Nucleic Acids Research, 2024, **52**, D980–D989
https://doi.org/10.1093/nar/gkad1080
Advance access publication date: 13 November 2023
Database issue

PigBiobank: a valuable resource for understanding genetic and biological mechanisms of diverse complex traits in pigs

Haonan Zeng[1,†], Wenjing Zhang[1,†], Qing Lin[1,†], Yahui Gao[1,†], Jinyan Teng[1], Zhiting Xu[1], Xiaodian Cai[1], Zhanming Zhong[1], Jun Wu[1], Yuqiang Liu[1], Shuqi Diao[1], Chen Wei[1], Wentao Gong[1], Xiangchun Pan[1], Zedong Li[1], Xiaoyu Huang[1], Xifan Chen[1], Jinshi Du[1], The PigGTEx Consortium, Fuping Zhao[2], Yunxiang Zhao[3], Maria Ballester[4], Daniel Crespo-Piazuelo[4], Marcel Amills[5,6], Alex Clop[5,7], Peter Karlskov-Mortensen[8], Merete Fredholm[8], Pinghua Li[9,10], Ruihua Huang[9,10], Guoqing Tang[11], Mingzhou Li[11], Xiaohong Liu[12], Yaosheng Chen[12], Qin Zhang[13], Jiaqi Li[1], Xiaolong Yuan[1], Xiangdong Ding[14,*], Lingzhao Fang[15,*] and Zhe Zhang[1,*]

[1]State 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
[2]Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China
[3]College of Animal Science and Technology, Guangxi University, Nanning 530004, China
[4]Animal Breeding and Genetics Programme, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Torre Marimon, Caldes de Montbui, Spain
[5]Department of Animal Genetics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus de la Universitat Autònoma de Barcelona, Bellaterra 08193, Spain
[6]Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain
[7]Consejo Superior de Investigaciones Científicas, Barcelona, Catalonia, Spain
[8]Animal Genetics, Bioinformatics and Breeding, Department of Veterinary and Animal Sciences, University of Copenhagen, Frederiksberg C1870, Denmark
[9]Institute of Swine Science, Nanjing Agricultural University, Nanjing 210095, China
[10]Key Laboratory in Nanjing for Evaluation and Utilization of Livestock and Poultry (Pigs) Resources, Ministry of Agriculture and Rural Areas, China, Nanjing 210095, China
[11]State Key Laboratory of Swine and Poultry Breeding Industry, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu 611130, China
[12]State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, China
[13]College of Animal Science and Technology, Shandong Agricultural University, Tai'an 271018, China
[14]College of Animal Science and Technology, China Agricultural University, Beijing 100193, China
[15]Center for Quantitative Genetics and Genomics (QGG), Aarhus University, Aarhus, Denmark

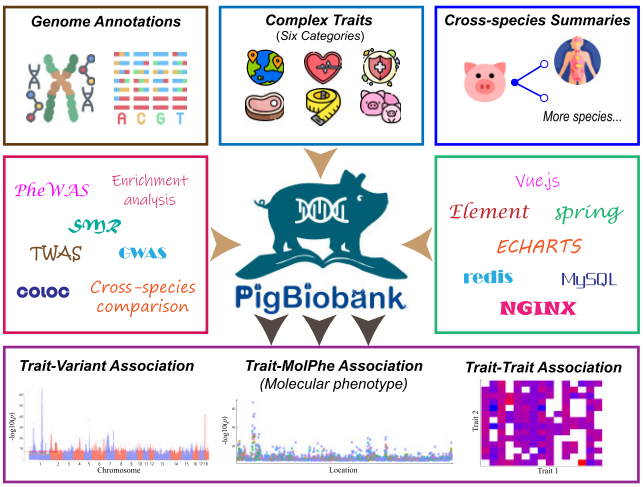
[*]To whom correspondence should be addressed. Tel +86 20 85282019; Email: zhezhang@scau.edu.cn
Correspondence may be also addressed to Lingzhao Fang. Tel +45 89991301; Email: lingzhao.fang@qgg.au.dk
Correspondence may be also addressed to Xiangdong Ding. Tel +86 10 62734277; Email: xding@cau.edu.cn
[†]The authors wish it to be known that, in their opinion, the first four authors should be regarded as Joint First Authors.

Abstract

To fully unlock the potential of pigs as both agricultural species for animal-based protein food and biomedical models for human biology and disease, a comprehensive understanding of molecular and cellular mechanisms underlying various complex phenotypes in pigs and how the findings can be translated to other species, especially humans, are urgently needed. Here, within the Farm animal Genotype-Tissue Expression (FarmGTEx) project, we build the PigBiobank (http://pigbiobank.farmgtex.org) to systematically investigate the relationships among genomic variants, regulatory elements, genes, molecular networks, tissues and complex traits in pigs. This first version of the PigBiobank curates 71 885 pigs with both genotypes and phenotypes from over 100 pig breeds worldwide, covering 264 distinct complex traits. The PigBiobank has the following functions: (i) imputed sequence-based genotype-phenotype associations via a standardized and uniform pipeline, (ii) molecular and cellular mechanisms underlying trait-associations via integrating multi-omics data, (iii) cross-species gene mapping of complex traits via transcriptome-wide association studies, and (iv) high-quality results display and visualization. The PigBiobank will be updated timely with the development of the FarmGTEx-PigGTEx project, serving as an open-access and easy-to-use resource for genetically and biologically dissecting complex traits in pigs and translating the findings to other species.

Received: August 14, 2023. **Revised:** October 13, 2023. **Editorial Decision:** October 24, 2023. **Accepted:** October 27, 2023

Graphical abstract



Introduction

The immense value of pigs in the realms of agriculture and human biomedicine has been widely acknowledged. Pork is the second-largest meat source worldwide and plays an important role in ending the global hunger and malnutrition crisis, as reported by the Food and Agriculture Organization of the United Nations (FAO) (1). Additionally, pigs are considered to be more similar to humans in terms of anatomical characteristics, physiology, immunology, and genome than other model organisms such as rodents (2). Lunney *et al.* (3) comprehensively reviewed the recent applications of pigs as human biomedical models in reproduction and fetal development, brain and neurodegenerative disease, and xenotransplantation, shedding light on the myriad advantages that pigs can offer in biomedical research (4–6). Hence, it becomes crucial to thoroughly explore and elucidate the underlying molecular mechanisms responsible for diverse complex traits in pigs, which will substantially contribute to both expediting the genetic gain via selective breeding and establishing the pig as an invaluable biomedical model.

