SCIENCE CHINA
Life Sciences

•RESEARCH PAPER•

August 2022 Vol.65 No.8: 1517–1534
https://doi.org/10.1007/s11427-021-2034-5

Mapping and analysis of a spatiotemporal H3K27ac and gene expression spectrum in pigs

Yaling Zhu[1,2†], Zhimin Zhou[1†], Tao Huang[1†], Zhen Zhang[1†], Wanbo Li[1], Ziqi Ling[1], Tao Jiang[1], Jiawen Yang[1], Siyu Yang[1], Yanyuan Xiao[1], Carole Charlier[3,1], Michel Georges[3,1], Bin Yang[1*] & Lusheng Huang[1*]

[1]*State Key Laboratory of Pig Genetic Improvement and Production Technology, Jiangxi Agricultural University, Nanchang 330045, China;*
[2]*Laboratory Animal Research Center, School of Basic Medical Sciences, Anhui Medical University, Hefei 230032, China;*
[3]*Unit of Animal Genomics, GIGA-Institute and Faculty of Veterinary Medicine, University of Liege, 4000 Liege, Belgium*

Received September 9, 2021; accepted October 29, 2021; published online January 27, 2022

The limited knowledge of genomic noncoding and regulatory regions has restricted our ability to decipher the genetic mechanisms underlying complex traits in pigs. In this study, we characterized the spatiotemporal landscape of putative enhancers and promoters and their target genes by combining H3K27ac-targeted ChIP-Seq and RNA-Seq in fetal (prenatal days 74–75) and adult (postnatal days 132–150) tissues (brain, liver, heart, muscle and small intestine) sampled from Asian aboriginal Bama Xiang and European highly selected Large White pigs of both sexes. We identified 101,290 H3K27ac peaks, marking 18,521 promoters and 82,769 enhancers, including peaks that were active across all tissues and developmental stages (which could indicate safe harbor locus for exogenous gene insertion) and tissue- and developmental stage-specific peaks (which regulate gene pathways matching tissue- and developmental stage-specific physiological functions). We found that H3K27ac and DNA methylation in the promoter region of the *XIST* gene may be involved in X chromosome inactivation and demonstrated the utility of the present resource for revealing the regulatory patterns of known causal genes and prioritizing candidate causal variants for complex traits in pigs. In addition, we identified an average of 1,124 super-enhancers per sample and found that they were more likely to show tissue-specific activity than ordinary peaks. We have developed a web browser to improve the accessibility of the results (http://segtp.jxau.edu.cn/pencode/?genome=susScr11).

H3K27ac, ChIP-Seq, RNA-Seq, pig, XIST, complex traits

Citation: Zhu, Y., Zhou, Z., Huang, T., Zhang, Z., Li, W., Ling, Z., Jiang, T., Yang, J., Yang, S., Xiao, Y., et al. (2022). Mapping and analysis of a spatiotemporal H3K27ac and gene expression spectrum in pigs. Sci China Life Sci 65, 1517–1534. https://doi.org/10.1007/s11427-021-2034-5

INTRODUCTION

It is increasingly apparent that a substantial proportion of the heritability of quantitative traits, including economically important traits in livestock (Georges et al., 2019), is attributable to variants affecting gene regulatory elements, including proximal promoters, distant enhancers and silencers. These regulatory elements have been extensively explored in multiple tissues/cell types at different developmental stages in humans (Kundaje et al., 2015) and mice (Stamatoyannopoulos et al., 2012) through approaches targeting whole-genome open chromatin (DNase I and ATAC-Seq) and histone modifications (ChIP-Seq) followed by next-generation sequencing. These data have contributed to a better understanding of gene regulation underlying tissue or cell lineage differentiation (Kundaje et al., 2015) and development (Nord et al., 2013). Moreover, the combined analysis of the land-

†Contributed equally to this work
*Corresponding authors (Bin Yang, email: binyang@live.cn; Lusheng Huang, email: lushenghuang@hotmail.com)

scape of regulatory regions with risk loci identified by genome-wide association studies (GWAS) revealed tissues and cell types relevant to disease traits (Kundaje et al., 2015) as well as regulatory mechanisms underlying large-effect disease-associated loci (Nasrallah et al., 2020).

A recent study showed that ubiquitous chromatin opening elements (UCOEs) marked by modifications such as histone 3 lysine 27 acetylation (H3K27ac) and DNase I hypersensitivity marks can support the stable expression of nearby genes (Neville et al., 2017). Thus, UCOEs are informative for the identification of safe harbor locus, preferably in intergenic regions, in which exogenous genes can be inserted and stably expressed across tissues and developmental stages. Identifying putative safe harbor locus would be very helpful for the production of transgenic animals to manufacture therapeutic proteins and for other biomedical studies (Neville et al., 2017). Currently, the major safe harbor locus used in swine models are Rosa26 (Li et al., 2014) and H11 (Ruan et al., 2015), which were first identified in mice (Friedrich and Soriano, 1991; Tasic et al., 2011). No study has been implemented to scan the swine genome for candidate safe harbor locus based on chromatin modification and gene expression data from pigs.

Some mammalian regulatory elements reside within ~5% of the genome under evolutionary constraint (Lindblad-Toh et al., 2011). Thus, interspecies sequence conservation effectively contributes to the identification of at least a portion of mammalian regulatory space. However, distant regulatory elements in particular have been shown to be capable of rapid evolutionary turnover despite their functional importance (Nord et al., 2013; Villar et al., 2015). Therefore, it is essential to identify gene regulatory elements at the genome-wide scale independent of interspecies conservation. Towards that goal, collaborative initiatives are now underway in livestock species, which are collectively referred to as the Functional Annotation of ANimal Genome project (Andersson et al., 2015). In pigs, studies have profiled H3K27ac and H3K4me3 in bulk liver tissue (Villar et al., 2015), placental tissues of fetuses from gestational days 50 and 95 (Han et al., 2019a), and open chromatin regions (ATAC-Seq) and the genome 3D topology (Hi-C) in $CD4^+$ and $CD8^+$ T cells, and dissociated liver cells (Foissac et al., 2019). A recent study performed H3K27ac and H3K4me4 ChIP-Seq of 12 tissues and conducted ATAC-seq in muscle and fat tissues from four pig breeds, revealing more than 220,000 *cis*-regulatory elements in the pig genome (Zhao et al., 2021). However, further efforts are needed to enrich the catalogue of known pig regulatory elements in multiple tissues at different developmental stages.

Herein, we describe the characterization of H3K27ac profiles across the porcine genome. H3K27ac is an established histone mark of active promoters and enhancers that has been shown to be highly variable across tissues and significantly enriched for disease-associated variants in humans (Kundaje et al., 2015). Assays were performed on 70 samples from five tissues (brain, heart, liver, muscle and small intestine), two developmental stages (fetal, from 74–75 days prenatal, and adult, from 132–150 days postnatal), two breeds (European Large White and Chinese Bama Xiang) and both sexes of pigs (Figure 1A). Sixty-two of these samples were also profiled by RNA sequencing. We obtained a total of 101,290 H3K27ac peaks the were present in at least two samples, which marked 18,521 active promoters and 82,769 active enhancers. These peaks were further characterized for their (i) genomic distribution, (ii) association with the expression of nearby genes, (iii) constitutive activity across different tissues and developmental stages and utilization in revealing potential safe harbor locus for genome editing, (iv) specific activities in different tissues and developmental stages, and (v) differential activity between the two breeds and two sexes. Additionally, we identified an average of 1,124 (471–2,076) super-enhancers per sample, revealing their tissue-specific activity and associated genes. We also demonstrated the utility of H3K27ac peaks for identifying the regulatory patterns of candidate genes and causative variants associated with complex traits.

RESULTS

The H3K27ac landscape in five tissues, two sexes, two breeds and two developmental stages of pigs

Two Large White full sisters (sow1 and sow2) were mated to the same Large White boar. Sow 1 was slaughtered at day 75 post mating, and two male and two female fetuses were harvested. Sow2 was raised until delivery, and two male and two female offspring were slaughtered at 150 d after birth for sampling. The same design was applied to the Chinese aboriginal Bama Xiang pig breed, except that only 3 fetuses were harvested at day 74, and the adult Bama Xiang pigs were slaughtered at days 132–141. We sampled the brain, liver, heart, muscle and small intestine from all fetuses and adult animals (Figure 1A).

We performed H3K27ac ChIP-Seq on 70/75 samples and RNA-Seq on 62/75 samples (Figure 1A; Tables S1 and S2 in Supporting Information). In the ChIP-Seq data, an average of 31.2 and 30.3 million single-end 50-bp reads were obtained for the input and immunoprecipitated (IP) samples, respectively, among which 87.5% and 89.5% were mapped to *Sus scrofa* genome assembly 11.1 (see MATERIAL AND METHODS; Table S1 in Supporting Information). Sequence enrichment analysis, the fraction of reads in peaks (FRiP), and library complexity statistics including the nonredundant fraction (NRF), PCR bottlenecking coefficient 1 (PBC1), and PBC2 supported the validity of the ChIP-Seq data (see MATERIAL AND METHODS; Table S1 in Supporting In-

