animals

MDPI

Article

Utilizing Variants Identified with Multiple Genome-Wide Association Study Methods Optimizes Genomic Selection for Growth Traits in Pigs

Ruifeng Zhang [1], Yi Zhang [2], Tongni Liu [3], Bo Jiang [1], Zhenyang Li [1], Youping Qu [4], Yaosheng Chen [1] and Zhengcao Li [1,*]

[1] State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510006, China
[2] Institute of Neuroscience, Panzhihua University, Panzhihua 617000, China
[3] Genetic Data Center, Faculty of Forestry, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[4] Guangdong IPIG Technology Co., Ltd., Guangzhou 510006, China
* Correspondence: lizhc7@mail.sysu.edu.cn

Simple Summary: The accurate prediction of growth traits in genomic selection (GS) is essential for pig breeding. Here, we performed GS using variants identified with three genome-wide association study methods on four growth-related traits in Yorkshire and Landrace pigs. A total of 1485 loci related to these traits and 24 candidate genes were mapped. Compared with using 60K SNP-chip data, GS with the pre-selected variants significantly improved prediction accuracies by 4 to 46% in genomic best linear unbiased prediction (GBLUP) models, and 5 to 27% in a two-kernel based GBLUP model for the four traits.

Abstract: Improving the prediction accuracies of economically important traits in genomic selection (GS) is a main objective for researchers and breeders in the livestock industry. This study aims at utilizing potentially functional SNPs and QTLs identified with various genome-wide association study (GWAS) models in GS of pig growth traits. We used three well-established GWAS methods, including the mixed linear model, Bayesian model and meta-analysis, as well as 60K SNP-chip and whole genome sequence (WGS) data from 1734 Yorkshire and 1123 Landrace pigs to detect SNPs related to four growth traits: average daily gain, backfat thickness, body weight and birth weight. A total of 1485 significant loci and 24 candidate genes which are involved in skeletal muscle development, fatty deposition, lipid metabolism and insulin resistance were identified. Compared with using all SNP-chip data, GS with the pre-selected functional SNPs in the standard genomic best linear unbiased prediction (GBLUP), and a two-kernel based GBLUP model yielded average gains in accuracy by 4 to 46% (from 0.19 ± 0.07 to 0.56 ± 0.07) and 5 to 27% (from 0.16 ± 0.06 to 0.57 ± 0.05) for the four traits, respectively, suggesting that the prioritization of preselected functional markers in GS models had the potential to improve prediction accuracies for certain traits in livestock breeding.

Keywords: GWAS; genomic selection; growth traits; pigs

check for updates

Citation: Zhang, R.; Zhang, Y.; Liu, T.; Jiang, B.; Li, Z.; Qu, Y.; Chen, Y.; Li, Z. Utilizing Variants Identified with Multiple Genome-Wide Association Study Methods Optimizes Genomic Selection for Growth Traits in Pigs. *Animals* **2023**, *13*, 722. https://doi.org/10.3390/ani13040722

Academic Editors: Gregório Miguel Ferreira De Camargo and Raphael Bermal Costa

Received: 25 December 2022
Revised: 9 February 2023
Accepted: 15 February 2023
Published: 17 February 2023

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

1. Introduction

Genomic selection (GS) has been a routine method for genetic evaluation in animal breeding schemes [1]. It can increase the rate of genetic improvement, by reducing generation interval and providing higher accuracy of estimated breeding values. Due to the reduced cost of SNP chips and genome sequencing [2], the continuous escalation of the density of marker panels is generating both theoretical and implementation problems in breeding practice. Although improvements on prediction accuracies for some traits were occasionally observed by increasing the density of marker panels [3–5], in most scenarios the increase of marker density did not transform into an improved prediction accuracy, especially with a small population size. Theoretically, increasing the number of SNPs in

a linear model may magnify collinearity and reduce the effects of causal variants, which will hamper the ability to prioritize relevant variants [6]. Recent studies reported that the prediction accuracy was increased by using only a limited number of SNPs associated with specific traits [7–9], in comparison to using low or moderate SNP-chip. In addition, SNPs around genes or potential causal variants can be preselected and integrated into genomic prediction [10]. A study highlighted that using only SNPs within quantitative trait loci (QTLs) showed a higher accuracy of the prediction of reproductive traits in two pig breeds [11]. A similar result was also observed for the prediction of residual feed intake in Duroc pigs, in which the predictive ability increased when only utilizing SNPs located in genes [12]. Moreover, combining pre-selected markers from QTLs with 50k SNP panels could result in a better prediction performance of production traits in Nordic Holstein cattle [13].

Genome-wide association analysis (GWAS) is a powerful tool to identify potential causal mutations associated with traits of interest [14,15]. It can be implemented with various statistical models, such as a mixed linear model and Bayesian model [16,17]. The mixed linear model takes one SNP into the regression model every time to analyze the relationship between marker and target traits. Significant signals might fall into a broad-range genomic region owing to the strong linkage disequilibrium between SNPs. However, the interaction of markers from an underlying QTL would be ignored [18,19]. As an alternative, Bayesian methods simultaneously fit multiple SNPs effects, mapping the genomic region on the complex trait and taking LD effect into account [20]. Performing GWAS independently on multiple populations may reduce the false positive rate and increase statistical power to find true causal mutations. However, combining summary statistics from these independent GWAS studies using meta-analysis can increase the power to find effects that are homogeneous across populations, and can elucidate between-population heterogeneity [21]. This method has gained popularity for the discovery of new candidate loci and for improving mapping precision using data from multi-breed populations in the field of livestock breeding [22,23].

Pigs are a primary source of animal protein for humans [24]. Improving pork production has always been a never-ending goal for livestock breeding. Growth-related traits such as average daily gain (ADG), backfat thickness (BFT), body weight (BW) and birth weight (BTHWT) are commonly treated as selection indices in breeding programs [25,26]. Although the genetic architecture of growth traits is largely polygenic in nature, which may be affected by a substantial number of variants with small effects, significant variants with large effects have also been discovered [14,27,28]. In this study, we use three GWAS methods including mixed linear model, Bayesian analysis and meta-analysis, with 60K SNP-chip data and whole genome sequence data from 1734 Yorkshire and 1123 Landrace pigs, to detect as many potential causal mutations as possible and assess whether utilizing these pre-selected SNPs in GS can improve the prediction accuracies of growth traits in pigs.

2. Materials and Methods

2.1. Phenotype Data

In this study, a total of 1734 Yorkshire and 1123 Landrace pigs were sampled from the nucleus farm of Guangdong IPIG Technology Co., Ltd. (Guangzhou, Guangdong, China). The 153 male and 2704 female pigs of two populations were born between 2015 and 2021. All fattening pigs were reared under uniform feeding conditions and consistent management. Phenotypic observations included 4 growth-related traits: average daily gain (ADG kg/d), backfat thickness (BFT mm), birth weight (BTHWT kg) and body weight (BW kg). When the body weight of pigs reached approximately 115 kg, backfat thickness and body weight were recorded in the test station. Backfat thickness was measured by ultrasound between the 10th and 11th rib in the live pigs. To standardize the phenotypic records, BFT and ADG

were both adjusted to 100 kg and calculated by corrected equations [29]. The equation for corrected 100 kg BFT was shown below:

$$BFT\ (\text{mm}) = Measured\ BFT \times \frac{A}{A + B \times (Measured\ body\ weight - 100\ \text{kg})}$$

where the values of A and B for sire and dam were shown as follows:

$$Sire: A = 13.47;\ B = 0.1115$$
$$Dam: A = 15.65;\ B = 0.156$$

The ADG was adjusted to 100 kg by the formula:

$$ADG\ (\text{kg/day}) = \frac{100\ \text{kg}}{AGE}$$

where age to 100 kg (AGE) was calculated with the following formula:

$$AGE\ (day) = Measured\ age - \frac{Measured\ body\ weight - 100\ \text{kg}}{CF}$$

where correction factors (CF) were as follows:

$$Sire: CF = 1.826 \times \frac{Measured\ body\ weight}{Measured\ age}$$
$$Dam: CF = 1.715 \times \frac{Measured\ body\ weight}{Measured\ age}$$

Piglets were weighed within 24 h after the delivery of each individual to measure and record the birth weight. In Yorkshire and Landrace pigs, the numbers of phenotypic records for ADG, BFT and BW were 1734 and 1123, and for BTHWT were 887 and 405, respectively. Animals with missing records were excluded from the following analysis.

2.2. Genotype Data

Ear tissues collected from target swine were used for genomic DNA extraction with a standard phenol/chloroform method. The A260/A280 ratio of all DNA samples was qualified by electrophoresis and a light absorption. The samples with ratio in the range of 1.8–2.0 were diluted to a final concentration of 50 ng/μL. Genotyping was conducting with the KPS Porcine 60K SNP Chip (Beijing Compass Biotechnology Co., Ltd., Beijing, China), which contains 57,466 genome-wide SNPs across 18 autosomes and sex chromosomes. SNPs with call rate less than 95% and the minor allele frequency less than 1% were filtered out using PLINK [30]. Individuals with more than 5% missing genotype were removed. Unmapped or sex-chromosome located SNPs were also excluded. The quality controls for the two populations were implemented under the same criteria, with 44,119 and 44,142 SNPs remaining for Yorkshire and Landrace pigs, respectively. In meta-analysis of GWAS, a union set of SNP-chip from two populations that passed the same quality control procedure mentioned above were later used.