Over the past decades, genome-wide association study (GWAS) has proven to be a valuable tool for understanding the genetic basis of complex traits and diseases. However, compared with large-scale public GWAS summary statistics of various complex traits in humans like GWAS Catalog (7), GWASdb (8),GWAS ATLAS (9), Brain Catalog (10) and Open Target (11), GWASs of complex traits in pigs are typically performed using SNP arrays in small populations with highly related individuals, where the linkage disequilibrium (LD) among SNPs is high. What is even worse: the access to full GWAS summary statistics in the public domain is limited, posing a hindrance to the aggregation of multiple independent GWAS datasets for performing meta-analyses. Moreover, the majority of these GWAS loci are primarily located in non-coding regions of the genome (12,13), implying that they exert their effects on complex traits via altering gene regulation and expression (14–16). Consequently, the underlying biological mechanisms that explain the effects of these non-coding variants on complex traits and diseases are still not well understood. Although the AnimalQTLdb (17) curated tens of thousands of QTLs of 279 traits from 800 publications, it lacks full summary statistics of each individual GWAS. Recently, Teng et al. (14) released the Pig Genotype-Tissue Expression (PigGTEx) resource, a highly valuable catalog of regulatory variants across multiple pig tissues. It provides an extensive collection of millions of molecular QTLs associated with five distinct types of molecular phenotypes (i.e. protein coding gene expression, exon expression, lncRNA expression, enhancer expression and alternative splicing) derived from 34 different pig tissues. The advent and future development of PigGTEx will greatly enhance our understanding of the regulatory mechanisms underlying complex traits in pigs and it will also facilitate cross-species gene mapping via transcriptome-wide association studies (TWAS) such as pigs vs. humans (14), chickens vs. humans (18), and rats vs. humans (19).

Here, to fully unleash the potential of pigs as an agricultural species and for human biomedical applications, we constructed the PigBiobank database (http://pigbiobank.farmgtex.org) by integrating large-scale imputed sequence-based GWAS of 264 distinct complex traits with the PigGTEx resource. To ensure the reliability and credibility of our data and findings, we implemented rigorous quality control measures and a standardized analytical pipeline throughout the entire process, from raw data collection to data analysis within the PigBiobank. It will serve as the most extensive resource for discerning candidate causal variants, genes, molecular networks, and tissues underlying complex traits in pigs. In addition, the PigBiobank not only allows the users to explore the genetic relationships between pig traits, but also establish connections with traits in other species (e.g. humans) via TWAS to uncover the shared genetic basis of homologous traits (e.g. human body weight vs. pig body weight) across species.

Materials and methods

Data collection and pre-processing

GWAS resources and pre-processing

The original data of PigBiobank is defined as a set of GWAS summary statistics generated from a large-scale meta-analysis of GWAS (metaGWAS, dataset 1) (20), and GWAS

on the eigenvector composition (eigenGWAS) and environmental phenotypes (envGWAS) using PGRP from PigGTEx project (14) (dataset 2). A total of 71 885 pigs with both genotypes and phenotypes from over 100 pig breeds, containing up to 264 complex traits that are classified into six main trait categories (17) (Supplementary Table S1, S2). To obtain accurate and valuable information, a standardized protocol was set up to deal with such complex datasets.

Dataset 1 consists of 70328 pigs from 59 populations in 14 pig breeds (20). SNPs genotyped with six types of SNP array from previous pig reference genome version were lifted over to the current Sscrofa11.1 (Ensembl v100) using R (v.3.6). Only the successfully mapped autosomal bi-allelic SNPs were retained. Afterwards, the SNP array data of each population was imputed to the sequence level via Beagle (v5.1) (21) using the Pig Genomics Reference Panel (PGRP) version 1 in PigGTEx project (14), which consists of 1602 WGS samples and 42 523 218 SNPs. For primary phenotypes, we manually removed phenotypic outliers and ensured consistency of the same trait among populations by the phenotypic descriptive statistics using R (v3.6) (https://www.R-project.org). Finally, a total of 232 complex traits were included in the subsequent analyses after quality control (QC), which were classified into five main trait categories (i.e. Reproduction, Meat and Carcass, Production, Health, Exterior).

Dataset 2 consisting of 1557 pigs from over 100 pig breeds was extracted from PGRP v1 (14). To elucidate the evolutionary patterns of environmental adaptability, a total of 30 environment-related phenotypes for envGWAS were obtained from the WorldSIM 2 (22), the High-resolution gridded datasets (23), and the GLOBMAP Leaf Area Index (LAI) (24) using the information on latitude and longitude for each indigenous pig breed. Moreover, to interpret the evolution and adaptation of the characteristics, two genetic differentiation phenotypes derived from the genotypic data are used to perform GWAS with eigenvector composition (eigenGWAS) within Asian–European pig populations and within Asian south-north domestic pig populations. In total, dataset 2 encompasses 32 traits that are further classified under the main trait category of 'Adaptation'.

Cis-molQTLs, chromatin states, and phastCons score resources. We downloaded 15 chromatin states from 14 pig tissues (25) and conservation scores (phastCons) from UCSC (http://hgdownload.cse.ucsc.edu/goldenpath/hg38/phastCons100way/hg38.100way.phastCons/) (phastCons100way). Moreover, we generated five different types of *cis*-molecular quantitative trait locus (*cis*-molQTL, <1 Mb to the TSS of genes) data involved in 34 tissues (i.e. tissues, and organ systems) from the PigGTEx project (14). To be specific, it included *cis*-eQTL for protein coding gene expression, *cis*-eeQTL for exon expression, *cis*-lncQTL for lncRNA expression, *cis*-enQTL for enhancer expression and *cis*-sQTL for alternative splicing.

Summary-based analysis

SNP-level and gene-level association analysis

For each of the populations in dataset 1, we performed GWAS on binary and continuous traits with fastGWA-GLMM and fastGWA models implemented in GCTA (v1.94.0beta) (26–28), respectively. For reproductive traits (with repeated recordings), association analysis was conducted using MMAP (released on 2021–08-19) (https://mmap.github.io/) with deregressed estimated breeding value (dEBV) as the phenotype. The metaGWAS was performed using the inverse variance-weighted fixed effects model in METAL (released on 2011-03-25) (29) after QC for GWAS on the basis of sample size, effect size and standard error. Ultimately, the 232 complex traits comprised of 268 studies were further classified into five main trait categories (i.e. Reproduction, Meat/Carcass, Production, Health, and Exterior). For more details of processing raw individual data, we referred to Xu *et al.* (20). For dataset 2, GWAS analysis of environment-related phenotypes was conducted with GEMMA (v0.98.5) (30), including five genotypic principal components as covariates. In addition, we conducted eigenGWAS for two genetic differentiation phenotypes using default parameters in GEAR (v0.919) (31). The analysis yielded 32 summary statistics falling under the category of 'Adaptation'. Lastly, leveraging SNP-level GWAS, we conducted the gene-level association analysis using MAGMA (v1.0) (32).

Estimation of heritability and genetic correlations of complex traits

We conducted the SNP heritability estimation for all the GWAS summary statistics by LDSC (v1.0.1) (33) with default parameters. In order to build the landscape of genetic correlations of pig complex traits, we used LDSC (v1.0.1) (33) to calculate the global genetic correlation for each pair of the aforementioned GWAS studies. By default, linkage disequilibrium (LD) pattern was calculated from biallelic SNPs in the reference panel PGRP v1. We also tested whether genetic correlations significantly deviated from 0 with the chi-square test (df = 1) using the Wald statistic.

