

RESEARCH ARTICLE

Open Access



Weighted single-step GWAS and gene network analysis reveal new candidate genes for semen traits in pigs

Daniele B. D. Marques¹, John W. M. Bastiaansen^{2*}, Marleen L. W. J. Broekhuijse³, Marcos S. Lopes^{3,4}, Egbert F. Knol³, Barbara Harlizius³, Simone E. F. Guimarães¹, Fabyano F. Silva¹ and Paulo S. Lopes¹

Abstract

Background: In recent years, there has been increased interest in the study of the molecular processes that affect semen traits. In this study, our aim was to identify quantitative trait loci (QTL) regions associated with four semen traits (motility, progressive motility, number of sperm cells per ejaculate and total morphological defects) in two commercial pig lines (L1: Large White type and L2: Landrace type). Since the number of animals with both phenotypes and genotypes was relatively small in our dataset, we conducted a weighted single-step genome-wide association study, which also allows unequal variances for single nucleotide polymorphisms. In addition, our aim was also to identify candidate genes within QTL regions that explained the highest proportions of genetic variance. Subsequently, we performed gene network analyses to investigate the biological processes shared by genes that were identified for the same semen traits across lines.

Results: We identified QTL regions that explained up to 10.8% of the genetic variance of the semen traits on 12 chromosomes in L1 and 11 chromosomes in L2. Sixteen QTL regions in L1 and six QTL regions in L2 were associated with two or more traits within the population. Candidate genes *SCN8A*, *PTGS2*, *PLA2G4A*, *DNAI2*, *IQCG* and *LOC102167830* were identified in L1 and *NME5*, *AZIN2*, *SPATA7*, *METTL3* and *HPGDS* in L2. No regions overlapped between these two lines. However, the gene network analysis for progressive motility revealed two genes in L1 (*PLA2G4A* and *PTGS2*) and one gene in L2 (*HPGDS*) that were involved in two biological processes i.e. eicosanoid biosynthesis and arachidonic acid metabolism. *PTGS2* and *HPGDS* were also involved in the cyclooxygenase pathway.

Conclusions: We identified several QTL regions associated with semen traits in two pig lines, which confirms the assumption of a complex genetic determinism for these traits. A large part of the genetic variance of the semen traits under study was explained by different genes in the two evaluated lines. Nevertheless, the gene network analysis revealed candidate genes that are involved in shared biological pathways that occur in mammalian testes, in both lines.

Background

Artificial insemination (AI) pig industry focuses mainly on maximizing the number of insemination doses produced from each boar ejaculate. To achieve this goal, the ability of boars to produce high-quality semen (high motility and progressive motility with low levels of

morphological defects) in sufficient quantity (large number of sperm cells per ejaculate) is decisive [1].

In recent years, with the fast advances in high-throughput genotyping and in molecular techniques in general, there is an increased interest in the study of the molecular processes and genetic mechanisms that affect semen traits. Genes and markers associated with pig semen traits have been described in the literature [2–8]. However, very few studies analyze large datasets to identify novel quantitative trait loci (QTL) and to provide a deeper knowledge of the genes that control boar semen

*Correspondence: john.bastiaansen@wur.nl

² Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands

Full list of author information is available at the end of the article



traits. One reason for this is the genetic complexity of the process for the production and maturation of sperm cells. Mammalian spermatogenesis requires coordination among different genes and cell types (germ cells, Sertoli cells, and Leydig cells) [9] and occurs in the seminiferous tubules of the testes in three steps: mitotic phase, meiotic phase, and spermiogenesis [10]. In the first step (mitosis), spermatogonia produce primary spermatocytes, which enter the first stage of meiosis (meiosis I) and produce secondary spermatocytes. Then, the second step of meiosis (meiosis II) leads to the generation of haploid round spermatids. In the last phase, i.e. spermiogenesis, the spermatids undergo morphological transformations and acquire the spermatozoa shape. Then, the new pre-formed spermatozoa go through the epididymis to mature and acquire motility [10]. Mutations and impaired expression of genes that control the whole process of spermatogenesis and sperm maturation can lead to problems in semen quality and fertility.

Genome-wide association studies (GWAS) are commonly used to identify single nucleotide polymorphisms (SNPs) that are associated with QTL with major effects [11]. The weighted single-step GWAS (WssGWAS), proposed by Wang et al. [12], is a method that allows estimation of SNP effects using genomic estimated breeding values (GEBV) from single-step genomic best linear unbiased prediction (ssGBLUP, [13]) based on all phenotyped, genotyped and pedigree-related animals. In addition, it allows unequal variances for SNPs, which results in improved precision of the estimation of SNP effects [12]. Therefore, when the number of animals with both phenotypes and genotypes is small and the traits are controlled by QTL with large effects, the WssGWAS may perform better than traditional GWAS methods. Recent studies have used this method for production, carcass and reproductive traits in livestock [14–23].

In a post-GWAS study, a gene network analysis can be performed for candidate genes related to QTL regions identified in GWAS. The gene network is used to investigate pathways and biological processes that are shared by these genes [24]. In addition, the biological information provided by these gene networks are helpful to understand genetic differences between populations for the same trait [25].

In this study, our aim was to identify QTL regions that are associated with four semen traits (motility, progressive motility, number of sperm cells per ejaculate and total morphological defects) in pigs. In addition, our aim was to identify candidate genes within those QTL regions that explained the highest proportions of genetic variance. To achieve our goal, we performed a WssGWAS in two commercial pig lines (L1: Large White type and L2: Landrace type), followed by gene network analyses to

investigate the biological processes shared by genes that were identified for the same semen traits in these two lines.

Methods

Phenotypic, genotypic and pedigree data

Phenotypic data were available from two commercial pig lines, a Large White type line (L1) and a Landrace type line (L2), on ejaculate samples that were collected between January 2007 and October 2014. The evaluated traits were: (1) sperm motility (MOT), which is the proportion of moving sperm cells in an ejaculate; (2) sperm progressive motility (PROMOT), defined as the proportion of sperm cells that move in a straight line; (3) abnormal sperm cell number (ABN), which is the total number of sperm cells with morphological abnormalities; and (4) the total number of sperm cells in the ejaculate (Ncells per 10^6 sperm cells). MOT and PROMOT were evaluated using the UltiMate™ CASA system (Hamilton Thorne Inc., Beverly, MA, USA). Ncells was calculated as the product of the semen volume (mL) and concentration (10^6 mL⁻¹, measured by the CASA system). The values of this measure were not normally distributed, and thus log-transformed before further analyses (lnNcells). Ejaculates evaluated for ABN were analyzed microscopically at a 1000× magnification by a trained technician with a phase contrast microscope and a thermal plate (BH-2, Olympus, Tokyo, Japan), counting 100 sperm cells per assessment. All semen traits were assessed on the day of semen collection using fresh semen.

The phenotypic data for MOT, PROMOT and lnNcells consisted of 43,455 ejaculates for L1 (866 boars) and 39,161 ejaculates for L2 (900 boars). For ABN, the phenotypic data consisted of 13,366 ejaculates for L1 (849 boars) and 9853 ejaculates for L2 (886 boars). The average number of ejaculates per boar (with standard deviations in parenthesis) for MOT, PROMOT and lnNcells were 50.18 (38.12) for L1 and 43.51 (36.37) for L2. For ABN, the average number of ejaculates per boar were 15.74 (11.19) for L1 and 11.12 (8.99) for L2. Number of boars with phenotypic data, mean, standard deviation, minimum and maximum values of semen traits in L1 and L2 are in Table 1.

