

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

²Department of Animal and Veterinary Sciences, The Roslin Institute Building, Scotland's Rural College (SRUC), Easter Bush, Midlothian EH25 9RG, UK; ³Animal Genomics and Improvement Laboratory, Henry A. Wallace Beltsville Agricultural Research Center, Agricultural Research Service (ARS), U.S. Department of Agriculture (USDA), Beltsville, Maryland 20705, USA;

⁴Department of Animal and Avian Sciences, University of Maryland, College Park, Maryland 20742, USA;

(Continued on next page)

MOLECULAR BIOLOGY & GENETICS

Integrating large-scale meta-GWAS and PigGTEx resources to decipher the genetic basis of 232 complex traits in pigs

Zhiting Xu ^{1,†}, Qing Lin ^{1,†}, Xiaodian Cai^{1,†}, Zhanming Zhong^{1,†}, Jinyan Teng^{1,†}, Bingjie Li ², Haonan Zeng¹, Yahui Gao^{1,3,4}, Zexi Cai ⁵, Xiaoqing Wang⁶, Liangyu Shi⁶, Xue Wang⁷, Yi Wang⁷, Zipeng Zhang⁷, Yu Lin⁸, Shuli Liu ⁹, Hongwei Yin¹⁰, Zhonghao Bai⁵, Chen Wei¹, Jun Zhou¹, Wenjing Zhang¹, Xiaoke Zhang ¹, Shaolei Shi¹, Jun Wu ¹, Shuqi Diao¹, Yuqiang Liu¹, Xiangchun Pan¹, Xueyan Feng¹, Ruiqi Liu¹, Zhanqin Su¹, Chengjie Chang¹, Qianghui Zhu¹, Yuwei Wu¹, The PigGTEx Consortium, Zhongyin Zhou¹¹, Lijing Bai¹⁰, Kui Li¹⁰, Qishan Wang ¹², Yuchun Pan ¹², Zhong Xu¹³, Xianwen Peng¹³, Shuqi Mei¹³, Delin Mo ¹⁴, Xiaohong Liu¹⁴, Hao Zhang¹, Xiaolong Yuan ¹, Yang Liu¹⁵, George E. Liu³, Guosheng Su⁵, Goutam Sahana⁵, Mogens Sandø Lund⁵, Li Ma⁴, Ruidong Xiang^{16,17}, Xia Shen ^{18,19,20}, Pinghua Li^{21,22}, Ruihua Huang^{21,22}, Maria Ballester²³, Daniel Crespo-Piazuelo²³, Marcel Amills^{24,25}, Alex Clop ²⁴, Peter Karlskov-Mortensen²⁶, Merete Fredholm²⁶, Guoqing Tang⁸, Mingzhou Li ⁸, Xuewei Li⁸, Xiangdong Ding⁷, Jiaqi Li¹, Yaosheng Chen ¹⁴, Qin Zhang ²⁷, Yunxiang Zhao^{28,*}, Fuping Zhao ^{6,*}, Lingzhao Fang^{5,*} and Zhe Zhang ^{1,*}

ABSTRACT

Understanding the molecular and cellular mechanisms underlying complex traits in pigs is crucial for enhancing genetic gain via artificial selection and utilizing pigs as models for human disease and biology. Here, we conducted comprehensive genome-wide association studies (GWAS) followed by a cross-breed meta-analysis for 232 complex traits and a within-breed meta-analysis for 12 traits, using 28.3 million imputed sequence variants in 70 328 animals across 14 pig breeds. We identified 6878 quantitative trait loci (QTL) for 139 complex traits. Leveraging the Pig Genotype-Tissue Expression resource, we systematically investigated the biological context and regulatory mechanisms behind these trait-QTLs, ultimately prioritizing 14 829 variant-gene-tissue-trait regulatory circuits. For instance, rs344053754 regulates *UGT2B31* expression in the liver and intestines, potentially by modulating enhancer activity, ultimately influencing litter weight at weaning in pigs. Furthermore, we observed conservation of certain genetic and regulatory mechanisms underlying complex traits between humans and pigs. Overall, our cross-breed meta-GWAS in pigs provides invaluable resources and novel insights into the genetic regulatory and evolutionary mechanisms of complex traits in mammals.

Keywords: pig, meta-GWAS, PigGTEx, molQTL, complex traits

*Corresponding authors. E-mails: zhezhang@scau.edu.cn; lingzhao.fang@qqg.au.dk; zhaofuping@caas.cn; yunxiangzhao@126.com

[†]Equally contributed to this work.

Received 10 July 2024; Revised 13 January 2025; Accepted 26 January 2025

INTRODUCTION

Pigs are one of the most important sources of animal protein worldwide, with an global pork production reached 124.5 million tons in 2023 [1]. In the past decade, genomic selection has been extensively employed to improve pig breeding [2]. Recently, genome editing has been increasingly applied to protect pigs against porcine reproductive and respiratory syndrome virus [3] and transmissible gastroenteritis virus [4]. A deeper comprehension of the genetic underpinnings of complex traits in pigs will further enable us to genetically maximize their production efficiency [5] and improve their health and welfare [6,7], while minimizing environmental impacts [8,9] through developing more advanced precision breeding techniques. Besides serving as a primary source of animal protein for humans [10], pigs are widely accepted as a valuable model for studying human biology and diseases, including Alzheimer's disease [11], cardiovascular disease [12], wound healing [13,14], human reproduction [15], human gastrointestinal tract [16], dry eye [17] and immunological studies [18–20]. Therefore, investigating the genetic and biological architecture of complex traits and diseases in pigs is crucial for both agricultural advancements and biomedical research.

A genome-wide association study (GWAS) is a prevalent approach for dissecting the genetic basis of complex traits and diseases [21–23]. As of 28 April 2024, the Pig quantitative trait loci (QTL) database (QTLdb) has gathered 48 875 QTL for 673 distinct traits [24]. However, the causal variants and genes underlying most of these QTL regions remain unknown due to extensive linkage disequilibrium (LD) among genetic variants and the lack of functional annotations [25]. Cross-ancestry meta-analysis of GWAS (meta-GWAS) has emerged as an effective strategy to uncover trait-associated variants shared across populations and accelerate statistical fine mapping of causal variants and genes through reducing LD [23,26–28]. In addition, the majority of genetic variants identified in GWAS are located in non-coding regions [29] and are notably enriched in *cis*-regulatory elements such as promoters and enhancers [30,31], as well as in gene expression QTL (eQTL) within relevant tissues [32]. Therefore, it is of great interest to prioritize causal variants, genes, pathways and tissues of complex traits in livestock through systematically integrating GWAS results with relevant functional annotation data, including those from the Functional Annotation of ANimal Genomes (FAANG) project [33] and the Farm Animal Genotype-Tissue Expression (FarmGTEx) project [34].

Here, we collected and analyzed phenotypes and genotypes from 70 328 pigs across 59 pop-

ulations, representing 14 distinct pig breeds, to identify genetic variants underlying various complex traits in pigs. After imputing genotypes to the sequence level using a multi-breed genotype imputation reference panel [34], we conducted a comprehensive cross-breed meta-GWAS for 232 complex traits and a within-breed meta-GWAS for 12 traits. Subsequently, we integrated regulatory elements from the FAANG project [33] and molecular QTL (molQTL) from the Pig Genotype-Tissue Expression (PigGTEx) project [34] to systematically elucidate the molecular and cellular mechanisms underlying complex traits in pigs. To explore the potential of pigs as models for human complex traits and diseases, we compared the genetic regulation of 136 human complex phenotypes with our 126 pig meta-GWAS, focusing on those with significant GWAS loci and sample sizes greater than 1000. All the resources are freely available in the PigBiobank web portal (<https://pigbiobank.farmgtex.org>) (Fig. 1a).

RESULTS

Summary of genotypes and phenotypes

After excluding ancestral outliers within each of the 59 populations based on population structure analysis (for details see Methods), we retained 69 242 out of 70 328 pigs genotyped by single-nucleotide polymorphism (SNP) arrays (averaging 45 418 autosomal SNPs) or low-coverage whole-genome sequencing (WGS) (averaging 198 178 autosomal SNPs) for subsequent analyses. This cohort included 20 706 Duroc, 9159 Landrace, 34 540 Yorkshire and 4837 animals from 11 other pig breeds (Table S1). We imputed genotypes for all 69 242 animals to the sequence level using the multi-breed Pig Genomics Reference Panel (PGRP version 1) from the PigGTEx project [34] (Fig. 1b and c). We kept 28 297 602 imputed SNPs with dosage R-squared (DR^2) > 0.8 and minor allele frequency (MAF) $\geq$ 0.01 for subsequent analyses (Fig. S1a–e). Through 20 cross-validations, we estimated average concordance rates and Pearson's correlations between imputed and observed genotypes across 14 breeds to be 96.67% and 0.939, respectively (Fig. 1d, Fig. S1f–h, Tables S1–S3). The population structure of GWAS samples, estimated from imputed genotypes, closely matched that from raw genotypes (Pearson's correlation = 0.996), and the distribution of imputed SNPs aligned well with all the SNPs in the genotype reference panel across diverse genomic features (Fig. S1i–l). Altogether, these results supported the reliability of our imputed genotype data.

We collected 271 continuous and 15 binary traits across 59 populations, representing 14 pig breeds worldwide (Fig. S2a). After excluding

(Continued from previous page)

⁵Center for Quantitative Genetics and Genomics (QGG), Aarhus University, Aarhus 8000, Denmark; ⁶Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China; ⁷College of Animal Science and Technology, China Agricultural University, Beijing 100193, China; ⁸Key Laboratory of Agricultural Bioinformatics, Ministry of Education, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu 611130, China; ⁹Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou 310024, China; ¹⁰Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124, China; ¹¹State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; ¹²Department of Animal Science, College of Animal Sciences, Zhejiang University, Hangzhou 310058, China;