Integrative analysis

TWAS, colocalization and SMR analysis

We performed gene prioritization through integrating the GWAS with five *cis*-molQTLs from 34 tissues using the following three complementary strategies, colocalization (COLOC), summary-based mendelian randomization (SMR), and transcriptome-wide association study (TWAS). To explore whether a genetic variant affects both the intermediate molecular phenotypes and the complex trait of interest, we used fastENLOC (v1.0) (34) to quantify the regional colocalization probability (RCP) for each independent molQTL signal clusters and GWAS hits, and considered a gene to be significant if its RCP $\geq$ 0.9 in the COLOC analysis. TWAS was to test associations between a complex trait of interest and genetically predicted gene expression levels. We applied S-PrediXcan (35) in single tissues and S-MultiXcan (36) in multiple tissues to detect transcriptionally regulated genes underlying complex traits with a stringent Bonferroni multiple-testing correction. Furthermore, SMR was performed to identify molecular phenotypes that are associated with a complex trait because of a shared candidate causal variant (i.e. pleiotropy or causality) (37). A Benjamini–Hochberg method correction (FDR < 0.05) was used in each SMR analysis and the heterogeneity in dependent instruments (HEIDI) test was applied to distinguish pleiotropy from linkage with a threshold of 0.05 (HEIDI > 0.05) (37).

Enrichment analysis

Identification of trait-relevant functional elements and molQTLs

We utilized SnpEff (v.4.3) (38) to annotate the biallelic SNPs in the PGRP v1 VCF file, resulting in 20 functional categories. For each functional category, the enrichment fold of molQTL was calculated by R package fmsb (v0.7.5). In addition, we utilized the stratified LD score regression (S-LDSC) (39) to comprehensively explore the heritability enrichment of GWAS with five molQTL annotations across all 34 tissues in dataset 1.

Inference for trait-relevant tissues

We applied four computational strategies to detect the tissues associated with complex traits. The first was QTLEnrich (v2) (40), as a rank- & permutation-based method, which aims to test for enrichment of trait-associations in molQTLs (e.g. eQTL, eeQTL, IncQTL, enQTL, sQTL) specific in each tissue. For each type of *cis*-molQTLs, QTLEnrich was used to test whether the molQTLs in a given tissue were significantly enriched for given traits ($P < 0.05$). Second, we extracted genomic regions by expanding 100 kb windows around the top 1000 genes that were highly expressed in each of the 34 tissues. Subsequently, we utilized BEDTools (v2.25.0) (41) to calculate the enrichment fold of these regions with trait-associations. Permutation tests with 10000 replicates were conducted to determine the P values using the R package regioneR (v1.24.0) (42). Third, we applied stratified LD score regression (S-LDSC) (43) to the above-mentioned genomic regions from 34 tissues to evaluate whether the heritability of each of the 232 traits in dataset 1 was significantly enriched in tissue-specific expressed gene regions. Lastly, we utilized BEDTools (v2.25.0) (41) to calculate the enrichment fold of the genomic region of each of 15 chromatin states in 14 tissues, and the permutation test with 10000 replicates was conducted to obtain the P values using the R package regioneR (v1.24.0) (42).

Cross-species comparison analysis

To explore the sharing patterns of the genetic architecture of complex traits between species, we further investigated the trait similarity between pigs and humans on the level of GWAS and TWAS summary statistics. On the one hand, we calculated the Pearson's correlation of the *z*-scores of GWAS for homologous variants between pigs (Sus scrofa11.1) and humans (GRCh38/hg38). On another hand, we obtained the TWAS summaries statistics from PigGTEx project (14) and calculated the Pearson's correlation of the absolute standardized effect size of TWAS for orthologous genes between pigs and humans.

Database design

The PigBiobank was designed with a decomposing framework using Vue (https://github.com/vuejs/core) as the front-end and Spring Boot in Java as the back-end. NGINX was used as the reverse proxy server for balancing the network load. To develop the user-friendly interface, we used Element (https://github.com/ElemeFE/element) for beautifying the page layout, ECharts (https://github.com/apache/echarts) and IGV (44) for data visualization. To fit and invocate the multi-omics data, we used MySQL as an engine for both data storage and data querying.

Results

Overview of PigBiobank

Herein, the current version of PigBiobank is of six perspectives: Trait, Resource, Biology, Analysis, Search and Download (Figure 1). With each perspective, a general exploration of the database is presented. (i) Trait. PigBiobank encompasses a comprehensive collection of 264 complex pig traits, which are meticulously classified into six main trait categories, including adaptation, exterior, health, meat and carcass, production, and reproduction (17). (ii) Resource. PigBiobank systematically collects, processes and consolidates the data resource from multiple databases, i.e. the PigGTEx-portal, Functional Annotation of Animal Genomes (FAANG) project, UK Biobank, Human GTEx project, Ensembl, UCSC and The National Center for Biotechnology Information (NCBI). (iii) Biology. PigBiobank integrates the aforementioned large-scale multi-omics data to facilitate users to effortlessly explore the regulatory mechanisms underlying various complex traits in pigs. The platform also strives to establish connections with traits in other species (currently only humans) to unveil the shared genetic basis of homologous traits. (iv) Analysis. PigBiobank provides comprehensive features collected from multi-layer analyses, including trait-variant association, trait-molecular QTL association, and trait-trait association. (v) Search. Users can utilize the user-friendly quick search function to explore specific traits, genomic variants, genes, or genomic regions of interest by entering relevant keywords. They can also jump seamlessly from one page to another to explore connections and interactions among traits, genes and genomic variants. (vi) Download. Users can intuitively visualize and freely download all the results of association analysis or query data from the PigBiobank.

Web interface and usage

We developed a user-friendly interface allowing users to access all the information from any device and location by means of searching, browsing, visualizing and downloading. The current version of the PigBiobank mainly contains six menus, namely Home, Trait Browser, Module, Download, Contact and Help (Figure 2A). The homepage provides the basic summary of the database and a search box on a very prominent position. Users can utilize the quick search function in search of traits, genomic variants, genes, or genomic regions of interest via typing relevant keywords (Figure 2A). Searching by trait is always a central point of database construction and web design, so we have made a separate overview interface for these traits in the menu of 'Trait browser'. The page displays concise summaries of 264 traits comprised of 300 studies, such as trait names, trait types, synonyms, main and sub trait categories, breeds, total sample size, total SNPs, lead SNPs, and single GWAS population. Users can click the items of interesting traits and then go into the specific web presentation for each trait that is similar to use the search function on the homepage. The menu of 'Module' provides detailed functions for point-to-point searching and analyzing such as GWAS and phenome-wide association study (PheWAS). The menu of 'Download' provides the download entry for the list of available files. The remaining menus (i.e. Contact, Help) allow users to get detailed information and documentation about the PigBiobank and convenient communications with us.

