

RESEARCH

Open Access



Deciphering genetic characteristics of South China and North China indigenous pigs through selection signatures

Yahui Gao^{1†}, Xueyan Feng^{1†}, Shuqi Diao¹, Yuqiang Liu¹, Zhanming Zhong¹, Xiaotian Cai¹, Guangzhen Li¹, Jinyan Teng¹, Xiaohong Liu², Jiaqi Li¹ and Zhe Zhang^{1*}

Abstract

Background Indigenous pig breeds in China have accumulated significant genetic diversity due to regional selection pressures. Investigating the selection signatures of these populations helps to understand their adaptive evolution and contributes to genetic improvement programs.

Results We collected whole-genome sequencing data from 133 individuals, including South China and North China indigenous pigs and Asian wild boars. After data filtering, we retained 31,521,978 high-quality SNPs. Population structure analysis using PCA revealed distinct genetic clustering among these populations. Selection signature detection identified 5,227 loci under selection in South China indigenous pigs and 5,800 in North China indigenous pigs compared to Asian wild boars. Candidate genes were enriched in immune response pathways, reproductive traits, and pigmentation pathways. South China indigenous pigs exhibited selection signals for fat deposition and immune responses, while North China indigenous pigs showed stronger signals related to growth, blood physiology, and reproductive performance. Additionally, key genes such as *MC1R* and *KIT* were associated with coat color variation, and *IGF1R* and *IGF2R* were linked to growth regulation.

Conclusion Our results demonstrate that indigenous pigs in China have undergone selection for distinct traits aligned with their regional environments and farming systems. South China indigenous pigs have been selected for traits related to fat deposition and immunity, while North China indigenous pigs have been selected for growth and reproductive traits. The findings offer crucial insights into the genetic architecture of indigenous pig breeds, providing a valuable foundation for future genetic breeding programs.

Keywords China indigenous pigs, Selection signatures, Whole genome sequencing

[†]Yahui Gao and Xueyan Feng contributed equally to this work.

*Correspondence:

Zhe Zhang
zhezhang@scau.edu.cn

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

²State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, China



Introduction

Pigs serve as a consistent source of high-quality animal protein in human diets and play a vital role in human development and agricultural civilization. As one of the earliest domesticated animals in the shift from hunter-gatherer societies to agricultural civilizations, pigs probably underwent stages similar to other domesticated animals, including being hunted by humans, coexisting with humans, and eventually being domesticated [1–3]. Existing research has found that Eurasian wild boars started to diverge around one million years ago [4], with pig domestication beginning approximately 10,000 years ago [5]. It is now widely recognized that there are two independent centers of pig domestication globally, East Asia and the Near East [6], a conclusion supported by archaeological evidence [7–9]. China has a rich history of domestic pig breeding, with an abundance of indigenous pig resources. Over time, significant genetic variation has accumulated in traits and morphological characteristics, such as body size and skin color. This has led to the development of numerous indigenous pig breeds, each adapted to different environmental conditions, with varied body types and appearances, and distinct economic traits. Based on the current classification system, China indigenous pigs can be divided into six major types considering factors such as their origin distribution, body characteristics, and production performance: North China type, South China type, Jianghai type, Central China type, Southwest China type, and Plateau type [10]. North China indigenous pigs have medium-sized heads, black hair, large drooping ears, strong constitution, and adaptability to extensive farming, while the South type is smaller with sagging bellies and backs, thin skin, sparse hair, mostly black or black-and-white, and short broad bodies [10]. These two types differ in growth rate, feed efficiency, and lean meat rate. For instance, the weight and height of adult boars and sows of the Min pig breed, a North China type indigenous pig, were 227.10 ± 8.7 kg and 181.40 ± 10.27 kg, 89.10 ± 0.71 cm and 84.00 ± 0.42 cm, respectively [10]. For the breed of Luchuan pig, a South China type indigenous pig, the weight and height of adult boars and sows were 79.32 ± 2.94 kg and 78.52 ± 0.52 kg, 54.83 ± 0.81 cm, and 53.72 ± 0.18 cm, respectively [10]. Analyzing these differences aids in understanding the adaptive evolution history and breeding process, which provides a foundation and reference for genetic breeding work. At present, selection signature detection is a commonly used method that has been extensively applied in the genetic analysis of economic traits and the exploration of adaptive evolution in indigenous pigs and European commercial pig breeds.

In recent years, with the initiation and ongoing development of the Functional Annotation of Animal Genomes (FAANG) project [11], the FarmGTE_x project

[12], and the PigGTE_x project [13], researchers have gradually deepened their understanding and research from the genomic to the levels of gene expression, single-cell expression, and the regulation of functional elements within the genome. These projects have provided deeper biological insights into the tissue expression level and genetic regulatory mechanisms of important economic traits in pigs. By utilizing the advanced multi-omics findings from these livestock studies, the analysis of selection signature and the study of complex phenotypic regulation in livestock genomes are being advanced. In this study, we focused on the regions under selection and genetic variations affecting important economic traits in South China and North China indigenous pigs. Utilizing biological information from large-scale, multi-omics databases across different species, we performed signature mining analysis to thoroughly investigate the genetic basis under selection and the biological functions of key candidate genes in China indigenous pig populations.

Materials and methods

Sample collection and genotyping

We collected a total of 133 short-read whole-genome sequence (WGS) individual datasets from the PigGTE_x project [13]. This dataset included 63 South China indigenous pigs of three breeds, 40 North China indigenous pigs of four breeds, and 30 Asian wild boars (Table 1). Among them, the Luchuan pig and Guangdongxiaoerhua pig were two subgroups of the breed Liangguangxiaoerhua pigs, and the Tunchang pig and Ding'an pig were two subgroups of the breed Hainan pigs.

The initial genotype file contained 42,523,218 single nucleotide polymorphisms (SNPs). We used PLINK v1.90 [14] to perform the following data quality control on the initial genotype: (1) retained SNPs with a minor allele frequency > 0.01 using the command “--maf 0.01”; (2) retained SNPs on autosomes using the command “--autosome”. Finally, 11,001,240 SNP loci were filtered out, leaving 31,521,978 SNPs for subsequent analysis after quality control.

Population genetic structure analysis

Principal component analysis

We applied PLINK v1.90 [14] to conduct principal component analysis (PCA) on the study population and calculate the eigenvalues and eigenvectors of the first ten principal components using the command “--pca 10”. The scatter plot of the first two principal components was generated using the R package ggplot2 v3.3.6 [15].

Phylogenetic tree construction

To investigate the phylogenetic relationships among the study populations, we constructed a phylogenetic tree based on genetic distances. Using PLINK v1.90 [14], we

Table 1 Sample information on South China indigenous pigs, North China indigenous pigs and Asian wild boars

Population	Breed	Sample size	Location/origin
North China indigenous pig breeds	Min	16	Suihua City, Heilongjiang Province
	Laiwu	6	Laiwu City, Shandong Province
	Hetaodaer	7	Bayannur City, Inner Mongolia Autonomous Region
	Bamei	11	Yulin City, Shaanxi Province
South China indigenous pig breeds	Liangguangxiaerhua	17	Luchuan County, Yulin City, Guangxi Zhuang Autonomous Region
	Luchuan		
	Guangdongxiaerhua	10	Suixi County, Zhanjiang City, Guangdong Province
	Hainan	10	Tunchang County, Hainan Province
		10	Ding'an County, Hainan Province
	Bamaxiang	16	Bama County, Guangxi Zhuang Autonomous Region
Asian wild boars		30	/
Sum.		133	

calculated the identity by state (IBS) distances between individuals from 31,521,978 SNPs with the command "--genome", representing genetic distances as 1-IBS. We used MEGA v7.0.14 [16] for genetic distance file format conversion, and constructed the phylogenetic tree using the Neighbor-joining (NJ) method. The tree was then visualized with the iTOL v6.7.2 tool [17].

Linkage disequilibrium decay analysis

We utilized PopLDdecay v3.40 [18] to perform LD decay analysis on three populations. Firstly, we extracted single-chromosome data for each population using BCFtools v1.12 [19] to generate genotype input files in *.vcf.gz format. We divided SNPs within 1 Mb into intervals as follows: "10 bp intervals for distances within 500 bp and 100 bp intervals for distances over 500 bp", and calculated the average LD coefficient for all SNPs within these intervals. We then calculated LD per chromosome for a 1 Mb range with parameters "-MaxDist 1000 -bin1 10 -bin2 100 -break 500". Finally, we merged the single-chromosome results for genome-wide LD calculations and visualized the results by plotting the LD decay curves for multiple populations.

Population genetic structure analysis

We performed population genetic structure analysis using admixture v1.3.0 [20]. We predefined ancestors (K) from 2 to 8 and conducted cross-validation analysis to compare the reliability of each K.

Genome-wide detection of selection signatures

Cross-population extended haplotype homozygosity (XP-EHH)

We carried out genome-wide selection signature detection on pairwise groups of South China indigenous pigs, North China indigenous pigs, and Asian wild boars. Using PLINK v1.90 [14], we extracted single-chromosome genotype files and conducted XP-EHH analysis with Selscan v1.3.0 [21], merging the results for all

chromosomes. Genetic distance files for genome-wide sites were obtained by converting physical positions, with 1 cM assumed to be equal to 1 Mb [22]. We applied two-tailed tests to the XP-EHH results, with SNPs in the top 1% considered significant. SNPs with scores below the 0.5th percentile indicated selection in population A, while scores above the 99.5th percentile indicated selection in population B. In the groupings of South/North China indigenous pigs with Asian wild boars, the wild boars were treated as population A, and the indigenous pigs as population B. In the grouping of South China vs. North China indigenous pigs, the North China indigenous pigs were designated as population A, and the South China indigenous pigs as population B.

Pairwise fixation index (F_{ST})

We employed VCFtools v0.1.13 [23] to calculate the per-site F_{ST} statistics for three groups. Subsequently, we ordered the F_{ST} statistics for all loci in each group from highest to lowest, and SNP loci exceeding the top 0.1% quantile were considered significant loci detected by this method.

