

RESEARCH

Open Access



Genome-wide identification of functional enhancers and their potential roles in pig breeding

Yinqiao Wu^{1,2†}, Yuedong Zhang^{1,2,3,4†}, Hang Liu^{1,2}, Yun Gao¹, Yuyan Liu^{3,4}, Ling Chen^{3,4}, Lu Liu^{1,5}, David M. Irwin⁶, Chunhui Hou^{1,7}, Zhongyin Zhou^{1*} and Yaping Zhang^{1,8*}

Abstract

Background: The pig is an economically important livestock species and is a widely applied large animal model in medical research. Enhancers are critical regulatory elements that have fundamental functions in evolution, development and disease. Genome-wide quantification of functional enhancers in the pig is needed.

Results: We performed self-transcribing active regulatory region sequencing (STARR-seq) in the porcine kidney epithelial PK15 and testicular ST cell lines, and reliably identified 2576 functional enhancers. Most of these enhancers were located in repetitive sequences and were enriched within silent and lowly expressed genes. Enhancers poorly overlapped with chromatin accessibility regions and were highly enriched in chromatin with the repressive histone modification H3K9me3, which is different from predicted pig enhancers detected using ChIP-seq for H3K27ac or/and H3K4me1 modified histones. This suggests that most pig enhancers identified with STARR-seq are endogenously repressed at the chromatin level and may function during cell type-specific development or at specific developmental stages. Additionally, the *PPP3CA* gene is associated with the loin muscle area trait and the *QKI* gene is associated with alkaline phosphatase activity that may be regulated by distal functional enhancers.

Conclusions: In summary, we generated the first functional enhancer map in PK15 and ST cells for the pig genome and highlight its potential roles in pig breeding.

Keywords: Breeding, Functional enhancers, Pig, STARR-seq

Background

A fundamental goal in modern biology is to identify and characterize non-coding regulatory elements that control gene expression in development and disease [1, 2]. Enhancers, as genomic non-coding sequences, are critical regulatory elements that control spatial and temporal

gene expression [3]. Variation at enhancers is associated with complex traits and diseases [4, 5]. Previous studies found that enhancers have crucial roles in sexual development in mammals and that duplication or deletion of some of these enhancers could result in sex reversal [6, 7]. ZRS is a well-characterized long-range enhancer directing limb-specific sonic hedgehog (*Shh*) expression. Kvon et al. [8] employed a combination of comparative genomics and genetic engineering to introduce snake-specific deletions into the orthologous genomic region of the ZRS enhancer in the mouse that resulted in severe limb reduction. Thus, it is important to identify enhancers and reveal their biological functions.

[†]Yinqiao Wu and Yuedong Zhang contributed equally to this work.

* Correspondence: zhouzhongyin@mail.kiz.ac.cn; zhangyp@mail.kiz.ac.cn

¹State Key Laboratory of Genetic Resources and Evolution, and Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, Yunnan, China

⁸Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, Yunnan, China

Full list of author information is available at the end of the article



© The Author(s). 2022 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

The domestic pig (*Sus scrofa*) is an economically important livestock species and is widely used as a large animal model for preclinical studies due to its human-like size and physiology [9, 10]. A number of studies performed in the pig, based on sequence homology information [11] and chromatin marks, including DNA accessibility, histone modification and transcription factor or cofactor binding, have been conducted to characterize regulatory elements such as enhancers [12–14]. However, these currently used predictive methods do not provide direct functional evidence or quantify the enhancer activity. More recently, the development of the self-transcribing active regulatory region sequencing (STARR-seq) strategy has allowed direct measurement of enhancer activity across the whole genome [15]. This method was first applied in *Drosophila* [15, 16] and later in human cells [17–19] and other species [20–23]. However, to date, no genome-wide studies have been performed to detect functional enhancers in the pig genome.

Here, we performed STARR-seq in the porcine kidney epithelial PK15 and testicular ST cell lines to generate the first whole genome-wide enhancer activity quantification map for the pig. We identified 2576 functional enhancers, most of which are located in repetitive sequences and enriched within silent and lowly expressed genes. We compared our map to ATAC-seq and ChIP-seq data and found that our enhancers poorly overlap with accessible chromatin regions and were highly enriched in chromatin with the repressive histone modification H3K9me3. Moreover, mapping of enhancers to complex traits in the pig found that the *PPP3CA* gene, which is associated with the loin muscle area trait, and the *QKI* gene, which is associated with alkaline phosphatase activity, might be regulated by distal enhancers. In summary, we provide the first functional enhancer map for the pig that should promote the identification of causal mutations for complex traits in the pig.

Methods

Cell culture

Pig (*Sus scrofa*) PK15 cells were maintained in DMEM media (Gibco, Thermo Fisher Scientific, Shanghai, China) and ST cells were maintained in MEM media (Gibco, Thermo Fisher Scientific, Shanghai, China) supplied with 10% FBS (Gibco, Thermo Fisher Scientific, Shanghai, China) and 1% penicillin/streptomycin (Thermo Fisher Scientific, Shanghai, China). All cells were tested for mycoplasma using Myco-Blue Mycoplasma Detector (Vazyme, Nanjing, China).

Construction of STARR-seq input plasmid libraries

STARR-seq input plasmid libraries were generated as described in a previous study [15]. Briefly, genomic

DNA was extracted from an ear tissue sample from one Diannan small-ear pig. After sonication with a Scientz-II D ultrasonic generator (NingBo Scientz Biotechnology, Ningbo, China), 500–800 bp sheared DNA was selected by cutting from a gel and was recovered using E.Z.N.A.® Gel Extraction Kit (Omega Bio-Tek, Norcross, GA, USA). About 6 µg of DNA fragments were end repaired, 5'-phosphorylated, 3'dA-tailed and ligated with Illumina adaptor using the VAHTS Mate Pair Library Prep Kit for Illumina® (Vazyme, Nanjing, China) following the manufacturer's protocol. Ligated DNA was amplified with TransStart FastPfu Fly DNA Polymerase (Transgen, Beijing, China) with homology arms primers (Forward primer: 5'-ACA CTC TTT CCC TAC ACG ACG-3' and Reverse primer: 5'-GAC TGG AGT TCA GAC GTG TGC-3'). The vector backbone was modified from pGL3-basic (Promega, Beijing, China). The Super Core Promoter 1 (SCP1) promoter was inserted upstream of the luciferase site. The luciferase sequence was replaced with a GFP sequence containing a synthetic intron and homology arms. The homology arms were used to insert the genomic DNA fragments. The vector backbone was linearized using PCR (95 °C for 5 min; then 25 cycles of 95 °C for 20 s, 57 °C for 20 s and 72 °C for 4 min; Forward primer: 5'-CGT CGT GTA GGG AAA GAG TGT-3' and Reverse primer: 5'-GCA CAC GTC TGA ACT CCA GTC-3'). Further, circular plasmids were removed using DpnI (New England BioLabs, Ipswich, MA, USA) at 37 °C for 30 min. Linearized vector was separated using a 1% agarose gel and purified with E.Z.N.A.® Gel Extraction Kit. We then recombined the genomic DNA fragments into the linearized vector using ClonExpress II One Step Cloning Kit (Vazyme, Nanjing, China). The 10 µL of the recombined plasmids transformed into Trans1-T1 Phage Resistant Chemically Competent Cell (Transgen, Beijing, China). A total of 140 transformation reactions were pooled together in 4 L LB medium with 10 µg/mL Ampicillin, which was grown to an Optical Density (OD) of 0.8. Plasmids were purified using E.Z.N.A.® Endo-Free Plasmid Maxi Kit (Omega Bio-Tek, Norcross, GA, USA).

Transfection of STARR-seq input plasmid libraries into cells

The STARR-seq input plasmid library was transfected into pig PK15 and ST cells using TurboFect™ Transfection Reagent (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol. The inhibitors BX-795 (MedChemExpress, Shanghai, China) and C16 (Sigma Aldrich, Shanghai, China) were used to suppress the expression of immune-related genes in transfected PK15 and ST cells. For PK15 cells, 1 µmol/L of both inhibitors was added to the culture medium, while for ST cells, 0.5 µmol/L of both inhibitors was used. We

generated two biological replicates of the PK15 and ST cells.