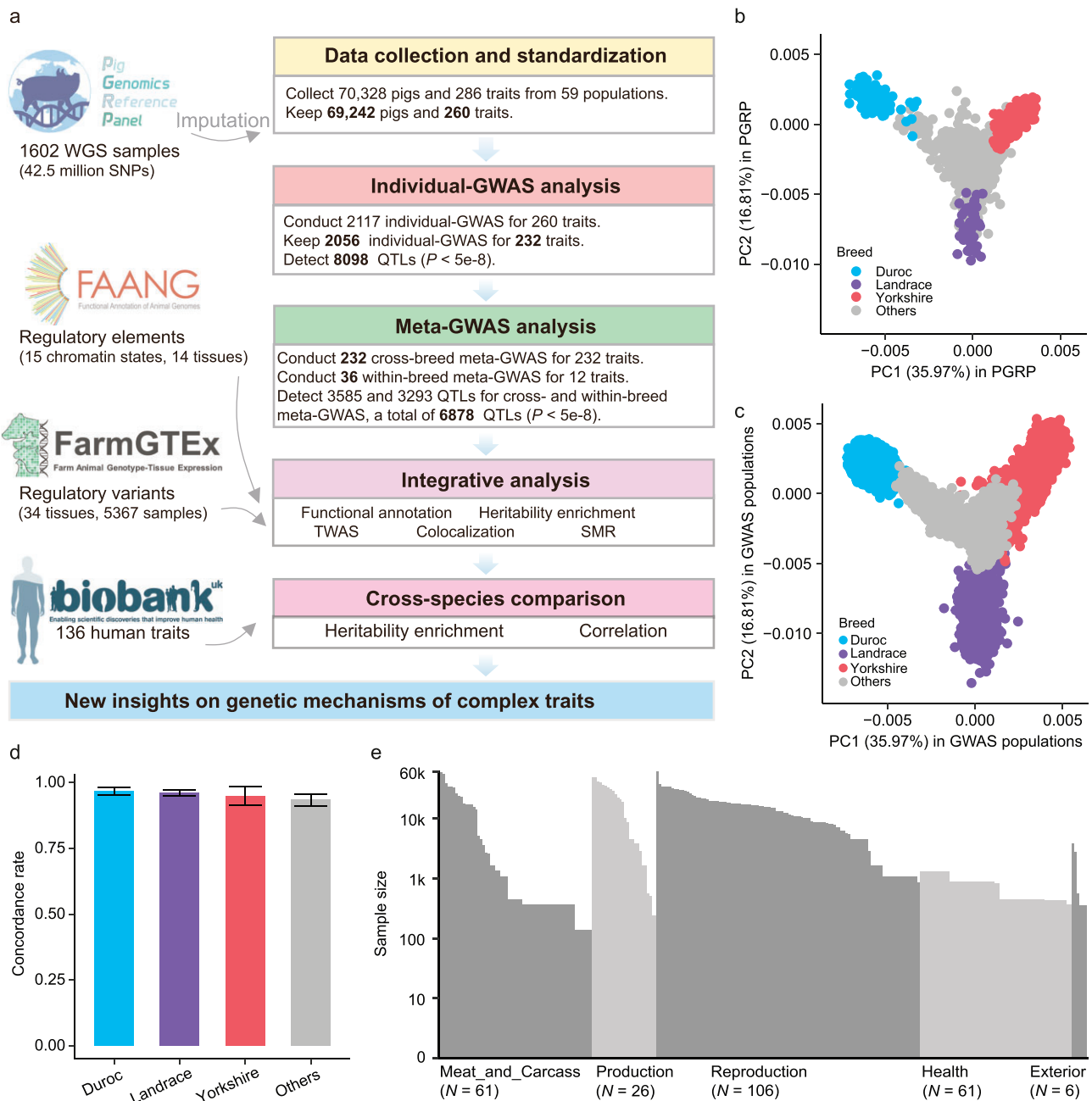


Figure 1. Overall study design and summary of genotypes and phenotypes. (a) Overview of study design. WGS: whole genome sequence. GWAS: genome-wide association study. TWAS: transcriptome-wide association study. SMR: summary data-based Mendelian randomization. (b and c) Principal component (PC) analysis of Pig Genomics Reference Panel (PGRP) (b) and GWAS populations (c), which were conducted based on 57 600 individuals (samples with raw genotype data) and a total of 1603 shared array SNPs using PLINK (v1.90) [84]. The first two PCs were plotted using the *geom_point* function from the ggplot2 (v3.3.6) package in R (v4.1.2). (d) The imputation accuracy (concordance rate between the imputed and observed genotypes) of PGRP in GWAS data. This was evaluated on chromosome 6 with 20 times 5-fold cross-validation using Chip array data from 60 720 pigs in 53 GWAS populations that had individual-level genotype data. Error bars indicate the standard deviation. (e) The total sample size was collected for each trait. Traits were classified into five main trait-categories.

low-quality phenotypes (Methods), we retained 249 continuous traits and 11 binary traits for subsequent analyses (Table S4). These 260 traits represent **5 main trait-categories** (with 17 sub trait-

categories): **Production** ($n = 57\,612$), comprising Feed intake ($n = 240$), Growth ($n = 57\,612$) and Feed conversion ($n = 19\,095$); **Meat and Carcass** ($n = 65\,883$), comprising Fatness ($n = 60\,203$),

Anatomy ($n = 52\,470$), Chemistry ($n = 368$), Fatty acid content ($n = 368$), Texture ($n = 140$), Meat color ($n = 140$) and pH ($n = 140$); **Health** ($n = 2139$), including Immune capacity ($n = 1317$) and Blood parameters ($n = 2139$); **Reproduction** ($n = 49\,018$), including Reproductive traits ($n = 30\,207$), Litter traits ($n = 41\,907$) and Reproductive organs ($n = 40\,049$); and **Exterior** ($n = 6625$), including Behavioral ($n = 2797$) and Conformation ($n = 3828$) (Fig. 1e). The average heritability across all traits was 0.27, varying from 0.09 for the number of mummified pigs to 0.91 for lysozyme level (Table S4, Fig. S2b and c). Notably, **Reproduction** traits exhibited a lower heritability (an average of 0.18) compared to **Health** traits (average of 0.50) and **Meat and Carcass** traits (average of 0.32). This aligns with previous studies indicating low heritability for reproductive traits [35–39] and medium to high heritability for health traits [40–45].

Sharing patterns of QTL between populations revealed by meta-GWAS

We conducted a total of 2117 individual GWAS for the 260 traits across 59 pig populations (Table S4). To ensure the quality and reliability of these individual GWAS results for subsequent meta-analyses, we implemented stringent quality control measures across four different strategies: SE-N (i.e., inverse of the median standard error versus the square root of the sample size) plot, P-Z (i.e., for each SNP, compares the reported P-values with the P-values computed from the Z-statistics based on reported beta-estimate and standard error) plot, effect allele frequencies (EAF) plot and λ_{GC} (Methods) (Fig. S2d–j). This process yielded 2056 high-quality summary-level GWAS results for 232 traits (Table S4), with 78 traits across all 5 main trait-categories not previously included in the Pig QTLdb (release 46) [24] (<https://www.animalgenome.org/cgi-bin/QTLdb/SS/index>) (Table S5). Generally, Pearson's correlations of SNP effects on the same traits across populations were significantly higher than those on different traits within populations (Fig. S3a and b). We identified 8098 QTL (i.e. 8098 significant trait-associations) ($P < 5 \times 10^{-8}$), representing 7011 distinct lead SNPs. Among them, 45.66% were detected solely in one population (Methods) (Fig. S3c), where the MAF was higher in the respective populations compared to others (Fig. S3d). For instance, rs323720776, associated with average daily gain (birth–100 kg), was only identified in one Yorkshire population ($n = 4383$), with a MAF of 0.46 versus an average MAF of 0.28 across all 20 studied populations ($n = 19\,824$) (Fig. S3e). The findings suggest that differences

in the MAF of genetic variants among populations could contribute to population-specific associations.

To enhance the detection of QTL and identify shared associations across populations with small effects often overlooked by individual GWAS [46], we conducted cross-breed meta-GWAS for each of the 232 complex traits using the 2056 high-quality GWAS summary statistics (Methods, Table S4). Additionally, we conducted 36 within-breed meta-GWAS analyses for 12 complex traits with large sample sizes within the Duroc, Landrace and Yorkshire breeds (prefixed as 'D_', 'L_' and 'Y_', respectively), to explore potential breed-specific genetic regulatory mechanisms underlying complex traits. In total, we obtained 268 meta-GWAS results, with an average sample size of 10 147 individuals, ranging from 140 for dressing percentage to 58 725 for backfat thickness (BFT) (Table S5). We identified a total of 6878 QTL ($P < 5 \times 10^{-8}$), representing 6233 distinct lead SNPs (Table S6, Fig. 2a). These lead SNPs were distributed across all 18 autosomes (Fig. 2b) and exhibited a lower MAF compared to random background SNPs in the respective populations (Fig. S4a). The number of QTL detected in both cross-breed and within-breed meta-GWAS showed a positive correlation with sample size (Pearson's $r = 0.69$, $P = 1.36 \times 10^{-25}$) and trait heritability (Pearson's $r = 0.49$, $P = 9.59 \times 10^{-3}$) (Fig. 2c and d), aligning with human genetic studies [47,48].

To validate the detected QTL, we implemented three strategies. First, we conducted cross-breed and within-breed meta-GWAS for average daily gain (ADG) across 9 independent populations [49], including 42 790 Duroc pigs, 88 984 Landrace pigs and 69 606 Yorkshire pigs. The association signals identified in these validation populations exhibited significant enrichment (mean fold = 6.84, $P < 1 \times 10^{-300}$) and replicated an average of 48.49% of the QTL identified in our study (Fig. S4b). Second, we utilized 447 and 379 lead SNPs of suggestive significance ($P < 1 \times 10^{-5}$) to estimate polygenic scores for the number born alive and teat number, respectively, across 7 pig breeds. Notably, Jiaxinghei, Erhualian and Meishan pigs displayed higher estimated polygenic scores compared to Landrace and Yorkshire (Fig. 2e, Fig. S4c), in line with previous findings that Jiaxinghei, Erhualian and Meishan pigs have greater prolificacy than Landrace and Yorkshire pigs [50–52]. Third, we used lead SNPs ($P < 5 \times 10^{-8}$) identified from within-breed and cross-breed meta-GWAS of ADG and BFT to perform genomic prediction in 3 independent populations including 93 Duroc, 1510 Landrace and 2844 Yorkshire pigs. We found that the prediction accuracy of the model using lead SNPs from within-breed and cross-breed meta-GWAS was higher than that

(Continued from previous page)

¹³Hubei Key Laboratory of Animal Embryo and Molecular Breeding, Institute of Animal Husbandry and Veterinary, Hubei Provincial Academy of Agricultural Sciences, Wuhan 430064, China; ¹⁴State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China; ¹⁵College of Animal Science and Technology, Nanjing Agricultural University, Nanjing 210095, China; ¹⁶Faculty of Veterinary & Agricultural Science, University of Melbourne, Parkville, VIC 3010, Australia; ¹⁷Agriculture Victoria Research, AgriBio Centre for AgriBiosciences, Bundoora, VIC 3083, Australia; ¹⁸State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai 200438, China; ¹⁹Center for Intelligent Medicine Research, Greater Bay Area Institute of Precision Medicine (Guangzhou), Fudan University, Guangzhou 510000, China; ²⁰Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh EH16 4UX, UK; ²¹Institute of Swine Science, Nanjing Agricultural University, Nanjing 210095, China;

