Research article

The landscape of super-enhancer regulates remote target gene transcription through loop domains in adipose tissue of pig

Lin Yu [a,1], Tengda Huang [a,1], Siqi Liu [a], Jingsu Yu [a], Menglong Hou [a], Songtao Su [a], Tianyu Jiang [a], Xiangling Li [a], Yixing Li [b], Turtushikh Damba [c], Lei Zhou [a,*], Yunxiao Liang [a,**]

[a] Institute of Digestive Disease, Guangxi Academy of Medical Sciences, The People's Hospital of Guangxi Zhuang Autonomous Region, Nanning, China
[b] College of Animal Science and Technology, Guangxi University, Nanning, China
[c] School of Pharmacy, Mongolian National University of Medical Sciences, Ulan Bator, Mongolia

ARTICLE INFO

ABSTRACT

Background: A super-enhancer (SE) is a huge cluster of multiple enhancers that control the key genes for cell identity and function. The rise of advanced chromatin immunoprecipitation sequencing (ChIP-seq) technology such as Cleavage Under Targets and Tagmentation (CUT&Tag) allows more SEs to be discovered. However, SE studies in Luchuan and Duroc pigs are very rare in animal husbandry.
Results: We used the CUT&Tag technique to identify 145 and 378 SEs from the adipose tissues of Luchuan and Duroc pigs, respectively. There were significant differences in the peak coverage ratio of SE peaks in the gene promoter region between the two breeds. Not only that, peak signals at the start and end point of the SE peak profile showed obvious spikes. The proximal target genes of SE were highly expressed compared with the background genes and the typical enhancer target genes. Subsequently, in conjoint analysis with high-throughput chromosome conformation capture sequencing (Hi-C seq) data, we predicted the remote regulatory genes of SE and found that their expression level was related to the distance of SE extended to the loop's anchor, but not the length of loops. According to our prediction model, SEs can maintain promoter accessibility of partial remote target genes through loop domains. Finally, a batch of SEs closely related to fat metabolism traits were obtained by performing a coalition analysis of quantitative trait loci and SE data.
Conclusions: This work enabled us to obtain hundreds of SEs from Luchuan and Duroc pigs. Our model provides a new method for predicting the SE remote target genes based on loop domains, and to further explore the potential role of super-enhancer in the regulation of fat metabolism.

* Corresponding author. Guangxi Academy of Medical Sciences, the People's Hospital of Guangxi Zhuang Autonomous Region, Nanning, 530021, China.
** Corresponding author. Guangxi Academy of Medical Sciences, the People's Hospital of Guangxi Zhuang Autonomous Region, Nanning, 530021, China.
E-mail addresses: zhoulei@gxu.edu.cn (L. Zhou), doctor_liang1977@sina.com (Y. Liang).
[1] These authors contributed equally to this work.

1. Introduction

A super-enhancer (SE) is a cluster of transcriptionally active enhancers with enhancer-associated chromatin characteristics [1,2]. Super-enhancers (SEs) usually contain a large number of master transcription factors, cofactor and enhancer histone modification marks [3,4]. Identification of an SE depends on chromatin immunoprecipitation sequencing (ChIP-seq) data of H3K27ac or Med1 and other proteins, which is essentially "stitching" and filtration of the obtained typical enhancers. Therefore, SE discovery is strongly associated with the development of ChIP-seq technology. Among the multiple advanced ChIP-seq technologies, Cleavage Under Targets and Tagmentation (CUT&Tag) technology has a high average mapping rate [5]. It uses Tn5 transposase as a tool for DNA fragmentation, and has the advantages of a short preparation time, low background noise, strong specificity and high sensitivity [6]. Therefore, the CUT&Tag technique was thought to be the best choice for the identification of SEs in the current study.

Increasing research on SEs has been carried out in the field of oncology. For example, Blinka et al. found that a distal SE differentially regulates several neighboring genes. Furthermore, the enhancer RNAs produced by SE selectively activate the expression of one neighboring gene [7]. It has also been demonstrated that cell differentiation and growth require SEs, which are more sensitive to perturbation than typical enhancers [4,8,9].

However, published research investigating SEs in pigs is very scarce. Currently, only Peng et al. have identified SEs and broad H3K4me3 domains in various tissues from six pigs by analyzing public ChIP-seq data, and discussed their conserved regions between human and mouse SEs [10]. On the other hand, Luan et al. identified and analyzed the conserved *cis*-regulatory elements in pig liver [11]. Zhou identified 2025 non-redundant SEs by analyzing ChIP-seq data of H3K27ac in the pituitary of Bama pig and Large White pig, and found that 302 SEs were associated with biological processes of pituitary function [12].

Luchuan pig is a typical obese pig breed, which is produced in Guangxi Province, China [13]. Duroc is a classic lean pig breed, with marked differences in backfat thickness and fat deposition compared with Luchuan pig [14]. Both breeds are good models for investigating the regulatory mechanism of fat metabolism [15].

Fat deposition is an important economic factor in pig farming. Fat deposition in pigs can seriously affect reproductive performance and productivity. Changes in meat quality can also affect consumers' choices of pork [16,17]. However, to date, there is little analysis of SEs in lean and obese pig breed's fat deposition. Therefore, in order to identify SEs in these two pig breeds, we used the CUT&Tag method to analyze ChIP-seq data of H3K27ac in the adipose tissue of pigs. In addition, we integrated Hi-C loops, ATAC-seq and RNA-seq data to analyze the unique driving patterns of proximal and remote target genes of SE. Importantly, through collaborative analysis of quantitative trait loci (QTL) data, we obtained a batch of SEs associated with fat metabolism. These results will allow better understanding of the mechanism of SE regulation of fat metabolism and fat deposition genes in the adipose tissue of these two pig breeds, which will expand our horizon of epigenetics during fat metabolism.

2. Materials and methods

2.1. Animals and sample collection

Three 1-month-old Duroc boars were purchased from the Guangxi State Farms Yongxin Animal Husbandry Group Co. Ltd., Nanning, China, and three 1-month-old Luchuan boars were purchased from the Animal Husbandry Research Institute of Guangxi Zhuang Autonomous Region, Nanning, China. All pigs were fed a basic diet, containing 11.98 MkJ/kg energy, 16% crude protein, 6% crude fiber, 7% crude ash, 0.6–1.2% calcium, 0.4–1.0% total phosphorus, 0.2–0.8% NaCl, and 0.8% lysine. All pigs had free access to food and water, and were fed under identical conditions for 6 months. When the pigs were slaughtered, fresh backfat was isolated corresponding to the fourth rib. The fat tissue was immediately placed in liquid nitrogen and temporarily frozen, and then transferred to an ultra-low temperature refrigerator for storage at −80 °C.

