Genome-wide association study and genomic prediction for intramuscular fat content in Suhuai pigs using imputed whole-genome sequencing data

Binbin Wang1,2,3 | Pinghua Li1,3 | Liming Hou1,3 | Wuduo Zhou1 | Wei Tao1,3 | Chenxi Liu1,3 | Kaiyue Liu1,3 | Peipei Niu3 | Zongping Zhang3 | Qiang Li4 | Guosheng Su2 | Ruihua Huang1,3 

1Key Laboratory in Nanjing for Evaluation and Utilization of Pigs Resources, Ministry of Agriculture and Rural Areas of China, Institute of Swine Science, Nanjing Agricultural University, Nanjing, China
2Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark
3Huaian Academy, Nanjing Agricultural University, China
4Huaiyin Xinhuai Pig Breeding Farm of Huaian City, China

Abstract

Integrating the single-nucleotide polymorphisms (SNPs) significantly affecting target traits from imputed whole-genome sequencing (iWGS) data into the genomic prediction (GP) model is an economic, efficient, and feasible strategy to improve prediction accuracy. The objective was to dissect the genetic architecture of intramuscular fat content (IFC) by genome wide association studies (GWAS) and to investigate the accuracy of GP based on pedigree-based BLUP (PBLUP) model, genomic best linear unbiased prediction (GBLUP) models and Bayesian mixture (BayesMix) models under different strategies. A total of 482 Suhuai pigs were genotyped using an 80K SNP chip. Furthermore, 30 key samples were selected for resequencing and were used as a reference panel to impute the 80K chip data to the WGS dataset. The 80K data and iWGS data were used to perform GWAS and test GP accuracies under different scenarios. GWAS results revealed that there were four major regions affecting IFC. Two important functional candidate genes were found in the two most significant regions, including protein kinase C epsilon (PRKCE) and myosin light chain 2 (MYL2). The results of the predictions showed that the PBLUP model had the lowest reliability (0.096 ± 0.032). The reliability (0.229 ± 0.035) was improved by replacing pedigree information with 80K chip data. Compared with using 80K SNPs alone, pruning iWGS SNPs with the R-squared cutoff of linkage disequilibrium (0.55) led to a slight improvement (0.006), adding significant iWGS SNPs led to an improvement of reliability by 0.050 when using a one-component GBLUP, a further increase of 0.033 when using...
INTRODUCTION

With the improvement of people's living standards, pork quality has attracted more and more attention. Intramuscular fat content (IFC) is an important determinant of pork quality affecting multiple other meat quality indicators, such as flavor, shearing force, and texture (Cho et al., 2015). Moreover, the appearance of pork, such as marbling and meat color, will strongly influence the purchase intention of consumers (Ngapo, 2017). It is generally accepted that a higher level of IFC has a positive influence on the sensory experience associated with eating (Wood et al., 2004). Thus, there is an increased requirement for genetic improvement of IFC in pig breeding programs. However, genetic improvement of IFC by a traditional breeding method is challenging because the trait can only be measured accurately after slaughter.

Animal breeding has been revolutionized by genomic prediction (GP). In recent years, genomic information has been widely applied in livestock breeding, which greatly increases selection accuracy, particularly for those traits with low heritability, difficult or expensive to measure (e.g. meat quality) (Jia & Jannink, 2012). In GP, the selection of candidate animals is based on the genomic estimated breeding value (GEBV) using genome-wide dense single-nucleotide polymorphisms (SNPs). Some popular genomic evaluation approaches, such as genomic best linear unbiased prediction (GBLUP) and Bayesian variable selection methods, have been widely used in practical genomic evaluation (Salek Ardestani et al., 2021).

Genomic prediction using SNP chip genotype data depends on linkage disequilibrium (LD) between causal genes and SNP markers. It is expected that the GP using whole-genome sequence (WGS) data will increase prediction accuracy because WGS data included causal SNPs affecting target traits, and the prediction is much less dependent on LD between SNPs and causal SNPs (Meuwissen & Goddard, 2010; Song et al., 2019). Despite the cost of WGS is rapidly decreasing, resequencing a large number of individuals is still expensive. An efficient approach to increase the number of animals with WGS SNP is to impute genotypes of chips to WGS (Larmer et al., 2017). Imputation is the process of inferring missing genotypes, such as inferring unknown genotypes for animals genotyped at lower SNP density by using a reference set of animals genotyped at a higher SNP density. However, a previous study showed that the direct use of imputed WGS (iWGS) data for prediction did not improve prediction accuracy (van Binsbergen et al., 2015). On the other hand, several studies have demonstrated that the accuracy of GP could be improved by adding significant SNPs which were detected by genome-wide association studies (GWAS) from WGS data (Brøndum et al., 2015; Warburton et al., 2020). Genetic variation affecting quantitative traits is ubiquitous in nature yet. Evolutionary force changes the allele frequency of a series of variations and results in changes in specific traits. GWAS is the most commonly used method to detect these genetic variations (Josephs et al., 2017). We speculated that the phenotypic variation of IFC in Suhuai pigs might also be affected by artificial selection, resulting in the change of allele frequency at several SNPs, which were significantly associated with IFC. A previous study found that the accuracy of genomic prediction could be improved by pruning the LD between SNPs in iWGS data (Ye et al., 2019). Moreover, various models to integrate important SNPs for GP have been used, such as one- or two-component GBLUP, Bayesian mixture model and so on (Liu et al., 2020).

The accuracy of genomic selection depends on size of reference population. For the trait such as IFC, it is difficult to have a large reference population. It is unclear whether iWGS data can give an extra contribution to accuracy of genomic prediction compared with SNP chip data, in the case of a small reference population. The objectives of this study were (1) to identify significant SNPs associated with IFC and (2) to improve the accuracy of genomic prediction by integrating significant SNPs into the prediction models.

MATERIALS AND METHODS

Ethics approval

All experimental animals in this study were carried out in accordance with the Guidelines for the Care and Use of Laboratory Animals prepared by the Institutional Animal Welfare and Ethics Committee of Nanjing Agricultural University, Nanjing, China (Certification No: SYXK [Su] 2017–0007).

Pig population

A total of 482 Suhuai pigs from Huaiyin Xinhui Pig Breeding Farm (Huaiian, China) were randomly selected for this study, including 291 barrows and 191 females. These 482 pigs included 330 pigs used in our previous study (Wang et al., 2019). Briefly, these 482 Suhuai pigs mainly originated from 80 sires and 226 dams in 15 consanguinities.
All growing-finishing pigs were fed same diet and ad libitum with free access to water under standard indoor conditions. Finally, they were slaughtered at 80–90 kg (about 220 days) live weight in four batches at the same slaughterhouse from 2017 to 2019. After slaughter, longissimus dorsi muscle samples from the last rib of the left side carcass and ear samples were collected.