2.3. Imputation of SNP Chips

To obtain high density SNP data, genotype phasing and imputation from 60K SNP to whole genome sequencing (WGS) data were performed by Beagle using two reference panels [31], including WGS data of 19 Yorkshire pigs and 12 Landrace pigs from NCBI Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/sra/ accessed on 1 July 2020) and European Nucleotide Archive (ENA, https://www.ebi.ac.uk/ena accessed on 3 July 2020) under project PRJNA260763 (Table S5). SNPs calling was performed under the following pipeline. The raw reads were trimmed by filtering adapter and low-quality bases using Trimmomatic (version 0.40) [32]. Clean reads were aligned to the pig reference genome *Sus Scrofa* (v11.1.104) by BWA, and then the aligned reads were used to detect short variants [33]. The raw SNPs were generated by employing GATK under the instruction of

GATK best practice online documentation [34,35]. The GATK VariantFiltration parameter was set as "QUAL < 30.0 || QD < 2.0 || FS > 60.0 || MQ < 40.0 || SOR > 4.0 || ReadPosRankSum < −8.0 || MQRankSum < −12.5"to filter raw SNPs, followed by a further quality control step with the "–maf 0.05 –min-alleles 2 –max-alleles 2 –hwe 1e-6 –min-meanDP 5 –max-missing 0.9" option in Vcftools [36]. After the filtration, the autosome-only VCF files which contained 9,075,894 and 11,268,962 highly confident SNPs for Yorkshire and Landrace pigs, respectively, were used as reference panels of corresponding populations for genotype-imputation from the 60K SNP panel to WGS data. The average imputation accuracy based on Beagle R^2 was 0.71 and 0.69 for the two populations. The imputed SNPs with minor allele frequency (MAF) < 0.01 and Hardy–Weinberg Equilibrium < 1×10^{-6} were discarded. Finally, the remaining 1,468,003 and 2,221,629 SNPs for Yorkshire and Landrace pigs were available for the following analyses.

2.4. Statistical Models for GWAS

We used three statistic models including the mixed linear model, Bayesian model, and meta-analysis for GWAS, including (1) linear model based on 60K chip data (CL_GWAS); (2) linear model based on imputed WGS data (IL_GWAS); (3) Bayesian model based on 60K chip data (CB_GWAS); and (4) meta-analysis based on 60K chip data.

2.4.1. Mixed Linear Model

We performed CL_GWAS and IL_GWAS by fitting a univariate Mixed Linear Model (MLM) in GEMMA [37]:

$$y = W\alpha + X\beta + u + \varepsilon \quad (1)$$

where y denotes the vector of phenotypic values; W is the incident matrices of multiple covariates, comprised of sex, days to measure and top five eigenvectors of PCA which were generated by GCTA tool [38]; α denotes the vector of corresponding coefficients including the intercept; X is the vector of SNP genotypes; β denotes the marker effect; $u \sim \mathrm{MVN_n}\left(0, \lambda\tau^{-1}\mathrm{K}\right)$ and $\varepsilon \sim \mathrm{MVN_n}\left(0, \tau^{-1}\mathrm{I_n}\right)$ represent the vector of random effects and random residuals, respectively; τ^{-1} is the variance of the residual errors; λ is the ratio of variance components of random effects to random residuals; K is a genomic relatedness matrix; $\mathrm{I_n}$ is an n × n identity matrix; and $\mathrm{MVN_n}$ denotes the multivariare normal distribution with n-dimension. Bonferroni correction was used to determine the genome-wide and suggestive significance threshold values [39]. The estimated significant levels were, respectively, set as 1.13×10^{-6} and 2.25×10^{-5} for CL_GWAS, and 3.40×10^{-8} and 6.81×10^{-7} for imputed data-based GWAS.

2.4.2. Meta-Analysis

Combining the CL_GWAS results from two breeds by the same trait, a meta-analysis using an inverse variance weighting method was implemented by METAL [40], in which p values and effects of common SNPs were used as input data. Bonferroni correction threshold for meta-analysis was identical to CL_GWAS.

2.4.3. Bayesian Model

Bayes C model was implemented by GenSel, the prior assumption of which was that most markers were set as zero effect, with only a few SNPs explaining large variances [41]. Prior variance components of genetic effects and residuals were obtained using AIREMLF90 from BLUPF90 [42,43]. Gibbs-sampling chains for 50,000 iterations were run, and the first 40,000 burn-in iterations were discarded. The probFixed parameter of π was set to 0.999, which meant that one of 1000 SNPs were taken into the Bayesian analysis iterations and calculated with the given none-zero variance. The windowsBV option in GenSel was set to 1 Mb, which means that every consecutive 1 Mb SNPs region that followed the physical map order across the genome was divided into a genomic window. The percentage of genome-wide genetic variance explained by windows was utilized to identify

the informative regions. Windows with the proportion of genomic variance (GV) greater than 1% were considered as significant regions, which were then taken to the downstream analyses. Genetic variances of SNP effects were estimated with the same model as in the single-marker GWAS.

2.5. Candidate Genes Annotation

Sus scrofa reference genome (version 11.1) which served as the gene location map was downloaded from Ensembl database [44]. Genes that contained or were located nearby significant SNPs which were identified using three single-marker approaches (CL_GWAS, IL_GWAS and meta-analyses) were considered as candidate genes associated with growth traits. In addition, genes which overlapped significant genomic regions (1 Mb) that explained more than 1% genetic variance proportion from CB_GWAS were also considered to be candidate genes. These genes were annotated with KOBAS 2.0 [45]. The annotated KEGG pathways and GO biological processes were confirmed under Fisher's exact test with a significant threshold of $p < 0.05$.

2.6. Statistical Models for GS

2.6.1. GBLUP

The model of benchmark best linear unbiased prediction (GBLUP) is [46]:

$$y = W\alpha + Xg + \varepsilon \tag{2}$$

where y is a vector of phenotypic values; α denotes the vector of fixed effects; $g \sim N\left(0, G\sigma_g^2\right)$ is a vector of genomic estimated breeding value (GEBV); $\varepsilon \sim N\left(0, I\sigma_e^2\right)$ represents the vector of residual effects; W and X are incident matrices for α and g. Genomic relationship matrix G was calculated as $G = \frac{ZZ'}{2\sum p_i(1-p_i)}$, where p_i and Z represent MAF of markers and the adjusted MAF matrix, respectively. The G was constructed by different datasets of SNP markers depending on different strategies used [47].

2.6.2. Two-Kernel Based GBLUP

The model of two-kernel based GBLUP is:

$$y = W\alpha + X_1g_1 + X_2g_2 + \varepsilon \tag{3}$$

where $g_1 \sim N\left(0, G_1\sigma_{g_1}^2\right)$ represents the matrix of $GEBV_1$ estimated with G_1 constructed by all significant SNPs resulting from CL_GWAS, CB_GWAS IL_GWAS, and meta-analysis; $g_2 \sim N\left(0, G_2\sigma_{g_2}^2\right)$ represents the matrix of $GEBV_2$ estimated with G_2 constructed by the remaining SNPs in 60K SNP-chip. The sum of $GEBV_1$ and $GEBV_2$ were GEBV for the method.

2.7. Evaluation of the Accuracy of GS

Prediction accuracy (GPA) was calculated as:

$$GPA = \frac{\mathrm{r}(GEBV, y^*)}{\sqrt{h^2}} \tag{4}$$

where $\mathrm{r}(GEBV, y^*)$ represents the Pearson's correlation between GEBV and phenotypes corrected by fixed effects (y^*); h^2 stands for heritability of trait. All models above were implemented by BLUPF90 [43]. The predictive abilities of the defined models were measured with 20 replications of a 5-fold random cross-validation [47]. The average (SD) accuracies were presented as the results across replications.

3. Results and Discussion

3.1. Phenotypic Statistics and Heritability Estimation

Phenotypic summary statistics and heritability based on the 60K SNP-chip for four traits across two breeds are listed in Table 1. The coefficients of variation (CV) were in different ranges from 7.95% to 30.12% and 8.05% to 33.47%, which indicated that two breeds were in different variation levels. The results showed that ADG and BFT were moderately heritable for Yorkshire and Landrace, 0.44, 0.43 and 0.44, 0.40, which were similar to previous studies [14,48]. Heritability estimates of BW were also moderate, with 0.42 and 0.38 for two breeds. However, the additive genetic variance accounted for a small proportion of phenotypic inheritance in BTHWT. The heritability of Yorkshire was 0.16, consistent with the findings of [49], and the estimate of Landrace was 0.09 higher. As a low heritable trait, a great breeding improvement of BTHWT would be made by implementing marker-assisted selection or genomic selection.

Table 1. Phenotypic statistics and SNP chip-based heritability estimations for four growth-related traits in two breeds.

Breed	Trait	N	Mean ± SD	CV	Max	Min	h^2
Yorkshire	ADG	1734	565.32 ± 44.92	7.95	792.59	394.51	0.44
	BFT	1734	11.09 ± 3.34	30.12	28.04	5.17	0.43
	BTHWT	887	1.39 ± 0.25	17.99	2.42	0.60	0.16
	BW	1734	112.45 ± 14.64	13.02	167.10	74.01	0.42
Landrace	ADG	1123	608.19 ± 53.69	8.83	856.78	421.54	0.44
	BFT	1123	14.58 ± 4.88	33.47	34.95	5.46	0.40
	BTHWT	405	1.42 ± 0.26	18.31	2.17	0.71	0.27
	BW	1123	117.12 ± 16.25	13.87	160.12	80.09	0.38

ADG, average daily gain; BFT, backfat thickness; BTHWT, birth weight; BW, body weight.

3.2. Population Structure Analysis

Principal component analysis (PCA) was performed using a 60K SNP panel. Almost all individuals were divided into two clear clusters representing two different genetic backgrounds of target animals, with only a few individuals overlapping (Figure 1). The first two principal components explained 13.68% and 6.26% of the total variations. The top five eigenvectors were added in the GWAS models as covariates to correct for existing population stratification.

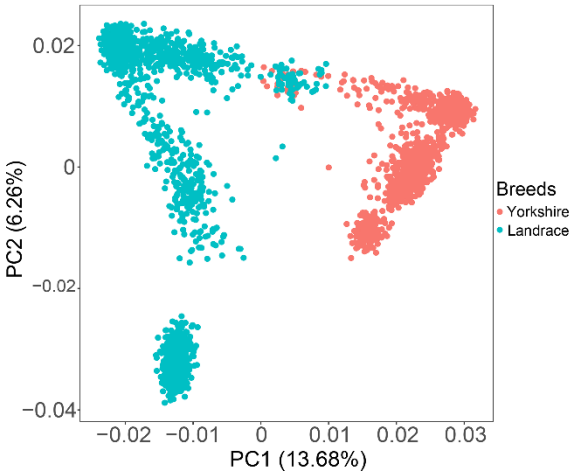

Figure 1. Principal component analysis shows the population structure between Landrace and Yorkshire pigs. Top two principal components: PC1 = 13.68%; PC2 = 6.26%.