Genotyping data for 3737 animals (856 males and 2881 females) from L1 and 3307 animals (953 males and 2354 females) from L2 were available. A majority of the animals (2718 for L1 and 2394 for L2) were genotyped using the Illumina PorcineSNP60 BeadChip (Illumina, Inc., San Diego, CA) but part of the animals from L1 ($n=1019$) and L2 ($n=913$) were genotyped using the (Illumina, Inc.) GeneSeek Custom 80 K SNP chip (GeneSeek Inc., Lincoln, NE). Quality control was performed by

Table 1 Descriptive statistics of semen traits

	Number of boars ^a	Mean ^b	SD ^c	Min ^d	Max ^e
MOT ^f					
L1 ^g	866	86.5	7.1	10	100
L2	900	87.1	6.5	14	100
PROMOT					
L1	866	78.6	8.3	0	100
L2	900	77.4	7.9	0	100
lnN _{cells}					
L1	866	25.0	0.4	24.0	26.4
L2	900	24.9	0.4	22.5	26.4
ABN					
L1	849	19.3	14.8	1	98
L2	886	14.4	12.6	1	99

^a Number of boars with phenotypic data
^b Mean values of semen traits for each pig line
^c Standard deviation
^d Minimum trait value of semen traits
^e Maximum trait value of semen traits
^f Semen traits: MOT: sperm motility; PROMOT: sperm progressive motility; lnN_{cells}: number of sperm cells per ejaculate; ABN: total morphological abnormalities
^g Pig lines: L1 = Large White and L2 = Landrace

excluding SNPs that had an unknown position on the pig genome build 10.2 [26], that were located on sex chromosomes, that had a call rate lower than 0.95 or a minor allele frequency lower than 0.01, or that were in strong deviation from Hardy–Weinberg equilibrium ($\chi^2 > 600$). Animals for which the frequency of missing genotypes was higher than 0.05 were also excluded. After quality control, missing genotypes from animals genotyped with the SNP60 BeadChip were imputed with the software Beagle version 3.3.2 [27] to the set of SNPs on the SNP60 BeadChip that passed quality control. In addition, genotypes from the animals genotyped with the GeneSeek Custom 80 K SNP chip were imputed to the set of SNPs on the SNP60 BeadChip that passed quality control. After quality control, 39,945 and 41,478 SNPs remained for L1 and L2, respectively, and were used for the GWAS.

The complete pedigree included 8352 animals for L1 and 8271 animals for L2. The total number of animals that remained after pedigree pruning was 6724 for L1 and 6502 for L2. Most animals had either phenotypic or genotypic data. The number of animals with both phenotypes and genotypes was 349 for L1 and 446 for L2.

Statistical analyses

The weighted ssGBLUP analysis was conducted within line using the BLUPF90 software family [28] adapted for genomic analyses [29]. First, variance components were

estimated using AIREMLF90, which were then used in BLUPF90 to predict GEBV. SNP effects were then calculated using postGSf90 software.

The single-trait animal model for ssGBLUP was as follows:

$$y = X\beta + Za + Wp + \epsilon,$$

where **y** is the vector of phenotypic observations; **β** is the vector of fixed effects (combined effects of AI station, year and month of semen collection; the laboratory where the samples were analyzed and the covariates of interval between two subsequent semen collections in days and age of the boar at the time of collection in months); **a** is the vector of random additive genetic effects; **p** is the vector of random permanent environmental effects; **ε** is the vector of random residuals; and **X**, **Z** and **W** are the incidence matrices of **β**, **a** and **p**, respectively.

It was assumed that $\mathbf{a} \sim N(0, \mathbf{H}\sigma_a^2)$, $\mathbf{p} \sim N(0, \mathbf{I}\sigma_p^2)$ and $\epsilon \sim N(0, \mathbf{I}\sigma_e^2)$, where σ_a^2 , σ_p^2 and σ_e^2 are the additive genetic, permanent environmental and residual variances, respectively; **H** is the matrix that combines pedigree and genomic information [13], and **I** is an identity matrix. The inverse of **H** needed for mixed model equations is given by:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix},$$

where **A** is the numerator relationship matrix based on pedigree for all animals; **A**₂₂ is the numerator relationship matrix for genotyped animals; and **G** is the genomic relationship matrix [30]. This matrix was obtained as follows:

$$\mathbf{G} = \frac{\mathbf{ZDZ}'}{\sum_{i=1}^M 2p_i(1-p_i)},$$

where **Z** is a matrix of gene content adjusted for allele frequencies (0, 1 or 2 for *aa*, *Aa* and *AA*, respectively); **D** is a diagonal matrix of weights for SNP variances (initially **D** = **I**); *M* is the number of SNPs, and *p*_{*i*} is the minor allele frequency of *i*-th SNP.

Estimates of SNP effects and weights for WssGWAS were obtained according to Wang et al. [12] by the following steps:

1. In the first iteration (*t* = 1): **D** = **I**; **G**_(*t*) = **D**_(*t*)**Z'** λ , where $\lambda = \frac{1}{\sum_{i=1}^M 2p_i(1-p_i)}$ [30];
2. GEBV were calculated for the entire dataset using ssGBLUP;

3. GEBV were converted to estimates of SNP effects ($\hat{\mathbf{u}}$): $\hat{\mathbf{u}}_{(t)} = \lambda \mathbf{D}_{(t)} \mathbf{Z}' \mathbf{G}_{(t)}^{-1} \hat{\mathbf{a}}_g$, where $\hat{\mathbf{a}}_g$ is the GEBV of animals that were also genotyped;
4. The weight for each SNP to be used in the next iteration was calculated as: $d_{i(t+1)} = \hat{u}_{i(t)}^2 2p_i(1 - p_i)$, where i is the i -th SNP;
5. The SNP weights were normalized to keep the total genetic variance constant:

$$\mathbf{D}_{(t+1)} = \frac{\text{tr}(\mathbf{D}_{(1)})}{\text{tr}(\mathbf{D}_{(t+1)})} \mathbf{D}_{(t+1)};$$

6. $\mathbf{G}_{(t+1)} = \mathbf{ZD}_{(t+1)}\mathbf{Z}'\lambda$ was calculated;
7. $t = t + 1$ and loop to step 2.

This procedure was run for three iterations based on the realized accuracies of GEBV according to Legarra et al. [31] and performed by Wang et al. [14]. At each iteration, the weights for SNPs were updated (steps 4 and 5), used to construct the \mathbf{G} matrices (step 6), update the GEBV (step 2) and, consequently, the estimated SNP effects (step 3). The percentage of genetic variance explained by the i -th set of consecutive SNPs (i -th SNP window) was calculated as described by Wang et al. [14] as:

$$\frac{\text{Var}(a_i)}{\sigma_a^2} \times 100\% = \frac{\text{Var}\left(\sum_{j=1}^x \mathbf{Z}_j \hat{u}_j\right)}{\sigma_a^2} \times 100\%,$$

where a_i is the genetic value of the i -th SNP window that consists of a region of consecutive SNPs located within 0.4 Mb, which is the average haplotype block size in commercial pig lines [32], including the lines considered in the present study; σ_a^2 is the total additive genetic variance; \mathbf{Z}_j is the vector of gene content of the j -th SNP for all individuals and \hat{u}_j is the effect of the j -th SNP within the i -th window. Manhattan plots showing these windows were created using the R software [33].

Selection of relevant SNP windows, search for candidate genes, and gene network analyses

The selection of relevant SNP windows and the search for genes within these QTL regions were performed in three steps. First, for the four traits (MOT, PROMOT, lnNcells and ABN), the SNP windows that explained 1% or more of the total genetic variance based on the Wss-GWAS were selected within each line (L1 and L2). The threshold of 1% was chosen based on the literature [19, 20, 34] and on the expected contribution of SNP windows [35]. For example, assuming an equal contribution

of all windows (on average, we had 4223 and 4229 windows for each trait in L1 and L2, respectively), the expected proportion of genetic variance explained by each window was 0.02% for both lines (100/4223 for L1 and 100/4229 for L2). The threshold of 1% used in the present study is equal to 50 times the expected variance (0.02% \times 50 = 1%) and considered a suitable threshold for our purposes. Then, after selecting the windows that explained more than 1% of the genetic variance, a search for overlapping windows for two or more traits of the same line was performed. Windows were considered to overlap if their midpoints were less than 0.4 Mb apart. Second, the three most important windows (that explained the highest proportion of genetic variance) for each trait and the 0.4 Mb region on either side of these windows midpoints were also identified. Third, based on the windows selected in the first step (> 1% of variance explained), overlapping windows for the same traits, but across lines, were investigated (also considering a maximum distance of 0.4 Mb between midpoints).