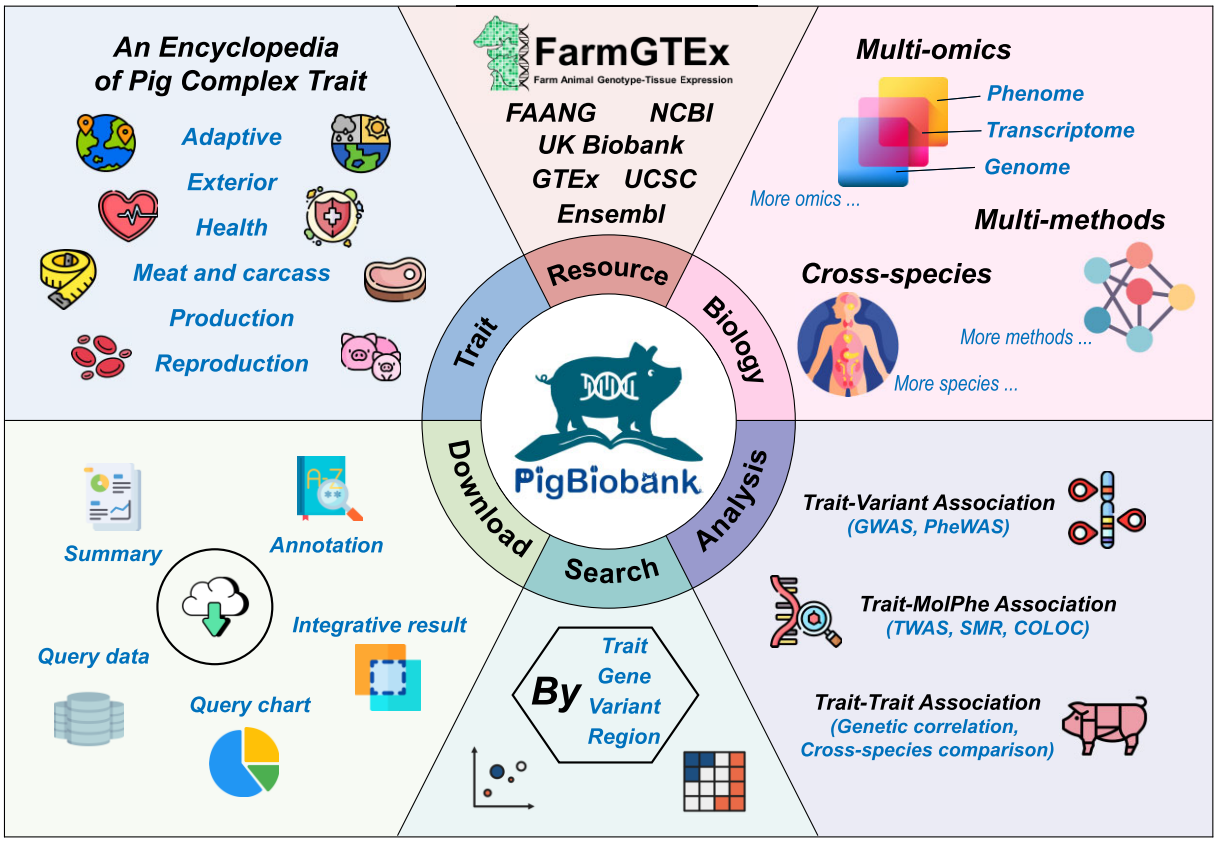


Figure 1. Schematic overview of the PigBiobank. It consists of six components. (i) Trait: 264 complex pig traits derived from six main trait categories. (ii) Resource: data collected from multiple relevant international projects. (iii) Biology: biological elucidation using multi-omics data, multi-methods analysis, and cross-species comparison. (iv) Analysis: seven online analysis modules/tools are available in the PigBiobank web portal. (v) Search: users can query the database in four different ways (by trait, gene, variant, or region) in a user-friendly manner. (vi) Download: data files or results are available for free download.

To help users elucidating the potential regulatory mechanism of complex traits in pigs, we categorize the research contents into six sections (Figure 2B), including (i) 'Summary-based Analysis' containing SNP-based GWAS, gene-based GWAS, genetic parameter estimation and PheWAS. (ii) 'Content of Region' by genomics viewer. (iii) 'Integrative Analysis' presented by TWAS, SMR, and COLOC. (iv) List of 'associated Variants/Genes' for target region or variant based on various methods. (v) 'Enrichment Analysis' for detecting trait-relevant functional elements and tissues. (vi) 'Cross-species Analysis' based on GWAS and TWAS. Moreover, we use interactive or static visualizations for the results from most analyses. For instance, as shown in Figure 2B, PigBiobank provides a variety of diagrams such as the scatter plot to show trait-associated SNPs and genes for the GWAS, PheWAS, TWAS, SMR, COLOC, and the heatmap plot is mainly utilized to visually depict the results of enrichment analysis and cross-species comparisons (i.e. human). For the genetic correlation analysis, the histogram is available for single trait search and heatmap for multiple traits search in the database. It is worth noting that PigBiobank provide all the accession of the images presented in the website.

Case study for complex trait search module

To demonstrate the utility of the PigBiobank in deciphering the molecular mechanisms behind complex traits in pigs and to showcase the functionality of the trait search module within PigBiobank, we present the study named 'MetaGWAS_M_BFT' focusing on the trait of 'Average backfat thickness' as a case study. Upon querying the database using the keyword 'MetaGWAS_M_BFT', following five sections are presented: (i) Summary of Trait, (ii) Summary-based Analysis, (iii) Integrative Analysis, (iv) Enrichment Analysis and (v) Cross-species analysis.

The section titled 'Summary of Trait' presents a table that provides extra information about the study named 'MetaGWAS_M_BFT' (Figure 3A). The table includes the following information: (i) trait details, including the trait name, trait type, trait category and breeds involved; (ii) association analysis details, which consist of the number of GWASs used for meta-analysis, phenotype records, tested SNPs, genome-wide significant SNPs, lead SNPs and the analysis software used and (iii) heritability estimated along with the standard error.

The section of 'Summary-based Analysis' presents the result of metaGWAS, gene-based GWAS and genetic correlation (Figure 3B). As an example, the 'SNP-based metaGWAS' panel displays significant signals of metaGWAS for 'MetaGWAS_M_BFT'. Among these metaGWAS signals, the top QTL (chr1: 160114440–161114440) with the lead SNP of 1_160614440_C_A ($P = 4.1620 \times 10^{-65}$) was identified. Notably, the putative gene *MC4R*, which has been supported by laboratory experiments and multiple association studies (45,46), is located within the QTL region. In the gene-based GWAS, we also find *MC4R* significantly associ-

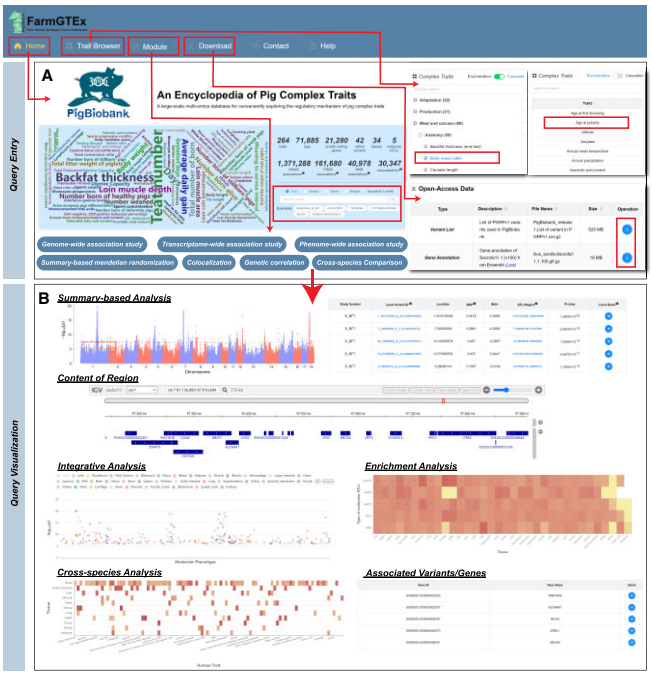


Figure 2. The web interface of the PigBiobank. (**A**) Query entries for the PigBiobank database. The menu options 'Home', 'Trait Browser', and 'Module' provide access to the quick search, detailed trait browses, and online analysis modules, respectively. (**B**) Query visualization. A total of six sections namely 'Summary-based Analysis', 'Content of Region', 'Integrative Analysis', 'Enrichment Analysis', 'Cross-species Analysis', and 'Associated Variants/Genes' are presented in the database.

