Selection signature reveals genes associated with susceptibility loci affecting respiratory disease due to pleiotropic and hitchhiking effect in Chinese indigenous pigs

Zhong Xu¹, Hao Sun¹, Zhe Zhang¹, Cheng-Yue Zhang¹, Qing-bo Zhao¹, Qian Xiao¹, Babatunde Shittu Olasege¹,
Pei-Wei Ma¹, Xiang-Zhe Zhang¹, Qi-Shan Wang¹, Yu-Chun Pan¹.²

1 Department of Animal Science, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, PR China

2 Shanghai Key Laboratory of Veterinary Bio-technology, Shanghai 200240, PR China

*Corresponding author: Yuchun Pan; Qishan Wang.
E-mail: panyuchun1963@aliyun.com; wangqishan@sjtu.edu.cn.

Short title: The research of swine EP in pigs
Abstract

**Background:** Porcine respiratory disease is one of the most important health problems which causes significant economic losses.

**Objective:** To understand the genetic basis for susceptibility to swine enzootic pneumonia (EP) in pigs, we detected 102,809 SNPs in a total of 249 individuals based on genome-wide sequencing data.

**Methods:** Genome comparison of three susceptibility to swine EP pig breeds (Jinhua, Erhualian and Meishan) with two western lines that are considered more resistant (Duroc and Landrace) using XP-EHH and FST statistical approaches identified 691 positively selected genes. Based on QTLs, GO terms and literature search, we selected 14 candidate genes that have convincible biological functions associated with swine EP or human asthma.

**Results:** Most of these genes were tested by several methods including transcription analysis and candidate genes association study. Among these genes: CYP1A1 and CTNNB1 are involved in fertility; TGFBR3 plays a role in meat quality traits; WNT2, CTNNB1 and TCF7 take part in adipogenesis and fat deposition simultaneously; PLAUR (completely linked to AXL, r²=1) plays an essential role in the successful ovulation of matured oocytes in pigs; CLPSL2 (strongly linked to SPDEF, r²=0.848) is involved in male fertility.

**Conclusion:** These adverse genes susceptible to swine EP may be selected while selecting for economic traits (especially reproduction traits) due to pleiotropic and hitchhiking effect of linked genes. Our study provided a completely new point of view to understand the genetic basis for susceptibility or resistance to swine EP in pigs thereby, provide insight for designing sustainable breed selection programs. Finally, the candidate genes are crucial due to their potential roles in respiratory diseases in a large number of species, including human.
INTRODUCTION

Porcine respiratory disease is one of the most important health problems associated with swine production [1]. It can be caused by various infectious pathogens, such as viruses, mycoplasma, bacteria and parasites. Swine enzootic pneumonia (EP) caused by Mycoplasma hyopneumoniae (Mhp) is a chronic and endemic respiratory disease, which has similar pathogenesis and clinical symptoms to human asthma [2]. Mhp infections are highly prevalent in almost all swine producing areas, causing significant economic losses due to decreased growth rate, poor feed efficiency and increased cost of healthcare. Both the control and eradication of EP were difficult because of its complicated mechanisms of infection and its co-infection with other respiratory pathogens such as Mycoplasma hyorhinis, Pasteurella multocida, porcine respiratory and reproductive syndrome virus (PRRSV), and porcine circovirus type 2 (PCV2). Many efforts including vaccination, medication and sanitation have been devoted to controlling its prevalence. However, the frequent application of antibiotics can increase the incidence of drug resistant bacteria and also raise the level of antibiotic residue in food.