2.3 Phenotypic data

IFC accumulates both in (intramyocellular) and out (extramyocellular) of the muscle fibers. The IFC of 482 Suhuai pigs was measured by the Soxhlet extraction method according to a standard procedure, as described by Supakankul and Mekchay (Supakankul & Mekchay, 2016). In addition to the original phenotypes, corrected phenotypes were calculated as genomic estimated breeding values (GEBV) plus the estimated residual from a GBLUP model with 80K phenotypes were calculated as genomic estimated breeding values (GEBV) plus the estimated residual from a GBLUP model with 80K SNP chip data (see the model below). The corrected phenotypes were used for subsequent GWAS and calculated the reliability of prediction in the validation population.

2.4 Genotype data

The genomic DNA was isolated from ear tissue using conventional phenol-chloroform extraction method (Sambrook et al., 1982). Only high-quality genomic DNAs from 482 samples were genotyped using GeneSeek GGP Porcine 80K SNP chip (Neogen), which contained a total of 68,516 SNPs (Banerjee et al., 2020). The physical positions of SNPs in 80K SNP chip were based on the Sus scrofa 10.2 genome of Duroc breed. To facilitate the subsequent genotype imputation, we converted the physical positions of 80K SNP chip data to the Sus scrofa 11.1 by USCS website (https://genome.ucsc.edu/cgi-bin/hgLiftOver). These 80K chip dataset of 482 samples included 96 sample SNP data in our previous study (Wang et al., 2021). Quality control of genotype data was conducted using PLINK (v1.90 beta) (Purcell et al., 2007) to detect and exclude unreliable genotypes. The SNPs were removed if any of the following criteria were involved: (a) no chromosomal or physical location, (b) non-autosomal variants, (c) minor allele frequency (MAF) < 0.01, or (d) call rate < 0.9. In addition, the individuals missing more than 10% of genotypes were removed. After filtering, 50,482 SNPs for 482 individuals were retained in the dataset. The number of SNPs after filtering was similar to other studies (Banerjee et al., 2020).

2.5 Whole-genome resequencing and variant calling

According to the pedigree data of all Suhuai pigs in the breeding farm, which contain 15 lineages, one male and one female from each lineage (a total of 30 individuals) were selected for resequencing. The resequencing was performed by the MGISEQ-2000 sequencing platform. The details of 30 sequenced pigs are shown in Table S1.

In this study, about 1.0 Tb of clean data was generated from the 30 individuals using Fastp (v0.23.1) with the command ‘-c -q 20 -u 50 -n 15 -5 20 -3 20 -w 10’ and 978.2 Gb were mapped to the pig reference genome using BWA (v0.7.8) software with the command ‘mem -t 10 -M -R’. In this study, Sus scrofa 11.1 genome of Duroc breed was selected as the reference genome. Suhuai pig contains Chinese local pig and European pig lineage, however, compared with the reference genome of Chinese local pigs (such as Meishan pig) (Zhou et al., 2021), Luchuan pig (Yang et al., 2019), and European wild boar (Groenen et al., 2012), the reference genome of Duroc is the standard reference genome by NCBI (https://www.ncbi.nlm.nih.gov/databank/). The annotation information of the Duroc’s reference genome is more complete and highly recognized, and is maintained in real time by NCBI (Warr et al., 2020). Furthermore, an important point was that the physical positions of the converted 80K chip data were based on Duroc’s reference genome, in order to improve the accuracy of subsequent imputation, the reference genome of Duroc was selected finally. Among all selected individuals, the average read mapping rates was 95.81%, the minimum value and maximum value were 94.98% and 96.55%, respectively. The possible reason was that the quality of several bases in front of each read was relatively low, which makes the mapping rate slightly low. The average uniquely mapped reads rates were 94.17% (from 93.07% to 95.87%), and the average depth of coverage was 13.30-fold (from 10.43-fold to 15.44-fold). SNP calling and filtering were implemented according to the GATK Best Practices pipeline (Van der Auwera et al., 2013) and VariantFiltration options. The filter parameters were as follows: QualByDepth (QD) ≥ 2.0, Fisher-Strand (FS) > 60.0, strand odds ratio (SOR) > 3.0, mapping qualities of reads (MQRankSum) < −12.5, ranked sum test for the distance of alleles from the end of the reads (ReadPosRankSum) < −8.0. A total of 23,765,813 SNPs were retained in the 30 Suhuai pigs. Finally, the 80K SNP chip data and resequencing data were phased by Beagle 5.1 (Browning et al., 2018) for subsequent imputation.

2.6 Imputation

The imputation from the 80K SNP chip data to WGS genotypes was performed by Beagle 5.1 (Browning et al., 2018) under the default parameter settings. During the imputing process, 30 individuals with resequencing data were used as the reference panel, and 482 individuals with 80K SNP chip data were used as the target panel. Imputation quality control of iWGS data was based on the dosage R-squared (DR2) in Beagle, which is an estimate of the squared correlation between the estimated allele dose and the true allele dose. Meanwhile, we found that some SNPs originated from 80K data were missing after imputation, so, the genotypes of 80K SNP chip were integrated into the iWGS data and replaced the imputed ones. To validate imputation accuracy, we calculated correlation between imputed and observed genotypes from five replicates under the condition that DR2 was greater than 0.9. In each replicate, 2000 SNPs in the 80K SNP chip data of 482 individuals were randomly masked, and then imputed them. The imputed genotypes were compared
with chip panel genotypes to calculate the correlation coefficient. Finally, SNPs with MAF less than 0.01 and DR² less than 0.9 were removed, and 8,103,716 SNPs with high imputing accuracy were remained for subsequent GWAS.