3.3. GWAS Based on Mixed Linear Model and Bayesian Model

Results of significant genome-wide associations for four growth traits were shown in Figures 2 and 3 and Supplementary Figures S1 to S4. The number of identified SNPs for CL_GWAS and IL_GWAS were 48 and 142 (Tables S1 and S3), along with 44 genomic regions that explained more than 1% of the genetic variance from CB_GWAS (Table S2). Within the 65 genomic regions, several candidate genes were found to be related to growth traits (Table 2).

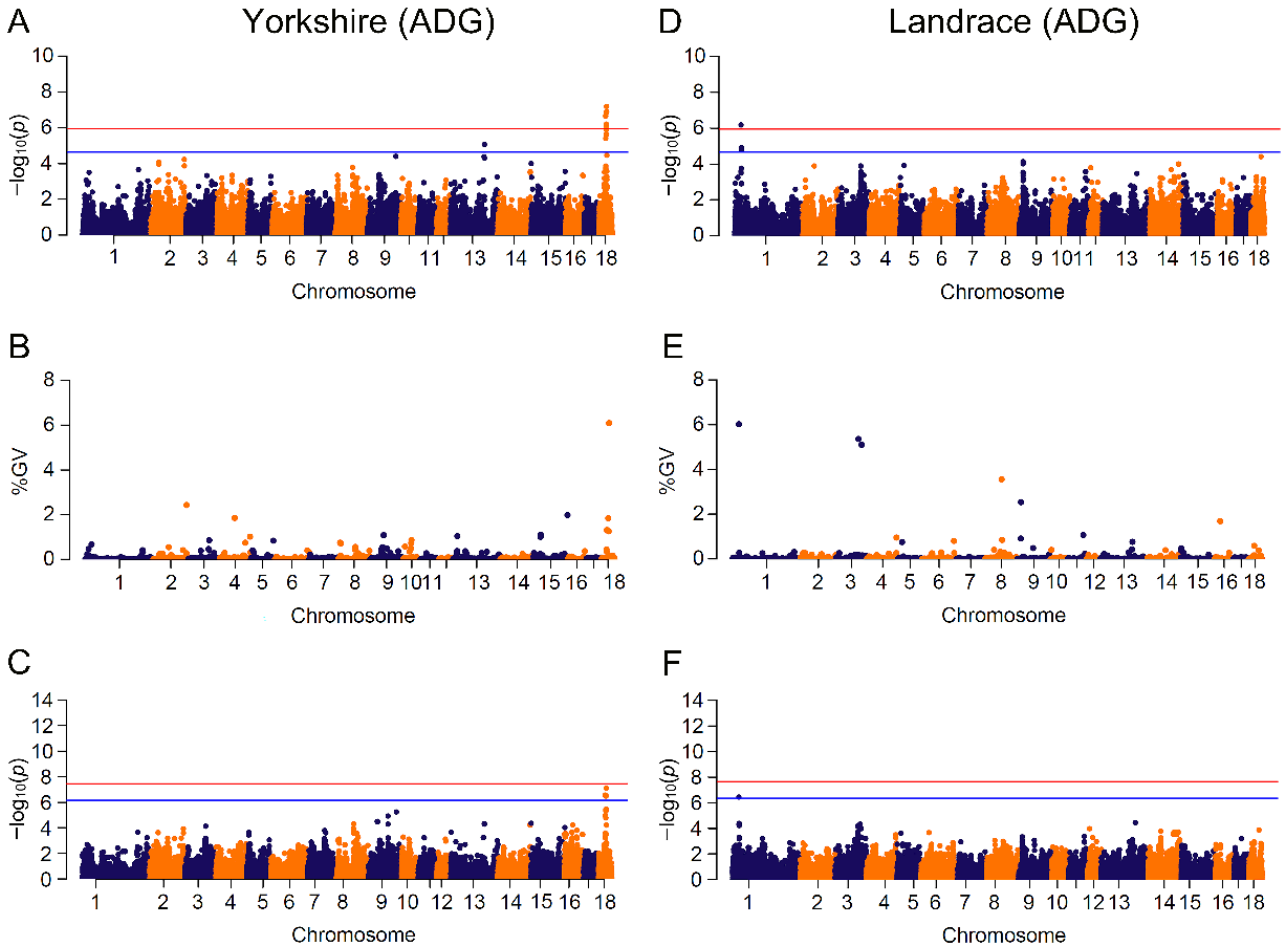

Figure 2. Manhattan plots for ADG (average daily gain) by breed in different GWAS approaches. The red and blue lines in plots represent genome-wide and suggestive significance thresholds, respectively. Results in Yorkshire population for: (**A**) CL_GWAS; (**B**) CB_GWAS; (**C**) IL_GWAS. Results in Landrace population for: (**D**) CL_GWAS; (**E**) CB_GWAS; (**F**) IL_GWAS. %GV denotes the proportion of genomic variance explained by the region identified.

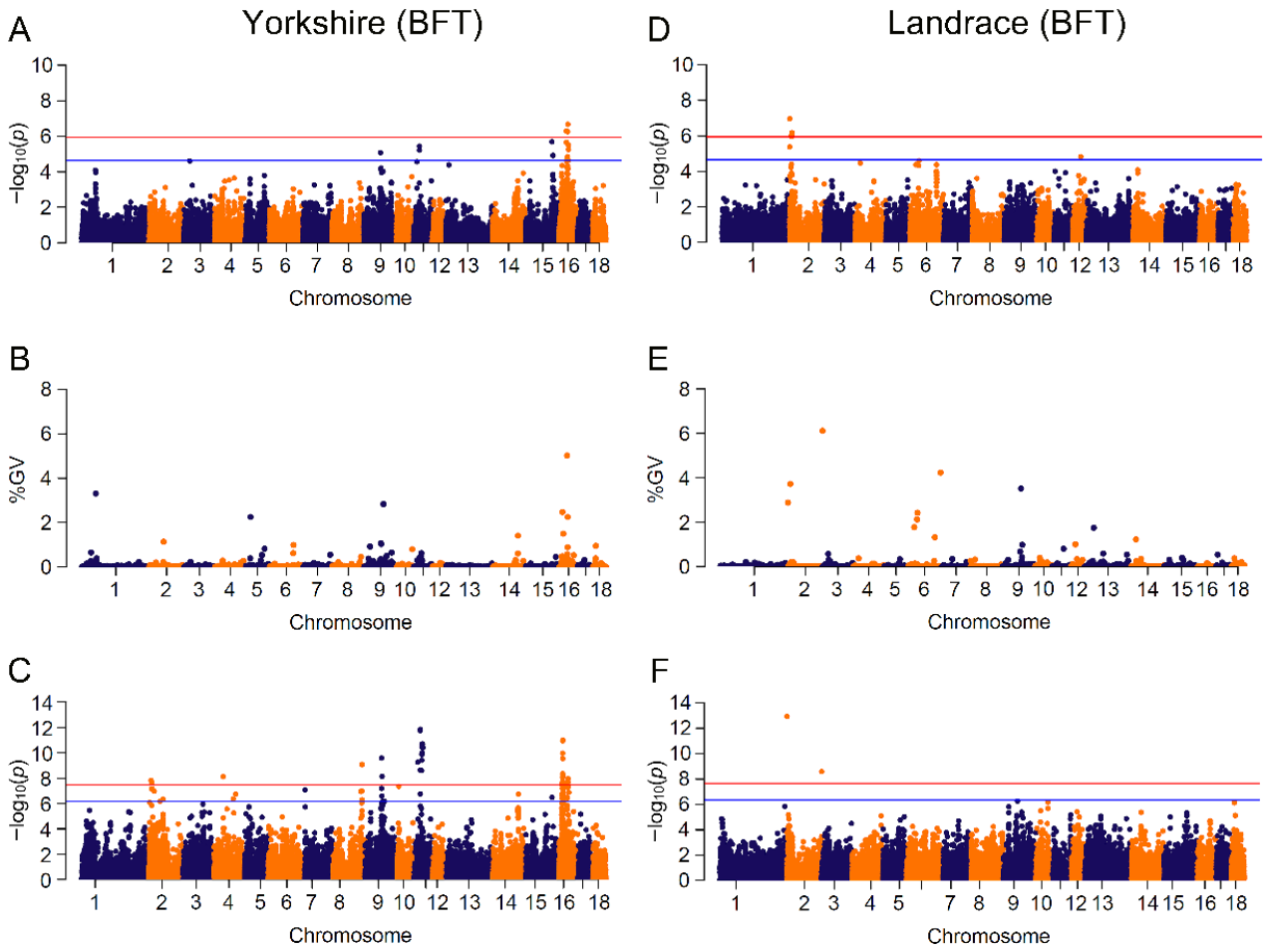

Figure 3. Manhattan plots for BFT (backfat thickness) by breed in different GWAS approaches. The red and blue lines in plots represent genome-wide and suggestive significance thresholds, respectively. Results in Yorkshire population for: (**A**) CL_GWAS; (**B**) CB_GWAS; (**C**) IL_GWAS. Results in Landrace population for: (**D**) CL_GWAS; (**E**) CB_GWAS; (**F**) IL_GWAS. %GV denotes the proportion of genomic variance explained by the region identified.

Four candidate genes (*MDFIC*, *RPS12*, *PDE4D*, *AQP4*) were putatively linked to ADG and BW in pigs, which played supporting roles in the biological process of physical or muscular growth. As a myogenic regulatory factor, the post-transcription downregulation of *MDFIC* (MyoD family inhibitor domain containing) controls the promotion of myogenic differentiation, which is beneficial for skeletal muscle development in bovines [50]. The gene *MDFIC* was putatively linked to the birth weight of piglets [51] and the meat to fat ratio in F2 × cross pigs [52], and harbored two SNPs which were associated with ADG on SSC18:31.02~31.04 Mb. Ribosomal protein S12 (*RPS12*), mapped by two SNPs on SSC1:30.88~30.92 Mb, was examined to regulate the process of nucleic acid translation and physical growth in Drosophila by controlling *Xrp1* expressing levels [53]. Induced by fibroblast growth factor 1 (*FGF1*), phosphodiesterase 4D (*PDE4D*) was associated with growth and fertility impairment [54] and lipolysis suppression [55]. Assigned to the most significant SNP on SSC16:111.36 Mb for trait BW, aquaporin 4 (*AQP4*) was identified to widely express in the digestive tract of guinea pigs, and the mediation of gastric acid secretion in mice [56,57], which affected the physical growth in animals.