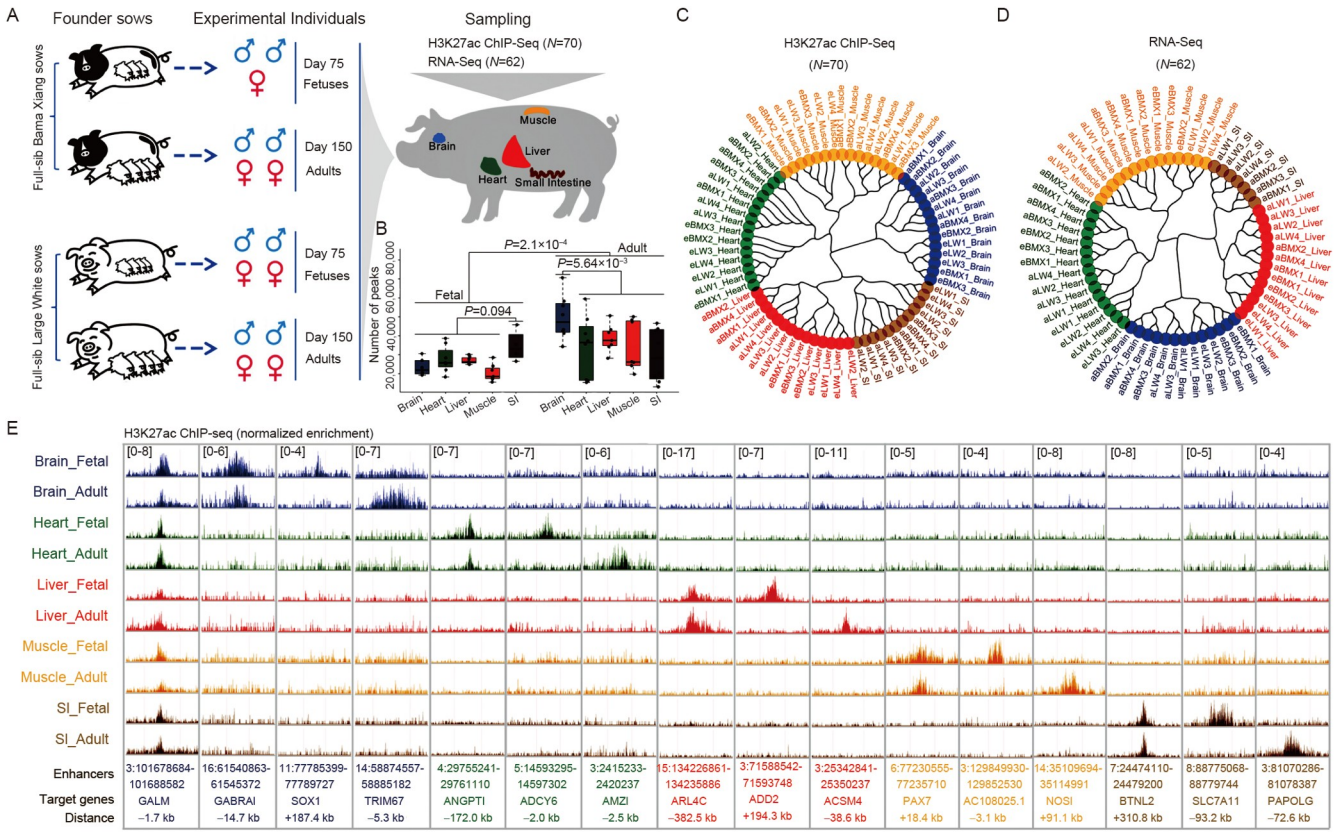


Figure 1 Identifying the activity landscape of active promoters and enhancers via H3K27ac ChIP-Seq in pigs. A, Schematic diagram showing the design, samples and data collected in this study. B, Boxplot showing the distribution of the number of peaks identified in samples sorted by tissue type and developmental stage. C and D, Neighbor joining tree of samples inferred from peak activity and gene expression separates the ten tissues-developmental stage groups (5 tissues×2 developmental stages). E, Representative examples of H3K27ac peaks showing (i) constitutive activity: near *GALM*, a gene encodes an enzyme involved in the fundamental cellular function of galactose metabolism (Vaillancourt et al., 2008). (ii) 5 tissue specific activity with or without additional developmental stage specific activity: brain specific peaks locate close to *GABRA1* (neurotransmitter in mammalian brain), *SOX1* (embryonic and forebrain neuron development) and *TRIM67* (Brain development and behavior) with function matching the biology of brain; heart specific peaks locate close to *ANGPT1I* (angiopoietin and heart development), *ADCY6* (adenylyl cyclase related to vascular sensitivity) and *AMZ1* (heart development); liver specific peaks near *ARL4C* (cholesterol transport), *ADD2* (hematopoietic protein) and *ACSM4* (fatty acid ligase); muscle specific peaks near *PAX7* (transcription factor regulating myogenesis), *AC1080251* (a transcript with unknown function) and *NOS1* (Skeletal muscle integrity and contractile); small intestine specific peaks near *BTNL2* (immune surveillance and ulcerative colitis), *SLC7A11* (amino acid transmembrane transport) and *PAPOLG* (immune process). Each track represents average number of base-pairs per million mapped H3K27ac ChIP-Seq reads in 20-bp bin across all samples in corresponding tissue-developmental stage. The scale of a peak track was set to be consistent across different tissue-developmental stages.

formation). From these data, we identified an average of 31,121 peaks per sample (range: 13,064 to 70,649), with an average peak length of 2,527 bp (range: 1,224 to 5,480 bp), which was comparable to previous reports (Villar et al., 2015) (Table S1 and Figure S1 in Supporting Information). The adult samples showed significantly higher peak numbers than fetal samples (P=2.1×10^{-4}) (Figure 1B), which was in agreement with the observation that the number of H3K27ac peaks in the postnatal stage was greater than that in the prenatal stage among the forebrain, heart and liver of mice (Nord et al., 2013) (Figure S2 in Supporting Information). Moreover, the adult brain samples exhibited more peaks on average than the other adult tissues (P=5.6×10^{-3}) (Figure 1B), which may reflect the greater cellular heterogeneity of the brain relative to other tissues.

In the RNA-Seq data, an average of 43.3 million reads were generated for the 62 samples, among which 93.8% mapped to the reference genome on average (Table S2 in Supporting Information). Transcripts were assigned to a reference set of 44,415 genes (34,355 from Ensembl and 10,060 from *de novo* assembly with StringTie) (Table S3 in Supporting Information). After quality control procedures (MATERIALS AND METHODS), 20,319 genes were retained for further analysis. Gene expression levels were normalized using the variance stabilizing transformation (VST) approach in DESeq2 (Love et al., 2014).

To facilitate the comparison of H3K27ac across samples, we combined peaks across samples and removed the top 5% in terms of peak length, yielding a set of 170,371 merged peaks, with an average length of 2,667 bp, which was only marginally larger than the average length of the peaks (2,527 bp) identified in individual samples before merging.

To assess the sequence and functional conservation of the peaks, a cross-species comparison of H3K27ac peaks in the brain, heart and liver was performed between pigs and humans based on data downloaded from the ENCODE consortium (https://www.encodeproject.org/) (Figure S3 in Supporting Information). The results showed that the sequences of 89.9%–91.8% of the H3K27ac peaks were conserved and that 35.7%–50.3% of the peaks were functionally conserved in the brain, heart and liver tissues of pigs and humans (Figure S3A–C in Supporting Information). First, we used a binary approach to define the status of the peaks as "active" or "inactive" and considered a merged peak to be active in a given sample if it overlapped by at least 1 bp with any peak identified in that sample. We further quantitatively measured the activities of the merged peaks in each sample as the average read depth per million IP reads minus the average read depth per million input reads in the peak region. It was noteworthy that the binarily inactive peaks in an individual could be covered by sequence reads and that their quantitative activity could therefore be greater than zero. Based on the binary peak activity analysis, 85.3% of the peaks were active in at least one of our adult male liver samples overlapped with the adult male liver peaks identified in an independent study (Villar et al., 2015) (Figure S4 in Supporting Information), supporting the validity of the peaks revealed in this study. A clustering analysis based on the quantitative activities of the 170,371 peaks showed that the samples were grouped first by tissue, followed by developmental stage, without exceptions (Figure 1C; Figure S5A in Supporting Information). A similar clustering pattern was observed based on the expression levels of 20,319 genes identified from RNA-Seq data, except in the heart, for which the breed effect appeared stronger than the developmental stage effect (Figure 1D; Figure S5B in Supporting Information). Thus, both the H3K27ac and gene expression profiles accurately predicted the tissue types and developmental stages of the samples. We provide examples of peaks that were ubiquitously active across all tissues and developmental stages, peaks that were tissue specific but not developmental stage specific, and peaks that were both tissue and developmental stage specific (Figure 1E).

We next focused on 101,290 peaks that were active (binary phenotype) in at least two samples among the entire set of samples. Following Nord et al. (2013), we classified 18,521 peaks intersecting with the 5′ end of any transcript (+/− kb), as inferred from the RNA-Seq data or provided by the Ensembl database (release 98), as proximal peaks/putative promoters. The remaining 82,769 peaks were considered to be distal peaks/putative enhancers. The quantitative activity of the 18,521 proximal peaks was significantly correlated with the expression levels of their corresponding genes within each sample (average Pearson correlation coefficient=0.37, P value$<1\times10^{-300}$) (Figure S6 in Supporting Information), reflecting a link between a gene's expression level and the degree of H3K27 acetylation in its promoter region.

We observed that the greatest number of distal peaks were located in introns (59.5%), followed by intergenic (34.4%) regions. Among the intronic peaks, more than half were located in the first three introns of the corresponding genes (Figure S7 in Supporting Information). This underscores the important role of intronic regions close to the 5′ end of genes in the regulation of gene transcription, in agreement with reports in other species, such as rats (Chan et al., 1999) and *Arabidopsis* (Gallegos and Rose, 2017). Compared to proximal peaks, distal peaks were more likely to be active in a specific tissue in both fetal and adult samples (Figure S8 in Supporting Information). Similar results were also observed in mice (Nord et al., 2013), reflecting the important roles of enhancers in controlling tissue-specific gene expression. Moreover, distal peaks exhibited a significantly lower GC content, the extent of sequence conservation across species (GERP score), peak length and peak activity than proximal peaks, reflecting their distinct biological identities (Figure S9 in Supporting Information).

Identifying target genes of H3K27ac regions across the whole genome and the potential silencing role of H3K27ac in gene transcription

To investigate the associations of H3K27ac elements with nearby gene expression, we calculated the correlations of each H3K27ac peak with the expression levels of genes within a 500 kb region from the corresponding peak across 62 samples, given that 75% of promoter-based DNA interactions revealed by HiC data were reported to be located within 500 kb (Javierre et al., 2016). We identified an estimated 41,105 correlations, including 35,644 positive and 5,461 negative correlations, among 1,089,891 tests at a q value threshold of 0.05. We next investigated the distribution of positive and negative peak-gene correlations against peak-gene distances (Figure 2). The number of significant correlations of peaks located upstream and downstream of their associated genes appeared to be approximately symmetric, suggesting that H3K27ac regions can function equally well upstream and downstream of the transcription start sites (TSSs) of corresponding genes (Figure 2A). Interestingly, the positive correlations were enriched in peaks and genes located nearby (Figure 2A), while the negative peak-gene correlations appeared to be underrepresented in peaks located close to their target genes (Figure 2B), reflecting a mechanism by which H3K27ac in promoter regions is always associated with increased expression of corresponding genes. Moreover, the peak length was positively associated with the strength of the peak-gene correlations, particularly for distal peaks (Figure 2C–F). We reasoned that the peaks with greater lengths could include extended open chromatin

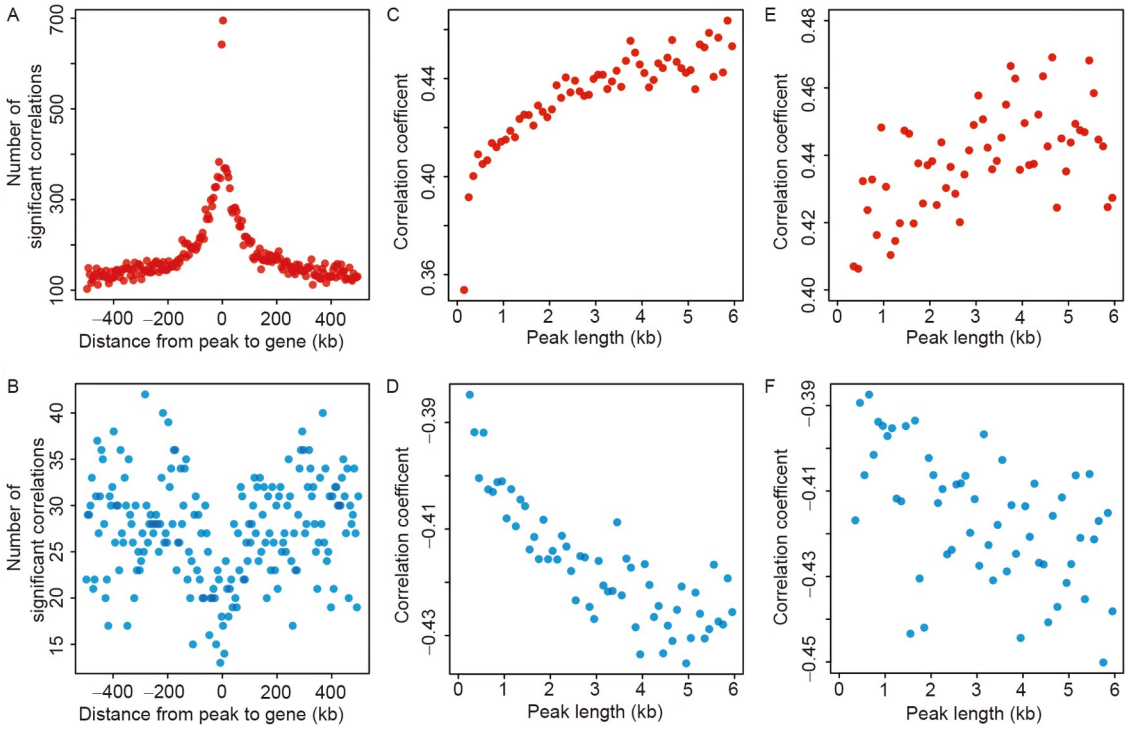


Figure 2 Characterizing the correlations between peak activity and expression of nearby genes (within 500 kb). A, Distribution of significant positive peak-gene correlations as a function of peak-gene distance in a window of 1 Mb. B, Distribution of significant negative peak-gene correlations as a function of peak-gene distance. The dots represent the number of significant correlations in 5 kb sliding windows. C and D, The distribution of positive and negative correlation coefficients as a function of peak size (distal peaks) in a window of 1 Mb. E and F, The distribution of positive and negative correlation coefficients as a function of peak size (proximal peaks). C–F, The results were visualized by 100 bp sliding windows in terms of peak size.

regions that better support the transcription of corresponding genes, thus contributing to stronger peak-gene correlations. An alternative explanation is that peak activity is better estimated for large than for small peaks, as more reads are used to estimate activity for large peaks than for small peaks.

H3K27ac peaks that are ubiquitously active may serve as safe harbor locus for swine genome editing

We next searched for peaks that are ubiquitously active across tissues, ages, sexes and breeds. We assumed that such peaks may play important roles in maintaining fundamental biological functions and may reveal a safe harbor locus where transgenes can be inserted and stably expressed. We defined a peak as ubiquitously active if it was active in at least half of the samples in each of the 10 tissue-developmental stage groups (Figure 1C and METHODS). Based on this, we identified a total of 2,679 ubiquitously active peaks, most of which (96.7%) were proximal. In agreement with our expectations, the genes whose TSSs overlapped with ubiquitously active peaks were enriched for basic housekeeping functions such as protein metabolism (q=1.2×10^{-58}), RNA metabolism (q=7.1×10^{-24}) and splicing (q=1.6×10^{-12}) (Figure 3A; Table S4 in Supporting Information).