Construction of cDNA and input plasmid libraries for Illumina sequencing

Total RNA was extracted from the transfected PK15 and ST cells using the TransZol™ Up Plus RNA Kit (Transgen, Beijing, China). The poly(A)⁺ RNA fraction was enriched using VAHTS mRNA Capture Beads (Transgen, Beijing, China). Genomic DNA was digested with 5 U DNase I (New England BioLabs, Ipswich, MA, USA) at 37 °C for 20 min. First strand cDNA was synthesized using TransScript One-Step gDNA Removal and cDNA synthesis SuperMix kit (Transgen, Beijing, China) at 50 °C for 30 min and 85 °C for 5 s with the library specific primer (5'-GAC TGG AGT TCA GAC GTG TGC-3'). We used 10–30 ng cDNA as template in a 50-μL PCR reaction with TransStart FastPfu Fly DNA Polymerase with the reporter gene specific primers (Forward primer: 5'-AAC AAG AAT TGG GAC AAC TCC AGT GAA-3' and Reverse primer: 5'-GAC TGG AGT TCA GAC GTG TGC-3'). PCR products were purified by GeneJET PCR Purification Kit (Thermo Fisher Scientific, Waltham, MA, USA). Purified PCR products were used as templates for the second round PCR (98 °C for 2 min; followed by 6–10 cycles of 98 °C for 30 s, 65 °C for 30 s, 72 °C for 30 s; and then 72 °C for 5 min) with TransStart FastPfu Fly DNA Polymerase and the index primers in the VAHTS Multiplex Oligos set1/set2 for Illumina Kit (Vazyme, Nanjing, China). Second round PCR products were then purified using a GeneJET PCR Purification Kit.

After capture of poly(A)⁺ RNA, the remaining aqueous solution was treated with 10 μL RNase A (Transgen, Beijing, China) at 37 °C for 60 min to remove any RNA in the solution. Input plasmid templates were purified with a GeneJET PCR Purification Kit. Purified plasmid DNA was amplified with the TransStart FastPfu Fly DNA Polymerase and the index primers from the VAHTS Multiplex Oligos set1/set2 for Illumina.

Both the cDNA and input plasmid libraries were sequenced on the Illumina HiSeq X Ten platform at BerryGenomics (Beijing, China).

Reporter assay

To verify the enhancers identified in this study, we used the input plasmid library backbone and inserted candidate regions. Plasmids were co-transfected with luciferase control vector (pGL3-promoter (Promega, Beijing, China)) into PK15 cells with TurboFect™ Transfection Reagent following the manufacturer's protocol. Primers of candidate regions are reported in Additional file 1: Table S1.

RT-qPCR

RNAs were extracted using TransZol™ Up Plus RNA Kit. Contaminating genomic DNA was digested with 5 U DNase I and cDNAs was then synthesized using the HiScript III 1st Strand cDNA Synthesis Kit (Vazyme, Nanjing, China). RT-qPCR used ChamQ Universal SYBR qPCR Master Mix (Vazyme, Nanjing, China) with 95 °C for 5 min, followed by 40 cycles of 95 °C for 15 s, 60 °C for 30 s with the melting curve program. Primers used for RT-qPCR are described in Additional file 2: Table S2.

Computational processing of STARR-seq data

To remove low-quality reads and adaptor sequences, the STARR-seq data was processed using Trimmomatic v0.39 with following parameters: TruSeq2-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36 [24]. Reads were then aligned to the pig reference genome (Sscrofa11.1) using Bowtie2 v2.3.5.1 with parameter "--no-discordant -X 2000" [25]. Mapped reads were filtered using SAMtools v1.3.1 with parameter "view -bS -f 2 -q 5" to obtain uniquely mapped reads [26]. PCR duplicates were removed using the Picard toolkit [27]. Reproducibility of the two independent biological replicates in PK15 and ST cells were evaluated using Pearson correlation coefficients, which were calculated using multiBamSummary and plotCorrelation in deepTools v3.5.1 [28]. Enhancers were identified using BasicSTARRseq as described in previous studies [15, 21] with a strength > 1.0, *P* value < 0.001 and FDR < 0.1.

RNA-seq library construction and analysis

RNA was extracted from PK15 cells using the TransZol™ Up Plus RNA Kit. Libraries construction and sequencing were performed at Novogene.

Trimmomatic v0.39 [24] was used to filter low-quality reads with default parameter. Reads were aligned to the pig genome (Sscrofa11.1) using HISAT2 v2.2.1 [29]. The expression values of each gene was calculated using StringTie v2.1.4 [30].

Potential biological function analysis

To obtain potential biological functions of the enhancers identified in this study, each enhancer was assigned to a putative gene based on the closest genomic distance. Gene Ontology (GO) enrichment analysis of the putative genes was performed using DAVID [31]. In addition, we identified potential transcription factor binding sites in the putative enhancers using the MEME-ChIP web server [32].

ATAC-seq library construction and analysis

ATAC-seq library preparation was performed as previously described [33]. PK15 cell nuclei (~100,000) were

extracted using lysis buffer (10 mmol/L Tris-HCl (pH 7.4), 10 mmol/L NaCl, 3 mmol/L MgCl₂, 0.1% IGEPAL CA-630) and incubated with Tn5 transposase (Vazyme, Nanjing, China). Next, the transposed DNA was amplified through 14 PCR cycles with indexed primers according to the manufacturer's protocol for the TruePrep Index Kit V4 for Illumina (Vazyme, Nanjing, China). We sequenced two independent libraries using the HiSeq X ten platform (BerryGenomics, Beijing, China).

Trimmomatic 0.39 [24] with default parameter was used to filter low-quality reads and adaptor sequences. Reads were then aligned to the pig reference genome (Sscrofa11.1) using BWA-MEM v0.7.17 [34]. Mapped reads were processed with SAMtools v1.3.1 [26] to keep uniquely mapped reads. PCR duplicates were removed using the Picard toolkit [27]. Effective reads were generated by removing mitochondrial reads using BEDTools v2.25.0 [35]. Peaks were called using MACS v2.1.0 (setting: -f BAMPE -q 0.001 --shift -75 --extsize 150 --nomodel -B --SPMR -g 2.5e9) [36].

ChIP-seq library construction and analysis

ChIP-Seq libraries were prepared following the ENCODE guidelines [37]. ChIP-seq data was mapped to the pig genome (Sscrofa11.1) using Bowtie2 v2.3.5.1 [25]. MACS v2.1.0 [36] was used to identify peaks (default settings for CTCF, H3K9me₃; broad peaks model for H3K4me₁, H3K4me₃, H3K27ac and H3K27me₃) with *P* values less than 0.001.

Hi-C analysis

We identified TADs and Hi-C contacts from the Hi-C data, which was download from the SRA database (accessions number: PRJEB40576 [38] and PRJNA482496 [39]). We then aligned the reads to the pig genome (Sscrofa11.1) using BWA-MEM v0.7.17 [34] and applied HiCEXplorer v3.4.2 [40] to build Hi-C contact matrixes with 40 Kb resolution. We identified TADs using HiCEXplorer v3.4.2 [40].

Results

Quantifying genome-wide enhancer activity in the pig using STARR-seq

To comprehensively identify putative enhancers with activity in the pig genome, we constructed libraries of randomly fragmented genomic DNA from one Diannan small-ear pig. Furthermore, we used porcine PK15 and ST cells to perform STARR-seq to quantify the activity of the enhancers. Since transfection of most mammalian cells with a plasmid DNA causes a striking increase in the expression of immune-related genes, through the activation of their enhancers, leading to false positive signals with STARR-seq methods [18], thus, we first assessed the immunoreaction of PK15 and ST cells after

plasmid transfection. As expected, we found that immune-related genes were highly expressed after transfection with plasmids (Fig. 1A and Additional file 3: Figs. S1 and S2). Therefore, we treated these cell lines with the TBK1/IKKe inhibitor BX-79521 and the PKR inhibitor C1622 during plasmid transfection to decrease expression of immune system-related genes (Fig. 1A and Additional file 3: Figs. S1 and S2). We found that 1 μmol/L of both BX-7951 and C1662 inhibitors in PK15 cells and 0.5 μmol/L of both BX-7951 and C1662 in ST cells reduced the expression of immune system-related genes (Fig. 1A and Additional file 3: Figs. S1 and S2).