Recently, the genetic improvement of disease resistance in animals has been attracting attention. Several studies have explored the genetic mechanisms of susceptibility to EP. Okamura et al. (2012) improved a Landrace breed by selecting for swine EP resistance and meat production over five generations [3]. They identified a significant quantitative trait locus (QTL) for EP between microsatellite markers SW1650 and SW240 on S.scrofa chromosome 2. Huang et al. (2017) conducted a genome-wide association studies (GWAS) on 332 Chinese Erhualian pigs which were genotyped using Illumina Porcine 60K SNP chips [4], and found CXCL6, CXCL8, KIT and CTBP2 as candidate genes that might play important roles in determining resistance or susceptibility to swine EP-like respiratory disease.
In the previous researches and practical production, we noticed that Chinese indigenous pig breeds such as Meishan pigs, Erhualian pigs and Jinhua pigs are more sensitive to *Mhp* compared with commercial Western pig breeds in the same rearing environment [4, 5]. By coincidence, these pig breeds show an excellent fertility and meat quality compared to European breeds and other Chinese breeds under the natural and artificial selective pressure. We proposed two hypotheses: One possibility is that the susceptible loci affecting swine EP are closely linked to the genes or regulatory elements involved in fertility and meat quality traits, and these genes or regulatory elements may have a hitch-hiking effect during the selection for fertility and meat quality traits; another possibility is that the pleiotropic genes or regulatory elements involved in fertility and meat quality traits might simultaneously play important roles in determining susceptibility to swine EP.

To test these hypotheses, we compared the genomes of three Chinese indigenous pig breeds which are very sensitive to swine EP with two commercial pig breeds using cross-population extended haplotype homozygosity (XP-EHH) and $F_{ST}$ approaches [6, 7] to detect genomic regions under positive selection that are related to susceptibility to swine EP. Recently, extensive researches have been carried to identify the positive selection of genes in relation to different specific traits using a distorted pattern of genetic variation between populations. For example, positive selection identified several candidate genes involved in Berkshire meat quality [8]. Taye et al. (2017) also discovered positively selected genes in African cattle responsible for thermotolerance [9], and Yuan et al. (2017) revealed several genes associated with tail type in Chinese indigenous sheep [10].

The aim of this study is to identify genes associated with susceptibility to respiratory disease like swine EP in Chinese indigenous populations using a genomic scan of selective sweep signatures, and to explore how it happened. Understanding the genetic mechanisms in pigs could help to design sustainable breed selection programs.
MATERIALS AND METHODS

Animals and genotyping

A total of 249 individuals were considered in this study. There were three pig breeds sensitive to swine EP consist of 53 Jinhua pigs, 31 Erhualian pigs and 80 Meishan pigs (50 Middle Meishan and 30 Small Meishan pigs), and two relatively insensitive to swine EP pig breeds including 48 Duroc pigs and 37 Landrace pigs. The sequencing data of Jinhua pigs were obtained from Li et al. [11], Erhualian and Meishan pigs were retrieved from Wang et al. [12], Duroc and Landrace pigs were acquired from Zhang et al. [13]. All individuals were genotyped using the genotyping by genome reducing and sequencing (GGRS) protocol [14].

The Sscrofa10.2 pig reference genome was used for calling SNP via SAMtools. Quality control of the fastq file was performed using NGS QC Toolkit v2.3, and the parameters were set according to the report from [14]. Clean sequencing reads were subsequently mapped to the pig reference genome (Sscrofa10.2) using BWA v0.7.5 with default settings for single-end mapping. The SNP calling was performed using SAMTOOLS software (version 0.1.19), and the missing genotypes were imputed using BEAGLE v4.1. Some filters were applied to SNPs as follows: (1) minor allele frequency ≤ 0.05; (2) minimum number of individuals with genotyping ≤ 35%; and (3) SNPs on chromosome X. We filtered the SNPs on chromosome X in consideration of the extremely ancient interspecies introgression and low rate of recombination.

Population structure analysis

To assess the relationships between the animals and breeds under investigation, several procedures were carried out. To generate a pruned subset of SNPs that are in approximate linkage equilibrium with each other in
multidimensional scaling (MDS) analysis, the “--indep 50 5 2” command was used in PLINK v1.07 [15].

Principal component analysis (PCA) was performed using GCTA software (version 1.24), which represented the population structure based on genetic correlations between individuals. Pairwise identity-by-state (IBS) distances between all individuals were calculated using PLINK.

**Genome scans for selection signatures**

Before statistical analysis, Jinhua, Erhualian and Meishan pigs were clustered into Group1 (observed group), whereas Duroc and Landrace pigs were clustered into Group2 (reference groups). To determine a genome wide pattern of positive selection between Group1 and Group2, $F_{ST}$ values per-SNP were calculated based on the formulae proposed by Weir and Cockerham[7]. All $F_{ST}$ values in this study are for a single locus. To determine if Group1 has undergone selection, we computed the XP-EHH values using haplotype information in the xpehh program [6]. The phased haplotype data was reconstructed using FASTPHASE (http://stephenslab.uchicago.edu/software.html). The XP-EHH values were estimated by calculating EHH and log-ratio iHH by comparing haplotypes between Group1 and Group2. EHH and REHH values were calculated using SWEEP v.1.1 (http://www.broadinstitute.org/mpg/sweep/) software.