To determine whether the ubiquitously active peaks could aid in the identification of potential safe harbor locus in pigs, we compared the expression means and coefficients of variation of genes adjacent to the ubiquitously active peaks with those of all other genes and found that the former showed significantly higher average expression levels (P value<4.5×10^{-13}) and lower CVs (P value<2.2×10^{-16}) across the 62 samples investigated (Figure 3C and D), suggesting that the ubiquitously active peaks could support high, stable expression of neighbouring genes. We further examined the peak activity and gene expression of two known safe harbor locus (Papapetrou and Schambach, 2016), Rosa26 and H11, and found that both the Rosa26 and H11 regions contained peaks and genes that were active or highly expressed across all tissues and developmental stages (Figure 3B; Figure S10 in Supporting Information). Encouraged by these results, we further filtered out 110 ubiquitously active peak regions containing genes with higher mean expression and lower CVs of expression than the *THUMPD3* gene in the Rosa26 region and the *DRG1* gene in the H11 region as candidate safe harbor locus (Table S5 in Supporting Information). Two such regions are exemplified in Figure S11 in Supporting Information. Overall, H3K27ac and RNA-Seq data across tissues, ages, sexes and breeds provide a valuable resource for identifying genome-wide potential safe harbor locus for exogenous gene insertion in the genome engineering of pigs.

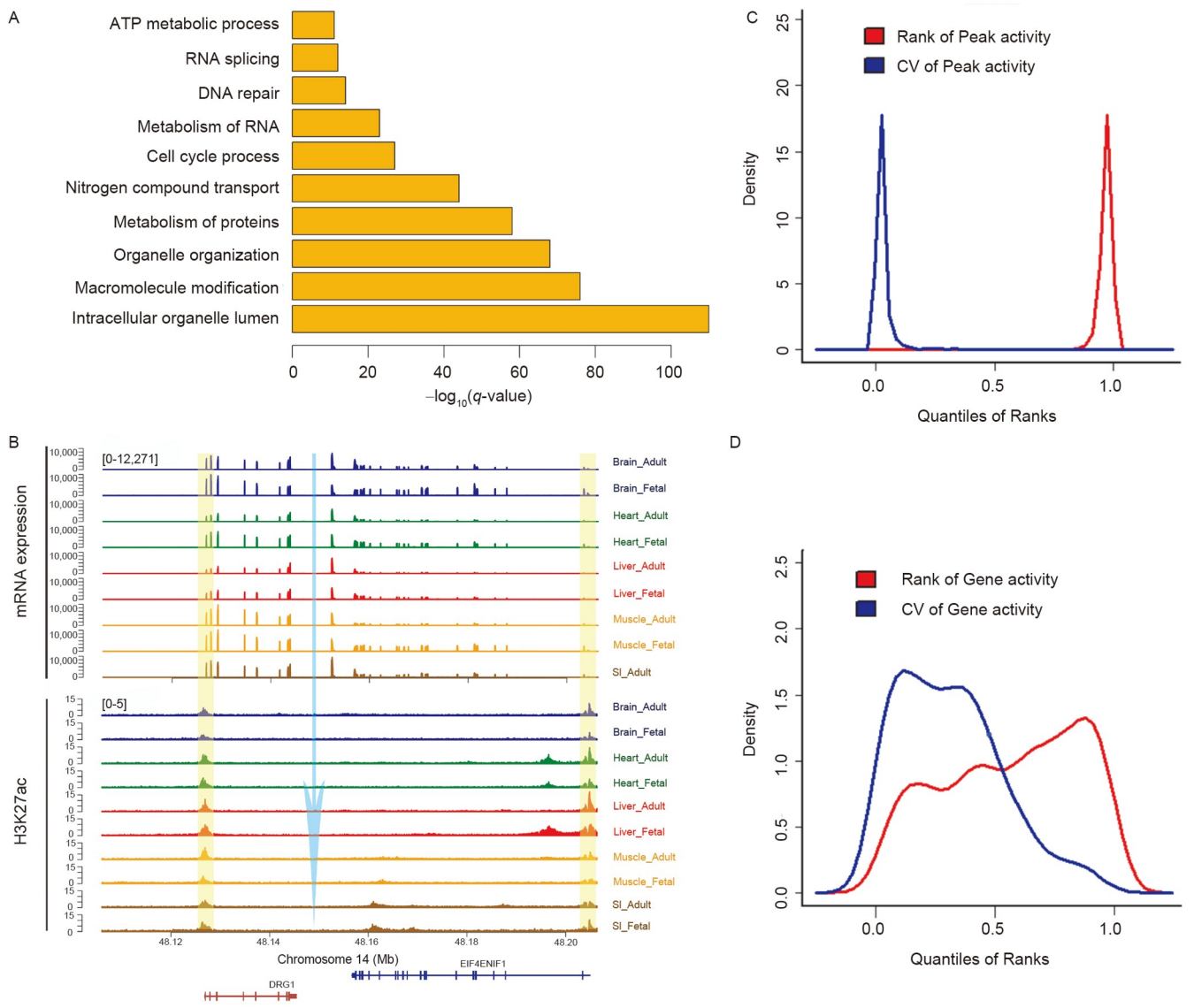


Figure 3 Properties of ubiquitously active H3K27ac peaks and their potential to identify safe harbor locus. A, Biological processes enriched in cognate genes of the ubiquitously active peaks (the genes whose TSS overlap with the peaks). B, The tracks of mRNA expression and H3K27ac across the tissue-developmental stages at H11 safe harbor locus on pig genome. The transparent yellow bars mark the peaks that show constitutive activity across tissue-developmental stages. The blue arrow marks a potential intergenic safe harbor, a region between two ubiquitously expressed genes, for exogenous gene insertion. Each track represents the average number of base-pairs per million mapped RNA-Seq/H3K27ac ChIP-Seq reads in 20-bp bin across all samples in corresponding tissue-developmental stage. The scale of a gene/peak track was set to be consistent across different tissue-developmental stages. C, The distributions of peak activity and coefficient of variations of activity for the 2,679 ubiquitously active peaks defined in this study. D, The distributions of expression levels and coefficient of variation of expression for genes cognate to the 2,679 ubiquitously active peaks.

Tissue-specific peaks are associated with biological processes and transcription factors matching tissue-specific physiological functions

We next investigated the impacts of tissue, development stage, breed and sex on peak activity using a linear mixed model. Tissue, developmental stage, breed and sex explained 21.2%, 4.0%, 0.01% and 0.35% of the variation in peak activity and 51.1%, 6.1%, 0.94% and 0.02% of the variation in mRNA expression on average, respectively (Figure S12 in Supporting Information). This demonstrates that the tissues are the largest source of variation underlying both the peak activity and gene expression results, followed by the development stage, while the impacts of breed and sex are much lower.

To identify H3K27ac peaks that potentially regulate tissue-specific gene expression and, in turn, tissue-specific physiological functions, we searched for peaks that were active in one specific tissue in both fetal and adult stages. We selected peaks that were active (binary) in at least half of both fetal and adult samples of the target tissue without being active (binary) in any other sample type, and Nasrallah et al.

(2020) showed an average activity (quantitative) value that was at least four times higher in the target tissue than in all other tissues. We identified a total of 790, 420, 940, 156 and 259 tissue-specific peaks in the brain, heart, liver, muscle and small intestine, respectively (Figure 4A). All of them were significantly associated with their respective tissue type according to a linear regression analysis at a q value threshold of 0.05 and showed extreme Q and Tau values (Figure S13 in Supporting Information; MATERIALS AND METHODS), which indicated strong tissue-specific activity of peaks, supporting the validity of our filtering approach. Compared with the ubiquitously active peaks, the tissue-specific peaks were mainly distal (Figure S14 in Supporting Information), reflecting an important role of distal peaks in determining tissue-specific biology.

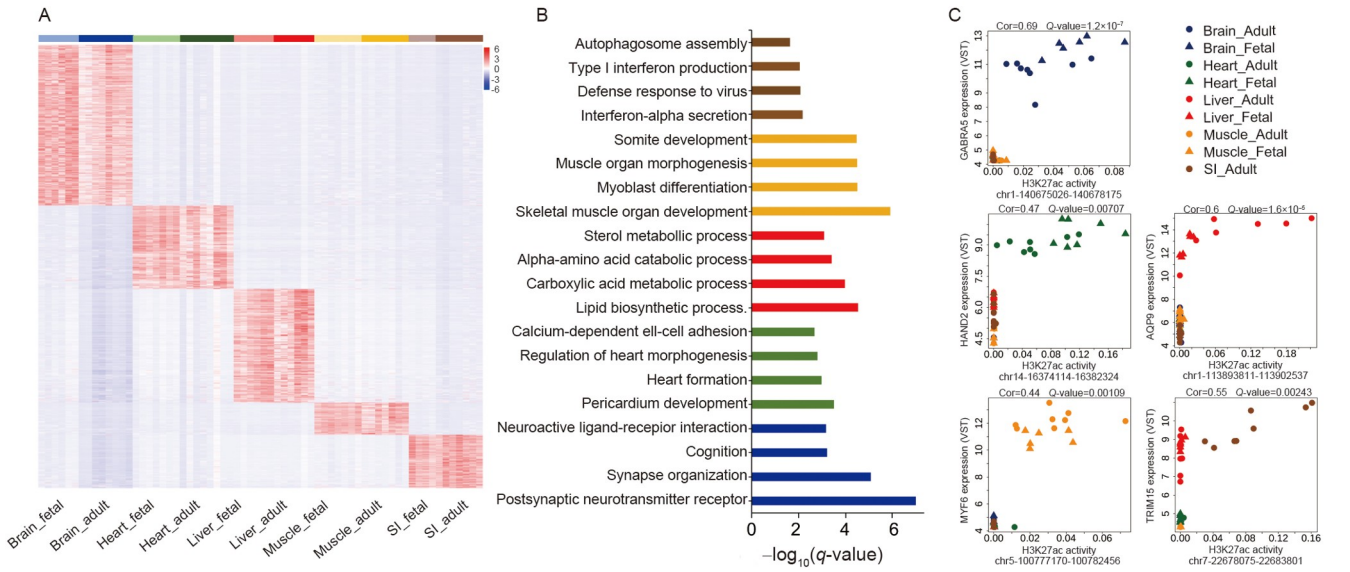
Gene ontology enrichment analysis of the target genes of the tissue-specific peaks (i.e., the genes located within 500 kb from the peak with the highest correlation between peak activity and expression) revealed biological processes such as postsynaptic neurotransmitter receptor activity (q=9.8×10^{-8}) related to brain-specific peaks, heart formation (q=0.001) related to heart-specific peaks, lipid biosynthetic process (q=2.8×10^{-5}) and cellular amino acid metabolic process (q=9.9×10^{-5}) related to liver-specific peaks, skeletal muscle organ development (q=1.2×10^{-6}) related to muscle-specific peaks, and interferon-alpha secretion (q=6.7×10^{-3}) related to SI-specific peaks (Figure 4B and C; Figure S15 and Table S6 in Supporting Information). Transcription factors play central roles in tissue differentiation and development (Almalki and Agrawal, 2016). We further performed an enrichment analysis of transcription factor binding sites (TFBS) in the tissue-specific peaks, which revealed an overrepresentation of transcription factors (TFs) related to the biology of corresponding tissues. These included RFX3 and RFX2 in the brain, MEF2C, MEF2D, GATA2, GATA4, MEF2A and ESRRG in the heart, CUX2 and NR2f2 in the liver, and MEF2D and MEF2A in muscle (Figure 4D; Table S7 in Supporting Information). Among these TFs, RFX3 showed the highest expression in the brain, GATA2, GATA4 and MEF2A showed the highest expression in the heart, and MEF2D showed the highest expression in muscle (Figure 4E; Figure S16 in Supporting Information). Taken together, the results showed that tissue-specific peaks were associated with genes that are involved in biological processes and bound by TFs related to the biology of corresponding tissues, emphasizing the critical roles of tissue-specific H3K27ac peaks in driving tissue-specific biological functions.