2.2. CUT&Tag assay and library preparation

Six adipose samples and a ChIP-grade rabbit polyclonal anti-H3K27Ac antibody (Sigma, Cat#07-360-S) were used for DNA library construction. The CUT&Tag assay protocol was modified by referring to Hatice [6]. Briefly, native nuclei were purified from six frozen adipose samples [18]. Nuclei samples were gently washed twice with wash buffer (20 mM HEPES pH 7.5; 150 mM NaCl; 0.5 mM Spermidine; 1 × Protease inhibitor cocktail). Then, 10 μL Concanavalin A coated magnetic beads (Bangs Laboratories) were added to each sample and incubated at room temperature for 10 min. The unbound supernatant was removed and bead-bound cells were resuspended in Dig wash buffer (20 mM HEPES pH 7.5; 150 mM NaCl; 0.5 mM Spermidine; 1 × Protease inhibitor cocktail; 0.05% Digitonin; 2 mM EDTA) and a 1:50 dilution of ChIP grade rabbit polyclonal anti-H3K27Ac antibody (Sigma, Cat#07-360-S) or normal rabbit IgG antibody (Millipore, Cat#12–370) and incubated on a rotating platform overnight at 4 °C. The primary antibody was removed using a magnet stand. Goat monoclonal anti-rabbit IgG antibody (Millipore, Cat# AP132) was diluted 1:100 in Dig wash buffer and samples were incubated at room temperature for 60 min. The samples were washed using the magnet stand 2–3 times in Dig wash buffer. A 1:100 dilution of pA-Tn5 adapter complex was prepared in Dig-med buffer (0.01% Digitonin; 20 mM HEPES pH 7.5; 300 mM NaCl; 0.5 mM Spermidine; 1 × Protease inhibitor cocktail) and incubated with the samples at room temperature for 1 h. The samples were washed 2–3 times for 5 min each time in 1 mL Dig-med buffer, then resuspended in tagmentation buffer (10 mM MgCl2 in Dig-med Buffer) and incubated at 37 °C for 1 h. DNA was purified using phenol-chloroform-isoamyl alcohol extraction and ethanol precipitation. To amplify DNA libraries, 35.5 μL DNA was mixed with 0.5 μL of Phusion® DNA polymerase (ThermoFisher, Shanghai, China), 10 μL of 5 × Phusion® HF buffer, 1 μL dNTPs, 1.5 μL of universal PCR primer and 1.5 μL of index primer. The mixture was
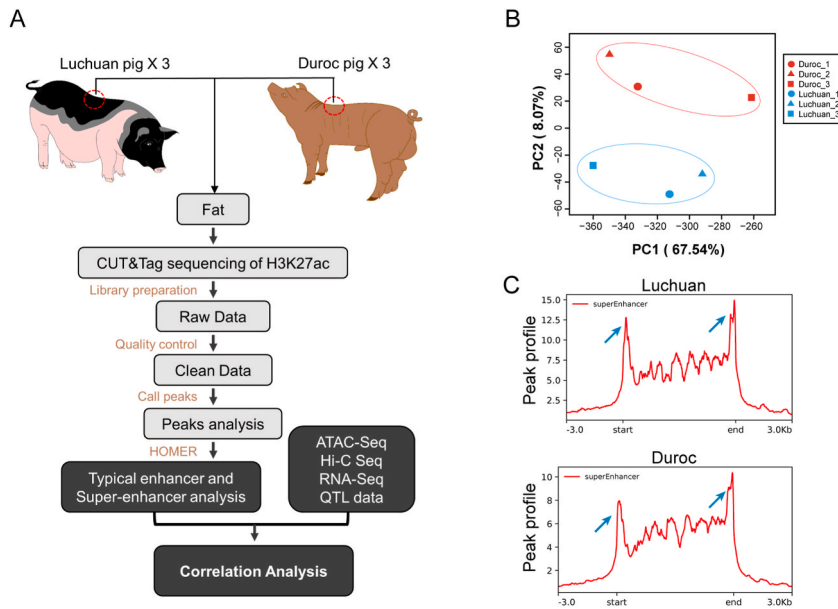
Fig. 1. Overview of SE data and peak characteristics. **A:** The work flow chart of SEs identification. **B:** The PCA analysis of all samples. **C:** The SE peak profiles of Luchuan and Duroc pigs. All of the SE's length was homogenized in the horizontal axis. The "start" and "end" represent SE's start point and end point, respectively.

placed in a Thermocycler with a heated lid using the following cycling conditions: 98 °C for 30 s, 98 °C for 10 s, 60 °C for 20 s and 72 °C for 30 s, 9 cycles of step 2 to 4. Then extension at 72 °C for 5 min, final hold at 4 °C for 2 min. Library clean-up was performed using SPRI Magnetic Beads (Beckman Coulter, Beverly, MA, USA). The libraries were stored at −20 °C. The concentration was assessed using Qubit. The library (2 μL) was diluted 10-fold with 10 mM Tris-HCl or 0.1 × TE and the library quality was assessed on an Agilent 2100 Bioanalyzer® (Agilent Technologies, Inc., California, USA) high sensitivity chip. The result was confirmed to show a narrow length distribution of approximately 300–600 bp.

2.3. Sequencing and quality control

DNA sequencing was performed in Illumina Novaseq 6000 using a 150 bp paired-end following the manufacturer's instructions. Raw data (raw reads) were processed by an in-house Perl script. Briefly, reads containing adapter, ploy-N and low-quality reads were removed. Then, the Q20, Q30 and GC content of clean data were calculated.

2.4. Reads mapping to the reference genome and call peak

Clean data were aligned to the reference genome (*Sus scrofa 10.2*) using the BWA-MEME program [19]. MACS2 software was used to perform the call peak operation on the data and q-value was set at < 0.05.

2.5. Peak annotation and analysis

Peaks were annotated using HOMER's annotatePeaks.pl script [20]. The results of the annotations were counted and the distribution results were plotted using R. The bedtools were used for sample peak file merging, calculation and other analysis [21].

2.6. Identification of super-enhancer and motif analysis

According to Young Lab's definition of SE [8], they were obtained using the default parameters of ROSE software. Using DESeq2 software [22] for the calculation of different SEs, the threshold was |log2FC| > 1 and p value < 0.05. The HOMER's findMotifsGenome. pl tool [20] was used for motif analysis.

2.7. Target gene prediction

Proximal target gene prediction of SEs was calculated by the ROSE_geneMapper.py script of ROSE software using default parameters. Remote target gene prediction of SEs was calculated by loci comparison utilizing in-house Python scripts written by ourselves. In brief, a preliminary screening was performed for loops with p < 10^{-6} and the genes located near the loop's anchor were
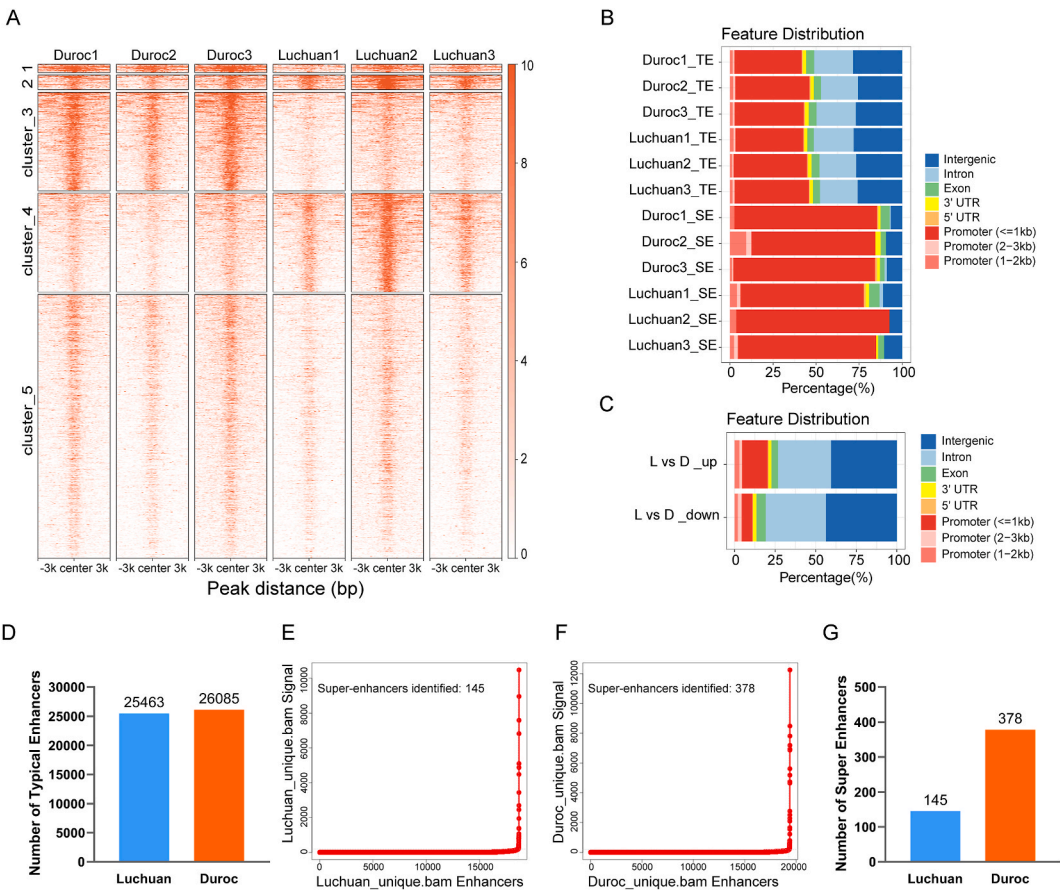
Fig. 2. Differential peak analysis and SEs identification. **A:** The cluster heat maps of the SE central signal. The horizontal axis represents the normalized coordinates of the peak region, "center" represents the peak central point, "-3" represents 3 kb upstream of the peak center, and "3" represents 3 kb downstream of the peak center. Each line on the vertical axis represents the signal value of a peak, and the darker the color, the higher the signal value. **B–C:** Distribution of peaks in gene functional regions. "L" represents the Luchuan pigs; "D" represents the Duroc pigs. **D:** Statistics of the typical enhancer numbers. **E-F:** Schematic of SE identification in Luchuan and Duroc pigs. The abscissa represents the peak number and the ordinate represents the enhancer signal value. All points are sorted by signal value, with slopes >1 denoted as SE and the rest as typical enhancers. **G:** Statistics of the SE numbers.