Genome-wide association study with eigenvector decomposition (EigenGWAS)

We utilized GEAR v0.919 [24] with default parameters to perform EigenGWAS analysis on South China and North China indigenous pigs, aiming to screen for selection signals of population differentiation. First, we extracted the first principal component information from three groups: "South China indigenous pigs vs. Asian wild boars", "North China indigenous pigs vs. Asian wild boars", and "South vs. North China indigenous pigs" as the input phenotype. We then conducted calculations for each chromosome file and merged the results from individual chromosomes to achieve genome-wide site analysis results. To enhance the statistical power of this selection signature analysis, we utilized GC (Genetic Correction)-adjusted P values (P_{GC}) as the indicator for significant

loci. We applied the Bonferroni correction method, using “0.05 / total number of loci” as the significance threshold.

Biological annotation of significantly selected SNPs

Definition of significantly selected SNPs and intervals

We defined significantly selected SNPs as those detected as significant by the XP-EHH method and simultaneously detected as significant by either the EigenGWAS or F_{ST} methods, i.e., $(\text{Sig_XP-EHH} \cap \text{Sig_}F_{ST}) \cup (\text{Sig_XP-EHH} \cap \text{Sig_EigenGWAS})$. Additionally, we defined a potential selective region as the interval extending 50 Kb upstream and downstream from the significantly selected SNPs. Genes and associated QTLs located in candidate regions based on their chromosomal physical positions (*Sus scrofa* 11.1, Release 100) were regarded as candidate genes for those regions.

QTL region enrichment analysis

To determine if significant SNPs were significantly enriched in specific trait types, we conducted QTL region enrichment analysis on significant loci from each group using gff files downloaded from Animal QTLdb (Release 45) [25]. The steps were as follows: (1) To ensure the reliability of the results, we removed QTLs from the gff files that were insignificant, lacked a clear physical location, or had QTL intervals > 1 Mb; (2) Trait classification was based on the Trait Type from Animal QTLdb (Release 45) [25], and we excluded trait categories with fewer than 100 reports; (3) We used custom R scripts to extract and create input files: bed files with the physical location information of significant loci; (4) We performed permutation tests using the R package *regioneR* v1.26.1 [26], with 10,000 permutations for each set of significant SNPs. The significance threshold for permutation tests was set at P value < 0.05.

Candidate gene pathway enrichment analysis

For a deeper understanding of the biological functions of candidate genes annotated by selection signatures in each group, we utilized the R package *clusterProfiler* v4.6.0 [27] and DAVID tool [28] for enrichment analysis, specifying the gene background species as pig (*Sus scrofa*). The significance criterion for enriched pathways was a P value < 0.05. Specifically, we focused on the analysis of Biological Process terms from the GO analysis and the KEGG pathway categories.

Chromatin state enrichment analysis

To examine the roles of significant loci in different tissues and functional genomic layers, we performed chromatin state enrichment analysis on significant selected loci. The chromatin state information was obtained from the FANNG project [11]. These 15 chromatin states for 14 pig tissues (adipose, cecum, cerebellum, colon, cortex,

duodenum, hypothalamus, ileum, jejunum, liver, lung, muscle, spleen, and stomach) were divided into six categories: promoter-associated states (TssA, TssAHet, TssBiv), states related to proximal transcription regions near Transcription Start Sites (TSS) (TxFlnk, TxFlnkWk, TxFlnkHet), enhancer-associated states (EnhA, EnhAMe, EnhAWk, EnhAHet, EnhPois), ATAC island regions (ATAC_Is), repressive states (Repr, ReprWk), and quiescent states (Quiescent). For the chromatin state enrichment analysis, we excluded the Quiescent state due to its inactivity. We performed enrichment analysis using the R package *LOLA* v1.22.0 [29], and the significance threshold was set to $P_{FDR} < 0.05$ and enrichment fold > 1.

Enrichment analysis of complex traits in pigs

To further investigate the relationship between significant selected loci in South China and North China indigenous pigs and complex phenotypic traits in domestic pigs, we used 268 GWAS meta-analysis data from the PigGTEX project [13] to conduct enrichment analysis on the significantly selected SNPs for pig complex traits. We used the R package *LOLA* v1.22.0 [29] to conduct the analysis and perform Fisher’s exact test, with the significance threshold set at $P_{FDR} < 0.05$ and enrichment fold > 1. Traits with odds ratio > 0 were plotted and converted to enrichment fold using the formula ‘log2(odds ratio + 1)’.

Multi-omics functional annotation analysis of candidate genes

We used large multi-omics databases such as HPA v22.0 [30], IMPC [31], GWASATLAS [32], and the PigGTEX project [13] to perform multi-level cross-species biological function annotation for key candidate genes, including transcriptomic expression levels in humans, pig, and mouse tissues, single-cell transcriptomic expression levels and protein expression levels in human, and associated phenotypes in human and mouse.

Results

Characteristics of the genome datasets

We collected 133 whole genome sequencing data of pigs from the pig genomics reference panel (PGRP) [13], including 30 Asian wild boars, 63 South and 40 North China indigenous pigs (Fig. 1a; Table 1). We aligned the clean reads to the *Sus scrofa* 11.1 reference genome [33] using BWA [34]. The sequence depth for each sample ranged from 5.24X to 69.40X, with an average of 19.94X (Table S1). Subsequently, we called SNPs on a population level and obtained 42,523,218 SNPs. After filtration of the raw variants, we kept 31,521,978 high-quality SNPs belonging to autosomes for subsequent analyses.

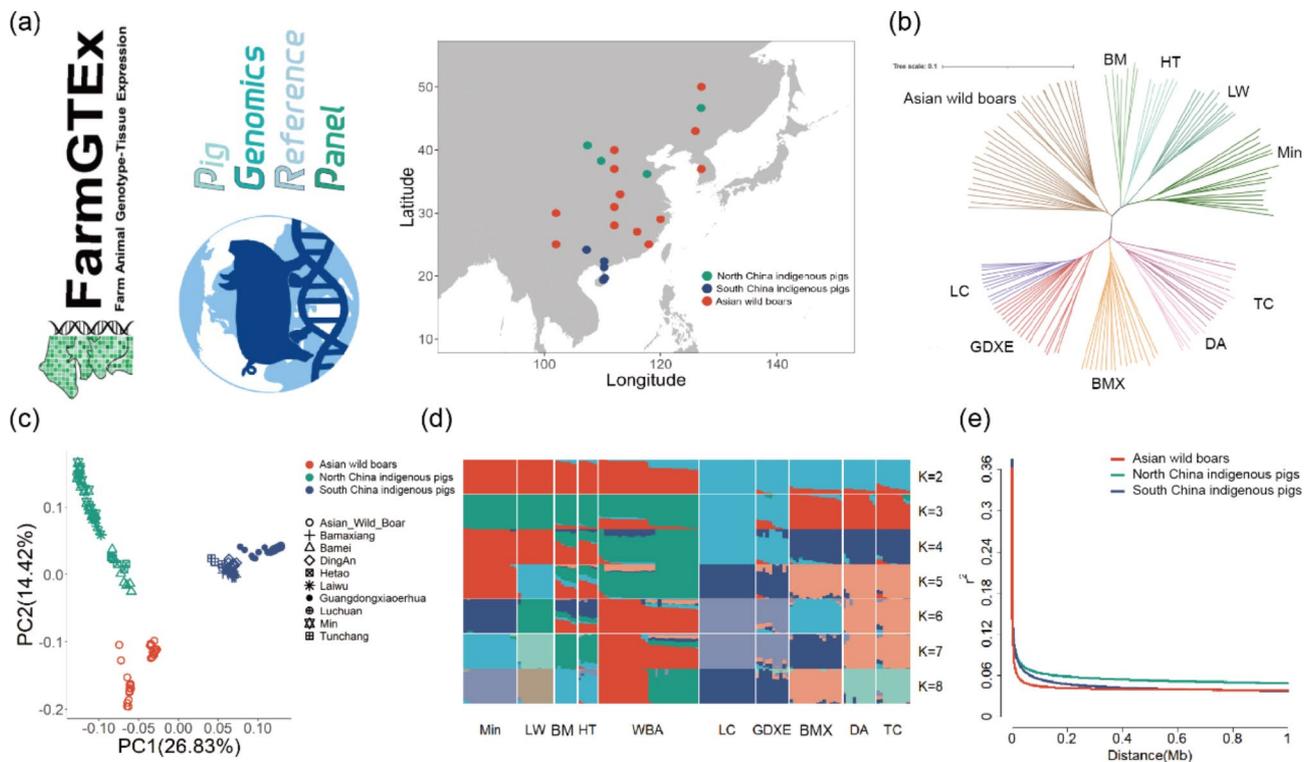


Fig. 1 Samples location and population genetic structures of South China and North China indigenous pigs and Asian wild boars. **a** Locations of the samples, which were collected from the PigGTEx project. **b** Neighbor-joining phylogenetic tree of 133 pigs. **c** Principal component analysis (PCA) result of 133 pigs on the first two PCs. **d** Genetic ancestry compositions with the assumed number of ancestries from $K=2$ to $K=8$. **e** Linkage disequilibrium decay in the distance of 1 Mb. LC, Luchuan pigs; GDXE, Guangdongxiaohua pigs; BMX, Bamaxiang pigs; DA, Ding'an pigs; TC, Tunchang pigs; LW, Laiwu pigs; HT, Hetao pigs; BM, Bamei pigs

Population structure analyses

The top two principal components (PCs) of the PCA contributed 26.83% and 14.42% genetic variance, respectively (Fig. 1c). PC1 and PC2 divided all the individuals into three groups defined by geographical distribution. Consistent with the relationships in the PCA, the Neighbor-joining tree showed that all individuals were clustered together according to their breeds (Fig. 1b). The assumed ancestral lineage compositions of all individuals were determined with a range of K values, where K represents the number of assumed ancestries (Fig. 1d). With an increasing K value, the populations within the three groups were gradually distinguished from each other. As depicted in Fig. 1e, the linkage disequilibrium (LD) decay among the three groups showed a similar trend, where r^2 decreased with increasing SNP marker distance, but the decay rates differed. Asian wild boars had lower LD levels and a faster decay rate than the other two populations. North China indigenous pigs exhibited a slower decay rate and higher LD levels compared to the South type. The maximum r^2 values were 0.3749 for South China indigenous pigs, 0.3737 for North type, and 0.3623 for Asian wild boars, with corresponding maximum LD distances of approximately 300 Kb, 200 Kb, and 50 Kb.