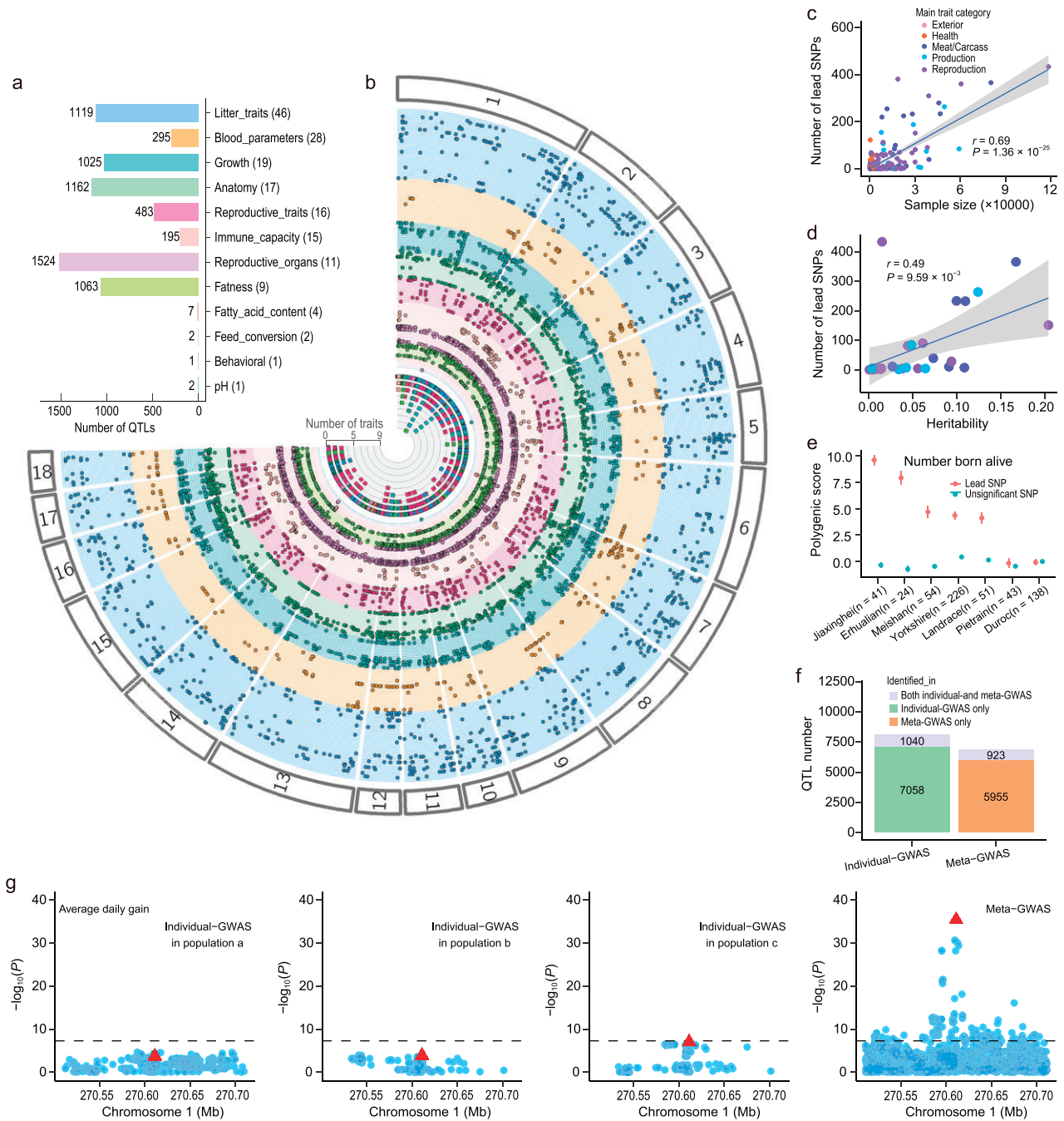


Figure 2. Summary and validation of quantitative trait loci (QTL) for pig complex traits. (a) The total number of QTL detected in 12 sub-trait-categories. The number of traits included in each trait-category is shown in parentheses. (b) A Fuji-plot summarizes the 6878 lead SNPs ($P < 5 \times 10^{-8}$) identified in 169 within-breed and cross-breed meta-GWAS analyses. The plot was created using the Fuji-plot script developed by Kanai *et al.* [85]. The innermost ring (ring 1) indicates the number of traits associated with each lead SNP. Rings 2–170 indicate the 169 meta-GWAS. The order of traits is shown in Table S5 (starting with the innermost ring). The points indicate the genomic position of the 6878 lead SNPs in chromosomes 1–18. The colors represent the trait-category, consistent with (a). (c) Pearson correlation between sample size and the number of lead SNPs in 169 meta-GWAS. (d) Pearson correlation between heritability and the number of lead SNPs in 27 meta-GWAS (sample size > 15000). Heritability was estimated using linkage disequilibrium score regression (LDSC) [86]. The Pearson correlation coefficient in (c) and (d) was calculated by the *cor.test* function in R. (e) Results of genomic predictions for individuals from several pig breeds in the PGRP with large phenotype differences, based on a linear mixed model and effect information from suggestive lead variants ($P < 1 \times 10^{-5}$) and the same number of random non-significant variants in the total number born alive. The x-axis labels indicate the different pig breeds. The y-axis labels indicate the genomic estimated breeding values (GEBVs). The error bars are the standard errors of GEBVs. (f) The number of different categories of QTL detected in individual GWAS and meta-GWAS. Both individual- and meta-GWAS QTL refer to the QTL that exhibit physical position overlap in both individual GWAS and meta-GWAS for the same trait. (g) rs320375241 associated with average daily gain (ADG) in individual GWAS and cross-breed meta-GWAS, respectively. (a), (b) and (c) were three random populations for ADG.

of the models using lead SNPs (and random SNPs) only from within-breed meta-GWAS (Fig. S4d and e). QTL with higher significance showed a greater overlap with those reported in Pig QTLdb (release 46) [24] (Fig. S4f). These results together underscored the reliability of the detected QTL and indicated a substantial degree of sharing across breeds.

In comparison to individual GWAS results, 5955 out of the 6878 QTL identified in cross-breed and within-breed meta-GWAS were novel across 147 traits (Fig. 2f). For ADG, rs320375241 exhibited non-significance in any individual GWAS but demonstrated significance in cross-breed meta-GWAS (Fig. 2g). In contrast, 7058 out of the 8098 QTL identified in individual GWAS were not replicated in either cross-breed or within-breed meta-GWAS (QTL_{individual_GWAS_only}) (Fig. 2f). Lead SNPs of these QTL_{individual_GWAS_only} exhibited a higher proportion of effects in the opposite direction across populations compared to those detected in both individual GWAS and meta-GWAS (Fig. S4g). Our findings suggested that QTL identified in cross-breed and within-breed meta-analyses were more likely to represent shared genetic associations underlying complex traits across populations, in contrast to QTL identified in individual GWAS. Therefore, in subsequent analyses, we focused solely on the 6878 QTL identified in cross-breed and within-breed meta-GWAS to elucidate the genetic regulatory mechanisms underlying complex traits.

Breed specificity and pleiotropy of QTL

To explore the breed-specific regulatory effects of genetic variants on complex traits, we first conducted within-breed meta-GWAS analyses for Duroc, Landrace and Yorkshire, respectively, resulting in 3293 QTL in at least one breed across 12 complex traits (Fig. 3a). Of these QTL, 89% exhibited breed-specific effects (see Methods) (Fig. 3a). For instance, we identified 68, 192 and 211 QTL with breed-specific effects for BFT in Duroc, Landrace and Yorkshire, respectively (Fig. 3b). The observation of breed-specific QTL is likely to be a result of the differences in allele frequencies or sample sizes among the three breeds (Fig. 3c and d, Fig. S5) [27]. Additionally, we observed a negative correlation between the effect sizes of lead SNPs and their MAFs across different breeds (Fig. S6a–l). To further investigate the potential regulatory mechanism of breed-specific QTL, we identified genes with specific high expression levels in each of the breeds (breed-specific expressed genes), based on multiple tissue samples (see Methods, Fig. 3e). Enrichment analyses revealed that breed-specific QTL exhibited a higher enrichment in breed-specific expressed genes compared to non-breed-specific QTL present

in at least two breeds (Fig. 3f, Fig. S6m–q). These findings suggested that breed-specific effects on complex traits may arise from gene regulatory differences, potentially resulting from gene–environment interactions and genetic drift [53].

The cross-breed meta-analysis replicated 28% of the QTL identified in within-breed meta-analyses across traits (Fig. S6r), while the remaining 72% of QTL often exhibited varying directions of effects for lead SNPs among the three breeds (Fig. S6s–u). Taking ADG as an example, the QTL with lead SNP rs320375241 was consistently identified in the within-breed analyses for Landrace and Yorkshire but not for Duroc, as well as in the cross-breed analysis with increased statistical power (Fig. 3g, Table S6) [27]. Furthermore, utilizing the effect information of breed-shared lead SNPs, the 36 within-breed meta-GWAS summary results can be clustered by traits rather than breeds (Fig. 3h).

We further explored the pleiotropic effects of variants across 232 traits based on the cross-breed meta-GWAS results using PLACO [54]. As expected, we observed a higher proportion of shared effects within the same trait-category compared to those shared between different trait-categories (Fig. S7a, Table S5). Additionally, we estimated the posterior probability of lead SNP effects on each trait in 25 complex traits with the largest sample sizes using METASOFT [55]. Our analysis revealed that 8.24% (174 SNPs) of lead SNPs exhibited an effect on at least two complex traits (m -value > 0.9 from METASOFT). For instance, rs336748894 on chromosome 7 was simultaneously associated with Loin muscle area ($P = 7.15 \times 10^{-24}$), Loin muscle depth ($P = 5.87 \times 10^{-18}$), BFT ($P = 3.26 \times 10^{-15}$), Days ($P = 2.23 \times 10^{-6}$) and ADG ($P = 1.59 \times 10^{-7}$) (Fig. S7b). Notably, lead SNP tended to have a smaller effect if it showed effects on more traits (Fig. S7c).

Functional characterization of GWAS loci by integrating multi-omics data

To explore the biological context in which the meta-GWAS loci act and their underlying regulatory circuits, we annotated the GWAS loci from both cross-breed and within-breed meta-GWAS using multi-layered biological data, including the functions of genomic variants and 14 different chromatin states previously annotated by Pan *et al.* [56]. We focused on 5837 unique lead SNPs identified in 176 meta-GWAS with sample sizes >1000. Notably, 99.04% of the 5837 lead SNPs were located in non-coding regions (Fig. 4a and Fig. S8a), with 9.41% and 52.73% in promoters and enhancers, respectively. This aligns with previous human studies (Pearson's $r = 0.91$, $P = 4.81 \times 10^{-11}$) (Fig. 4b).

(Continued from previous page)

²²Key Laboratory in Nanjing for Evaluation and Utilization of Livestock and Poultry (Pigs) Resources, Ministry of Agriculture and Rural Areas, Nanjing 210095, China; ²³Animal Breeding and Genetics Programme, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Torre Marimon, Caldes de Montbui 08140, Spain; ²⁴Department of Animal Genetics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus de la Universitat Autònoma de Barcelona, Bellaterra 08193, Spain; ²⁵Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain; ²⁶Animal Genetics and Breeding, Department of Veterinary and Animal Sciences, University of Copenhagen, Frederiksberg C 1870, Denmark; ²⁷College of Animal Science and Technology, Shandong Agricultural University, Tai'an 271018, China and ²⁸College of Animal Science and Technology, Guangxi University, Nanning 530004, China