Developmental stage-specific H3K27ac peaks in the five tissues

We next sought to identify peaks showing developmental stage-specific activity and their potential associations with pathways underlying the functional switch from the fetal to the adult stage (Figure 5A). We defined the fetal/adult-specific peaks of a tissue as those that were active (binary) in at least half of fetal/adult samples but inactive in all adult/fetal samples and (Nasrallah et al., 2020) showed at least a four-fold difference in average quantitative activity between the two developmental stages in the corresponding tissue. Based on this, we identified 428, 949, 2,145, 1,202 and 2,705 fetal-specific peaks and 4,328, 1,376, 5,778, 2,522 and 973 adult-specific peaks in the brain, heart, liver, muscle and duodenum, respectively (Figure 5A and B). The developmental stage-specific peaks were mostly tissue specific (95.9% for fetal and 89.6% for adult-specific peaks) (Figure 5B), as only 1 fetal-specific peak and 17 adult-specific peaks were shared among the four tissues (Figure 5B), and none were shared across the five tissues.

We focused on a region containing six myosin family genes (*MYH13*, *MYH8*, *MYH4*, *MYH1*, *MYH2* and *MYH3*) that are regulated in a developmental stage-specific manner in muscle. This region was previously reported to be significantly associated with intramuscular fat content (IMF) in Korean native pigs, and a 6-bp deletion upstream of the *MYH3* gene was proposed to be the causative mutation (Cho et al., 2019). We observed fetal-specific expression of *MYH8* and *MYH3* in muscle, coinciding with fetal-specific H3K27ac peaks in the promoter regions of these two genes. Higher expression of the *MYH4*, *MYH1* and *MYH2* genes in adult muscle was associated with adult-specific peaks located close to these genes (Figure 5C). The fetal-specific expression of *MYH3* may suggest that *MYH3* is not a causal gene of the loci for IMF or that the effect of *MYH3* on IMF could result from a mechanism operating in the fetuses.

Gene ontology enrichment analysis showed that the target genes of the fetal-specific peaks were enriched in biological processes such as the cell cycle, tissue development and other pathways that correspond to the biology of the fetal status of the corresponding tissues. For example, gene sets related to myelination (Jakovcevski et al., 2009) (q=0.005) and GABAergic synaptic transmission (Kirmse et al., 2018) (q=0.0098) (which are important for fetal brain development) were enriched in target genes of fetal brain-specific peaks, while gene sets related to haematopoietic stem cell differentiation (q=0.052) (which correspond to the haematopoiesis function of the fetal liver) (Popescu et al., 2019) were enriched in target genes of fetal liver-specific peaks (Figure 5D; Table S8 in Supporting Information). Among adult-specific peaks, we identified biological processes including choline transport (q=0.007) and cardiac muscle contraction (q=0.008) in the heart, oxo-acid metabolic processes (q=5.8×10^{-9}) in the liver, glycogen catabolic process (q=3.8×10^{-4}) and skeletal muscle contraction (q=0.011) in muscle, and T cell selection (q=0.022) and intestinal



D

Group	Motif	Name	*q*-value (Benjamini)	% of Peaks with Motif
Brain_Specific		RFX3	0.0002	2.28%
Brain_Specific		RFX2	0.0002	2.41%
Heart_Specific		MEF2C	0.0075	12.87%
Heart_Specific		GATA4	0.0353	18.07%
Liver_Specific		CUX2	0.0362	10.13%
Muscle_Specific		MEF2D	0.0141	8.97%
Muscle_Specific		MEF2A	0.0141	14.10%

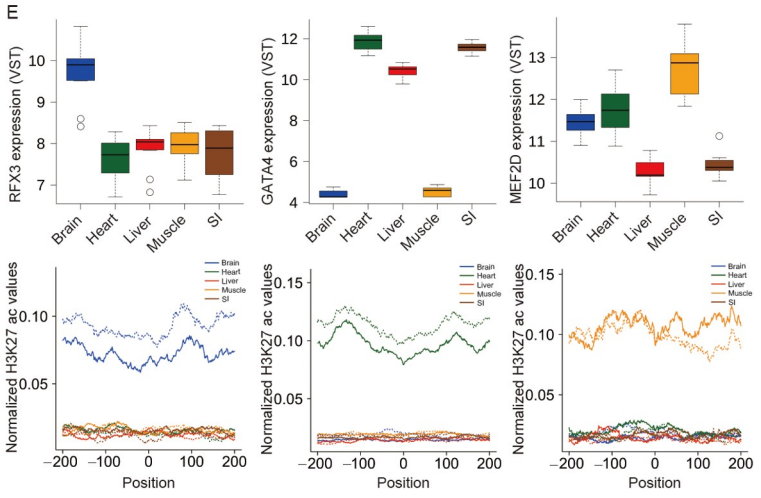


Figure 4 Functional annotation of tissue specific H3K27ac peaks in the five tissues. A, Heatmap of tissue specific peaks in the five tissues. Each row of the heatmap shows the normalized density of read-depth at one peak. B, Representative examples of enriched biological processes in target genes of tissue-specific peaks. C, Representative examples of correlation of tissue specific peak activity with nearby gene expression. D, List of transcription factors with binding motifs enriched in tissue specific peaks. E, Transcription factors whose binding sites are enriched in tissue-specific peaks are highly expressed in the corresponding tissue: examples. The blue, green, red, orange and tan colors indicate signals in brain, heart, liver, muscle and small intestine, respectively. The light colors in the lower three graphs indicate fetal stages of corresponding tissues.

absorption (q=0.027) in the small intestine (Figure 5D). These results underscore the important roles of H3K27ac in the fetal-to-adult functional switch in the corresponding tissues by regulating genes and pathways with developmental stage-specific roles.

We also found significant overrepresentation of 140 TFBSs in the developmental stage-specific peaks (Table S9 in Supporting Information). Many of the enriched TFs have physiological functions that are relevant to the biology of the cognate tissues at the respective developmental stage. The most enriched TFs included FOXD3 (neural crest development), SOX17 (cell cycle in the mouse brain) and SOX2 (adult mouse brain neurodegeneration) among adult brain-specific peaks and RFX3 (early brain development) among fetal brain-specific peaks; GATA4 (myocardial regeneration in neonatal mice), GATA6 (cardiac development) and TBX20 (cardiac progenitor formation) among fetal heart-specific peaks and HNF4a (cell differentiation and proliferation in the liver), FOXK1 (liver cancer), and FOXO3 (regulate hepatic triglyceride metabolism) among adult liver-specific peaks; GATA2 and GATA1 (haematopoietic factors in fetal liver cells) among fetal liver-specific peaks, and MEF2D, MEF2C, MEF2b, SIX1 and SIX2, which are essential for muscle differentiation and development, among adult muscle-specific peaks, and myogenic regulatory factors (MYF5, MYOG and MYOD) that are relevant to embryonic development among fetal muscle-specific peaks; and IRF2 (regulate IL-7 in intestinal epithelial cells) among adult

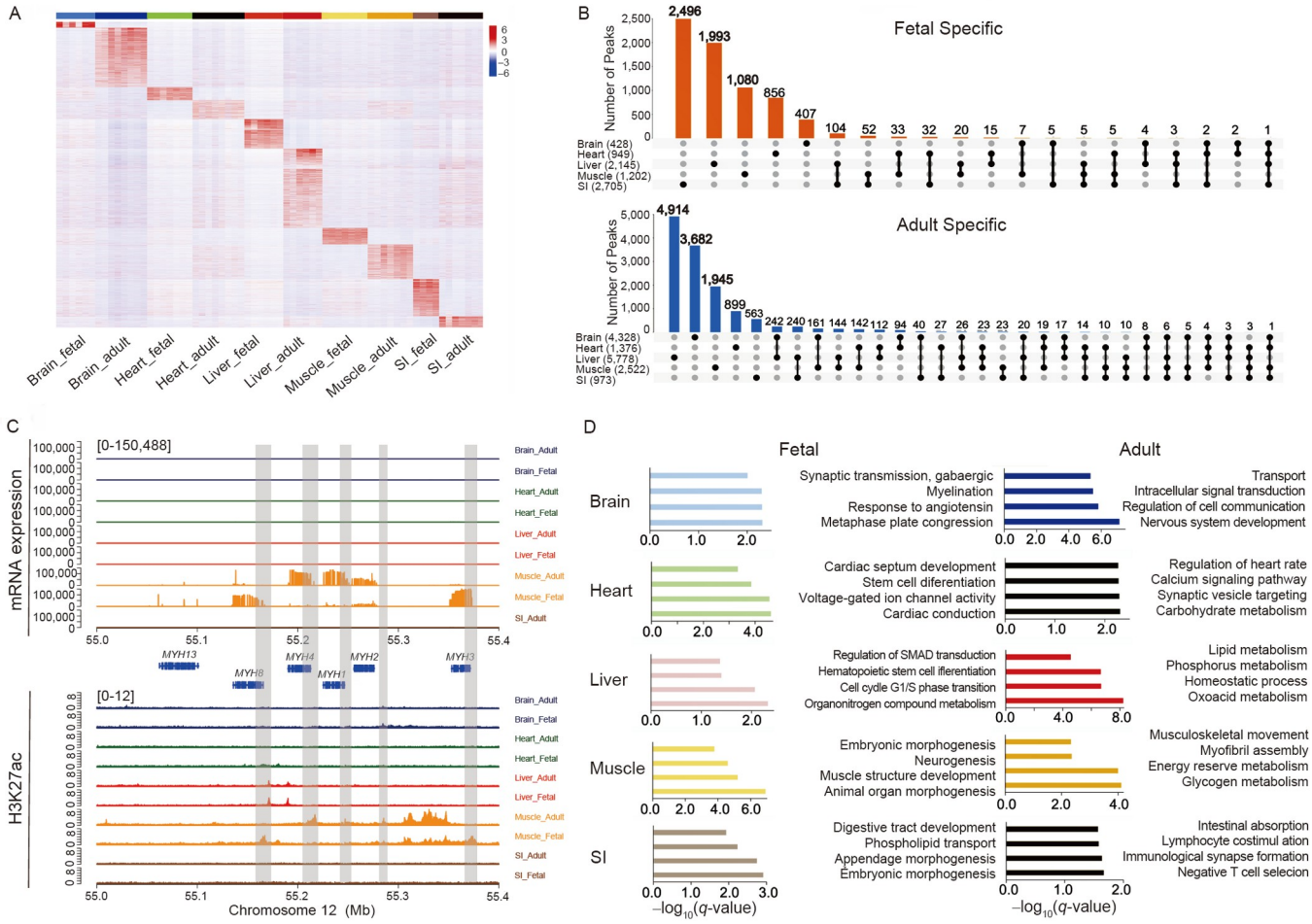


Figure 5 Functional annotation of developmental stage specific H3K27ac peaks in the five tissues. A, Heatmap of all developmental stage specific peaks in the five tissues. Each row of the heatmap shows normalized density of read-depth at one peak. B, Intervene plot showing the number of adult or fetal specific peaks that were shared across the five tissues. C, The H3K27ac activity and mRNA expression tracks in the genomic region on chromosome 12 containing the myosin heavy chain gene family. Each track represents average number of base-pairs of RNA-Seq/H3K27ac ChIP-Seq per million mapped reads in 20-bp bin across all samples in corresponding tissue-developmental stage. The scale of a gene/peak track was set to be consistent across different tissue-developmental stages. D, Bar plot showing representative biological processes enriched in target genes of the developmental stage specific peaks in the five tissues.

intestine specific peaks, and HNF4a (maturation of fetal intestine) among fetal intestine specific peaks (Figure S17 in Supporting Information). Further independent analyses or experiments are required to validate these findings.

Identifying H3K27ac peaks with differential activity between Bama Xiang and Large White pigs

To reveal peaks differing in activity between the two investigated breeds, we performed a linear regression analysis and revealed 264 peaks with significant differential activity between Large White and Bama Xiang pigs at a q value threshold of 0.05 (Table S10 in Supporting Information). These peaks were associated with genes involved in the IL-6 signalling pathway (Benjamini-Hochberg corrected q=0.011), ovarian follicle development (q=0.012) and response to thyroid hormone (q=0.013) (Table S11 in Supporting Information). They were enriched for the binding motifs of three TFs, ERG (q=0.016), GAPBA (q=0.033) and CEBP (q=0.033) (Table S12 in Supporting Information). Among the 264 peaks, 52 and 13 consistently showed higher activity in samples across all tissue-development stages in Bama Xiang or Large White pigs, respectively (Figure S18A in Supporting Information). Two representative examples are shown in Figure S18B–E in Supporting Information. The relevance of these peaks and their target genes to the phenotypic differences between the two breeds requires further investigation.

