



Heterozygosity and homozygosity regions affect reproductive success and the loss of reproduction: A case study with litter traits in pigs



Zitao Chen^a, Zhenyang Zhang^a, Zhen Wang^a, Zhe Zhang^a, Qishan Wang^{a,b,*}, Yuchun Pan^{a,b}

^a Department of Animal Science, College of Animal Science, Zhejiang University, 866# Yuhangtang Road, Hangzhou East 310058, China

^b Hainan Institute, Zhejiang University, Yongyou Industry Park, Yazhou Bay Sci-Tech City, Sanya South 572000, China

ARTICLE INFO

Article history:

Received 1 May 2022

Received in revised form 21 July 2022

Accepted 22 July 2022

Available online 26 July 2022

Keywords:

Association analysis

Mortality

Reproductive success

Runs of heterozygosity

Runs of homozygosity

ABSTRACT

Runs of heterozygosity (ROHet) and homozygosity (ROH) harbor useful information related to traits of interest. There is a lack of investigating the effect of ROHet and ROH on reproductive success and the loss of reproduction in mammals. Here, we detected and characterized the ROHet and ROH patterns in the genomes of Chinese indigenous pigs (i.e., Jinhua, Chun'an, Longyou Black, and Shengxian Spotted pigs), revealing the similar genetic characteristics of indigenous pigs. Later, we highlighted the underlying litter traits-related ROHet and ROH using association analysis with linear model in these four indigenous pig breeds. To pinpoint the promising candidate genes associated with litter traits, we further in-depth explore the selection patterns of other five pig breeds (i.e., Erhualian, Meishan, Minzhu, Rongchang, and Diqing pigs) with different levels of reproduction performance at the underlying litter traits-related ROHet and ROH using F_{ST} and genetic diversity ratio. Then, we identified a set of known and novel candidate genes associated with reproductive performance in pigs. For the novel candidate genes (i.e., *CCDC91*, *SASH1*, *SAMD5*, *MACF1*, *MFSD2A*, *EPC2*, and *MBD5*), we obtained public available datasets and performed multi-omics analyses integrating transcriptome-wide association studies and comparative single-cell RNA-seq analyses to uncover the roles of them in mammalian reproductive performance. The genes have not been widely reported to be fertility-related genes and can be complementally considered as prior biological information to modify genomic selections models that benefits pig genetic improvement of litter traits. Besides, our findings provide new insights into the function of ROHet and ROH in mammals.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Litter traits are one of the most important traits for breeding in pig production [1,2]. Improving total number born (TNB) and number born alive (NBA) traits are of great interest as enhancing the economic profit, and become the main breeding goals in the modern pig industry [1,3]. Meanwhile, sow reproductive performances are limited by embryonic mortality and fetal losses during the pregnancy period [4]. Recently, the genetic mechanisms of the loss of reproduction have gradually received attention, and a few studies have explored underlying genetic architecture of the number of piglets born dead traits in pigs [3,5,6]. Moreover, the extreme litter sizes reduce welfare and survival of the piglets at weaning, and decreasing the variation in litter size will lead to more sustainable

breeding in terms of increasing piglet survival [7]. Generally, litter traits have low heritability [8,9], therefore, genetic analyses should be systematically conducted for the traits to generate the in-depth biological knowledge that can significantly improve the selection efficiency of genomic selection and benefit genetic improvements of low heritability traits [10,11]. In addition, swine can be used as an alternative animal model, which likely overcome the limitations in human and other models, one reason is that swine are highly similar to humans in many aspects, e.g., hormonal cycle and physiology [12]. For this reason, pig is considered as the larger animal model for human infertility in the field of reproduction research [13].

The rapid development of high-throughput sequencing has opened up the possibility of characterizing the genome in terms of homozygosity or heterozygosity. Runs of heterozygosity (ROHet), known as heterozygosity-rich regions, emerged as a more recent concept, and have been used to identify genomic regions that are under gene introgression or admixture [14,15]. ROHet

* Corresponding author at: Department of Animal Science, College of Animal Science, Zhejiang University, 866# Yuhangtang Road, Hangzhou East 310058, China.
E-mail address: wangqishan@zju.edu.cn (Q. Wang).

avoid the deleterious effects of harmful homozygous genotype aggregation on specific traits and maintain heterozygote advantage at immune-related genes [16], benefiting reproduction performance. In the meantime, runs of homozygosity (ROH) have been a helpful strategy to detect regions under selection [17]. A higher incidence of ROH could lead to congenital anomalies [18], resulting in the loss of reproduction. Previous studies mainly used the distribution of ROHet or ROH and the function of genes within ROHet or ROH to investigate the population characteristics [19,20]. Based on this, association analyses using ROHet and ROH provide a powerful strategy to identify genomic regions associated with traits of interest in human [21] and livestock [16].

China has the most genetic resources of domestic pigs [22,23]. Jinhua (JH), Chun'an (CA), Longyou Black (LY), and Shengxian Spotted (SX) pigs originated in eastern China and have similar characteristics to other indigenous breeds, e.g., improved meat quality, disease resistance, and high fertility in comparison with intensively-reared commercial pig breeds (e.g., Large white, Landrace, and Duroc pig breeds) [24–26]. These excellent characteristics mark them as the best material to uncover the genetic mechanism of economic traits and support the sustainable development of the pig industry [27]. It is well reported that Chinese indigenous pigs have excellent litter sizes and contribute to the improvement of fertility in intensively-reared commercial pig breeds [28,29]. However, there is a lack of investigating the effect of ROHet and ROH on litter traits in Chinese indigenous pigs.

In this study, we performed integrated genomic analyses to explore the genetic mechanisms of reproductive success and the loss of reproduction in mammal. Firstly, we systematically investigated the distribution of ROHet and ROH in the genomes of indigenous pigs to detect and characterize the ROHet and ROH patterns, as well as to reveal ROHet and ROH islands that contain the candidate genes related to breed-specific traits of Chinese indigenous pig breeds. Then, we conducted association analyses to identify the litter traits-related regions, and uncovered the effects of the identified regions and promising candidate genes on reproductive performance in other pig breeds via traits-specific selection signatures. Furthermore, we performed comparative transcriptome analysis to identify important fertility-related genes in human, pig, and mouse.

2. Materials and methods

2.1. Ethics statement

Animal care and experiments were conducted in accordance with the Chinese guidelines for animal welfare and were approved by the Animal Care and Use Committee of Zhejiang University, Hangzhou, China (permit number: ZJU20160346).

2.2. Animal resources, phenotypes, and SNP genotyping

Blood samples of JH (N = 193: 41 sires and 152 dams), CA (N = 98: 31 sires and 67 dams), LY (N = 94: 32 sires and 62 dams), and SX (N = 174: 43 sires and 131 dams) were collected from the national or provincial conservation farms in Zhejiang province of China. We collected all sires and dams except for full siblings according to the pedigree records in conservation farms, thus these individuals represent the most comprehensive genetic diversity of each pig breed. Litter traits included TNB, NBA, and total number of piglets born dead (TND). Herein, NBA was identified as the piglets alive after the farrowing; TND included the number of stillborn piglets and the mummified piglets at birth. We removed records if TNB was equal to zero, or if only one record per sow was available. After data editing, 4742 litter records from 412 sows were

available. After grouping records according to sow parity, the farrowing records were 1236, 624, 537, 507, 645, 543, 489, and 161 for parity groups 1, 2, 3, 4, 5, 6, 7, and 8, respectively. The descriptive statistics of TNB, NBA, and TND per parity were listed in Supplementary Table 1.

To get a better understanding of the genetic architecture of litter size variability traits, we obtained log-transformed variance of residuals of TNB (LnVarTNB), NBA (LnVarNBA), and TND (LnVarTND) traits from the records of each trait following [9,30]: firstly, we used the formula (1) to calculate a residual for available observation for TNB, NBA, and TND traits; secondly, a variance of these residuals was log-transformed to generate a unique value of LnVar for each of TNB, NBA, and TND traits per sow.

$$y = W\alpha + Zb + Upe + e \quad (1)$$

where y is the TNB, NBA, and TND of the individuals; α is a vector of fixed effects, including farm-year-season of the farrowing (64 levels) and parity of the sow (8 levels); b is a vector of additive genetic effects and is set as $b \sim N(0, G\sigma_b^2)$, where G is the genomic relationship matrix [31]; pe is a vector of permanent environmental sow effects and is set as $pe \sim N(0, I\sigma_{pe}^2)$; W , Z , and U are incidence matrices for α , b , and pe ; e is a vector of residuals and is set as $e \sim N(0, I\sigma_e^2)$; I is an identity matrix.

Genomic DNA was extracted from blood samples with a standard phenol–chloroform method and genotyped using the GeneSeek Genomic Profiler Porcine SNP BeadChip (Neogen Corporation, Lansing, MI, USA), which contains 50,697 SNPs. Subsequently, quality control was performed using PLINK (v1.9) software [32], of which SNPs were filtered out with a call rate less than 90 % [33]. A final set of 44,901 informative SNPs were used for further ROHet and ROH analysis.

Furthermore, we obtained whole-genome re-sequencing (WGS) data of five pig breeds (Erhualian, EHL, N = 12; Minzhu, MZ, N = 10; Meishan, MS, N = 14; Rongchang, RC, N = 12; Diqing, DQ, N = 12) from the NCBI SRA database to uncover the evolutionary history of the regions and potential effects on litter traits in pigs (Supplementary Table 2). These pig breeds contained two extremely prolific breeds (EHL and MS), two prolific breeds (MZ and RC), and one breed with the lowest litter size (DQ). Herein, raw reads were filtered using fastp software with default parameters. Filtered reads were aligned to the latest pig reference genome (*Scrofa11.1*) using the Burrows–Wheeler alignment (v0.7.17-r1188) tool [34] with the parameters for paired-end reads. Subsequently, SAM files were merged and sorted using SAMtools (v1.9) software [35]. SNP calling for each individual was implemented using the HaplotypeCaller program and further joint calling using GenotypeGVCFs program of GATK4 (v4.1.6.0) software [36]. We filtered SNPs using VariantFiltration program with the parameters: --filter-expression “QD less than 2.0 || FS > 60.0 || MQ less than 40.0 || SOR > 3.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0.” Quality control was performed using PLINK (v1.9) software [32], of which SNPs were filtered out with a call rate less than 90 % and minor allele frequency (MAF) less than 0.01. The remaining 34,265,460 informative SNPs were used for signatures of selection detection. In addition, the performance of TNB, NBA, and TND traits for each available breed was obtain from *Animal Genetic Resources in China: Pigs*.