Table 2. Summary of genomic regions significantly associated with growth-related traits of three statistical analyses.

Traits [1]	Breeds [2]	The Number of Significant SNPs			Candidate Genes
		CL_GWAS [3]	CB_GWAS [4]	IL_GWAS [5]	
ADG	YY	14	242 (12)	6	*MDFIC*, *FOXP2*, *DOCK4*, *IMMP2L*, *ZPLD1*, *CYP7B1*
	LL	3	164 (7)	1	*ALDH8A1*, *RPS12*
BFT	YY	20	219 (11)	71	*UMAD1*, *GLCCI1*, *PDE4D*, *ZSWIM6*, *RNF180*, *ANKRD55*, *NDUFS4*, *NDUFA4*
	LL	5	193 (12)	2	*ODF3*, *DEAF1*, *PACS1*, *ZNF300*, *MS4A8*, *MS4A13*
BTHWT	YY	4		20	*EDRF1*, *DHX32*, *GMNN*, *MPP7*, *CUBN*, *ITGA8*, *RPP38*, *UCMA*
	LL		104 (5)		*ASAP1*, *NAV3*, *MROH5*, *PTP4A3*, *GPR20*
BW	YY	6	245 (12)	40	*TAF4B*, *AQP4*, *RORB*, *ATXN1*, *TAFA5*, *SELENOI*, *TMEM104*
	LL	1	174 (10)	2	*AMER2*, *MTMR6*, *NUP58*, *ATP8A2*, *SHISA2*, *FAM171A1*

[1] ADG: average daily gain; BFT: backfat thickness; BTHWT: birth weight; BW: body weight. [2] YY: Yorkshire population; LL: Landrace population. [3] The number of significant SNPs identified in CL_GWAS with a threshold of 2.25×10^{-5}. [4] The total number of SNPs located in genomic regions identified in CB_GWAS, explaining more than 1% of genetic variance. The digit in brackets represents the number of genomic regions identified in CB_GWAS, explaining more than 1% of genetic variance. [5] The number of significant SNPs identified in IL_GWAS with a threshold of 6.81×10^{-7}.

Seven candidate genes (*PNPLA2*, *DEAF1*, *PLCD3*, *ANKRD55*, *RORB*, *NDUFS4*, *ALDH8A1*) were found to potentially affect the trait of BFT in pigs, and are involved in pathways of fatty acid metabolism, adipogenesis and insulin signaling. Candidate gene *PNPLA2* (patatin like phospholipase domain containing 2), overlapping a genomic region with GV of 2.88% on SSC2:0.01~0.96 Mb, controls the initial step in triglyceride hydrolysis of long-chain fatty acid esters in adipocyte and non-adipocyte, and regulates adiposomes size [58]. Crespo-Piazuelo et al. identified that *PNPLA2* was embedded in a region on SSC2:0~12.76 Mb which was associated with the abundance of three fatty acids in backfat in pigs [59]. Gene *DEAF1* (deformed epidermal autoregulatory factor 1 homolog) mapped an identified variant on SSC2, is involved in fatty acid deposition, and regulated the expression of peripheral tissue antigen in lymph node stromal cells among diabetic mice and humans [52,60,61]. Phospholipase c delta 3 (*PLCD3*) played a key role in phosphatidylinositol catabolism by hydrolyzing phospholipids into fatty acids [62], and was linked to backfat thickness in Chinese native pigs [63] and intramuscular fat deposition in African Ankole cattle [64]. Assigned to a significant SNP on SSC16, *ANKRD55* (ankyrin repeat domain 55) showed a high correlation with insulin resistance and lipid metabolism, which may result in adiposity and have a risk of Type 2 Diabetes in humans [65,66]. With low expression in hyperglycemic donors, retinoid-related orphan receptor beta (*RORB*) positively regulated insulin secretion in pancreatic β cells [67] and acted as a regulator of osteogenesis [68]. It was closed to a variant on SSC1 and reported that backfat thickness showed a positive correlation with intramuscular fatness, which is a key factor affecting the flavor and juiciness of pork [69–71]. Gene *NDUFS4* (NADH: ubiquinone oxidoreductase subunit S4), harboring two identified markers on SSC16:23.93~23.97 Mb, was reported to highly express in high intramuscular fat in pigs [72], and *ALDH8A1* (aldehyde dehydrogenase 8 family member A1) was shown to be significantly associated with fatness in large white pigs [73].

The annotation of candidate genes supported that cellular processes in early stages may impact birth weight of piglets, and consequently affect many post-natal traits including growth performance and meat quality. For BTHWT in Yorkshire, three SNPs located

in SSC14:135.03~135.18 Mb that were shared in two single-marker models completely coincided with the QTL identified in the birth weight of Yorkshire piglets [74]. Closed to these imprinted 13 kb regions, erythroid differentiation regulatory factor 1 (*EDRF1*) mediated organ development and histological differentiation and controlled erythroid development and differentiation in primitive cells or organs, such as fetal liver and fetal kidney [75]. The gene DEAD-box helicase 32 (*DHX32*), overlapping the two most prominent SNPs, has an ability to unwind DNA and RNA, and is involved in embryogenesis or cellular growth and division by changing RNA structure [76]. Geminin DNA Replication Inhibitor (*GMNN*), closed to the BTHWT-linked marker on SSC7:19.60 Mb, plays a crucial role in the S and G2 phases of the cell cycle. The gene *GMNN* is involved in the regulation of DNA replication and cell fate control during embryonic development [77]. Moreover, *ZNF300* (zinc finger protein 300) served as a zinc finger-domain protein regulating embryonic development and transcription. Gene *TAF4B* (TATA-box binding protein associated factor 4b) was essential for the maintenance of spermatogonia stem cells in mammals [78].

In our study, 48 SNPs related to the four traits were identified in the two populations. In each single-marker analysis, we found several significant SNPs that were in high linkage disequilibrium within a small region. For example, 13 ADG-associated SNPs on SSC18 from 31 Mb to 35.1 Mb were partitioned into four blocks with high pairwise LDs of 0.86, 0.58, 0.99 and 0.80. However, single-marker analysis only identified the association between a single SNP and traits of interest, neglecting the LD effect of adjoining SNPs [18,19]. Since *p*-value was used to test the statistical significance of SNPs [79], markers were likely to lack the power to reach the stringent threshold when association analysis came to small sample size test or low heritability of trait. Bayesian methods, which derived from the genomic estimation of breeding value [1,79,80], simultaneously estimated SNP effects without any significance testing. To take the genetic interaction effect into account, these models investigated the variation of genomic regions by fitting multiple SNPs. The Bayes C model used in this study assumed that most variants had a non-zero effect, which was consistent with MLM method, but only a small part of SNPs had zero effect [80]. From the comparison between the Bayes C model and single-marker regression analysis, most of the significant SNPs were embedded in the genomic regions with large variances, indicating that the linked SNPs did have large effects and confirming that the results of the two methods were in high accordance. On the other hand, numerous regions covered no significant SNP in single-marker analyses and also accounted for over 1% of genetic variations, such as the genomic region on SSC2:162.08~162.31 Mb explaining 6.11% of genetic variance with no identified SNP in CL_GWAS for BFT in Landrace. The single variant within the SSC2:162.08~162.31 Mb region was unable to surpass the significance threshold in CL_GWAS, but the total effects of these variants in this region might be substantial due to loci interactions. It has been reported that non-significant SNPs might have large effects for phenotype variability, and putatively causal variants that failed to surpass the threshold could be addressed by Bayesian analyses [81]. However, IL_GWAS identified a BFT-related locus ($-\log(p\text{-value}) > 8$) within that region, suggesting that increased marker density might construct a more precise genomic relatedness matrix, and was able to highlight non-significant SNPs among genomic regions capturing large genetic variance. GWAS based on genomic regions with a Bayesian model was able to effectively discover putative QTLs and candidate genes [82,83], but to date few studies performed Bayesian GWAS on growth traits.

Performing GWAS with imputed genotypes is an effective way to enhance SNP density at a low cost [84]. We identified 142 loci associated with four traits across two breeds, composed of 122 novel loci and 20 common loci identified with CL_GWAS. With the IL_GWAS, some of the significant loci obtained a much stronger significance to surpass a more stringent threshold. For instance, the suggestive-significant marker CNC10110504 in CL_GWAS became the peak variant across the genome. A possible reason is that the overall effect of the genomic relatedness matrix in IL_GWAS was different from that in

CL_GWAS [85,86]. However, the advantage of higher SNP density was not observed in Landrace.

3.4. Meta-Analyses

Meta-analysis uplifted the detection power by taking all informative SNPs into account, including those SNPs filtered out by quality control in single population GWAS [28]. We used meta-analysis to uncover potential associated loci by pooling CL_GWAS of Yorkshire and Landrace breeds. We detected 40 significant SNPs associated with four traits: fourteen for ADG, twelve for BFT, three for BTHWT and five for BW (Figure 4 and Table S4), 30 of which were consistent with CL_GWAS by single population. Among the ten novel SNPs (Table S4), three neighboring loci associated with ADG in SSC18:25.13~25.23 Mb were in the range of *PTPRZ1* (protein tyrosine phosphatase receptor type Z1), which encodes protein tyrosine phosphatase that regulates the cell proliferation on the embryonic spinal cord. Another ADG-associated SNP located in SSC15:0.17 Mb was closely adjacent to candidate gene signal transducing adaptor molecule 2 (*STAM2*) which plays a role in intracellular signal transduction induced by growth factors and cytokines [87]. Yang et al. detected seven genetic variants of the *STAM2* gene and demonstrated that these variants had significant associations with growth performance in Chinese Wuchuan Black cattle [88]. Moreover, two candidate genes, *ADAMTS6* (a disintegrin and metalloproteinase with thrombospondin motifs 6) and *BMP2K* (bone morphogenetic protein 2 inducible protein kinase), were assigned to the biological pathway of skeletal development and patterning. The lack of gene *ADAMTS6* leads to a drastic reduction in aggrecan and cartilage link protein, the impairment of bone morphogenetic proteins (BMP) signaling in cartilage and the detention of growth differentiation factors, which consequently result in impaired skeletal development [89]. BMP plays a key role in skeletal development and patterning, and protein encoded by *BMP2K* is considered as a kinase potentially regulating the attenuation of osteoblast differentiation [90]. Since body fatness was reported to negatively correlate with bone weight, bone mineral content and density [91], *ADAMTS6* and *BMP2K* may play roles in body weight in pigs.