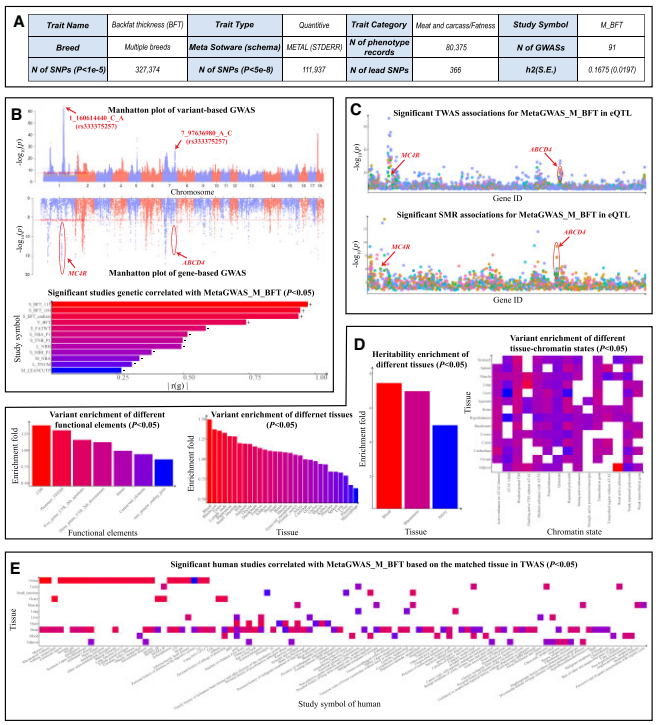


Figure 3. Case study of trait search module (MetaGWAS_M_BFT). (**A**) The summary information of the study. (**B**) The manhatton plot of variant-based and gene-based GWAS, and the list of traits significantly genetic correlated ($P < 0.05$) with study. (**C**) The scatter plot of significant associations of TWAS, and SMR (corrected $P < 0.05$). (**D**) The bar plot and heatmap show significant results of enrichment analysis for study, containing the variant enrichment trait-relevant functional elements, variant enrichment of trait-relevant tissue, heritability enrichment of trait-relevant tissue, and variant enrichment of trait-relevant tissue-chromatin states ($P < 0.05$). (**E**) The heatmap of human traits significantly correlated with pig study based on TWAS ($P < 0.05$).

ated ($P = 1.2741 \times 10^{-11}$) with BFT. Additionally, the 'genetic correlation' panel reveal genetic correlations between the searched trait and other traits. For instance, BFT exhibit a significant genetic correlation with reproductive trait such as the number born alive (NBA), the total number of born (TNB), and the number born of healthy pigs (NBH). This observation supports that fatness has the potential to impact farrowing performance (47–49).

The 'Integrative Analysis' section presents the result of integrating regulatory variants from PigGTEx with GWAS of complex traits to identify the candidate causal genes for the searched trait (Figure 3C). The in-line panels display the plot and table showcasing significant results from TWAS, SMR and colocalization analyses between GWAS of complex traits and five types of molecular phenotypes/QTLs. For 'MetaGWAS_M_BFT', we detect that *MC4R* was a significant gene in TWAS ($P = 1.1408 \times 10^{-17}$) and SMR ($P = 1.0128 \times 10^{-6}$) with eQTL data in the frontal cortex. In addition, the expression of *ABCD4* is significantly associated with BFT in multiple tissues, especially in the small intestine ($P = 3.1800 \times 10^{-22}$ in TWAS, $P = 1.7848 \times 10^{-8}$ in SMR).

The section of 'Enrichment Analysis' presents the results of enrichment of GWAS signals of complex traits and different biological annotations (Figure 3D) including functional elements, chromatin states, molQTLs and tissues. We find that significant variants of metaGWAS in 'MetaGWAS_M_BFT' were significantly enriched in the functional element of 'CDS' (enrichment fold $= 1.3778$, $P = 9.9990 \times 10^{-5}$), chromatin states of 'weak active enhancer' in adipose (enrichment fold $= 2.0848$, $P = 9.9990 \times 10^{-4}$) and 'bivalent/poised TSS' in 12 out of 14 tissues. Both molQTL and heritability enrichment analysis showed that the tissue of 'blood' was highly relevant with BFT, with enrichment fold $= 1.4952$ ($P = 9.9990 \times 10^{-4}$) and 7.4395 ($P = 3.3016 \times 10^{-2}$), respectively.

The section titled 'Cross-species analysis' presents the results of comparative analyses of complex traits between humans and pigs to aid in the understanding of genetic similarities of complex traits between species. PigBiobank provides the Pearson's correlation of the absolute standardized effect sizes of homologous variants from GWAS summary. For comparison in GWAS, we obtained only eight traits of humans significantly correlated with BFT of pig at a low level (the absolute of Pearson's correlation < 0.06). By integrating PigGTEx resources with GWAS of complex traits, PigBiobank also provide the Pearson's correlation of the absolute standardized effect sizes of orthologous genes from TWAS summary for the matched tissue between species (Figure 3E). The comparison of TWAS results showed that BFT of pig was significantly correlated with 100 human traits in at least one tissue, with 14 traits in human being highly correlated ($r > 0.1500$) such as the trait of 'memory' ($r = 0.2019$, $P = 1.210 \times 10^{-3}$) and 'weight' ($r = 0.1740$, $P = 5.1600 \times 10^{-3}$).

Case study for gene and variant search module

To explore the regulatory mechanism of gene/variant towards complex traits, PigBiobank not only provide gene or vari-