2.3. Runs of heterozygosity and homozygosity detection

ROHet and ROH were detected separately for each individual using detectRUNS R package v0.9.6. A ROHet was identified based on a consecutive strategy [37] with the following parameters: (i) a minimum ROHet length of 1 Mb; (ii) a minimum number of SNPs

within a ROHet > 15; (iii) no homozygous and missing genotypes; (iv) the maximum gap of 500 kb between consecutive heterozygous SNPs. In addition, a ROH was defined based on a sliding window strategy [32] with the following parameters: (i) a sliding window of 50 SNPs; (ii) a minimum ROH length of 1 Mb; (iii) no heterozygous and missing genotypes; (iv) the maximum gap of 500 kb between consecutive homozygous SNPs; (v) a minimum SNP density of two SNP per Mb; (vi) a minimum number of SNPs within a ROH at least 85, 74, 50, and 84 in JH, CA, LY, and SX pigs, which were computed by the formula (2) [38]. Next, we filtered out the ROHet and ROH with less than five SNPs and present in less than 5 % of the individuals.

$$l = \frac{\ln \frac{\alpha}{n_s \times n_i}}{\ln(1 - het)} \quad (2)$$

where α is the percentage of false positive ROH (set to 0.05), n_s is the number of SNPs per individual, n_i is the number of individuals, and het is the proportion of heterozygosity across all SNPs.

2.4. ROhet and ROH size categories

We categorized the ROHet and ROH for each breed into five length classes (1–2 Mb, 2–4 Mb, 4–8 Mb, 8–16 Mb, and > 16 Mb). In addition, we estimated ROH-based inbreeding coefficients (F_{ROH}) using the formula (3) [39].

$$F_{ROH} = \frac{\sum L_{ROH}}{L_{AUTO}} \quad (3)$$

where $\sum L_{ROH}$ is the total length of all the ROH detected in an individual, L_{AUTO} is the total length of the autosomes covered by SNPs (2.452 Gb in this study).

Furthermore, we generated the false discovery rate (FDR) adjusted empirical P -values [40,41] by genome-wide ranking of the percentage of SNP occurrences to identify highly heterozygous regions (ROhet island) or highly homozygous regions (ROH island) via extracting SNPs with the FDR adjusted empirical P -values less than 0.01 located in a ROhet or ROH.

2.5. Genome-wide ROhet and ROH analysis

We fitted linear model to investigate the association between unique heterozygous or homozygous regions (presence or absence) and six reproduction traits using lm function in R v4.0.5 [42] via the formula (4).

$$y_{jkli} = breed_j + farmyearseason_k + region_i + e_{jkli} \quad (4)$$

where y_{jkli} is the phenotype of the i individual, i.e., TNB, NBA, and TND per parity, LnVarTNB, LnVarNBA, and LnVarTND; $breed_j$ is the fixed effect of the j breed; $farmyearseason_k$ is the fixed effect of the k farm-year-season of the farrowing; $region_i$ is the fixed effect with two levels of ROhet or ROH (presence as one and absence as zero); e_{jkli} is the random residual. Here, we used the ROhet ($N = 95$) and ROH ($N = 203$) detected in >1 % of the individuals that had phenotypic records. Then, the Bonferroni correction was applied to determine the threshold of ROhet or ROH association analysis. ROhet or ROH with P -values lower or equal than $0.05/N$, were regarded as genome-wide significant ROhet or ROH. ROhet or ROH with a P -value higher than $0.05/N$ but lower than $1/N$ were considered as genome-wide suggestive significant ROhet or ROH.

2.6. Signatures of selection detection

We investigated selection signatures across the highest frequency ROhet and ROH using VCFtools version 0.1.14 [43] via F_{ST} [44] and genetic diversity ratio (θ_π statistic). We calculated θ_π

statistic across the highest frequency ROhet and ROH regions in JH, CA, LY, and SX pigs to measure the variability of polymorphism levels across the regions between these pigs. To get more comprehensive results of genetic differentiation among the extremely prolific breeds, prolific breeds and breeds with the lowest litter size, five pig breeds (EHL, MZ, MS, RC, and DQ) were further compared with each other using F_{ST} and θ_π statistic with the use of a sliding window method (50 kb window and 10 kb step). The θ_π statistic between two breeds was calculated as $\ln(\theta_{\pi|breed1}/\theta_{\pi|breed2})$. Moreover, we divided five pig breeds into two groups (group 1: MS, RC, and DQ; group 2: EHL, MZ, and DQ) and separately conducted F_{ST} analysis. Here, we uncovered the selection patterns of the highest frequency ROhet and ROH with the breeds of group 1, then the breeds of group 2 were further used to check the patterns.

2.7. Gene annotation

Genome annotations were based on *Sscrofa11.1*, genes overlapping with ROhet and ROH islands were treated as candidate genes. We downloaded the information of tissue-specific and human-pig homologous genes from the Pig RNA Atlas [45]. Furthermore, we collected the results of genome-wide association studies (GWAS) summary statistic and transcriptome-wide association studies (TWAS) research for fertility-related traits in human from the webTWAS database [46] to further explore the potential roles of the novel genes in female mammalian reproductive function.

2.8. RNA-seq data processing and bioinformatics analysis

The single cell RNA-seq (scRNA-seq) data of in vivo embryos (oocyte, 2-cell, 4-cell, and 8-cell) from human, mouse [47], and pig [48] were collected from the GEO public database (GSE44183 and GSE139512) to conduct comparative transcriptome analysis. Furthermore, the RNA-seq data of gestational diabetes mellitus (GDM) and healthy control placentas was collected from the database (GSE154414). We downloaded the datasets and further processed that using FastQC (v0.11.4) software to obtain clean reads with default parameters. Clean reads were then mapped to the reference genome and annotated transcripts (human: GRCh38; mouse: GRCm39; pig: *Sscrofa11.1*) by Hisat2 (v2.1.0) software [49]. The read numbers mapped to each gene were counted and quantified as fragments per kilobase million mapped reads (FPKM) using featureCounts (v2.0.3) software [50]. To reduce the bias of genes and gene expression levels potentially affected by different library preparation and sequencing platforms, we retained the transcripts with FPKM larger than zero.

2.9. Functional enrichment analyses

In addition, all ROhet and ROH islands were annotated using the Animal QTL Database (Hu et al., 2019). We conducted QTL enrichment analyses using GALLO (Fonseca et al., 2020). We used ClusterProfiler package [51] to perform Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. Here, QTLs and terms with the FDR less than 0.05 were retained.

3. Results

3.1. Runs of heterozygosity detection and annotation

In total, we detected 117, 353, 1,505, and 676 ROhet in the CA, JH, LY, and SX populations, with an average of 1.19, 1.83, 16.01, and 3.89 ROhet for each pig breed, respectively (Fig. 1A). The average number of SNPs in each ROhet was 18.68 ± 4.11 , 18.69 ± 3.55 ,

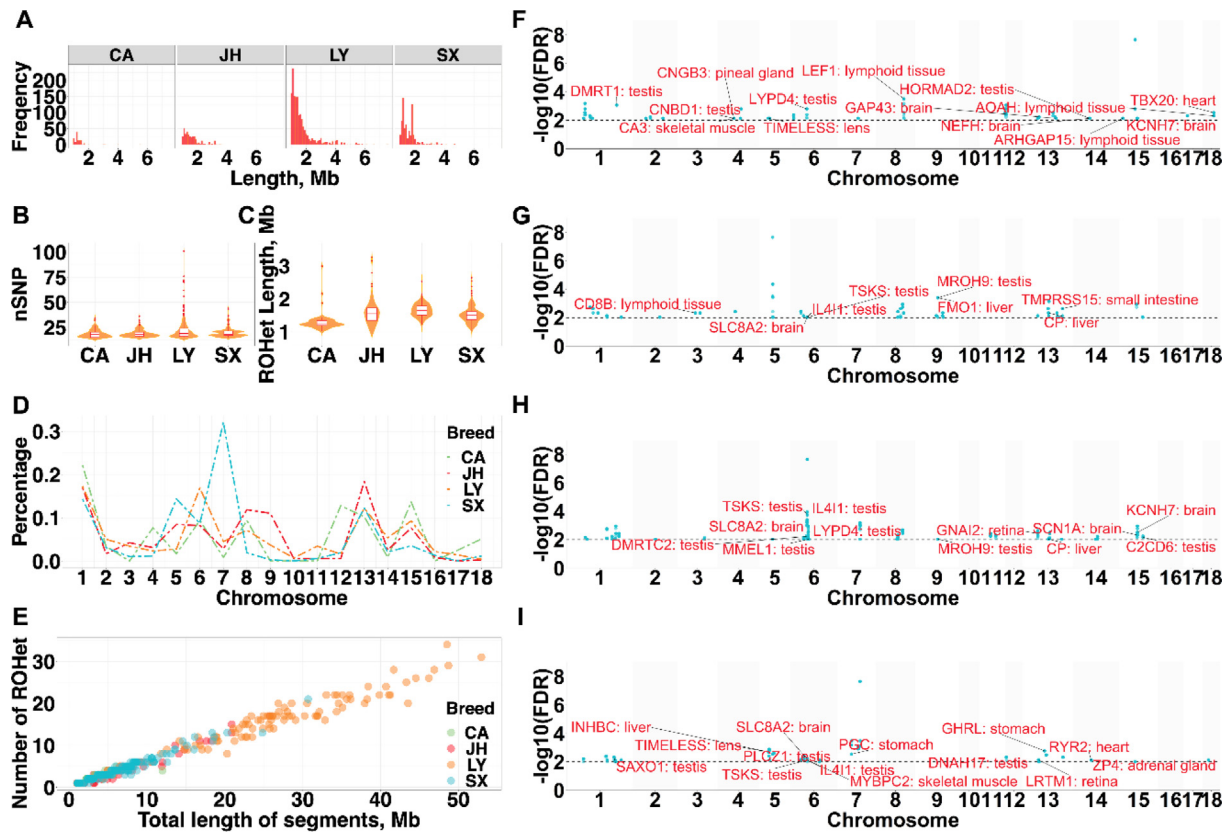


Fig. 1. ROHet in the genomes of Chun'an (CA), Jinhua (JH), Longyou Black (LY), and Shengxian Spotted (SX) pig populations. (A) Number of ROHet; (B) Average number of SNPs in ROHet; (C) Average ROHet length; (D) Number of total ROHet per chromosome; (E) Number and total length of ROHet per individual; Manhattan plot of the incidence of heterozygotes within ROHet in CA (F), JH (G), LY (H), and SX (I) populations.