The STARR-seq libraries generated from these two cell lines contained between 43.7 and 46.0 million unique fragments in the input plasmid libraries and from 11.9 to 21.0 million unique fragments in their cDNA libraries (Additional file 4: Table S3). The median fragment length of the input plasmid and cDNA libraries were about 600 bp (Additional file 3: Fig. S3). cDNA and input plasmid libraries covered between 78.2% and 92.5% of the non-repetitive sequence of the pig genome (Additional file 3: Fig. S4). The GC-content for each library showed that the sequencing libraries were unbiased (Additional file 3: Fig. S5). Correlation analysis of the two biological replicates of the plasmid and cDNA libraries showed high correlation for both PK15 and ST cells (Additional file 3: Fig. S6A and B). From the above, these results indicated that STARR-seq could be used to quantify pig enhancers activity at a genome-wide scale.

We calculated the enrichment of cDNA reads normalized by input plasmid for 600 bp bins across the whole genome. Potential enhancers were identified as described [15, 21] using BasicSTARRseq with a binomial test *P* < 0.001, enrichment > 1.0, FDR < 0.1. Between 1015 to 2217 enhancers were detected in the pig genome in the PK15 and ST replicates (Fig. 1B and Additional file 5: Table S4). Moreover, we found that 601 of the enhancers (> 50% reciprocal overlap) were shared between the two cell lines (Additional file 5: Table S4). An enhancer located on chromosome 10 is shown in Fig. 1C as an example. The identified enhancers showed a wide range of strengths (Additional file 3: Fig. S6C and D). The Pearson correlation coefficient of the strengths for the two replicates was 0.620 and 0.707 in the PK15 and ST cells, respectively (Fig. 1D and Additional file 3: Fig. S7), demonstrating that the STARR-seq libraries were reproducible in the pig genome. Combining the enhancers found in the PK15 and ST cells resulted in the identification of 2576 non-redundant enhancers in the pig genome (Fig. 1B and Additional file 5: Table S4), which display low overlap with the enhancers identified by ChIP-seq for the H3K27ac or/and H3K4me₁ chromatin modifications (Additional file 3:

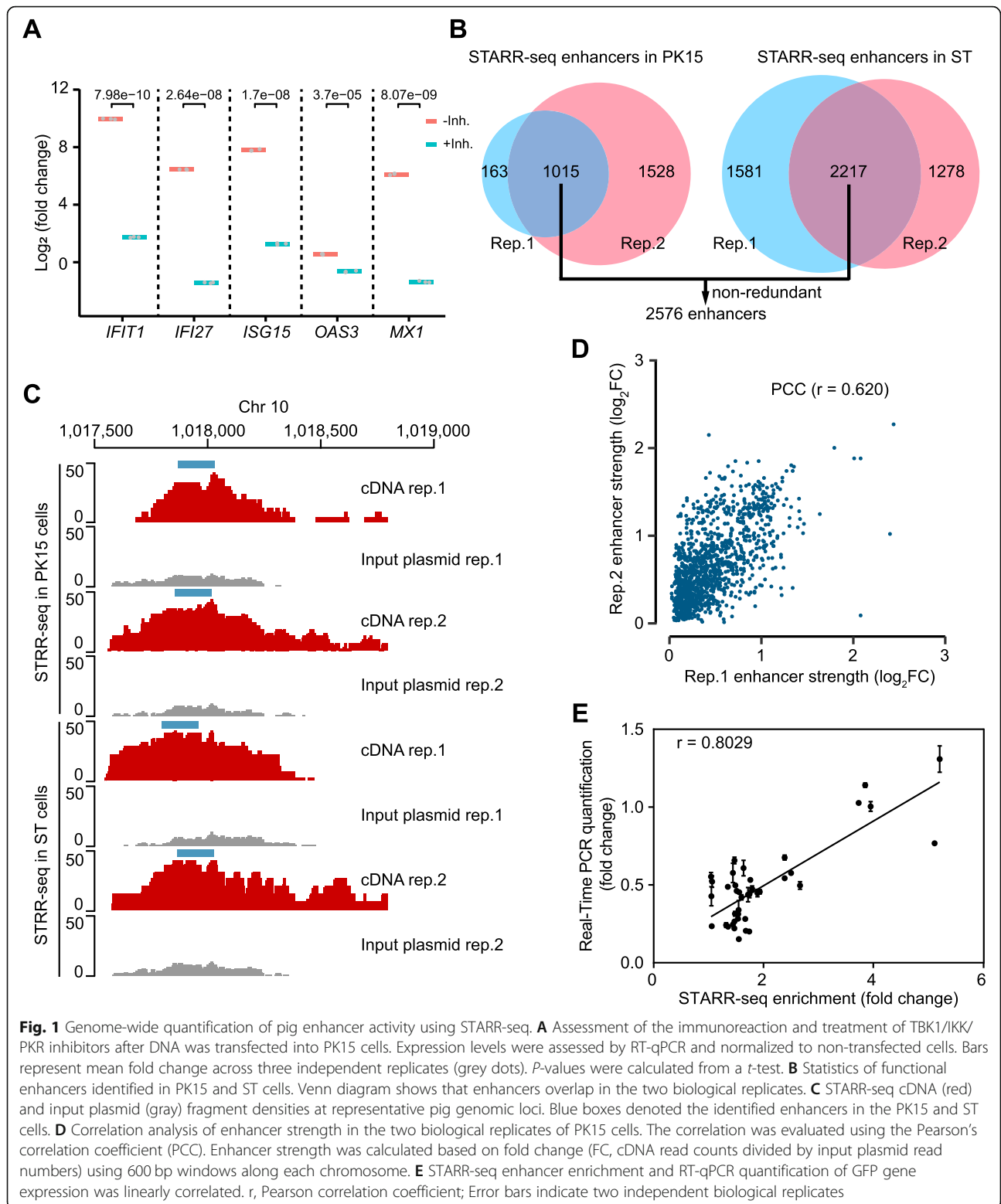


Fig. S8) [12, 14]. To verify the strength of the enhancers identified using STARR-seq, 40 enhancers were selected (Additional file 1: Table S1), with a wide range of strengths, and measured using RT-

qPCR. STARR-seq enhancer strength and RT-qPCR of reporter gene expression levels were highly correlated ($r = 0.8029$, $***P < 0.0001$, Fig. 1E). These results demonstrate that the enhancers identified in this

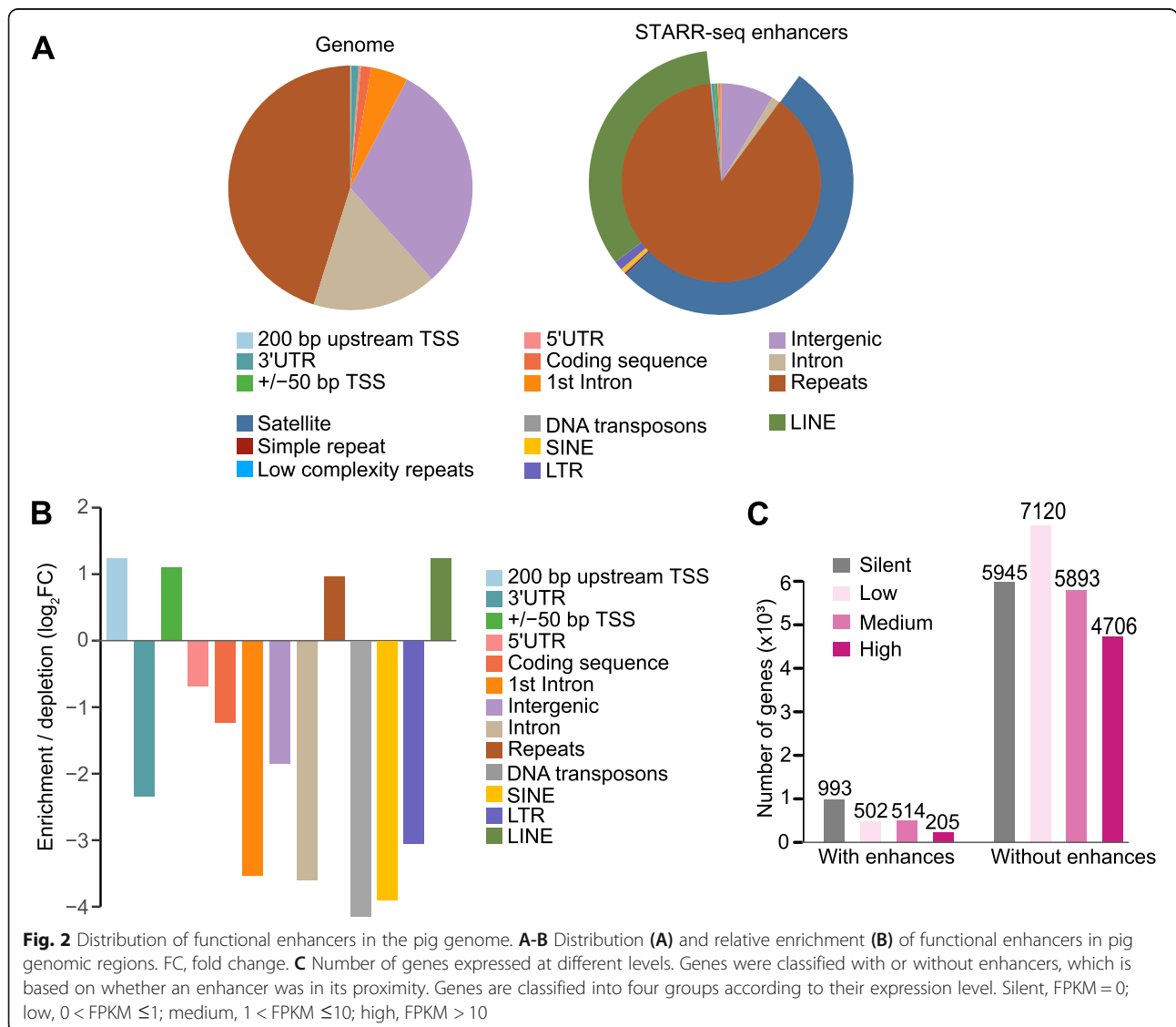
study are reliable and that pig enhancers identified using histone modifications need to be verified using methods such as STARR-seq.