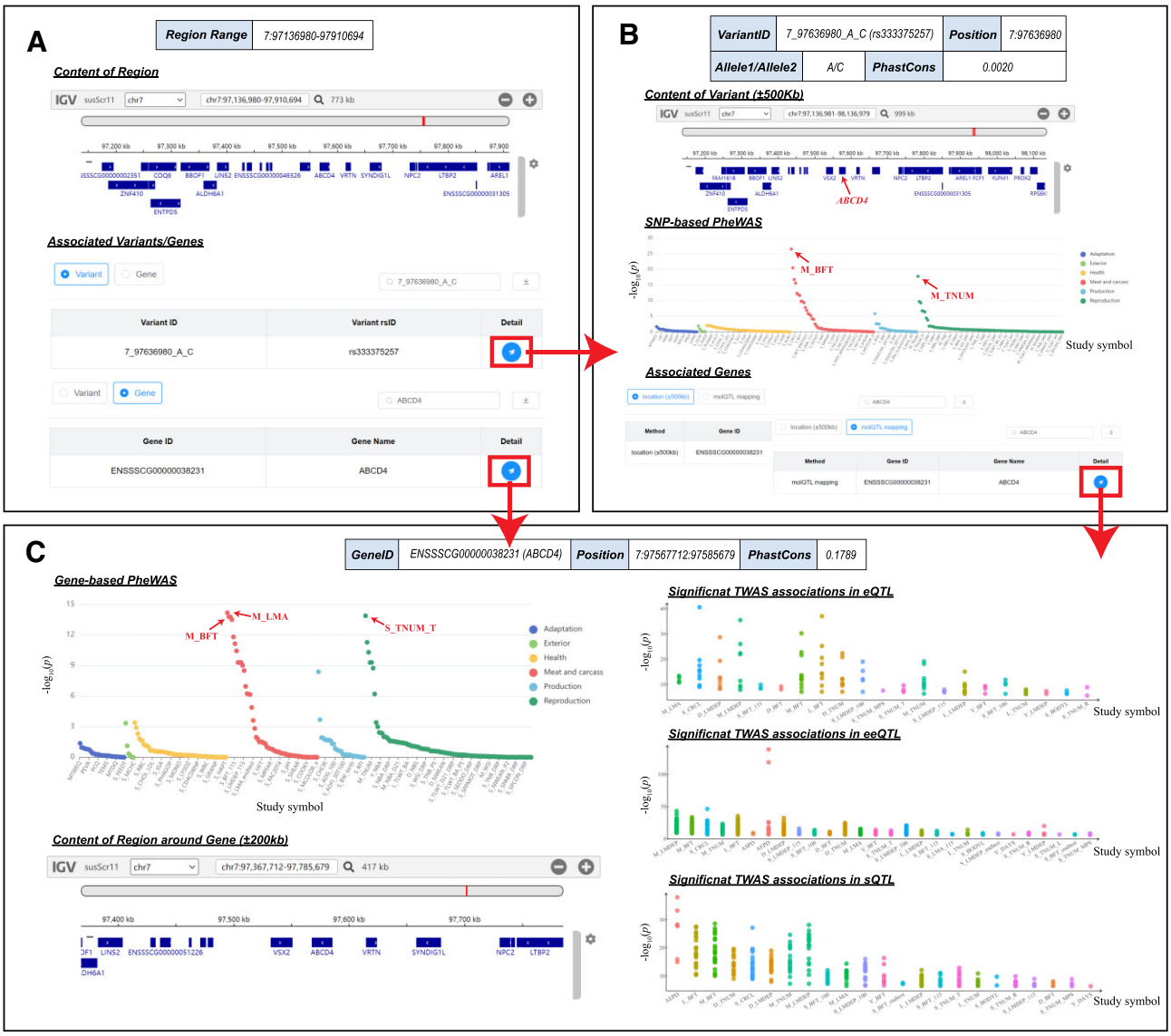


Figure 4. Case study of region search (SSC7: 97136980~97910694), variant search (rs333375257), and gene search (*ABCD4*). (**A**) The result of the region search, including the summary information of the target region, the genomics viewer of the region and the list of significant variants in GWAS within the target region, or genes overlapped within the target region. (**B**) The result of the variant search, containing the summary information of the target variant, the genomics viewer of the region (bp ±500 kb), and the scatter plot of SNP-based PheWAS and the list of associated genes based on physical location (±500 kb), or molecular QTL mapping from top to down, respectively. (**C**) The result of the gene search, containing the summary information of the target gene, gene-based PheWAS, the genomics viewer of the region around the gene (gene body ±200 kb), and scatter plot of significant TWAS in eQTL, eeQTL and sQTL.

ant search module enable users to perform an exact search, but also provide region search module to narrow down a genome region to candidate target genes or variants for users to perform a fuzzy search. We presents the QTL region (SSC7: 97136980–97910694), the lead SNP (rs333375257) and gene (*ABCD4*) as case studies to explore the usage of gene/variant search in the PigBiobank.

In the region search module, three sections are presented: (i) 'Summary of Region', (ii) 'Content of Region' and (iii) 'Associated Gene List' (Figure 4A). The sections of 'Summary of Region' and 'Content of Region' show the information of the genomic region being queried and genes within the genomic region. The section of 'Associated Variant/Gene List' shows a list of significant GWAS variants within the genomic region, and genes whose range of gene body overlapped with the region. For instance, the region of 'SSC7: 97136980–97910694' presents a total of 2324 significant GWAS variants and 16 genes in the third section.

In the variant search module, the following four sections are presented: (i) 'Summary of Variant', (ii) 'Content of Variant', (iii) 'Summary-based Analysis', and (iv) 'Associated Genes' (Figure 4B). The section of 'Summary of Variant' shows the detailed information of the variant such as the physical position, two alleles, and the phastCons score across 100 vertebrate species. The section titled 'Content of Variant' visualizes a series of genes located around the variant using IGV. The section of 'Summary-based Analysis' shows the result of PheWAS to explore which traits are associated with this genomic variant. As an example, the variant of 'rs333375257' was significantly associated with multiple meat and carcass traits (Figure

4B). In the section of 'Associated Genes', PigBiobank provides a user defined list of genes either by 'location' or 'molQTL mapping'. For instance, 21 and six genes can be obtained by querying 'rs333375257' for 'location' and 'molQTL mapping' search, respectively.

In the gene search module, the following four sections are presented: (i) 'Summary of Gene', (ii) 'Content of Region around Gene', (iii) 'Summary-based Analysis', and (iv) 'Integrative Analysis' (Figure 4C). The section of 'Summary of Gene' shows the information of queried gene including gene ensembl ID, gene name, the genomic range of gene body, and the phastCons score across 100 vertebrate species. The section of 'Content of Region around Gene' visualizes a series of genes around the queried gene using IGV. The section of 'Summary-based Analysis' presents the associations across all complex traits within the PigBiobank named as 'PheWAS' module. The section of 'Integrative analysis' provides diagrams and lists of associated gene-tissue-trait pairs based on TWAS, SMR, and COLOC. For instance, in PheWAS, *ABCD4* is found to be significantly associated with both the meat and carcass traits (e.g. backfat thickness) and the reproductive traits (e.g. total teat number) ($P < 5.0000 \times 10^{-8}$) (Figure 4C). In TWAS, we identify that *ABCD4* is significantly associated with a variety of traits across multiple tissues (Figure 4C). All these results illustrate the potential of PigBiobank in elucidating regulatory mechanism of complex traits, such as pleiotropy.

Discussion

To the best of our knowledge, the PigBiobank is the largest and most comprehensive database that integrates large-scale multi-omics data to deeply resolve the molecular and cellular mechanisms underlying complex traits in pigs. The PigBiobank is freely available to the public, and easy-to-accessible without logging in or registering. We would be anticipating more collaborators from around the world (pigbiobank@farmgtex.org). The first version of the PigBiobank, an encyclopedia of pig complex traits, curates a dataset of 71 885 pigs with genotypes and phenotypes from over 100 breeds, representing 264 distinct complex traits. The PigBiobank has three main features: (i) data standardization and sharing. The PigBiobank has undergone strict quality controls from raw data collection to data integrative analysis and adopted a uniform pipeline for analysis to ensure the quality of the data and the reliability of the results. Additionally, all the 264 traits in the PigBiobank have been classified into six main categories and 22 sub categories with unique and standardized indexes and names. The PigBiobank stored and shared detailed outcomes of the meta-GWAS and multi-omics integrative analysis; (ii) novel biological insights into the complex traits in pigs. By integrating the functional annotations collected from multiple consortia, such as FarmGTEx and FAANG projects, the PigBiobank provides a prototype for researchers to validate and annotate their GWAS findings, shedding light on the novel biological mechanisms underlying complex traits. Leveraging the resource of the PigBiobank, users could help identify candidate causal genes and variants underlying economically important traits in pigs, which will accelerate selective breeding to ensure food security for the growing global population in an environmentally sustainable way. (iii) Continuous update. Given that the GWAS data and molQTL data will be expanding and updated regularly in the coming years, we will keep the PigBiobank data updated continuously along with the PigGTEx project and functionality updated continuously annually. In the coming version, the PigBiobank will include more complex traits from more populations and breeds, as well as more diverse omics data. It will also provide user-friendly tools and a stable backend framework to support interactive and real-time analysis. We expect it will be a state-of-the-art, easy-to-use and open-access resource for biologically and genetically deciphering complex traits in pigs.