21.09 ± 7.33, and 19.57 ± 4.66, and the mean length of ROHet was 1.35 ± 0.55, 1.58 ± 0.55, 1.67 ± 0.84, and 1.55 ± 0.51 Mb in the CA, JH, LY, and SX populations (Fig. 1B, C). Further, we observed that a higher number of ROHet on *sus scrofa* chromosome (SSC) 1 and 13, which had the longest chromosome length. To note, the SX population had the highest number of ROHet on SSC 7, showing an extremely different pattern with other populations (Fig. 1D). The short fragment (1–2 Mb) of ROHet accounted for the largest fraction of ROHet, and no ROHet with the length > 8 Mb was observed (Table 1). Meanwhile, the number of ROHet decreased with the increasing of ROHet length (Fig. 1E).

To explore the function of ROHet islands, we conducted tissue-specific genes annotation and QTL enrichment analysis (Fig. 1F-I).

For CA pigs, 17 out of 167 candidate genes were specifically expressed in a single tissue. The genome regions overlapping with ROHet islands were mostly significantly related to days to 100 kg, carcass weight (cold), umbilical hernia, backfat at last rib, and shear force at first peak traits (Supplementary Fig. 1A). For JH pigs, a total of 137 candidate genes were located in ROHet islands, and involved in production (e.g., days to 100 kg), meat and carcass (e.g., intramuscular fat content), and health (e.g., actinobacillus pleuropneumoniae susceptibility) traits (Supplementary Fig. 1B). For LY pigs, we detected 179 candidate genes, of which had 12 tissue-specific genes. These genes were linked with basophil number, intramuscular fat content, days to 100 kg, blood non-esterified fatty acid level, and fat androstenone level traits (Supplementary

Table 1
Number, percentage, and average length for ROHet categories.

Categories	Jinhua			Chun'an			Longyou Black			Shengxian Spotted		
	Count	Percentage (%)	Average (Mb)	Count	Percentage (%)	Average (Mb)	Count	Percentage (%)	Average (Mb)	Count	Percentage (%)	Average (Mb)
ROHet: 1-2 Mb	301	85.27	1.39	113	96.58	1.26	1219	81	1.34	625	92.46	1.44
ROHet: 2-4 Mb	52	14.73	2.67	2	1.71	3.08	228	15.15	2.66	48	7.1	2.74
ROHet: 4-8 Mb	-	-	-	2	1.71	4.75	58	3.85	4.68	3	0.44	4.51
ROH: 1-2 Mb	55	0.73	1.81	20	0.98	1.64	171	11.25	1.64	28	0.46	1.86
ROH: 2-4 Mb	1205	15.94	3.23	419	20.55	3.06	563	37.04	2.95	726	11.84	3.18
ROH: 4-8 Mb	2870	37.97	5.69	687	33.69	5.77	417	27.43	5.71	1999	32.63	5.9
ROH: 8-16 Mb	1960	25.93	11.22	528	25.89	11.08	236	15.53	10.65	1685	27.5	11.38
ROH: > 16 Mb	1469	19.43	33.54	385	18.89	35.95	133	8.75	31.79	1689	27.57	35.45

Fig. 1C). For SX pigs, we detected 146 candidate genes, which were associated with average backfat thickness, meat color a^* , shear force at first peak, fat androstenone level, and backfat at last rib traits (Supplementary Fig. 1D). Furthermore, we found that there were 66 ROHet shared by at least two pig breeds, and 41, 99, 658, and 116 unique ROHet detected in the CA, JH, LY, and SX populations, respectively (Supplementary Fig. 2A).

3.2. Runs of homozygosity detection and annotation

In total, we identified 17,245 ROH in the 559 individuals considered (Table 1). In terms of average number of ROH per individual, the JH population had the largest number (39.17 ROH per animal), followed by the SX (35.21 ROH), CA (20.81 ROH), and LY (16.17 ROH) populations (Fig. 2A). The average number of SNPs in each ROH was 238.66 ± 241.91 , 230.93 ± 228.94 , 132.84 ± 173.59 , and 300.21 ± 275.92 for the CA, JH, LY, and SX populations (Fig. 2B). Meanwhile, the LY population had the lowest average value (7.28 Mb) of the total ROH length, and the SX, CA, and JH populations were 15.21, 12.25, and 12.12 Mb, respectively (Fig. 2C). Further, we found that the patterns of the number of ROH on each chromosome were mostly similar in the four populations considered, whereas the SX population showed slightly different trends on SSC 7, 15, 16 (Fig. 2D). In addition, the number and length of ROH had negative correlations (Fig. 2E). The F_{ROH} showed that the SX and JH populations had the highest level of inbreeding (0.24 and 0.21, respectively), followed by the CA (0.11) and LY populations (0.05) (Supplementary Table 3). Nevertheless, we observed the similar trends of the F_{ROH} on each chromosome in the studied populations, except for the LY population (Supplementary Fig. 3).

There were 136 ROH shared by at least two pig breeds, and 1372, 5355, 1184, and 4112 unique ROH detected in the CA, JH, LY, and SX populations, respectively (Supplementary Fig. 2B). We found several shared candidate genes overlapping with the ROH islands of the studies breeds, e.g., *UHRF1*, *UBA6*, *GC*, *AFP* and *CHN3*. In addition, the QTLs overlapping with the ROH islands were mostly related to meat and carcass traits (e.g., backfat at last rib, intramuscular fat content, and fat androstenone level traits). Nevertheless, we found some QTLs associated with exterior traits (e.g., ear weight and thoracic vertebra number traits) reached the significant threshold (Supplementary Fig. 4).

3.3. Litter traits-related ROHet and ROH detection

The association analysis revealed 17, 7, 14, and 1 genome-wide significant ROHet for TNB, NBA, TND, and LnVarNBA traits, and 37, 39, 16, 4, and 2 ROHet reached the suggestive significance level for TNB, NBA, TND, LnVarNBA, and LnVarTND traits, respectively (Supplementary Table 4). Later, we counted the genome-wide significant ROHet and the ROHet reached the suggestive significance level, 44.22–46.11 Mb on SSC 5, 61.31–63.40 Mb on SSC 6, and 107.32–109.23 Mb on SSC 8 had the highest frequency (Fig. 3A). Herein, 13 out of 40 promising candidate genes were located on the highest frequency ROHet and overlapped with ROHet islands completely or partly, e.g., *CAMK2D*, *UGT8*, *ARSJ*, and *OVCH1* (Supplementary Table 5).

Meanwhile, we identified 7, 6, 17, 1, and 2 genome-wide significant ROH for TNB, NBA, TND, LnVarTNB, and LnVarNBA traits. There were 40, 23, 37, 6, 2, and 1 ROH reached the suggestive significance level for TNB, NBA, TND, LnVarTNB, LnVarNBA, and LnVarTND traits (Supplementary Table 6). In addition, 94.43–

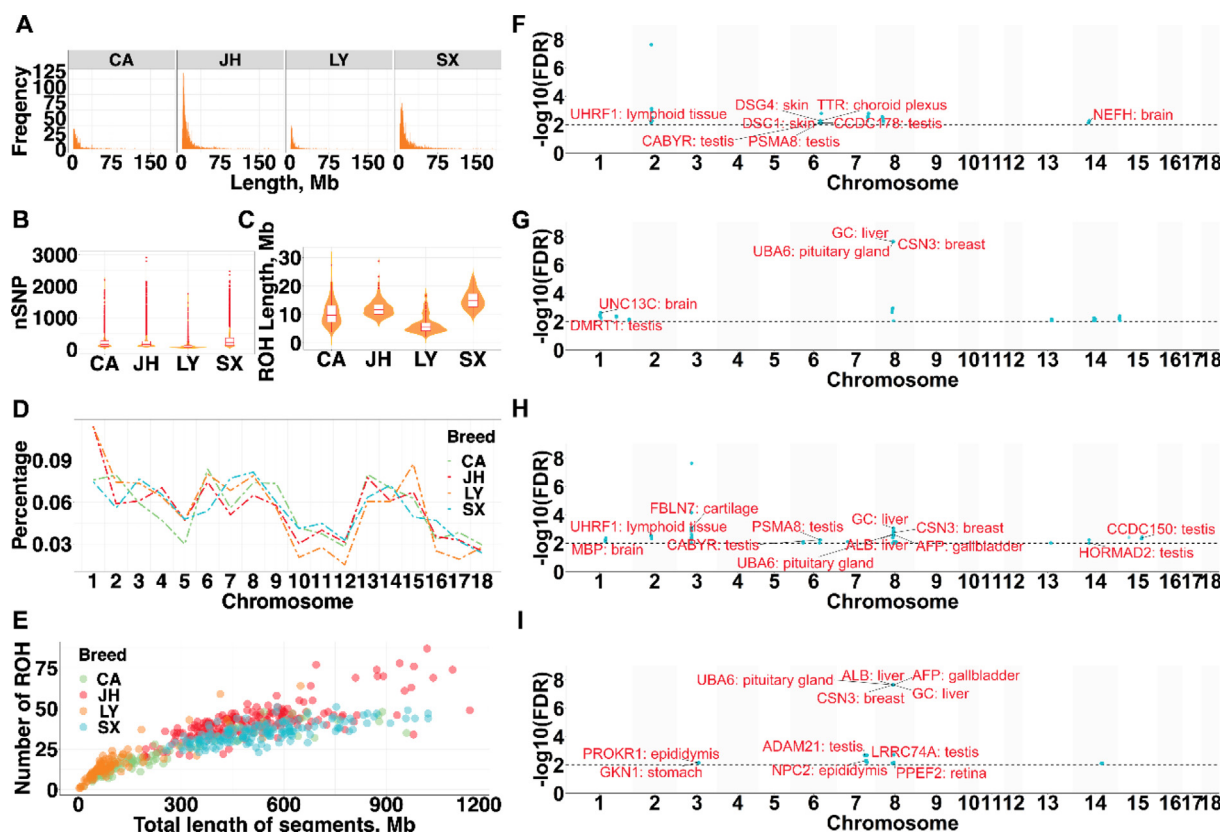


Fig. 2. ROH in the genomes of Chun'an (CA), Jinhua (JH), Longyou Black (LY), and Shengxian Spotted (SX) pig populations. (A) Number of ROH; (B) Average number of SNPs in ROH; (C) Average ROH length; (D) Number of total ROH per chromosome; (E) Number and total length of ROH per individual; Manhattan plot of the incidence of homozygous within ROH in CA (F), JH (G), LY (H), and SX (I) populations.