In addition, in preparation for using the EigenGWAS method in subsequent research, which requires population eigenvector files as input, we performed PCA on each pair of the three groups. As illustrated in Fig. S1, PC1 in each group successfully differentiated the corresponding two populations. Consequently, we extracted the PC1 values from these groups which will be used as input phenotype data for subsequent EigenGWAS.

Selection signatures in South China indigenous pig and Asian wild boar

Overview of selection signatures detection

In the XP-EHH test results for the grouping of South China indigenous pigs and Asian wild boars, the scores ranged from 1.2633 to 2.1607, with an average value of 0.1318. The 0.5% and 99.5% percentile XP-EHH values were -0.3793 and 1.1205 , respectively (Fig. S2a). Loci with XP-EHH values below the 0.5th percentile (XP-EHH score < -0.3793) were considered candidate loci under selection in Asian wild boars, while loci with XP-EHH values above the 99.5th percentile (XP-EHH score > 1.1205) were considered candidate loci under selection in South China indigenous pigs. There were 157,561 and 157,567 candidate loci detected in South China indigenous pigs and Asian wild boars, respectively,

with significant loci distributed across all chromosomes (Fig. 2a). The results of the F_{ST} test indicated that the F_{ST} statistics ranged from -0.0252 to 1 , with an average value of 0.0886 (Fig. S2a). In total, we detected $30,349$ significant loci ($F_{ST} > 0.7314$) (Fig. 2b). From the EigenGWAS findings, we discovered $7,651$ significant loci ($P < 1.6430 \times 10^{-9}$, Fig. S2), with peaks appearing on multiple chromosomes such as SSC1, SSC4, and SSC8 (Fig. 2c).

By using the XP-EHH method and confirming significance with either F_{ST} or EigenGWAS, we detected $5,227$ loci as significantly selected (Fig. 2d). These loci were unevenly distributed across the autosomes, with the strongest signature peaks found on SSC1 and SSC3. The highest concentrations of significant loci were on SSC1, SSC15, and SSC2, with counts of $1,721$, 745 , and 533 , respectively (Fig. 2d).

Biological annotation of significant selected loci

By extending 50 Kb both upstream and downstream from the significant selected loci to identify the selected regions, we conducted gene annotation for these regions. In South China indigenous pigs, we obtained 304 candidate genes, compared to 44 in Asian wild boars (Table S2). The pathway analysis indicated that the candidate genes in Asian wild boars were significantly enriched in a few pathways such as regulation of ventricular cardiac muscle cell membrane repolarization, cochlea development, integrin-mediated cell adhesion, prion diseases, and focal adhesion (Table S3). For South China indigenous pigs, candidate genes were significantly enriched in 21 GO biological processes and 17 KEGG pathways. Among these, the *DLX1* and *DLX2* genes were enriched in several brain nerve development pathways, including hippocampus development, fate commitment of GABAergic interneurons in the cerebral cortex, and subpallium development. Additionally, genes including *IKBKB*, *NFATC2*, *CD3E*, *CD3D*, *MAPK14*, and *NFATC2* were extensively enriched in pathways related to viral infection and immune response, such as Th1 and Th2 cell differentiation, C-type lectin receptor signaling pathway, and Chagas disease (Table S3).

QTL enrichment analysis for significant selected loci

Following the QTL enrichment analysis for significant selected loci in these two groups, it was observed that unlike Asian wild boars, which showed significant enrichment solely in blood parameters and meat texture traits ($P < 0.05$), South China indigenous pigs exhibited enrichment in five QTL trait categories: exterior, health, meat and carcass, production, and reproduction. Notably, in the health category, enrichment was noted in QTLs associated with disease resistance and immune capacity traits. In the reproduction category, the QTLs

related to reproductive traits, reproductive organs, and litter performance (Fig. 2f, Table S4).

Chromatin state enrichment analysis for significant selected loci

The enrichment revealed that differences between the two groups were concentrated in visceral organs, the digestive system, and cerebellum tissues. In Asian wild boars, selected loci were significantly enriched in enhancer regions of visceral organs (liver, spleen, lungs), with extreme enrichment in the weakly active enhancer chromatin state in the liver ($P_{FDR} < 0.001$, Table S5). For spleen, Asian wild boars showed significant enrichment in poised enhancer functional elements, while South China indigenous pigs were significantly enriched in weakly repressive Polycomb regions (ReprWk) ($P_{FDR} < 0.01$). Also, South China indigenous pigs' selected loci were enriched in the ATAC island state in cerebellar tissue and repressive states in the cecum and colon. Additionally, in the ileum, both were significantly enriched in weakly repressive Polycomb regions (ReprWk) and weak enhancer (EnhAWK) states, respectively (Fig. 2e).

Complex trait enrichment analysis for significant selected loci

In comparison to Asian wild boars, South China indigenous pigs showed significant enrichment in multiple complex traits within the growth, reproduction, and fat trait categories (Fig. 2g, Table S6). These included significant enrichment in growth traits such as days to reach 115 kg (DAYS_115) and average daily gain (ADG) ($P < 0.001$). In terms of reproductive traits, it mainly included the total number of piglets born (TNB) and total litter weight at weaning (TLWT_Weaning) ($P < 0.001$). There was also significant enrichment in backfat thickness (BFT) ($P < 0.001$).

Selection signatures in North China indigenous pig and Asian wild boar

Overview of selection signatures detection

The test results between this comparison group showed XP-EHH scores ranging from -1.0273 to 1.8237 , with an average of 0.0425 . The XP-EHH values at the 0.5% and 99.5% quantiles were -0.4107 and 0.7443 , respectively (Fig. 2). Loci with XP-EHH values less than the 0.5% quantile (XP-EHH score < -0.4107) were candidate selection loci in Asian wild boars, whereas those greater than the 99.5% quantile (XP-EHH score > 0.7443) were candidate in North China indigenous pigs. We found $157,530$ and $157,528$ significant loci in North China indigenous pigs and Asian wild boars. These loci were unevenly distributed across the chromosomes, with distinct selection signatures on SSC1, SSC5, SSC6, and SSC8 in the North China indigenous pigs (Fig. 3a). According to the F_{ST} test, the statistics ranged from -0.0300 to 0.9709 , with an

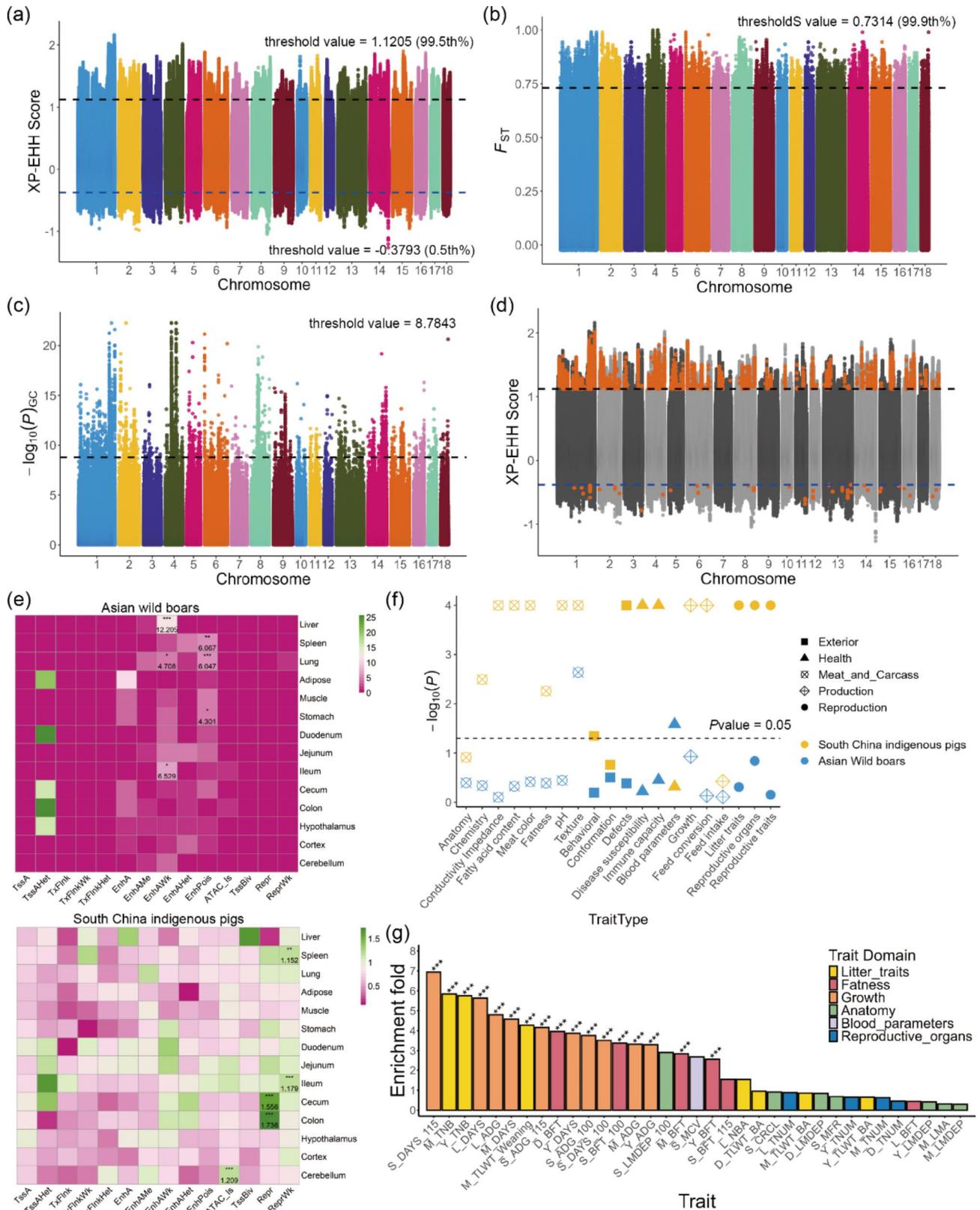


Fig. 2 Selection signatures and biological annotation of the paired South China indigenous pigs and Asian wild boars. **a-c** Manhattan plots of the selection signatures detected by XP-EHH, F_{ST} , and EigenGWAS. **d** Significant selected loci distribution in the whole genome. **e** Chromatin state enrichment analysis for significant selected loci in 14 major tissues of pigs. **f** QTL region enrichment analysis for the selected region. **g** Enrichment analysis of complex traits based on the selected SNP windows in pigs. The whole trait name showed in table S6