Based on the start and end positions of each identified and selected QTL region in the third step, we identified the genes in these QTL regions based on the Gene database for *Sus scrofa* available at "National Center for Biotechnology Information" (NCBI, [36]). For all the identified genes, we manually searched the literature if they had a previously identified relationship with the traits under study. In addition, based on human genes with the same description, we carried out four gene network analyses that described the biological processes and relations between the L1 and L2 sets of genes that were identified for the same traits by using the ClueGO and CluePedia Cytoscape plug-ins [37, 38]. The ClueGO plug-in combines Gene Ontology (GO) terms and KEGG/BioCarta pathways and develops a GO/pathway network. It also calculates enrichment and depletion tests for groups of genes based on the hypergeometric distribution and corrects the P-values for multiple testing [37]. Using the CluePedia plug-in, new associations between genes can be discovered with enrichments and added to the ClueGO pathways [38].

Results

Estimates of variance components for all semen traits in both lines are in Additional file 1: Table S1. Low to moderate heritabilities ranging from 0.10 to 0.31 were estimated (Table 2), with higher estimates in L1 than L2 for MOT, PROMOT and ABN.

For L1, the three most important windows for MOT, PROMOT, lnNcells and ABN explained 14.4, 18.2, 15.67 and 18.8% of the genetic variance, respectively (Fig. 1) and Additional file 2: Table S2. For L2, the three most important windows explained 21.9, 18.7, 18.3 and 13.8%

Table 2 Estimates (standard error) of heritabilities for the evaluated semen traits in two lines

Semen traits ^a	Pig lines ^b	
	L1	L2
MOT	0.21 (0.05)	0.12 (0.04)
PROMOT	0.31 (0.05)	0.12 (0.03)
lnNcells	0.10 (0.03)	0.13 (0.03)
ABN	0.22 (0.05)	0.14 (0.05)

^a MOT: sperm motility; PROMOT: sperm progressive motility; lnNcells: number of sperm cells per ejaculate; ABN: total morphological abnormalities

^b L1: Large White type and L2: Landrace type

of the genetic variance of each trait, respectively (Fig. 2) and Additional file 3: Table S3. For L1, 20 relevant QTL regions (single and overlapping) were found on SSC1 (SSC for *Sus scrofa* chromosome), 3, 4, 5, 6, 8, 9, 10, 12, 13, 14, and 15 (Table 3). For L2, 16 relevant QTL regions were located on SSC1, 2, 3, 6, 7, 8, 9, 11, 13, 15 and 18 (Table 4). For L1, 110 genes located in the relevant QTL regions were identified, of which six genes were described to be related to MOT, PROMOT, lnNcells and ABN in the literature (Table 3). For L2, 106 genes were found in the detected relevant regions, of which five genes were related to MOT, PROMOT, lnNcells and ABN based on literature (Table 4).

We found no overlapping QTL regions between the two lines for the same traits in this study. Nevertheless, the gene network analysis for PROMOT revealed two genes in L1 (*PLA2G4A* and *PTGS2*) and one gene in L2 (*HPGDS*) that shared the biological processes of eicosanoid biosynthesis and arachidonic acid metabolism. The genes *PTGS2* in L1 and *HPGDS* in L2 were also found to share the biological process of cyclooxygenase pathway (Fig. 3).

Discussion

In the past years, interest in investigating genomic regions that control boar semen traits has increased due to advances in molecular and genotyping techniques, statistical methods, and the ease of GWAS application. According to Wang et al. [12], two of the most used GWAS methods are (1) single-SNP GWAS, which fits each SNP separately as a fixed effect in a model that accounts for population stratification; and (2) Bayesian methods, which analyze all SNPs at the same time. Nonetheless, when the number of animals that are both phenotyped and genotyped is relatively small, the application of single-SNP GWAS and Bayesian methods is limited, since the calculation of pseudo-phenotypes (i.e. deregressed breeding values [39]) is necessary [14]. An alternative method is single-step GWAS (ssGWAS), which was proposed by Wang et al. [12]

and converts the GEBV of genotyped animals obtained by ssGBLUP into SNP effects. However, the ssGWAS model, which assumes equal variance for all marker effects, may be limited when traits are affected by large QTL [12, 40]. To overcome this limitation, Wang et al. [12] proposed the WssGWAS approach, in which SNP effects are weighted according to their importance for the trait of interest, which improves QTL detection [12].

To perform the WssGWAS, first, the \mathbf{H}^{-1} matrix is calculated, by combining all known pedigree and genotype information, and then used in the ssGBLUP procedure to estimate GEBV for all animals. Then, the GEBV of the genotyped animals are used to estimate effects for the SNPs. Finally, SNP effects are used to calculate the percentage of genetic variance that is explained by sets of consecutive SNPs (SNP windows). In this approach, the SNP effects are not directly estimated from the model and there are no measures of uncertainty for the statistical tests. However, the WssGWAS provides information on the most important SNP windows, based on the proportion of explained genetic variance, which is a general concept that is widely accepted in QTL detection analyses [12–14] but does not allow for a formal testing of significance. In this study, we chose the WssGWAS approach because: (1) it can integrate all phenotypic, genotypic and pedigree data simultaneously, thus avoiding the need to calculate pseudo-phenotypes for genotyped animals to incorporate all phenotypic information; (2) it allows the use of different weights for SNPs according to their importance, which is a deviation from the non-realistic GBLUP assumption of the infinitesimal model and improves the precision of estimates of SNP effects; and (3) it provides the possibility to work with SNP windows, since a window of consecutive SNPs in the GWAS may be more successful in finding QTL regions compared to individual SNP analysis because of linkage disequilibrium (LD). In general, analyses that consider the association of individual SNPs may overestimate the number of detected QTL [41].

In spite of the small number of animals that were both genotyped and phenotyped (349 for L1 and 446 for L2), all the information of the 3737 and 3307 genotyped animals, the 866 and 900 phenotyped boars (849 and 886 for ABN), and the 6724 and 6502 animals in the pedigree file, for L1 and L2, respectively, were used to calculate GEBV, and consequently, to estimate SNP effects in the WssGWAS. Obtaining a large dataset of both genotyped and phenotyped animals is unlikely or difficult for semen traits because phenotypes can be recorded only for animals that are used for AI. The number of males and females genotyped in this study also shows the relatively small number of boars, 856 and 953 for L1 and L2, respectively, compared to the number of females, 2881 and 2354. In this context, the advantage of WssGWAS