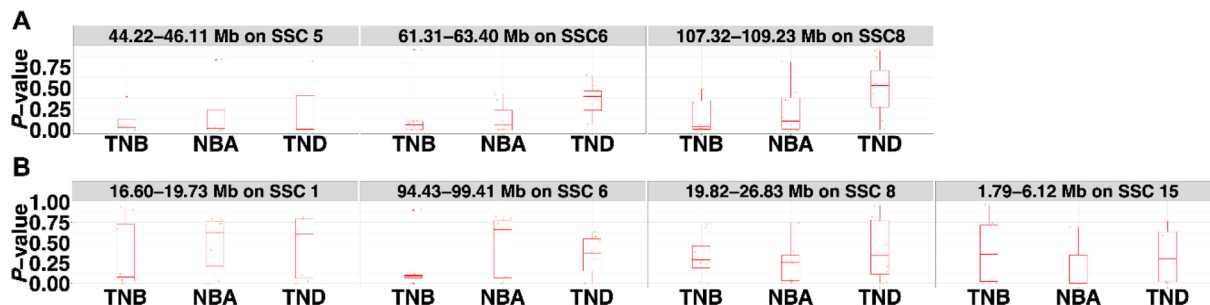


Fig. 3. The association degree between the highest frequency (A) ROHet or (B) ROH and litter traits among parities.

99.41 Mb on SSC 6 got the highest frequency, followed by 16.60–19.73 Mb on SSC 1, 19.82–26.83 Mb on SSC 8, and 1.79–6.12 Mb on SSC 15 (Fig. 3B). We found 5 out of 63 promising candidate genes were located on the highest frequency ROH and ROH islands, i.e., *EPC2*, *MMADHC*, *MBD5*, *KIF5C*, and *ACVR2A* (Supplementary Table 7).

Furthermore, we observed that the different association degree between the highest frequency ROHet (or ROH) and litter traits among parities (Fig. 3), which revealed that the genomic structures of piglets born alive and dead were different among parity groups.

3.4. Pattern of selection at the highest frequency litter traits-related ROHet

We calculated θ_{π} statistic across the highest frequency ROHet in JH, CA, LY, and SX pigs, however, no trend was observed in comparison of these pigs (Supplementary Fig. 5). These might be caused by the similar reproduction performance of these pigs. Thus, we used the high-density genotypes of other pig breeds contained the extremely prolific (MS and EHL), prolific (MZ and RC), and lowest litter sized (DQ) breeds to detect selection signatures. Later, we correlated the median θ_{π} statistic across the highest frequency litter traits-related ROHet for each breed comparison to the difference in mean breed value for TNB, NBA, and TND traits. One ROHet (44.22–46.11 Mb on SSC 5) was significantly associated with TNB and NBA traits (Supplementary Fig. 6). It was also detected in the F_{ST} analysis with the breeds of group 1 and contains nine genes: *OVCH1*, *ERGIC2*, *FAR2*, *CCDC91*, *PTHLH*, *KLHL42*, *MANSC4*, *MRPS35*, and *PPFIBP1*. Herein, *OVCH1* and *CCDC91* were two strong candidates with the peak of selection falling (Fig. 4). The similar selection patterns were also shown in the analysis with the breeds of group 2 (Supplementary Fig. 7), precisely revealing the genetic

differentiation at the ROHet between the prolific pig breeds and pig breeds with the lowest litter size. Meanwhile, the peak of selection was overlapped with the QTLs that were related to reproduction traits (i.e., teat number, number of stillborn, reproductive tract weight, uterine horn length, and uterine horn weight traits).

3.5. Pattern of selection at the highest frequency litter traits-related ROH

In addition, there was no trend of selection patterns across the highest frequency ROH has been observed in comparison of JH, CA, LY, and SX pigs (Supplementary Fig. 5). Likewise, we investigated the correlation between the median θ_{π} statistic across the highest frequency litter traits-related ROH for each breed and difference in mean breed value for TNB, NBA, and TND traits with other five pig breeds (Supplementary Fig. 8). The median θ_{π} statistic at the ROH: 16.60–19.73 Mb on SSC 1 had a significant correlation with difference in mean of TNB and NBA (Fig. 5A). There were eight promising candidate genes (i.e., *TAB2*, *UST*, *SASH1*, *SAMD5*, *STXBP5*, *RAB32*, *GRM1*, and *SHPRH*) located on the ROH: 16.60–19.73 Mb on SSC 1 (Fig. 5B and Supplementary Fig. 7). The peak of selection across the ROH were overlapped with QTLs of number of stillborn and teat number traits. In the meanwhile, we observed the significant correlation between the median θ_{π} statistic for each breed (Fig. 5C) and TND at the ROH: 94.43–99.41 Mb on SSC 6 where 43 promising candidate genes (e.g., *MACF1*, *MFS2A*, *CAP1*, *PPT1*, *ZMPSTE24*, *MC2R*, and *PTPRM*) and a set of QTLs of reproduction traits (i.e., age at puberty, gestation length, litter size, non-functional nipples, number of mummified pigs, and teat number traits) were located (Fig. 5D). The further detection of selection signatures with group 2 proved the accuracy of the genetic differentiation between the extremely, prolific pig breeds, and pig breeds

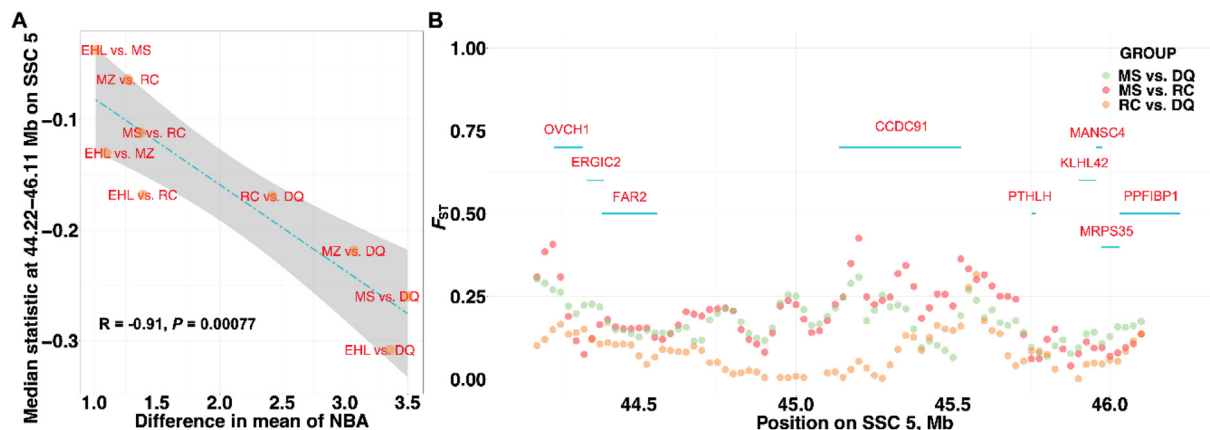


Fig. 4. Patterns of selection at the ROHet: 44.22–46.11 Mb on SSC 5. (A) The correlation between the median genetic diversity ratio at the ROHet and the corresponding differences in number born alive trait between each pair of pig breeds; (B) The F_{ST} statistics at the ROHet in selected breed comparisons.

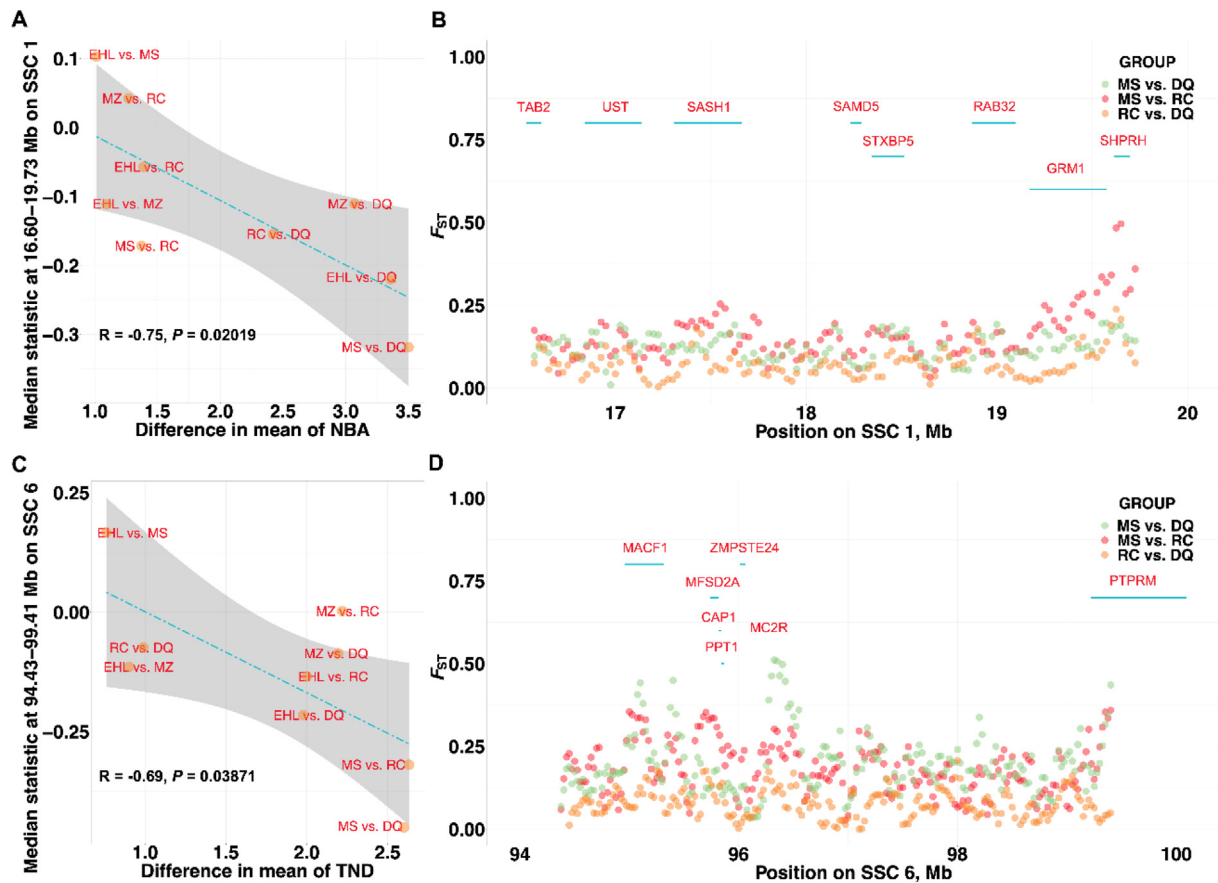


Fig. 5. Patterns of selection at the ROH: 16.60–19.73 Mb on SSC 1 and 94.43–99.41 Mb on SSC 6. (A) The correlation between the median genetic diversity ratio at the ROH 16.60–19.73 Mb on SSC 1 and the corresponding differences in number born alive trait between each pair of pig breeds; (B) The F_{ST} statistics at the ROH 16.60–19.73 Mb on SSC 1 in selected breed comparisons; (C) The correlation between the median genetic diversity ratio at the ROH 94.43–99.41 Mb on SSC 6 and the corresponding differences in total number of piglets born dead trait between each pair of pig breeds; (D) The F_{ST} statistics at the ROH 94.43–99.41 Mb on SSC 6 in selected breed comparisons.