Table 1
Partial Super-enhancer information.

chr	start	end	Super-enhancer ID	signal value	sample
1	78172598	78224985	8_peaks659_lociStitched	151644.65	Luchuan pig
3	97234641	97334311	14_peaks15167_lociStitched	91865.84	Luchuan pig
6	84696568	84828400	23_peaks27436_lociStitched	216072.65	Luchuan pig
7	20839860	20911905	10_peaks19477_lociStitched	65928.38	Luchuan pig
12	3180845	3309679	22_peaks5640_lociStitched	202282.26	Luchuan pig
18	45412053	45495429	13_peaks16400_lociStitched	114266.81	Luchuan pig
1	21547128	21577205	8_peaks377_lociStitched	44432.75	Duroc pig
2	463946	534654	11_peaks16747_lociStitched	96169.95	Duroc pig
3	132341383	132379324	7_peaks21598_lociStitched	41067.34	Duroc pig
4	44592525	44639871	7_peaks22184_lociStitched	52042.72	Duroc pig
6	166524847	166534587	2_peaks28374_lociStitched	51660.96	Duroc pig
7	4177076	4227408	10_peaks28542_lociStitched	37391.64	Duroc pig

found. The SE ends were then extended by 12.5 kb in length. Finally, the anchor of the loop far from the SE end point was overlapped to the SE and its extension segment. If the overlap was successful, the SE and loop were identified as valid entries, thereby obtaining the corresponding remote target gene.
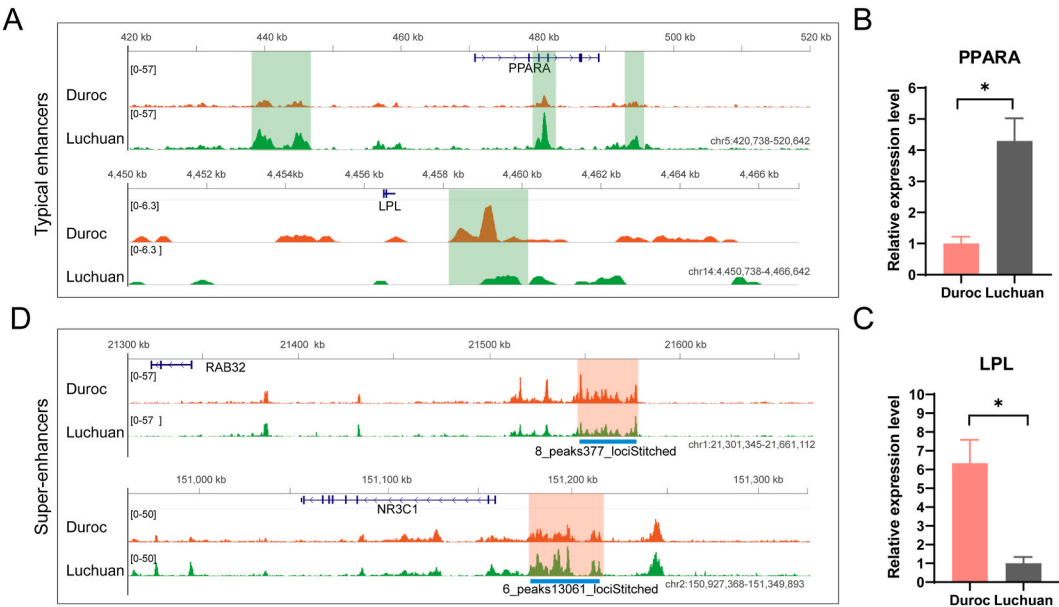
Fig. 3. Visualization of different typical enhancers and SEs. **A:** The typical enhancers near the *PPARA* and *LPL* genes were visualized using the IGV tool, with green color representing the peak region of difference. **B–C:** The gene expression levels of *PPARA* and *LPL* were verified by qPCR. **D:** Visualization of SEs near the *RAB32* and *NR3C1* genes, with pink shadows representing the peak region of difference.

Table 2
The motif analysis of different peaks (up).

Rank	motif	Name/Source of reference data	p_value	% of targets sequences with motif	Luchuan vs Duroc peak
1		**AT2G01818** (PLATZ)/col-AT2G01818-DAP-Seq (GSE60143)	1.00E-04	3.79%	up
2		**HSF6** (HSF)/col-HSF6-DAP-Seq (GSE60143)	1.00E-04	4.87%	up
3		**HRE** (HSF)/Striatum-HSF1-ChIP-Seq (GSE38000)	1.00E-04	4.19%	up
4		**NF1-halfsite** (CTF)/LNCaP-NF1-ChIP-Seq (Unpublished)	1.00E-03	23.14%	up
5		**HSFA1E** (HSF)/col-HSFA1E-DAP-Seq (GSE60143)	1.00E-03	2.30%	up
6		**NFIL3** (bZIP)/HepG2-NFIL3-ChIP-Seq (Encode)	1.00E-03	8.39%	up
7		**HLF** (bZIP)/HSC-HLF.Flag-ChIP-Seq (GSE69817)	1.00E-03	10.28%	up
8		**HRE** (HSF)/HepG2-HSF1-ChIP-Seq (GSE31477)	1.00E-03	2.98%	up
9		**AGL95** (ND)/col-AGL95-DAP-Seq (GSE60143)	1.00E-03	2.57%	up
10		**HSF7** (HSF)/colamp-HSF7-DAP-Seq (GSE60143)	1.00E-03	4.06%	up

2.8. QTL data and frequency analysis of pigs

Pig QTL data were download from the Animal Quantitative Trait Loci (QTL) Database (Animal QTLdb, https://www.animalgenome.org/cgi-bin/QTLdb/SS/index) [23]. The joint analysis of SE and QTL was performed using an in-house Python

Table 3
The motif analysis of different peaks (down).

Rank	motif	Name/Source of reference data	p_value	% of targets sequences with motif	Luchuan vs Duroc peak
1		**ERG** (ETS)/VCaP-ERG-ChIP-Seq (GSE14097)	1.00E-06	21.69%	down
2		**Etv2** (ETS)/ES-ER71-ChIP-Seq (GSE59402)	1.00E-05	12.85%	down
3		**EWS:ERG-fusion** (ETS)/CADO_ES1-EWS:ERG-ChIP-Seq (SRA014231)	1.00E-05	10.69%	down
4		**NF1** (CTF)/LNCAP-NF1-ChIP-Seq (Unpublished)	1.00E-04	6.58%	down
5		**Atf3** (bZIP)/GBM-ATF3-ChIP-Seq (GSE33912)	1.00E-04	8.94%	down
6		**Jun-AP1** (bZIP)/K562-cJun-ChIP-Seq (GSE31477)	1.00E-04	3.91%	down
7		**ETV4** (ETS)/HepG2-ETV4-ChIP-Seq (ENCODE)	1.00E-04	13.87%	down
8		**ZFX** (Zf)/mES-Zfx-ChIP-Seq (GSE11431)	1.00E-04	17.27%	down
9		**Fra2** (bZIP)/Striatum-Fra2-ChIP-Seq (GSE43429)	1.00E-04	6.68%	down
10		**AP-1** (bZIP)/ThioMac-PU.1-ChIP-Seq (GSE21512)	1.00E-04	9.76%	down

script. In brief, when the SE falls into the control range of QTL, it was considered to have a mapping relationship with this trait, and the cumulative frequency of mapping different SEs on this trait was counted.