3.5. Genomic Prediction

Genomic prediction using GBLUP and a 60K chip panel (CHIP) was set as a benchmark (Figure 5A,C). GPAs for three traits (ADG, BFT and BW) with moderate heritability ranged from 0.42 to 0.50 in Yorkshire, and from 0.43 to 0.55 in Landrace (Table S6). The prediction accuracies of the trait BTHWT with low heritability were 0.13 ± 0.05 and 0.21 ± 0.07 in the two breeds. The heritabilities and prediction accuracies based on the 60K chip panel for the four traits were highly positively correlated in the two breeds ($r = 0.99$) (Figure 5B,D).

Compared with using SNP chip data, the prediction accuracy using imputed WGS data had almost no improvement for ADG and BW in both breeds, which was contrary to the report for ADG in pigs [2], while for BTHWT prediction accuracies in Yorkshire and Landrace were increased by 22% and 10%, respectively. For the prediction accuracies of BFT, no improvement was observed in Landrace, which was similar to studies for backfat thickness in pigs reported by Zhang et al. [2] and Pérez-Enciso et al. [92], but a 7% improvement was yielded with the WGS dataset in Yorkshire. Genomic prediction improved with a large amount of variants identified for traits of BFT and BTHWT in Yorkshire, which was congruent with previous reports that genomic prediction with imputed data was affected by the genetic architecture of traits [2,5,92], although it can also be influenced by factors such as the statistical models used [92,93] and LD pruning before imputation [85,94].

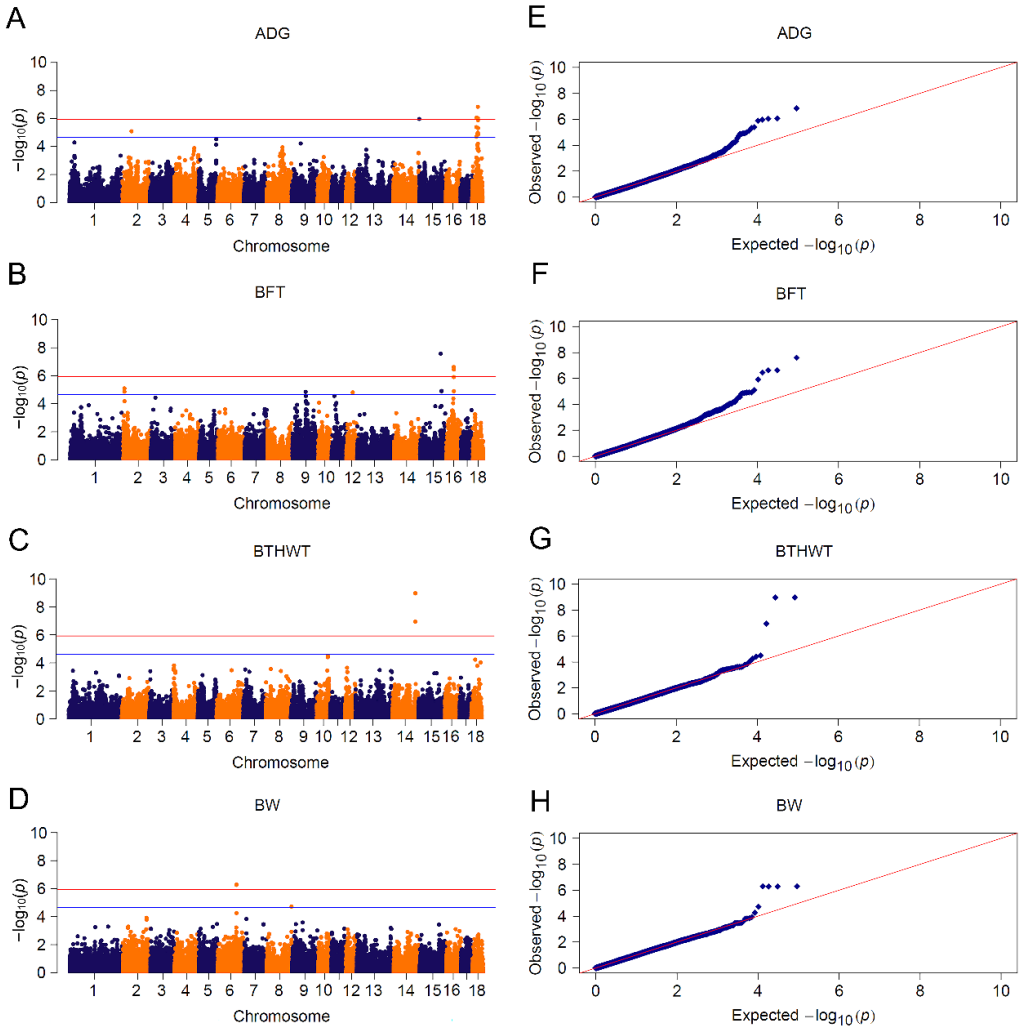

Figure 4. Meta-analyses for 4 growth traits. Manhattan plots for: (**A**) ADG: average daily gain; (**B**) BFT: backfat thickness; (**C**) BTHWT: birth weight and (**D**) BW: body weight. Q-Q plots for: (**E**) ADG, (**F**) BFT, (**G**) BTHWT and (**H**) BW. The red and blue lines in plots represent genome-wide and suggestive significance thresholds, respectively.

We used significant variants (Table 2 and Table S4) preselected from CL_GWAS, IL_GWAS, meta-analysis, and SNPs located in genomic regions with high GV% from CB_GWAS in GS (SIG). In the comparison with using the complete chip data, prediction accuracies resulting from preselected SNPs for BFT in Yorkshire were 0.56 ± 0.08, which increased by 10%, and likewise for BW the GPAs increased by 8% and 4% in Yorkshire and Landrace, respectively (Figure 5A,C). The improvement was even more significant for the trait with low heritability, such as BTHWT, with GPA increasing by 45% in Yorkshire and 46% in Landrace. Taken together, the use of preselected SNPs had the potential to improve the GPA by reason of being free from the interference of non-significant loci [7,8]. Corredor et al. found that genomic prediction with SNPs identified to be linked to the vulva size in pigs had a better performance than that using all SNPs in the chip panel [11]. Van den Berg et al. suggested using only variants closed to the causal variants, and rare sequence variants closed to rare causal variants could improve prediction accuracies in dairy cattle, with accuracies ranging from 0.73 to 0.95 [95]. Such improvements were not observed in some studies which only used one of the GWAS methods to identify significant SNPs [96,97]. We then hypothesized that prediction accuracy using preselected data could be determined by the number of causal SNPs detected [98]. We pooled all SNPs identified by three different GWAS methods in genomic prediction. For example, significant causal variants missed

by using a single GWAS method might be identified by other GWAS methods, which can be considered as supplements to the main model used, although there may be more false positive loci with no effects identified. For BFT in Yorkshire, GPAs estimated by two-kernel based GBLUP (TK) saw a slight improvement compared to SIG (Figure 5A,C). Compared with CHIP, TK improved by 23% and 29% for BTHWT in the two breeds, and by 4% for BW in Landrace. A similar study [9] revealed that TK outperformed the single kernel-based model with pre-selected variants by 0.11, and with all markers on the SNP panel by 0.06 in bulls, suggesting that the prioritization or appropriate weighting of pre-selected functional markers have advantages in GS for some traits.

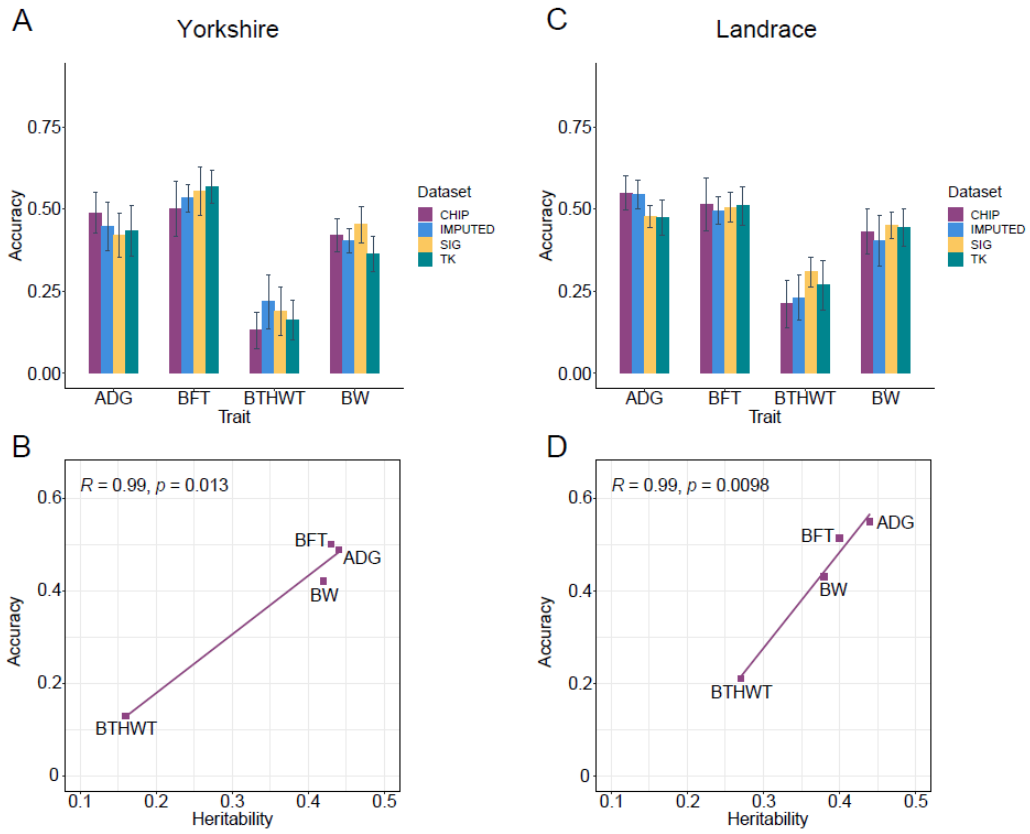

Figure 5. Genomic prediction accuracies (GPA) of GS using three sets of SNPs for average daily gain (ADG), backfat thickness (BFT), birth weight (BTHWT) and body weight (BW). Results for Yorkshire and Landrace are in (**A**,**C**), respectively. CHIP and IMPUTED represent GPAs estimated by GBLUP model based on all loci in 60K chip and imputed data, respectively. SIG represents GPAs estimated by GBLUP model based on the combination of significant variants from CL_GWAS, CB_GWAS and IL_GWAS. TK represents GPAs estimated by two-kernel based GBLUP based on the combination of significant variants from CL_GWAS, CB_GWAS and IL_GWAS. (**B**,**D**) represent correlation fitting curves between heritabilities and GPAs of 60K chip SNPs in Yorkshire and Landrace, respectively. Error bars represent the standard deviation of the accuracy across replications.