with the lowest litter size detected by F_{ST} analysis with group 1 (Supplementary Fig. 7).

3.6. The potential roles of the promising candidate genes in reproductive function

We further explored the potential roles of the genes (i.e., *CCDC91*, *SASH1*, *SAMD5*, *MACF1*, *MFSD2A*, *EPC2*, and *MBD5*) that have been recorded in the webTWAS database and not been widely reported to be fertility-related genes (Table 2). Interestingly, we found that the promising candidate genes were associated with polycystic ovary syndrome, thyroid problem, and diabetes.

According to the associated diseases and tissues given in the database, we further explored the changes in expression level of the genes from oocyte to 8-cell in three mammalian species (i.e., human, pig, and mouse), and the differences in expression level of the genes between GDM and healthy control placentas. We observed that the expression level of *CCDC91*, *MACF1*, *MFSD2A*, *EPC2*, and *MBD5* has a significant fluctuation during the development of early embryo in at least one out of three species (Fig. 6A–C). In addition, *SASH1* and *SAMD5* were only detected from oocyte to 8-cell in mouse and showed a significantly high expression level at the period of oocyte comparison with other periods (Supplementary Fig. 9). The results elucidated the potential roles of the promising candidate genes in the early embryonic develop-

Table 2
The potential roles of the genes that have not been widely reported to be fertility-related genes.

Gene symbol	Pig gene ID	Human Gene ID	Associated disease	Associated tissues
<i>CCDC91</i>	ENSSSCG00000024232	ENSG00000123106	Polycystic ovary syndrome [52], Type 2 Diabetes [53], Osteoarthritis [54]	Ovary, testis, brain hippocampus
<i>SASH1</i>	ENSSSCG00000033175	ENSG00000123106	Thyroid problem, Hypothyroidism/myxoedema [55]	Thyroid, brain hippocampus
<i>SAMD5</i>	ENSSSCG00000004112	ENSG00000203727	Thyroid problem, Local infections of skin and subcutaneous tissue, Diseases of veins, lymphatic vessels and lymph nodes [55]	Brain hypothalamus, spleen, testis
<i>MACF1</i>	ENSSSCG00000003654	ENSG00000127603	Type 2 Diabetes, Diabetes [56], hypertensive diseases [55]	Ovary, pancreas, thyroid, brain hippocampus, pituitary, spleen, liver
<i>MFSD2A</i>	ENSSSCG00000003669	ENSG00000168389	Type 2 Diabetes [53], Diabetes mellitus, hypertension [55], diabetes [56]	Uterus
<i>EPC2</i>	ENSSSCG00000027211	ENSG00000135999	Diabetes [56]	Liver
<i>MBD5</i>	ENSSSCG00000015667	ENSG00000204406	Diabetes mellitus [55]	Colon Sigmoid

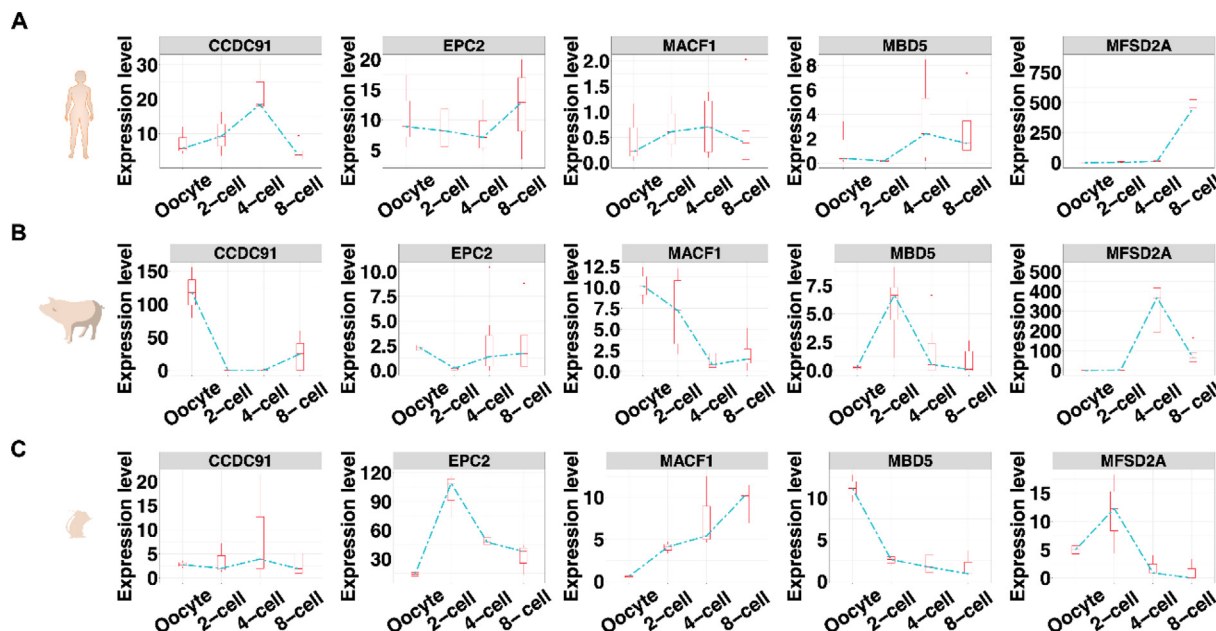


Fig. 6. Comparative transcriptome analysis on the promising candidate genes. The expression level of *CCDC91*, *MACF1*, *MFSD2A*, *EPC2*, and *MBD5* during the development of early embryo in human (A), pig (B), and mouse (C).

ment of mammals. Furthermore, the expression level of *MBD5* and *SAMD5* were significantly different (t -test, P -value = 0.03549 and 0.01884, respectively) between GDM and healthy control placentas (Supplementary Fig. 10).

4. Discussion

4.1. The factors affecting ROHet and ROH detection

In general, genotyping errors, the minimal SNP density per segment, the minimum length of a ROHet or ROH, the number of genotyping errors allowed and the minimum number of SNPs included in a ROHet or ROH are likely to affect the detection of ROHet and ROH [17,57–59]. These factors are especially relevant for populations with fewer number of individuals, which may lack pedigree records, therefore, genomic analyses help fill the gap to their genetic characterization. The density of the SNP chip is related to the minimum ROH length, fewer and shorter ROH will be detected with higher density of SNP [57]. Furthermore, the minimal SNP density affect genome coverage of the ROH analysis and is crucial for medium density genotypes [33]. With medium density SNP panel, one heterozygous call is allowed since genotyping errors in SNP chip data may occur [57], and allowing a number of genotyping errors in long ROH may minimize the underestimation of them [58]. In addition, Meyermans et al. [33] has shown that MAF and LD pruning severely impact ROH analyses and recommend to avoid MAF and LD pruning prior to ROH analyses based on medium density genotypes. Meanwhile, ROHet has far less characterized than ROH in livestock, especially in pigs [16,60]. Biscarini et al. [57] presented the results of the first sensitivity analysis for ROHet and varied the number of missing and homozygous SNPs allowed inside ROHet. In comparison with the detection of ROH, they found a significant increase in the number and average size of the detected ROHet when increasing numbers of missing and/or homozygous SNPs are allowed.

4.2. Distribution of heterozygous and homozygous loci

According to previous studies [61,62], indigenous pigs studied in this study had lower heterozygosity than intensively-reared

commercial pig breeds. The results of ROHet detection showed that a few ROHet were detected in indigenous pigs, except LY pigs, which might have gene introgression from intensively-reared commercial pig breeds (data not shown). A study reported that Mangalarga Marchador horses were inbred, leading to reduced heterozygosity and only a few ROHet were formed [19]. In this study, the number of ROHet were lower than that of ROH, which reflected the breed conservation of the studied pig populations. In addition, the number and length of ROH or ROHet had negative correlations, which have been observed in previous studies [57,63]. In spite of this heterogeneity of methodology and data, all studies found that heterozygosity-regions were much rarer and shorter compared to ROH.

4.3. The potential roles of the shared ROHet and ROH

The shared ROHet showed the possibility of similar selection patterns on growth, meat quality, health, and reproduction traits in Chinese indigenous pigs. Here, we noticed the tissue-specific genes located on the shared ROHet, which might reflect the same breed characteristics of indigenous pigs more typically. The *ceruloplasmin (CP)* gene potential participated in iron transport during spleen development [64] and the growth of piglets [65]. The *timeless circadian regulator (TIMELESS)* gene was linked to cell survival after damage or stress [66], which might help disease resistance of indigenous pigs. Two testis tissue-specific genes, the *testis specific serine kinase substrate (TSKS)* and *dynein axonemal heavy chain 17 (DNAH17)* genes, were essential for spermatogenesis and fertility in mammals [67,68].