**Gene annotation**

We searched for positional candidate genes in a 20-kb region centering on the selected SNPs using the NCBI database and Ensembl pig genome databases (http://uswest.ensembl.org/Sus_scrofa/Info/Index). To further analyze the functions of identified genes, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) enrichment analyses were performed using the Database for Annotation, Visualization
and Integrated Discovery (DAVID). Due to the sparsity porcine gene database, we corresponded these genes with the human genome by aligning human ensemble ID. Only terms with a p-value (Benjamini correction) less than 0.05 were considered as significant and listed.

Selection strategy for candidate genes involved in the susceptibility to swine EP

To select genes that contribute to the susceptibility to swine EP from the total genes with strong signatures of selection, we used the following selection strategies: 1) the pig QTL (SS_10.2) database was downloaded from the Animal QTL database website (https://www.animalgenome.org/cgi-bin/QTLdb/SS/index). Because the number of QTLs of Mycoplasmal pneumonia susceptibility (MPS) trait is still relatively few, the keyword “Mycoplasma” was used to find the MPS relevant traits, and then the genes overlapped with the QTLs were selected as groupA; 2) Meanwhile, the genes matched to GO terms associated with lung, immune and inflammatory response were selected as groupB; 3) We merged these two groups and then the plausible genes were manually selected based on literature search, GeneCard and their biological function.

Transcription analysis based on GEO

To complement and verify the selected candidate genes, transcription analysis based on Gene Expression Omnibus (GEO) was conducted. The only one published Expression profiling microarray (GSE49882, GPL17577) of the porcine alveolar macrophage infected by porcine M. hyopneumoniae was downloaded from public GEO database[16], and differentially expressed genes were identified by GEO2R using default parameters. A threshold was applied based on a fold change (FC) of 2.0 (p≤0.05) between infected (6hMhp) and control (6hControl) cells.
In this analysis, all 171 Jinhua pigs were treated similarly. The individuals were genotyped using the same genotyping by genome reducing and sequencing (GGRS) protocol. The phenotypes were recorded based on Huang et al. (2017) [4]. Briefly, one enzootic pneumonia (EP) score was recorded for an individual which was found with a dry cough or obvious abdominal fluctuations and suffering from fast breathing and loss of appetite per day during 100 to 200 days of age. According to their EP scores, these pigs were grouped into controls (EP score = 0) and cases (EP score ≥ 1). Then we calculated the allele frequencies in the candidate genes between case and control groups and the p-values of SNPs were calculated by Fisher’s exact test.

Calculation of $r^2$

In order to search for potential genes linked to these candidate genes involved in the susceptibility to swine EP owing to the hitchhiking effect or lack of recombination, we calculated the linkage disequilibrium (LD) scores between SNPs located in these candidate genes and the other SNPs in selected regions. The $r^2$ between two SNPs was then computed using PLINK v1.07 [15]. In this study, two genes were considered to be connected if the SNPs located in these two genes were physically close to each other and were subject to strong LD (usually measured by $r^2 \geq 0.8$).

RESULTS AND DISCUSSION

SNP quality control

After quality control, there were 249 subjects and 102,809 SNPs in the analyzed dataset. The variants were
distributed on each chromosome in a relatively uniform fashion, with the exception of some isolated regions on some chromosomes (Figure 1). The SNPs positions within the chromosomes were based on the pig reference genome (SGSC Sscrofa10.2).

Population structure

To assess the relationships between the animals and breeds under investigation, we applied MDS to analyze 249 individuals using the pruned 35 104 SNPs with low LD extents. The PCA results showed that the first two principal components (PC1 and PC2) explained 19.8% and 6.5% of the variance respectively. The PC1 separated all individuals into two non-overlapping clusters (Figure 2). Our clustering results showed that the Jinhua, Erhualian and Meishan pigs were grouped together as one group and that the Duroc, Landrace pigs belonged to another group.