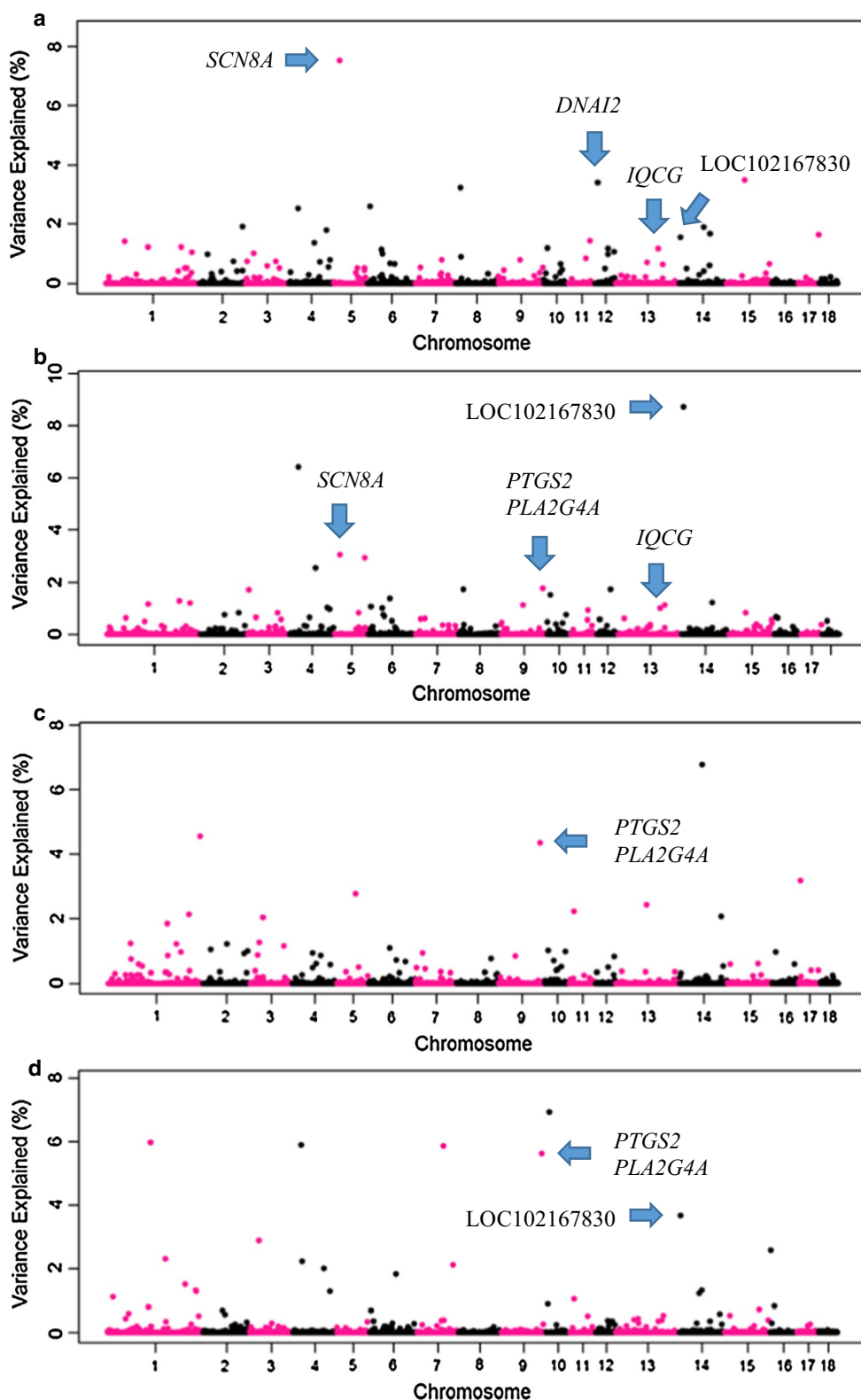


Fig. 1 GWAS results of semen traits in the Large White type line (L1). **a** Sperm motility, **b** sperm progressive motility, **c** number of sperm cells per ejaculate, **d** total morphological abnormalities. Each dot represents one SNP window of 0.4 Mb. On the y-axis is the percentage of genetic variance explained by windows

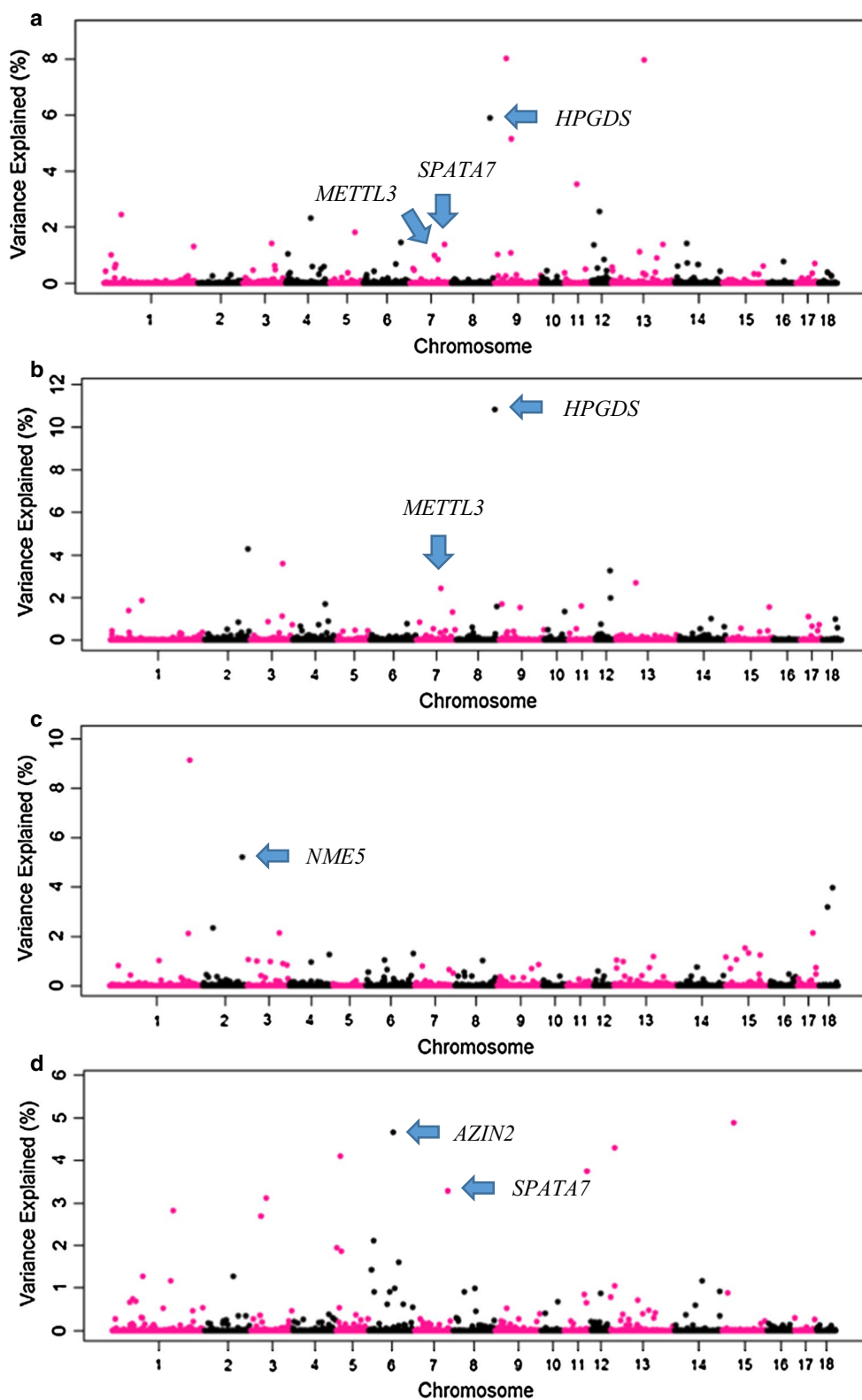


Fig. 2 GWAS results of semen traits in the Landrace type line (L2). **a** Sperm motility, **b** sperm progressive motility, **c** number of sperm cells per ejaculate, **d** total morphological abnormalities. Each dot represents one SNP window of 0.4 Mb. On the y-axis is the percentage of genetic variance explained by windows

Table 3 Individual and overlapping QTL regions for semen traits in L1

Chr ^a	QTL region (Mb) ^b	Nb SNP ^c	Var (%) ^d MOT ^e	Var (%) PROMOT	Var (%) InNcells	Var (%) ABN	Candidate gene ^g
1	135.51–136.31	17	1.2	1.2	– ^f	6.0	– ^h
1	255.48–256.28	17	1.2	–	–	1.5	–
1	290.90–291.84	25	1.0	–	–	1.3	–
1	305.18–305.98	23	–	–	4.6	–	–
3	28.53–29.33	18	1.0	–	–	2.9	–
4	28.25–29.05	12	2.5	6.4	–	5.9	–
4	84.90–85.73	21	1.4	2.5	–	–	–
4	123.12–124.20	30	1.8	1.0	–	1.3	–
5	17.61–18.47	26	7.5	3.0	–	–	SCN8A
6	8.24–9.13	20	2.6	1.1	–	–	–
8	16.06–16.86	20	3.2	1.7	–	–	–
9	139.53–140.63	23	–	1.8	4.3	5.6	PTGS2, PLA2G4A
10	10.58–11.45	23	1.2	1.5	1.0	6.9	–
12	6.23–7.03	32	3.4	–	–	–	DNAI2
12	40.76–41.56	17	1.2	1.7	–	–	–
13	143.61–144.69	13	1.2	1.0	–	–	IQCG
14	4.13–5.22	19	1.5	8.7	–	3.7	LOC102167830
14	72.83–73.63	16	–	–	6.8	–	–
14	99.70–100.51	25	1.7	1.2	–	–	–
15	61.93–62.73	15	3.5	–	–	–	–

^a Chromosome^b Position of QTL region^c Number of SNPs within the QTL region^d Percentage of genetic variance explained by the QTL region^e Semen traits: MOT: sperm motility; PROMOT: sperm progressive motility; InNcells: number of sperm cells per ejaculate; ABN: total morphological abnormalities^f The percentage of genetic variance explained by the QTL region is < 1%. When the variance is reported for more than one trait, the QTL region is overlapping across traits^g Best candidate gene(s) in the region^h No candidate genes associated with the trait

is to supply additional information on relatives without genotypes in order to improve the statistical power of QTL detection [12].