Likewise, we investigated the tissue-specific genes at the shared ROH to uncover the same breed characteristics of the studied populations. The *ubiquitin like with PHD and ring finger domains 1 (UHRF1)* gene involved in porcine adipogenesis [69]. The *GC vitamin D binding protein (GC)* gene was associated with bone metabolic pathways [70]. The *alpha fetoprotein (AFP)* gene encodes a major fetal protein and may promote fetal ovarian follicular development in Chinese indigenous pigs [71]. To sum up, the shared ROH might cause broadly similar fat deposition, litter size and body length characteristics in indigenous pigs.

4.4. Litter traits-related ROHet and ROH detection

In this study, we highlighted the genetic characteristics of litter traits in indigenous pigs. The highest frequency litter traits-related ROHet within the genome might be able to achieve optimal production performance in indigenous pigs. We identified three highest frequency litter traits-related ROHet among parity groups and each litter traits. Several promising candidate genes at the ROHet involved in reproductive pathways in mammals. The *calcium/calmodulin dependent protein kinase II delta (CAMK2D)* gene participated in GnRH signalling pathway, as one of the most important reproductive pathways [72]. The *UDP glycosyltransferase 8 (UGT8)* gene was down-regulated in the endometrium of women affected by implantation failure [73]. Whereas, we noticed that these genes had pleiotropic effects on the growth of muscle and fat deposition in adipose tissues [74,75]. This implied these genes might lead to lighter piglets born, and give birth to more piglets.

High levels of inbreeding are directly related to a higher incidence of ROH, which could result in inbreeding depression [76]. Previous studies reported several genes likely caused fetal mortality in mutant homozygous pigs [77]. Here, the roles of the *activin A receptor type 2A (ACVR2A)*, *kinesin family member 5C (KIF5C)* and *Metabolism of cobalamin associated D (MMADHC)* gene in reproductive pathways have been previously reported [78–80]. In the meantime, we reported for the first time the potential effects of other two promising candidate genes (*EPC2* and *MBD5*) on fertility.

In addition, the analyses of association degree between the highest frequency ROHet (or ROH) and litter traits revealed that the genomic structures of piglets born alive and dead were different among parity groups, which were consistent with previous studies [3,81]. In view of the vary genetic background of traits of born alive and dead among parities, the different parities should be considered as different traits to proceed selection and breeding works.

4.5. Porcine litter traits-related regions detection via selection signatures

Furthermore, we deeply investigated the roles of the highest frequency litter traits-related ROHet and ROH in pigs via litter traits-specific selection signatures. There was no trend of selection patterns has been observed in comparison of the θ_{pi} of JH, CA, LY, and SX pigs across the highest frequency ROHet and ROH. These might be caused by the similar reproduction performance of these pigs. Thus, we used the high-density genotypes of other five pig breeds with different levels of reproduction performance to detect selection signatures. A set of known and novel candidate genes associated with litter size were located on the peak of selection. The *ovochoymase 1 (OVCH1)* and *coiled-coil domain containing 91 (CCDC91)* genes were two strong candidates with the peak of selection across the ROHet that were related to TNB and NBA traits. Herein, The *OVCH1* gene encodes oocyte extracellular polyproteins and has pleiotropic effects on fertility and the development of muscle [82,83]. And, it has been established that the *CCDC91* gene involved in skeletal development [84], but its potential role in litter traits remains unknown.

As mentioned above, the genes of which were homozygous might lead to fetal mortality. The ROH: 16.60–19.73 Mb on SSC 1 and ROH: 94.43–99.41 Mb on SSC 6 were overlapped with several QTLs of the loss of reproduction traits, e.g., non-functional nipples, number of mummified pigs, and number of stillborn. In addition, the promising candidate genes located across the ROH were associated with embryonic and fetal development. The *microtubule actin crosslinking factor 1 (MACF1)* gene has been linked to ovarian function [85]. The *major facilitator superfamily domain containing 2a (MFSD2A)* gene was needed in body growth, lipid metabolism and

brain integrity [86]. The TGF-beta activated kinase 1 binding protein 2 (TAB2) gene were associated with non-heterotaxic fetuses [87].

4.6. The promising candidate genes with reproductive function

Previous studies have showed that the homozygous variants could lead to human infertility with early embryonic arrest [88] and zygotic cleavage failure [89]. In this study, we deeply explored the potential fertility-related function of the promising candidate genes that have not been widely investigated. The expression level of *CCDC91*, *MACF1*, *MFSD2A*, *EPC2*, and *MBD5* has a significant fluctuation during the development of early embryo, but showed different trends in human, pig, and mouse. The phenomenon was consistent with previous studies, which stated that the molecular mechanisms of embryonic development shared many similarities, but also several marked differences in mammals [90,91]. In addition, the infants born to GDM females had reduced size at birth and persisting reductions in adiposity [92]. *MBD5* and *SAMD5* likely involved in the occurrence of GDM. The results showed that the promising candidate genes played potential roles in mammalian fertility-related pathways.

4.7. The use of the promising biological priors

To our knowledge, few ROHet and ROH analyses have been systematically conducted in pigs. And amounts of “omics” data were generated and released by some consortiums (e.g., the FarmGTEx [93] and FAANG consortium [94]), providing opportunities to explore the potential roles of the causal genome regions and promising candidate genes. Here, we integrated the findings of ROHet and ROH with transcriptome analyses to better elucidate the potential causal genes that determine the genetic and phenotypic differences among animals. Many studies have proved that modifying the models to incorporate prior knowledge (e.g., the potential causal genes) improves the performance of genomic prediction or selection [95–97]. Meanwhile, animal models have been widely used for the understanding of human physiological processes. The porcine anatomy and physiology are similar to human anatomy and physiology (e.g., all-eating and hormonal cycle), therefore, some breeds with size and weight similar to human adults making the breeds attractive in biomedical research [98,99]. In this study, the in-depth cross-species transcriptome comparison reveals the conserved and divergent features of the promising candidate genes in human, mouse, and pig, the results will benefit animal model selection for fertility-related traits and/or diseases.

Taken together, we present a multi-omics analyses integrating transcriptome and genome analyses that captures the underlying chromosome regions of litter traits. The in-depth analyses provided the evidence for the roles of the promising candidate genes that can be used as biological priors to improve the accuracies of genomic selection and further benefit pig genetic improvement. Besides, the results offer a genome-wide understanding of the relationships between heterozygosity or homozygosity regions and reproductive success and the loss of reproduction in mammals.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (Grant number: U21A20249, 31941007, 31972534), Zhejiang Science and Technology Major Program on Agricultural New Variety Breeding (Grant number: 2021C02068-1), and National Key Research and Development Program of China (2021YFD1200802).

Author statement

Q.S.W. and Y.C.P. conceived the study. Z.W. and Z.Z. collected samples. All authors contributed to experimental design or data interpretation. Z.T.C. and Z.Y.Z. carried out experiments and performed data analysis. Z.T.C. wrote the manuscript. All authors edited and approved the final draft.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2022.07.039>.