Positive selection signature of swine EP sensitive populations

In this study, selection signature analysis was conducted using XP-EHH and FST approaches. We used an empirical procedure simultaneously with significantly high XP-EHH (10% right tail, where XP-EHH is 0.50) and FST values (10% right tail, where FST is 0.54) of the empirical distribution to clarify regions with strong selective sweep signals along the genome, which should harbor genes that underwent selective sweep. Consequently, we identified a subset of 1 479 SNPs with strong selective sweep signals (Figure 3A), which exhibited significant differences ($p < 2.2 \times 10^{-16}$, MannWhitney U test) in XP-EHH and FST values when compared to whole genomic background (Figure 3B). Finally, these SNPs represented 691 genes with strong signatures of selection (Supplementary Table S1), and only two genes, POLR3H and ACO2, were embedded in the most
significantly selected regions simultaneously with XP-EHH values (1% right tail) and \( F_{ST} \) values (1% right tail) (Figure 3C). Aconitase 2 (ACO2) was involved in tricarboxylic acid (TCA) cycle, and hypoxia, a fundamental characteristic of respiratory disease, can regulate the gene expression of mitochondrial aconitase [17].

Using the default settings within DAVID, GO terms and KEGG pathways with p-value <0.05 were enriched and shown in Table 1. The terms overrepresented were related to cell adhesion, membrane part, membrane. In the KEGG analysis, 10 significant pathways were also identified: such as retinol metabolism, pentose and glucuronate interconversions, metabolism of xenobiotics by cytochrome P450, androgen and estrogen metabolism, starch and sucrose metabolism and porphyrin and chlorophyll metabolism.

Positive selection signature related to susceptibility to swine EP in Chinese indigenous pigs

While selective sweep signals are likely to be detected among various regions, in this paper, we focused on and discussed the genes and pathways that putatively contribute to the swine EP-susceptibility mechanisms of Chinese indigenous pigs. Among the 691 genes with strong signatures of selection, there are 50 genes located in seven QTLs of three traits (Mycoplasma pneumonia susceptibility, Mycoplasma hyopneumoniae antibody titer and Change in Mycoplasma hyopneumoniae antibody titer) (Supplementary Table S2), and 33 genes matched the GO terms associated with lung, immune, inflammatory response and drug metabolism (Supplementary Table S3), and then we merged genes of these two groups together as preliminary candidate genes. After literature search based on GeneCard and their biological functions, a set of 14 genes that have convincible biological functions associated with swine EP or human asthma come into the priority candidate genes finally (Table 2 and Supplementary Table S4). Among these 14 genes, previous researches have reported Cytochrome P450 1A1 (CYP1A1) and Toll-like receptor 2 (TLR2) to play an important role in the immune response to \( Mhp \) infection [18, 19].
Innate immune response and inflammatory response play an important role in pathogenicity of swine EP. AXL regulates inhibition of Toll-like receptors (TLRs)-mediated innate immune response and the DNA methylation of AXL at birth was associated with higher risk for asthma-related symptoms in early childhood [20]. CCL11, an inflammatory cytokines, was significantly related to increased risk of asthma in adults [21]. IL7R may control the adaptive immune response to PRRSV vaccine in Pietrain pig by transcriptome data. Kurz et al (2006) identified five asthma susceptibility loci including IL7R in German and Hutterite populations [22].

The abnormality of Wnt/β-catenin signaling was believed to be associated with the development and pathogenesis of lung diseases. WNT2 and CTNNB1 (β-catenin) play a crucial role in the Wnt/β-catenin signaling pathway [23, 24]. Transcription factor 7 (TCF7) is one of important transcription factors for T cell development and differentiation, tumorogenesis, or embryonic development [25]. It is suggested to be involved in immune responses to lung diseases such as pulmonary infection, asthma, acute lung injury, emphysema and lung cancer through several signal pathways, especially the canonical Wnt/β-catenin pathway [25].

The histochemistry and morphology of porcine airway cells would change in response to infection with swine EP. SPDEF selectively expressed in respiratory epithelia cells is an important transcriptional factor of airway goblet cell differentiation and pulmonary inflammation in response to aeroallergens, and it can regulate a transcriptional network mediating mucus secretion in chronic airway disease [26]. MMP2, involved in the inflammatory response, can mediate IL-13-induced suppression of elastin expression in airway fibroblasts [27]. CXCL2 and TGFBR3 play a role in the pathogenesis of airway remodeling in asthma [28].