2.7 | Genome-wide association study

A single marker GWAS was performed with a linear model using LDAK version 5.1 (Speed et al., 2020). The GWAS model was as follows:

\[
y = \mathbf{1} \mu + \mathbf{X} \beta + Z \alpha + e,
\]

where \(y\) is the vector of corrected phenotypes; \(\mathbf{1}\) is a vector of ones; \(\mu\) is the overall mean; \(\beta\) is the unknown allele substitution effect of the SNP tested for the association; \(\mathbf{X}\) is the vector containing the genotype score for the tested SNP; \(\alpha\) is the vector of the random polygenic effects which follow a normal distribution \(\alpha \sim \mathcal{N}(0, \sigma^2_{\alpha})\), where \(\mathbf{G}\) is the genomic relationship matrix (GRM) which is built using 80K or iWGS data excluding the chromosome of the SNP tested, and \(\sigma^2_{\alpha}\) is the additive genetic variance; \(Z\) is the incidence matrix for \(\alpha\) and \(e\) is the vector of residual effects with \(e \sim \mathcal{N}(0, \sigma^2_e)\), where \(I\) is an identity matrix and \(\sigma^2_e\) is the residual variance. The SNP data sets used for GWAS were 80K SNP chip data and iWGS data, respectively. The Bonferroni correction method was used to determine the threshold values of significance in this study. The genome-wide significance level (0.05/N) and suggestive significance level (1/N) (Bland & Altman, 1995) were used in all GWAS, where \(N\) is the number of analyzed SNPs. Consequently, the genome-wide significance level and suggestive significance level were \(p = 9.90 \times 10^{-7}\) and \(p = 1.99 \times 10^{-5}\) for 80K SNP chip data, and the genome-wide significance level and suggestive significance level were \(p = 6.18 \times 10^{-9}\) and \(p = 1.24 \times 10^{-7}\) for WGS data. In addition, we defined QTL region as an interval between significant SNPs located on the upstream and downstream of physical location. Manhattan plots were obtained using the CMplot package (Yin et al., 2021) within the R software (http://www.r-project.org/). For the public annotation databases, clusterProfiler (Yu et al., 2012), an R package, were used to obtain Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in all significant regions.

The variance components of IFC were estimated and the IFC phenotype was adjusted for fixed effects (sex, batch, age and carcass weight) using the mixed linear model implemented in the DMU software (Madsen et al., 2014) based on 80K SNP chip data, and the model was described previously (Wang et al., 2021). The phenotypic variance explained (PVE) by each locus was calculated using the equation \(\text{PVE}_i = \frac{2q_i(1-q_i)b_i^2}{\sigma^2_Y}\), where \(q_i\) is the minor allele frequency of the \(i\)th SNP, \(b_i\) is the allele substitution effect of the \(i\)th SNP and \(\sigma^2_Y\) is the phenotypic variance of IFC. Heritability \(h^2\) of IFC was defined as the ratio of the additive genetic variance to phenotypic variance \(h^2 = \frac{\sigma^2_a}{\sigma^2_Y + \sigma^2_e}\).

2.8 | Pre-selection of SNPs for genomic prediction

To avoid automatically using information of validation animals to predict breeding value of the validation animal, the GWAS was performed in each generated reference population using iWGS (GWAS_iWGS) data. Briefly, all genotyped Suhuai pigs \((n = 482)\) with phenotype data were randomly divided into five separate groups, four groups were selected randomly as reference population, and the phenotype of the remaining one group was masked and used as the validation population. According to the above-mentioned method, five reference populations were formed, and GWAS was carried out based on each reference population using Model 1, respectively. Then, the significant SNPs affecting IFC in each reference population were selected into the models to predict breeding value when using this reference population.

2.9 | Statistical model for predicting breeding value

The statistical methods used for predicting breeding value in this study included linear mixed model using pedigree-based relationship matrix (PBLUP), linear mixed models using genome-wide marker-based relationships matrix (GBLUP) and Bayesian mixture model (BayesMix). For GBLUP and BayesMix models, the genotype data sets include 80K data, iWGS data, and pre-selected iWGS SNPs data based on GWAS results. We assessed a one-component model and a two-component model on their efficiency to use the pre-selected iWGS SNPs in these models. A one-component model considered 80K data and pre-selected iWGS SNPs as one genetic component, and a two-component model considered 80K data and pre-selected iWGS SNPs as two separate genetic components. It is worth noting that when one- or two-component model was used for GP, the duplicate SNPs in 80K data and pre-selected SNPs data will be deleted from 80K data.

In order to investigate the impact of LD-based SNP pruning on genomic prediction using iWGS data, twenty different R-squared cutoffs of LD \((1.00, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70, 0.65, 0.60, 0.55, 0.50, 0.45, 0.40, 0.35, 0.30, 0.25, 0.20, 0.15, 0.10, 0.05)\) were used to prune variants. LD was calculated based on the option --indep-pairwise PLINK (v1.90 beta) in a 50-kb sliding window with 10 variants. The number of SNPs after pruning at different cutoff levels are shown in Table S2.

2.10 | PBLUP model

The PBLUP model is

\[
y = \mathbf{1} \mu + \mathbf{X} \beta + Z \alpha + e,
\]

where \(y\) is the vector of phenotype value; \(\mathbf{1}\) is a vector of ones; \(\mu\) is the overall mean; \(\beta\) is the vector of fixed effects (sex and batch as
class variables, age and carcass as regression covariates; $u$ is the vector of additive genetic effects, and assumed that $u \sim N(0, \Lambda \sigma^2_u)$, in which $\Lambda$ is the matrix of additive genetic relationship constructed based on the pedigree; $\sigma^2_u$ is the additive genetic variance; $X$ and $Z$ are incidence matrices relating the fixed effects and the additive genetic values to phenotype values; $e$ represents random residual effects. The prediction of breeding value with PBLUP model was performed using the HIBLUP software (https://www.hiblup.com/).

### 2.11 | GBLUP model

The models for predicting breeding value using genotype information are as follows:

The one-component GBLUP model is

$$y = 1\mu + Xb + Zg + e,$$  

(3)

The two-component GBLUP model is

$$y = 1\mu + Xb + Z1_g1 + Z2_g2 + e,$$  

(4)

where $y$, $\mu$, $X$, $b$, and $e$ are the same as in Model 2; $g$, $g1$, $g2$ are vectors of additive genetic values; $Z, Z1, Z2$ are the incidence matrix to assign $y$ to $g$, $g1$, $g2$. We assumed that the additive genetic value is $g \sim N(0, G \sigma^2_g)$, $g1 \sim N(0, G1 \sigma^2_{g1})$, and $g2 \sim N(0, G2 \sigma^2_{g2})$, where $G$ is a relationship matrix built with either 80K, iWGS data, or 80K together with pre-selected SNPs; $G1$ is a relationship matrix built with 80K data; $G2$ is a relationship matrix built with pre-selected SNPs data; $\sigma^2_g$, $\sigma^2_{g1}$, and $\sigma^2_{g2}$ represent corresponding additive genetic variance, respectively. The estimated breeding value in Model 4 was defined as sum of $g1$ and $g2$. The prediction of breeding value with GBLUP model was performed using the HIBLUP software (https://www.hiblup.com/).