H3K27ac and DNA methylation in the *XIST* gene promoter region regulate X chromosome inactivation

We next investigated the impact of sex on peak activity. At a q value threshold of 0.05, we identified 13 sex-biased peaks, including two located on the X chromosome and 11 located on the Y chromosome (Figure 6A). The H3K27ac peak

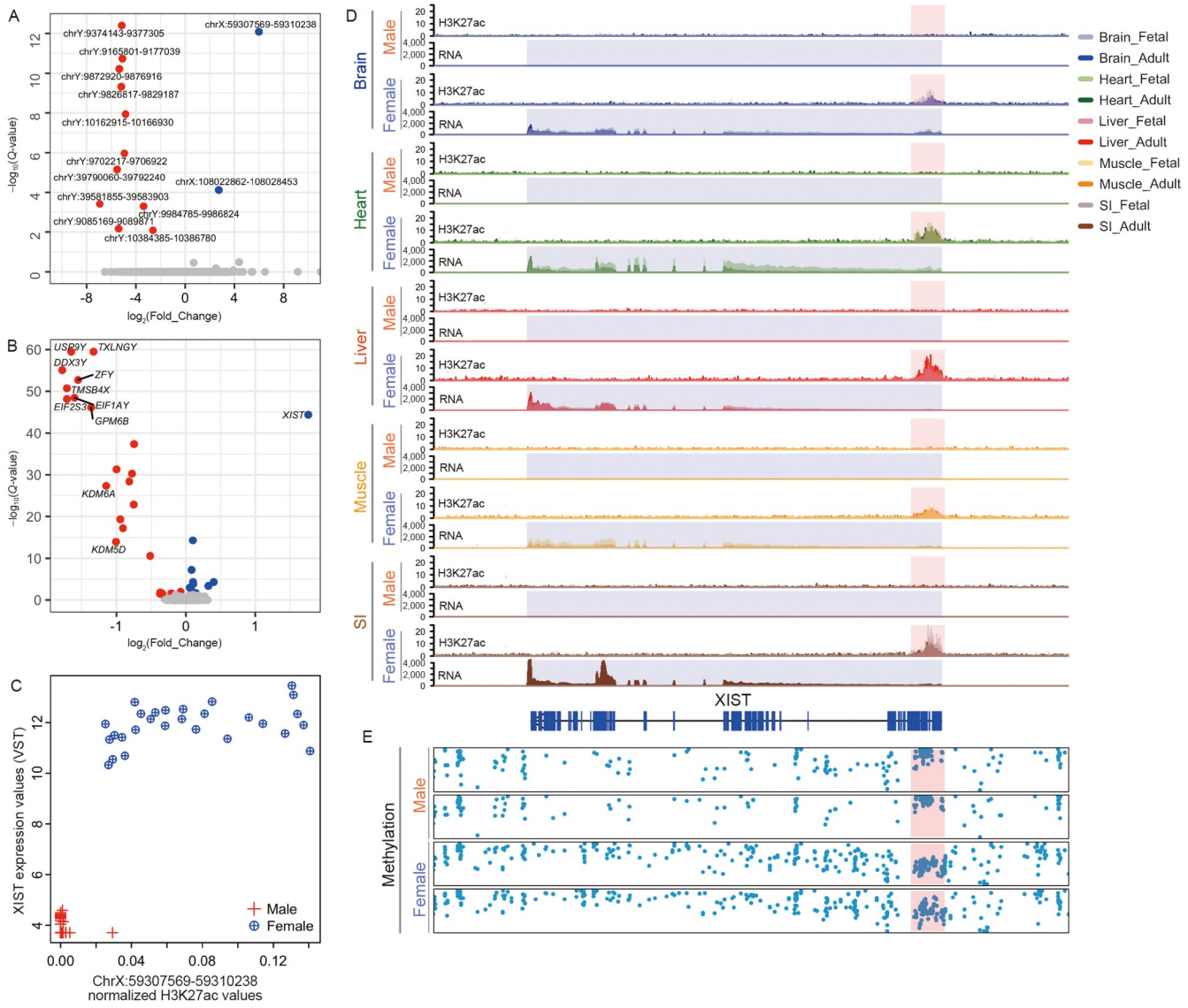


Figure 6 Identification of H3K27ac peak involved in X chromosome inactivation. A, Volcano plot of peaks exhibiting significant gender specific activity. B, Volcano plot of differentially expressed genes between the two genders, the annotated genes shown higher expression in males are all located on Y chromosome. C, Scatter plot showing the correlation between the peak at chrX: 59,307,569–59,310,238 with expression of *XIST* gene. D, Comparison of the H3K27ac ChIP-Seq and RNA-Seq tracks at *XIST* gene regions in males and females across the tissue-developmental stages; *Y* axis corresponding to the normalized read depths of ChIP-Seq or RNA-Seq data. Each track represents average number of base-pairs of H3K27ac ChIP-Seq/RNA-Seq per million mapped reads in 20-bp bin across all samples in corresponding tissue-developmental stage. The scale of a peak/gene track was set to be consistent across different tissue-developmental stages or sexes. E, DNA methylation signals at *XIST* gene region in two female and two male Large White pigs. *Y* axis indicates the percentage of reads covering corresponding sequence bases that were methylated, and a dot represents a CpG island.

region from 59,307,569–59,310,238 on chromosome X coincided with the promoter region of *XIST* and exhibited female-specific activity across all five tissues and two developmental stages (Figure 6B–D). This coincided with the female-specific expression of the *XIST* gene, which is a well-studied noncoding gene driving X chromosome inactivation (Marahrens et al., 1998). As DNA methylation is highly related to H3K27ac (Charlet et al., 2016), we further examined the distribution of muscle DNA methylation based on whole-genome bisulphite sequencing (WGBS) in the *XIST* gene region in two female and two male adult Large White pigs (Zheng et al., 2020) and found that the peak region (chrX: 59,307,569–59,310,238) was fully methylated in males and hemimethylated in females (Figure 6E). We assume that this complete methylation in males is accompanied by the deacetylation of H3K27 in the promoter region of *XIST* and the silencing of the *XIST* gene, while in females, the promoter region of *XIST* from the active X chromosome is fully methylated, and the corresponding region on the inactivate X chromosome is demethylated and associated with the H3K27 acetylation and transcription of the *XIST* gene, which in turn initiates the

downstream pathways of X chromosome inactivation. Taken together, these analyses support the involvement of H3K27ac and DNA methylation in the promoter region of the *XIST* gene in the regulation of X chromosome inactivation. To the best of our knowledge, this has not yet been reported elsewhere, although the initiation, spreading and maintenance of X inactivation in mammals are well studied phenomena (Marks et al., 2009). Therefore, the analysis presented herein provides complementary information about this fundamental biological process conserved across placental mammals.

Characterization of super-enhancers from H3K27ac ChIP-Seq data

Super-enhancers are groups of enhancers located in close genomic proximity to each other that are associated with strong enrichment for the binding of TFs and mediators and have important functions in the maintenance of cell identity (Pott and Lieb, 2015; Sengupta and George, 2017). We identified an average of 1,124 super-enhancers within each H3K27ac ChIP-Seq sample using the ROSE program (Whyte et al., 2013), which were grouped into a total of 6,713 merged super-enhancers. We then compared the activity of the 6,173 merged super-enhancers among different tissues, development stages, breeds and sexes using the same procedures applied to the ordinary H3K27ac peaks. Among the 6,713 super-enhancers, 6.3% (424/6,713) and 16.0% (1,071/6,713) were identified as tissue specific or developmental stage specific, respectively (Figures S19A and S20 in Supporting Information). In comparison, 2.1% (2,162/101,290) and 22.1% (22,406/101,290) of the ordinary H3K27ac peaks were classified as tissue specific or developmental stage specific, respectively, suggesting that the super-enhancers were more likely to play tissue-specific roles than the ordinary peaks, consistent with the tissue-specific activity of super-enhancers reported previously (Zhao et al., 2021). For example, genes including DPYSL4 (regulator of hippocampal neuron development) (Quach et al., 2007), MYLK3 (regulation of cardiac myocyte contraction) (Tsukamoto and Kitakaze, 2013), ARG1 (a protein predominantly released from hepatocytes) (Poillet-Perez et al., 2018) and SIX4 (skeletal myogenesis) (Chakroun et al., 2015), which have functions related to the biology of corresponding tissues, were located within the regions of brain-, heart-, liver- and muscle-specific super-enhancers, respectively (Figure S19B in Supporting Information). We identified 9 and 0 super-enhancers with significant differential activity between different breeds or sexes, respectively, at thresholds of Q value=0.05 and fold change=2 (Figure S21 in Supporting Information), suggesting that breed and sex have a negligible influence on the activity of super-enhancers.

Application of the H3K27ac profile for the genetic dissection of complex traits in pigs

Characterizing the regulatory elements that control relevant genes for complex traits is a key step in understanding the regulatory mechanisms underlying variations in quantitative traits. Here, we investigated the mRNA expression patterns of three previously reported complex trait-related genes (*PHKG1*, *LDLR* and *MSTN*) and the nearby H3K27ac landscape across ten tissue developmental stages. We showed that *PHKG1*, which harbor locus a splicing mutation associated with the pH level and water holding capacity of meat (Ma et al., 2014), displayed adult muscle-restricted expression that was associated with adult muscle-specific activity of an H3K27ac peak in the first intron of this gene (Figure S22 in Supporting Information). The *LDLR* gene, encoding the low-density lipoprotein receptor, which has been reported to be associated with hypercholesterolemia in pigs (Hasler-Rapacz et al., 1998), was also found to show liver-biased expression associated with two liver-biased active H3K27ac peaks located in the *LDLR* gene body (Figure S22 in Supporting Information). *MSTN*, a gene that is well established to be associated with muscle growth in various livestock species, including pigs (Aiello et al., 2018), displayed muscle-specific expression, which was concordant with a muscle-specific H3K27ac peak in this gene (Figure S22 in Supporting Information). These examples demonstrate that the present resources provide helpful information for understanding the regulatory basis of the tissue-specific expression of complex trait-related genes.

Identifying causative mutations underlying the variations in complex traits is a key objective of genetic studies. To examine whether the H3K27ac data could help to prioritize causal variants for complex traits, we examined the genomic region harboring a well-known causal mutation (chr2:1483817) in the first intron of the *IGF2* gene that increases muscle growth in pigs (Van Laere et al., 2003). We found that the causal mutation was located within an H3K27ac peak that showed greater activity in fetal liver and muscle; correspondingly, the *IGF2* gene displayed higher expression in fetal liver and muscle tissues (Figure 7A and B). The causal mutation is known to abrogate the binding of an inhibitor that is normally expressed in postnatal skeletal muscle. Notably, the exact position of the mutation coincides with a ~300 bp trough in the H3K27ac peaks found in fetal muscle and liver. The precise meaning of this observation remains to be elucidated (Figure 7C). Overall, these results suggest that the H3K27ac signal data generated herein are valuable resources that will be helpful for prioritizing causal variants and understanding the regulatory mechanisms of causal genes for quantitative traits in pigs.