Hypoxia is a fundamental component of respiratory disease, and the pressurese of pulmonary and hematologic O2 and CO2 were decreased because of less oxygen consumption in swine with Mycoplasma hyopneumoniae
pneumonia [29]. *EPAS1*, also known as *HIF2a*, is a transcription factor that responds to hypoxia-responsive under high-altitude conditions in Tibetan pigs. Interestingly, a previous study reported that physiological hypoxia arises from lung alveolar suffering progressive airflow limitation which increases with COPD severity and *EPAS1* is a key regulator of COPD through responding to hypoxia induced by airflow limitation as proved by integrative analysis of DNA methylation and gene expression data [30]. *CYP1A2* is predominantly expressed in the liver which has been shown to modulate pulmonary oxygen toxicity. It plays a critical role in the susceptibility to hyperoxic lung injury by decreasing oxidative stress and lipid peroxidation in mice [31].

**Transcription analysis based on GEO**

To complement and verify the candidate genes we selected, the allele frequencies of loci in these candidate genes between two groups were shown in Table 2. And the only one published microarray of the porcine alveolar macrophage infected by swine EP was collected from GEO database, and the candidate genes were analysed. The p-value and log₂ (fold change) were listed in Table 2. Among 14 candidate genes, only six genes were found in the expression profiling array (Platforms GPL17577). Four of the six genes, *EPAS1*, *CXCL2*, *TLR2* and *IL7R*, were significantly up-regulated (Figure 4).

**Allele frequencies in candidate genes**

After quality control, we got 125233 SNPs. Of the 171 Jinhua pigs, 93 pigs were grouped into controls (EP score = 0) and 78 pigs were grouped into cases (EP score ≥ 1). The allele frequencies in the 14 candidate genes were calculated between case and control groups. We found that the allele frequencies in six genes (*TCF7*, *TGFBR3*, *SPDEF*, *CCL11*, *IL7R* and *WNT2*) were significantly different between those two groups (Table S5). These results demonstrate the reliability of our results from other sides.
The pleiotropic genes maybe related to susceptibility to swine EP

Among above genes, some have other biological function involved in fertility, meat quality traits, or adipogenesis and fat deposition simultaneously. *CYP1A1* can alter the activity of estrogen in the porcine ovary through estrogen receptor pathway [32]. *CYP1A1* plays an important role in the human placenta during pregnancy. *CTNNB1* as a molecular transcription factor is very important for the remodeling of uterine mucosa during development, the estrous cycle and early pregnancy in pigs [33]. *CTNNB1* contributes to intercellular adhesion in large antral follicles and might influence the normal follicle development and pig fertility [34]. *TGFBR3* plays a role in the muscular and adipose tissue development and it is a candidate gene involved in meat quality traits in pigs [8].

*WNT2*, *CTNNB1* and *TCF7* are all involved in the Wnt/β-catenin signaling pathway, and this pathway plays a critical role in regulating porcine adipogenesis and fat deposition. Chinese indigenous breeds display a higher intramuscular fat content and more abdominal fat deposition than commercial Western pig breeds. And thus, these genes involved in fat deposition may be selected during the natural and artificial selection and they are involved in susceptibility to the EP disease in pigs simultaneously. In a previous study, Huang et al also suggested that genes involved in fat deposition may play an important role in susceptibility or resistance to the enzootic pneumonia disease in pigs[4]. Additionally, many epidemiologic studies have found that abdominal obesity increases the risk of developing asthma in human [35].

Linked genes with candidate genes maybe related to susceptibility to swine EP

The LD scores between SNPs located in these candidate genes and the other SNPs in selected regions were
calculated. Three genes subjected to strong LD (r² ≥ 0.8) were listed in Table 3. Specifically, seven and two loci on chromosomes 6 were completely linked to AXL and MMP2 respectively. In addition, two loci on chromosomes 7 were strongly linked to SPDEF (Table 3).