2.9. Cell culture

Pre-adipocytes were isolated from the subcutaneous abdominal tissue of 1-week-old Duroc pig. Cells cultured at 37 $^\circ$C and 5% CO_2 using F12/DMEM medium (Thermo Fisher Scientific Inc., CAS#11765054) supplemented with 20% fetal bovine serum. Once the cells reached a confluence of 70–80% and became confluent, (+)JQ-1 inhibitor was added at a final concentration of 250 nM (Med Chem Express, CAS#1268524-70-4). After culturing for 24 h, total RNA was extracted, cDNA was synthesized through reverse transcription, and gene expression was detected by qPCR.

2.10. Data collection

The CUT&Tag sequencing data, ATAC-seq analysis data, loops of Hi-C sequencing data and RNA-seq expression data related to this research project is available through the "Data Availability Statement" section of this article.

2.11. Statistical analysis

The data were input into GraphPad Prism v8.0.2 software for calculation and drawing. Data were expressed as the mean plus standard error of the mean (SEM). Statistical analysis was performed using IBM SPSS Statistics 22 software (IBM, Chicago, IL, United States). Two groups of data were analyzed using the unpaired two-tailed *t*-test. Significance analysis of length and expression level between samples was performed by the Wilcoxon test. $p < 0.05$ (*) was considered significant, $p < 0.01$ (**) and $p < 0.001$ (***) were considered to be extremely significant.

3. Results

### 3.1. Overview of CUT&Tag sequencing data

To examine genome-wide SEs in Luchuan and Duroc pigs, we profiled the H3K27ac histone modifications in adipose tissue by CUT&Tag sequencing. The specific process is shown in Fig. 1A. The procedures are recorded in the Methods section. CUT&Tag technology relies on the Tn5 enzyme's ability to cut the DNA chain. Analyzing the size distribution of DNA fragments after enzyme

Table 4
Remote target genes of Super-enhancer.

| chr	start	end	SE_ID	SEsignal_value	chr	start	end	loop_ID	p_value	chr	start	end	gene_ID	name	sample
1	165072134	165124629	7_peaks1437_lociStitched	75419.5665	1	164540000	165080000	loop_38930	1.85E-13	1	164524502	164540118	ENSSSCG00000004947	LCTL	Luchuan pig
3	8446171	8532295	11_peaks19469_lociStitched	109368.868	3	8540000	8700000	loop_107286	2.97E-153	3	8711215	8716871	ENSSSCG00000027486	TRIP6	Luchuan pig
4	95894359	95958064	11_peaks22900_lociStitched	116293.478	4	95900000	96200000	loop_163461	4.08E-22	4	96233327	96239285	ENSSSCG00000006588	S100A9	Luchuan pig
6	48562359	48606306	6_peaks18929_lociStitched	158380.593	6	48500000	48560000	loop_212964	1.40E-23	6	48811148	48830976	ENSSSCG00000033836	SHKBP1	Luchuan pig
9	104844783	104913338	6_peaks24055_lociStitched	86166.3054	9	104880000	106760000	loop_334082	1.94E-21	9	105004987	105265595	ENSSSCG00000026110	SRPK2	Luchuan pig
10	14381636	14432274	6_peaks2662_lociStitched	80792.929	10	14380000	14440000	loop_342607	5.76E-11	10	14594599	14649619	ENSSSCG00000010861	COQ8A	Luchuan pig
1	165072134	165124629	7_peaks1437_lociStitched	75419.5665	1	164540000	165080000	loop_38930	1.85E-13	1	164778679	164781055	ENSSSCG00000004949	SMAD6	Duroc pig
14	74837199	74922375	7_peaks8157_lociStitched	102080.4936	14	73780000	74860000	loop_468176	4.38E-08	14	73798761	73807281	ENSSSCG00000010275	PCBD1	Duroc pig
6	49546360	49562767	4_peaks17845_lociStitched	59677.1811	6	48500000	49540000	loop_216287	1.19E-205	6	48553842	48606167	ENSSSCG00000002989	AKT2	Duroc pig
15	7494562	7523385	6_peaks8445_lociStitched	119768.212	15	7520000	9200000	loop_489604	4.52E-07	15	8190607	8812049	ENSSSCG00000015671	ARHGAP15	Duroc pig
4	95894359	95958064	11_peaks22900_lociStitched	116293.478	4	95900000	96220000	loop_163549	2.41E-10	4	96302070	96305110	ENSSSCG00000037666	PGLYRP3	Duroc pig
9	104844783	104913338	6_peaks24055_lociStitched	86166.3054	9	104880000	105160000	loop_330903	1.99E-90	9	105004987	105265595	ENSSSCG00000026110	SRPK2	Duroc pig
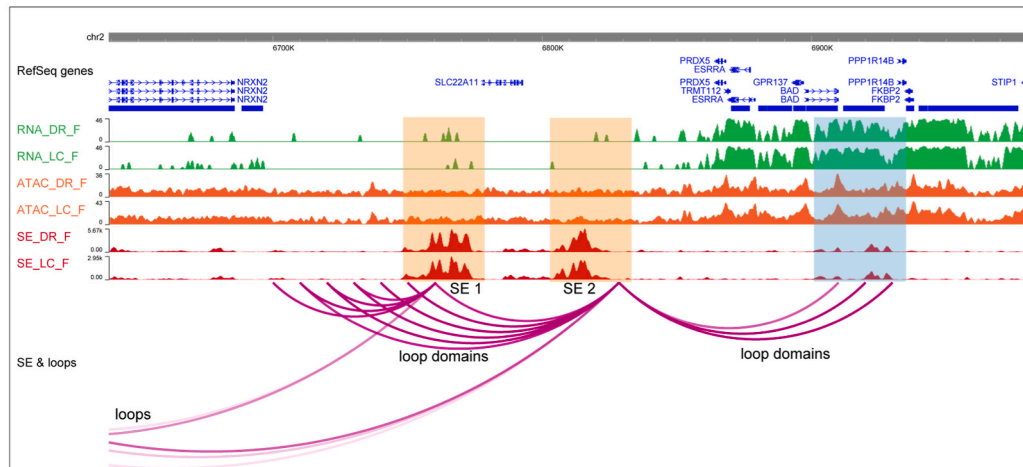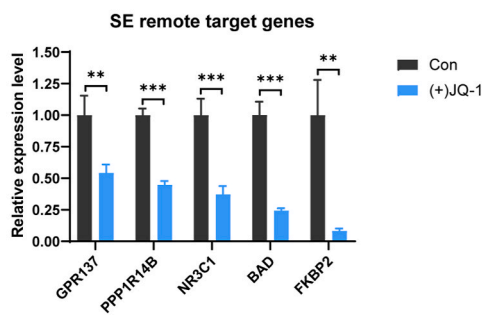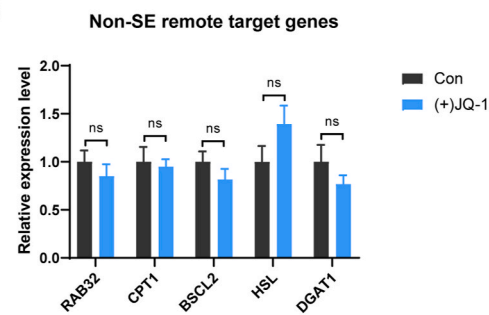
Fig. 4. Example diagram of SE driving remote genes with loop domains. **A:** Green is the RNA-seq expression track, orange is the ATAC-seq signal track, red is the SE signal track, and the purple curve represents loops. The yellow color represents the SE location, and the blue color represents the remote gene region activated near the loop anchor. **B–C:** Duroc porcine pre-adipocytes were treated with (+)JQ-1 inhibitor, then the expression of SE remote target genes was detected by qPCR.

digestion is helpful to master the cutting efficiency of Tn5 enzyme. Three distinct peaks can be observed in Fig. S1 (Supplemental file 1). It represents the length distribution of chromatin accessible regions, mononucleosome DNA fragments, and dinucleosome DNA fragments, respectively. These fragments were consistent with the expected length of nucleosome DNA. The mononucleosome length was generally about 147 bp, indicating that samples were of good quality and the digestion efficiency of Tn5 enzyme was appropriate. When all libraries were constructed and qualified, they underwent sequencing using the Illumina NovaSeq 6000 platform.