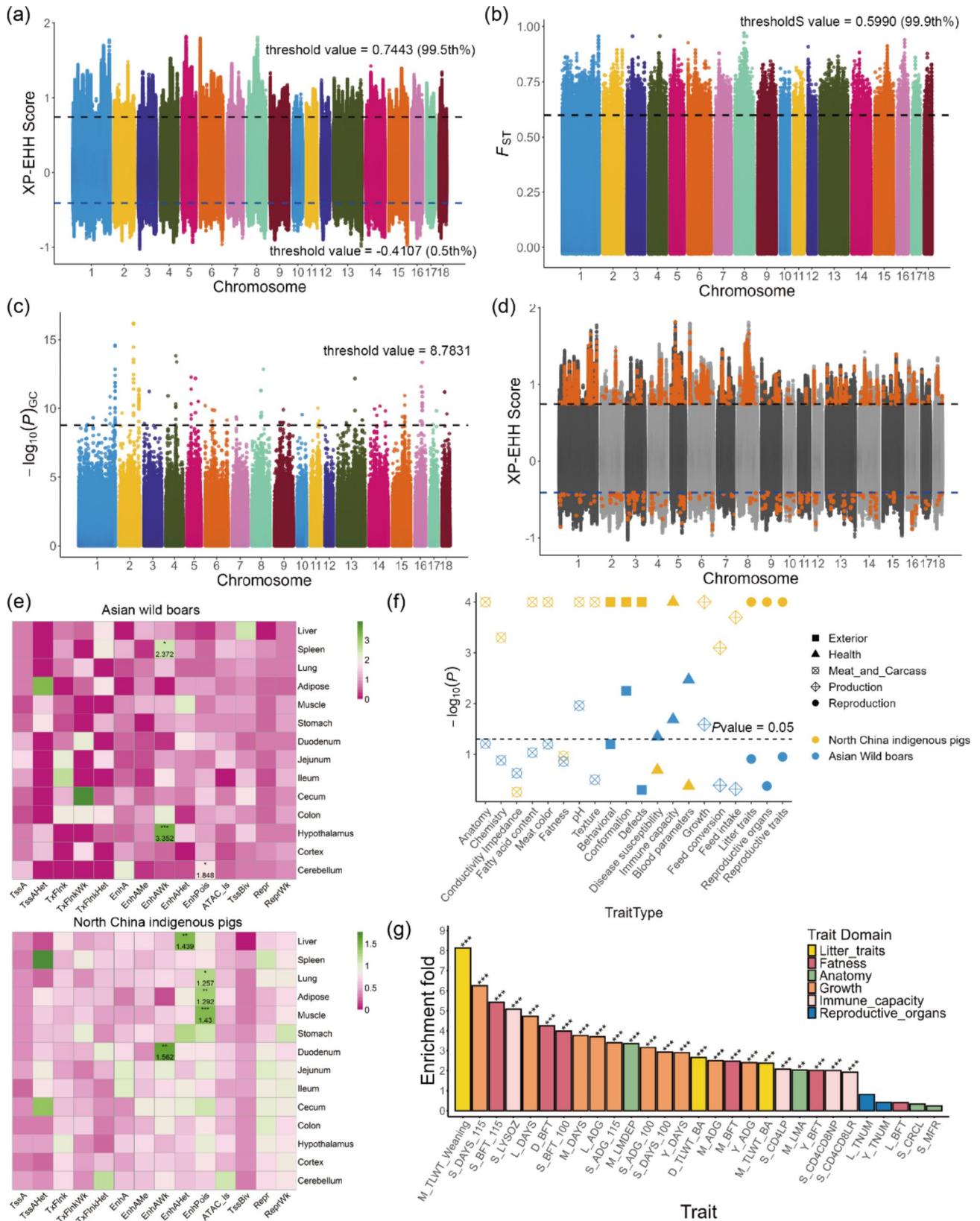


Fig. 3 Selection signatures and biological annotation of the paired North China indigenous pigs and Asian wild boars. **a-c** Manhattan plots of the selection signatures detected by XP-EHH, F_{ST} , and EigenGWAS. **d** Significant selected loci distribution in the whole genome. **e** Chromatin state enrichment analysis for significant selected loci in 14 major tissues of pigs. **f** QTL region enrichment analysis for the selected region. **g** Enrichment analysis of complex traits based on the selected SNP windows in pigs. The whole trait name showed in table S6

average of 0.0670. The F_{ST} value at the 0.1% highest quantile was 0.5990 and we detected 30,346 significant loci using this method ($F_{ST} > 0.5990$) (Fig. 3b). The EigenG-WAS results showed that $-\log_{10}P_{GC}$ values ranged from 0 to 16.1805, with a mean value of 0.4223, and 191 significant loci were detected ($P_{GC} = 1.6476 \times 10^{-9}$). The most significant signature was found on SSC2 with peaks also present on SSC1, SSC2, and SSC16 (Fig. 3c).

Finally, we found 5,800 significant selection loci in the North China indigenous pig. These loci were distributed across all 18 chromosomes, showing an uneven distribution (Fig. 3d). The locus with the highest XP-EHH score was located on SSC5. SSC1 and SSC8 had the highest numbers of significant selection loci, with 1,305 and 1,526 loci, respectively. In the Asian wild boar group, 489 significant selection loci were detected.

Biological annotation of significant selected loci

In the regions showing selection signatures, we identified 363 and 157 candidate genes in North China indigenous pigs and Asian wild boars, respectively (Table S7). Pathway analysis of candidate genes indicated that those from the Asian wild boar group were significantly enriched in seven GO biological process terms and five KEGG pathways (Table S8). In North China indigenous pigs, candidate genes were significantly enriched in 26 GO biological process terms and 34 KEGG pathways, mainly related to immune response, inflammatory response, and viral infection. Among these, *RNF114*, *GAL3ST1*, *LIMK2*, *KIT*, *MAEL*, *SPATA2*, *AP3B1*, *CABS1*, *PATZ1*, and *ACVR2A* genes were significantly enriched in the spermatogenesis pathway, *GNAQ*, *KCNMA1*, *ITPR2*, *ITPR3*, *LYZ*, and *ADRA1A* genes were enriched in salivary secretion, and *MC1R*, *KIT*, and *SNAI2* genes were significantly enriched in the pigmentation pathway (Table S8).

QTL enrichment analysis for significant selected loci

The selection signatures in the North China indigenous pig population were significantly enriched in QTLs related to production trait class (growth, feed conversion ratio, and feed intake), reproduction traits class (reproductive traits, reproductive organs), and meat and carcass class (fatty acid content and meat color) (Fig. 3f). On the other hand, the significant signatures in Asian wild boars were predominantly enriched in QTLs associated with immunity and health.

Chromatin state enrichment analysis for significant selected loci

As depicted in Fig. 3e, the significant selection signatures in Asian wild boars were significantly enriched in the enhancer states of tissues such as the hypothalamus, cerebellum, and spleen. Meanwhile, the significant selection signatures in North China indigenous pigs were

prominently enriched in the enhancer states of visceral tissues (liver and lungs), muscle, fat, and duodenum (Fig. 3e, Table S5).

Complex trait enrichment analysis for significant selected loci

Selection signatures in North China indigenous pigs were enriched in various complex traits related to growth, reproduction, immunity, and fat characteristics (Fig. 3g, Table S6). Specifically, for growth traits, there was significant enrichment in traits like DAYS_115, ADG, and BFT ($P < 0.001$). For immune traits, significant enrichment was observed in lysozyme levels and the percentage of CD4-positive leukocytes ($P < 0.001$). TLWT_Weaning showed the highest level of enrichment. In contrast, the selection signatures in Asian wild boars were significantly enriched in the phenotype of the number of teats (TNUM) ($P < 0.001$, Table S6).

Selection signatures in South China and North China indigenous pig

Overview of selection signatures detection

In this comparison group, XP-EHH scores ranged from -1.1870 to 2.0498 , with an average of 0.0778 . The XP-EHH values at the 0.5% and 99.5% quantiles were -0.4868 and 0.8738 , respectively. Loci with XP-EHH values below the 0.5% quantile (XP-EHH score < -0.4868) were candidate selection loci in North China indigenous pigs, while those above the 99.5% quantile (XP-EHH score > 0.8738) were candidate in South types. We identified 157,569 and 157,567 significant loci in South China and North China indigenous pigs, respectively. These loci were distributed across every chromosome, with significant ones on SSC1, SSC4, and SSC12 in South China indigenous pigs, and on SSC8 and SSC11 in North types (Fig. 4a). The F_{ST} statistics ranged from -0.0209 to 0.9701 , with an average value of 0.1017 . The F_{ST} statistic at the 0.1% highest quantile was 0.7265 . This method identified 30,612 significant loci with F_{ST} values exceeding 0.7265 (Fig. 4b). The EigenG-WAS results showed $-\log_{10}P_{GC}$ values ranging from 0 to 23.9567 , with a mean value of 0.4456 . There were 2,165 significant loci with P_{GC} values greater than the threshold ($P_{GC} = 1.6331 \times 10^{-9}$), with significant peaks appearing on SSC1, SSC4, SSC14, and SSC16 (Fig. 4c).