Data availability

The PigBiobank is freely available without registration at http://pigbiobank.farmgtex.org.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

The authors would like to thank the users, data providers and open-source contributors to the database. We thank National Supercomputer Center in Guangzhou China for its support in providing computing resources. We are grateful for the technical support provided by Guangdong iPig Technology Co. Ltd. We also appreciate the icons made by Surang, xnimrodx, Freepike, Becris, Marz Gallery, Smashicons, Iconic Panda, Pixel perfect, photo3idea_studio from www.flaticon.com, and Luciano Jung from www.iconfinder.com.

Author contributions: Haonan Zeng: Data curation, Visualization, Project administration, Writing-original draft. Wenjing Zhang: Formal Analysis, Validation, Writing-Original draft. Qing Lin: Formal Analysis, Writing-Original draft. Yahui Gao: Conceptualization, Writing-review & editing. Jinyan Teng: Formal Analysis, Visualization. Zhiting Xu: Formal Analysis. Xiaodian Cai: Formal Analysis. Zhanming Zhong: Formal Analysis. Jun Wu: Formal Analysis. Yuqiang Liu: Formal Analysis. Shuqi Diao: Formal Analysis. Chen Wei: Validation. Wentao Gong: Validation. Xiangchun Pan: Validation. Zedong Li: Investigation. Xiaoyu Huang: Investigation. Xifan Chen: Visualization. Jinshi Du: Visualization. Fuping Zhao: Resource. Yunxiang Zhao: Resource. Maria Ballester: Resource. Daniel Crespo-Piazuelo: Resource. Marcel Amills: Resource. Alex Clop: Resource. Peter Karlskov-Mortensen: Resource. Merete Fredholm: Resource. Pinghua Li: Resource. Ruihua Huang: Resource. Guoqing Tang: Resource. Mingzhou Li: Resource. Xiaohong Liu: Resource. Yaosheng Chen: Resource. Qin Zhang: Resource. Jiaqi Li: Resource, Funding acquisition. Xiaolong Yuan: Resource. Xiangdong Ding: Resource Conceptualization, Writing-review & editing. Lingzhao Fang: Resource, Conceptualization, Writing-review & editing. Zhe Zhang: Resource, Conceptualization, Funding acquisition, Supervision, Writing-review & editing.

Funding

National Natural Science Foundation of China [32022078]; National Key R&D Program of China [2022YFF1000900]; Local Innovative and Research Teams Project of Guangdong Province [2019BT02N630]; China Agriculture Research Sys-

tem [CARS-35]. Funding for open access charge: National Natural Science Foundation of China [32022078].

Conflict of interest statement

None declared.

References

1. FAO. (2022) *Meat Market Review: Emerging Trends and Outlook 2022*, Rome.
2. Pabst,R. (2020) The pig as a model for immunology research. *Cell Tissue Res.*, **380**, 287–304.
3. Lunney,J.K., van Goor,A., Walker,K.E., Hailstock,T., Franklin,J. and Dai,C. (2021) Importance of the pig as a human biomedical model. *Sci. Transl. Med.*, **13**, eabd5758.
4. Zhao,J., Ross,J.W., Hao,Y., Spate,L.D., Walters,E.M., Samuel,M.S., Rieke,A., Murphy,C.N. and Prather,R.S. (2009) Significant improvement in cloning efficiency of an inbred miniature pig by histone deacetylase inhibitor treatment after somatic cell nuclear transfer. *Biol. Reprod.*, **81**, 525–530.
5. Lind,N.M., Moustgaard,A., Jelsing,J., Vajta,G., Cumming,P. and Hansen,A.K. (2007) The use of pigs in neuroscience: modeling brain disorders. *Neurosci. Biobehav. Rev.*, **31**, 728–751.
6. Rogers,C.S. (2016) Genetically engineered livestock for biomedical models. *Transgenic Res.*, **25**, 345–359.
7. Buniello,A., MacArthur,J.A.L., Cerezo,M., Harris,L.W., Hayhurst,J., Malangone,C., McMahon,A., Morales,J., Mountjoy,E., Sollis,E., *et al.* (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, **47**, D1005–D1012.
8. Li,M.J., Liu,Z., Wang,P., Wong,M.P., Nelson,M.R., Kocher,J.-P.A., Yeager,M., Sham,P.C., Chanock,S.J., Xia,Z., *et al.* (2016) GWASdb v2: an update database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.*, **44**, D869–D76.
9. Watanabe,K., Stringer,S., Frei,O., Umićević Mirkov,M., Leeuw,C.d., Polderman,T.J.C., van der Sluis,S., Andreassen,O.A., Neale,B.M. and Posthuma,D. (2019) A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.*, **51**, 1339–1348.
10. Pan,S., Kang,H., Liu,X., Lin,S., Yuan,N., Zhang,Z., Bao,Y. and Jia,P. (2023) Brain Catalog: a comprehensive resource for the genetic landscape of brain-related traits. *Nucleic Acids Res.*, **51**, D835–D844.
11. Ghoussaini,M., Mountjoy,E., Carmona,M., Peat,G., Schmidt,E.M., Hercules,A., Fumis,L., Miranda,A., Carvalho-Silva,D., Buniello,A., *et al.* (2021) Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.*, **49**, D1311–D1320.
12. Maurano,M.T., Humbert,R., Rynes,E., Thurman,R.E., Haugen,E., Wang,H., Reynolds,A.P., Sandstrom,R., Qu,H., Brody,J., *et al.* (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science*, **337**, 1190–1195.
13. Chen,S., Liu,S., Shi,S., Jiang,Y., Cao,M., Tang,Y., Li,W., Liu,J., Fang,L., Yu,Y., *et al.* (2022) Comparative epigenomics reveals the impact of ruminant-specific regulatory elements on complex traits. *BMC Biol.*, **20**, 273.
14. Teng,J., Gao,Y., Yin,H., Bai,Z., Liu,S., Zeng,H., Bai,L., Cai,Z., Zhao,B., Li,X., *et al.* (2022) A compendium of genetic regulatory effects across pig tissues. bioRxiv doi: https://doi.org/10.1101/2022.11.11.516073, 11 November 2022, preprint: not peer reviewed.
15. Albert,F.W. and Kruglyak,L. (2015) The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.*, **16**, 197–212.
16. Umans,B.D., Battle,A. and Gilad,Y. (2021) Where are the disease-associated eQTLs?*Trends Genet.*, **37**, 109–124.
17. Hu,Z.-L., Park,C.A. and Reecy,J.M. (2022) Bringing the animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. *Nucleic Acids Res.*, **50**, D956–D961.
18. Guan,D., Bai,Z., Zhu,X., Zhong,C., Hou,Y., Lan,F., Diao,S., Yao,Y., Zhao,B., Zhu,D., *et al.*2023) The ChickenGTEx pilot analysis: a reference of regulatory variants across 28 chicken tissues. bioRxiv doi: https://doi.org/10.1101/2023.06.27.54667029, 29 June 2023, preprint: not peer reviewed.
19. Santhanam,N., Sanchez-Roige,S., Liang,Y., Chitre,A.S., Munro,D., Chen,D., Cheng,R., Nyasimi,F., Perry,M., Gao,J., *et al.* (2022) RatXcan: framework for translating genetic results between species via transcriptome-wide association analyses. bioRxiv doi: https://doi.org/10.1101/2022.06.03.494719, 05 June 2022, preprint: not peer reviewed.
20. Xu,Z., Lin,Q., Cai,X., Zhong,Z., Li,B., Teng,J., Zeng,H., Gao,Y., Cai,Z., Wang,X., *et al.* (2023) Integrating large-scale meta-GWAS and PigGTEx resources to decipher the genetic basis of complex traits in pig. bioRxiv doi: https://doi.org/10.1101/2023.10.09.561393, 11 October 2023, preprint: not peer reviewed.
21. Browning,B.L., Zhou,Y. and Browning,S.R. (2018) A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.*, **103**, 338–348.
22. Fick,S.E. and Hijmans,R.J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.*, **37**, 4302–4315.
23. Harris,I., Osborn,T.J., Jones,P. and Lister,D. (2020) Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data.*, **7**, 109.
24. Liu,Y., Liu,R. and Chen,J.M. (2012) Retrospective retrieval of long-term consistent global leaf area index (1981–2011) from combined AVHRR and MODIS data. *J. Geophys. Res.*, **117**, G04003.
25. Pan,Z., Yao,Y., Yin,H., Cai,Z., Wang,Y., Bai,L., Kern,C., Halstead,M., Chanthavixay,G., Trakooljul,N., *et al.* (2021) Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. *Nat. Commun.*, **12**, 5848.
26. Yang,J., Lee,S.H., Goddard,M.E. and Visscher,P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
27. Jiang,L., Zheng,Z., Qi,T., Kemper,K.E., Wray,N.R., Visscher,P.M. and Yang,J. (2019) A resource-efficient tool for mixed model association analysis of large-scale data. *Nat. Genet.*, **51**, 1749–1755.
28. Jiang,L., Zheng,Z., Fang,H. and Yang,J. (2021) A generalized linear mixed model association tool for biobank-scale data. *Nat. Genet.*, **53**, 1616–1621.
29. Willer,C.J., Li,Y. and Abecasis,G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
30. Zhou,X. and Stephens,M. (2012) Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.*, **44**, 821–824.
31. Chen,G.-B., Lee,S.H., Zhu,Z.-X., Benyamin,B. and Robinson,M.R. (2016) EigenGWAS: finding loci under selection through genome-wide association studies of eigenvectors in structured populations. *Heredity (Edinb)*, **117**, 51–61.
32. Leeuw,C.A.d., Mooij,J.M., Heskes,T. and Posthuma,D. (2015) MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.*, **11**, e1004219.
33. Bulik-Sullivan,B.K., Loh,P.-R., Finucane,H.K., Ripke,S., Yang,J., Patterson,N., Daly,M.J., Price,A.L. and Neale,B.M. (2015) LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.*, **47**, 291–295.
34. GTEx Consortium, Pividori,M., Rajagopal,P.S., Barbeira,A., Liang,Y., Melia,O., Bastarache,L., Park,Y., Wen,X. and Im,H.K.