Pig enhancers are enriched in repetitive sequences

To reveal the distribution pattern of enhancers in the pig genome, we annotated and calculated the percentage of enhancers in different functional genomic regions. The majority (2272/2576 = 88.19%) of the enhancers were located in repetitive sequences (Sscrofa11.1), and secondly in intergenic regions (222/2576 = 8.61%) (Fig. 2A and Additional file 5: Table S4). Previous studies indicated that transposable elements (TEs) are a widespread class of repetitive sequences in genomes [41, 42]. Thus, we analyzed the percentage of enhancers located in TE regions and found a high proportion of the enhancers were located in TEs (907/2576 = 35.20%), including short interspersed

nuclear elements (SINEs) (29/2576 = 1.13%), long interspersed nuclear elements (LINEs) (863/2576 = 33.50%), long terminal repeats (LTRs) (11/2576 = 0.43%) and DNA transposons (4/2576 = 0.15%) (Fig. 2A). There is also a relative enrichment of the enhancers in LINEs compared to other TE classes, which was not due to LINEs abundance in the pig genome (Fig. 2B). Our observations are in line with previous reports of STARR-seq in human ESCs (Embryonic Stem Cells) [19], human HeLa-S3 cells [18], mouse ESCs [23] and rice [21]. These results indicate that certain families of TEs contribute to enhancer function [19].

Additionally, enhancers were overrepresented in the proximity of transcription start sites (TSSs) (Fig. 2B), emphasizing the importance of these regions for transcriptional regulation. This observation is consistent with the enhancers identified in *Drosophila* [15]. We



also found that the pig enhancers were underrepresented in 1st intron, 5'UTR and 3'UTR (Fig. 2B), which is strikingly different with previous studies [15, 17, 21], although this could be a species-specific difference in enhancer distribution.

Enhancers are enriched with silent and lowly expressed genes

To investigate whether enhancers are preferentially enriched in active or silent genes, we separated genes into four groups based on gene expression levels (silent: FPKM = 0; low: $0 < \text{FPKM} \leq 1$; medium: $1 < \text{FPKM} \leq 10$; and high: FPKM > 10) detected by RNA-seq (Additional file 4: Table S3) and assigned each enhancer to a putative gene based on closest genomic distance. A total of 2214 genes were assigned as targets for these enhancers (Fig. 2C). Enhancers were preferentially enriched in silent and lowly expressed genes (Fig. 2C), in line with a previous report in rice [21] and indicating that STARR-seq enhancers are not necessarily enriched within actively transcribed genes in vivo [15, 21].

The potential biological function of enhancers

Assigning an enhancer to its target gene is challenging. For genes that were located close to enhancers, we further examined if these genes were enriched with specific biological functions. GO analysis for the closest genes of enhancers found that they were strikingly enriched in the cell projection morphogenesis, gene transcription and positive regulation of developmental biological processes (Additional file 3: Fig. S9). Furthermore, the enhancers were enriched in motifs of transcription factor (TF) binding sites, including Myod1, Tcfap2c and Zbtb3 (Additional file 6: Table S5). Myod1 is required for myogenin activation and affects the terminal differentiation of muscle cells [43], which indicated that the enhancers may participate in muscle development in the pig.

Pig enhancers poorly overlap with accessible chromatin regions

To examine the enhancers with endogenous chromatin accessibility, we performed ATAC-seq in the PK15 cells (Additional file 4: Table S3) and found that 19.41% (500/2576) of the enhancers were accessible (Additional file 3: Fig. S10A and Additional file 7: Table S6). The low chromatin accessibility of the STARR-seq enhancers is also reported in previous studies [17, 21, 23]. Additionally, the enrichment of accessible chromatin regions was poorly correlated with STARR-seq enhancer strength (Additional file 3: Fig. S10B), indicating that pig enhancer strength cannot easily be predicted from chromatin accessibility features. From the above, this indicates that enhancers predicted based on chromatin accessibility may not be reliable in the pig genome.

Enhancers correlate with active and repressive chromatin states

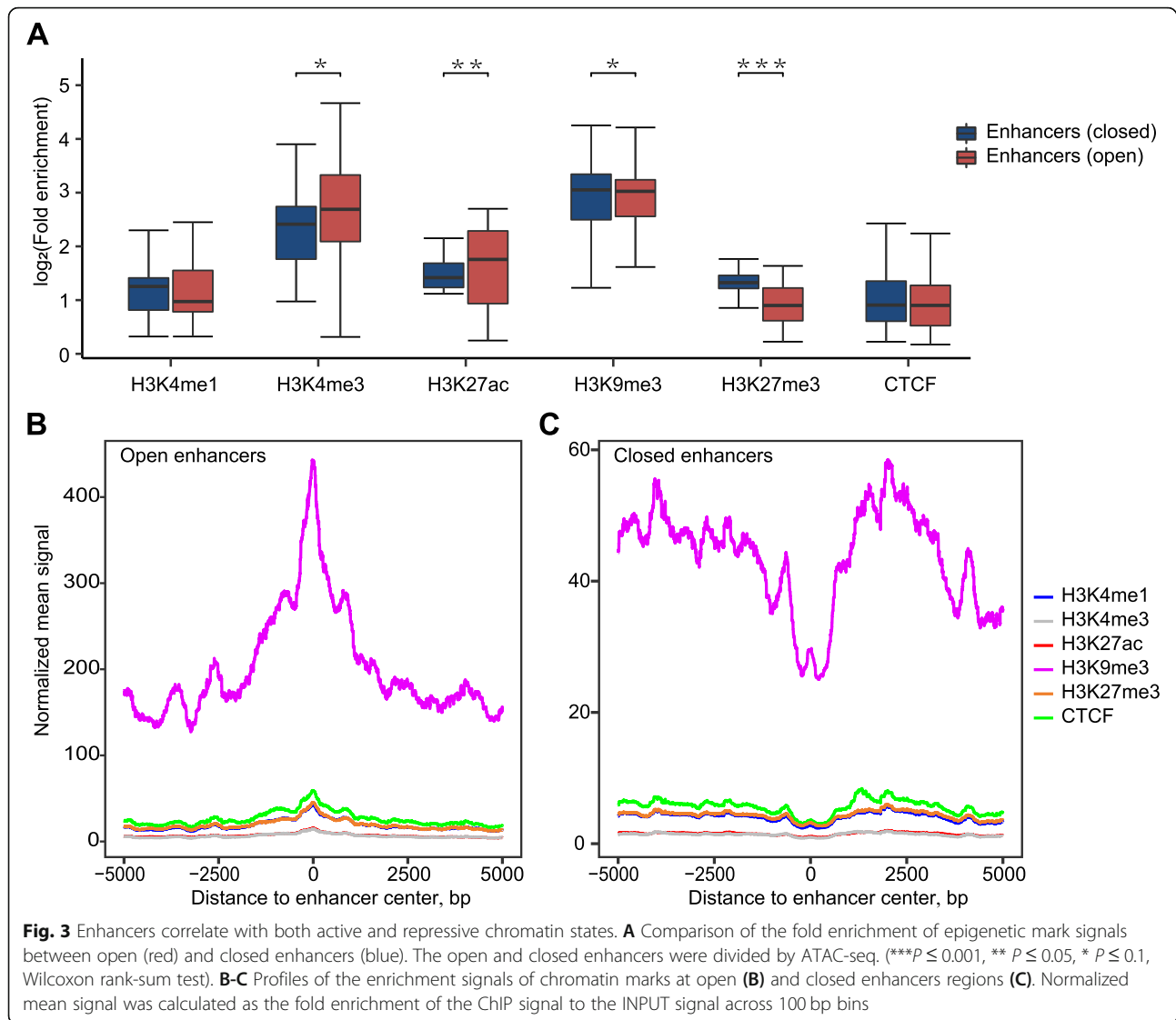
We divided the pig enhancers into open enhancers, which overlap with ATAC-seq accessible chromatin regions, and closed enhancers, which did not locate in accessible chromatin regions, as defined in previous studies (Additional file 3: Fig. S10A and Additional file 7: Table S6) [15, 17, 21]. To compare the epigenetic marks between the open and closed enhancers, we performed ChIP-seq in PK15 cells (Additional file 4: Table S3) and integrated an analysis between the enhancer and ChIP-seq data (Fig. 3A-C and Additional file 7: Table S6).