4. Conclusions

In this study, we employed three different statistical methods as well as 60K SNP-chip and whole genome sequence data (CL_GWAS, CB_GWAS, IL_GWAS and meta-analysis based on chip data) for GWAS, which allowed us to discover more candidate loci linked to the four growth traits. A total of 1485 candidate loci, and 24 candidate genes which are involved in skeletal muscle development, fatty deposition, lipid metabolism and insulin resistance, were identified. Using the pre-selected functional SNPs in GS outperformed using all 60K SNP-chip and imputed WGS data for some traits, suggesting

that the prioritization of preselected functional markers in GS models has a potential to improve prediction accuracies in livestock breeding.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ani13040722/s1. Table S1: Significant SNPs related to growth traits by breed in CL_GWAS. Table S2: Genomic regions explained ≥ 1% genomic variance from CB_GWAS in traits by breed. Table S3: Significant SNPs related to growth traits by breed in IL_GWAS. Table S4: Significant SNPs identified in meta-analysis for four traits. Table S5: Sequencing depth of WGS data of reference panel. Table S6: Genomic prediction accuracies of different strategies by traits. Figure S1: Manhattan plots for BTHWT (birth weight) by breed in different GWAS approaches. Figure S2: Manhattan plots for BW (body weight) by breed in different GWAS approaches. Figure S3: Q-Q plots for CL_GWAS of four growth traits in Yorkshire and Landrace population. Figure S4: Q-Q plots for IL_GWAS of four growth traits in Yorkshire and Landrace population.

Author Contributions: Conceptualization, Z.L. (Zhengcao Li); investigation, Y.Z., B.J., Z.L. (Zhenyang Li) and Y.Q.; visualization, R.Z. and T.L.; formal analysis, R.Z.; writing—original draft preparation, R.Z.; writing—review and editing, Z.L. (Zhengcao Li); supervision, Y.C. and Z.L. (Zhengcao Li); All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Pig Industry Technology System (CARS-35), and the Special Project for Research and Development in Key Areas of Guangdong Province (Nos. 2018B020203003).

Institutional Review Board Statement: The animal study protocol was approved by the Ministry of Agricultural of China and the Institutional Animal Care and Use Committee (IACUC) of Sun Yat-sen University (Approval number SYXK-2017-0175).

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be available upon request.

Acknowledgments: We would like to thank Guangdong IPIG Technology Co., Ltd. (Guangdong, China) for providing all phenotypic data and ear tissue samples.

Conflicts of Interest: The authors declare no competing interest. No financial and personal relationships with other people or organizations that can inappropriately influence the study. No professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

References

1. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **2001**, *157*, 1819–1829. [CrossRef] [PubMed]
2. Zhang, C.; Kemp, R.A.; Stothard, P.; Wang, Z.; Boddicker, N.; Krivushin, K.; Dekkers, J.; Plastow, G. Genomic evaluation of feed efficiency component traits in Duroc pigs using 80K, 650K and whole-genome sequence variants. *Genet. Sel. Evol.* **2018**, *50*, 14. [CrossRef] [PubMed]
3. Meuwissen, T.; Goddard, M. Accurate Prediction of Genetic Values for Complex Traits by Whole-Genome Resequencing. *Genetics* **2010**, *185*, 623–631. [CrossRef] [PubMed]
4. Druet, T.; Macleod, I.M.; Hayes, B.J. Toward genomic prediction from whole-genome sequence data: Impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity* **2014**, *112*, 39–47. [CrossRef] [PubMed]
5. Iheshiulor, O.O.M.; Woolliams, J.A.; Yu, X.; Wellmann, R.; Meuwissen, T.H.E. Within- and across-breed genomic prediction using whole-genome sequence and single nucleotide polymorphism panels. *Genet. Sel. Evol.* **2016**, *48*, 15. [CrossRef] [PubMed]
6. Chang, L.-Y.; Toghiani, S.; Ling, A.; Aggrey, S.E.; Rekaya, R. High density marker panels, SNPs prioritizing and accuracy of genomic selection. *BMC Genet.* **2018**, *19*, 4. [CrossRef]
7. Ni, G.; Cavero, D.; Fangmann, A.; Erbe, M.; Simianer, H. Whole-genome sequence-based genomic prediction in laying chickens with different genomic relationship matrices to account for genetic architecture. *Genet. Sel. Evol.* **2017**, *49*, 8. [CrossRef]
8. Teissier, M.; Sanchez, M.P.; Boussaha, M.; Barbat, A.; Hoze, C.; Robert-Granie, C.; Croiseau, P. Use of meta-analyses and joint analyses to select variants in whole genome sequences for genomic evaluation: An application in milk production of French dairy cattle breeds. *J. Dairy Sci.* **2018**, *101*, 3126–3139. [CrossRef]
9. Raymond, B.; Bouwman, A.C.; Wientjes, Y.C.J.; Schrooten, C.; Houwing-Duistermaat, J.; Veerkamp, R.F. Genomic prediction for numerically small breeds, using models with pre-selected and differentially weighted markers. *Genet. Sel. Evol.* **2018**, *50*, 49. [CrossRef]

10. MacLeod, I.M.; Bowman, P.J.; Vander Jagt, C.J.; Haile-Mariam, M.; Kemper, K.E.; Chamberlain, A.J.; Schrooten, C.; Hayes, B.J.; Goddard, M.E. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genom.* **2016**, *17*, 144. [CrossRef]
11. Corredor, F.-A.; Sanglard, L.P.; Leach, R.J.; Ross, J.W.; Keating, A.F.; Serão, N.V.L. Genetic and genomic characterization of vulva size traits in Yorkshire and Landrace gilts. *BMC Genet.* **2020**, *21*, 28. [CrossRef] [PubMed]
12. Do, D.N.; Janss, L.L.G.; Jensen, J.; Kadarmideen, H.N. SNP annotation-based whole genomic prediction and selection: An application to feed efficiency and its component traits in pigs. *J. Anim. Sci.* **2015**, *93*, 2056–2063. [CrossRef] [PubMed]
13. Brøndum, R.F.; Su, G.; Janss, L.; Sahana, G.; Guldbrandtsen, B.; Boichard, D.; Lund, M.S. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. *J. Dairy Sci.* **2015**, *98*, 4107–4116. [CrossRef] [PubMed]
14. Gozalo-Marcilla, M.; Buntjer, J.; Johnsson, M.; Batista, L.; Diez, F.; Werner, C.R.; Chen, C.-Y.; Gorjanc, G.; Mellanby, R.J.; Hickey, J.M.; et al. Genetic architecture and major genes for backfat thickness in pig lines of diverse genetic backgrounds. *Genet. Sel. Evol.* **2021**, *53*, 76. [CrossRef]
15. Qiao, R.; Gao, J.; Zhang, Z.; Li, L.; Xie, X.; Fan, Y.; Cui, L.; Ma, J.; Ai, H.; Ren, J.; et al. Genome-wide association analyses reveal significant loci and strong candidate genes for growth and fatness traits in two pig populations. *Genet. Sel. Evol.* **2015**, *47*, 17. [CrossRef]
16. Sahana, G.; Guldbrandtsen, B.; Janss, L.; Lund, M.S. Comparison of association mapping methods in a complex pedigreed population. *Genet. Epidemiol.* **2010**, *34*, 455–462. [CrossRef]
17. Dashab, G.R.; Kadri, N.K.; Shariati, M.M.; Sahana, G. Comparison of linear mixed model analysis and genealogy-based haplotype clustering with a Bayesian approach for association mapping in a pedigreed population. *BMC Proc.* **2012**, *6*, S4. [CrossRef]
18. Akey, J.; Jin, L.; Xiong, M. Haplotypes vs single marker linkage disequilibrium tests: What do we gain? *Eur. J. Hum. Genet.* **2001**, *9*, 291–300. [CrossRef]
19. Wang, X.; Morris, N.J.; Schaid, D.J.; Elston, R.C. Power of Single- vs. Multi-Marker Tests of Association. *Genet. Epidemiol.* **2012**, *36*, 480–487. [CrossRef]
20. Kemper, K.E.; Reich, C.M.; Bowman, P.J.; vander Jagt, C.J.; Chamberlain, A.J.; Mason, B.A.; Hayes, B.J.; Goddard, M.E. Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. *Genet. Sel. Evol.* **2015**, *47*, 29. [CrossRef]
21. Evangelou, E.; Ioannidis, J.P.A. Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* **2013**, *14*, 379–389. [CrossRef] [PubMed]
22. Leroy, G. Inbreeding depression in livestock species: Review and meta-analysis. *Anim. Genet.* **2014**, *45*, 618–628. [CrossRef] [PubMed]
23. van den Berg, I.; Boichard, D.; Lund, M.S. Comparing power and precision of within-breed and multibreed genome-wide association studies of production traits using whole-genome sequence data for 5 French and Danish dairy cattle breeds. *J. Dairy Sci.* **2016**, *99*, 8932–8945. [CrossRef] [PubMed]
24. Ritchie, H.; Rosado, P.; Roser, M. Meat and Dairy Production. *Our World in Data*. 2017. Available online: https://ourworldindata.org/meat-production (accessed on 8 February 2023).
25. Ponzoni, R.W.; Davies, R.L. An evaluation of biological and conventional pig selection indices. *Aust. J. Exp. Agric.* **1989**, *29*, 775–779. [CrossRef]
26. McKay, R.M. Responses to index selection for reduced backfat thickness and increased growth rate in swine. *Can. J. Anim. Sci.* **1990**, *70*, 973–977. [CrossRef]
27. Martínez-Montes, Á.M.; Fernández, A.; Muñoz, M.; Noguera, J.L.; Folch, J.M.; Fernández, A.I. Using genome wide association studies to identify common QTL regions in three different genetic backgrounds based on Iberian pig breed. *PLoS ONE* **2018**, *13*, e0190184. [CrossRef]
28. Guo, Y.; Qiu, H.; Xiao, S.; Wu, Z.; Yang, M.; Yang, J.; Ren, J.; Huang, L. A genome-wide association study identifies genomic loci associated with backfat thickness, carcass weight, and body weight in two commercial pig populations. *J. Appl. Genet.* **2017**, *58*, 499–508. [CrossRef]
29. Zhou, S.; Ding, R.; Meng, F.; Wang, X.; Zhuang, Z.; Quan, J.; Geng, Q.; Wu, J.; Zheng, E.; Wu, Z.; et al. A meta-analysis of genome-wide association studies for average daily gain and lean meat percentage in two Duroc pig populations. *BMC Genom.* **2021**, *22*, 12. [CrossRef]
30. Chang, C.C.; Chow, C.C.; Tellier, L.C.A.M.; Vattikuti, S.; Purcell, S.M.; Lee, J.J. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* **2015**, *4*, s13742-015. [CrossRef]
31. Browning, B.L.; Zhou, Y.; Browning, S.R. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* **2018**, *103*, 338–348. [CrossRef]
32. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
33. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.
34. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 11.10.1–11.10.33. [CrossRef]

35. DePristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V.; Maguire, J.R.; Hartl, C.; Philippakis, A.A.; del Angel, G.; Rivas, M.A.; Hanna, M.; et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **2011**, *43*, 491–498. [CrossRef] [PubMed]
36. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef] [PubMed]
37. Zhou, X.; Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **2012**, *44*, 821–824. [CrossRef] [PubMed]
38. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **2011**, *88*, 76–82. [CrossRef]
39. Bland, J.M.; Altman, D.G. Multiple significance tests: The Bonferroni method. *BMJ* **1995**, *310*, 170. [CrossRef]
40. Willer, C.J.; Li, Y.; Abecasis, G.R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **2010**, *26*, 2190–2191. [CrossRef]
41. Verbyla, K.L.; Hayes, B.J.; Bowman, P.J.; Goddard, M.E. Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. *Genet. Res.* **2009**, *91*, 307–311. [CrossRef]
42. Misztal, I.; Tsuruta, S.; Lourenco, D.; Masuda, Y.; Aguilar, I.; Legarra, A.; Vitezica, Z. *Manual for BLUPF90 Family Programs*; University of Georgia: Athens, Greek, 2018.
43. Masuda, Y. *Introduction to BLUPF90 Suite Programs*; University of Georgia: Athens, Greek, 2018.
44. Yates, A.D.; Achuthan, P.; Akanni, W.; Allen, J.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Azov, A.G.; Bennett, R.; et al. Ensembl 2020. *Nucleic Acids Res.* **2020**, *48*, D682–D688. [CrossRef] [PubMed]
45. Xie, C.; Mao, X.; Huang, J.; Ding, Y.; Wu, J.; Dong, S.; Kong, L.; Gao, G.; Li, C.-Y.; Wei, L. KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **2011**, *39*, W316–W322. [CrossRef] [PubMed]
46. VanRaden, P.M. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [CrossRef] [PubMed]
47. Li, Z.; Gao, N.; Martini, J.W.R.; Simianer, H. Integrating Gene Expression Data Into Genomic Prediction. *Front. Genet.* **2019**, *10*, 126. [CrossRef]
48. Hong, J.K.; Jeong, Y.D.; Cho, E.S.; Choi, T.J.; Kim, Y.M.; Cho, K.H.; Lee, J.B.; Lim, H.T.; Lee, D.H. A genome-wide association study of social genetic effects in Landrace pigs. *Asian-Australas. J. Anim. Sci.* **2018**, *31*, 784–790. [CrossRef]
49. Li, Y.; Li, B.; Yang, M.; Han, H.; Chen, T.; Wei, Q.; Miao, Z.; Yin, L.; Wang, R.; Shen, J.; et al. Genome-Wide Association Study and Fine Mapping Reveals Candidate Genes for Birth Weight of Yorkshire and Landrace Pigs. *Front. Genet.* **2020**, *11*, 183. [CrossRef]
50. Hu, X.; Xing, Y.; Ren, L.; Wang, Y.; Li, Q.; Yang, Q.; Du, M.; Xu, L.; Willems, L.; Li, J.; et al. bta-miR-23a Regulates the Myogenic Differentiation of Fetal Bovine Skeletal Muscle-Derived Progenitor Cells by Targeting MDFIC Gene. *Genes* **2020**, *11*, 1232. [CrossRef]
51. Zhang, L.; Xiang, Z.; Michal, J.J.; Ding, B.; Rui, L.; Jiang, Z. Genome Wide Screening of Candidate Genes for Improving Piglet Birth Weight Using High and Low Estimated Breeding Value Populations. *Int. J. Biol. Sci.* **2014**, *10*, 236–244. [CrossRef]
52. Falker-Gieske, C.; Blaj, I.; Preuß, S.; Bennewitz, J.; Thaller, G.; Tetens, J. GWAS for Meat and Carcass Traits Using Imputed Sequence Level Genotypes in Pooled F2-Designs in Pigs. *G3 Genes | Genomes | Genet.* **2019**, *9*, 2823–2834. [CrossRef]
53. Ji, Z.; Kiparaki, M.; Folgado, V.; Kumar, A.; Blanco, J.; Rimesso, G.; Chuen, J.; Liu, Y.; Zheng, D.; Baker, N.E. Drosophila RpS12 controls translation, growth, and cell competition through Xrp1. *PLoS Genet.* **2019**, *15*, e1008513. [CrossRef]
54. Jin, S.L.C.; Richard, F.J.; Kuo, W.-P.; D'Ercole, A.J.; Conti, M. Impaired growth and fertility of cAMP-specific phosphodiesterase PDE4D-deficient mice. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11998–12003. [CrossRef] [PubMed]
55. Sancar, G.; Liu, S.; Gasser, E.; Alvarez, J.G.; Moutos, C.; Kim, K.; van Zutphen, T.; Wang, Y.; Huddy, T.F.; Ross, B.; et al. FGF1 and insulin control lipolysis by convergent pathways. *Cell Metab.* **2022**, *34*, 171–183.e6. [CrossRef]
56. Jiang, L.; Li, J.; Liu, X.; Burnstock, G.; Xiang, Z. Expression of aquaporin-4 water channels in the digestive tract of the guinea pig. *J. Mol. Histol.* **2014**, *45*, 229–241. [CrossRef] [PubMed]
57. Wang, K.S.; Komar, A.R.; Ma, T.; Filiz, F.; McLeroy, J.; Hoda, K.; Verkman, A.S.; Bastidas, J.A. Gastric acid secretion in aquaporin-4 knockout mice. *Am. J. Physiol. Gastrointest. Liver Physiol.* **2000**, *279*, G448–G453. [CrossRef] [PubMed]
58. Smirnova, E.; Goldberg, E.B.; Makarova, K.S.; Lin, L.; Brown, W.J.; Jackson, C.L. ATGL has a key role in lipid droplet/adiposome degradation in mammalian cells. *EMBO Rep.* **2006**, *7*, 106–113. [CrossRef]
59. Crespo-Piazuelo, D.; Criado-Mesas, L.; Revilla, M.; Castelló, A.; Noguera, J.L.; Fernández, A.I.; Ballester, M.; Folch, J.M. Identification of strong candidate genes for backfat and intramuscular fatty acid composition in three crosses based on the Iberian pig. *Sci. Rep.* **2020**, *10*, 13962. [CrossRef]
60. Yip, L.; Fathman, C.G. Type 1 diabetes in mice and men: Gene expression profiling to investigate disease pathogenesis. *Immunol. Res.* **2014**, *58*, 340–350. [CrossRef]
61. Lukić, B.; Ferenčaković, M.; Šalamon, D.; Čačić, M.; Orehovački, V.; Iacolina, L.; Curik, I.; Cubric-Curik, V. Conservation Genomic Analysis of the Croatian Indigenous Black Slavonian and Turopolje Pig Breeds. *Front. Genet.* **2020**, *11*, 261. [CrossRef]
62. Nakamura, Y.; Kanemarum, K.; Fukami, K. Physiological functions of phospholipase Cδ1 and phospholipase Cδ3. *Adv. Biol. Regul.* **2013**, *53*, 356–362. [CrossRef]
63. Qiao, R.; Zhang, M.; Zhang, B.; Li, X.; Han, X.; Wang, K.; Li, X.; Yang, F.; Hu, P. Population genetic structure analysis and identification of backfat thickness loci of Chinese synthetic Yunan pigs. *Front. Genet.* **2022**, *13*, 1039838. [CrossRef]