(2020) PhenomeXcan: mapping the genome to the phenome through the transcriptome. *Sci. Adv.*, **6**, eaba2083.
35. Barbeira,A.N., Dickinson,S.P., Bonazzola,R., Zheng,J., Wheeler,H.E., Torres,J.M., Torstenson,E.S., Shah,K.P., Garcia,T., Edwards,T.L., *et al.* (2018) Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.*, **9**, 1825.
36. Barbeira,A.N., Pividori,M., Zheng,J., Wheeler,H.E., Nicolae,D.L. and Im,H.K. (2019) Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet.*, **15**, e1007889.
37. Zhu,Z., Zhang,F., Hu,H., Bakshi,A., Robinson,M.R., Powell,J.E., Montgomery,G.W., Goddard,M.E., Wray,N.R., Visscher,P.M., *et al.* (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.*, **48**, 481–487.
38. Cingolani,P., Platts,A., Le Wang,L., Coon,M., Nguyen,T., Wang,L., Land,S.J., Lu,X. and Ruden,D.M. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: sNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*, **6**, 80–92.
39. Finucane,H.K., Bulik-Sullivan,B., Gusev,A., Trynka,G., Reshef,Y., Loh,P.-R., Anttila,V., Xu,H., Zang,C., Farh,K., *et al.* (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.*, **47**, 1228–1235.
40. Consortium,T.G.T.E., Aguet,F., Anand,S., Ardlie,K.G., Gabriel,S., Getz,G.A., Graubert,A., Hadley,K., Handsaker,R.E., Huang,K.H, *et al.* (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318–1330.
41. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
42. Gel,B., Díez-Villanueva,A., Serra,E., Buschbeck,M., Peinado,M.A. and Malinverni,R. (2016) regioneR: an R/bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics*, **32**, 289–291.
43. Finucane,H.K., Reshef,Y.A., Anttila,V., Slowikowski,K., Gusev,A., Byrnes,A., Gazal,S., Loh,P.-R., Lareau,C., Shoresh,N., *et al.* (2018) Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.*, **50**, 621–629.
44. Robinson,J.T., Thorvaldsdottir,H., Turner,D. and Mesirov,J.P. (2023) igv.Js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics*, **39**, btac830.
45. Kim,K.S., Larsen,N., Short,T., Plastow,G. and Rothschild,M.F. (2000) A missense variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake traits. *Mamm. Genome*, **11**, 131–135.
46. Kim,K.-S., Reecy,J.M., Hsu,W.H., Anderson,L.L. and Rothschild,M.F. (2004) Functional and phylogenetic analyses of a melanocortin-4 receptor mutation in domestic pigs. *Domest. Anim. Endocrinol.*, **26**, 75–86.
47. Thongkhuy,S., Chuaychu,S.B., Burarnrak,P., Ruangjoy,P., Juthamanee,P., Nuntapaitoon,M. and Tummaruk,P. (2020) Effect of backfat thickness during late gestation on farrowing duration, piglet birth weight, colostrum yield, milk yield and reproductive performance of sows. *Livest. Sci.*, **234**, 103983.
48. Cheng,C., Wu,X., Zhang,X., Zhang,X. and Peng,J. (2019) Obesity of sows at late pregnancy aggravates metabolic disorder of perinatal sows and affects performance and intestinal health of piglets. *Animals*, **10**, 49.
49. Hu,J. and Yan,P. (2022) Effects of backfat thickness on oxidative stress and inflammation of placenta in large white pigs. *Vet. Sci.*, **9**, 302.

Received: August 14, 2023. **Revised:** October 13, 2023. **Editorial Decision:** October 24, 2023. **Accepted:** October 27, 2023