Both the open and closed enhancers correlated with active and repressive chromatin states (Fig. 3A-C and Additional file 7: Table S6). Moreover, the open enhancers showed a higher fold enrichment compared with closed enhancers in active histone modification H3K4me3 ($*P = 0.076$, Wilcoxon rank-sum test) and H3K27ac ($**P = 0.047$, Wilcoxon rank-sum test) (Fig. 3A). In contrast, closed enhancers showed a significantly higher fold enrichment compared with open enhancers in repressive chromatin marks H3K9me3 ($*P = 0.056$, Wilcoxon rank-sum test) and H3K27me3 ($***P = 3.8e-14$, Wilcoxon rank-sum test) (Fig. 3A). Both closed and open enhancers displayed similar enrichment in H3K4me1 and insulator protein CTCF (Fig. 3A).

We found a subset of enhancers ($133/2576 = 5.16\%$) that were enriched in active modifications and not in repressive modifications (Additional file 3: Fig. S11A and Additional file 7: Table S6). In addition, we also found a subset of enhancers ($743/2576 = 28.84\%$) that were contrarily enriched in repressive modifications and not active modifications (Additional file 3: Fig. S11B and Additional file 7: Table S6). Interesting, a subset of enhancers ($520/2576 = 20.19\%$) that correlated with active chromatin modifications was also enriched with repressive chromatin modifications (Additional file 3: Fig. S11C and Additional file 7: Table S6). These results indicate that enhancers correlated with both active and repressive chromatin modifications (Additional file 3: Fig. S11A-C). The apparently conflicting combination (Additional file 3: Fig. S11C) was also reported in rice [21] and it is assumed that enhancers were modified differentially in different subpopulations of the cells. Further studies are needed to test this hypothesis in different cell subpopulations.

Most pig enhancers are endogenously repressed at the chromatin level

Additional, 43.83% (1129/2576) of the enhancers were enriched with the repressive histone signal H3K9me3 (Additional file 7: Table S6). Profiles along ± 5 -kb regions flanking the enhancers exhibited significantly higher



signals for repressive modifications such as H3K9me3, CTCF and H3K27me3 than for active histone modifications including H3K4me1, H3K4me3 and H3K27ac in both open and closed enhancers (Fig. 3B and C). This result is consistent with majority of pig enhancers being located in repetitive sequences ($2272/2576 = 88.19\%$) (Fig. 2A and Additional file 5: Table S4) and with repetitive sequences subjected to repressive epigenetic modifications [44–46]. From the above, we assume that most pig enhancers are endogenously repressed at the chromatin level and may function during cell type-specific development or at different developmental stages.

Map enhancers to pig complex traits

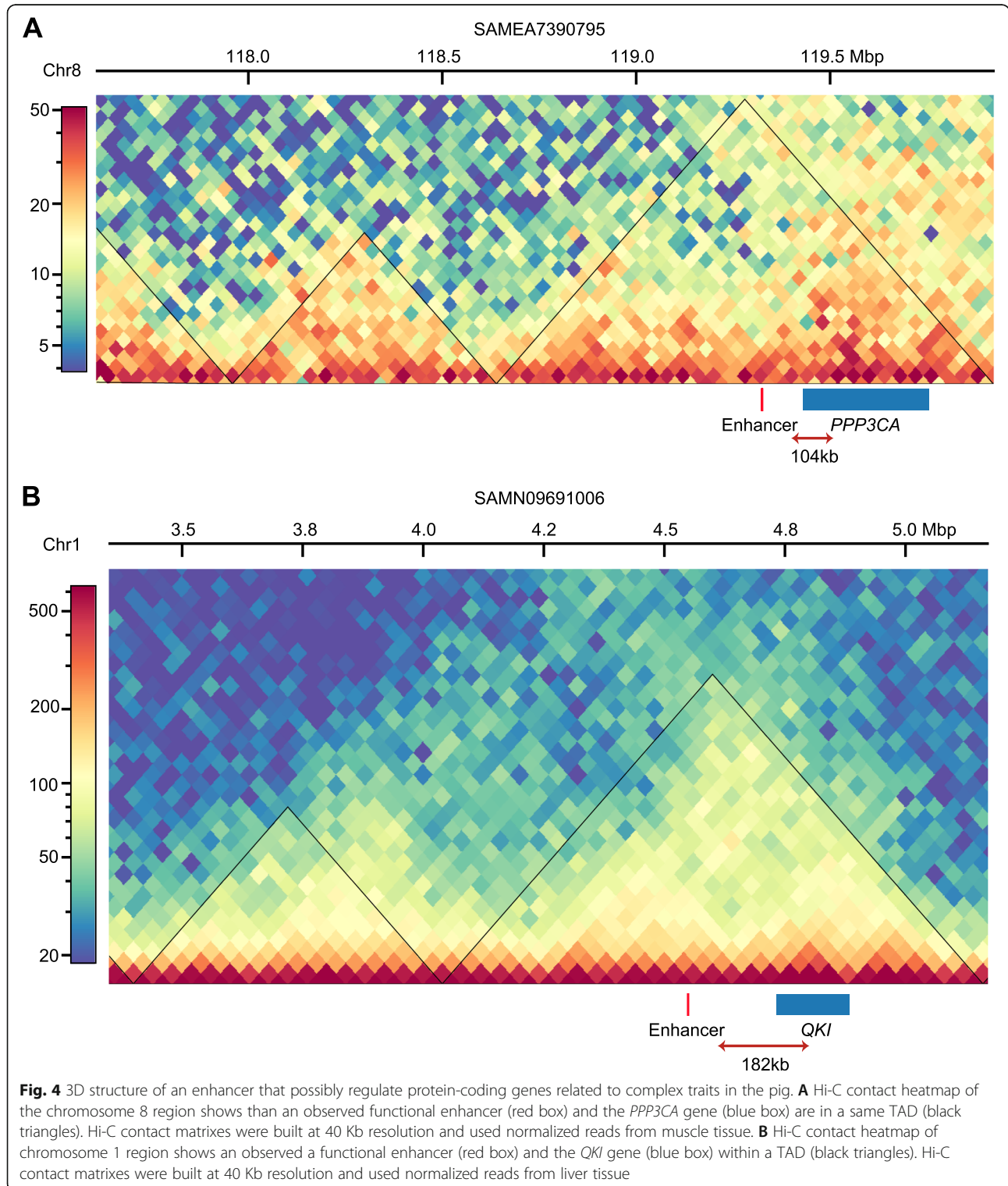
Quantitative trait locus (QTL) and genome wide association study (GWAS) mapping for important economical traits have been widely applied in pigs [47–49]. Due to a large proportion of causal mutations being located in

non-coding regions and a lack of annotation for non-coding regulatory elements in the pig, it is difficult to identify causal mutations for QTL and GWAS regions. In this study, we first annotated 2576 functional enhancers in the pig genome (Additional file 5: Table S4). We then integrated Hi-C data [38, 39] and QTL datasets [47] to identify enhancers that have biological associations with complex traits in the pig.

