expression in Human Asthenozoospermic Semen. *Cells*. **2021**, *10*, 1956. [[CrossRef](#)]
51. Wang, Y.N.; Wang, B.; Liang, M.; Han, C.Y.; Zhang, B.; Cai, J.; Sun, W.; Xing, G.G. Down-regulation of CatSper1 channel in epididymal spermatozoa contributes to the pathogenesis of asthenozoospermia, whereas up-regulation of the channel by Sheng-Jing-San treatment improves the sperm motility of asthenozoospermia in rats. *Fertil. Steril.* **2013**, *99*, 579–587. [[CrossRef](#)]
52. Yu, Q.; Mei, X.Q.; Ding, X.F.; Dong, T.T.; Dong, W.W.; Li, H.G. Construction of a catsper1 DNA vaccine and its antifertility effect on male mice. *PLoS ONE* **2015**, *10*, e0127508. [[CrossRef](#)] [[PubMed](#)]
53. Forero-Forero, A.; Lopez-Ramirez, S.; Felix, R.; Hernandez-Sanchez, J.; Tesoro-Cruz, E.; Orozco-Suarez, S.; Murbartian, J.; Soria-Castro, E.; Olivares, A.; Bekker-Mendez, C.; et al. Down Regulation of Catsper1 Expression by Calmodulin Inhibitor (Calmidazolium): Possible Implications for Fertility. *Int. J. Mol. Sci.* **2022**, *23*, 8070. [[CrossRef](#)] [[PubMed](#)]
54. Hannigan, M.M.; Zagore, L.L.; Licatalosi, D.D. Ptpb2 Controls an Alternative Splicing Network Required for Cell Communication during Spermatogenesis. *Cell. Rep.* **2017**, *19*, 2598–2612. [[CrossRef](#)] [[PubMed](#)]
55. Zagore, L.L.; Grabinski, S.E.; Sweet, T.J.; Hannigan, M.M.; Sramkoski, R.M.; Li, Q.; Licatalosi, D.D. RNA Binding Protein Ptpb2 Is Essential for Male Germ Cell Development. *Mol. Cell. Biol.* **2015**, *35*, 4030–4042. [[CrossRef](#)]
56. Liu, W.; Zhao, Y.; Liu, X.; Zhang, X.; Ding, J.; Li, Y.; Tian, Y.; Wang, H.; Liu, W.; Lu, Z. A Novel Meiosis-Related lncRNA, Rbakdn, Contributes to Spermatogenesis by Stabilizing Ptpb2. *Front. Genet.* **2021**, *12*, 752495. [[CrossRef](#)]
57. Lin, Y.; Liang, A.; He, Y.; Li, Z.; Li, Z.; Wang, G.; Sun, F. Proteomic analysis of seminal extracellular vesicle proteins involved in asthenozoospermia by iTRAQ. *Mol. Reprod. Dev.* **2019**, *86*, 1094–1105. [[CrossRef](#)]
58. Xu, K.; Lu, T.; Zhou, H.; Bai, L.; Xiang, Y. The role of MSH5 C85T and MLH3 C2531T polymorphisms in the risk of male infertility with azoospermia or severe oligozoospermia. *Clin. Chim. Acta.* **2010**, *411*, 49–52. [[CrossRef](#)]
59. Nawaz, S.; Ullah, M.I.; Hamid, B.S.; Nargis, J.; Nawaz, M.; Hussain, S.; Ahmad, W. A loss-of-function variant in DNA mismatch repair gene MLH3 underlies severe oligozoospermia. *J. Hum. Genet.* **2021**, *66*, 725–730. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.