

A whole-body atlas of enhancer and super-enhancer regulatory networks reveals tissue-specific hub enhancers and hierarchical regulatory logic in pigs

Tao Jiang^{1,#,*}, Jiahong Zhong^{1,#}, Siyi Liu¹, Qiang Yang¹, Sha Liu¹, Na Li¹, Jianzhen Huang^{1,*}

¹ *National Key Laboratory for Swine Genetic Improvement and Germplasm Innovation, Ministry of Science and Technology of China, Jiangxi Agricultural University, Nanchang, 330045, China*

[#]These authors contributed equally: Tao Jiang, Jiahong Zhong.

^{*} Corresponding authors: Tao Jiang, Jianzhen Huang

Email addresses:

TJ: jiangtao19951@outlook.com

JZ: zhongjh1129@foxmail.com

SL: aprillsy@hotmail.com

QY: yangqiangjxau@outlook.com

SL: liusha202201@outlook.com

NL: nali1046@hotmail.com

JH: huang813813@jxau.edu.cn

Abstract

Enhancer regulatory networks are central determinants of tissue identity and function, yet their hierarchical organization across pig tissues remains poorly understood, largely due to incomplete tissue coverage and genetic heterogeneity. Here, leveraging genetically matched full-sibling pig pair, we generated epigenomic profiles from 151 anatomical sites across 58 tissues and reconstructed 20,468 typical enhancer regulatory networks at whole-body scale. Tissue-specific enhancers were preferentially enriched in more complex networks, and 1,535 networks showed tissue-specific functional relevance. In parallel, we identified 477 super-enhancer-mediated core regulatory circuitries involving 368 transcription factors, revealing an additional layer of tissue-specific regulation. These two frameworks were highly complementary, identifying distinct sets of hub enhancers with limited overlap. Hub enhancers identified by both frameworks were strongly enriched for CTCF, implicating three-dimensional chromatin architecture in tissue-specific regulation. Integrating single-cell chromatin accessibility and transcriptomic data from the thymus further resolved hub enhancer activity to cellular resolution. Together, this whole-body atlas provides a valuable resource for decoding enhancer regulatory hierarchies and tissue-specific gene regulatory programs in pigs.

Keywords: Enhancer regulatory networks; Hub enhancer; Super-enhancer; Pig; Thymus; Single cell

INTRODUCTION

Gene transcription is orchestrated by complex cis-regulatory programs that are essential for tissue identity, function, and homeostasis (Zhang et al., 2024). Among these, enhancers play a central role by activating target genes in a tissue-specific manner, either individually or through coordinated interactions with other regulatory elements (Zhu et al., 2022). These interactions can give rise to higher-order enhancer regulatory networks that buffer transcriptional outputs and stabilize tissue-specific gene expression programs (Lin et al., 2025). Super-enhancers (SEs), which consist of densely clustered enhancers with exceptionally strong regulatory activity, represent a specialized regulatory layer with prominent roles in establishing cell and tissue identity (Hnisz et al., 2013; Whyte et al., 2013). In parallel, SEs can mediate core regulatory circuitries (CRCs), in which master transcription factors (TFs) autoregulate themselves and each other through SE-associated feedback loops to maintain lineage-specific transcriptional states (Chen et al., 2020a; Saint-André et al., 2016). Together, typical enhancers and SEs provide complementary regulatory frameworks for organizing hierarchical gene control across tissues.

The abundance and functional complexity of enhancer elements in mammalian genomes have motivated the development of multiple computational frameworks for reconstructing enhancer-centered regulatory networks. These include ANANSE, which predicts key TFs involved in cell fate determination (Xu et al., 2021), the eNet algorithm, which identifies hub enhancers associated with cell identity through



enhancer network analysis (Hong et al., 2023), and *CRC mapper*, which infers SE-mediated core TF circuits that govern tissue- and lineage-specific transcriptional programs (Saint-André et al., 2016). In parallel, functional studies have shown that enhancers within clusters can exhibit additive, redundant, or synergistic effects in diverse developmental and tissue contexts, including the brain (Vermunt et al., 2014), spinal cord (Shu et al., 2022), and limb (Osterwalder et al., 2018). Genome-editing approaches have further dissected enhancer clusters at single-locus resolution, revealing distinct contributions of individual constituent elements to transcriptional robustness and phenotypic output (Blayney et al., 2023; Thomas et al., 2021).