### 2.12 | Bayesian mixture model

The one-component BayesMix model is

$$y = 1\mu + Xb + Zg + e,$$  

(5)

The two-component BayesMix model is

$$y = 1\mu + Xb + Z_{S1}a1 + Z_{S2}a2 + e,$$  

(6)

where $y$, $\mu$, $X$, $b$, and $e$ are the same as in Model 2; $a$ is the vector of SNP effects for all SNPs in the model, $a1$, and $a2$ are the vector of effects for 80K SNPs and the vector of effects for pre-selected iWGS SNPs, respectively; $Z_{S1}, Z_{S1}$, and $Z_{S2}$ represent corresponding genotype matrices, respectively. For iWGS data, only pruned iWGS SNPs of scenario LD<0.55 were used for genomic prediction using the BayesMix model. We assumed the distribution of SNP effects ($a$, $a1$, or $a2$) follows a mixture of four normal distributions:

$$a_i \sim x_{i1}N(0, \sigma^2_{a1}) + x_{i2}N(0, \sigma^2_{a2}) + x_{i3}N(0, \sigma^2_{a3}) + x_{i4}N(0, \sigma^2_{a4}),$$

where $a_i$ is a particular vector of SNP effects ($a, a1$, or $a2$); $x_{ij}$ ($j = 1, 2, 3$ and 4) is the probability of an SNP belongs to the $j$th distribution within the $i$th component in the model, and $\sigma^2_{aj}$ is the variance for $j$th distribution within the $i$th component. In the present study, $x_{ij}$ was sampled from the Dirichlet distribution $x_{ij} = (x_{i1}, x_{i2}, x_{i3}, x_{i4}) \sim \text{dir}(125, 25, 5, 1)$ with prior $x_{i1} = 0.889, x_{i2} = 0.1, x_{i3} = 0.01, \text{ and } x_{i4} = 0.001$. $\sigma^2_{aj}$ is assumed to have a scaled inverse $\chi^2$ distribution and have a fixed ratio of 1000 $\sigma^2_{a1} = 10 \sigma^2_{a2} = \sigma^2_{a3} = \sigma^2_{a4}$; thus, only one of them is required to be estimated within each genetic component. In the present study, the BayesMix model was run using single chain Gibbs sampler with a total length of 50,000 samples, where the first 10,000 samples were discarded as burn-in. The analyses with BayesMix models were carried out using the Bayz software (http://www.bayz.biz).

In total, nine approaches were used to estimate breeding values for IFC. (1) PBLUP with Model 2 which used the pedigree-based relationship matrix ($A$). (2) GBLUP-80K with Model 3 which used the 80K SNP chip data to calculate the GRM ($G$). (3) GBLUP-iWGS with Model 3 which used the iWGS data pruned at different LD levels to calculate the GRM ($G$) separately. (4) GBLUP-80K-GWAS_{iWGS} one with Model 3 which pooled the pre-selected iWGS SNPs and the 80K SNPs together as one component to construct the GRM ($G$). (5) GBLUP-80K-GWAS_{iWGS} two with Model 4 which used the pre-selected iWGS SNPs to construct $G2$ and the 80K SNPs to construct $G1$; (6) Bayesian-80K with Model 5 which used the 80K SNP chip data only; (7) Bayesian-iWGS with Model 5 which used the pruned iWGS data (LD<0.55) only. (8) Bayesian-80K-GWAS_{iWGS} one with Model 6 which use the 80K SNPs and pre-selected iWGS SNPs as one genetic component; (9) Bayesian-80K-GWAS_{iWGS} two with Model 6 which used the 80K SNPs and the pre-selected iWGS SNPs as two separate genetic components.

### 2.13 | Validation of prediction

Accuracy of genomic prediction were assessed using a 5-fold cross validation procedure. The predictability for estimating breeding values were assessed by reliability. The reliability of the prediction was measured as squared correlation between estimated breeding values and corrected phenotypes in the validation groups divided by the $h^2$ of IFC. The standard error (SE) was calculated as the standard deviation of the 5 calculated reliability from the 5 fold cross-validation divided by the square root of 5.

### 3 | RESULTS

#### 3.1 | Heritability of IFC and accuracy of genotype imputation

As shown in Table 1, the mean and standard error of IFC in 482 Suhuai pigs was 1.91 ± 0.03%, and the coefficient of variation was 31.73%, indicating a large variation of IFC phenotypes in this
population. The heritability \((h^2)\) of IFC was 0.26, estimated using a model 1. Heritability estimated from other models was also around this size.

The number of SNPs after imputation and imputation accuracy for each autosome are shown in Table 2. For iWGS genotype data, after removing the SNPs with \(D^2 < 0.9\), the corresponding values of average imputation accuracy \((D^2)\) was 0.96, and 8 103,716 SNPs were remained for subsequent GWAS. The average correlation between imputed and observed genotypes was 0.93 (0.86–0.96) in the validation, which was lower than \(D^2\).

### 3.2 | Genome-wide association study for IFC in 482 Suhuai pigs

This study conducted GWAS in two scenarios. In the first scenario, GWAS was conducted using the 80K data. We identified 7 QTL regions including 80 SNPs which surpassed the suggestive significance level \((p < 1.98E-05)\) and 20 SNPs which reached the genome-wide significant level \((p < 9.90E-07)\) (Figure 1a, Table 3). These significant SNPs (suggestive or genome-wide significance) were mainly distributed on Sus scrofa chromosome (SSC) 1 (124.19–125.56 Mb), SSC3 (94.40–104.56 Mb), SSC6 (79.75–89.16 Mb), SSC14 (25.80–40.17 Mb) and SSC16 (75.66–76.27 Mb). Compared to the first scenario with iWGS data, 4 QTL regions were identified. A total of 1217 suggestive-significant SNPs were identified \((p < 1.23659E-07)\), while 517 SNPs achieved a genome-wide significant level \((p < 6.18296E-09)\) in these QTL regions (Figure 1b, Table 3). These significant SNPs were also mainly distributed on SSC3 (94.32–105.54 Mb), SSC6 (79.75–89.16 Mb), SSC14 (29.21–35.16 Mb) and SSC16 (75.66–76.27 Mb).
scenario, GWAS based on iWGS dataset could reduce the range of significant regions and increase the number of significant SNPs in these significant regions, especially on SSC14 and SSC16. The number of QTL regions was fewer using the iWGS than using 80 K data because significant level was at much smaller \( p \) value (after multiple testing correction) when using iWGS data.