PLAUR (completely linked to AXL) also known as CD87, is a part of the plasminogen activation system, which regulates the conversion of plasminogen to plasmin. PLAUR plays an essential role in the successful ovulation of matured oocytes together with follicle stimulating hormone (FSH) in pigs [36]. CLPSL2 (strongly linked to SPDEF, r²=0.848), specifically expressed in epididymis, was involved in the regulation of acrosomal integrity, spermatozoa motility, and male fertility. Knockdown of CLPSL2 expression in mouse epididymis results in the decreased number of sperm with intact acrosome, attenuated sperm motility, and reduced the fertility [37].

We noticed that these Chinese indigenous pig breeds show an excellent fertility and meat quality, higher intramuscular fat content and more abdominal fat deposition compared to European breeds under the natural and artificial selective pressure. These adverse genes of susceptibility to swine EP may be selected simultaneously while selecting economic traits due to pleiotropic effects and hitchhiking effect of linked genes. These results support our hypothesis to a great extent and provide a completely new point of view to explain why Chinese indigenous pig breeds are more sensitive to swine EP.

In conclusion, this study revealed several candidate genes that are involved in the susceptibility to swine EP including TCF7, EPAS1, TGFBR3, MMP2, AXL, SPDEF, CYP1A1, CYP1A2, CXCL2, TLR2, CCL11, CTNNB1, IL7R and WNT2 from positive selection signature. Most of these genes are involved in inflammatory response, hypoxia-responsive or Wnt/β-catenin signaling pathway. The susceptibility loci affecting swine EP have increased during the selection for fertility and meat quality traits because of pleiotropic and hitch-hiking effect.
These findings would help in increasing our understanding of the genetic basis for susceptibility or resistance to EP and other respiratory diseases in pigs thereby, provide insight for designing sustainable breed selection programs (such as marker assisted selection). Moreover, we provided a completely new point of view to explain why Chinese indigenous pig breeds are more sensitive to swine EP. In addition, exploring the molecular mechanisms of the susceptibility to swine EP maybe crucial for its potential role in human asthma and other respiratory diseases.

CONFLICT OF INTEREST

We certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

ACKNOWLEDGEMENTS

This work was supported by Zhejiang province agriculture (livestock) varieties breeding Key Technology R&D Program (Grant No. 2016C02054-2), National Natural Science Foundation of China (No. 31872976, U1402266).

Author contributions: Y.C.P and Q.S.W designed and supervised the study, while Z.X analyzed the data with the help of H.S, Z.Z, Q.B.Z and C.Y.Z. Z.X wrote the article with the help of Q.X, B.S.O, P.P.M and X.Z.Z. All authors read and edited the article.

REFERENCES


35. Diaz J, Warren L, Helfner L, et al. Obesity shifts house dust mite-induced airway cellular infiltration from...


Table 1. GO terms and KEGG pathways enriched with candidate genes.

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>Gene number</th>
<th>P-value (Benjamini)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0000267</td>
<td>cell fraction</td>
<td>65</td>
<td>0.0022</td>
</tr>
<tr>
<td>GO:0044425</td>
<td>membrane part</td>
<td>274</td>
<td>0.0127</td>
</tr>
<tr>
<td>GO:0016020</td>
<td>membrane</td>
<td>296</td>
<td>0.0194</td>
</tr>
<tr>
<td>KEGG:hsa00830</td>
<td>Retinol metabolism</td>
<td>13</td>
<td>0.0001</td>
</tr>
<tr>
<td>KEGG:hsa00040</td>
<td>Pentose and glucuronate interconversions</td>
<td>7</td>
<td>0.0023</td>
</tr>
<tr>
<td>KEGG:hsa00053</td>
<td>Ascorbate and aldarate metabolism</td>
<td>7</td>
<td>0.0023</td>
</tr>
<tr>
<td>KEGG:hsa00150</td>
<td>Androgen and estrogen metabolism</td>
<td>8</td>
<td>0.0155</td>
</tr>
<tr>
<td>KEGG:hsa00980</td>
<td>Metabolism of xenobiotics by cytochrome P450</td>
<td>10</td>
<td>0.0190</td>
</tr>
<tr>
<td>KEGG:hsa00983</td>
<td>Drug metabolism</td>
<td>8</td>
<td>0.0282</td>
</tr>
<tr>
<td>KEGG:hsa00500</td>
<td>Starch and sucrose metabolism</td>
<td>8</td>
<td>0.0284</td>
</tr>
<tr>
<td>KEGG:hsa00860</td>
<td>Porphyrin and chlorophyll metabolism</td>
<td>7</td>
<td>0.0304</td>
</tr>
<tr>
<td>KEGG:hsa00140</td>
<td>Steroid hormone biosynthesis</td>
<td>8</td>
<td>0.0329</td>
</tr>
<tr>
<td>KEGG:hsa00982</td>
<td>Drug metabolism</td>
<td>9</td>
<td>0.0419</td>
</tr>
</tbody>
</table>
Table 2. Transcription analysis based on GEO.