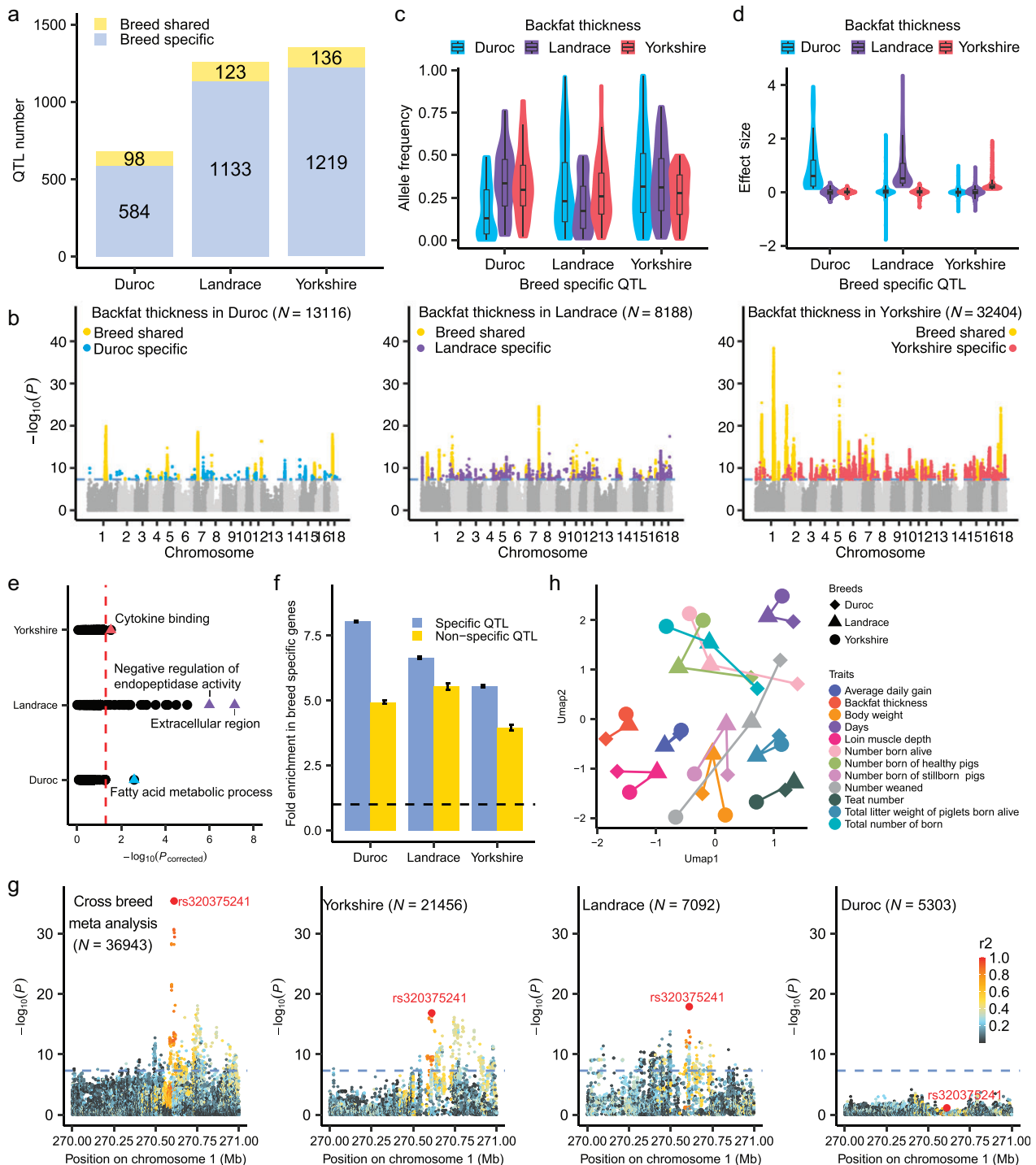


Figure 3. Identified QTL across breeds. (a) The number of QTL from the within-breed meta-analysis for 12 complex traits in Duroc, Landrace and Yorkshire. Breed-specific QTL refers to the QTL that does not exhibit physical overlap with QTL from other breeds for the same trait. Breed-shared QTL refers to the QTL that exhibits physical overlap with QTL from the same trait in at least one breed. (b) Manhattan plots of within-breed meta-analysis for backfat thickness in Duroc (left), Landrace (middle) and Yorkshire (right). The blue ($n = 68$), purple ($n = 192$) and red ($n = 211$) points represent the breed-specific QTL identified in Duroc, Landrace and Yorkshire, respectively. (c and d) The distribution of allele frequencies (c) and effect sizes (d) of lead SNPs of breed-specific QTL for backfat thickness in the three breeds. Boxplots depict the median value as the center, and first and third quartiles as box boundaries. Sample sizes for each breed are provided in Table S5. (e) The Gene Ontology (GO) and Kyoto Encyclopedia of Gene and Genomes (KEGG) enrichment results of breed-specific highly expressed genes in three breeds identified from muscle RNA samples of PigGTEx [34]. The enrichment was conducted using KOBAS [87]. The sample sizes of muscle for detecting breed-specific highly expressed genes in Duroc, Landrace and Yorkshire

Figure 3. (Continued) were 157, 49 and 119, respectively. The x-axis represents the false discovery rate (FDR)-corrected P -value. (f) Enrichment of breed-specific QTL and non-breed-specific QTL for loin muscle depth in the breed-specific highly expressed genes in muscle. The error bar represents the standard error of the enrichment fold. (g) Local Manhattan plots of the cross-breed meta-analysis of ADG in all breeds (left 1), as well as a within-breed meta-analysis of Yorkshire (left 2), Landrace (left 3) and Duroc (left 4) on chromosome 1. The color indicates the magnitude of the linkage disequilibrium (LD) between rs320375241 and other SNPs. (h) Clustering for a 36 within-breed meta-analysis of 12 complex traits based on the Z-score of breed shared lead SNPs (lead SNPs with m -values > 0.9 in all the three breeds). Clustering was conducted using the umap package in R.

Furthermore, lead SNPs exhibited a significant enrichment in regulatory elements, particularly active promoters (TssA) (12.01-fold, $P < 1.0 \times 10^{-30}$) and enhancers (EnhA) (2.02-fold, $P = 1.24 \times 10^{-5}$) (Fig. 4a and Fig. S8b). Notably, the coding sequence (CDS) showed a high enrichment fold (20.16-fold, $P < 1.0 \times 10^{-30}$), despite only 1.01% of lead SNPs mapping to this region. Additionally, compared to non-significant SNPs, lead SNPs showed a greater density near the transcription start sites (TSSs) of protein-coding genes (Fig. S8c) and lower evolutionary constraints across species (lower PhastCons scores, Fig. S8d).

To further investigate the regulatory role of genetic variants on complex traits, we integrated molQTL corresponding to five molecular phenotypes (molPhe, including cis-eQTL for protein-coding genes expression, cis-eeQTL for exon expression, cis-lncQTL for lncRNA expression, cis-enQTL for enhancer expression and cis-sQTL for alternative splicing) from 34 tissues in PigGTEx [34]. We performed summary-based heritability enrichment analyses for these molQTL in cross-breed and within-breed meta-GWAS with sample size exceeding 1000, and detected 296 significantly enriched molQTL-trait pairs across 73 meta-GWAS (false discovery rate, FDR < 0.05) (Fig. S8e, Table S7). In general, at least one type of molQTL exhibited significant enrichment for trait heritability in 14 out of the 25 complex traits with large sample sizes (FDR < 0.05 , mean enrichment fold = 1.37 ± 0.21) (Fig. 4c, Table S7). We further explored the heritability contributions of independent eQTL and sQTL to complex traits and found that they explained a higher proportion of heritability compared to MAF-matched SNPs in muscle and liver tissues (Fig. 4d). These results suggested that regulatory variants contributed to dissecting genetic and molecular architecture underlying complex traits in pigs.

Furthermore, we observed that different trait categories exhibited heritability enrichment across eGenes (genes with at least one significant variant in at least one tissue) with distinct tissue-sharing patterns (Table S8, Fig. 4e). For instance, BFT (fatness of sub trait-category) showed a notable enrichment for tissue-specific eGenes, while ADG (growth of sub trait-category) was enriched in tissue-sharing eGenes (Fig. 4e). The lead SNP (rs1108824455) of BFT was identified as an eQTL that regulates the expression of *LGALS13*, a gene with specific expression

in a few tissues, including the duodenum, lung and adipose (Fig. 4f, Fig. S8f). *LGALS13* has been recognized as a serum biomarker during early pregnancy [57,58]. In contrast, the lead SNP (rs324200444) of ADG was also an eQTL that regulated the expression of *ABCC10*, a gene displaying extensive tissue sharing regarding both gene expression and eQTL (Fig. 4g, Fig. S8g). *ABCC10* is known to act as a genetic marker for pig growth [59].

Detecting trait-relevant tissues

To further investigate the tissue-mediated patterns of genetic regulation of complex traits, we conducted enrichment analyses for meta-GWAS signals within tissue-specific expressed genes across 34 pig tissues (Table S9, Fig. S9). Trait-associated SNPs were significantly enriched in tissue-specific genes of trait-relevant tissues (Fig. 5a and Table S10). For example, both litter weight (weaning) (M_TLWT_Weaning) (10.14-fold, $P = 0.001$) and body weight (M_BW) (6.84-fold, $P = 0.001$) exhibited the highest enrichment in the liver compared to other tissues. Likewise, gestation length exhibited the highest enrichment in the ovary (5.62-fold, $P = 0.001$) (Fig. 5a).

Based on the results of the transcriptome-wide association study (TWAS), summary-based mendelian randomization analysis (SMR) and co-localization analyses [34] conducted on within-breed and cross-breed meta-GWAS with a sample size greater than 1000, we identified 56 003 variant-molPhe-tissue-trait circuits, including 14 829 variant-gene-tissue-trait regulatory circuits (Table S11). For instance, *UGT2B31*, which exhibits the highest expression in the liver compared to other tissues (Fig. 5b), was significantly associated with litter weight (weaning) through gene-based GWAS ($P = 1.48 \times 10^{-9}$) (Fig. 5c) and TWAS ($P = 2.46 \times 10^{-9}$) (Fig. 5d and e). Furthermore, a cis-eQTL (rs344053754) regulated litter weight (weaning) through influencing *UGT2B31* mRNA expression in the liver ($P = 5.62 \times 10^{-7}$, SMR) (Fig. 5f). The SNP rs344053754 is located in active enhancers in the liver and intestines, but not in other tissues (Fig. 5g). *UGT2B31* is an important drug metabolizing enzyme in the liver across various animal species [60–63]. In summary, these analyses provided many testable biological hypotheses of gene regulation underlying complex traits for functional follow-ups.