We performed functional enrichment analysis in all significant regions, a total of 160 genes were annotated. Functional analysis of these genes revealed that there were 99 GO terms and 20 KEGG pathways significantly enriched (false discovery rates <0.05) based on Sus scrofa 11.1 genome (Tables S4 and S5). Moreover, through gene function analysis, protein kinase C epsilon (PRKCE) and myosin light
chain 2 (MYL2) gene were suggested to be two important candidate genes affecting IFC. It is worth noting that, a total of 355 significant SNPs were located in PRKCE gene on SSC3, and the most significant locus rs338155853 was located near 0.33 Mb downstream of PRKCE gene. Moreover, the most significant SNP rs331439214 on SSC14 was located near 0.36 Mb upstream of MYL2 gene.

3.3 Pre-selection of SNPs based on GWAS in five reference populations

We performed GWAS using iWGS data in the generated five reference populations, respectively. The summary of significant SNPs is shown in Table S3. Similar to the above GWAS results based on the whole data of 482 animals, these significant SNPs were also mainly distributed on SSC3 (94.32–105.54 Mb) and SSC14 (31.79–32.59 Mb). The most significant SNPs on SSC3 were near 95.00 Mb, while rs331439214 on SSC14 31,793,530 bp was always the most significant SNP associating with IFC in the five reference populations. The number of significant SNPs in GWAS results based on different reference populations were different. Finally, based on five GWAS results, 1292, 1128, 596, 153, and 595 suggestive significant SNPs were pre-selected for subsequent genomic prediction in the 5-fold cross-validation procedure.

3.4 Genomic prediction

The reliabilities of prediction using PBLUP, GBLUP and BayesMix in validation populations are presented in Figures 2 and 3. In this study, the PBLUP model had the lowest prediction reliability (0.096 ± 0.032). The reliability (0.229 ± 0.035) of genome prediction was significantly improved by replacing pedigree information with 80K SNP chip data. Compared with 80K SNP chip data in GBLUP model, the reliability of prediction was reduced by 0.011 when whole iWGS SNPs were used. When using pruned iWGS SNPs, pruning using the R-squared cutoff of LD at 0.40 to 0.60 led to higher accuracies with the highest accuracy for cutoff of LD at 0.55, which led to a slight improvement (0.006) compared with GBLUP-80K. For GBLUP models, compared with using the 80K SNPs alone, adding additional pre-selected iWGS SNPs to the 80K SNPs and treating all the SNPs as one genetic component (GBLUP-80K-GWAS-iWGS-one) resulted in a slight increase of reliability which was 0.033 higher than that of the one-component GBLUP model. Moreover, the GBLUP model treating the two sets of SNPs as two genetic components (GBLUP-80K-GWAS-iWGS-two) resulted in a significant improvement of reliability by 0.05. However, the GBLUP model treating the two sets of SNPs as two genetic components (GBLUP-80K-GWAS-iWGS-two) resulted in a significant improvement of reliability by 0.05. Moreover, the GBLUP model treating the two sets of SNPs as two genetic components (GBLUP-80K-GWAS-iWGS-two) resulted in a significant improvement of reliability by 0.05. Moreover, the GBLUP model treating the two sets of SNPs as two genetic components (GBLUP-80K-GWAS-iWGS-two) resulted in a significant improvement of reliability by 0.05.

FIGURE 2 The reliability of prediction using GBLUP-iWGS models with different R-squared cutoffs of linkage disequilibrium to prune iWGS data.

FIGURE 3 The reliability of predictions using PBLUP, GBLUP and Bayesian mixture models.
of GBLUP model, compared with using 80 K SNPs, Bayesian-iWGS (LD ≤ 0.55) model led to slight improvement (0.008) of reliability. In this study, the prediction reliabilities of GBLUP models were better than those of BayesMix models when using the same SNPs dataset.

4 | DISCUSSION

4.1 | Genome-wide association study

In the present study, GWAS based on 80 K SNP chip data and iWGS data were carried out to explore key variants affecting IFC in 482 Suhuai pigs. The consistency of these two GWAS results was relatively high, and the identified significant regions affecting IFC mainly distributed on SSC3, SSC6, SSC14 and SSC16. In this study, the use of iWGS data instead of 80K data improved the detection power of SNPs of interest, association analyses using iWGS data rather than 80K data could highlight the peaks and increase the phenotypic variation explained by the peak SNP in each QTL. Besides, GWAS using iWGS data rather than 80K data could reduce the mapping noise. Several significant SNPs on SSC1, SSC17 and SSC18 were detected based on 80K dataset, however, the p values of these SNPs were larger than those in GWAS based on iWGS data. One possible reason was that the genomic relationship matrices in the models were constructed with different SNP data sets. The higher p values could be also resulted from larger sample size, in our previous studies, a QTL on SSC5 affecting IFC was identified by using GWAS based on 80K dataset in high (n = 48) and low (n = 48) IFC groups (Wang et al., 2021), and the rs1110770079 SNP located on FABP3 was significantly associated with IFC by association analysis in 330 Suhuai pigs (Wang et al., 2019). However, this significant signal on SSC5 and rs1110770079 SNP were not detected in current study, we speculated that the significant signals of early research were detected based on a limited sample size, and the phenotypic variations explained were small. When we used a larger sample size and iWGS data, four novel QTLs that could explain larger phenotypic variations were detected, and the signal peaks in our previous studies were disappeared. Therefore, it was essential to increase the sample size and the density of SNPs (e.g. iWGS data) in GWAS. Meanwhile, these significant signals with large effects derived from iWGS data might be beneficial to improve the accuracy of genomic prediction for IFC.

In the results of GWAS based on iWGS, a total of 932 SNPs significantly associated with IFC were identified based on 80 K SNP chip data and iWGS data, four novel QTLs that could explain larger phenotypic variations were detected, and the signal peaks in our previous studies were disappeared. Therefore, it was essential to increase the sample size and the density of SNPs (e.g. iWGS data) in GWAS. Meanwhile, these significant signals with large effects derived from iWGS data might be beneficial to improve the accuracy of genomic prediction for IFC.

In the second significant region based on iWGS data, SSC14 (29.21–35.16 Mb), a total of 208 significant SNPs were identified. The most significant SNP rs331349214 on this region was located near 0.36 Mb upstream of myosin might chain 2 (MYL2) gene. MYL2 is a major sarcomeric protein in mammalian striated muscle, and plays a pivotal role in the development and function of the heart (Sheikhi et al., 2015). MYL2, MYL3, MYH7 and TPM3 were considered to be the four best hits coded for typical slow muscle proteins (Amann et al., 2014), previous research found that slow fibers exhibit a higher intramyocellular lipid content than fast glycolytic fibers (Essén-Gustavsson et al., 1994). Pan et al. (2020) performed transcriptome sequencing in longissimus dorsi muscle tissue of Luchuan pig and Duroc pig and found forty differentially expressed genes (DEGs) related to muscle development, including MYL2. Given that MYL2 might affect fatty acid metabolism, we speculated that it might also be considered as a candidate gene for IFC in pigs.