<table>
<thead>
<tr>
<th>Candidate genes</th>
<th>Chr</th>
<th>Physical position</th>
<th>Group1</th>
<th>Group2</th>
<th>Microarray</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>AF1</td>
<td>AF2</td>
<td>p-value</td>
</tr>
<tr>
<td>TCF7</td>
<td>2</td>
<td>142047175</td>
<td>0</td>
<td>1</td>
<td>0.488</td>
</tr>
<tr>
<td>EPAS1</td>
<td>3</td>
<td>100164175</td>
<td>0.024</td>
<td>0.976</td>
<td>0.694</td>
</tr>
<tr>
<td>TGFBR3</td>
<td>4</td>
<td>136718160</td>
<td>0.018</td>
<td>0.982</td>
<td>0.806</td>
</tr>
<tr>
<td>MMP2</td>
<td>6</td>
<td>27552563</td>
<td>0.003</td>
<td>0.997</td>
<td>0.5</td>
</tr>
<tr>
<td>AXL</td>
<td>6</td>
<td>44979578</td>
<td>0.324</td>
<td>0.676</td>
<td>0.997</td>
</tr>
<tr>
<td>SPDEF</td>
<td>7</td>
<td>35242328</td>
<td>0.128</td>
<td>0.872</td>
<td>0.8</td>
</tr>
<tr>
<td>CYP1A1</td>
<td>7</td>
<td>63474840</td>
<td>0.003</td>
<td>0.997</td>
<td>0.741</td>
</tr>
<tr>
<td>CYP1A2</td>
<td>7</td>
<td>63474840</td>
<td>0.003</td>
<td>0.997</td>
<td>0.741</td>
</tr>
<tr>
<td>CXCL2</td>
<td>8</td>
<td>74284832</td>
<td>0</td>
<td>1</td>
<td>0.529</td>
</tr>
<tr>
<td>TLR2</td>
<td>8</td>
<td>79818077</td>
<td>0.04</td>
<td>0.96</td>
<td>0.871</td>
</tr>
<tr>
<td>CCL11</td>
<td>12</td>
<td>42446570</td>
<td>0.476</td>
<td>0.524</td>
<td>0</td>
</tr>
<tr>
<td>CTNNB1</td>
<td>13</td>
<td>27684730</td>
<td>0</td>
<td>1</td>
<td>0.512</td>
</tr>
<tr>
<td>IL7R</td>
<td>16</td>
<td>22326963</td>
<td>0.318</td>
<td>0.682</td>
<td>1</td>
</tr>
<tr>
<td>WNT2</td>
<td>18</td>
<td>30813337</td>
<td>0.11</td>
<td>0.89</td>
<td>0.771</td>
</tr>
</tbody>
</table>

Group1, Jinhua, Erhualian and Meishan pigs; Group2, Duroc and Landrace pigs; AF, allele frequency; FC, fold change; NF: Not found.
Table 3. Strong LD between SNPs located in candidate genes and the other SNPs in selected regions.