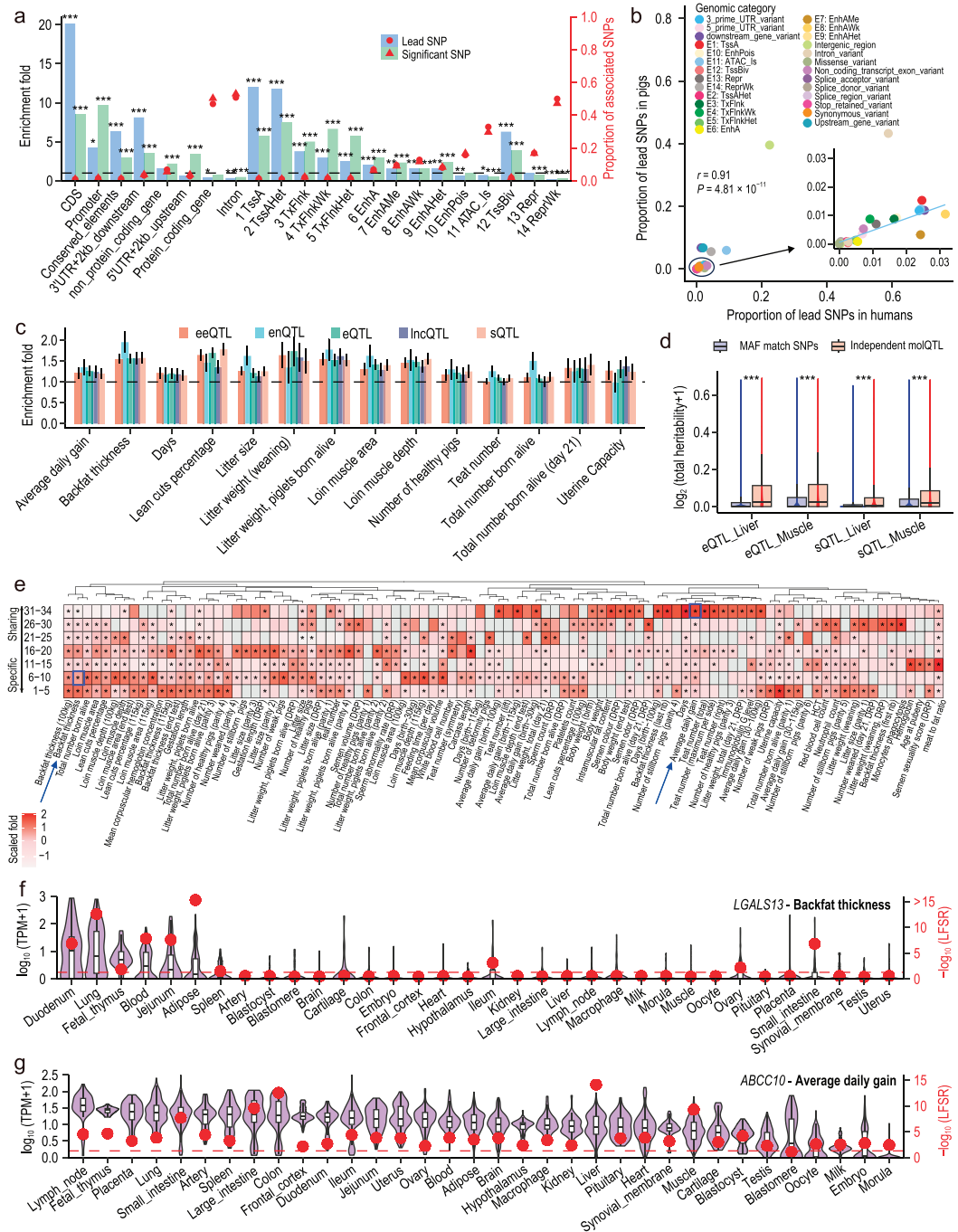


Figure 4. Exploiting the PigGTEx resource to decipher regulatory mechanisms of GWAS loci. (a) Results of annotation and enrichment of lead SNPs and genome-wide-level significant SNPs from within- and cross-breed meta-GWAS with a sample size greater than 1000 in different categories of genomic regions, conserved elements and regulatory elements. The enrichment analyses were conducted using resampled SNPs that matched the MAF (within 0.02) and LD (within 0.1) of the significant SNPs. The red dots and bars indicate the proportion and the enrichment fold of associated SNPs in the category. Significance of enrichment was indicated by *, ** and *** for $P < 0.05$, 0.01 and 0.001 , respectively. (b) Annotation results of the lead SNPs identified in this study (y-axis) and downloaded from the human GWAS catalog (<https://www.ebi.ac.uk/gwas/>) (x-axis) in each category of SNPeff and chromatin states (https://figshare.com/articles/dataset/6_type_of_regulator_hg19_zip/13480425). Each dot represents a genomic category. The Pearson correlation coefficient and significance were calculated by the *cor.test* function in R. (c) The heritability enrichment for 5 types of molecular *cis*-QTL (molQTL) in 14 complex traits with large sample size. The dashed line represents enrichment fold = 1. The error bar reflects the standard error of the enrichment fold. *cis*-eQTL: gene expression QTL, *cis*-sQTL: splicing QTL, *cis*-eeQTL: exon expression, *cis*-IncQTL: lncRNA expression QTL, *cis*-enQTL: enhancer expression QTL. (d) The estimated total SNP heritability contributed by independent molQTL in the liver

Figure 4. (Continued) and muscle for 232 complex traits. Significance is indicated by *** for $P < 0.001$, which was obtained by the *Wolcox.test* of *ggsignif* package in R. (e) The heritability enrichment for the genes of 7 tissue-sharing gradients in 94 complex traits with a sample size > 1000 . The red color represents the scaled heritability enrichment fold. The bright blue boxes indicate the position of examples (f and g). The ** indicates heritability enrichment fold > 1 and $P < 0.05$. Column clusters were produced by the *dist* function with the 'euclidean' method and the *hclust* function with the 'complete' method in R. The heatmap was plotted by *ggplot2* package (v3.3.2) in R (v4.2.1). (f) Expression (left, black) and the overall tissue-sharing pattern at LFSR obtained by MashR (v0.2–6) (right, red) of *LGALS13* in 34 tissues. (g) Expression (left, black) and the overall tissue-sharing pattern at LFSR obtained by MashR (v0.2–6) (right, red) of *ABCC10* in 34 tissues. The red dashed line in (f and g) indicates LFSR $< 5\%$.

Exploring shared genetic regulation of complex traits between pigs and humans

To explore the shared genetic regulatory mechanisms underlying complex traits between humans and pigs, we conducted heritability enrichment analyses on pig GWAS loci from each of the 126 pig meta-GWAS with a sample size exceeding 1000, in relation to orthologous QTL ($P < 5 \times 10^{-8}$) from 136 human diseases/traits across 18 trait domains (Table S12). These analyses revealed 62 significant human–pig trait pairs (enrichment fold > 1 and $P < 0.05/126 \times 136$), representing 21 pig traits and 47 human traits (Fig. 6a and b, Table S13), including body weight (end test) in pigs and esophagitis in humans (enrichment fold = 8.30, $P = 3.13 \times 10^{-73}$) (Fig. 6a and b). For these 62 human–pig trait pairs, the corresponding conserved, homologous and random loci exhibited lower but conserved pig GWAS loci showing a similar heritability and enrichment fold (Fig. S10, Table S13). We also calculated Spearman's correlations for human–pig trait pairs based on the absolute Z-score (effect size) of homologous variants from GWAS summary statistics. We observed significant correlations for physiologically relevant traits between the two species (Fig. 6c, Table S14). For example, the Days to 100 kg in pigs showed a significant correlation with Coxarthrosis in humans (Spearman's $r = -0.07$, $P = 2.70 \times 10^{-4}$), litter size in pigs correlated with type 2 diabetes in humans (Spearman's $r = 0.10$, $P = 6.16 \times 10^{-5}$), and body weight at birth in pigs correlated with personal history of malignant neoplasm of digestive organs in humans (Spearman's $r = 0.08$, $P = 1.32 \times 10^{-4}$) (Fig. 6c). These findings complement established discoveries in human research [64–67]. Meanwhile, we introduced novel insights, such as the number weaned in pigs is correlated with memory and unstable angina in humans, and loin muscle area in pigs is associated with the presence of orthopedic joint implants in humans (Fig. 6c). Furthermore, we identified potential conserved regulatory mechanisms between humans and pigs. For example, rs322242884, associated with ADG in Landrace pigs (Z-score = -4.53 , $P = 5.80 \times 10^{-6}$), has a homologous variant rs11877146 associated with body fat percentage in humans (Z-score = 6.06,

$P = 1.33 \times 10^{-9}$) (Fig. 6d). Notably, these homologous variants were eQTLs for *NPC1* in the muscle ($P_{\text{human}} = 6.70 \times 10^{-5}$, $P_{\text{pig}} = 7.20 \times 10^{-7}$), and for *TMEM241* in the brain ($P_{\text{human}} = 1.80 \times 10^{-7}$ and $P_{\text{pig}} = 1.78 \times 10^{-4}$) in both species (Fig. 6e and f). Consistently, previous studies have proposed that *NPC1* was associated with body weight and adipocyte processes across various animals [68–71] and *TMEM241* was associated with bone degeneration and osteoporosis [72]. These results together provided evidence for shared regulatory mechanisms underlying complex traits between humans and pigs, supporting the use of pigs as models for studying human biology and diseases.

DISCUSSION

In summary, to maximize the power of QTL discovery in pigs, we harmonized population-scale data sets from 59 diverse populations, encompassing a total of 70 328 pigs. We then carried out a large-scale cross-breed meta-GWAS analysis for 232 complex traits and within-breed meta-GWAS for 12 complex traits, representing 22 production traits, 54 meat and carcass traits, 61 health traits, 89 reproduction traits and 6 exterior traits (Fig. 1e). In total, we reported 6878 QTL for 169 traits. We annotated most of these QTL by examining regulatory elements from the FAANG project [33] and molQTL from the PigGTEx project [34]. We obtained 14 829 variant–gene–tissue–trait circuits, in which regulatory variants affect complex traits by modulating gene expression in specific tissues. Finally, by comparing with 136 publicly available human GWAS summary statistics, we revealed a certain conservation of genetic and regulatory mechanisms of complex traits between pigs and humans. Our findings provided evidence for using pigs as biomedical models to investigate the underlying genetic mechanisms of human biology and diseases [19,73], particularly for those with challenging phenotypic accessibility in humans.

While the QTLdb has documented 48 875 QTL from 426 publications on pig complex traits [24], the applicability of these findings across populations/breeds is limited as most of them were derived from individual populations. The multi-ancestry

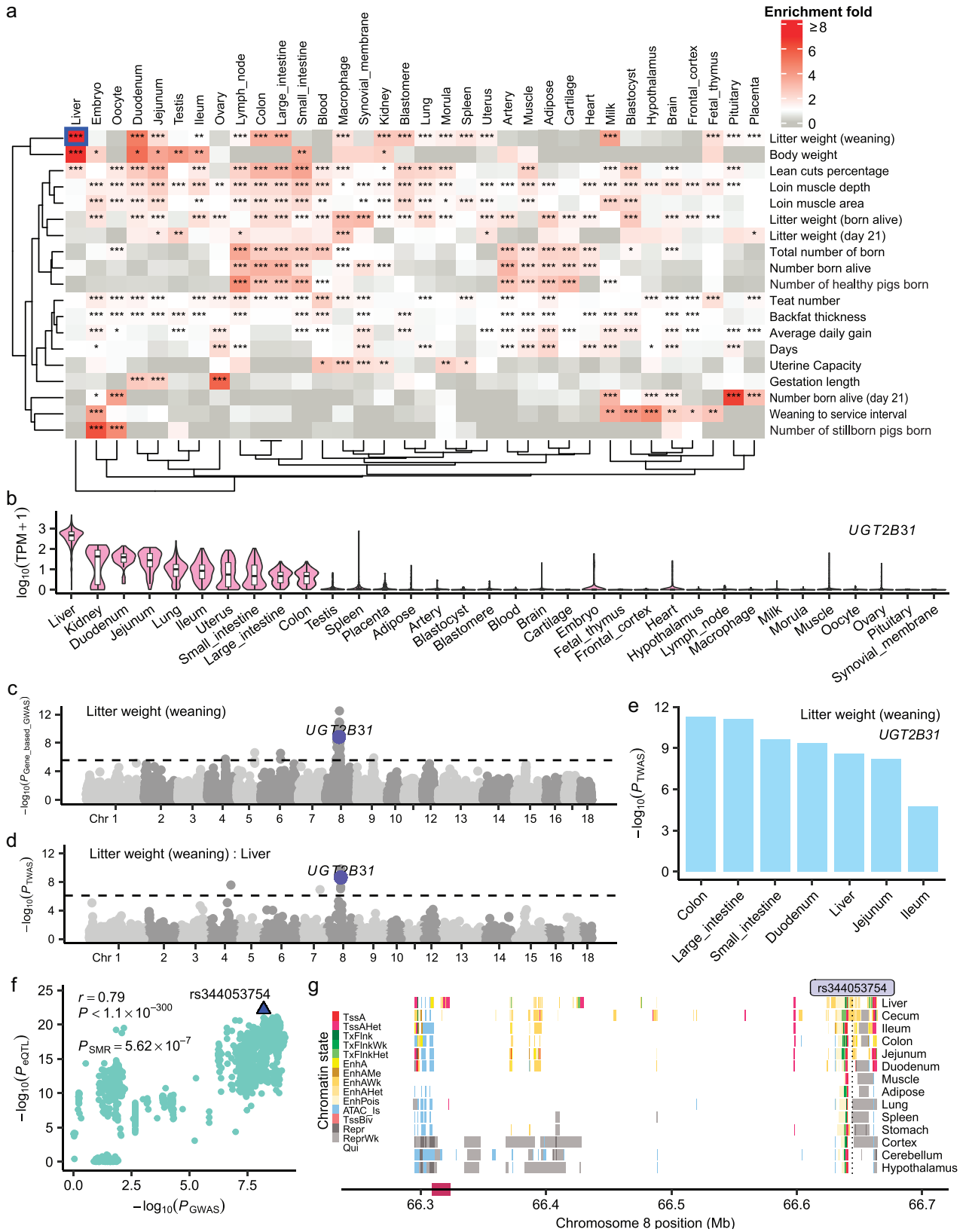


Figure 5. Tissue-specific regulation of GWAS loci. (a) Enrichment results for significantly associated SNPs of 19 cross-breed meta-GWAS in tissue-specific functional regions (the top 1000 tissue-specific highly expressed genes per tissue and their upstream and downstream 100 kb regions) in 34 tissues. Colors indicate enrichment fold. Rows indicate traits and columns indicate tissues. The bright blue box indicates the position of examples (b–g). Enrichment for trait-tissue pairs $E_T = p_T$ (proportion of significant SNPs for trait T located in functional regions of tissue T)/ q_T (proportion of all SNPs