4.2 | Genomic prediction

This study applied PBLUP, GBLUP and Bayesian four-distribution mixture model with different sets of marker genotypes, for genetic evaluation of IFC in Suhuai pigs. The predictive abilities of these three methods were compared in terms of prediction reliability. Our results indicated that the methods based on marker information were more accurate than the method based on pedigree alone. Similarly, Uemoto et al. (2017) reported that the benefits of genomic prediction in terms of increased accuracy of prediction by about 21% over a pedigree-based model for IFC in a closed line of Duroc pigs. This is likely due to the fact that the GBLUP and Bayesian methods have a better capture of the Mendelian sampling terms in comparison with the pedigree-based prediction method (Chen et al., 2015).

Using WGS data in genomic prediction is expected to bring higher predictive reliability, because WGS data included causal SNPs affecting target traits (Meuwissen & Goddard, 2010; Song et al., 2019). Meuwissen and Goddard (2010) demonstrated in simulations that genomic predictions based on WGS data were 5%–10% more accurate than predictions based even on high dense markers, because the causal mutations were used in prediction. However, in the current
study, the prediction reliability of model with GRM constructed using whole iWGS data were slightly lower than using 80K data. Similar to our results, previous studies on reproduction and production traits in pig (Song et al., 2019; Van Binsbergen et al., 2015), reproduction and conformation trait in cattle (Frischknecht et al., 2018), carcass and meat quality in sheep (Moghaddar et al., 2019) suggested no improvement or only marginal increase in genomic prediction accuracy when using of iWGS, compared with chip data. Several factors could explain the unsatisfied results of genomic prediction using iWGS data. First, in iWGS data, a larger number of SNPs do not link to the QTLs affecting the traits of interest, thus, more noise might be introduced to prediction. Second, in the current study, thirty key individuals based on pedigree relationship were selected as the reference panel for resequencing, and the imputation from clean 80K data to WGS data was performed. However, potential imputation errors were difficult to detect and eliminate, which also affected the accuracy of genomic prediction. Calus et al. (2014) have reported that low imputation accuracy had a negative impact on genomic prediction accuracy, thereby reducing the response to selection.

Several studies on genomic prediction for multiple traits often pruned WGS data with fixed LD value (Brøndum et al., 2015; Song et al., 2019). In this study, we investigated the impact of pruning at different levels on genomic prediction. The results showed that moderate LD (LD<0.55) resulted in higher prediction reliability when using all iWGS data. However, Ye et al. (2019) found that the prediction accuracies of most traits using iWGS data were the highest when lower LD (0.10) was used to prune markers. We speculated that an important reason could be LD degree in different population. The number of remaining SNPs was large in low LD population, and small in high LD population under the same cutoff threshold. Another possible explanation is due to the difference in genetic structure of traits.

Many studies have demonstrated that adding significant SNPs derived from high-density marker panels or WGS data into medium density panel data could improve the accuracy of GP (Brøndum et al., 2015; Lopes et al., 2017). Brøndum et al. (2015) found that the prediction reliability of production traits in cattle could be improved by 3 to 5 percentage points by including markers significant from GWAS based on WGS data alongside the 54K SNP data sets. Similarly, in current study the significant SNPs were preselected from GWAS_iWGS and added to the 80K SNPs in GBLUP models and Bayesian models, the predicted reliabilities were significantly increased. We speculated that the pre-selected markers obtained from iWGS data in our study might be closer to the causal mutations for IFC.

To select a model to use the information of selected iWGS SNPs efficiently, we compared GBLUP models with BayesMix models by polling the selected SNP with the 80K data or taking them as an independent component. Bayesian method is expected to perform better than GBLUP for traits controlled by loci with large effects (Lee et al., 2017), otherwise, GBLUP performs as well as Bayesian models. In this study, the prediction reliabilities of GBLUP models for IFC were better than that of Bayesian four-distribution model when using the same SNPs dataset. Similar to our results, Chen et al. (2015) reported that the GBLUP models outperformed Bayesian for carcass marbling score trait in 543 Angus and 400 Charolais beef cattle. A possible reason may be due to the IFC trait was controlled by multiple SNPs with small effects. Another reason could be that the reference population in the current study is small and the Bayesian model is not sufficient to distinguish the SNPs with large or small effects appropriately. Karaman et al. (2016) reported that Bayesian variable selection models have no advantage over GBLUP when the size of reference population was small.

In a conventional GBLUP model, an important assumption is that all SNPs (whether or not associated with specific traits or not) have the same variance and follow the same normal distribution. However, the assumption may be not appropriate since some SNPs have large effect on the target traits than the others. In contrast, previous studies reported that a method called weighted GBLUP (WGGLUP) which put unequal weights on different SNPs led to higher accuracies than the unweighted counterpart (Su et al., 2014; Zhang et al., 2010). The WGGLUP model has similar computation costs as the regular GBLUP model but is able to reach similar reliabilities as the Bayesian mixture model. Similar to the WGGLUP concept (Zhang et al., 2010), a two-component GBLUP model including an additional G matrix for the highly informative SNPs can differentiate the highly informative SNPs from the other SNPs. In this study, the two-component GBLUP model (GBLUP-80K-GWAS_iWGS-two) achieved higher reliabilities than the one component GBLUP model. Liu et al. (2020) also reported that the use of selected WGS SNPs together with the 54K SNP chip in a two-component GBLUP model increased the prediction reliability for milk production traits in dairy cattle. Similar to the GBLUP model, the advantage of two component approach was observed when using the Bayesian mixture model in the current study, though the Bayesian model is able to differentiate SNPs with large or small effects. The reason could be that dividing the SNPs into two sets provides good priors to the model, which could be important, especially for small reference data.