In our previous study [7], a classical GWAS (single-SNP GWAS) was performed for sperm motility in a subset of the L1 and L2 populations that were used in the present study: 645 and 1886 phenotyped and genotyped animals for L1, respectively, and 760 and 1972 animals for L2. Due to the small number of animals that were both genotyped and phenotyped, deregressed breeding values [39] were calculated and included as response variables in the GWAS model, a polygenic effect was added in the model to account for population structure, and a genome-wide false discovery rate (FDR) was applied (q-values) to avoid false positives due to multiple testing. The results showed no SNPs with a significant association (q-value ≤ 0.05) with sperm motility for L1 but six SNPs associated with the trait for L2 (SSC1, from 117.26 to 119.50 Mb). In this QTL region on SSC1, a novel candidate gene

(*MTFMT*) that affects translation efficiency of proteins in sperm cells, was identified. The number of relevant QTL regions that was identified for MOT in the present study was greater than in [7], which indicates that the WssGWAS was more successful in detecting QTL. All the recommended steps for the GWAS method used were performed in each study and reliable candidate genes with biological meaning were found. In addition, Marques et al. [42] reported high genetic correlations between MOT, PROMOT and ABN, which corroborate our findings regarding the overlapping QTL regions that explained more than 1% of the genetic variance for these traits. Therefore, we believe that the QTL results of the present and previous studies should not be considered as false positives and we conclude that the single-SNP method with pseudo-phenotypes and the amount of data in the previous study were not able to identify some of the relevant QTL related to MOT. For L1, we identified overlapping QTL regions for MOT and other semen

Table 4 Individual and overlapping QTL regions for semen traits in L2

Chr ^a	QTL region (Mb) ^b	Nb SNP ^c	Var (%) ^d MOT ^e	Var (%) PROMOT	Var (%) InNcells	Var (%) ABN	Candidate gene ^g
1	270.94–271.74	10	< 1 ^f	–	9.1	–	– ^h
1	55.61–56.47	26	2.4	1.4	–	–	–
2	145.69–146.49	16	–	–	5.2	–	<i>NME5</i>
2	154.03–154.83	23	–	4.3	–	–	–
3	110.29–111.09	19	–	3.6	–	–	–
6	83.32–84.12	12	–	–	–	4.7	<i>AZIN2</i>
7	116.37–117.28	25	1.4	–	–	3.3	<i>SPATA7</i>
7	82.56–83.36	10	1.0	2.4	–	–	<i>METTL3</i>
8	133.90–134.94	20	5.9	10.8	–	–	<i>HPGDS</i>
9	36.46–37.26	12	8.0	–	–	–	–
9	9.32–10.31	19	1.0	1.7	–	–	–
11	41.05–41.85	11	3.5	1.6	–	–	–
13	11.35–12.15	19	–	–	–	4.3	–
13	107.48–108.28	10	8.0	–	–	–	–
15	37.17–37.97	20	–	–	–	4.9	–
18	42.80–43.60	19	–	–	4.0	–	–

^a Chromosome

^b Position of QTL region

^c Number of SNPs within the QTL region

^d Percentage of genetic variance explained by the QTL region

^e Semen traits: MOT: sperm motility; PROMOT: sperm progressive motility; InNcells: number of sperm cells per ejaculate; ABN: total morphological abnormalities

^f The percentage of genetic variance explained by the QTL region is < 1%. When the variance is reported for more than one trait, the QTL region is overlapping across traits

^g Best candidate gene(s) in the region

^h No candidate genes associated with the trait

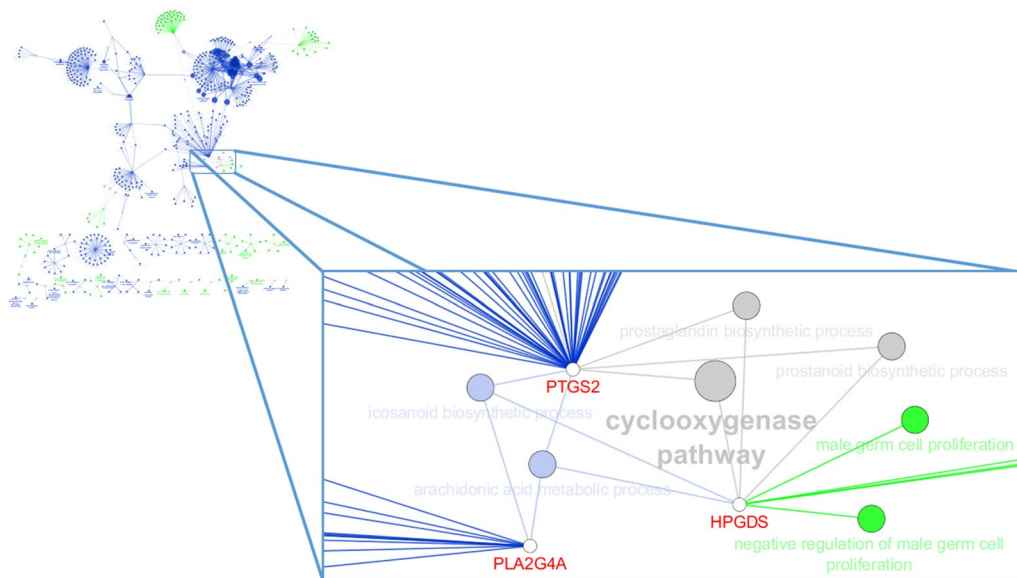


Fig. 3 Gene network of biological processes for progressive motility. Complete network and important shared pathways (with zoom) are shown. Blue color indicates pathways for the Large White type line (L1) and green color indicates pathways for the Landrace type line (L2). Processes shared by *PTGS2* and *PLA2G4A* genes (L1) and *HPGDS* gene (L2) are connected by blue nodes. Processes shared by *PTGS2* and *HPGDS* are connected by grey nodes. Green dots are biological processes for *HPGDS*

traits on SSC1 (Table 3), but they did not include the *MTFMT* gene described in our previous study [7], likely as a result of the larger number of animals used in the current study and the different statistical models used in the two studies, i.e. WssGWAS based on the **H** relationship matrix in the current study and single-SNP GWAS using the **A** relationship matrix and deregressed phenotypes in the previous study.

In the WssGWAS, we identified several relevant QTL regions associated with the traits under study, which confirms the assumption that these traits have a complex genetic determinism. In this study, the region used to search for candidate genes was not limited to the SNP window, but also included the upstream and downstream flanking regions. The use of a larger genomic region for the identification of genes is important because SNPs within a window may be in high LD with QTL in the surrounding area.

We detected several candidate genes for the semen traits in both L1 and L2 pig lines (Tables 3 and 4). For L1, the *dynein axonemal intermediate chain 2 (DNAI2)* gene, located on SSC12, was considered the best candidate gene for MOT in the QTL region. The *DNAI2* gene encodes axonemal dyneins; the axoneme is a microtubular structure located in the center of all motile cilia and flagella, including sperm flagella [43]. Dyneins are large multisubunit ATPases that interact with microtubules to generate the driving force for flagellar motility [44]. Mutations in the human *DNAI2* are involved in defects of the axoneme [45].