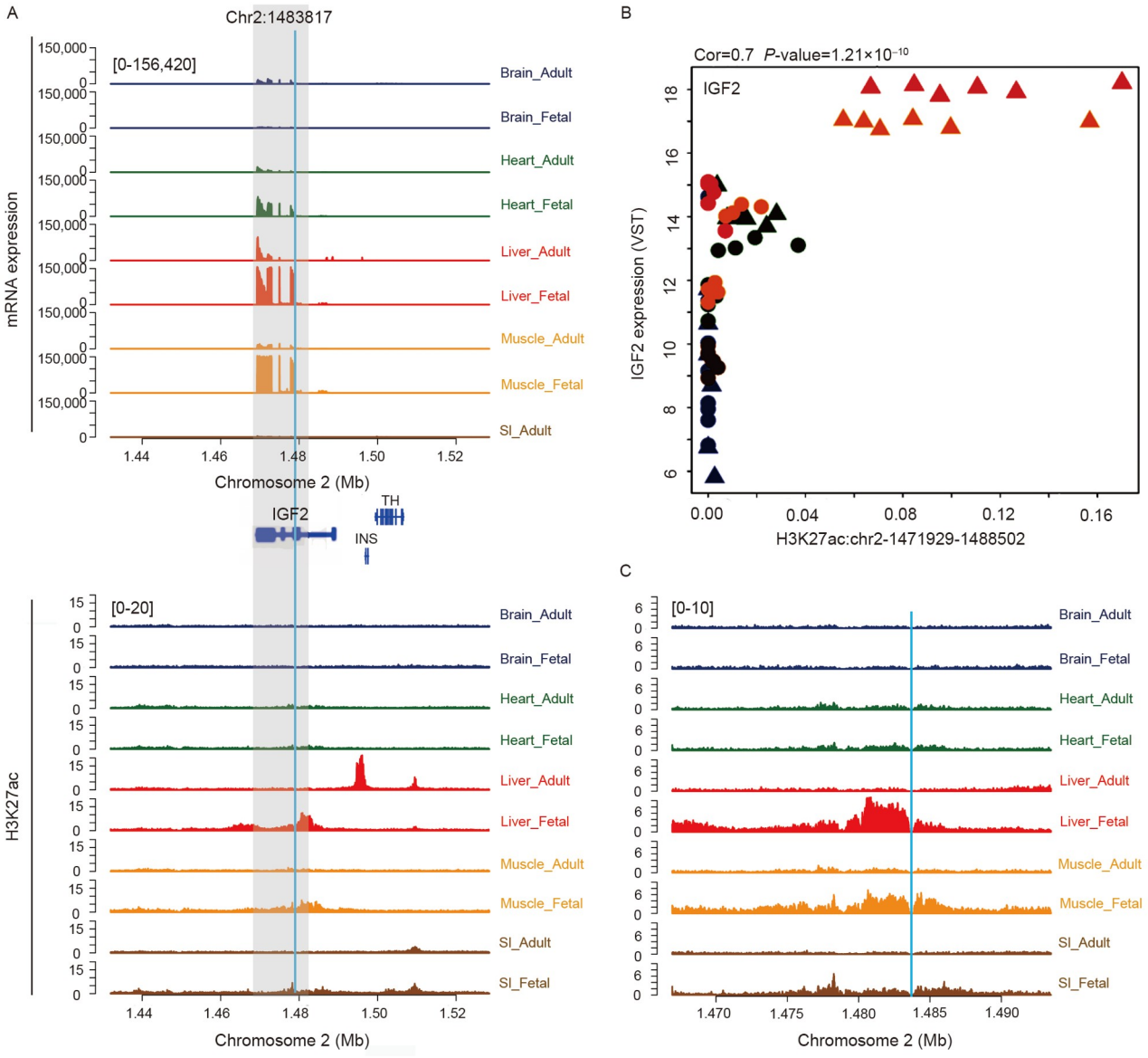


Figure 7 Prioritizing causative mutation using H3K27ac ChIP-Seq data. A, The causal mutation for muscle growth at first intron of *IGF2* (chr2:1483817) genes located within an H3K27ac peaks in first intron of *IGF2* gene exhibiting greater activity in fetal liver and muscle in pigs. B, Correlation of the intronic peak harboring the causal mutation with expression of IGF2 across samples. C, A zoom-in plot of the H3K27ac ChIP-Seq track reveals the exact position of the mutation coincides with an ~300 bp trough in the H3K27ac peaks in fetal muscle and liver. Each track (A and C) represents the average number of base-pairs of RNA-Seq/H3K27ac ChIP-Seq per million mapped reads in 20-bp bin across all samples in corresponding tissue-developmental stage. The scale of a gene/peak track was set to be consistent across different tissue-developmental stages.

A web-based browser providing easy access to these H3K27ac peak and gene expression data

To improve the accessibility of the data and results generated in this study, we developed a web-based browser that allows the fast navigation of (i) the H3K27ac ChIP-Seq and RNA-Seq read coverage of the ten different tissue-developmental stages across the genome, and (ii) the links between H3K27ac peaks and nearby genes based on significant correlations. The browser is largely self-explanatory and can be easily customized to visualize the subset of ChIP-Seq or RNA-Seq tracks. The colors of the tracks can be manually specified. Moreover, the visualized tracks can be saved as high-resolution publishable figures. We expect that these attributes of the browser will be particularly helpful for colleagues who are interested in examining candidate regions associated with a certain complex trait or regions that are subject to artificial or natural selection (http://segtp.jxau.edu.cn/pencode/?genome=susScr11).

DISCUSSION

Active promoters and enhancers, marked by H3K27ac

chromatin modification, are important sequences that regulate gene transcription (Creyghton et al., 2010). H3K27ac marks show highly dynamic activity across different tissues (Kundaje et al., 2015) and developmental stages (Gorkin et al., 2020) and are enriched in disease-associated variants (Kundaje et al., 2015). In this study, H3K27ac ChIP-Seq was performed on samples from individuals on prenatal day 75 and postnatal day 150, and the results were considered alongside the data of Zhao et al. (2021) obtained from samples analyzed at two weeks postnatal. Moreover, the Bama Xiang pigs investigated in this study were not included in the study of Zhao et al. (2021). Overall, our study is largely complementary to that of Zhao et al. (2021). There are several lines of evidence that support the quality of our H3K27ac ChIP-Seq data. These include (i) the high correlation of peak activities between technical replicates (Figure S23 in Supporting Information); (ii) the cumulative distribution of read depths drawn by plotFingerprint from deepTools (http://deeptools.ie-freiburg.mpg.de), which showed that the sequences from the ChIP-Seq experiment were significantly enriched in certain regions of the genome (Figure S24 in Supporting Information); (iii) FRiP and library complexity values, including NRF, PBC1, and PBC2 statistics, for most of the samples greater than the values recommended by the ENCODE project (the FRiP score of all samples was greater than 1%; in terms of library complexity, NRF and PBC1 were greater than 0.7, and PBC2 was greater than 3); (iv) the high proportion of liver peaks overlapping with those reported in independent studies; (v) the highly significant correlations of proximal peaks with their cognate genes within each sample; and (vi) the accurate classification of samples according to their tissue and developmental stage based on the quantified peak activity.

A total of 62 samples covering all tissues, ages, sexes and breeds were simultaneously analyzed by both H3K27ac ChIP-Seq and RNA-Seq, allowing us to reveal the landscape of the associations between H3K27ac activity and gene expression. Notably, we observed considerable numbers of significant negative peak-gene correlations, which showed different distributions in terms of peak-gene distances relative to positive peak-gene correlations, reflecting a different regulatory mechanism between positive and negative peak-gene correlations. The observed negative peak-gene correlations could also be attributed to the effects of TFs or other histone modifications that had not been assayed. Studies have been implemented to identify silencers through the high-throughput screening of genomic sequences that repress the transcription of cell death proteins (Pang and Snyder, 2020) or the correlation of H3K27me3-DNase I hypersensitive site (DHS) peaks with nearby genes across diverse cell lines (Huang et al., 2019). Nevertheless, very few studies have reported negative correlations of H3K27ac activity with gene expression. Further experimental validation is required to confirm the assumed silencing effect of H3K27ac on gene transcription.

We revealed a small proportion (2.6%, 2,679 in 101,290) of peaks showing constitutive activity across the ten tissue developmental stages, most of which were proximal peaks. These peaks were demonstrated to be associated with the high, stable expression of genes across tissue stages, indicating that the intergenic or intronic regions located close to ubiquitously active H3K27ac sites are good candidate safe harbor regions for genomic engineering in pigs. Thus far, studies aimed at characterizing such safe harbor locus in pigs have focused on well-known housekeeping genes such as GAPDH (Han et al., 2019b) and ACTB (Xiong et al., 2020). This study provides over one hundred candidate safe harbor locus containing genes that showed higher and more stable expression relative to two known safe harbor locus (H11 and Rosa26), hence greatly expanding the list of candidate safe harbor regions in pigs.

Consistent with reports in humans and mice that enhancers marked by H3K27ac show highly variable activity across tissues and developmental stages (Gorkin et al., 2020; Kundaje et al., 2015; Nord et al., 2013), we also identified a large number of tissue- and developmental stage-specific distal peaks. We found that these peaks were enriched for TF binding sites and regulated genes with functions matching the biology of the corresponding tissues and stages, such as TFs and gene pathways related to haematopoiesis in fetal liver and immune responses in the adult small intestine, indicating that these ChIP-Seq and RNA-Seq data covering different tissues and developmental stages provide a valuable resource is helpful for understanding the regulatory mechanisms underlying the tissue differentiation and development of pigs. Nevertheless, further experimental validation, such as those in Wang et al. (2019), would be needed to consolidate the observation in this study. The developed browsers further improve the accessibility of the data generated in this study.

The analysis of sex-differentiated activity suggested the involvement of H3K27ac and methylation in the promoter region of the *XIST* gene in regulating the expression of *XIST*, which in turn initiates the pathways underlying X chromosome inactivation. A number of studies have been performed to characterize the downstream processes regulated by *XIST* genes, such as histone modifications along the X chromosome (Żylicz et al., 2019) and the expression patterns of genes such as those escaping X chromosome inactivation (Tukiainen et al., 2017). To our knowledge, this is the first study to report the association of histone acetylation and DNA methylation with *XIST* gene expression and, in turn, X chromosome inactivation in pigs, and very few if any studies have reported the links of H3K27ac and DNA methylation with the regulation of the *XIST* gene in humans and other model organisms.

We also explored the utility of the data for interpreting genes and mutations associated with economically important traits in pigs. We showed that the expression of trait-associated genes such as *MYH3*, *PHKG1*, *MSTN* and *LDLR* was correlated with H3K27ac peaks and displayed tissue- or stage-specific expression, thus enhancing our knowledge of the epigenetic regulatory basis of trait-associated genes, which are usually investigated at gene and protein expression levels (Zhang et al., 2021; Zong et al., 2021). Moreover, we have provided a proof-of-principle example that a well-known causal mutation related to muscle growth was located within a peak showing higher activity in fetal muscle and liver; this was associated with higher expression of the *IGF2* gene in fetal muscle and liver tissues, in accord with a recent report on the gene expression pattern of the *IGF2* gene in mice (Younis et al., 2018). These results demonstrate that the data obtained in this study are valuable for characterizing the regulatory pattern of candidate genes and prioritizing the causal variants associated with complex traits in pigs. However, the extensive intersection of the peaks identified herein with GWAS loci for various traits in pigs (e.g., from the pig QTL database) (Hu et al., 2013) revealed little significant enrichment (data not shown). We reasoned that a large proportion of the association studies conducted in pigs to date were based on SNP arrays and involved a single or limited number of populations and that many of the identified loci reside in regions with high linkage disequilibrium (Gong et al., 2019; Zhang et al., 2017), which presents a challenge in the enrichment analysis of trait-associated loci. We expect that with the growth of genome sequence-based GWAS locus catalogues in pigs, our data will provide a valuable map for characterizing the regulatory mutations underlying complex traits in pigs.

MATERIALS AND METHODS

Samples

Five tissues, including brain (cortex), heart (left ventricle), liver, small intestine (duodenum), and skeletal muscle (longissimus dorsi) samples, were collected from seven fetuses (two males and two females of the Large White breed at day 75 and two males and one female of the Bama Xiang breed at day 74 post insemination) and eight adult pigs (two males and two females of the Large White breed at day 150 and two males and two females of the Bama Xiang at days 132–141 postnatal) (Figure 1; Table S1 in Supporting Information). We carefully dissected the tissues following the standardized sample collection protocols of the FAANG Project (https://www.faang.org/bbs?s=protocols.txt). The tissues were flash frozen in liquid nitrogen immediately after collection and then stored at −80°C.

These samples represent tissues developing from all three germ layers: ectoderm (brain cortex), mesoderm (heart left ventricle and *longissimus* muscle) and endoderm (liver and duodenum) and are relevant for meat production and biomedical research. The brain is frequently studied to better understand psychiatric and mental disorders in humans (Dedova et al., 2009) and is also a potentially useful target for studying pig behavior. The heart serves as a pump that circulates blood and is essential for survival, and pigs are potential heart donors for humans (McGregor and Byrne, 2017). *Longissimus* muscle mass is directly relevant to meat production. The liver is central to lipid and protein metabolism and detoxification, and plays a major role in fetal haematopoiesis (Zhao and Duncan, 2005). The small intestine is an organ involved in food digestion and may affect feed conversion in pigs (Vigors et al., 2016). The inclusion of two developmental stages in the study allowed investigation of the regulatory basis underlying the functional switch in a tissue from the prenatal to the postnatal stage. Moreover, the two investigated breeds, Large white and Bama Xiang, represent the two major swine domestication centers (Near East and China) (Larson et al., 2005) and differ considerably in a variety of phenotypes, including body size, coat colour and fat deposition. The wild ancestors of the two breeds diverged approximately 1.2 million years ago (Frantz et al., 2013).