In conclusion, we implemented GWAS using iWGS data and identified two candidate genes (PRKCE and MYL2) and a number of significant SNPs for IFC, providing useful knowledge for further identification of the causal genes affecting IFC deposition. The validation of prediction showed that the models with genomic marker information could have higher prediction abilities than the methods based on pedigree alone. Compared with genomic prediction only using the chip genotype data, using all iWGS SNPs led to lower accuracy, but adding important SNPs selected from GWAS based on iWGS data to the 80K SNPs increased accuracy. When including the important iWGS SNPs, the two-components models performed better than the one-component models. In addition, the 80K chip data was imputed into WGS data based on LD, which increased the SNP density, and the causal variants affecting IFC could be more accurately identified through GWAS based on the iWGS dataset. This provides opportunity to improve pork quality by gene editing technology in the future.
ACKNOWLEDGMENTS
We thank the researchers at our laboratories in Nanjing and Huaian for their dedication and hard work. The first author acknowledges a scholarship provided by the China Scholarship Council (CSC). This work was financially supported by the National Natural Science Foundation of China (grant no. 31872318, 31601923 and 32002149), the Key Project of Jiangsu Agricultural New Variety Innovation (grant no. PZCZ201732), the Ministry of Agriculture and Rural Affairs Joint Projects for the National High Quality and Lean Pig Breeding (grant no. 19190540), the Project of Jiangsu Agricultural (pig) Industry Technology System (grant no. JATS[2020]179 and JATS [2020]399), the Fundamental Research Funds for the Central Universities (grant no. KJQN202129), and the Cultivation Project of NSFC-Henan United Fund (grant no. U1904115).

CONFLICT OF INTEREST
The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT
The WGS data for this study can be found in the NCBI Sequence Read Archive (SRA) under Bioproject: PRJNA791712. The 80K SNP chip data and phenotype data for 482 Suhuai pigs used in this study were deposited at the figshare repository (https://doi.org/10.6084/m9.figshare.19129349).

ORCID
Ruihua Huang https://orcid.org/0000-0002-4722-2271

REFERENCES
Amann, R., Wyder, S., Slavotinek, A. M., & Trueb, B. (2014). The FgfrL1 receptor is required for development of slow muscle fibers. Developmental Biology, 394(2), 228–241. https://doi.org/10.1016/j.ydbio.2014.08.016
Banerjee, P., Carmelo, V. A. O., & Kadarmideen, H. N. (2020). Genome-wide epistatic interaction networks affecting feed efficiency in Duroc and landrace pigs. Frontiers in Genetics, 11, 121–135. https://doi.org/10.3389/fgene.2020.00121
Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. British Medical Journal, 310(6973), 170. https://doi.org/10.1136/bmj.310.6973.170
Brandum, R. F., Su, G., Janss, L., Sahana, G., Guldbrachtsen, B., Boichard, D., & Lund, M. S. (2015). Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. Journal of Dairy Science, 98(6), 4107–4116. https://doi.org/10.3168/jds.2014-9005
Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. The American Journal of Human Genetics, 103(3), 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015
Calus, M. P. L., Bouwman, A. C., Hickey, J. M., Veerkamp, R. F., & Mulder, H. A. (2014). Evaluation of measures of correctness of genotype imputation in the context of genomic prediction: A review of several methods. Animal, 8(11), 1743–1753. https://doi.org/10.1017/S1751731114001803
Chen, L., Vinsky, M., & Li, C. (2015). Accuracy of predicting genomic breeding values for carcass merit traits in Angus and Charolais beef cattle. Animal Genetics, 46(1), 55–59. https://doi.org/10.1111/age.12238
Cho, I. C., Yoo, C. K., Lee, J. B., Jung, E. J., Han, S. H., Lee, S. S., Ko, M. S., Lim, H. T., & Park, H.-B. (2015). Genome-wide QTL analysis of meat-quality-related traits in a large F2 intercross between landrace and Korean native pigs. Genetics Selection Evolution, 47(1), 7–14. https://doi.org/10.1186/s12711-014-0080-6
Essén-Gustavsson, B., Karlsson, A., Lundström, K., & Enfält, A.-C. (1994). Intramuscular fat and muscle fibre lipid contents in halothane-gene-free pigs fed high or low protein diets and its relation to meat quality. Meat Science, 38(2), 269–277. https://doi.org/10.1016/0309-1749(94)00116-3
Frischknecht, M., Meeuwissen, T. H. E., Bapst, B., Seefried, F. R., Flury, C., Garrick, D., Signer-Hasler, H., Stricker, C., Bieber, A., Fries, R., Russ, I., Sölkner, J., Bagnoato, A., & Gredler-Grandi, B. (2018). Short communication: Genomic prediction using imputed whole-genome sequence variants in Brown swiss cattle. Journal of Dairy Science, 101(2), 1292–1296. https://doi.org/10.3168/jds.2017-12890
Groenen, M. A. M., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.-J., Li, S., Larkin, D. M., Kim, H., Frantz, L. A. F., Caccamo, M., Ahn, H., Aken, B. L., Anselmo, A., Anthon, C., ... Schook, L. B. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. Nature, 491(7424), 393–398. https://doi.org/10.1038/nature11622
Gyamfi, J., Yeo, J. H., Kwon, D., Min, B. S., Cha, Y. J., Koo, J. S., Jeong, J., Lee, J., & Choi, J. (2021). Interaction between CD36 and FABP4 modulates adipocyte-induced fatty acid import and metabolism in breast cancer. NPJ Breast Cancer, 7(1), 1–18. https://doi.org/10.1038/s41523-021-00324-7
Jia, Y., & Jannink, J.-L. (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics, 192(4), 1513–1522. https://doi.org/10.1534/genetics.111.144246
Josephs, E. B., Stinchcombe, J. R., & Wright, S. I. (2017). What can genome-wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits? New Phytologist, 214(1), 21–33. https://doi.org/10.1111/nph.14410
Karaman, E., Cheng, H., Firat, M. Z., Garrick, D. J., & Fernando, R. L. (2016). An upper bound for accuracy of prediction using GBLUP. PLoS One, 11(8), e0161054.
Larimer, S. G., Sargolzaei, M., Brito, L. F., Ventura, R. V., & Schenkel, F. S. (2017). Novel methods for genotype imputation to whole-genome sequence and a simple linear model to predict imputation accuracy. BMC Genetics, 18(1), 1–12. https://doi.org/10.1186/s12861-017-0588-1
Lee, J., Cheng, H., Garrick, D., Golden, B., Dekkers, J., Park, K., Lee, D., & Fernando, R. (2017). Comparison of alternative approaches to single-trait genomic prediction using genotyped and non-genotyped Hanwoo beef cattle. Genetics Selection Evolution, 49(1), 1–9. https://doi.org/10.1186/s12711-016-0279-9
Liu, A., Lund, M. S., Boichard, D., Karaman, E., Fritz, S., Aamand, G. P., Nielsen, U. S., Wang, Y., & Su, G. (2020). Improvement of genomic prediction by integrating additional single nucleotide polymorphisms selected from imputed whole genome sequencing data. Heredity, 124(1), 37–49. https://doi.org/10.1038/s41433-019-0246-7
Lopes, M. S., Bovenhuis, H., van Son, M., Nordbo, Ø., Grindflek, E. H., Knol, E. F., & Bastiaansen, J. W. M. (2017). Using markers with large effect in genetic and genomic predictions. Journal of Animal Science, 95(1), 59–71. https://doi.org/10.2527/jas2016-0754
Madsen, P., Jensen, J., Labouriau, R., Christensen, O.F., & Sahana, G. (2014). DMU-a package for analyzing multivariate mixed models with random effects in genetic and evolution. BMC Genetics, 95(2), 228–241. https://doi.org/10.1111/j.2045-7199.2014.00001.x
Meuwissen, T., & Goddard, M. (2010). Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics, 185(2), 623–631. https://doi.org/10.1093/genetics/185.2.NP
Moghadar, N., Khaneefid, M., van der Werf, J. H. J., Bolormaa, S., Duijvestein, N., Clark, S. A., Swan, A. A., Daetwyler, H. D., & MacLeod, I. M. (2019). Genomic prediction based on selected variants from imputed whole-genome sequence data in Australian sheep populations. Genetics Selection Evolution, 51(1), 1–14. https://doi.org/10.1186/s12711-019-0514-2