Figure 5. (Continued) located in functional regions of tissue T_i , as calculated with BEDTools v2.25.0 [88]. Associated SNPs were resampled 1000 times with MAF within 0.02 and LD within 0.1, matched to calculate enrichment significance. $E_T > 1$ and $P < 0.05$ indicates the significant trait-tissue pair. Significance of enrichment is indicated by *, ** and *** for $P < 0.05$, 0.01 and 0.001, respectively. (b) Expression of gene *UGT2B31* in 34 tissues. (c) A Manhattan plot representing the gene-based GWAS results of litter weight (weaning). (d) A Manhattan plot representing the single-tissue TWAS results of litter weight (weaning) in the liver. (e) The single-tissue TWAS results of *UGT2B31* for litter weight (weaning) from S-PrediXcan [89]. (f) The summary data-based Mendelian randomization (SMR) results of litter weight (weaning) GWAS (x-axis) and *cis*-eQTL (y-axis) of *UGT2B31* on chromosome 8 in the liver. (g) The chromatin states around *UGT2B31* on chromosome 8.

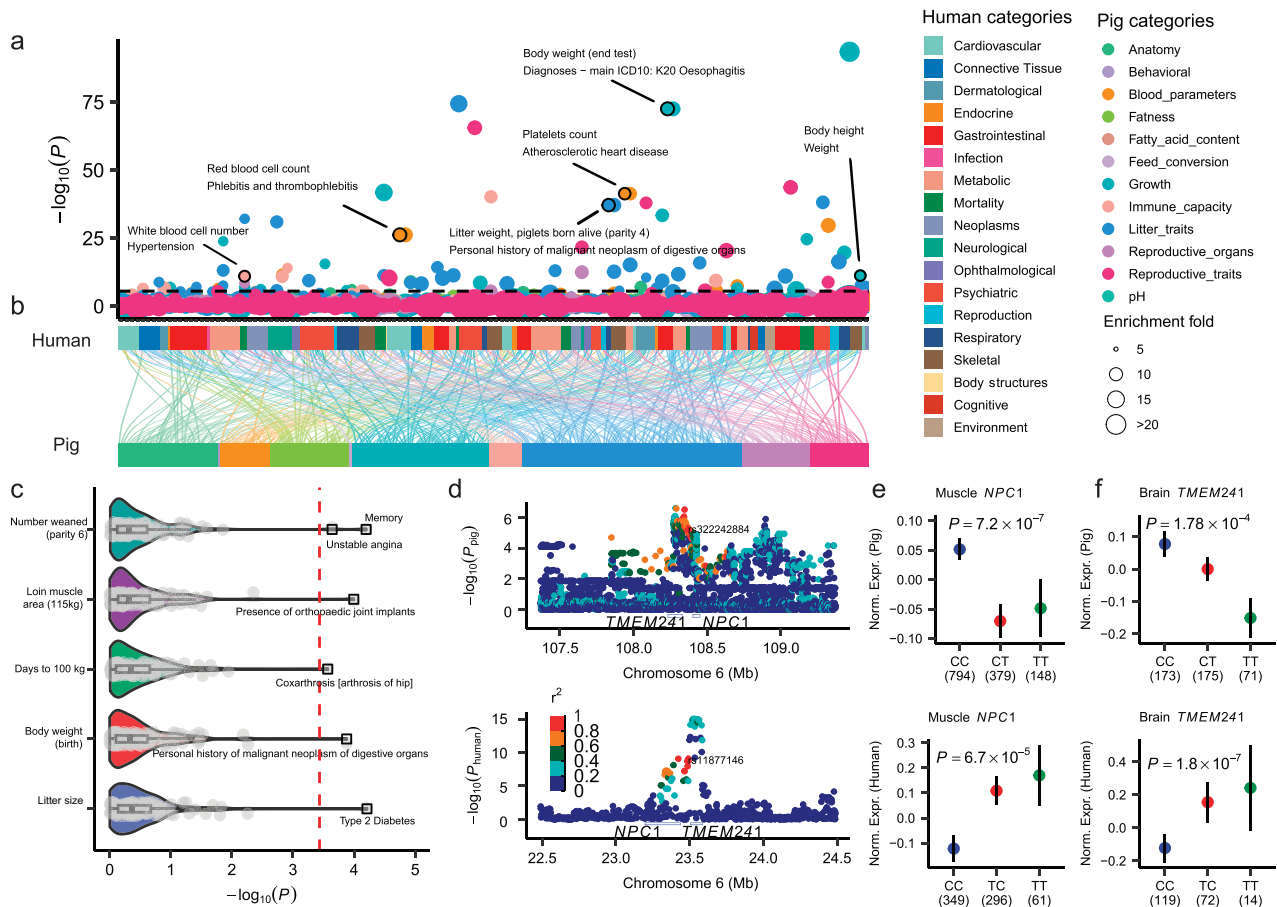


Figure 6. Comparison of complex trait genetics between humans and pigs. (a) The heritability enrichment of 126 pig GWAS in 136 human GWAS. The x-axis indicates traits and trait categories of human GWAS. The y-axis indicates the significance of heritability enrichment. The color of the dots indicates the trait category of pig GWAS. The size of the dots indicates the fold of heritability enrichment. The dashed line indicates the significant threshold ($P = 0.05/126 \times 136$ pig traits). (b) An alluvium-stratum plot showing the correlation between human and pig GWAS. Colors indicate trait categories. (c) Spearman's correlations of traits between humans and pigs, which were estimated by the absolute Z score of homologous variants from GWAS. The x-axis indicates the significance of Spearman's correlations. The red dashed line indicates the significant threshold ($P = 0.05/136$ human traits). The top trait pairs are labeled. (d–f) Similar regulatory mechanisms between body fatness rate (BFR) in humans and ADG in pigs. (d) A local Manhattan plot of GWAS for ADG in pigs (top) and BFR in humans (bottom). The red triangles represent homozygous variants of humans (rs11877146) and pigs (rs322242884). The color indicates the magnitude of LD between the homologous variants and other SNPs. (e) The effects of homozygous variants in (d) on the expression of homozygous gene *NPC1* in the muscle of pigs (top) and humans (bottom). (f) The effects of homologous variants in (d) on the expression of the homologous gene *TMEM241* in the brain of pigs (top) and humans (bottom). The black error bars in (e and f) represent the standard errors. The significance tests in (e and f) were performed with the *wilcox.test* function of the *ggsignif* package in R (v4.2.1). The effect size of eQTL in humans and pigs in (e and f) was derived from the GTEx portal (<https://www.gtexportal.org/>) and PigGTEx portal (<https://piggtex.farmgtex.org/>), respectively.

meta-analyses in humans have shown increased statistical power for detecting novel and shared genetic associations [23,26–28]. To the best of our knowledge, in the present study, we generated the largest-

scale genetic diversity resource to date for studying complex traits in pigs. Our results demonstrated that cross-breed meta-GWAS analysis enhances the transferability of predictions across breeds [28]. In

addition, by comparing within-breed meta-analyses in Duroc, Landrace and Yorkshire, we identified 2936 breed-specific QTL for 12 complex traits, including 3 production and 7 reproduction traits. The breed-specific lead SNPs tended to exhibit lower allele frequencies and higher effect sizes in the discovered breed than in other breeds. This aligns with human research and may be partly a result of the difference in sample size [27]. Our findings also implied that breed-shared QTL provide greater insight into the genetic architecture of traits, as breed-shared lead SNPs can cluster GWAS results according to traits rather than breeds. Conversely, we observed a higher proportion of QTL exhibiting pleiotropy within the same trait category compared to between trait categories, implying that multi-trait joint analysis could elucidate complex regulatory mechanisms shared among traits, as indicated in [74].

Functional annotation of GWAS results facilitates the identification of putative causal variants and guides subsequent empirical bench-based approaches to elucidate the molecular and cellular mechanisms behind GWAS loci [75,76]. By examining multi-omics data and functional annotation of various genomic features, we found that GWAS loci were significantly enriched in regulatory elements, including promoter and enhancer, as well as regulatory variants, such as eQTL and sQTL (QTL for alternative splicing). We observed distinct gene regulation patterns of complex traits. Specifically, the backfat thickness of fatness exhibited a significant enrichment in tissue-specific genes. On the other hand, the average daily gain of growth showed enrichment in ubiquitously expressed genes. This is in line with a previous study that demonstrated changes in mitochondrial DNA content and mitochondrial gene expression levels across multiple tissues during growth [77]. Furthermore, we proposed that rs344053754 has an impact on litter weight at weaning through influencing the expression of *UGT2B31* in the liver. Nevertheless, further validation of this hypothesis necessitates functional laboratory research. These findings provide valuable insights for the design of subsequent functional experiments and the advancement of animal breeding biotechnology. We also recognized several limitations within the current study. First, the sample sizes for many traits are modest, leading to reduced statistical power for QTL discovery. In the future, we will increase the sample size for these traits by combining data from more populations via meta-GWAS analyses. In addition, we will explore the impact of other types of genomic variants, such as copy number variants and InDels (insertions and deletions), on molecular and complex traits by employing long-read sequencing

and pangenome data. Although our integrative analysis has shed light on the genes and variants potentially involved in the underlying mechanisms of complex traits, future functional validation remains essential.

MATERIALS AND METHODS

Data collection and processing

We collected genotype and phenotype data in 70 328 pigs from 59 study populations, representing 14 pig breeds (details in Table S1). The phenotypes included 286 complex traits categorized into Reproduction, Meat and Carcass, Health, Production and Exterior (details in Table S4). To ensure data consistency, we implemented a standardized pipeline for processing individual-level genotype and phenotype data across populations. Ultimately, we retained 28 297 603 unique SNPs in 69 242 individuals, encompassing 20 706 Duroc, 34 540 Yorkshire, 9159 Landrace and 4837 from other pig breeds. Additionally, we retained 260 complex traits for downstream analyses (see Supplementary Data for details).