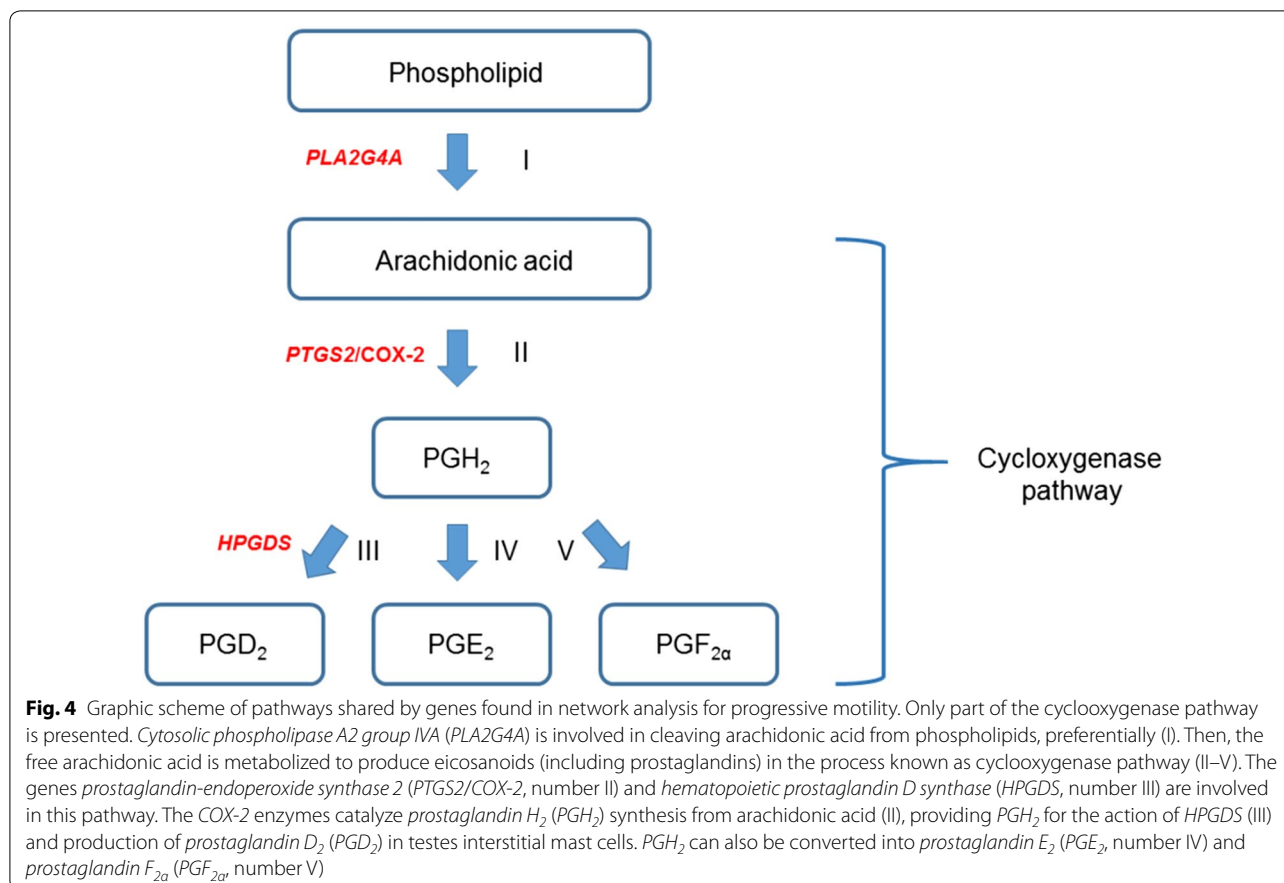
For L1, some overlapping QTL regions were identified between MOT and PROMOT. On SSC5, we identified one candidate gene, i.e. *sodium voltage-gated channel alpha subunit 8 (SCN8A)*. Pinto et al. [46] described the expression of *SCN8A* in human sperm flagellum principal piece and showed that sodium channels contribute to the regulation of human sperm motility. Another candidate gene for both traits was identified on SSC13, i.e. *Iq motif containing G (IQCG)*. Li et al. [10] showed that *Iqcg* knockout mice presented severe malformation and total immobility of their spermatozoa because of disorganized sperm flagellum axoneme. Harris et al. [47] reported that mice with mutations in the *Iqcg* gene presented spermiogenesis defects, with incomplete sperm tail formation.

We detected one overlapping QTL region in L1 for MOT, PROMOT and ABN on SSC14. This QTL region includes the gene *LOC102167830*, which is described as a *spermatogenesis associated (SPATA) protein 31E1-like*. *SPATA* genes form a large gene family that plays a very important role in testis development and spermatogenesis [48]. In humans, *SPATA31E1* is a subfamily of the *SPATA31* large gene family. In *Mus musculus*, only the *Spata31* gene has been described. Wu et al. [49]

demonstrated that the protein encoded by this mouse gene is located in the acrosome of round and elongated spermatids and *Spata31* knockout mice showed disorganized testis morphology and aberrant spermatogenic cells in seminiferous tubules.

The gene network analysis was very useful to investigate the shared biological processes between the candidate genes that were identified for the same traits between the two lines. For PROMOT, we found two genes in L1 (*PLA2G4A* and *PTGS2*) and one gene in L2 (*HPGDS*) that are involved in eicosanoid biosynthesis and arachidonic acid metabolism, of which *PTGS2* and *HPGDS* are also involved in the cyclooxygenase pathway (Figs. 3 and 4). Kaewmala et al. [50] described the presence of *PTGS2 (COX-2)* protein in boar Leydig cells, spermatogonium and spermatids, which suggests that it may have a role in the spermatogenic process in pigs. They also showed that the levels of *COX-2* mRNA and enzyme tended to be higher in animals with low sperm motility, which indicates that it has a negative effect on boar sperm quality. Frungieri et al. [51] showed that the *COX-2* enzyme is abundant in the interstitial cells of male seminiferous tubules with impaired spermatogenesis. The *COX-2* enzyme provides a precursor for the action of *HPGDS* in testes interstitial mast cells (Fig. 4), producing *prostaglandin D₂ (PGD₂)*, which regulates the function of Leydig cells [52]. Yamamoto et al. [53] and Saharkhiz et al. [54] showed that, after treatment with mast cell blockers, sperm motility increased in men. *HPGDS* is also involved in negative regulation of male germ cell proliferation (Fig. 3), which is linked to *PGD₂* production. Moniot et al. [55] identified the *PGD₂* pathway as one of the earliest signaling pathways involved in male germ cell differentiation in fetal testes. In the cyclooxygenase pathway, *PGH₂* can be converted into *prostaglandin E₂ (PGE₂)* or into *prostaglandin F_{2α} (PGF_{2α})* (Fig. 4, steps IV and V). Schlegel et al. [56] stated that seminal plasma is the richest natural source of prostaglandins, which are synthesized in the seminal vesicles. The authors showed that high concentrations of prostaglandins (especially *PGF_{2α}*) in the human seminal fluid were associated with poor sperm motility. Rios et al. [57] demonstrated that low physiological levels of *PGE₂* and *PGF_{2α}* were able to increase and prolong human progressive sperm motility.

For L2, the *NME/NM23 family member 5* gene (*NME5*) on SSC2 was considered the best candidate for lnNcells (Table 4). According to Munier et al. [58], this gene is highly and specifically expressed in testis and the encoded protein is important for the initial stages of spermatogenesis (before meiosis I). Choi et al. [59] reported that, when expression of murine *Nm23-M5* (which shares 86% identity with its human homolog *NME5*) is reduced, the round and elongated spermatids in the testes become



more sensitive to oxidative stress, leading to DNA damage and reduced cell numbers, showing that this gene is a critical factor for spermiogenesis.

The *antizyme inhibitor 2* gene (*AZIN2*) is located in a region on SSC6 that explained 4.7% of the genetic variance for ABN in L2 (Table 4). Lopez-Contreras [60] showed that *AZIN2* expression is critical for spermiogenesis in mice. Its expression starts in the testis at the beginning of spermiogenesis and its mRNA level increases in more differentiated spermatids. During spermiogenesis, round spermatids elongate, develop an acrosome in the sperm head, form a flagellum, and dispose of the excessive cytoplasm [10]. Therefore, if a mutation in *AZIN2* causes impaired spermiogenesis, it may lead to morphological defects in the spermatozoa.