A total of 55.61 GB of raw data were obtained from six fat samples. After strict quality control, 46.65 GB clean data were generated. The Q30 value of each sample remained above 90% (Supplemental file 2: Table S1), indicating that the data quality of CUT&Tag sequencing was very good. BWA-MEME software [19] was then used to compare clean data to the genome. The unique alignment rate of each immunoprecipitation sample was over 89.5% (Supplemental file 3: Table S2); therefore, the quality of this batch of sequencing data fully met the need for subsequent analysis. Subsequently, MACS2 software was used to determine call peak. The peak number of each sample is shown in Supplemental file 3: Table S3. The deepTools software [24] was used to calculate the normalized signal values in the peak center area, and the results confirmed that the signal distribution patterns of Luchuan pig and Duroc pig were of quality and concentrated (Supplemental file 1: Fig. S2). In addition, PCA analysis of the samples showed that the groups were independent (Fig. 1B), suggesting that there might be some differences in SEs between the two pig breeds. Therefore, we first investigated the peak profiles of Luchuan and Duroc pigs and found that they were basically consistent with the results reported by Whyte [8]. Peak characteristic analysis showed that most peak lengths were in the 200–1600 bp range (Supplemental file 1: Fig. S3). However, there were special spikes at the start and end sites of SEs in the pig breeds (Fig. 1C), and signal depression in the main area. In addition, the signal intensity of SEs in Luchuan pig was stronger than that in Duroc pig. Following clustering analysis of peak signals and construction of a heat map (Fig. 2A), it was found that the different peaks of the two pigs were clustered into five clusters, suggesting that there were huge differences in the SEs of these peak sites. Thus, we calculated the proportion of these peaks in functional regions across the genome (Fig. 2B). It was found that the up-regulated peaks were distributed in the promoter (≤1 kb) region (Fig. 2C), which were much higher than the down-regulated peaks. These results suggest that some SEs in the promoter region of adipose tissue of Luchuan pig may be relatively active.
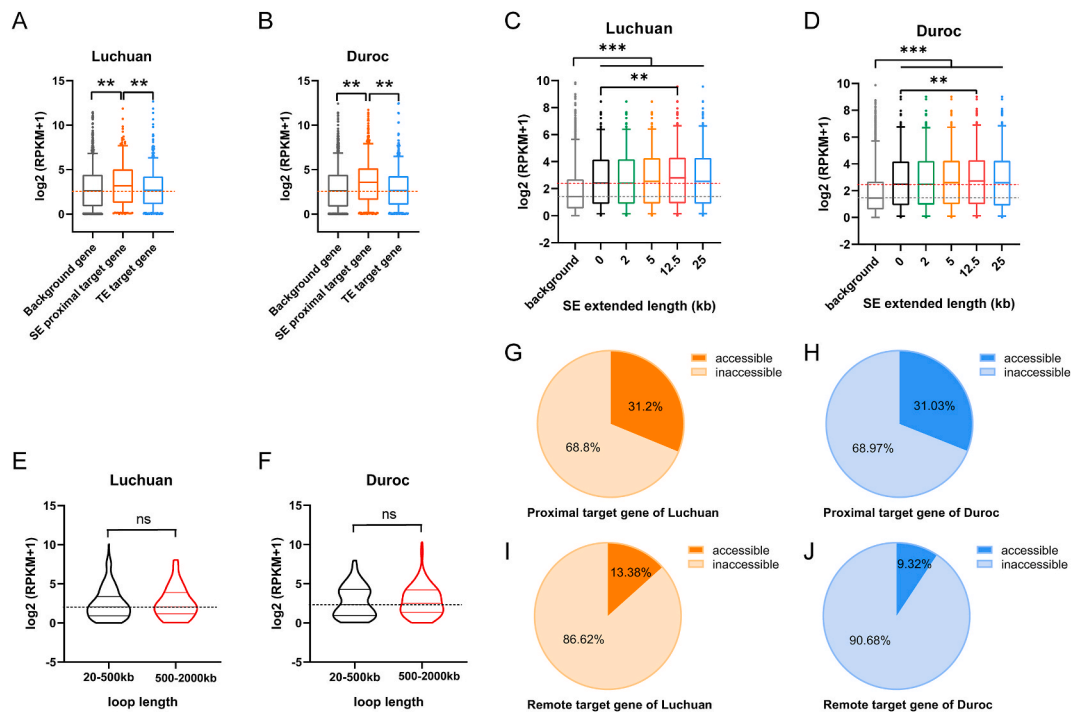
Fig. 5. Expression level calculation and promoter accessibility analysis of proximal and remote genes. **A-B:** Different types of prediction gene expression statistics of Luchuan and Duroc pigs. A total of 1500 background genes were randomly selected from all gene pools. The gene closest to the SE was the proximal target gene. Typical enhancer target genes were predicted by ROSE software. **C-D:** Changes in the expression level of remote target genes were analyzed when the SE extended to the loop anchor point with different lengths. **E-F:** Analysis of the expression levels of distal target genes under different loop lengths. **G-J:** Accessibility analysis of proximal genes and remote genes using ATAC-seq data. The predicted promoter was defined in the range of 2000 bp upstream to 200 bp downstream of the transcription start sites (TSS) site. Significance analysis of expression was performed by the Wilcoxon test. *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$.

3.2. Identification and characterization of super-enhancers

Using ROSE software, 25,463 and 26,085 typical enhancers were obtained from the adipose tissue of Luchuan and Duroc pigs, respectively (Fig. 2D and Supplemental file 4: Tables S4–5). According to Young Lab's definition of SE [8], we performed stitching and sorting of these enhancers (Fig. 2E–G). When the slope was greater than 1, Luchuan pigs had 145 SEs, and Duroc pigs generated 378 SEs (Table 1). The details of these SEs are shown in Supplemental file 5: Tables S6–7. We observed peak signal patterns of Luchuan and Duroc pigs, and found that there were some differences in the typical enhancers signal near some genes related to lipid metabolism, such as *PPARA, LPL* (Fig. 3A). Additionally, the expression levels of these genes also varied (Fig. 3B–C and Supplemental file 6: Table S8). Interestingly, the SE at the same position in the two pig breeds also showed differences in signal intensity (Fig. 3B). These results confirmed the previous speculation of large differences in peak clusters (Fig. 2A).

Transcription factors tend to bind to specific DNA sequences that often have highly conservative nucleotide arrangement patterns. Such sequence patterns are known as motifs of the transcription factor binding site. In order to explore the differences in transcription factors bound to the SEs of Luchuan and Duroc pigs, we used the findMotifsGenome.pl tool for prediction analysis, and the results showed that the main motifs corresponding to the transcription factors of Luchuan vs Duroc up-regulated peaks were AT2G01818, HSF6, HRE, NF1-halfsite and so on (Table 2). The motifs of down-regulated peaks were mainly ERG, Etv2, EWS and NF1 (Table 3). Interestingly, AGL95 and Atf3 were also found in the results. AGL is known to be involved in the glycogen degradation process [25,26], while Atf3 has recently been reported to be involved in fat metabolism and energy metabolism [27–29].