Individual GWAS and quality control

We conducted individual GWAS for each trait in each population. We performed association analysis using a logistic mixed model with fastGWA-GLMM in GCTA (v1.94.0) [78] for binary traits, and using a mixed linear model with fastGWA in GCTA (v1.94.0) [79] for quantitative traits. The model for each trait might be different for each population, but in general, we considered farms, sex, year and season as fixed effects and included the first five genotype principal components from within each population after removing outliers as covariates when necessary. Specifically, for 49 traits that have phenotypic records in multiple time points for the same individual (e.g. sperm traits and litter sizes, detailed in Table S4), we employed MMAP (v2021_08_19_22_30.intel) (<https://mmap.github.io/>) to perform association analysis based on their de-regressed proofs and weights calculated by DMU software (v6-R5-2-EM64T) [80]. Details are provided in Supplementary Data.

To enable individual GWAS from different populations to be comparable in the meta-analysis, we checked all summary-level GWAS results based on EasyQC [81]. After stringent quality control procedures, we ultimately retained 2056 out of 2117 individual GWAS for subsequent analysis, involving 232 complex traits. Details are provided in Supplementary Data.

Meta-analysis of GWAS

To identify the QTL shared across populations, as in [23], we performed cross-breed meta-analyses on the filtered individual GWAS results for each of the 232 traits using METAL (v2011-03-25) [82]. In addition, we performed 36 within-breed meta-analyses for 12 complex traits with large sample sizes in Duroc, Landrace and Yorkshire, respectively. In addition, we conducted gene-based GWAS for each trait using MAGMA (v1.10) [83]. Details are provided in [Supplementary Data](#).

Definition and validation of QTL

For both individual GWAS and meta-GWAS, we used $P < 5.0 \times 10^{-8}$ as the genome-wide significance threshold and defined lead SNPs and QTL based on genomic position by referring to the research of cattle stature [23]. If a QTL for trait A overlaps but does not fully coincide with a QTL for trait B, they are counted as two separate QTL, i.e. two trait-associations. To validate the identified QTL regions, we used three strategies. First, we validated the QTL regions in independent populations. Second, we used information on suggestive lead SNPs ($P < 1.0 \times 10^{-5}$) for breed-level genomic prediction. Third, we tested genomic prediction in the independent populations. Details are provided in [Supplementary Data](#).

DATA AVAILABILITY

The summary statistics of cross-breed and within-breed meta-GWAS underlying this article are available in ScienceDB (<https://www.scidb.cn/en/s/ZRjEve>), and can be queried and visualized in the PigBiobank web portal (<https://pigbiobank.farmgtex.org/>). The integrated results of meta-GWAS and molQTL from PigGTEx, such as TWAS, SMR and co-localization, are also available in ScienceDB (<https://www.scidb.cn/en/s/zU7rIj>). The BioProject of WGS samples in the genotype imputation reference panel PGRP are listed in Supplementary Table 2. Chromatin states of pigs are available through the UCSC Genome Browser (http://genome.ucsc.edu/s/zhypan/susScr11_15_state_14_tissues_new). The raw genotype and phenotype data of 14 populations downloaded from the public database are shown in Supplementary Table 15. The raw genotype and phenotype data of the remaining populations are not publicly available since they are commercial breeding data, but they are available from the corresponding author upon reasonable request. The codes for genotype imputation, GWAS, functional enrichment, heritability enrichment and cross-species comparison are stored on GitHub (<https://github.com/SCAU-AnimalGenetics/Pig-metaGWAS>). Scripts and code for co-localization, SMR and TWAS are available at the FarmGTEx GitHub website (<https://github.com/FarmGTEx/PigGTEx-Pipeline-v0>).

SUPPLEMENTARY DATA

Supplementary data are available at [NSR](#) online.

ACKNOWLEDGEMENTS

We are very grateful to all the researchers who contributed to the publicly available data used in this research. We thank the National Supercomputer Center in Guangzhou, China, for their technical support. We are also grateful to the National Supercomputing Center in Wuxi for doing the numerical calculations in this paper on their supercomputer system. We thank the editor and two anonymous reviewers for their constructive comments.

FUNDING

This work was supported by the National Key R&D Program of China (2022YFF1000900 to Zhe Zhang), the National Natural Science Foundation of China (32022078 to Zhe Zhang), the China Agriculture Research System (CARS-35 to Zhe Zhang, J.L., Y.C., X. Liu., S.M. and X.D.), the Local Innovative and Research Teams Project of Guangdong Province (2019BT02N630 to Zhe Zhang), the Agriculture and Food Research Initiative Competitive grants (2022-67015-36215 to L.F.) from the USDA National Institute of Food and Agriculture, the USDA NIFA AFRI grant (2019-67015-29321 and 2021-67015-33409 to G.E.L.), 'Accelerating Genetic Improvement of Ruminants Through Enhanced Genome Assembly, Annotation, and Selection' of the USDA Agricultural Research Service (ARS) (appropriated project 8042-31000-112-00-D to G.E.L.), the SCINet project of the USDA ARS project (0500-00093-001-00-D), the United States Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) (in part by AFRI grant numbers 2020-67015-31398 and 2021-67015-33409 to L.M.), the Australian Research Council's Discovery Projects (DP200100499 to R.X.), the grant from the Danish Research Council for Independent Research (DFF 1335-00127 to M.F. and P.K.M.), PID2020-112677RB-C21 awarded by MCIN/AEI/<https://doi.org/10.13039/501100011033> (to M.B.) and the European Union's Horizon 2020 research and innovation program (the GENE-SWitCH project under grant agreement no. 817998 to D.C.P.). Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. The USDA is an equal opportunity provider and employer. All the funders had no role in study design, data collection, analysis, decision to publish or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Study design: Zhe Zhang, L.F., F.Z., Y.Z.; genotype and phenotype data preprocessing: Zhiting Xu, Z. Zhong, H. Zeng, Xiaoping Wang, L.S., Xue Wang, Y. Wang, Zipeng Zhang, Y. Lin, C.W., J.Z., X.Z., Q.L., J.T., S.D., Yuqiang Liu, X. Pan, X.F., R.L., Z.S., C.C., Q. Zhu; genotype imputation: Z. Zhong, Zhiting Xu, H. Zeng, Xiaoping Wang, H.Y.; individual GWAS and meta-GWAS analyses:

Zhiting Xu, Z. Zhong, B.L., H. Zeng, C.W., Xiaoqing Wang; QTL validation: Q.L., Z.C., Zhiting Xu; GWAS and multi-omics data integration: X.C., Q.L., J.T., J.W., Zhiting Xu; genetic parameter estimation: Q.L., W.Z., S.S., Y. Wu; comparison of pig GWAS with human GWAS: Q.L., Y. Wu, Z.B.; critical interpretation of analytical results: L.F., Zhe Zhang, B.L., J.T., Y.G., G.E.L., P.K.M., M.F., M.A., S.L., F.Z., Y.Z., Q. Zhang, G.E.L., X.S., R.X., L.M., M.S.L., G. Sahana, G. Su, Yang Liu, P.L., M.B.; contribution of data and computational resources: Zhe Zhang, L.F., F.Z., Y.Z., Q. Zhang, Y.C., J.L., X.D., X. Li, M.L., G.T., M.F., P.K.M., A.C., M.A., D.C.P., M.B., R.H., P.L., X.Y., H. Zhang, X. Liu, D.M., Y.P., Q.W., K.L., L.B., Z. Zhou, Zhong Xu, X. Peng, S.M.; drafting the manuscript: Zhe Zhang, L.F., Zhiting Xu, Q.L. All authors read, edited and approved the final manuscript.

Conflict of interest statement. None declared.

REFERENCES

- FAO. 2023. *Meat Market Review: Emerging Trends and Outlook*. 2023. <https://openknowledge.fao.org/handle/20.500.14283/cd0465en>
- Tribout T, Larzul C, Phocas F. Efficiency of genomic selection in a purebred pig male line. *J Anim Sci* 2012; **90**: 4164–76.
- Whitworth KM, Rowland RRR, Ewen CL *et al*. Gene-edited pigs are protected from porcine reproductive and respiratory syndrome virus. *Nat Biotechnol* 2016; **34**: 20–2.
- Xu K, Zhou Y, Mu Y *et al*. CD163 and pAPN double-knockout pigs are resistant to PRRSV and TGEV and exhibit decreased susceptibility to PDCoV while maintaining normal production performance. *eLife* 2020; **9**: e57132.
- Gutierrez-Reinoso MA, Aponte PM, Garcia-Herreros M. Genomic analysis, progress and future perspectives in dairy cattle selection: a review. *Animals* 2021; **11**: 599.
- Vukasovic N, Bacciu N, Przybyla CA *et al*. Development of genetic and genomic evaluation for wellness traits in US Holstein cows. *J Dairy Sci* 2017; **100**: 428–38.
- McNeel AK, Reiter BC, Weigel D *et al*. Validation of genomic predictions for wellness traits in US Holstein cows. *J Dairy Sci* 2017; **100**: 9115–24.
- Strandén I, Kantanen J, Lidauer MH *et al*. Animal board invited review: genomic-based improvement of cattle in response to climate change. *Animal* 2022; **16**: 100673.
- Hayes BJ, Lewin HA, Goddard ME. The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends Genet* 2013; **29**: 206–14.
- de Almeida AM and Bendixen E. Pig proteomics: a review of a species in the crossroad between biomedical and food sciences. *J Proteomics* 2012; **75**: 4296–314.
- Jakobsen JE, Johansen MG, Schmidt M *et al*. Expression of the Alzheimer's disease mutations A β PP695sw and PSEN1M146I in double-transgenic Göttingen minipigs. *J Alzheimer's Dis* 2016; **53**: 1617–30.
- Barallobre-Barreiro J, Didangelos A, Schoendube FA *et al*. Proteomics analysis of cardiac extracellular matrix remodeling in a porcine model of ischemia/reperfusion injury. *Circulation* 2012; **125**: 789–802.
- Brasileiro ACL, de Oliveira DC, da Silva PB *et al*. Impact of topical nifedipine on wound healing in animal model (pig). *J Vasc Bras* 2020; **19**: e20190092.
- Hwang J, Jeong H, Lee N *et al*. Ex vivo live full-thickness porcine skin model as a versatile in vitro testing method for skin barrier research. *Int J Mol Sci* 2021; **22**: 657.
- Mordhorst BR and Prather RS. *Pig Models of Reproduction. Animal Models and Human Reproduction*. Hoboken, New Jersey: Wiley, 2017, 213–34.
- Burrin D, Sangild PT, Stoll B *et al*. Translational advances in pediatric nutrition and gastroenterology: new insights from pig models. *Annu Rev Anim Biosci* 2020; **8**: 321–54.
- Menduni F, Davies LN, Madrid-Costa D *et al*. Characterisation of the porcine eyeball as an in-vitro model for dry eye. *Contact Lens and Anterior Eye* 2018; **41**: 13–7.
- Pabst R. The pig as a model for immunology research. *Cell Tissue Res* 2020; **380**: 287–304.
- Lunney JK, Van Goor A, Walker KE *et al*. Importance of the pig as a human biomedical model. *Sci Transl Med* 2021; **13**: eabd5758.
- Griffith BP, Goerlich CE, Singh AK *et al*. Genetically modified porcine-to-human cardiac xenotransplantation. *N Engl J Med* 2022; **387**: 35–44.
- Wang Q, Tang J, Han B *et al*. Advances in genome-wide association studies of complex traits in rice. *Theor Appl Genet* 2020; **133**: 1415–25.
- Wang X, Wang H, Liu S *et al*. Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. *Nat Genet* 2016; **48**: 1233–41.
- Bouwman AC, Daetwyler HD, Chamberlain AJ *et al*. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat Genet* 2018; **50**: 362–7.
- Hu Z-L, Park CA, Reecy JM. Bringing the animal QTLdb and Cor-DB into the future: meeting new challenges and providing updated services. *Nucleic Acids Res* 2022; **50**: D956–61.
- Visscher PM, Brown MA, McCarthy MI *et al*. Five years of GWAS discovery. *Am Hum Genet* 2012; **90**: 7–24.
- Shu X, Long J, Cai Q *et al*. Identification of novel breast cancer susceptibility loci in meta-analyses conducted among Asian and European descendants. *Nat Commun* 2020; **11**: 1217.
- Graham SE, Clarke SL, Wu K-HH *et al*. The power of genetic diversity in genome-wide association studies of lipids. *Nature* 2021; **600**: 675–9.
- Mahajan A, Spracklen CN, Zhang W *et al*. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet* 2022; **54**: 560–72.
- Maurano MT, Humbert R, Rynes E *et al*. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012; **337**: 1190–5.
- Farh KK-H, Marson A, Zhu J *et al*. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 2015; **518**: 337–43.
- Finucane HK, Bulik-Sullivan B, Gusev A *et al*. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* 2015; **47**: 1228–35.