The *methyltransferase like 3* gene (*METTL3*) is located on SSC7, in an overlapping QTL region for MOT and PROMOT in L2. According to Liu et al. [61], the *METTL3* protein catalyzes the methylation and formation of N⁶-methyladenosine (m⁶A), which is the most prevalent and reversible RNA epigenetic modification in mammalian mRNA. Yang et al. [62] detected increased m⁶A contents in sperm RNA from patients with reduced

sperm progressive motility, which was related to a higher expression of *METTL3*.

An overlapping QTL region for MOT and ABN was detected for L2 on SSC7 (Table 4), which includes the *spermatogenesis associated 7* gene (*SPATA7*). This gene was first identified in rat and human spermatocytes and may be involved in preparing chromatin for the initiation of meiotic recombination [63]. According to Ferguson et al. [64], meiotic recombination ties homologous chromosomes together and facilitates proper segregation of chromosomes during meiosis. Errors in the formation of crossovers can result in the production of aneuploid gametes. Sun et al. [65] reported a higher frequency of sperm aneuploidies for some chromosomes in men with sperm morphological defects compared to men with normal sperm concentrations. Thus, the *SPATA7* gene in this QTL region on SSC7 is considered the best candidate gene for MOT and ABN.

Conclusions

We identified several QTL regions that are associated with semen traits in two pig lines using the weighted single-step GWAS, which allowed detection of QTL in spite

of the small number of animals having both phenotypes and genotypes. A large part of the genetic variance of the semen traits was explained by different genes in the two lines but the gene network analysis revealed candidate genes for these two lines that are involved in shared biological pathways in the mammalian testes. These results can be used to search for causative mutations and for marker-assisted selection to enhance the production and quality of semen for a more efficient use of AI in pig breeding and production.

Additional files

Additional file 1: Table S1. Variance components and standard errors (in parenthesis) for semen traits. The data provided represent the values of variance components (additive genetic, environmental and residual) and the respective standard errors for each semen trait and pig line.

Additional file 2: Table S2. The three SNP windows that explained the highest proportion of genetic variance of each semen trait in L1. The data provided represent the chromosome, average window position (in basepairs) and the percentage of variance explained by the three SNP windows that explained the highest proportion of genetic variance for each semen trait in L1.

Additional file 3: Table S3. Three SNP windows that explained the highest proportion of genetic variance of each semen trait in L2. The data provided represent the chromosome, average window position (in basepairs) and the percentage of variance explained by the three SNP windows that explained the highest proportion of genetic variance for each semen trait in L2.

Authors' contributions

DBDM, MSL, EFK and SEFG conceived the experiment. All authors helped to design the experiment. EFK, MLWJB and MSL provided data for analysis. DBDM performed the analyses. DBDM, JWMB and BH interpreted the results. DBDM drafted the manuscript. All authors improved the writing, read and approved the final manuscript.

Author details

¹ Animal Science Department, Universidade Federal de Viçosa, Viçosa, MG 36.570-000, Brazil. ² Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands. ³ Topigs Norsvin Research Center B.V., P.O. Box 43, 6640 AA Beuningen, The Netherlands. ⁴ Topigs Norsvin, Curitiba, PR 80.420-210, Brazil.

Acknowledgements

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Topigs Norsvin.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Data availability

The datasets analysed during the current study are available from Egbert F. Knol (egbert.knol@topignorsvin.com) on reasonable request.

Ethics approval and consent to participate

The data used in this study were obtained as part of routine data recording in a commercial breeding program. Data recording and sample collection were

conducted strictly in accordance with the Dutch law on the protection of animals (Gezondheids- en welzijnswet voor dieren).

Funding

Not applicable.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 14 November 2017 Accepted: 11 July 2018