We deleted QTLs regions [47] that are longer than half a chromosome. And we analyzed filtered QTL regions that overlap with functional enhancer regions and found that 1508 non-redundant QTL regions associate with 440 traits enriched with enhancers (Additional file 8: Table S7). Among the identified regions, Lee et al. [50] reported a 1 Mb QTL region on *Sus scrofa* chromosome 8 that may impact the loin muscle area trait and contained the candidate gene *PPP3CA*. The QTL region overlaps with a functional pig enhancer (Sscrofa11.1, 8:

119,324,324-119,325,095) (Additional file 8: Table S7). Previous studies reported that *PPP3CA* (also called calcineurin A α) is actively involved in regulating the muscle fiber phenotype in pig [51, 52]. By reanalyzing the Hi-C data performed in pig fetal muscle tissues at 90 and 110

days of gestation [38], we found that the enhancer and *PPP3CA* gene were encompassed by a topologically associating domain (TAD) and linked by significant Hi-C contact, which indicates that the *PPP3CA* gene is associated with the loin muscle area trait in the pig and is



potentially regulated by a distal functional enhancer (Sscrofa11.1, 8:119,324,324-119,325,095) (Fig. 4A and Additional file 3: Fig. S12). Further experimental investigations are required to validate the regulatory relationship between *PPP3CA* and the functional enhancer in pig muscle tissue.

Additionally, Bovo et al. [53] reported a 0.42 Mb QTL region on *Sus scrofa* chromosome 1 that showed effects on alkaline phosphatase (ALP) levels in the pig. The QTL region contains a functional enhancer (Sscrofa11.1, 1:4,548,557-4,549,774), which is 182 Kb upstream of the KH domain containing RNA binding (*QKI*) gene (Fig. 4B). A previous study showed that *QKI* regulates the expression of ALP and functions as critical regulator for colon epithelial differentiation [54]. Moreover, ALP isoenzymes are derived from the bones and liver, which are of particular interest for the evaluation of disease state [53, 55]. By reanalyzing the Hi-C data of female pig adult and fetus liver tissue [39], we found that the enhancer and *QKI* gene are connected by chromatin interaction within a TAD (Fig. 4B and Additional file 3: Fig. S13). From the above, we assume that a distal functional enhancer (Sscrofa11.1, 1:4,548,557-4,549,774) regulates the expression level of *QKI* and thus, influences ALP activity levels. The influence of the causal mutations in the enhancer region for ALP activity needs further functional validation.

Discussion

This study is the first genome-wide identification of functional enhancers in the pig. We identified 2576 functional enhancers in the pig genome and revealed several different features of these functional enhancers compared with other species (Figs. 2 and 3 and Additional file 5: Table S4). Specifically, the number of pig functional enhancers detected in this study was less than seen in the human genome [17]. This might be due to our highly stringent selection criteria that included removing PCR replicates and a *q*-value cutoff to identify functional enhancers with greater confidence. Moreover, applying STARR-seq to the pig genome is still very challenging, due to the large size and complexity of this genome. Thus, new methods are needed for future studies on the detection of functional enhancers from highly complex mammalian genomes.

Previous studies used sequence homology information [11] and chromatin features, including DNA accessibility and active histone modifications, to identify potential enhancers [12–14]. Enhancer function in cell type-specific and developmental-stage-specific stage [56–58]. For this reason, methods based on DNA sequences or chromatin states can lead to many false positive predictions of enhancers that might not play functional roles in cells [15, 59]. Furthermore, these methods can only identify

putative enhancers based on chromatin features or sequences and therefore do not exhaustively detect functional enhancers genome-wide. Thus, enhancers identified by chromatin features or sequences should be confirmed using approaches such as STARR-seq or by a RT-qPCR method. In our study, we found that most functional enhancers poorly overlap with chromatin accessible regions (Additional file 3: Fig. S10A) and are highly enriched with the repressive histone modification H3K9me3 (Fig. 3B and C). These results were consistent with previous studies showing that STARR-seq can identify some closed chromatin functional enhancers [15, 17, 23] and those functional enhancers can be silent or active, depending on the cell type and the developmental stage. As such, identification and characterization of pig functional enhancers under various cell types, including embryonic stem cells and primary cell cultures, in the future will be important for understanding the genetic basis of development. Moreover, a previous study assessed the effects of cis and trans for regulatory elements in human and mouse [60]. Future studies are needed to analyze and compare the cis and trans effects of regulatory elements such as enhancers and promoters between pigs and other mammalian species.

Previous studies have indicated that strong selection signals and regions associated with complex traits in the pig genome are located in non-coding regions [9, 61–63]. Here, we analyzed the functional enhancers that have the potential to regulate complex traits by integrating Hi-C and QTL data. This analysis identified two examples of enhancers that likely have distal regulatory function: (1) one enhancer may regulate the expression of the *PPP3CA* gene and thus, influence the loin muscle area trait, and (2) a second enhancer may regulate the *QKI* gene associated with ALP activity levels. *PPP3CA* is regarded as a candidate gene that plays an important role in muscle fiber differentiation and affects meat quality of livestock [64]. The *QKI* gene has been identified as the culprit gene for a patient with intellectual disabilities and has important roles in broader biological systems, such as cardiovascular development, bone metabolism and cancer progression [65, 66]. Additional studies are needed to refine the causal mutations in the enhancers and clarify the role of *PPP3CA* and *QKI* genes in muscle development and physiological processes in the pigs, respectively. The comprehensively identified pig functional enhancers found in this study provide insights into the functional complexity of enhancers in the pig.

Conclusions

In all, we performed STARR-seq in pig PK15 and ST cells and identified 2576 functional enhancers in the pig genome. These enhancers poorly overlap with chromatin accessible regions and are highly enriched with the

repressive histone modification H3K9me3. By integrating functional enhancers with pig complex traits, we found that the *PPP3CA* gene is associated with the loin muscle area trait and the *QKI* gene is associated with ALP activity levels, with both genes regulated by distal functional enhancers. Our first map of functional enhancers in the pig genome provide an important resource for enhancer studies and supply the new regulatory elements for pig breeding and construction of human disease models.

Abbreviations

ALP: Alkaline phosphatase; ESCs: Embryonic stem cells; GO: Gene Ontology; GWAS: Genome-wide association study; Hi-C: High-throughput chromosome conformation capture; LINES: Long interspersed nuclear elements; LTRs: Long terminal repeats; OD: Optical density; QTL: Quantitative trait locus; SCP1: Super core promoter 1; Shh: Sonic hedgehog; SINES: Short interspersed nuclear elements; STARR-seq: Self-transcribing active regulatory region sequencing; TAD: Topologically associating domain; TEs: Transposable elements; TF: Transcription factor; TSSs: Transcription start sites.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40104-022-00726-y>.

Additional file 1: Table S1. Enhancers verified using the RT-qPCR method. Primers used to construct the reporter vectors.

Additional file 2: Table S2. Primers used for RT-qPCR.

Additional file 3: Fig. S1. DNA transfection induced immunoreaction and its suppression with treatment with immune inhibitors in ST cells. **Fig. S2.** Assessment of the immunoreaction (DMSO) and treatment effect of immune inhibitors after DNA is transfected into cells. **Fig. S3.** Distribution of fragment sizes in the STARR-seq libraries. **Fig. S4.** Genome coverage of STARR-seq libraries in pig non-repetitive regions. **Fig. S5.** GC-content analysis for the STARR-seq libraries. **Fig. S6.** Reproducibility of STARR-seq. **Fig. S7.** Correlation analysis of enhancers strength in two biological replicates of ST cells. **Fig. S8.** Venn diagram showing the overlap of enhancers between our STARR-seq approach and other published studies. **Fig. S9.** GO analysis of genes in proximity to enhancers. **Fig. S10.** ATAC-seq and STARR-seq enhancers integrated analysis. **Fig. S11.** Snapshots of signals at three types of chromatin state for the STARR-seq enhancer regions. **Fig. S12.** The enhancer (Sscrofa11.1, 8:119,324,324-119,325,095) and *PPP3CA* gene interacted by a TAD region in pig muscle tissues. **Fig. S13.** The enhancer (Sscrofa11.1, 1:4,548,557-4,549,774) and *QKI* gene interacted by a TAD region in pig liver tissues.