3.3. Prediction of the proximal target gene and regulation analysis of the remote target gene of the super-enhancer

Eukaryotic chromatin can be folded and wound to form complex 3D structures, including loop domains that bring two regions that are linearly far apart closer to each other. Michal [30] previously reported that two distant genes can achieve enhancer shared regulation through loop structure. Therefore, we assumed that SE may also realize the transcription of long-distance driving target genes with the help of loop domains. By joint analysis of the loops data of Hi-C sequencing, we defined the gene with the closest linear genomic distance to SE as the proximal target gene (Supplemental file 7: Tables S9–10), and the gene that was narrowed by loops was called the remote target gene (Table 4 and Supplemental file 8: Tables S11–12). The prediction method is shown in the Methods section. We observed that the SE can selectively activate a portion of remote target gene transcription through loop domains (Fig. 4A).
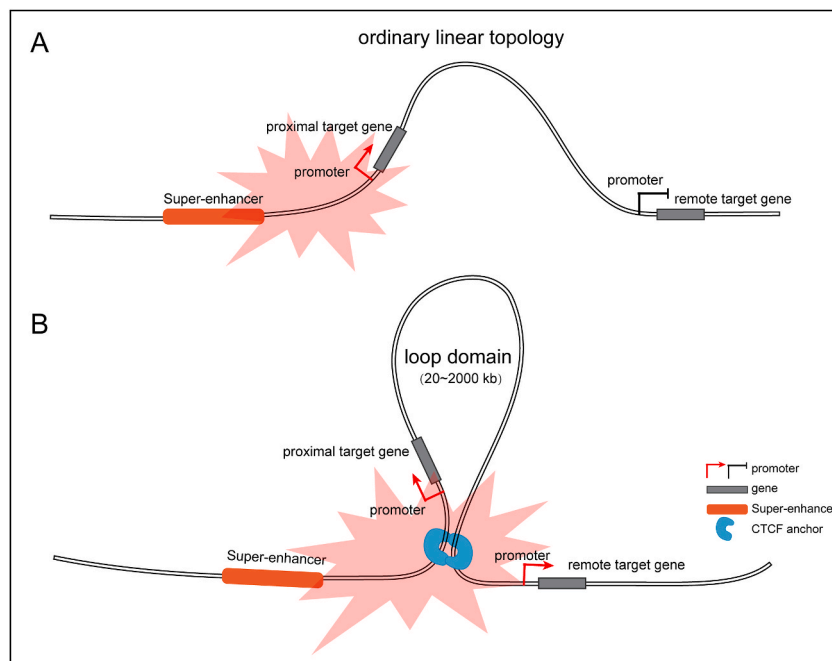
Fig. 6. The hypothetical model diagram of SE regulates remote genes through the loop domain. **A:** In the ordinary linear topology, SE can activate the proximal gene, but cannot affect the distal gene transcription. **B:** With the help of the loop domain, the SE can affect both proximal and distal gene.

We treated Duroc porcine pre-adipocytes with a (+)JQ-1 inhibitor (a specific and reversible BET bromodomain inhibitor) and observed significant inhibition of most Super-enhancer distal target genes (Fig. 4B), while non-SE target genes remained unaffected by the inhibitor (Fig. 4C). This result supports our hypothesis to a certain extent. In fact, SEs can easily improve the overall expression level of proximal target genes compared with the background genes and the typical enhancers target genes (Fig. 5A–B). When analyzing the data, we observed that SEs were often some distance from the loop anchor point. To investigate whether this distance affects the efficiency of SE in driving transcription, we evaluated the expression levels of remote target genes under different SE endpoint extension conditions. As the average stitched length of the SE formed by ROSE software was 12.5 kb, we took that as the expected value. We found that the expression level of the remote target gene was highest when the extended length was 12.5 kb, but the expression level of the remote target gene decreased when the length was lower or higher (Fig. 5C–D).

Considering that loop length might affect the SE driving effect, we measured the remote target gene expression at different loop lengths and found no significant change in overall remote target gene expression level in both Luchuan and Duroc pigs (Fig. 5E–F). Based on the 12.5 kb extension, we analyzed the chromatin accessibility of the promoter region of the SE remote target gene using ATAC-seq data, and found that 31.2% of Luchuan and 31.03% of Duroc proximal target genes showed chromatin accessibility of the promoter region (Fig. 5G–H). With regard to analysis of SE remote target genes, we were surprised to find that 13.8% and 9.32% promoter regions remained accessible (Fig. 5I–J). This means that even though SE and remote target genes were far apart in linear distance (the average distance was more than 600 kb), they can still become close to each other in space through the loop domains and play a certain role in regulating transcription. Based on the above results, we hypothesize and model the method for predicting remote target genes, as shown in Fig. 6A–B.

3.4. Joint analysis of super-enhancer and QTL

Quantitative trait loci (QTL) plays an important role in pig breeding and production. We analyzed the mapping relationship between SE and QTL. The method of SE frequency analysis is shown in the Methods section. Through mapping analysis, it was found that a large number of SEs were mapped to the adipose-related QTL (Fig. 7A–B). Interestingly, the top five QTLs of Luchuan and Duroc pigs were mainly enriched in several traits such as drip loss, average backfat thickness, loin muscle area, backfat at last rib, and teat number. We defined them as SEs associated with lipid metabolism, the details are shown in Table 5 and Supplemental file 9: Tables S13–14, and their corresponding remote target genes and proximal target genes are more likely to be driven by the SE.

4. Discussion

Luchuan pig is an obese and famous Chinese breed, while Duroc is a widely used lean pig breed. Both breeds are excellent animal models for studying fat deposition. Fat deposition in pigs is often regulated by multiple genes. According to previous reports, SEs often
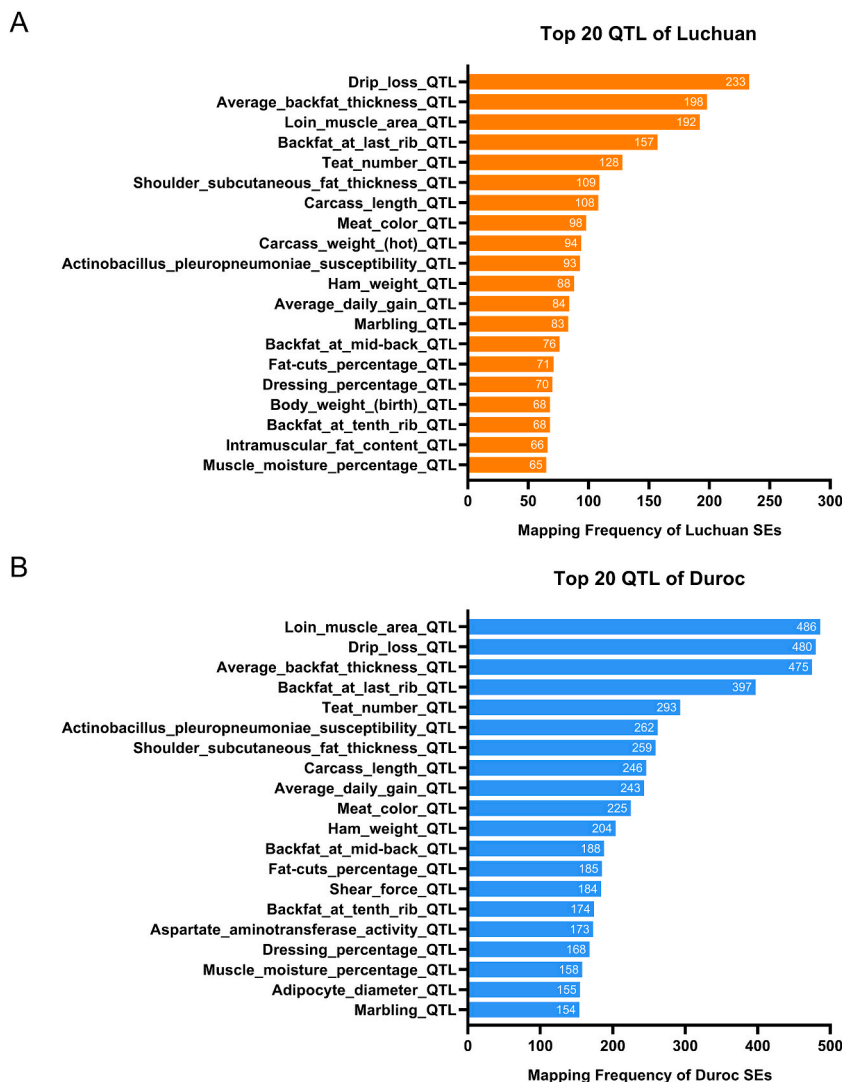
A