Ngapo, T. M. (2017). Consumer preferences for pork chops in five Canadian provinces. Meat Science, 129, 102–110. https://doi.org/10.1016/j.meatsci.2017.02.022

Pan, P., Qin, Z., Xie, W., Jiao, D., Chen, B., Guan, Z., & Xie, B. (2020). Identification of differentially expressed genes in the longissimus dorsi muscle of Luchuan and Duroc pigs by transcriptome sequencing. bioRxiv. https://doi.org/10.1101/2020.07.18.2159

Puig-Oliveras, A., Revilla, M., Castelló, A., Fernández, A. I., Folch, J. M., & Ballester, M. (2016). Expression-based GWAS identifies variants, gene interactions and key regulators affecting intramuscular fatty acid content and composition in porcine meat. Scientific Reports, 6(1), 1–12. https://doi.org/10.1038/srep31803

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81(3), 559–575. https://doi.org/10.1086/519795

Raddatz, K., Turner, N., Frangiolakis, G., Liao, B. M., Pedersen, D. J., Cantley, J., Wilks, D., Preston, E., Hegarty, B. D., Leitges, M., Raferty, M. J., Biden, T. J., & Schmitz-Peiffer, C. (2011). Time-dependent effects of Prkce deletion on glucose homeostasis and hepatic lipid metabolism on dietary lipid oversupply in mice. Diabetologia, 54(6), 1447–1456. https://doi.org/10.1007/s00125-011-2073-0

Salek Ardestani, S., Jafarikia, M., Sargolzaei, M., Sullivan, B., & Miari, Y. (2013). From FastQ data to high-confidence variant calls: The Genome Analysis Toolkit best practices pipeline. Current Protocols in Bioinformatics, 43(1), 10–11. https://doi.org/10.1002/0471250953.bi1110s43

Wang, B. B., Hou, L. M., Zhou, W. D., Liu, H., Tao, W., Wu, W., Niu, P. P., Zhang, Z. P., Zhou, J., & Li, Q. (2021). Genome-wide association study reveals a quantitative trait locus and two candidate genes on sus scrofa chromosome 5 affecting intramuscular fat content in Suhuai pigs. Animal, 15(9), 100341. https://doi.org/10.1016/j.animal.2021.100341

Wang, B. B., Li, P. H., Zhou, W. D., Gao, C., Liu, H., Li, H., Niu, P. P., Zhang, Z. P., Li, Q., Zhou, J., & Huang, R. H. (2019). Association of Twelve Candidate Gene Polymorphisms with the intramuscular fat content and average backfat thickness of Chinese Suhuai pigs. Animals, 9(11), 858–874. https://doi.org/10.3390/ani9110858

Warburton, C. L., Engle, B. N., Ross, E. M., Costilla, R., Moore, S. S., Corbet, N. J., Allen, J. M., Laing, A. R., Fordyce, G., Lyons, R. E., McGowan, M. R., Burns, B. M., & Hayes, B. J. (2020). Use of whole-genome sequence data and novel genomic selection strategies to improve selection for age at puberty in tropically-adapted beef heifers. Genetics Selection Evolution, 52(1), 1–13. https://doi.org/10.1186/s12711-020-00547-5

Warr, A., Affara, N., Aken, B., Beiki, H., Bickhart, D. M., Bills, K., Chow, W., Eory, L., Finlayson, H. A., & Flicek, P. (2020). An improved pig reference genome sequence to enable pig genetics and genomics research. GigaScience, 9(6), giaa051. https://doi.org/10.1093/gigascience/giaa051

Wood, J. D., Nute, G. R., Richardson, R. I., Whittington, F. M., Southwood, O., Plastow, G., Mansbridge, D., Costa, N., & Chang, K. C. (2004). Effects of breed, diet and muscle on fat deposition and eating quality in pigs. Meat Science, 67(4), 651–667. https://doi.org/10.1016/j.meatsci.2004.01.007

Yang, Y., Xie, J., Xie, B., Chen, M., Niu, Y., Li, Q., Liu, Y., Yi, G., Fan, X., & Tang, Y. (2019). Chromosome-scale de novo assembly and phasing of a Chinese indigenous pig genome. BioRxiv, 770958. https://doi.org/10.1101/770958

Ye, S. P., Gao, N., Zheng, R. R., Chen, Z. T., Teng, J., Yuan, X., Zhang, H., Chen, Z., Zhang, X., Li, J., & Zhang, Z. (2019). Strategies for obtaining and pruning imputed whole-genome sequence data for genomic prediction. Frontiers in Genetics, 10(6), 673. https://doi.org/10.3389/fgene.2019.00673

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., Yuan, X., Zhu, M., Zhao, S., & Li, X. (2021). Rmvp: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. Genomics, Proteomics & Bioinformatics, 19(4), 619–628. https://doi.org/10.1007/j.10039/gi.2020.10.007

Yu, G. C., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology, 16(5), 284–287. https://doi.org/10.1089/omi.2011.0118

Zhang, Z., Liu, J., Ding, X., Biju, P., de Koning, D.-J., & Zhang, Q. (2010). Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. PLoS One, 5(9), e12648. https://doi.org/10.1371/journal.pone.0012648
SUPPORTING INFORMATION
Additional supporting information can be found online in the Supporting Information section at the end of this article.