H3K27ac ChIP-Seq

The chromatin immunoprecipitation of the samples was performed using the SimpleChIP Plus Enzymatic Chromatin IP Kit (Magnetic Beads) (CST, USA). Briefly, approximately 200–300 mg of a sample was minced in 1 mL of PBS and cross-linked using 37% formaldehyde at room temperature for 10 min. Cross-linking was terminated by adding 10× glycine for 5 min, and the sample was lysed with buffer A and B in the IP kit. The DNA was then sheared to a target size of 100–300 bp by sonication in 500 μL of ChIP buffer in the IP kit, and 10 μL of the DNA solution was retained as the input. ChIP was performed by incubation with 5 μg of an H3K27ac antibody (Active Motif, USA) overnight. The DNA bound by the antibody was enriched using ChIP-Grade Protein G Magnetic Beads in the IP kit and then incubated with 2 μL of 20 mg mL^{-1} proteinase K and 6 μL of 5 mol L^{-1} NaCl at 65°C to reverse the cross-links. The immunoprecipitated DNA was then extracted. DNA sequencing was performed on an Illumina HiSeq 2500 in a single-end 50-bp run on the Illumina HiSeq 2500 platform, and the raw sequence reads were filtered by removing whole reads with (i) contaminated adaptor sequences; (ii) more than half of the bases with Phred quality scores <19; and (iii) >5% ambiguous/undetermined (N) bases.

H3K27ac peak calling and processing

The clean reads were mapped to the pig reference genome

Sscrofa 11.1 using the Burrows-Wheeler Aligner (BWA) (Abuín et al., 2015), allowing two mismatches. The Model-based Analysis for ChIP-Seq (MACS) version 2.1.0 peak caller was used to infer peaks in each individual with the "–broad and –broad-cutoff 0.1" option (Zhang et al., 2008). bedtools was used to combine the peaks identified in each individual into a reference peak set by ignoring the largest 5% of peaks, with peak lengths greater than 13.1 kb, as we considered these large peaks to potentially represent regions containing multiple adjacent peaks, in which we could not quantify individual peak activity accurately. The extent of developmental stage-specific peaks that were shared across tissues was visualized using the Intervene program (Khan and Mathelier, 2017). Fingerprint analysis was performed using the deepTools (http://deeptools.ie-freiburg.mpg.de) plotFingerprint module. Library complexity statistics, including NRF, PBC1, and PBC2, and FRiP values were calculated using bedtools bamtobed and multicov modules and the SAMtools flagstat module.

Identification of putative safe harbor locus from the ChIP-Seq data

The putative safe harbor locus indicated by ubiquitous activity were filtered out based on the 70 ChIP-Seq, 62 RNA-Seq and 4 WGBS datasets reported in this study following a series of carefully selected criteria: the peaks should (i) be proximal and located within 500 kb from the TSSs of genes, (ii) not be located within gene exons, (iii) not be located on the X or Y chromosome, (iv) contain at least one CpG island, (v) be active in at least 80% of samples, (vi) be ranked in the top 20% among all peaks according to their average activity, (vii) cover the promoter of gene that were expressed in at least 80% of samples, (viii) cover the promoter of gene that were ranked in the top 20% among all genes according to their average expression values, and (ix) show average expression values greater than any genes within the two known safe harbor locus (ROSA26: *ROSA* and *SEDT5*; H11: *DRG1* and *EIF4ENIF1*).

Functional annotations of ChIP-Seq peaks and GO enrichment analysis

The pig whole-genome GERP score was downloaded from the Ensembl Database, and we computed the GERP score of a peak by averaging the GERP scores of bases within that peak. The GC content of a peak was calculated using EMBOSS geecee software (Rice et al., 2000). The enrichment of different genomic features and TFBSs was conducted using the HOMER program. Gene ontology enrichment analysis was performed using ClueGO (Bindea et al., 2009). The *P* values for the enrichment of Gene Ontology (GO) terms were corrected using the Benjamini-Hochberg approach.

RNA-Seq

Total RNA was isolated from tissues using the TRIzol reagent (Invitrogen, USA), and mRNA was enriched using magnetic beads attached to poly-T oligos. cDNA was then synthesized using random hexamer primers. The cDNA was then purified, ligated to index adapters, size selected using AMPure XP beads, and PCR amplified to generate strand-specific cDNA libraries. The library was sequenced via paired-end 150 (PE150) sequencing run on the Illumina HiSeq 4000. The raw reads contaminated with adapter sequences and those with (i) more than 50% low-quality bases (Q<19) and (ii) a percentage of N bases>10% were removed. The clean reads were mapped to the pig reference genome Sscrofa 11.1 using STAR-2.5.3a (Dobin et al., 2013). The transcripts were assembled and merged with StringTie by referring to Ensembl GTF (98.111), which was further merged with Ensembl GTF (98.111) to obtain a customized GTF file for use as the reference for gene expression quantification, which contained a total of 44,415 genes/transcripts, including 24,129 protein-coding genes, 10,060 *de novo* assembled genes or transcripts, 6,985 long noncoding RNAs, 1,593 pseudogenes, 474 snoRNAs, 440 miRNAs and 773 other transcripts. The expression levels of genes were quantified using FeatureCounts 1.5.3 (Liao et al., 2014). We retained genes supported by at least 500 counts in 62 RNA-Seq samples. After removing mitochondrial genes, the VST algorithm from DESeq2 (Love et al., 2014) was used to normalize the expression values of each sample.

Statistical analyses

Unless otherwise stated, most of the statistical analyses were performed in the R program, version 3.5.1. The contributions of the tissue, development stage, breed and sex to the variance of the quantitative peak activities and gene expression levels were quantified with a linear mixed model as follows (Hoffman and Schadt, 2016): y=tissue+stage+breed+sex+ε, where y is a vector of the activities of a peak or the expression values of a gene; tissue, developmental stage, breed and sex were fitted as random effects; and ε represents the residuals, which were assumed to follow a normal distribution with variance of σ_{ε}^2. The total variance of a H3K27ac peak or mRNA expression trait (σ_y^2) was assumed to be the sum of the variance of the four random effects and the residuals. The variations in H3K27ac peak activity or mRNA expression explained by the four factors were averaged across all peaks or genes.

We used both entropy-based statistics (Q value) (Schug et al., 2005) and Tau values (Kryuchkova-Mostacci and Robinson-Rechavi, 2017) to assess the tissue, developmental stage specificity of the H3K27ac peaks. The Q value of a

peak in tissue i was computed as follows: $Q_i=H-\log_2(p_i)$, where H is the entropy of the peak activity distribution defined as $\sum_{i=1}^{N}-p_i\log_2(p_i)$, where $p_i=x_i/\sum_{i=1}^{N}x_i$; and the Tau value was defined as follows: $\sum_{i=1}^{N}\frac{1-\hat{x_i}}{N-1}$ (where $\hat{x_i}=x_i/\max(x_i)$, where x_i is the average activity of a peak in tissue i, and N is the total number of tissues under study).

In the correlation analyses of H3K27ac activity with gene expression traits, the H3K27ac peak activity and gene expression traits were first adjusted for the breed, developmental stage, and sex using the lm function in R, and the Spearman correlations between the residuals of peak activity and nearby genes were computed using the R program. We consider most of these long-distance peak-gene correlations to be background noise, although some of them could reflect long-distance enhancer-promoter interactions. To obtain a conservative estimate of peak-gene correlations, we pruned the number of significant correlations identified within 500 kb for each peak by the number of significant peak-gene correlations identified within 1.5–2 Mb regions from the corresponding peak and finally obtained 41,105 significant correlations (35,644 positive and 5,461 negative). To identify peaks showing differences in activity between the two breeds, we used the lm function in R and added tissue, developmental stage and sex as fixed effects. A similar analysis was performed to assess the effect of sex on peak activity. The q values corresponding to the P values from the correlation and lm analyses were determined using the R package qvalue (Storey et al., 2020).

Neighbour-joining tree

The Neighbour-joining (NJ) trees of the samples were constructed using the hclust function in the R program with the parameter "method=ward.D", where the distance between a pair of samples was defined as $1-r_{\text{spearman}}$, in which r_{spearman} is the Spearman correlation coefficient of genome-wide peak activities or gene expression. The NJ trees were plotted using the R package GGraph (Murtagh and Legendre, 2014).

Compliance and ethics *The author(s) declare that they have no conflict of interest.*

Acknowledgements *This work was supported by the National Natural Science Foundation of China (31790413, 31760657). We are grateful to colleagues in State Key Laboratory of Pig Genetic Improvement and Production Technology, Jiangxi Agricultural University for sample collection.*

References

Abuín, J.M., Pichel, J.C., Pena, T.F., and Amigo, J. (2015). BigBWA: approaching the Burrows-Wheeler aligner to Big Data technologies. Bioinformatics 31, 4003–4005.

Aiello, D., Patel, K., and Lasagna, E. (2018). The *myostatin* gene: an overview of mechanisms of action and its relevance to livestock animals. Anim Genet 49, 505–519.

Almalki, S.G., and Agrawal, D.K. (2016). Key transcription factors in the differentiation of mesenchymal stem cells. Differentiation 92, 41–51.

Andersson, L., Archibald, A.L., Bottema, C.D., Brauning, R., Burgess, S. C., Burt, D.W., Casas, E., Cheng, H.H., Clarke, L., Couldrey, C., et al. (2015). Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. Genome Biol 16, 57.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pagès, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 25, 1091–1093.

Chakroun, I., Yang, D., Girgis, J., Gunasekharan, A., Phenix, H., Kærn, M., and Blais, A. (2015). Genome-wide association between Six4, MyoD, and the histone demethylase Utx during myogenesis. FASEB J 29, 4738–4755.

Chan, R.Y.Y., Boudreau-Lariviere, C., Angus, L.M., Mankal, F.A., and Jasmin, B.J. (1999). An intronic enhancer containing an N-box motif is required for synapse- and tissue-specific expression of the acetylcholinesterase gene in skeletal muscle fibers. Proc Natl Acad Sci USA 96, 4627–4632.

Charlet, J., Duymich, C.E., Lay, F.D., Mundbjerg, K., Dalsgaard Sørensen, K., Liang, G., and Jones, P.A. (2016). Bivalent regions of cytosine methylation and H3K27 acetylation suggest an active role for DNA methylation at enhancers. Mol Cell 62, 422–431.

Cho, I.C., Park, H.B., Ahn, J.S., Han, S.H., Lee, J.B., Lim, H.T., Yoo, C.K., Jung, E.J., Kim, D.H., Sun, W.S., et al. (2019). A functional regulatory variant of *MYH3* influences muscle fiber-type composition and intramuscular fat content in pigs. PLoS Genet 15, e1008279.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci USA 107, 21931–21936.

Dedova, I., Harding, A., Sheedy, D., Garrick, T., Sundqvist, N., Hunt, C., Gillies, J., and Harper, C.G. (2009). The importance of brain banks for molecular neuropathological research: the New South Wales Tissue Resource Centre experience. Int J Mol Sci 10, 366–384.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Foissac, S., Djebali, S., Munyard, K., Vialaneix, N., Rau, A., Muret, K., Esquerré, D., Zytnicki, M., Derrien, T., Bardou, P., et al. (2019). Multi-species annotation of transcriptome and chromatin structure in domesticated animals. BMC Biol 17, 108.

Frantz, L.A.F., Schraiber, J.G., Madsen, O., Megens, H.J., Bosse, M., Paudel, Y., Semiadi, G., Meijaard, E., Li, N., Crooijmans, R.P.M.A., et al. (2013). Genome sequencing reveals fine scale diversification and reticulation history during speciation in *Sus*. Genome Biol 14, R107.

Friedrich, G., and Soriano, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes Dev 5, 1513–1523.

Gallegos, J.E., and Rose, A.B. (2017). Intron DNA sequences can be more important than the proximal promoter in determining the site of transcript initiation. Plant Cell 29, 843–853.