We detected 3,527 significant selection loci in South China indigenous pigs, distributed unevenly across the genome. Significant selection signatures were particularly evident on SSC1, SSC3, and SSC12, with SSC1 having the most significant one, totaling 1,285 (Fig. 4d). In North types, 958 significant selection loci were detected, with distinct selection signals on SSC1, SSC2, and SSC6. The highest numbers of significant loci were on SSC6 and SSC2, with 173 and 109 loci, respectively. The loci with the highest XP-EHH values were located on SSC8 (Fig. 4d).

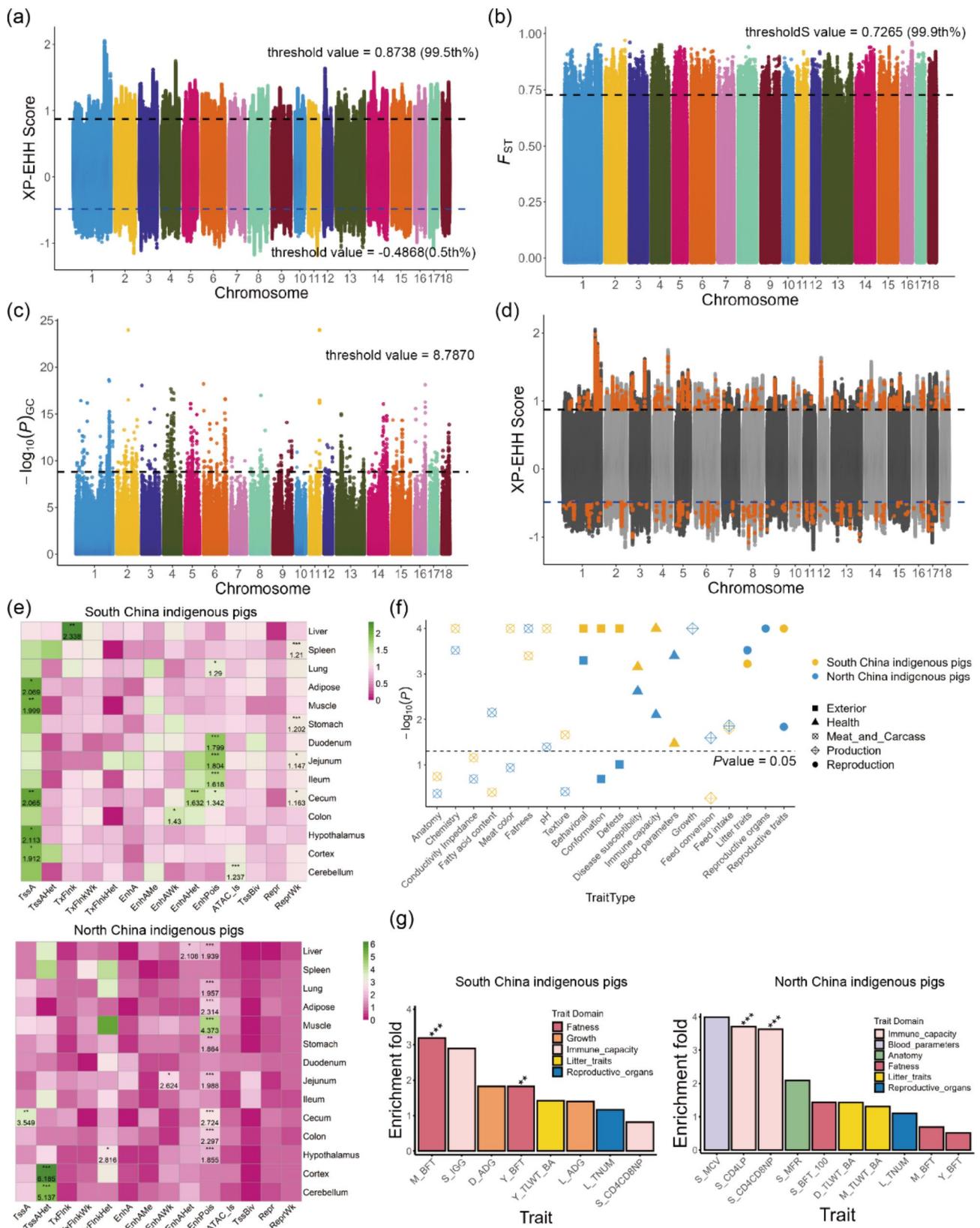


Fig. 4 Selection signatures and biological annotation of the paired South China indigenous pigs and North China indigenous pigs. **a-c** Manhattan plots of the selection signatures detected by XP-EHH, F_{ST} , and EigenGWAS. **d** Significant selected loci distribution in the whole genome. **e** Chromatin state enrichment analysis for significant selected loci in 14 major tissues of pigs. **f** QTL region enrichment analysis for the selected region. **g** Enrichment analysis of complex traits based on the selected SNP windows in pigs. The whole trait name showed in table S6

Biological annotation of significant selected loci

We annotated a total of 243 genes in South China indigenous pigs (Table S9), which were significantly enriched in 15 biological process pathways and one KEGG pathway, predominantly involved in complex developmental differentiation processes (Table S10). In North China indigenous pigs, 175 genes were annotated. These genes were enriched in 21 significant biological process pathways and 25 significant KEGG pathways, primarily linked to metabolic regulation (Table S10).

QTL enrichment analysis for significant selected loci

The enrichment analysis revealed that South China indigenous pigs' significant loci were enriched in meat quality, carcass, and appearance categories (Fig. 4f). In comparison, North China indigenous pigs' significant loci were more enriched in production and reproductive traits. For production traits, both populations showed enrichment in QTLs related to feed intake. Moreover, North China indigenous pigs' selection loci were enriched in QTLs related to growth traits and feed conversion efficiency (Fig. 4f).

Chromatin state enrichment analysis for significant selected loci

The results showed selection signatures enriched in the enhancer regions of the digestive system (stomach, small and large intestines) (Fig. 4e, Table S5). In South China indigenous pigs, significant selection signals were enriched in the EnhPois state of the small intestine (duodenum, jejunum, ileum) and cecum, and also enriched in the repressive functional regions of the stomach, jejunum, and spleen. The regulatory states in cecum tissue were diverse, with enrichment in TssA, EnhAHet, and ReprWk states. Additionally, there were differences between South China and North China indigenous pigs in the chromatin region enrichment of brain, muscle fat, and liver. Significant loci in South China indigenous pigs were enriched in the TssA of the cerebral cortex and hypothalamus and in the ATAC_Is of the cerebellum. In North China indigenous pigs, significant enrichment was found in the EnhPois and TxFlnkHet of the hypothalamus and in the TssAHet of the cerebellum and cerebral cortex. Both populations showed specific enrichment in the TssA and EnhPois states of adipose and muscle tissues, respectively. In the liver, the selected loci of South China indigenous pigs were enriched in TxFlnk, while those of North types were enriched in EnhAHet and EnhPois (Table S5). These differences in chromatin state enrichment in various tissues may be related to the different selection objectives and mechanisms for fat deposition, growth, and immune performance between these two populations.

Complex trait enrichment analysis for significant selected loci

The enrichment analysis for complex traits in pigs indicated that the most significant differences in selection direction between South China and North China indigenous pigs were observed in fat traits and immune traits (Fig. 4g). The candidate loci in South China indigenous pigs were enriched in the BFT trait, while those in North types were enriched in the CD4LP and CD4CD8NP traits (Fig. 4g).

Functional annotation analysis of genes under selection in indigenous pigs

For a more detailed analysis of the biological functions of selection signature in South China and North China indigenous pigs, all genes identified in the same type of pig population across three groups, South vs. North China indigenous pigs, South/North China indigenous pigs vs. Asian wild boars, were considered potential selected genes for that population. Repeatedly detected genes were classified as key selected genes.

Pathway analysis of selected genes in South China indigenous pigs

There were 439 potentially selected genes in the South China indigenous pig, of which 108 were key selected genes (Fig. 5a). These pathways, in which these genes were significantly enriched, formed clusters of biological function pathway networks, which are mainly associated with the activation, proliferation, and differentiation of immune cells (T cells, lymphocytes, leukocytes) as well as with the regulation of neural cell development and cell differentiation (Fig. 5b, Table S11). The enriched pathways for key selected genes were involved in functions such as cell development and metabolism, immune response, and others. *ABCA1*, in particular, was enriched in pathways that stabilize vascular endothelial cells and were associated with anti-atherosclerosis (Table S12).