Additional file 4: Table S3. Sequencing statistics of STARR-seq, RNA-seq, ATAC-seq and ChIP-seq libraries.

Additional file 5: Table S4. Characterization of a total of 2576 functional enhancers in the pig genome.

Additional file 6: Table S5. Potential transcription factor binding sites in the enhancers.

Additional file 7: Table S6. Epigenetic mark enrichment analysis of the pig enhancers.

Additional file 8: Table S7. QTL regions overlapping functional enhancers.

Acknowledgements

This work was supported by the Animal Branch of the Germplasm Bank of Wild Species, Chinese Academy of Sciences (the Large Research Infrastructure Funding), the Domestic Pig Molecular Breeding and Translational Medicine Research Center in Southwest China.

Authors' contributions

YPZ and ZYZ initiated the project. ZYZ designed the study. YQW performed data analysis and interpretation. HL participated in data analysis. YDZ, YYL and LC performed experiments. YQW, YDZ, HL, DMI, YG, ZYZ, CH and YPZ wrote and revised the manuscript. All authors read and approved the final version of the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (32100502), the Ministry of Agriculture of China (2016ZX08009003-006), Science & Technology Department of Yunnan Province (202102AE090039).

Availability of data and materials

The sequencing datasets generated in this study were submitted to the Genome Sequence Archive (GSA) with ID CRA006465.

Declarations

Ethics approval and consent to participate

Experiments were approved by the Institutional Animal Care and Use Committee at the Kunming Institute of Zoology, Chinese Academy of Sciences (approval ID No.: SMKX-2017023).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests

Author details

¹State Key Laboratory of Genetic Resources and Evolution, and Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, Yunnan, China. ²Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650204, Yunnan, China. ³State Key Laboratory for Conservation and Utilization of Bio-resource in Yunnan, Yunnan University, Kunming 650091, Yunnan, China. ⁴School of Life Science, Yunnan University, Kunming 650091, Yunnan, China. ⁵Institute of Physical Science and Information Technology, Anhui University, Hefei 230601, Anhui, China. ⁶Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada. ⁷Department of Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen 518055, China. ⁸Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, Yunnan, China.

Received: 14 December 2021 Accepted: 4 May 2022

Published online: 04 July 2022

References

- Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis C, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57–74. <https://doi.org/10.1038/nature11247>.
- Fulco CP, Munschauer M, Anyoha R, Munson G, Grossman SR, Perez EM, et al. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science*. 2016;354(6313):769–73. <https://doi.org/10.1126/science.aag2445>.
- Serfling E, Jasin M, Schaffner W. Enhancers and eukaryotic gene transcription. *Trends Genet*. 1985;1:224–30. [https://doi.org/10.1016/0168-9525\(85\)90088-5](https://doi.org/10.1016/0168-9525(85)90088-5).
- Pennacchio LA, Bickmore WA, Dean A, Nobrega MA, Bejerano G. Enhancers: five essential questions. *Nat Rev Genet*. 2013;14(4):288–95. <https://doi.org/10.1038/nrg3458>.
- Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, et al. Genome-wide enhancer maps link risk variants to disease genes. *Nature*. 2021;593(7858):238–43. <https://doi.org/10.1038/s41586-021-03446-x>.
- Gonen N, Futtner CR, Wood S, Garcia-Moreno SA, Salamone IM, Samson SC, et al. Sex reversal following deletion of a single distal enhancer of Sox9. *Science*. 2018;360(6396):1469–73. <https://doi.org/10.1126/science.aas9408>.
- Croft B, Ohnesorg T, Hewitt J, Bowles J, Quinn A, Tan J, et al. Human sex reversal is caused by duplication or deletion of core enhancers upstream of

- SOX9. *Nat Commun.* 2018;9(1):5319. <https://doi.org/10.1038/s41467-018-07784-9>.
8. Kwon EZ, Kamneva OK, Melo US, Barozzi I, Osterwalder M, Mannin BJ, et al. Progressive loss of function in a limb enhancer during Snake evolution. *Cell.* 2016;167(3):633–42 e11. <https://doi.org/10.1016/j.cell.2016.09.028>.
 9. Rubin C, Megens H, Barrio AM, Maqbool K, Sayyab S, Schwochow D, et al. Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci U S A.* 2012;109(48):19529–36. <https://doi.org/10.1073/pnas.1217149109>.
 10. Yan S, Tu Z, Liu Z, Fan N, Yang H, Yang S, et al. A huntingtin Knockin pig model recapitulates features of selective neurodegeneration in Huntington's disease. *Cell.* 2018;173(4):989–1002. <https://doi.org/10.1016/j.cell.2018.03.005>.
 11. Sun H, Liao Y, Wang Z, Zhang Z, Oyelami FO, Olasege BS, et al. ETph: enhancers and their targets in pig and human database. *Anim Genet.* 2020; 51(2):311–3. <https://doi.org/10.1111/age.12893>.
 12. Kern C, Wang Y, Xu X, Pan Z, Halstead M, Chanthavixay G, et al. Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. *Nat Commun.* 2021;12(1):1821. <https://doi.org/10.1038/s41467-021-22100-8>.
 13. Zhao Y, Hou Y, Xu Y, Luan Y, Zhou H, Qi X, et al. A compendium and comparative epigenomics analysis of cis-regulatory elements in the pig genome. *Nat Commun.* 2021;12(1):2217. <https://doi.org/10.1038/s41467-021-22448-x>.
 14. Pan Z, Yao Y, Yin H, Cai Z, Wang Y, Bai L, et al. Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. *Nat Commun.* 2021;12(1):5848. <https://doi.org/10.1038/s41467-021-26153-7>.
 15. Arnold CD, Gerlach D, Stelzer C, Boryn tM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science.* 2013; 339(6123):1074–7. <https://doi.org/10.1126/science.1232542>.
 16. Arnold CD, Gerlach D, Spies D, Matts JA, Sytnikova YA, Pagani M, et al. Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during cis-regulatory evolution. *Nat Genet.* 2014;46(7):685–92. <https://doi.org/10.1038/ng.3009>.
 17. Liu Y, Yu S, Dhiman VK, Brunetti T, Eckart H, White KP. Functional assessment of human enhancer activities using whole-genome STARR-sequencing. *Genome Biol.* 2017;18(1):219. <https://doi.org/10.1186/s13059-017-1345-5>.
 18. Muerdter F, Boryn LM, Woodfin AR, Neumayr C, Rath M, Zabidi MA, et al. Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat Methods.* 2018;15(2):141–9. <https://doi.org/10.1038/nmeth.4534>.
 19. Barakat TS, Halbritter F, Zhang M, Rendeiro AF, Perenthaler E, Bock C, et al. Functional dissection of the enhancer repertoire in human embryonic stem cells. *Cell Stem Cell.* 2018;23(2):276–88 e8. <https://doi.org/10.1016/j.stem.2018.06.014>.
 20. Vanhille L, Griffon A, Maqbool MA, Zacariascebeza J, Dao LTM, Fernandez NF, et al. High-throughput and quantitative assessment of enhancer activity in mammals by CapStar-seq. *Nat Commun.* 2015;6(1):6905. <https://doi.org/10.1038/ncomms7905>.
 21. Sun J, He N, Niu L, Huang Y, Shen W, Zhang Y, et al. Global quantitative mapping of enhancers in Rice by STARR-seq. *Genomics Proteomics Bioinformatics.* 2019;17(2):140–53. <https://doi.org/10.1016/j.gpb.2018.11.003>.
 22. Jores T, Tonniez J, Dorriy MW, Cuperus JT, Fields S, Queitsch C. Identification of plant enhancers and their constituent elements by STARR-seq in tobacco leaves. *Plant Cell.* 2020;32(7):2120–31. <https://doi.org/10.1105/tpc.20.00155>.
 23. Peng T, Zhai Y, Atlasi Y, Ter Huurne M, Marks H, Stunnenberg HG, et al. STARR-seq identifies active, chromatin-masked, and dormant enhancers in pluripotent mouse embryonic stem cells. *Genome Biol.* 2020;21(1):243. <https://doi.org/10.1186/s13059-020-02156-3>.
 24. Bolger A, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
 25. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923>.
 26. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16): 2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
 27. "Picard Toolkit." Broad Institute, GitHub Repository. 2019. <https://broadinstitute.github.io/picard/>.
 28. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 2016;44(W1):W160–5. <https://doi.org/10.1093/nar/gkw257>.
 29. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12(4):357–60. <https://doi.org/10.1038/nmeth.3317>.
 30. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015;33(3):290–5. <https://doi.org/10.1038/nbt.3122>.
 31. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1): 44–57. <https://doi.org/10.1038/nprot.2008.211>.
 32. Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics.* 2011;27(12):1696–7. <https://doi.org/10.1093/bioinformatics/btr189>.
 33. Buenostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods.* 2013;10(12):1213–8. <https://doi.org/10.1038/nmeth.2688>.
 34. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics.* 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
 35. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033>.
 36. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9(9):1–9. <https://doi.org/10.1186/gb-2008-9-9-r137>.
 37. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 2012;22(9):1813–31. <https://doi.org/10.1101/gr.136184.111>.
 38. Marti-Marimon M, Vialaneix N, Lahbib-Mansais Y, Zytnicki M, Camut S, Robelin D, et al. Major reorganization of chromosome conformation during muscle development in pig. *Front Genet.* 2021;12:748239. <https://doi.org/10.3389/fgene.2021.748239>.
 39. Tian X, Li R, Fu W, Li Y, Wang X, Li M, et al. Building a sequence map of the pig pan-genome from multiple de novo assemblies and hi-C data. *Sci China Life Sci.* 2020;63(5):750–63. <https://doi.org/10.1007/s11427-019-9551-7>.
 40. Wolff J, Rabbani L, Gilsbach R, Richard G, Manke T, Backofen R, et al. Galaxy HiExplorer 3: a web server for reproducible hi-C, capture hi-C and single-cell hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* 2020;48(W1):W177–W84. <https://doi.org/10.1093/nar/gkaa220>.
 41. Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. *Nature.* 1980;284(5757):601–3. <https://doi.org/10.1038/284601a0>.
 42. Orgel LE, Crick FH. Selfish DNA: the ultimate parasite. *Nature.* 1980; 284(5757):604–7. <https://doi.org/10.1038/284604a0>.
 43. Gerber AN, Klesert TR, Bergstrom DA, Tapscott SJ. Two domains of MyoD mediate transcriptional activation of genes in repressive chromatin: a mechanism for lineage determination in myogenesis. *Genes Dev.* 1997; 11(4):436–50. <https://doi.org/10.1101/gad.11.4.436>.
 44. Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, et al. Role of transposable elements in heterochromatin and epigenetic control. *Nature.* 2004;430(6998):471–6. <https://doi.org/10.1038/nature02651>.
 45. Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, et al. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* 2005;24(4):800–12. <https://doi.org/10.1038/sj.emboj.7600545>.
 46. Darby MM, Sabuncyan S. Repetitive elements and epigenetic marks in behavior and psychiatric disease. *Adv Genet.* 2014;86:185–252. <https://doi.org/10.1016/B978-0-12-800222-3.00009-7>.
 47. Hu Z-L, Park CA, Reecy JM. Building a livestock genetic and genomic information knowledgebase through integrative developments of animal QTLdb and CorrDB. *Nucleic Acids Res.* 2019;47(D1):D701–D10. <https://doi.org/10.1093/nar/gky1084>.
 48. Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat Genet.* 2015;47:217–25. <https://doi.org/10.1038/ng.3199>.