32. Dimas AS, Deutsch S, Stranger BE *et al*. Common regulatory variation impacts gene expression in a cell type–dependent manner. *Science* 2009; **325**: 1246–50.
33. Clark EL, Archibald AL, Daetwyler HD *et al*. From FAANG to fork: application of highly annotated genomes to improve farmed animal production. *Genome Biol* 2020; **21**: 285.
34. Teng J, Gao Y, Yin H *et al*. A compendium of genetic regulatory effects across pig tissues. *Nat Genet* 2024; **56**: 112–23.
35. Holm B, Bakken M, Klemetsdal G *et al*. Genetic correlations between reproduction and production traits in swine. *J Anim Sci* 2004; **82**: 3458–64.
36. Yang Y, Gan M, Yang X *et al*. Estimation of genetic parameters of pig reproductive traits. *Front Vet Sci* 2023; **10**: 1172287.
37. Yu G, Wang C, Wang Y. Genetic parameter analysis of reproductive traits in Large White pigs. *Anim Biosci* 2022; **35**: 1649–55.
38. Ye J, Tan C, Hu X *et al*. Genetic parameters for reproductive traits at different parities in Large White pigs. *J Anim Sci* 2018; **96**: 1215–20.
39. Cabling MM, Kang HS, Lopez BM *et al*. Estimation of genetic associations between production and meat quality traits in Duroc pigs. *Asian Australas J Anim Sci* 2015; **28**: 1061–5.
40. Ballester M, Ramayo-Caldas Y, González-Rodríguez O *et al*. Genetic parameters and associated genomic regions for global immunocompetence and other health-related traits in pigs. *Sci Rep* 2020; **10**: 1–15.
41. Ballester M, Jové-Juncà T, Pascual A *et al*. Genetic architecture of innate and adaptive immune cells in pigs. *Front Immunol* 2023; **14**: 1058346.
42. Edfors-Lilja I, Wattrang E, Magnusson U *et al*. Genetic variation in parameters reflecting immune competence of swine. *Vet Immunol Immunopathol* 1994; **40**: 1–16.
43. Clapperton M, Diack AB, Matika O *et al*. Traits associated with innate and adaptive immunity in pigs: heritability and associations with performance under different health status conditions. *Genet Sel Evol* 2009; **41**: 54.
44. Flori L, Gao Y, Laloë D *et al*. Immunity traits in pigs: substantial genetic variation and limited covariation. *PLoS One* 2011; **6**: e22717.
45. Clapperton M, Glass EJ, Bishop SC. Pig peripheral blood mononuclear leucocyte subsets are heritable and genetically correlated with performance. *Animal* 2008; **2**: 1575–84.
46. Marchini J, Cardon LR, Phillips MS *et al*. The effects of human population structure on large genetic association studies. *Nat Genet* 2004; **36**: 512–7.
47. Visscher PM, Wray NR, Zhang Q *et al*. 10 years of GWAS discovery: biology, function, and translation. *Am Hum Genet* 2017; **101**: 5–22.
48. Yengo L, Vedantam S, Marouli E *et al*. A saturated map of common genetic variants associated with human height. *Nature* 2022; **610**: 704–12.
49. Cai Z, Christensen OF, Lund MS *et al*. Large-scale association study on daily weight gain in pigs reveals overlap of genetic factors for growth in humans. *Bmc Genomics* 2022; **23**: 133.
50. Galvin JM, Wilmot I, Day BN *et al*. Reproductive performance in relation to uterine and embryonic traits during early gestation in Meishan, Large White and crossbred sows. *Reproduction* 1993; **98**: 377–84.
51. White BR, McLaren DG, Dziuk PJ *et al*. Age at puberty, ovulation rate, uterine length, prenatal survival and litter size in Chinese Meishan and Yorkshire females. *Theriogenology* 1993; **40**: 85–97.
52. Wang L, Wang A, Wang L *et al*. Animal genetic resources in China: pigs. *China Natl Comm Anim Genet Resour* 2011. Beijing: China Agriculture Press.
53. Wang, J, Gazal S. Ancestry-specific regulatory and disease architectures are likely due to cell-type-specific gene-by-environment interactions. *medRxiv* 2023; doi: 10.1101/2023.10.20.23297214.
54. Ray D, Venkataraghavan S, Zhang W *et al*. Pleiotropy method reveals genetic overlap between orofacial clefts at multiple novel loci from GWAS of multi-ethnic trios. *PLoS Genet* 2021; **17**: e1009584.
55. Han B and Eskin E. Interpreting meta-analyses of genome-wide association studies. *PLoS Genet* 2012; **8**: e1002555.
56. Pan Z, Yao Y, Yin H *et al*. Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. *Nat Commun* 2021; **12**: 5848.
57. Huppertz B, Meiri H, Gizurarson S *et al*. Placental protein 13 (PP13): a new biological target shifting individualized risk assessment to personalized drug design combating pre-eclampsia. *Hum Reprod Update* 2013; **19**: 391–405.
58. Than NG, Balogh A, Romero R *et al*. Placental protein 13 (PP13) - a placental immunoregulatory galectin protecting pregnancy. *Front Immunol* 2014; **5**: 348.
59. Li X, Kim S-W, Do K-T *et al*. Analyses of porcine public SNPs in coding-gene regions by re-sequencing and phenotypic association studies. *Mol Biol Rep* 2011; **38**: 3805–20.
60. Kojima M and Degawa M. Sex differences in the constitutive gene expression of sulfotransferases and UDP-glucuronosyltransferases in the pig liver: androgen-mediated regulation. *Drug Metab Pharmacokinet* 2014; **29**: 192–7.
61. Soars MG, Fettes M, O'Sullivan AC *et al*. Cloning and characterisation of the first drug-metabolising canine UDP-glucuronosyltransferase of the 2B subfamily. *Biochem Pharmacol* 2003; **65**: 1251–9.
62. Hamamoto-Hardman BD, Baden RW, McKemie DS *et al*. Equine uridine diphospho-glucuronosyltransferase 1A1, 2A1, 2B4, 2B31: cDNA cloning, expression and initial characterization of morphine metabolism. *Vet Anaesth Analg* 2020; **47**: 763–72.
63. Heikkinen AT, Friedlein A, Matondo M *et al*. Quantitative ADME proteomics—CYP and UGT enzymes in the beagle dog liver and intestine. *Pharm Res* 2015; **32**: 74–90.
64. Silverwood V, Blagojevic-Bucknall M, Jinks C *et al*. Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis. *Osteoarthritis Cartilage* 2015; **23**: 507–15.
65. Poulsen P, Grunnet LG, Pilgaard K *et al*. Increased risk of type 2 diabetes in elderly twins. *Diabetes* 2009; **58**: 1350–5.
66. Long L, He H, Shen Q *et al*. Birthweight, genetic risk, and gastrointestinal cancer incidence: a prospective cohort study. *Ann Med* 2023; **55**: 62–71.
67. Smith NR, Jensen BW, Zimmermann E *et al*. Associations between birth weight and colon and rectal cancer risk in adulthood. *Cancer Epidemiol* 2016; **42**: 181–5.
68. Altmann SW, Davis HR, Zhu L *et al*. Niemann-Pick C1 like 1 protein is critical for intestinal cholesterol absorption. *Science* 2004; **303**: 1201–4.
69. Lamri A, Pigeyre M, Garver WS *et al*. The extending spectrum of NPC1-related human disorders: from Niemann–Pick C1 disease to obesity. *Endocr Rev* 2018; **39**: 192–220.
70. Castillo JJ, Jelinek D, Wei H *et al*. The Niemann-Pick C1 gene interacts with a high-fat diet to promote weight gain through differential regulation of central energy metabolism pathways. *Am J Physiol Metab* 2017; **313**: E183–94.
71. Liu R, Zou Y, Hong J *et al*. Rare loss-of-function variants in NPC1 predispose to human obesity. *Diabetes* 2017; **66**: 935–47.
72. Teng Z, Zhu Y, Lin D *et al*. Deciphering the chromatin spatial organization landscapes during BMMSC differentiation. *J Genet Genomics* 2023; **50**: 264–75.
73. Meurens F, Summerfield A, Nauwynck H *et al*. The pig: a model for human infectious diseases. *Trends Microbiol* 2012; **20**: 50–7.

74. Shirai Y, Nakanishi Y, Suzuki A *et al*. Multi-trait and cross-population genome-wide association studies across autoimmune and allergic diseases identify shared and distinct genetic component. *Ann Rheum Dis* 2022; **81**: 1301–12.
75. Gallagher MD and AS C-P. The post-GWAS era: from association to function. *Am Hum Genet* 2018; **102**: 717–30.
76. Rasooly D, Peloso GM, Pereira AC *et al*. Genome-wide association analysis and Mendelian randomization proteomics identify drug targets for heart failure. *Nat Commun* 2023; **14**: 3826.
77. Xie YM, Jin L, Chen XJ *et al*. Quantitative changes in mitochondrial DNA copy number in various tissues of pigs during growth. *Genet Mol Res* 2015; **14**: 1662–70.
78. Jiang L, Zheng Z, Fang H *et al*. A generalized linear mixed model association tool for biobank-scale data. *Nat Genet* 2021; **53**: 1616–21.
79. Jiang L, Zheng Z, Qi T *et al*. A resource-efficient tool for mixed model association analysis of large-scale data. *Nat Genet* 2019; **51**: 1749–55.
80. Madsen P, Sørensen P, Su G *et al*. DMU—a package for analyzing multivariate mixed models. In: *8th World Congress on Genetics Applied to Livestock Production*, 2006, Brasil.
81. Winkler TW, Day FR, Croteau-Chonka DC *et al*. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* 2014; **9**: 1192–212.
82. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**: 2190–1.
83. de Leeuw CA, Mooij JM, Heskes T *et al*. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015; **11**: e1004219.
84. Chang CC, Chow CC, Tellier LC *et al*. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; **4**: 7.
85. Kanai M, Akiyama M, Takahashi A *et al*. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* 2018; **50**: 390–400.
86. Bulik-Sullivan BK, Loh P-R, Finucane HK *et al*. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; **47**: 291–5.
87. Bu D, Luo H, Huo P *et al*. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res* 2021; **49**: W317–25.
88. Quinlan AR and Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010; **26**: 841–2.
89. Barbeira AN, Dickinson SP, Bonazzola R *et al*. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* 2018; **9**: 1825.