References

- Chen Z, Li Y, Zhang Z, Zhao W, Zhang Z, Xiang Y, et al. Genome-wide epistatic interactions of litter size at birth in Chinese indigenous pigs. *Anim Genet* 2021;52:739–43. <https://doi.org/10.1111/age.13120>.
- Zheng X, Zhao P, Yang K, Ning C, Wang H, Zhou L, et al. CNV analysis of Meishan pig by next-generation sequencing and effects of AHR gene CNV on pig reproductive traits. *J Anim Sci Biotechnol* 2020. <https://doi.org/10.1186/s40104-020-00442-5>.
- Chen Z, Ye S, Teng J, Diao S, Yuan X, Chen Z, et al. Genome-wide association studies for the number of animals born alive and dead in duroc pigs. *Theriogenology* 2019;139:36–42. <https://doi.org/10.1016/j.theriogenology.2019.07.013>.
- Norseeda W, Liu G, Teltatham T, Supakankul P, Sringarm K, Naraballoh W, et al. Association of il-4 and il-4r polymorphisms with litter size traits in pigs. *Animals* 2021. <https://doi.org/10.3390/ani11041154>.
- Ding R, Qiu Y, Zhuang Z, Ruan D, Wu J, Zhou S, et al. Genome-wide association studies reveals polygenic genetic architecture of litter traits in Duroc pigs. *Theriogenology* 2021;173:269–78. <https://doi.org/10.1016/j.theriogenology.2021.08.012>.
- Wu P, Wang K, Zhou J, Yang Q, Yang X, Jiang A, et al. A genome wide association study for the number of animals born dead in domestic pigs. *BMC Genet* 2019. <https://doi.org/10.1186/s12863-018-0692-x>.
- Sell-Kubiak E, Duijvesteijn N, Lopes MS, Janss LLG, Knol EF, Bijma P, et al. Genome-wide association study reveals novel loci for litter size and its variability in a Large White pig population. *BMC Genomics* 2015. <https://doi.org/10.1186/s12864-015-2273-y>.
- Ye J, Tan C, Hu X, Wang A, Wu Z. Genetic parameters for reproductive traits at different parities in Large White Pigs. *J Anim Sci* 2018. <https://doi.org/10.1093/jas/sky066>.
- Sell-Kubiak E. Selection for litter size and litter birthweight in Large White pigs: Maximum, mean and variability of reproduction traits. *Animal* 2021. <https://doi.org/10.1016/j.animal.2021.100352>.
- Zhang C, Ye He, Huang Y, et al. Genome-wide association study for reproductive traits in a duroc pig population. *Animals* 2019;9:732. <https://doi.org/10.3390/ani9100732>.
- Zhang Z, Ober U, Erbe M, Zhang H, Gao N, He J, et al. Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. *PLoS ONE* 2014. <https://doi.org/10.1371/journal.pone.0093017>.
- Amaral AF, McQueen BE, Bellingham-Johnston K, Poston TB, Darville T, Nagarajan UM, et al. Host-pathogen interactions of chlamydia trachomatis in porcine oviduct epithelial cells. *Pathogens* 2021. <https://doi.org/10.3390/pathogens10101270>.
- Zhou X, He Y, Li N, Bai G, Pan X, Zhang Z, et al. DNA methylation mediated RSPO2 to promote follicular development in mammals. *Cell Death Dis* 2021. <https://doi.org/10.1038/s41419-021-03941-z>.
- Samuels DC, Wang J, Ye F, He J, Levinson RT, Sheng Q, et al. Heterozygosity ratio, a robust global genomic measure of autozygosity and its association with height and disease risk. *Genetics* 2016. <https://doi.org/10.1534/genetics.116.189936>.
- Marras G, Wood BJ, Mankajua B, Malchiodi F, Peeters K, Van AP, et al. Characterization of runs of homozygosity and heterozygosity-rich regions in a commercial turkey (Meleagris gallopavo) population. 11th World Congr Genet Appl to Livest Prod, 2018.
- Sanglard LP, Huang Y, Gray KA, Linhares DCL, Dekkers JCM, Niederwerder MC, et al. Further host-genomic characterization of total antibody response to PRRSV vaccination and its relationship with reproductive performance in commercial sows: genome-wide haplotype and zygosity analyses. *Genet Sel Evol* 2021;53:91. <https://doi.org/10.1186/s12711-021-00676-5>.
- Peripolli E, Munari DP, Silva MVGB, Lima ALF, Irgang R, Baldi F. Runs of homozygosity: current knowledge and applications in livestock. *Anim Genet* 2017. <https://doi.org/10.1111/age.12526>.
- Forutan M, Ansari Mahyari S, Baes C, Melzer N, Schenkel FS, Sargolzaei M. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. *BMC Genomics* 2018. <https://doi.org/10.1186/s12864-018-4453-z>.
- Bizarria dos Santos W, Pimenta Schettini G, Fonseca MG, Pereira GL, Loyola Chardulo LA, Rodrigues Machado Neto O, et al. Fine-scale estimation of inbreeding rates, runs of homozygosity and genome-wide heterozygosity levels in the Mangalarga Marchador horse breed. *J Anim Breed Genet* 2021. <https://doi.org/10.1111/jbg.12508>.
- Nosrati M, Asadollahpour Nanaei H, Javanmard A, Esmailizadeh A. The pattern of runs of homozygosity and genomic inbreeding in world-wide sheep populations. *Genomics* 2021. <https://doi.org/10.1016/j.ygeno.2021.03.005>.
- Yang HC, Chang LC, Liang YJ, Lin CH, Wang PL. A genome-wide homozygosity association study identifies runs of homozygosity associated with rheumatoid arthritis in the human major histocompatibility complex. *PLoS ONE* 2012. <https://doi.org/10.1371/journal.pone.0034840>.
- Wen J, Zheng Z, Gong M, Li D, Hu S, Cai Y, et al. Ancient genomes reveal the genetic inheritance and recent introgression in Chinese indigenous pigs. *Sci China Life Sci* 2022. <https://doi.org/10.1007/s11427-021-2066-y>.
- Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat Genet* 2015. <https://doi.org/10.1038/ng.3199>.
- Jiang Y, Li X, Liu J, Zhang W, Zhou M, Wang J, et al. Genome-wide detection of genetic structure and runs of homozygosity analysis in Anhui indigenous and Western commercial pig breeds using PorcineSNP80k data. *BMC Genomics* 2022;23:1–14. <https://doi.org/10.1186/s12864-022-08583-9>.
- Wu W, Zhan J, Tang X, Li T, Duan S. Characterization and identification of pork flavor compounds and their precursors in Chinese indigenous pig breeds by volatile profiling and multivariate analysis. *Food Chem* 2022;385:1–8. <https://doi.org/10.1016/j.foodchem.2022.132543>.
- Liu C, Li P, Zhou W, Ma X, Wang X, Xu Y, et al. Genome data uncover conservation status, historical relatedness and candidate genes under selection in Chinese indigenous pigs in the taihu lake region. *Front Genet* 2020. <https://doi.org/10.3389/fgene.2020.00591>.
- Huang M, Zhang H, Wu ZP, Wang XP, Li DS, Liu SJ, et al. Whole-genome resequencing reveals genetic structure and introgression in Pudong White pigs. *Animal* 2021;15. <https://doi.org/10.1016/j.animal.2021.100354>.
- Zhao P, Yu Y, Feng W, Du H, Yu J, Kang H, et al. Evidence of evolutionary history and selective sweeps in the genome of Meishan pig reveals its genetic and phenotypic characterization. *GigaScience* 2018. <https://doi.org/10.1093/gigascience/giy058>.
- Chen H, Huang M, Yang B, Wu Z, Deng Z, Hou Y, et al. Introgression of Eastern Chinese and Southern Chinese haplotypes contributes to the improvement of fertility and immunity in European modern pigs. *GigaScience* 2020. <https://doi.org/10.1093/gigascience/giaa014>.
- Dobrzański J, Mulder HA, Knol EF, Szwaczkowski T, Sell-Kubiak E. Estimation of litter size variability phenotypes in Large White sows. *J Anim Breed Genet* 2020. <https://doi.org/10.1111/jbg.12465>.
- VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci* 2008;91:4414–23. <https://doi.org/10.3168/jds.2007-0980>.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75. <https://doi.org/10.1086/519795>.
- Meyermans R, Gorssen W, Buys N, Janssens S. How to study runs of homozygosity using plink? a guide for analyzing medium density snp data in livestock and pet species. *BMC Genom*. 2020. <https://doi.org/10.1186/s12864-020-6463-x>.
- Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26:589–95. <https://doi.org/10.1093/bioinformatics/btp698>.
- Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 2011. <https://doi.org/10.1093/bioinformatics/btr509>.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303. <https://doi.org/10.1101/gr.107524.110>.
- Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsan P, Valentini A, et al. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim Genet* 2015. <https://doi.org/10.1111/age.12259>.
- Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, et al. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci U S A* 2007. <https://doi.org/10.1073/pnas.0710021104>.
- Dumitrescu AM, Liao XH, Abdullah MSY, Lado-Abel J, Majed FA, Moeller LC, et al. Mutations in SECISBP2 result in abnormal thyroid hormone metabolism. *Nat Genet* 2005. <https://doi.org/10.1038/ng1654>.