49. Moon S, Kim TH, Lee KT, Kwak W, Lee T, Lee SW, et al. A genome-wide scan for signatures of directional selection in domesticated pigs. *BMC Genomics*. 2015;16:130. <https://doi.org/10.1186/s12864-015-1330-x>.
50. Lee J, Kim Y, Cho E, Cho K, Sa S, Kim Y, et al. Genomic analysis using Bayesian methods under different genotyping platforms in Korean Duroc pigs. *Animals (Basel)*. 2020;10(5):752. <https://doi.org/10.3390/ani10050752>.
51. Depreux FFS, Scheffler JM, Grant AL, Bidwell CA, Gerrard DE. Molecular cloning and characterization of porcine calcineurin-alpha subunit expression in skeletal muscle. *J Anim Sci*. 2010;88(2):562–71. <https://doi.org/10.2527/jas.2009-1832>.
52. Wang H, Zhu Z, Wang H, Yang S, Mo D, Li K. Characterization of different expression patterns of calstarcin-1 and calstarcin-2 in porcine muscle. *Gene*. 2006;374:104–11. <https://doi.org/10.1016/j.gene.2006.01.035>.
53. Bovo S, Ballan M, Schiavo G, Gallo M, Dall'Olio S, Fontanesi L. Haplotype-based genome-wide association studies reveal new loci for haematological and clinical-biochemical parameters in large White pigs. *Anim Genet*. 2020; 51(4):601–6. <https://doi.org/10.1111/age.12959>.
54. Yang G, Fu H, Zhang J, Lu X, Yu F, Jin L, et al. RNA-binding protein quaking, a critical regulator of Colon epithelial differentiation and a suppressor of Colon Cancer. *Gastroenterology*. 2010;138(1):231–40. <https://doi.org/10.1053/j.gastro.2009.08.001>.
55. Fernandez NJ, Kidney BA. Alkaline phosphatase: beyond the liver. *Vet Clin Pathol*. 2007;36(3):223–33. <https://doi.org/10.1111/j.1939-165X.2007.tb00216.x>.
56. Tuan DY, Solomon WB, London IM, Lee DP. An erythroid-specific, developmental-stage-independent enhancer far upstream of the human "beta-like globin" genes. *Proc Natl Acad Sci U S A*. 1989;86(8):2554–8. <https://doi.org/10.1073/pnas.86.8.2554>.
57. Ellmeier W, Sunshine MJ, Losos K, Littman DR. Multiple developmental stage-specific enhancers regulate CD8 expression in developing thymocytes and in thymus-independent T cells. *Immunity*. 1998;9(4):485–96. [https://doi.org/10.1016/s1074-7613\(00\)80632-9](https://doi.org/10.1016/s1074-7613(00)80632-9).
58. Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol*. 2015;16(3):144–54. <https://doi.org/10.1038/nrm3949>.
59. Muerdter F, Boryn LM, Arnold CD. STARR-seq - principles and applications. *Genomics*. 2015;106(3):145–50. <https://doi.org/10.1016/j.jygeno.2015.06.001>.
60. Mattioli K, Oliveros W, Gerhardinger C, Andergassen D, Maass PG, Rinn JL, et al. Cis and trans effects differentially contribute to the evolution of promoters and enhancers. *Genome Biol*. 2020;21(1):210. <https://doi.org/10.1186/s13059-020-02110-3>.
61. Zhou ZY, Li AM, Adeola AC, Liu Y-H, Irwin DM, Xie H-B, et al. Genome-wide identification of long intergenic noncoding RNA genes and their potential association with domestication in pigs. *Genome Biol Evol*. 2014;6(6):1387–92. <https://doi.org/10.1093/gbe/evu113>.
62. Zhou ZY, Li AM, Wang LG, Irwin DM, Liu YH, Xu D, et al. DNA methylation signatures of long intergenic noncoding RNAs in porcine adipose and muscle tissues. *Sci Rep*. 2015;5(1):15435. <https://doi.org/10.1038/srep15435>.
63. Wu YQ, Zhao H, Li YJ, Khederzadeh S, Wei HJ, Zhou ZY, et al. Genome-wide identification of imprinted genes in pigs and their different imprinting status compared with other mammals. *Zool Res*. 2020;41(6):721–5. <https://doi.org/10.24272/j.issn.2095-8137.2020.072>.
64. Davoli R, Luise D, Mingazzini V, Zambonelli P, Braglia S, Serra A, et al. Genome-wide study on intramuscular fat in Italian large White pig breed using the PorcineSNP60 BeadChip. *J Anim Breed Genet*. 2016;133(4):277–82. <https://doi.org/10.1111/jbg.12189>.
65. Darbelli L, Richard S. Emerging functions of the quaking RNA-binding proteins and link to human diseases. *Wiley Interdiscip Rev RNA*. 2016;7(3): 399–412. <https://doi.org/10.1002/wrna.1344>.
66. Aberg K, Saetre P, Jareborg N, Jazin E. Human QKI, a potential regulator of mRNA expression of human oligodendrocyte-related genes involved in schizophrenia. *Proc Natl Acad Sci U S A*. 2006;103(19):7482–7. <https://doi.org/10.1073/pnas.0601213103>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