Georges, M., Charlier, C., and Hayes, B. (2019). Harnessing genomic information for livestock improvement. Nat Rev Genet 20, 135–156.

Gong, H., Xiao, S., Li, W., Huang, T., Huang, X., Yan, G., Huang, Y., Qiu, H., Jiang, K., Wang, X., et al. (2019). Unravelling the genetic loci for growth and carcass traits in Chinese Bamaxiang pigs based on a 1.4 million SNP array. J Anim Breed Genet 136, 3–14.

Gorkin, D.U., Barozzi, I., Zhao, Y., Zhang, Y., Huang, H., Lee, A.Y., Li, B., Chiou, J., Wildberg, A., Ding, B., et al. (2020). An atlas of dynamic chromatin landscapes in mouse fetal development. Nature 583, 744–751.

Han, K., Ren, R., Cao, J., Zhao, S., and Yu, M. (2019a). Genome-wide identification of histone modifications involved in placental development in pigs. Front Genet 10, 277.
Han, X., Xiong, Y., Zhao, C., Xie, S., Li, C., Li, X., Liu, X., Li, K., Zhao, S., and Ruan, J. (2019b). Identification of glyceraldehyde-3-phosphate dehydrogenase gene as an alternative safe harbor locus in pig genome. Genes 10, 660.
Hasler-Rapacz, J., Ellegren, H., Fridolfsson, A.K., Kirkpatrick, B., Kirk, S., Andersson, L., and Rapacz, J. (1998). Identification of a mutation in the low density lipoprotein receptor gene associated with recessive familial hypercholesterolemia in swine. Am J Med Genet 76, 379–386.
Hoffman, G.E., and Schadt, E.E. (2016). variancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinformatics 17, 483.
Hu, Z.L., Park, C.A., Wu, X.L., and Reecy, J.M. (2013). Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Res 41, D871–D879.
Huang, D., Petrykowska, H.M., Miller, B.F., Elnitski, L., and Ovcharenko, I. (2019). Identification of human silencers by correlating cross-tissue epigenetic profiles and gene expression. Genome Res 29, 657–667.
Jakovcevski, I., Filipovic, R., Mo, Z., Rakic, S., and Zecevic, N. (2009). Oligodendrocyte development and the onset of myelination in the human fetal brain. Front Neuroanat 3, 5.
Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Várnai, C., Thiecke, M.J., et al. (2016). Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell 167, 1369–1384. e19.
Khan, A., and Mathelier, A. (2017). Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. BMC Bioinformatics 18, 287.
Kirmse, K., Hübner, C.A., Isbrandt, D., Witte, O.W., and Holthoff, K. (2018). GABAergic transmission during brain development: multiple effects at multiple stages. Neuroscientist 24, 36–53.
Kryuchkova-Mostacci, N., and Robinson-Rechavi, M. (2017). A benchmark of gene expression tissue-specificity metrics. Brief Bioinform 18, 205–214.
Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330.
Larson, G., Dobney, K., Albarella, U., Fang, M., Matisoo-Smith, E., Robins, J., Lowden, S., Finlayson, H., Brand, T., Willerslev, E., et al. (2005). Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science 307, 1618–1621.
Li, X., Yang, Y., Bu, L., Guo, X., Tang, C., Song, J., Fan, N., Zhao, B., Ouyang, Z., Liu, Z., et al. (2014). Rosa26-targeted swine models for stable gene over-expression and Cre-mediated lineage tracing. Cell Res 24, 501–504.
Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930.
Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., et al. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. Nature 478, 476–482.
Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550.
Ma, J., Yang, J., Zhou, L., Ren, J., Liu, X., Zhang, H., Yang, B., Zhang, Z., Ma, H., Xie, X., et al. (2014). A splice mutation in the *PHKG1* gene causes high glycogen content and low meat quality in pig skeletal muscle. PLoS Genet 10, e1004710.
Marahrens, Y., Loring, J., and Jaenisch, R. (1998). Role of the *Xist* gene in X chromosome choosing. Cell 92, 657–664.
Marks, H., Chow, J.C., Denissov, S., Françoijs, K.J., Brockdorff, N., Heard, E., and Stunnenberg, H.G. (2009). High-resolution analysis of epigenetic changes associated with X inactivation. Genome Res 19, 1361–1373.
McGregor, C.G.A., and Byrne, G.W. (2017). Porcine to human heart transplantation: is clinical application now appropriate? J Immunol Res 2017, 1–11.
Murtagh, F., and Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? J Classif 31, 274–295.
Nasrallah, R., Imianowski, C.J., Bossini-Castillo, L., Grant, F.M., Dogan, M., Placek, L., Kozhaya, L., Kuo, P., Sadiyah, F., Whiteside, S.K., et al. (2020). A distal enhancer at risk locus 11q13.5 promotes suppression of colitis by Treg cells. Nature 583, 447–452.
Neville, J.J., Orlando, J., Mann, K., McCloskey, B., and Antoniou, M.N. (2017). Ubiquitous Chromatin-opening Elements (UCOEs): applications in biomanufacturing and gene therapy. Biotechnol Adv 35, 557–564.
Nord, A.S., Blow, M.J., Attanasio, C., Akiyama, J.A., Holt, A., Hosseini, R., Phouanenavong, S., Plajzer-Frick, I., Shoukry, M., Afzal, V., et al. (2013). Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. Cell 155, 1521–1531.
Pang, B., and Snyder, M.P. (2020). Systematic identification of silencers in human cells. Nat Genet 52, 254–263.
Papapetrou, E.P., and Schambach, A. (2016). Gene insertion into genomic safe harbors for human gene therapy. Mol Ther 24, 678–684.
Poillet-Perez, L., Xie, X., Zhan, L., Yang, Y., Sharp, D.W., Hu, Z.S., Su, X., Maganti, A., Jiang, C., Lu, W., et al. (2018). Autophagy maintains tumour growth through circulating arginine. Nature 563, 569–573.
Popescu, D.M., Botting, R.A., Stephenson, E., Green, K., Webb, S., Jardine, L., Calderbank, E.F., Polanski, K., Goh, I., Efremova, M., et al. (2019). Decoding human fetal liver haematopoiesis. Nature 574, 365–371.
Pott, S., and Lieb, J.D. (2015). What are super-enhancers? Nat Genet 47, 8–12.
Quach, T.T., Massicotte, G., Belin, M.F., Honnorat, J., Glasper, E.R., Devries, A.C., Jakeman, L.B., Baudry, M., Duchemin, A.M., and Kolattukudy, P.E. (2007). CRMP3 is required for hippocampal *CA1* dendritic organization and plasticity. FASEB J 22, 401–409.
Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16, 276–277.
Ruan, J., Li, H., Xu, K., Wu, T., Wei, J., Zhou, R., Liu, Z., Mu, Y., Yang, S., Ouyang, H., et al. (2015). Highly efficient CRISPR/Cas9-mediated transgene knockin at the H11 locus in pigs. Sci Rep 5, 14253.
Schug, J., Schuller, W.P., Kappen, C., Salbaum, J.M., Bucan, M., and Stoeckert Jr, C.J. (2005). Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biol 6, R33.
Sengupta, S., and George, R.E. (2017). Super-enhancer-driven transcriptional dependencies in cancer. Trends Cancer 3, 269–281.
Stamatoyannopoulos, J.A., Snyder, M., Hardison, R., Ren, B., Gingeras, T., Gilbert, D.M., Groudine, M., Bender, M., Kaul, R., Canfield, T., et al. (2012). An encyclopedia of mouse DNA elements (Mouse ENCODE). Genome Biol 13, 418.
Storey, J.D., Bass, A.J., Dabney, A., and Robinson, D. (2020). qvalue: Q-value estimation for false discovery rate control. R package version 2.22.0. Available from: URL: http://github.com/jdstorey/qvalue.
Tasic, B., Hippenmeyer, S., Wang, C., Gamboa, M., Zong, H., Chen-Tsai, Y., and Luo, L. (2011). Site-specific integrase-mediated transgenesis in mice via pronuclear injection. Proc Natl Acad Sci USA 108, 7902–7907.
Tsukamoto, O., and Kitakaze, M. (2013). Biochemical and physiological regulation of cardiac myocyte contraction by cardiac-specific myosin light chain kinase. Circ J 77, 2218–2225.
Tukiainen, T., Villani, A.C., Yen, A., Rivas, M.A., Marshall, J.L., Satija, R., Aguirre, M., Gauthier, L., Fleharty, M., Kirby, A., et al. (2017). Landscape of X chromosome inactivation across human tissues. Nature 550, 244–248.
Vaillancourt, K., Bedard, N., Bart, C., Tessier, M.., Robitaille, G., Turgeon,

N., Frenette, M., Moineau, S., and Vadeboncoeur, C. (2008). Role of *galK* and *galM* in galactose metabolism by *Streptococcus thermophilus*. Appl Environ Microbiol 74, 1264–1267.
Van Laere, A.S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., Archibald, A.L., Haley, C.S., Buys, N., Tally, M., et al. (2003). A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature 425, 832–836.
Vigors, S., Sweeney, T., O′Shea, C.J., Kelly, A.K., and O′Doherty, J.V. (2016). Pigs that are divergent in feed efficiency, differ in intestinal enzyme and nutrient transporter gene expression, nutrient digestibility and microbial activity. Animal 10, 1848–1855.
Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., et al. (2015). Enhancer evolution across 20 mammalian species. Cell 160, 554–566.
Wang, L., Zhu, F., Yang, H., Li, J., Li, Y., Ding, X., Xiong, X., Ji, F., Zhou, H., and Yin, Y. (2019). Epidermal growth factor improves intestinal morphology by stimulating proliferation and differentiation of enterocytes and mTOR signaling pathway in weaning piglets. Sci China Life Sci 63, 259–268.
Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319.
Xiong, Y., Han, X., Zhang, J., Zhao, G., Wang, Z., Zhuang, R., Nie, X., Xie, S., Li, C., Li, X., et al. (2020). Identification of *ACTB* gene as a potential safe harbor locus in pig genome. Mol Biotechnol 62, 589–597.
Younis, S., Schönke, M., Massart, J., Hjortebjerg, R., Sundström, E., Gustafson, U., Björnholm, M., Krook, A., Frystyk, J., Zierath, J.R., et al. (2018). The ZBED6-IGF2 axis has a major effect on growth of skeletal muscle and internal organs in placental mammals. Proc Natl Acad Sci USA 115, E2048–E2057.
Zhang, J., Zhang, Y., Gong, H., Cui, L., Huang, T., Ai, H., Ren, J., Huang, L., and Yang, B. (2017). Genetic mapping using 1.4M SNP array refined loci for fatty acid composition traits in Chinese Erhualian and Bamaxiang pigs. J Anim Breed Genet 134, 472–483.
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B. E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol 9, R137.
Zhang, Y., Sun, Y., Wu, Z., Xiong, X., Zhang, J., Ma, J., Xiao, S., Huang, L., and Yang, B. (2021). Subcutaneous and intramuscular fat transcriptomes show large differences in network organization and associations with adipose traits in pigs. Sci China Life Sci 64, 1732–1746.
Zhao, R., and Duncan, S.A. (2005). Embryonic development of the liver. Hepatology 41, 956–967.
Zhao, Y., Hou, Y., Xu, Y., Luan, Y., Zhou, H., Qi, X., Hu, M., Wang, D., Wang, Z., Fu, Y., et al. (2021). A compendium and comparative epigenomics analysis of *cis*-regulatory elements in the pig genome. Nat Commun 12, 2217.
Zheng, M., Xiao, S., Guo, T., Rao, L., Li, L., Zhang, Z., and Huang, L. (2020). DNA methylomic homogeneity and heterogeneity in muscles and testes throughout Pig Adulthood. Aging 12, 25412–25431.
Zong, X., Xiao, X., Jie, F., Cheng, Y., Jin, M., Yin, Y., and Wang, Y. (2021). YTHDF1 promotes NLRP3 translation to induce intestinal epithelial cell inflammatory injury during endotoxic shock. Sci China Life Sci 64, 1988–1991.
Żylicz, J.J., Bousard, A., Žumer, K., Dossin, F., Mohammad, E., da Rocha, S.T., Schwalb, B., Syx, L., Dingli, F., Loew, D., et al. (2019). The implication of early chromatin changes in X chromosome inactivation. Cell 176, 182–197.e23.

SUPPORTING INFORMATION

The supporting information is available online at https://doi.org/10.1007/s11427-021-2034-5. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.