Pathway analysis of selected genes in North China indigenous pigs

There were 503 potentially selected genes in the North China indigenous pig, with 35 identified as key selected genes (Fig. 5a). Clustering of functional pathway networks in which genes potentially under selection in North China indigenous pigs were significantly enriched showed that they were mainly associated with biological pathways related to blood circulation process, regulation of organism growth and development, and immune response regulation (Fig. 5c, Table S13), as well as reproductive physiological developmental pathways such as maternal processes during pregnancy, follicular development, and spermatogenesis (Table S11). The pathway enrichment result for key selected genes was shown in Table S11, with five genes (*LYN*, *KCNMA1*, *ITPR2*,

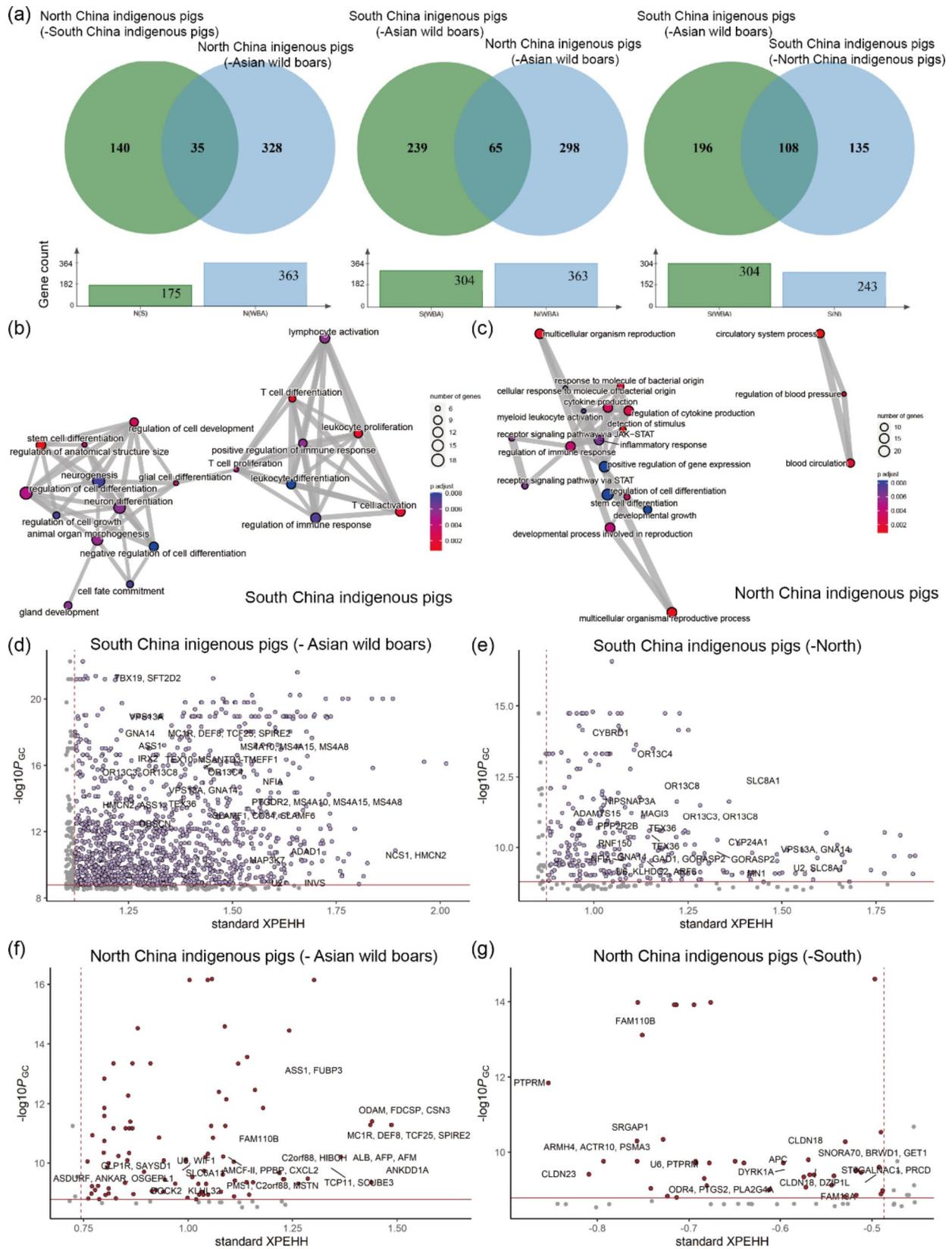


Fig. 5 The selected genes and their biological functions of South China and North China indigenous pigs. **a** Venn diagrams of the genes detected in different tests. **b-c** Network of the biological function terms where the selected genes in South China and North China indigenous pigs were enriched. **d-g** Key selected genes distribution in different tests

PLA2G4A, *PTGS2*) significantly enriched in six KEGG pathways, associated with neural cell signaling, immune response, and cardiovascular functions (Table S12).

Pathway analysis of shared selected genes in China indigenous pigs

In the selection signature detection analysis with Asian wild boars as the control group and indigenous pigs as the observation group, we identified 65 shared selected genes in both South China and North China indigenous pigs (Fig. 5a). These shared candidate genes were enriched in six GO biological processes and two KEGG pathways (Table S12), with several interferon (IFN) and interleukin (IL) gene family members enriched in pathways related to immune regulation processes and cytokine signaling.

Screening of differentially selected genes in China indigenous pigs

To delve deeper into the differential selection signals between South China and North China indigenous pigs, we kept the differential loci detected by XP-EHH, F_{ST} , and EigenGWAS, and defined candidate genes annotated with more than 10 of these loci as strongly selected genes. We found 33 and 15 strongly selected candidate genes in South China indigenous pigs (Fig. 5d-e, Fig. S3) and North China indigenous pigs (Fig. 5f-g, Fig. S3), respectively, of which 11 genes (*ASS1*, *FUBP3*, *MC1R*, *DEF8*, *TCF25*, *ODAM*, *C2orf88*, *FDCSP*, *CSN3*, *DOCK2*, *SPIRE2*) were commonly annotated as strongly selected (Table S14, Fig. 5d-g). These distinct genes were connected to different physiological functions, such as reproductive physiology, brain neurodevelopment, coat color, and immunity.

The transcriptome expression profiles from the HPA database indicated that genes such as *VPS13A* and *TEX36* were associated with male infertility in humans, with both showing specific high expression levels in the testicular tissue of pigs and humans. *EDRF1* was also highly expressed in the testicular tissue of pigs and humans, with single-cell transcriptome data showing significant enrichment in early and late-stage sperm cell clusters, suggesting a role in spermatogenesis. *ESR1* demonstrated high tissue-specific expression in reproductive organs such as the cervix and fallopian tubes (Fig. S4a-d). *CSN3* was involved in mammalian lactation, being expressed only in the salivary glands and mammary tissue, with particularly high expression in mammary tissue (Fig. S5a-c). Similarly, the pig transcriptome atlas showed that the *CSN3* gene had specific high expression in the lactating tissue of pigs (Fig. S5b).

NFIA, *BRWD1*, and *ST18* were associated with brain neurodevelopment. For example, *NFIA* was linked to brain development, brain malformations, and lethality in

mice. *BRWD1* was specifically enriched in human brain tissue and highly expressed in porcine fetal development tissues such as oocytes and blastomeres. *ST18* showed transcriptome-level expression only in human brain tissue but was highly expressed in various pig brain tissues.

The *MC1R* gene, associated with coat color, had an average expression level of TPM > 1 in pig brain tissues (frontal cortex, cerebrum, hypothalamus) (Fig. S5d-h). In humans, it showed tissue-specific high expression in the pituitary gland and testicular tissues (Fig. S5e). Additionally, in phenotype association tests, *MC1R* was linked to abnormal hair and hair pigmentation phenotypes in mice (Fig. S5f) and various skin and hair color phenotypes in humans (Fig. S5g).

Additionally, we identified several genes associated with mammalian immune function, such as *FDCSP*, *PIK3API*, and *DOCK2*. *FDCSP* and *ODAM* exhibited tissue-specific expression in lymphoid tissues and salivary glands and were linked to metabolic traits in humans (Fig. S4e-h). *PIK3API* showed high expression in lymphoid tissues, liver, and salivary glands in both human and pig transcriptome profiles and was associated with immune and metabolic complex traits in humans, and with increased neutrophil and monocyte counts and decreased lymphocyte counts in mice. *DOCK2* was expressed in various pig tissues, with higher expression levels in fetal thymus, lymph nodes, spleen, macrophages, and blood (Fig. 6). In humans, *DOCK2* had specific high expression in bone marrow, lungs, and lymphoid tissues (Fig. 6b), was specifically enriched in immune response clusters in lymphoid tissues (Fig. 6d), and was associated with decreased bone density and increased spleen weight phenotypes in mice (Fig. 6e).

Image credit: a, PigGTEEx-Portal [13], <http://piggtex.far.mgtex.org/>. b-d, Human Protein Atlas [30], www.protein-atlas.org. e, International Mouse Phenotyping Consortium [31], www.mousephenotype.org.

Discussion

Based on the WGS data, we employed three methods, i.e., F_{ST} , XP-EHH, and EigenGWAS, to detect genome-wide selection signatures in South China and North China indigenous pig breeds. We annotated the significant selected loci and functional genes using enrichment of chromatin state, QTL region, complex trait, and pathway enrichment. The results indicated that China indigenous pigs have been positively selected for traits related to feeding habits and feed conversion efficiency. Moreover, different types of indigenous pigs had distinct breeding directions: South China indigenous pigs showed selection signatures concentrated in traits related to fat deposition, meat quality, body shape, and immune function, while North China pigs exhibited signals related to