Published online: 06 August 2018

References

- Schulze M, Buder S, Rudiger K, Beyerbach M, Waberski D. Influences on semen traits used for selection of young AI boars. *Anim Reprod Sci.* 2014;148:164–70.
- Xing Y, Ren J, Ren D, Guo Y, Wu Y, Yang G, et al. A whole genome scanning for quantitative trait loci on traits related to sperm quality and ejaculation in pigs. *Anim Reprod Sci.* 2009;114:210–8.
- Sironen A, Uimari P, Nagy S, Paku S, Andersson M, Vilkki J. Knobbed acro-some defect is associated with a region containing the genes *STK17b* and *HECW2* on porcine chromosome 15. *BMC Genomics.* 2010;11:699.
- Kaewmala K, Uddin MJ, Cinar MU, Grosse-Brinkhaus C, Jonas E, Tesfaye D, et al. Association study and expression analysis of CD9 as candidate gene for boar sperm quality and fertility traits. *Anim Reprod Sci.* 2011;125:170–9.
- Gunawan A, Kaewmala K, Uddin MJ, Cinar MU, Tesfaye D, Phatsara C, et al. Association study and expression analysis of porcine ESR1 as a candidate gene for boar fertility and sperm quality. *Anim Reprod Sci.* 2011;128:11–21.
- Gunawan A, Cinar MU, Uddin MJ, Kaewmala K, Tesfaye D, Phatsara C, et al. Investigation on association and expression of ESR2 as a candidate gene for boar sperm quality and fertility. *Reprod Domest Anim.* 2012;47:782–90.
- Diniz DB, Lopes MS, Broekhuijse ML, Lopes PS, Harlizius B, Guimaraes SE, et al. A genome-wide association study reveals a novel candidate gene for sperm motility in pigs. *Anim Reprod Sci.* 2014;151:201–7.
- Zhao X, Zhao K, Ren J, Zhang F, Jiang C, Hong Y, et al. An imputation-based genome-wide association study on traits related to male reproduction in a White Duroc x Erhualian F2 population. *Anim Sci J.* 2016;87:646–54.
- Sarkar H, Arya S, Rai U, Majumdar SS. A study of differential expression of testicular genes in various reproductive phases of Hemidactylus flaviviridis (Wall Lizard) to derive their association with onset of spermatogenesis and its relevance to mammals. *PLoS One.* 2016;11:e0151150.
- Li RK, Tan JL, Chen LT, Feng JS, Liang WX, Guo XJ, et al. Iqcg is essential for sperm flagellum formation in mice. *PLoS One.* 2014;9:e98053.
- 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.
- Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet Res (Camb).* 2012;94:73–83.
- Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci.* 2010;93:743–52.
- Wang H, Misztal I, Aguilar I, Legarra A, Fernando RL, Vitezica Z, et al. 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.
- Fragomeni Bde O, Misztal I, Lourenco DL, Aguilar I, Okimoto R, Muir WM. Changes in variance explained by top SNP windows over generations for three traits in broiler chicken. *Front Genet.* 2014;5:332.
- Tiezzi F, Parker-Gaddis KL, Cole JB, Clay JS, 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.
17. Howard JT, Jiao S, Tiezzi F, Huang Y, Gray KA, 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.
 18. Valente TS, Baldi F, Sant'Anna AC, de Albuquerque LG, Paranhos da Costa MJ. Genome-wide association study between single nucleotide polymorphisms and flight speed in Nelore cattle. *PLoS One*. 2016;11:e0156956.
 19. Irano N, de Camargo GM, Costa RB, Terakado AP, Magalhaes AF, Silva RM, et al. Genome-wide association study for indicator traits of sexual precocity in Nelore cattle. *PLoS One*. 2016;11:e0159502.
 20. Lemos MV, Chiaia HL, Berton MP, Feitosa FL, Aboujaoud C, Camargo GM, et al. Genome-wide association between single nucleotide polymorphisms with beef fatty acid profile in Nelore cattle using the single step procedure. *BMC Genomics*. 2016;17:213.
 21. Medeiros de Oliveira Silva R, Stafuzza NB, de Oliveira Fragomeni B, de Camargo GMF, Ceacero TM, Cyrillo JNDSG, et al. Genome-wide association study for carcass traits in an experimental Nelore cattle population. *PLoS One*. 2017;12:e0169860.
 22. Melo TP, de Camargo GMF, de Albuquerque LG, Carvalheiro R. Genome-wide association study provides strong evidence of genes affecting the reproductive performance of Nelore beef cows. *PLoS One*. 2017;12:e0178551.
 23. Wu P, Yang Q, Wang K, Zhou J, Ma J, Tang Q, et al. Single step genome-wide association studies based on genotyping by sequence data reveals novel loci for the litter traits of domestic pigs. *Genomics*. 2018;110:171–9.
 24. Verardo LL, Silva FF, Varona L, Resende MD, Bastiaansen JWM, Lopes PS, et al. Bayesian GWAS and network analysis revealed new candidate genes for number of teats in pigs. *J Appl Genet*. 2015;56:123–32.
 25. Verardo LL, Lopes MS, Wijga S, Madsen O, Silva FF, Groenen MA, et al. After genome-wide association studies: gene networks elucidating candidate genes divergences for number of teats across two pig populations. *J Anim Sci*. 2016;94:1446–58.
 26. Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*. 2012;491:393–8.
 27. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*. 2007;81:1084–97.
 28. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH. BLUPF90 and related programs (BGF90). In Proceedings of the 7th world congress on genetics applied to livestock production, 19–23 Aug 2002; Montpellier; 2002.
 29. 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;128:422–8.
 30. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*. 2008;91:4414–23.
 31. Legarra A, Robert-Granié C, Manfredi E, Elsen JM. Performance of genomic selection in mice. *Genetics*. 2008;180:611–8.
 32. Veroneze R, Lopes PS, Guimarães SEF, Silva FF, Lopes MS, Harlizius B, et al. Linkage disequilibrium and haplotype block structure in six commercial pig lines. *J Anim Sci*. 2013;91:3493–501.
 33. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2017. <https://www.R-project.org/>. Accessed 5 July 2017.
 34. Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Kenney PB, et al. Genome-wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in rainbow trout (*Onchorhynchus mykiss*). *Front Genet*. 2016;7:203.
 35. Sollero BP, Junqueira VS, Gomes CCG, Caetano AR, Cardoso FF. Tag SNP selection for prediction of tick resistance in Brazilian Braford and Hereford cattle breeds using Bayesian methods. *Genet Sel Evol*. 2017;49:49.
 36. National Center for Biotechnology Information; 2017. <http://www.ncbi.nlm.nih.gov>. Accessed 17 Feb 2017.
 37. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009;25:1091–3.
 38. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics*. 2013;29:661–3.
 39. Garrick DJ, Taylor JF, Fernando RL. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol*. 2009;41:55.
 40. Melo TP, Takada L, Baldi F, Oliveira HN, Dias MM, Neves HH, et al. Assessing the value of phenotypic information from non-genotyped animals for QTL mapping of complex traits in real and simulated populations. *BMC Genet*. 2016;17:89.
 41. Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics*. 2011;12:186.
 42. Marques DBD, Lopes MS, Broekhuijsse MLWJ, Guimarães SEF, Knol EF, Bastiaansen JWM, et al. Genetic parameters for semen quality and quantity traits in five pig lines. *J Anim Sci*. 2017;95:4251–9.
 43. McLachlan RI, Ishikawa T, Osianlis T, Robinson P, Merriner DJ, Healy D, et al. Normal live birth after testicular sperm extraction and intracytoplasmic sperm injection in variant primary ciliary dyskinesia with completely immotile sperm and structurally abnormal sperm tails. *Fertil Steril*. 2012;97:313–8.
 44. Witman GB. Axonemal dyneins. *Curr Opin Cell Biol*. 1992;4:74–9.
 45. Loges NT, Olbrich H, Fenske L, Mussaffi H, Horvath J, Fliegauf M, et al. *DNAI2* mutations cause primary ciliary dyskinesia with defects in the outer dynein arm. *Am J Hum Genet*. 2008;83:547–58.
 46. Pinto FM, Ravina CG, Fernandez-Sanchez M, Gallardo-Castro M, Cejudo-Román A, Candenás L. Molecular and functional characterization of voltage-gated sodium channels in human sperm. *Reprod Biol Endocrinol*. 2009;7:71.
 47. Harris TP, Schimenti KJ, Munroe RJ, Schimenti JC. IQ motif-containing G (Iqcg) is required for mouse spermiogenesis. *G3 (Bethesda)*. 2014;4:367–72.
 48. Song H, Zhu L, Li Y, Ma C, Guan K, Xia X, et al. Exploiting RNA-sequencing data from the porcine testes to identify the key genes involved in spermatogenesis in Large White pigs. *Gene*. 2015;573:303–9.
 49. Wu YY, Yang Y, Xu YD, Yu HL. Targeted disruption of the spermatid-specific gene *Spat31* causes male infertility. *Mol Reprod Dev*. 2015;82:432–40.
 50. Kaewmala K, Uddin MJ, Cinar MU, Große-Brinkhaus C, Jonas E, Tesfaye D, et al. Investigation into association and expression of *PLCz* and *COX-2* as candidate genes for boar sperm quality and fertility. *Reprod Domest Anim*. 2012;47:213–23.
 51. Frungieri MB, Weidinger S, Meineke V, Köhn FM, Mayerhofer A. Proliferative action of mast-cell tryptase is mediated by *PAR2*, *COX2*, prostaglandins, and *PPARG*: possible relevance to human fibrotic disorders. *Proc Natl Acad Sci USA*. 2002;99:15072–7.
 52. Schell C, Frungieri MB, Albrecht M, Gonzalez-Calvar SI, Köhn FM, Calandra RS, et al. A prostaglandin D2 system in the human testis. *Fertil Steril*. 2007;88:233–6.
 53. Yamamoto M, Hibi H, Miyake K. New treatment of idiopathic severe oligozoospermia with mast cell blocker: results of a single-blind study. *Fertil Steril*. 1995;64:1221–3.
 54. Saharkhiz N, Nikbakht R, Hemadi M. Ketotifen, a mast cell blocker improves sperm motility in asthenospermic infertile men. *J Hum Reprod Sci*. 2013;6:19–22.
 55. Moniot B, Ujjan S, Champagne J, Hirai H, Aritake K, Nagata K, et al. Prostaglandin D2 acts through the *Dp2* receptor to influence male germ cell differentiation in the foetal mouse testis. *Development*. 2014;141:3561–71.
 56. Schlegel W, Rotermund S, Färber G, Nieschlag E. The influence of prostaglandins on sperm motility. *Prostaglandins*. 1981;21:87–99.
 57. Rios M, Carreño DV, Osés C, Barrera N, Kerr B, Villalón M. Low physiological levels of prostaglandins E2 and F2alpha improve human sperm functions. *Reprod Fertil Dev*. 2016;28:434–9.
 58. Munier A, Feral C, Milon L, Pinon VP, Gyapay G, Capeau J, et al. A new human *nm23* homologue (*nm23-H5*) specifically expressed in testis germinal cells. *FEBS Lett*. 1998;434:289–94.
 59. Choi YJ, Cho SK, Hwang KC, Park C, Kim JH, Park SB, et al. *Nm23-M5* mediates round and elongated spermatid survival by regulating *GPX-5* levels. *FEBS Lett*. 2009;583:1292–8.

60. López-Contreras AJ, Ramos-Molina B, Martínez-de-la-Torre M, Peñafiel-Verdú C, Puelles L, Cremades A, et al. Expression of antizyme inhibitor 2 in male haploid germinal cells suggests a role in spermiogenesis. *Int J Biochem Cell Biol.* 2009;41:1070–8.
61. Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L, et al. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat Chem Biol.* 2014;10:93–5.
62. Yang Y, Huang W, Huang JT, Shen F, Xiong J, Yuan EF, et al. Increased N6-methyladenosine in human sperm RNA as a risk factor for asthenozoospermia. *Sci Rep.* 2016;6:24345.
63. Zhang X, Liu H, Zhang Y, Qiao Y, Miao S, Wang L, et al. A novel gene, RSD-3/HSD-3.1, encodes a meiotic-related protein expressed in rat and human testis. *J Mol Med (Berl).* 2003;81:380–7.
64. Ferguson KA, Wong EC, Chow V, Nigro M, Ma S. Abnormal meiotic recombination in infertile men and its association with sperm aneuploidy. *Hum Mol Genet.* 2007;16:2870–9.
65. Sun F, Ko E, Martin RH. Is there a relationship between sperm chromosome abnormalities and sperm morphology? *Reprod Biol Endocrinol.* 2006;4:1.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