Top 20 QTL of Luchuan

B

Top 20 QTL of Duroc

Fig. 7. Mapping frequency statistics of SEs in different QTLs of pigs. **A-B:** The top 20 QTLs were calculated based on SEs cross mapping frequency of Luchuan and Duroc pigs.

have a multi-gene regulatory role due to their strong driving ability [31–33]. However, due to the lack of SE studies in pigs, we obtained 145 SEs in Luchuan pigs and 378 SEs in Duroc pigs using the CUT&Tag technique. Following a comparative analysis, it was found that there were significant differences in peak intensity and coverage ratio of these peaks in the gene promoter region between the two breeds. Subsequently, we also found that peak signals at the start and end point of the SE peak profile showed obvious spikes, while the body region showed an overall downward trend. These findings are similar to SE data obtained by Lovén et al. using MED1 and BRD4 [9]. We hypothesized that spikes at both ends might be characteristic of the SE obtained by ChIP-seq in H3K27ac of pigs. On the other hand, the analysis showed that the expression level of both proximal and remote target genes was higher than the background gene level. Notably, the proximal target gene expression was significantly higher than that of the typical enhancer gene. These results are consistent with those reported by Lovén et al. [9] and Dong et al. [34].

The expression of remote target genes was only related to the distance from the SE extended to the loop anchor, and the maximum expression was strongly correlated with the average SE length of 12.5 kb, independent of the loop length, which was an interesting finding. Therefore, we proposed a model of the SE's remote drive target gene (Fig. 6A–B), which explains why SE could drive remote target gene transcription by loops on a 3D spatial scale. However, the specific molecular mechanisms involved require confirmation by further research. Data from ATAC-seq assisted in verifying that the SE did indeed have an effect on driving remote target genes. Even at a long linear distance from the SE, approximately 10% of remote target gene promoters were accessible. Therefore, our method may reveal remote potential regulatory genes that have not yet been discovered. Of course, the molecular mechanism of SE remote activation requires further verification. Correlation analysis between pig QTL traits and SEs helped us to obtain a clearer picture of which SEs controlled fat deposition traits. Therefore, some SEs related to lipid metabolism were determined, which will provide a research

Table 5
Super-enhancers associated with fat metabolism traits.

chr	start	end	Super-enhancer_ID	value	chr	start	end	QTL	QTL_ID	sample
1	273934220	274012217	13_peaks3569_lociStitched	118376	1	227805980	275571750	Backfat_at_tenth_rib_QTL	23	Luchuan pig
4	95894359	95958064	11_peaks22900_lociStitched	116293.5	4	62054538	119527200	Average_backfat_thickness_QTL	162	Luchuan pig
5	19208215	19266337	10_peaks24176_lociStitched	148690.5	5	11124985	103928896	Backfat_at_mid-back_QTL	5210	Luchuan pig
10	39875246	39875967	1_peaks2920_lociStitched	132939.3	10	36817956	48807828	Backfat_at_last_rib_QTL	3787	Luchuan pig
14	123410154	123460911	10_peaks8663_lociStitched	137406.2	14	84693963	137406474	Fat-cuts_percentage_QTL	1162	Luchuan pig
16	71314390	71384532	13_peaks14615_lociStitched	129903	16	39864062	78708148	Average_backfat_thickness_QTL	5728	Luchuan pig
2	15185135	15245332	10_peaks16343_lociStitched	89326.3	2	668487	108072162	Average_backfat_thickness_QTL	21212	Duroc pig
5	21123364	21208677	10_peaks15696_lociStitched	116256	5	20096748	78807961	Backfat_at_mid-back_QTL	5209	Duroc pig
14	72379858	72433769	10_peaks11247_lociStitched	43684.1	14	58527320	84693963	Adipocyte_diameter_QTL	12850	Duroc pig
15	25328280	25382966	10_peaks12531_lociStitched	54341.5	15	15777650	137082629	Backfat_at_tenth_rib_QTL	5725	Duroc pig
16	51733294	51789416	10_peaks14598_lociStitched	47051.2	16	39864062	78708148	Average_backfat_thickness_QTL	5728	Duroc pig
17	51190187	51236886	10_peaks15704_lociStitched	46932.5	17	38860318	63780895	Backfat_at_last_rib_QTL	1191	Duroc pig

basis for the subsequent study of SE related genes driving lipid metabolism and fat deposition traits.

5. Conclusions

Overall, 145 SEs of Luchuan pigs and 378 SEs of Duroc pigs were determined in this study, and different characteristics of the SE peaks of the two pig breeds were found. The peak coverage ratio of SE peaks in the gene promoter region was significantly different between the two breeds. Peak signals at the start and end point of the SE peak profile showed obvious spikes. In addition, proximal target genes of the SE were highly expressed compared with the background genes and the typical enhancer target genes. Using loop data from Hi-C sequencing, we predicted the SE's remote target genes and found that their expression level was related to the distance the SE extended to the loop's anchor, but not the length of loops. Moreover, from our prediction method, some remote potential regulatory genes were revealed which had not been discovered before. The ATAC-seq data conjoint analysis confirmed that approximately 10% of the remote target gene's promoters were accessible. Finally, the QTL coalition analysis identified a batch of SEs associated with fat metabolism traits. These results will provide a new basis for the study of pig epigenetics and broaden the understanding of remote target gene regulation of SEs.

Data availability statement

The CUT&Tag sequencing data presented in the study are publicly released in the SRA: https://trace.ncbi.nlm.nih.gov/Traces/?view=run_browser&acc=SRR19392813&display=data-access, the accession number is SRR19392813 to SRR19392824. Bioproject number is PRJNA842197. The ATAC-seq analysis data, loops of Hi-C sequencing data and RNA-seq expression data can be found in the Supplemental files 10.

Ethics approval and consent to participate

All experimental protocols in the study were approved by the Institutional Animal Care and Use Committee of Guangxi University (Approval NO. GXU-2021-042), all methods were carried out in accordance with relevant guidelines and regulations. The study was conducted in accordance with ARRIVE guidelines (https://arriveguidelines.org).

Consent for publication

Not applicable.

CRediT authorship contribution statement

Lin Yu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Conceptualization. **Tengda Huang:** Formal analysis. **Siqi Liu:** Supervision. **Jingsu Yu:** Validation. **Menglong Hou:** Software. **Songtao Su:** Visualization, Conceptualization. **Tianyu Jiang:** Resources. **Xiangling Li:** Funding acquisition. **Yixing Li:** Resources. **Turtushikh Damba:** Formal analysis. **Lei Zhou:** Writing – review & editing, Funding acquisition, Conceptualization. **Yunxiao Liang:** Funding acquisition, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Abbreviation

SE   Super-enhancer
TE   Typical enhancer
CUT&Tag   Cleavage Under Targets and Tagmentation

TF	Transcription factor
QTL	Quantitative trait loci
ATAC-seq	Transposase-accessible chromatin with high-through sequencing
Hi-C seq	High-throughput chromosome conformation capture sequencing

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e25725.

References

[1] D. Hnisz, B.J. Abraham, T.I. Lee, A. Lau, V. Saint-André, A.A. Sigova, H.A. Hoke, R.A. Young, Super-enhancers in the control of cell identity and disease, Cell 155 (4) (2013) 934–947.

[2] Y. Jiang, F. Qian, X. Bai, Y. Liu, Q. Wang, B. Ai, X. Han, S. Shi, J. Zhang, X. Li, et al., SEdb: a comprehensive human super-enhancer database, Nucleic acids research 47 (D1) (2019) D235–d243.