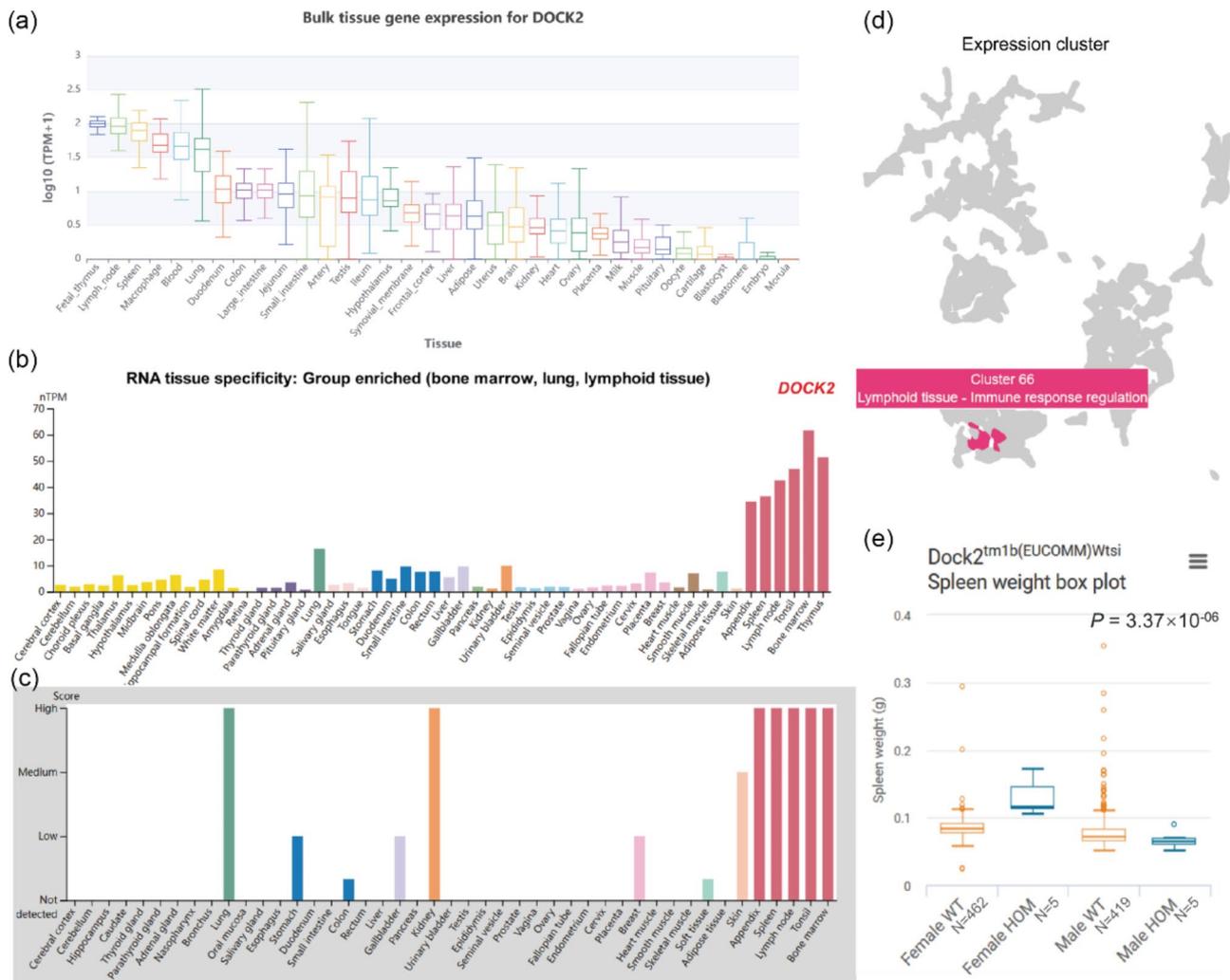


Fig. 6 Gene expression and biological phenotypes regulated by immune-associated gene *DOCK2* in mammals. **a** RNA expression overview across tissues on pigs. **b-c** RNA expression and protein level overview across tissues in humans. nTPM, normalized expression levels. Color coding is based on tissue groups, each consisting of tissues with functional features in common. **d** *DOCK2* is expressed specifically in the cluster Lymphoid tissue - Immue response regulation. **e** Related phenotype: increased spleen weight on *Dock2*^{-/-} female mice, compared with WT female mice

growth and development, blood physiology, and reproductive performance.

Firstly, by integrating the results from PCA, phylogenetic trees, and ancestry detection, we observed that North China indigenous pigs were genetically closer to Asian wild boars. This was hypothesized to be related to the geographical proximity of the Asian wild boar populations and the occurrence of gene flow events between them.

Secondly, through signature detection analysis of pairwise combinations among three groups, we identified candidate genes shared among China indigenous pigs, primarily associated with coat color, such as *MC1R*, *EDNRB*, and *KIT*. The *MC1R* gene was annotated in the SSC6: 0.120-0.28 Mb region in both South China and North China indigenous pigs. This gene was shown in humans to have multiple allelic mutations and was widely

reported to be associated with pigmentation, skin color [35], and susceptibility to melanoma [36]. Additionally, polymorphic mutations in this gene were related to changes in melanin synthesis in cattle coat color [37], horse coat color [38], chicken plumage color [39], and mice [40]. In pigs, Kijas et al. [41] were the first to reveal the role of *MC1R* variations in pig coat color diversity. These findings highlighted that *MC1R* played a central role in regulating the synthesis of eumelanin (black/brown) and pheomelanin (red/yellow) in mammalian melanocytes. Recent researchers used genetic engineering to create *MC1R* gene-edited pigs [42], manipulating pig coat color to cater to future consumer demands in the meat market, and evaluating the breeding of new pig breeds. We annotated the *KIT* gene in the SSC8: 41.46-41.56 Mb region in North China indigenous pigs. This gene was reported to have a clear association with the

white coat color phenotype in indigenous pigs [43] and was sensitive to melanocytes involved in the pigmentation of the epidermis and hair follicles [44]. Additionally, this gene was found to influence *MC1R* expression and was associated with the increased white spotting phenotype in horses [45]. Furthermore, we annotated the *EDNRB* gene in the SSC11: 50.04-50.14 Mb region in South China indigenous pig populations. Ai et al. [46] revealed through analysis of the “two-end-black” coat color in indigenous pig populations that *EDNRB* might be associated with the appearance of white coat color in indigenous pigs. This gene was also enriched in pigmentation and melanocyte differentiation pathways in Pudong White pigs [47]. In other mammals, the deletion of *EDNRB* in mice [48] and horses [49] results in a white banded coat color phenotype similar to the “two-end-black” color in China indigenous pigs. Additionally, an SNP site in this gene was reported by Yan et al. [50] as the causative mutation for albinism in canines, specifically in the Chinese raccoon dog.

Thirdly, the results of the study on the performance selection directions of South China and North China indigenous pigs demonstrated that North China indigenous pigs exhibited better growth characteristics, while South China indigenous pigs have a higher capacity for fat deposition. We annotated the *IGF1R* and *IGF2R* genes in South China and North China indigenous pigs, respectively. Zhan et al. [51] compared the expression patterns of *IGF1R* and *IGF2R* in different tissues and myocytes in Nanjiang Yellow goats, confirming that these two genes synergistically promoted muscle tissue development and myocyte proliferation and differentiation. The *IGF1R* gene was first discovered and confirmed to play a key role in the regulation of neuroendocrine functions and growth and development in animals [52, 53]. It was involved in regulating cell proliferation, migration, and organ formation during the developmental process of animals and played an important role in the insulin signaling pathway that regulates animal body size. Previous studies on selection gene loci in large-sized pigs and small-sized China indigenous pig breeds identified a missense mutation in the *IGF1R* gene that occurred at a low frequency only in large-sized pigs, suggesting that *IGF1R* may be a key candidate gene for regulating body size and organ development in domestic pigs [54]. Similar results were validated in mice [55]. Additionally, the *IGF1R* and *IGF2R* genes were reported by Wang et al. [56] to play an important role in pig growth and development.

Conclusion

In this study, we performed population genetic analysis and selection signature detection on Asian wild boars, and South China and North China indigenous pigs using whole-genome resequencing data. The population

genetic structure analysis showed that North China indigenous pigs are genetically closer to Asian wild boars than South China indigenous pigs. Both South China and North China indigenous pigs have been selected for growth, meat quality, and reproductive traits, but there are differences in the selection directions. South China indigenous pigs were more selected for fat deposition ability, neural development, small body size, and coat color, whereas selection signatures in North China indigenous pigs were more related to blood physiology, immunity, and reproductive performance. Furthermore, we identified multiple selected genes associated with reproductive physiology, pigmentation, and immune function in both South China and North China indigenous pigs.

Abbreviations

SNP	Single nucleotide polymorphism
WGS	Whole genome sequencing
QTL	Quantitative Trait Locus
Mb	Mega base pair
Kb	Kilo base pair
WBA	Asian wild boars
PCA	Principal component analysis
LD	Linkage disequilibrium
NJ	Neighbor-joining
EigenGWAS	Eigenvector genome-wide association study
XP-EHH	Cross population extended haplotype homozygosity
F_{ST}	Fixation index
GO	Gene ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
GWAS	Genome-wide association study
PigGTEx	Pig Genotype-Tissue Expression
IMPC	International Mouse Phenotyping Consortium
HPA	Human Protein Atlas
PGRP	Pig genomics reference panel
SSC	<i>Sus Scrofa</i> Chromosome
ADG	Average daily gain
DAYS_115	Days to reach 115 kg from birth
TNB	Total number of piglets born
TLWT_Weaning	Total litter weight at weaning
BFT	Backfat thickness
TNUM	Teat number
CD4LP	CD4 positive leukocyte percentage
CD4CD8NP	CD4 positive, CD8 negative leukocyte percentage

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-11119-y>.

Supplementary Material 1
 Supplementary Material 2
 Supplementary Material 3
 Supplementary Material 4
 Supplementary Material 5
 Supplementary Material 6
 Supplementary Material 7
 Supplementary Material 8
 Supplementary Material 9
 Supplementary Material 10

Supplementary Material 11
 Supplementary Material 12
 Supplementary Material 13
 Supplementary Material 14
 Supplementary Material 15
 Supplementary Material 16
 Supplementary Material 17
 Supplementary Material 18
 Supplementary Material 19

Acknowledgements

We also acknowledge technical support from the National Supercomputer Center in Guangzhou. We thank anonymous reviewers and editors for their constructive comments and suggestions.

Author contributions

ZZ, YG and XF conceived and designed the experiments. SD, YL, ZZ, XC, and GL provided technical assistance and revised the manuscript. ZZ, XC and JT helped to draw the figures. ZZ, JL and XL designed the analysis scheme and revised the manuscript. YG, XF and ZZ drafted the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by the China Agriculture Research System (CARS-35), Guangzhou Science and Technology Planning Project (2024A04J3806), Specific university discipline construction project (2023B10564001, 2023B10564003), Guangdong Province Rural Revitalization Strategy Special Fund Seed Industry Revitalization Project (2022-440000-43010101-9501), the Young Scientists Fund of the National Natural Science Foundation of China (32402714), the National Key R & D Program of China (2023YFD1300400), and Guangxi Science and Technology Program Project (GuikeJ23023003).