64. Taye, M.; Kim, J.; Yoon, S.H.; Lee, W.; Hanotte, O.; Dessie, T.; Kemp, S.; Mwai, O.A.; Caetano-Anolles, K.; Cho, S.; et al. Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher quality beef. *BMC Genet.* **2017**, *18*, 11. [CrossRef] [PubMed]
65. Li-Gao, R.; Hughes, D.A.; van Klinken, J.B.; de Mutsert, R.; Rosendaal, F.R.; Mook-Kanamori, D.O.; Timpson, N.J.; Willems van Dijk, K. Genetic Studies of Metabolomics Change After a Liquid Meal Illuminate Novel Pathways for Glucose and Lipid Metabolism. *Diabetes* **2021**, *70*, 2932–2946. [CrossRef] [PubMed]
66. Harder, M.N.; Ribel-Madsen, R.; Justesen, J.M.; Sparsø, T.; Andersson, E.A.; Grarup, N.; Jørgensen, T.; Linneberg, A.; Hansen, T.; Pedersen, O. Type 2 Diabetes Risk Alleles Near BCAR1 and in ANK1 Associate With Decreased β-Cell Function Whereas Risk Alleles Near ANKRD55 and GRB14 Associate With Decreased Insulin Sensitivity in the Danish Inter99 Cohort. *J. Clin. Endocrinol. Metab.* **2013**, *98*, E801–E806. [CrossRef] [PubMed]
67. Taneera, J.; Mohammed, A.K.; Dhaiban, S.; Hamad, M.; Prasad, R.B.; Sulaiman, N.; Salehi, A. RORB and RORC associate with human islet dysfunction and inhibit insulin secretion in INS-1 cells. *Islets* **2019**, *11*, 10–20. [CrossRef] [PubMed]
68. Roforth, M.M.; Liu, G.; Khosla, S.; Monroe, D.G. Examination of nuclear receptor expression in osteoblasts reveals Rorβ as an important regulator of osteogenesis. *J. Bone Miner. Res.* **2012**, *27*, 891–901. [CrossRef]
69. Hovenier, R.; Kanis, E.; van Asseldonk, T.; Westerink, N.G. Genetic parameters of pig meat quality traits in a halothane negative population. *Livest. Prod. Sci.* **1992**, *32*, 309–321. [CrossRef]
70. Suzuki, K.; Kadowaki, H.; Shibata, T.; Uchida, H.; Nishida, A. Selection for daily gain, loin-eye area, backfat thickness and intramuscular fat based on desired gains over seven generations of Duroc pigs. *Livest. Prod. Sci.* **2005**, *97*, 193–202. [CrossRef]
71. Suzuki, K.; Irie, M.; Kadowaki, H.; Shibata, T.; Kumagai, M.; Nishida, A. Genetic parameter estimates of meat quality traits in Duroc pigs selected for average daily gain, longissimus muscle area, backfat thickness, and intramuscular fat content. *J. Anim. Sci.* **2005**, *83*, 2058–2065. [CrossRef]
72. Chen, Q.; Zeng, Y.; Wang, H.; Yang, L.; Yang, Y.; Zhu, H.; Shi, Y.; Chen, W.; Hu, Y. Molecular characterization and expression analysis of NDUFS4 gene in m. longissimus dorsi of Laiwu pig (Sus scrofa). *Mol. Biol. Rep.* **2013**, *40*, 1599–1608. [CrossRef]
73. Grapes, L.; Rothschild, M.F. Investigation of a QTL region for loin eye area and fatness on pig Chromosome 1. *Mamm. Genome* **2006**, *17*, 657–668. [CrossRef]
74. Schneider, J.F.; Rempel, L.A.; Snelling, W.M.; Wiedmann, R.T.; Nonneman, D.J.; Rohrer, G.A. Genome-wide association study of swine farrowing traits. Part II: Bayesian analysis of marker data. *J. Anim. Sci.* **2012**, *90*, 3360–3367. [CrossRef] [PubMed]
75. Wang, X.; Wang, D.; Chen, X.; Hu, M.; Wang, J.a.; Li, Y.; Guo, N.; Shen, B. cDNA cloning and function analysis of two novel erythroid differentiation related genes. *Sci. China Ser. C Life Sci.* **2001**, *44*, 99–105. [CrossRef] [PubMed]
76. Alli, Z.; Ackerley, C.; Chen, Y.; Al-Saud, B.; Abdelhaleem, M. Nuclear and mitochondrial localization of the putative RNA helicase DHX32. *Exp. Mol. Pathol.* **2006**, *81*, 245–248. [CrossRef] [PubMed]
77. Kroll, K.L. Geminin in embryonic development: Coordinating transcription and the cell cycle during differentiation. *FBL* **2007**, *12*, 1395–1409. [CrossRef]
78. Oatley, J.M.; Brinster, R.L. Regulation of spermatogonial stem cell self-renewal in mammals. *Annu. Rev. Cell Dev. Biol.* **2008**, *24*, 263–286. [CrossRef]
79. Stephens, M.; Balding, D.J. Bayesian statistical methods for genetic association studies. *Nat. Rev. Genet.* **2009**, *10*, 681–690. [CrossRef]
80. Habier, D.; Fernando, R.L.; Kizilkaya, K.; Garrick, D.J. Extension of the bayesian alphabet for genomic selection. *BMC Bioinform.* **2011**, *12*, 186. [CrossRef]
81. Alpay, F.; Zare, Y.; Kamalludin, M.H.; Huang, X.; Shi, X.; Shook, G.; Collins, M.; Kirkpatrick, B. Genome-Wide Association Study of Susceptibility to Infection by Mycobacterium avium Subspecies paratuberculosis in Holstein Cattle. *PLoS ONE* **2014**, *9*, e111704. [CrossRef]
82. Guo, X.; Su, G.; Christensen, O.F.; Janss, L.; Lund, M.S. Genome-wide association analyses using a Bayesian approach for litter size and piglet mortality in Danish Landrace and Yorkshire pigs. *BMC Genom.* **2016**, *17*, 468. [CrossRef]
83. Zare, Y.; Shook, G.E.; Collins, M.T.; Kirkpatrick, B.W. Genome-Wide Association Analysis and Genomic Prediction of Mycobacterium avium Subspecies paratuberculosis Infection in US Jersey Cattle. *PLoS ONE* **2014**, *9*, e88380. [CrossRef]
84. Huang, Y.; Hickey, J.M.; Cleveland, M.A.; Maltecca, C. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. *Genet. Sel. Evol.* **2012**, *44*, 25. [CrossRef] [PubMed]
85. Esfandyari, H.; Sørensen, A.C.; Bijma, P. A crossbred reference population can improve the response to genomic selection for crossbred performance. *Genet. Sel. Evol.* **2015**, *47*, 76. [CrossRef] [PubMed]
86. Lee, S.H.; Clark, S.; van der Werf, J.H.J. Estimation of genomic prediction accuracy from reference populations with varying degrees of relationship. *PLoS ONE* **2017**, *12*, e0189775. [CrossRef] [PubMed]
87. Pandey, A.; Fernandez, M.M.; Steen, H.; Blagoev, B.; Nielsen, M.M.; Roche, S.; Mann, M.; Lodish, H.F. Identification of a Novel Immunoreceptor Tyrosine-based Activation Motif-containing Molecule, STAM2, by Mass Spectrometry and Its Involvement in Growth Factor and Cytokine Receptor Signaling Pathways. *J. Biol. Chem.* **2000**, *275*, 38633–38639. [CrossRef] [PubMed]
88. Yang, Y.Q.; Hui, Y.T.; Liu, R.Y.; Jiao, R.G.; Gong, Y.; Zhang, Y.Y.; Xie, H.Q. Molecular cloning, polymorphisms, and association analysis of the promoter region of the STAM2 gene in Wuchuan Black cattle. *Genet. Mol. Res.* **2013**, *12*, 3651–3661. [CrossRef]
89. Mead, T.J.; Martin, D.R.; Wang, L.W.; Cain, S.A.; Gulec, C.; Cahill, E.; Mauch, J.; Reinhardt, D.P.; Lo, C.W.; Baldock, C.; et al. Proteolysis of fibrillin-2 microfibrils is essential for normal skeletal development. *bioRxiv* **2021**. [CrossRef] [PubMed]

90. Kearns, A.E.; Donohue, M.M.; Sanyal, B.; Demay, M.B. Cloning and Characterization of a Novel Protein Kinase That Impairs Osteoblast Differentiation in Vitro. *J. Biol. Chem.* **2001**, *276*, 42213–42218. [CrossRef]
91. Skiba, G.; Sobol, M.; Raj, S. Bone mineralization, geometry and strength in pigs growing from 56 to 115 day of life as affected by body fatness. *J. Anim. Feed Sci.* **2016**, *25*, 302–308. [CrossRef]
92. Pérez-Enciso, M.; Forneris, N.; de los Campos, G.; Legarra, A. Evaluating Sequence-Based Genomic Prediction with an Efficient New Simulator. *Genetics* **2017**, *205*, 939–953. [CrossRef]
93. Song, H.; Ye, S.; Jiang, Y.; Zhang, Z.; Zhang, Q.; Ding, X. Using imputation-based whole-genome sequencing data to improve the accuracy of genomic prediction for combined populations in pigs. *Genet. Sel. Evol.* **2019**, *51*, 58. [CrossRef]
94. Calus, M.P.L.; Bouwman, A.C.; Schrooten, C.; Veerkamp, R.F. Efficient genomic prediction based on whole-genome sequence data using split-and-merge Bayesian variable selection. *Genet. Sel. Evol.* **2016**, *48*, 49. [CrossRef] [PubMed]
95. van den Berg, I.; Boichard, D.; Guldbrandtsen, B.; Lund, M.S. Using Sequence Variants in Linkage Disequilibrium with Causative Mutations to Improve Across-Breed Prediction in Dairy Cattle: A Simulation Study. *G3 Genes | Genomes | Genet.* **2016**, *6*, 2553–2561. [CrossRef] [PubMed]
96. Veerkamp, R.F.; Bouwman, A.C.; Schrooten, C.; Calus, M.P.L. Genomic prediction using preselected DNA variants from a GWAS with whole-genome sequence data in Holstein–Friesian cattle. *Genet. Sel. Evol.* **2016**, *48*, 95. [CrossRef] [PubMed]
97. Ye, S.; Gao, N.; Zheng, R.; Chen, Z.; Teng, J.; Yuan, X.; Zhang, H.; Chen, Z.; Zhang, X.; Li, J.; et al. Strategies for Obtaining and Pruning Imputed Whole-Genome Sequence Data for Genomic Prediction. *Front. Genet.* **2019**, *10*, 673. [CrossRef] [PubMed]
98. Ober, U.; Ayroles, J.F.; Stone, E.A.; Richards, S.; Zhu, D.; Gibbs, R.A.; Stricker, C.; Gianola, D.; Schlather, M.; Mackay, T.F.C.; et al. Using Whole-Genome Sequence Data to Predict Quantitative Trait Phenotypes in Drosophila melanogaster. *PLoS Genet.* **2012**, *8*, e1002685. [CrossRef]

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