<table>
<thead>
<tr>
<th>Chr</th>
<th>Physical position</th>
<th>Genes</th>
<th>Chr</th>
<th>Physical position</th>
<th>LD ($r^2$)</th>
<th>Associated genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>27552563</td>
<td>MMP2</td>
<td>6</td>
<td>27451454</td>
<td>1</td>
<td>LPCAT2, SLC6A2</td>
</tr>
<tr>
<td>6</td>
<td>27451485</td>
<td></td>
<td>6</td>
<td>27451485</td>
<td>1</td>
<td>LPCAT2, SLC6A2</td>
</tr>
<tr>
<td>6</td>
<td>44979578</td>
<td>AXL</td>
<td>6</td>
<td>45100366</td>
<td>1</td>
<td>CCDC97, HNRNPUL1</td>
</tr>
<tr>
<td>6</td>
<td>45487407</td>
<td></td>
<td>6</td>
<td>45487407</td>
<td>1</td>
<td>MEGF8</td>
</tr>
<tr>
<td>6</td>
<td>45601909</td>
<td></td>
<td>6</td>
<td>45601909</td>
<td>1</td>
<td>GSK3A, CYP2B6</td>
</tr>
<tr>
<td>6</td>
<td>46185803</td>
<td></td>
<td>6</td>
<td>46185803</td>
<td>1</td>
<td>IRGQ, XRCC1, ZNF576</td>
</tr>
<tr>
<td>6</td>
<td>46185804</td>
<td></td>
<td>6</td>
<td>46185804</td>
<td>1</td>
<td>IRGQ, XRCC1, ZNF576</td>
</tr>
<tr>
<td>6</td>
<td>46185819</td>
<td></td>
<td>6</td>
<td>46185819</td>
<td>1</td>
<td>IRGQ, XRCC1, ZNF576</td>
</tr>
<tr>
<td>6</td>
<td>46331807</td>
<td></td>
<td>6</td>
<td>46331807</td>
<td>1</td>
<td>PLAUR, CADM4</td>
</tr>
<tr>
<td>7</td>
<td>35242328</td>
<td>SPDEF</td>
<td>7</td>
<td>35722329</td>
<td>0.881</td>
<td>ANKS1A</td>
</tr>
<tr>
<td>7</td>
<td>36516210</td>
<td></td>
<td>7</td>
<td>36516210</td>
<td>0.848</td>
<td>CLPSL2, ARMC12</td>
</tr>
</tbody>
</table>
Figure 1. The distribution of the SNPs across the chromosomes. The x-axis denotes the chromosomal position (Mb), and the y-axis represents the chromosomes. The number of the SNPs present in each 1000 kb genome block is expressed via colours.

Figure 2. Population structure analysis of all individuals. Results of principal component analysis (PCA) of Duroc, Landrace, Jinhua, Erhualian and Meishan breeds. Eigenvector1 (x-axis) versus Eigenvector2 (y-axis). The first two principal components, PC1 and PC2, account for 19.8% and 6.5% of the total variance, respectively.

Figure 3. Genome-wide selective sweep analysis of group 1 (sensitive to swine EP). (A) Distribution of XP-EHH and FST. Data points located to the right of the vertical lines (10% right tails of the FST distribution, where FST is 0.54) and above the horizontal line (10% right tail of the XP-EHH distribution, where XP-EHH is 0.50) were identified as selected regions for group 1 pigs (red points); (B) Violin plot of XP-EHH and FST values for regions of group 1 pigs that have undergone positive selection versus the whole genome regions. The statistical significance was calculated by the Mann-Whitney U test. (C) 10%, 5% and 1% significance level of XP-EHH and FST. (D) Violin plot of XP-EHH and FST values for regions of group 1 pigs that have undergone positive selection versus the whole genome regions.

Figure 4. Expression of candidate genes in expression profiling array. (A) CXCL2, (B) IL7R, (C) TLR2 and (D) EPAS1. Transcription analysis of the porcine alveolar macrophage response to infection of Mhp between test (6hMhp) and control (6hControl) cells.
Fig2

**Breed**

- Duroc
- Landrace
- Meishan
- Erhualian
- Jinhua
Fig 3

(a) Distribution of XP-EHH values across chromosomes 1-18 for selected and whole genome regions. The XP-EHH value distribution is significantly different between the two regions, with a p-value of $<2.2 \times 10^{-16}$. The selected region has a lower range (-2.63 to 1.82) compared to the whole genome (0.50 to 1.52).

(b) Comparison of Fst values for whole genome and selected regions. The Fst distribution for the selected regions is significantly different from the whole genome, with a p-value of $<2.2 \times 10^{-16}$.

(c) Scatterplot showing the relationship between XP-EHH and Fst values. The red line indicates regions with high Fst values, highlighting potential selection signals. Key markers include POLR3H, ACO2, TGFBR3, and TLR7.