[3] S. Pott, J.D. Lieb, What are super-enhancers? Nature genetics 47 (1) (2015) 8–12.

[4] S. Sethi, I.E. Vorontsov, I.V. Kulakovskiy, S. Greenaway, J. Williams, V.J. Makeev, S.D.M. Brown, M.M. Simon, A.M. Mallon, A holistic view of mouse enhancer architectures reveals analogous pleiotropic effects and correlation with human disease, BMC genomics 21 (1) (2020) 754.

[5] R. Nakato, T. Sakata, Methods for ChIP-seq analysis: a practical workflow and advanced applications, Methods (San Diego, Calif) 187 (2021) 44–53.

[6] H.S. Kaya-Okur, S.J. Wu, C.A. Codomo, E.S. Pledger, T.D. Bryson, J.G. Henikoff, K. Ahmad, S. Henikoff, CUT&Tag for efficient epigenomic profiling of small samples and single cells, Nature communications 10 (1) (2019) 1930.

[7] S. Blinka, M.H. Reimer Jr., K. Pulakanti, S. Rao, Super-enhancers at the nanog locus differentially regulate neighboring pluripotency-associated genes, Cell reports 17 (1) (2016) 19–28.

[8] W.A. Whyte, D.A. Orlando, D. Hnisz, B.J. Abraham, C.Y. Lin, M.H. Kagey, P.B. Rahl, T.I. Lee, R.A. Young, Master transcription factors and mediator establish super-enhancers at key cell identity genes, Cell 153 (2) (2013) 307–319.

[9] J. Lovén, H.A. Hoke, C.Y. Lin, A. Lau, D.A. Orlando, C.R. Vakoc, J.E. Bradner, T.I. Lee, R.A. Young, Selective inhibition of tumor oncogenes by disruption of super-enhancers, Cell 153 (2) (2013) 320–334.

[10] Y. Peng, H. Kang, J. Luo, Y. Zhang, A comparative analysis of super-enhancers and broad H3K4me3 domains in pig, human, and mouse tissues, Frontiers in genetics 12 (2021) 701049.

[11] Y. Luan, L. Zhang, M. Hu, Y. Xu, Y. Hou, X. Li, S. Zhao, Y. Zhao, C. Li, Identification and conservation analysis of cis-regulatory elements in pig liver, Genes 10 (5) (2019).

[12] Z. Zhou, Y. Zhu, Z. Zhang, T. Jiang, Z. Ling, B. Yang, W. Li, Comparative analysis of promoters and enhancers in the pituitary glands of the Bama Xiang and large white pigs, Frontiers in genetics 12 (2021) 697994.

[13] L. Yu, L. Tai, L. Zhang, Y. Chu, Y. Li, L. Zhou, Comparative analyses of long non-coding RNA in lean and obese pig, Oncotarget 8 (25) (2017) 41440–41450.

[14] W. Miao, Z. Ma, Z. Tang, L. Yu, S. Liu, T. Huang, P. Wang, T. Wu, Z. Song, H. Zhang, et al., Integrative ATAC-seq and RNA-seq analysis of the longissimus muscle of luchuan and Duroc pigs, Frontiers in nutrition 8 (2021) 742672.

[15] L. Yu, L. Tai, J. Gao, M. Sun, S. Liu, T. Huang, J. Yu, Z. Zhang, W. Miao, Y. Li, et al., A new lncRNA, lnc-LLMA, regulates lipid metabolism in pig hepatocytes, DNA and cell biology 41 (2) (2022) 202–214.

[16] K. Xing, H. Liu, F. Zhang, Y. Liu, Y. Shi, X. Ding, C. Wang, Identification of key genes affecting porcine fat deposition based on co-expression network analysis of weighted genes, Journal of animal science and biotechnology 12 (1) (2021) 100.

[17] J. Yue, X. Hou, X. Liu, L. Wang, H. Gao, F. Zhao, L. Shi, L. Shi, H. Yan, T. Deng, et al., The landscape of chromatin accessibility in skeletal muscle during embryonic development in pigs, Journal of animal science and biotechnology 12 (1) (2021) 56.

[18] M.R. Corces, A.E. Trevino, E.G. Hamilton, P.G. Greenside, N.A. Sinnott-Armstrong, S. Vesuna, A.T. Satpathy, A.J. Rubin, K.S. Montine, B. Wu, et al., An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues, Nature methods 14 (10) (2017) 959–962.

[19] Y. Jung, D. Han, Bwa-Meme, BWA-MEM Emulated with a Machine Learning Approach, Bioinformatics, Oxford, England, 2022.

[20] S. Heinz, C. Benner, N. Spann, E. Bertolino, Y.C. Lin, P. Laslo, J.X. Cheng, C. Murre, H. Singh, C.K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities, Molecular cell 38 (4) (2010) 576–589.

[21] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics (Oxford, England) 26 (6) (2010) 841–842.

[22] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome biology 15 (12) (2014) 550.

[23] Z.L. Hu, C.A. Park, J.M. Reecy, Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services, Nucleic acids research 50 (D1) (2022) D956–d961.

[24] F. Ramírez, D.P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A.S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: a next generation web server for deep-sequencing data analysis, Nucleic acids research 44 (W1) (2016) W160–W165.

[25] M. Okubo, A. Horinishi, M. Takeuchi, Y. Suzuki, N. Sakura, Y. Hasegawa, T. Igarashi, K. Goto, H. Tahara, S. Uchimoto, et al., Heterogeneous mutations in the glycogen-debranching enzyme gene are responsible for glycogen storage disease type IIIa in Japan, Human genetics 106 (1) (2000) 108–115.

[26] I. Hussain, N. Patni, A. Garg, Lipodystrophies, dyslipidaemias and atherosclerotic cardiovascular disease, Pathology 51 (2) (2019) 202–212.

[27] Y. Xu, Y. Li, K. Jadhav, X. Pan, Y. Zhu, S. Hu, S. Chen, L. Chen, Y. Tang, H.H. Wang, et al., Hepatocyte ATF3 protects against atherosclerosis by regulating HDL and bile acid metabolism, Nature metabolism 3 (1) (2021) 59–74.

[28] C.F. Cheng, H.C. Ku, J.J. Cheng, S.W. Chao, H.F. Li, P.F. Lai, C.C. Chang, M.J. Don, H.H. Chen, H. Lin, Adipocyte browning and resistance to obesity in mice is induced by expression of ATF3, Communications biology 2 (2019) 389.

[29] M.K. Jang, Y. Son, M.H. Jung, ATF3 plays a role in adipocyte hypoxia-mediated mitochondria dysfunction in obesity, Biochemical and biophysical research communications 431 (3) (2013) 421–427.

[30] M. Levo, J. Raimundo, X.Y. Bing, Z. Sisco, P.J. Batut, S. Ryabichko, T. Gregor, M.S. Levine, Transcriptional coupling of distant regulatory genes in living embryos, Nature 605 (7911) (2022) 754–760.

[31] C. Zhang, S. Wei, W.P. Sun, K. Teng, M.M. Dai, F.W. Wang, J.W. Chen, H. Ling, X.D. Ma, Z.H. Feng, et al., Super-enhancer-driven AJUBA is activated by TCF4 and involved in epithelial-mesenchymal transition in the progression of Hepatocellular Carcinoma, Theranostics 10 (20) (2020) 9066–9082.

[32] X. Chen, C. Sun, C. Liu, J. Wu, lncRNA in hepatic glucose and lipid metabolism: a review, Chinese journal of biotechnology 37 (1) (2021) 40–52.

[33] S. Cui, Q. Wu, M. Liu, M. Su, S. Liu, L. Shao, X. Han, H. He, EphA2 super-enhancer promotes tumor progression by recruiting FOSL2 and TCF7L2 to activate the target gene EphA2, Cell death & disease 12 (3) (2021) 264.

[34] J. Dong, J. Li, Y. Li, Z. Ma, Y. Yu, C.Y. Wang, Transcriptional super-enhancers control cancer stemness and metastasis genes in squamous cell carcinoma, Nature communications 12 (1) (2021) 3974.