- [40] Storey JD, Xiao W, Leek JT, Tompkins RG, Davis RW. Significance analysis of time course microarray experiments. *Proc Natl Acad Sci U S A* 2005. <https://doi.org/10.1073/pnas.0504609102>.
- [41] International Encyclopedia of Statistical Science. 2011. <https://doi.org/10.1007/978-3-642-04898-2>.
- [42] Team RC. *R: A Language and Environment for Statistical Computing*. R Found Stat Comput 2021.
- [43] Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics* 2011;27:2156–8. <https://doi.org/10.1093/bioinformatics/btr330>.
- [44] Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution* (N Y) 1984;38:1358. <https://doi.org/10.2307/2408641>.
- [45] Karlsson M, Sjöstedt E, Oksvold P, Sivertsson A, Huang J, Álvarez MB, et al. Genome-wide annotation of protein-coding genes in pig. *BMC Biol* 2022;20:1–18. <https://doi.org/10.1186/s12915-022-01229-v>.
- [46] Cao C, Wang J, Kwok D, Cui F, Zhang Z, Zhao D, et al. WebTWAS: A resource for disease candidate susceptibility genes identified by transcriptome-wide association study. *Nucleic Acids Res* 2022;50. <https://doi.org/10.1093/nar/gkab957>. D1123–D1130.
- [47] Xue Z, Huang K, Cai C, Cai L, Jiang CY, Feng Y, et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 2013. <https://doi.org/10.1038/nature12364>.
- [48] Kong Q, Yang X, Zhang H, Liu S, Zhao J, Zhang J, et al. Lineage specification and pluripotency revealed by transcriptome analysis from oocyte to blastocyst in pig. *FASEB J* 2020. <https://doi.org/10.1096/fj.201901818RR>.
- [49] Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 2015. <https://doi.org/10.1038/nmeth.3317>.
- [50] Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014. <https://doi.org/10.1093/bioinformatics/btt656>.
- [51] Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omi A J Integr Biol* 2012;16:284–7. <https://doi.org/10.1089/omi.2011.0118>.
- [52] Day F, Karaderi T, Jones MR, Meun C, He C, Drong A, et al. Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *PLoS Genet* 2018. <https://doi.org/10.1371/journal.pgen.1007813>.
- [53] Xue A, Wu Y, Zhu Z, Zhang F, Kemper KE, Zheng Z, et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun* 2018. <https://doi.org/10.1038/s41467-018-04951-w>.
- [54] Tachmazidou I, Hatzikotoulas K, Southam L, Esparza-Gordillo J, Haberland V, Zheng J, et al. Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. *Nat Genet* 2019. <https://doi.org/10.1038/s41588-018-0327-1>.
- [55] Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nat Genet* 2018. <https://doi.org/10.1038/s41588-018-0248-z>.
- [56] Watanabe K, Stringer S, Frei O, Umičević Mirkov M, de Leeuw C, Polderman TJC, et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* 2019. <https://doi.org/10.1038/s41588-019-0481-0>.
- [57] Biscarini F, Mastrangelo S, Catillo G, Senczuk G, Ciampolini R. Insights into genetic diversity, runs of homozygosity and heterozygosity-rich regions in maresmanna semi-feral cattle using pedigree and genomic data. *Animals* 2020. <https://doi.org/10.3390/ani10122285>.
- [58] Ferencaković M, Sölkner J, Curik I. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet Sel Evol* 2013. <https://doi.org/10.1186/1297-9686-45-42>.
- [59] Zhang Z, Zhang Q, Xiao Q, Sun H, Gao H, Yang Y, et al. Distribution of runs of homozygosity in Chinese and Western pig breeds evaluated by reduced-representation sequencing data. *Anim Genet* 2018. <https://doi.org/10.1111/age.12730>.
- [60] Ruan D, Yang J, Zhuang Z, Ding R, Huang J, Quan J, et al. Assessment of heterozygosity and genome-wide analysis of heterozygosity regions in two duroc pig populations. *Front Genet* 2021;12. <https://doi.org/10.3389/fgene.2021.812456>.
- [61] Chen J, Peng J, Xiao Q, Pan Y, Zhang X, Lo LJ, et al. The genetic diversity and population structures of indigenous pig breeds in Zhejiang Province revealed by GGRS sequencing. *Anim Genet* 2018. <https://doi.org/10.1111/age.12625>.
- [62] Zhang Z, Xiao Q, Zhang Q, Sun H, Chen J, Li Z, et al. Genomic analysis reveals genes affecting distinct phenotypes among different Chinese and western pig breeds. *Sci Rep* 2018;8:13352. <https://doi.org/10.1038/s41598-018-31802-x>.
- [63] Huo JL, Zhang LQ, Zhang X, Wu XW, Ye XH, Sun YH, et al. Genome-wide single nucleotide polymorphism array and whole-genome sequencing reveal the inbreeding progression of Banna minipig inbred line. *Anim Genet* 2022. <https://doi.org/10.1111/age.13149>.
- [64] Che T, Li D, Jin L, Fu Y, Liu Y, Liu P, et al. Long non-coding RNAs and mRNAs profiling during spleen development in pig. *PLoS ONE* 2018. <https://doi.org/10.1371/journal.pone.0193552>.
- [65] Tao X, Xu Z, Men X. Transient changes of enzyme activities and expression of stress proteins in the small intestine of piglets after weaning. *Arch Anim Nutr* 2015. <https://doi.org/10.1080/1745039X.2015.1034828>.
- [66] Fabbri MC, Dadousis C, Tiezzi F, Maltecca C, Lozada-Soto E, Biffani S, et al. Genetic diversity and population history of eight Italian beef cattle breeds using measures of autozygosity. *PLoS ONE* 2021. <https://doi.org/10.1371/journal.pone.0248087>.
- [67] Chen L, Ouyang J, Li X, Xiao X, Sun W, Li S, et al. DNAH17 is essential for rat spermatogenesis and fertility. *J Genet* 2021. <https://doi.org/10.1007/s12041-021-01264-8>.
- [68] Hao Z, Jha KN, Kim YH, Vemuganti S, Westbrook VA, Chertihin O, et al. Expression analysis of the human testis-specific serine/threonine kinase (TSSK) homologues. A TSSK member is present in the equatorial segment of human sperm. *Mol Hum Reprod* 2004. <https://doi.org/10.1093/molehr/gah052>.
- [69] Stachecka J, Lemanska W, Noak M, Szczerbal I. Expression of key genes involved in DNA methylation during in vitro differentiation of porcine mesenchymal stem cells (MSCs) into adipocytes. *Biochem Biophys Res Commun* 2020. <https://doi.org/10.1016/j.bbrc.2019.11.175>.
- [70] Delanghe JR, Speeckaert R, Speeckaert MM. Behind the scenes of vitamin D binding protein: more than vitamin D binding. *Best Pract Res Clin Endocrinol Metab* 2015. <https://doi.org/10.1016/j.beem.2015.06.006>.
- [71] Xu M, Che L, Wang D, Yang Z, Zhang P, Lin Y, et al. Proteomic analysis of fetal ovary reveals that ovarian developmental potential is greater in Meishan pigs than in Yorkshire pigs. *PLoS ONE* 2015. <https://doi.org/10.1371/journal.pone.0135514>.
- [72] Tahir MS, Porto-Neto LR, Gondro C, Shittu OB, Wockner K, Tan AWL, et al. Meta-analysis of heifer traits identified reproductive pathways in bos indicus cattle. *Genes* (Basel) 2021. <https://doi.org/10.3390/genes12050768>.
- [73] Maekawa R, Taketani T, Mihara Y, Sato S, Okada M, Tamura I, et al. Thin endometrium transcriptome analysis reveals a potential mechanism of implantation failure. *Reprod Med Biol* 2017. <https://doi.org/10.1002/rmb2.12030>.
- [74] Lee YS, Shin D, Song KD. Dominance effects of ion transport and ion transport regulator genes on the final weight and backfat thickness of Landrace pigs by dominance deviation analysis. *Genes Genom*. 2018. <https://doi.org/10.1007/s13258-018-0728-7>.
- [75] Arora D, Srikanth K, Lee J, Lee D, Park N, Wy S, et al. Integration of multi-omics approaches for functional characterization of muscle related selective sweep genes in Nanchukmacdon. *Sci Rep* 2021. <https://doi.org/10.1038/s41598-021-86683-4>.
- [76] Curik I, Ferencaković M, Sölkner J. Inbreeding and runs of homozygosity: a solution to an old problem. *Livest Sci* 2014. <https://doi.org/10.1016/j.livsci.2014.05.034>.
- [77] Derks MFL, Lopes MS, Bosse M, Madsen O, Dibbits B, Harlizius B, et al. Balancing selection on a recessive lethal deletion with pleiotropic effects on two neighboring genes in the porcine genome. *PLoS Genet* 2018. <https://doi.org/10.1371/journal.pgen.1007661>.
- [78] Li X, Ye J, Han X, Qiao R, Li X, Lv G, et al. Whole-genome sequencing identifies potential candidate genes for reproductive traits in pigs. *Genomics* 2020. <https://doi.org/10.1016/j.ygeno.2019.01.014>.
- [79] Kant K, Tomar AK, Sharma P, Kundu B, Singh S, Yadav S. Human epididymis protein 4 quantification and interaction network analysis in seminal plasma. *Protein Pept Lett* 2019. <https://doi.org/10.2174/0929866526666190327124919>.
- [80] Pupavac M, Garcia MAM, Rosenblatt DS, Jerome-Majewska LA. Expression of mmachc and mmadhac during mouse organogenesis. *Mol Genet Metab* 2011. <https://doi.org/10.1016/j.ymgme.2011.04.004>.
- [81] Roehre R, Kennedy BW. Estimation of genetic parameters for litter size in Canadian Yorkshire and Landrace swine with each parity of farrowing treated as a different trait. *J Anim Sci* 1995. <https://doi.org/10.2527/1995.73102959x>.
- [82] Varade J, Wang N, Lim CK, Zhang T, Zhang Y, Liu X, et al. Novel genetic loci associated HLA-B*08:01 positive myasthenia gravis. *J Autoimmun* 2018. <https://doi.org/10.1016/j.jaut.2017.10.002>.
- [83] Mino M, Sawada H. Follicle cell trypsin-like protease HrOvochymase: its cDNA cloning, localization, and involvement in the late stage of oogenesis in the ascidian *Halocynthia roretzi*. *Mol Reprod Dev* 2016. <https://doi.org/10.1002/mrd.22627>.
- [84] Hsu YH, Estrada K, Evangelou E, Ackert-Bicknell C, Akesson K, Beck T, et al. Meta-analysis of genomewide association studies reveals genetic variants for hip bone geometry. *J Bone Miner Res* 2019. <https://doi.org/10.1002/jbmr.3698>.
- [85] Jaillard S, Bell K, Aklou L, Walton K, McElreavy K, Stocker WA, et al. New insights into the genetic basis of premature ovarian insufficiency: novel causative variants and candidate genes revealed by genomic sequencing. *Maturitas* 2020. <https://doi.org/10.1016/j.maturitas.2020.06.004>.
- [86] Gázquez A, Ruiz-Palacios M, Larqué E. DHA supplementation during pregnancy as phospholipids or TAG produces different placental uptake but similar fetal brain accretion in neonatal piglets. *Br J Nutr* 2017. <https://doi.org/10.1017/S0007114517002951>.
- [87] Liu H, Giguët-Valard AG, Simonet T, Szenker-Ravi E, Lambert L, Vincent-Delorme C, et al. Next-generation sequencing in a series of 80 fetuses with complex cardiac malformations and/or heterotaxy. *Hum Mutat* 2020. <https://doi.org/10.1002/humu.24132>.
- [88] Xu Y, Wang R, Pang Z, Wei Z, Sun L, Li S, et al. Novel homozygous PADI6 variants in infertile females with early embryonic arrest. *Front Cell Dev Biol* 2022;10:1–8. <https://doi.org/10.3389/fcell.2022.819667>.
- [89] Zheng W, Zhou Z, Sha Q, Niu X, Sun X, Shi J, et al. Homozygous mutations in BTG4 cause zygotic cleavage failure and female infertility. *Am J Hum Genet* 2020. <https://doi.org/10.1016/j.ajhg.2020.05.010>.
- [90] Oestrup O, Hall V, Petkov SG, Wolf XA, Hyldig S, Hyttel P. From zygote to implantation: morphological and molecular dynamics during embryo

- development in the pig. *Reprod Domest Anim* 2009. <https://doi.org/10.1111/j.1439-0531.2009.01482.x>.
- [91] Carreiro LE, Dos SGS, Luedke FE, Goissis MD. Cell differentiation events in pre-implantation mouse and bovine embryos. *Anim Reprod* 2021;18:1–15. <https://doi.org/10.1590/1984-3143-AR2021-0054>.
- [92] Prentice PM, Olga L, Petry CJ, Simmons D, Murphy HR, Hughes IA, et al. Reduced size at birth and persisting reductions in adiposity in recent, compared with earlier, cohorts of infants born to mothers with gestational diabetes mellitus. *Diabetologia* 2019. <https://doi.org/10.1007/s00125-019-4970-6>.
- [93] Liu S, Gao Y, Canela-Xandri O, Wang S, Yu Y, Cai W, et al. A comprehensive catalogue of regulatory variants in the cattle transcriptome. *BioRxiv* 2020. <https://doi.org/10.1101/2020.12.01.406280>.
- [94] Clark EL, Archibald AL, Daetwyler HD, Groenen MAM, Harrison PW, Houston RD, et al. From FAANG to fork: application of highly annotated genomes to improve farmed animal production. *Genome Biol* 2020. <https://doi.org/10.1186/s13059-020-02197-8>.
- [95] Teng J, Huang S, Chen Z, Gao N, Ye S, Diao S, et al. Optimizing genomic prediction model given causal genes in a dairy cattle population. *J Dairy Sci* 2020. <https://doi.org/10.3168/jds.2020-18233>.
- [96] Fang L, Sahana G, Ma P, Su G, Yu Y, Zhang S, et al. Use of biological priors enhances understanding of genetic architecture and genomic prediction of complex traits within and between dairy cattle breeds. *BMC Genomics* 2017. <https://doi.org/10.1186/s12864-017-4004-z>.
- [97] Teng J, Ye S, Gao N, Chen Z, Diao S, Li X, et al. Incorporating genomic annotation into single-step genomic prediction with imputed whole-genome sequence data. *J Integr Agric* 2022. [https://doi.org/10.1016/S2095-3119\(21\)63813-3](https://doi.org/10.1016/S2095-3119(21)63813-3).
- [98] Bertho N, Meurens F. The pig as a medical model for acquired respiratory diseases and dysfunctions: an immunological perspective. *Mol Immunol* 2021. <https://doi.org/10.1016/j.molimm.2021.03.014>.
- [99] Gabriel GC, Devine W, Redel BK, Whitworth KM, Samuel M, Spate LD, et al. Cardiovascular development and congenital heart disease modeling in the pig. *J Am Heart Assoc* 2021. <https://doi.org/10.1161/JAHA.121.021631>.