Data availability

All raw data analyzed in this study are publicly available from CNCB GSA (<https://ngdc.cnbc.ac.cn/>) and NCBI SRA (<https://www.ncbi.nlm.nih.gov/sra/>) databases. Details of WGS dataset can be found in Supplementary Table S1.

Declarations

Ethics approval and consent to participate

Not applicable. No animals or animal materials have been used in this study, and ethical approval for the use of animals was not necessary.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 26 September 2024 / Accepted: 3 December 2024

Published online: 18 December 2024

References

- Duarte CM, Marbá N, Holmer M, Ecology. Rapid domestication of marine species. *Science*. 2007;316:382–3.
- Marom N, Bar-Oz G. The prey pathway: a regional history of cattle (*Bos taurus*) and pig (*Sus scrofa*) domestication in the northern Jordan Valley, Israel. *PLoS ONE*. 2013;8:e55958.
- Zohary D, Tchernov E, Horwitz LK. The role of unconscious selection in the domestication of sheep and goats. *J Zool*. 1998;245:129–35.
- Kijas JM, Andersson L. A phylogenetic study of the origin of the domestic pig estimated from the near-complete mtDNA genome. *J Mol Evol*. 2001;52:302–8.
- Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*. 2012;491:393–8.
- Giuffra E, Kijas JM, Amarger V, Carlborg O, Jeon JT, Andersson L. The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics*. 2000;154:1785–91.
- Ervynck A, Dobney K, Hongo H, Meadow R. Born free? New evidence for the Status of *Sus scrofa* at Neolithic Çayönü Tepesi (Southeastern Anatolia, Turkey). *Paléorient*. 2001;27:47–73.
- Jing Y, Flad RK. Pig domestication in ancient China. *Antiquity*. 2002;76:724–32.
- Zeder MA, Emshwiler E, Smith BD, Bradley DG. Documenting domestication: the intersection of genetics and archaeology. *Trends Genet*. 2006;22:139–55.
- Wang LY, Wang AG, Wang LX, Li K, Yang GS, He RG, et al. Animal genetic resources in China: pigs. Beijing, China: China Agriculture; 2011. (in Chinese).
- Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW, et al. Coordinated international action to accelerate genome-to-phenome with FAANG, the functional annotation of animal genomes project. *Genome Biol*. 2015;16:57.
- Liu SL, Gao YH, Canela-Xandri O, Wang S, Yu Y, Cai WT, et al. A multi-tissue atlas of regulatory variants in cattle. *Nat Genet*. 2022;54:1438–47.
- Teng JY, Gao YH, Yin HW, Bai ZH, Liu SL, Zeng H, et al. A compendium of genetic regulatory effects across pig tissues. *Nat Genet*. 2024;56:112–23.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
- Wickham H. *ggplot2: elegant graphics for data analysis*. New York: Springer; 2016.
- Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33:1870–4.
- Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49:W293–6.
- Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*. 2019;35:1786–8.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10:giab008.
- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19:1655–64.
- Szpiech ZA, Hernandez RD. Selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol Biol Evol*. 2014;31.
- Ma YL, Wei JL, Zhang Q, Chen L, Wang JY, Liu JF, et al. A genome scan for selection signatures in pigs. *PLoS ONE*. 2015;10:e0116850.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–8.
- Chen GB, Lee SH, Zhu ZX, Benyamin B, Robinson MR. EigenGWAS: finding loci under selection through genome-wide association studies of eigenvectors in structured populations. *Heredity (Edinb)*. 2016;117:51–61.
- Hu ZL, Park CA, Reecy JM. Bringing the animal QTLdb and CorDB into the future: meeting new challenges and providing updated services. *Nucleic Acids Res*. 2022;50:D956–61.
- Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics*. 2016;32:289–91.
- Wu TZ, Hu EQ, Xu SB, Chen MJ, Guo PF, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innov (Camb)*. 2021;2:100141.
- Sherman BT, Hao M, Qiu J, Jiao XL, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res*. 2022;50:W216–21.
- Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics*. 2016;32:587–9.
- Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347:1260419.
- Koscielny G, Yaikhom G, Iyer V, Meehan TF, Morgan H, Atienza-Herrero J, et al. The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. *Nucleic Acids Res*. 2014;42:802–9. Database issue:D.

32. Watanabe K, Stringer S, Frei O, Umičević Mirkov M, de Leeuw C, Polderman TJC, et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet.* 2019;51:1339–48.
33. Martin FJ, Amode MR, Aneja A, Austine-Orimoloye O, Azov AG, Barnes I, et al. Ensembl 2023. *Nucleic Acids Res.* 2023;51:D933–41.
34. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25:1754–60.
35. Akey JM, Wang H, Xiong M, Wu H, Shriver WL. MD, Interaction between the melanocortin-1 receptor and P genes contributes to inter-individual variation in skin pigmentation phenotypes in a tibetan population. *Hum Genet.* 2001;108.
36. Shi H, Cheng Z. MC1R and melanin-based molecular probes for theranostic of melanoma and beyond. *Acta Pharmacol Sin.* 2022;43:3034–44.
37. Rouzaud F, Martin J, Gallet PF, Delourme D, Goulemot-Leger V, Amigues Y, et al. A first genotyping assay of French cattle breeds based on a new allele of the extension gene encoding the melanocortin-1 receptor (Mc1r). *Genet Sel Evol.* 2000;32:511–20.
38. Marklund L, Moller MJ, Sandberg K, Andersson L. A missense mutation in the gene for melanocyte-stimulating hormone receptor (MC1R) is associated with the chestnut coat color in horses. *Mamm Genome.* 1996;7:895–9.
39. Kerje S, Lind J, Schütz K, Jensen P, Andersson L. Melanocortin 1-receptor (MC1R) mutations are associated with plumage colour in chicken. *Anim Genet.* 2003;34:241–8.
40. April CS, Barsh GS. Skin layer-specific transcriptional profiles in normal and recessive yellow (Mc1re/Mc1re) mice. *Pigment Cell Res.* 2006;19:194–205.
41. Kijas JM, Wales R, Törnsten A, Chardon P, Moller M, Andersson L. Melanocortin receptor 1 (MC1R) mutations and coat color in pigs. *Genetics.* 1998;150:1177–85.
42. Zhong HW, Zhang J, Tan C, Shi JS, Yang J, Cai GY, et al. Pig Coat Color Manipulation by MC1R Gene Editing. *Int J Mol Sci.* 2022;23:10356.
43. Johansson Moller M, Chaudhary R, Hellmén E, Høyheim B, Chowdhary B, Andersson L. Pigs with the dominant white coat color phenotype carry a duplication of the KIT gene encoding the mast/stem cell growth factor receptor. *Mamm Genome.* 1996;7:822–30.
44. Aoki H, Yamada Y, Hara A, Kunisada T. Two distinct types of mouse melanocyte: differential signaling requirement for the maintenance of non-cutaneous and dermal versus epidermal melanocytes. *Development.* 2009;136:2511–21.
45. Patterson Rosa L, Martin K, Vierra M, Lundquist E, Foster G, Brooks SA, et al. A KIT variant Associated with increased White spotting epistatic to MC1R genotype in horses (*Equus caballus*). *Anim (Basel).* 2022;12:1958.
46. Ai HS, Huang LS, Ren J. Genetic diversity, linkage disequilibrium and selection signatures in Chinese and western pigs revealed by genome-wide SNP markers. *PLoS ONE.* 2013;8:e56001.
47. Zhang Z, Xiao Q, Zhang QQ, Sun H, Chen JC, Li Z-C, et al. Genomic analysis reveals genes affecting distinct phenotypes among different Chinese and western pig breeds. *Sci Rep.* 2018;8:13352.
48. Ceccherini I, Zhang AL, Matera I, Yang G, Devoto M, Romeo G, et al. Interstitial deletion of the endothelin-B receptor gene in the spotting lethal (sl) rat. *Hum Mol Genet.* 1995;4:2089–96.
49. Metallinos DL, Bowling AT, Rine J. A missense mutation in the endothelin-B receptor gene is associated with Lethal White Foal Syndrome: an equine version of Hirschsprung disease. *Mamm Genome.* 1998;9:426–31.
50. Yan SQ, Bai CY, Qi SM, Li ML, Si S, Li YM, et al. Cloning and association analysis of KIT and EDNRB polymorphisms with dominant white coat color in the Chinese raccoon dog (*Nyctereutes procyonoides procyonoides*). *Genet Mol Res.* 2015;14:6549–54.
51. Zhan SY, Ding X, Tao Zhong, Wang LJ, Li L, Zhang HP. Comparison of IGF1R and IGF2R expression patterns in different tissues and muscle cells of Nanjiang Brown Goats. *Acta Vet Et Zootechnica Sinica.* 2019;50:701–11. (in Chinese).
52. Baker J, Liu JP, Robertson EJ, Efstratiadis A. Role of insulin-like growth factors in embryonic and postnatal growth. *Cell.* 1993;75:73–82.
53. Zanou N, Gailly P. Skeletal muscle hypertrophy and regeneration: interplay between the myogenic regulatory factors (MRFs) and insulin-like growth factors (IGFs) pathways. *Cell Mol Life Sci.* 2013;70:4117–30.
54. Li WB, Zhu YL, Ai HS, Guo TF. Identifying signatures of selection related to small body size in pigs. *Acta Vet Et Zootechnica Sinica.* 2016;47:1977–85. (in Chinese).
55. Lee Y, Wang Y, James M, Jeong JH, You M. Inhibition of IGF1R signaling abrogates resistance to afatinib (BIBW2992) in EGFR T790M mutant lung cancer cells. *Mol Carcinog.* 2016;55:991–1001.
56. Wang K, Wu PX, Yang Q, Chen DJ, Zhou J, Jiang A, et al. Detection of selection signatures in Chinese landrace and Yorkshire pigs based on genotyping-by-sequencing data. *Front Genet.* 2018;9:119.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.