Despite these advances, most current insights into enhancer network organization have been derived from cell lines, isolated cell types, or locus-specific perturbation studies, leaving their hierarchical architecture across whole-body tissue systems poorly resolved. Existing resources, such as dbCoRC (Huang et al., 2018b) and CRCdb (Feng et al., 2023), provide enhancer network annotations for human and mouse cell lines or limited tissue collections, but comparable tissue-resolved enhancer network atlases remain scarce in large mammals. In pigs, this gap is particularly pronounced because previous studies have been constrained by incomplete tissue coverage, heterogeneous genetic backgrounds, and limited anatomical resolution, hindering systematic identification of tissue-specific hub enhancers and core regulatory genes. Moreover, how typical enhancer networks and SE-mediated CRCs jointly shape tissue-specific regulatory hierarchies across organs remains largely unknown.



To address these limitations, we leveraged a genetically matched full-sibling pig pair and comprehensive epigenomic maps spanning 151 anatomical sites across 58 tissues to construct a whole-body atlas of enhancer regulatory networks. Using two complementary frameworks—eNet algorithm (Hong et al., 2023) for typical enhancers and CRC *mapper* (Saint-André et al., 2016) for SEs—we systematically reconstructed tissue-resolved regulatory networks and delineated their hierarchical organization across organs. We identified tissue-specific enhancer networks, hub enhancers, and core genes, and further showed that hub enhancers shared by both frameworks were strongly enriched for CTCF binding, linking enhancer hierarchy to three-dimensional chromatin architecture. Finally, by integrating single-cell chromatin accessibility and transcriptomic data from the thymus, we refined hub enhancer activity at cellular resolution. Together, this study establishes a comprehensive resource and conceptual framework for decoding enhancer regulatory hierarchies and tissue-specific gene regulatory programs in pigs.

MATERIALS AND METHODS

Sample collection

Pigs used in this study were the twelfth (F12) generation of a mosaic population, created by crossbreeding four commercial European/American pig breeds (Landrace, Large White, Duroc and Piétrain) with four indigenous Chinese pig breeds (Erhualian, Bamaxiang, Tibetan, Laiwu). One healthy male pig and one healthy female full sibling, both aged eight months, were used for tissue sampling. Following euthanasia, tissue



samples were collected from 151 anatomical sites, immediately transferred into 2 mL sterile freezer tubes, flash-frozen in liquid nitrogen, and stored at -80°C . until further assay processing. The tissue sites analyzed in this study encompassed nearly all regions of the pig's body, with detailed information on all sampling sites provided in Supplementary Tables S1 and S2.

CUT&Tag library preparation and sequencing

CUT&Tag libraries were prepared following previously described protocols (Kaya-Okur et al., 2019) with minor adjustments. Briefly, after generating a nucleus suspension, 2×10^5 nuclei per reaction were incubated with 10 μL of concanavalin A-coated magnetic beads for 10 minutes at room temperature. Following this, the bead-bound cells were resuspended and incubated overnight with a primary antibody (ab4729 and ab8580, Abcam, UK) at 4°C for H3K4me3 and H3K27ac, respectively. Eight tissue sites were also incubated with a primary antibody against CTCF (A13272, Abclonal, China), and the thymus sample was additionally incubated with an antibody against RAD21 (ab217678, Abcam, UK). Then, a 1:100 dilution of ChIP-grade rabbit polyclonal antibody was added for 30 minutes at room temperature. Unbound antibodies were removed using a magnet stand, followed by 2-3 washes with Dig wash buffer. The pA-Tn5 adapter complex was added and incubated at room temperature for 1 hour, after which unbound pA-Tn5 protein was removed. The cells were resuspended in tagmentation buffer and incubated at 37°C for 1 hour, after which the immunoprecipitated DNA was collected.



129

130 To amplify libraries, 21 μ L of DNA was mixed with 2 μ L of a universal i5 primer and
131 a uniquely barcoded i7 primer, with a different barcode for each sample. A total of 25
132 μ L of NEBNext HiFi 2 $\times$ PCR Master mix was added, and the sample was thoroughly
133 mixed. PCR amplification was performed in a thermocycler with a heated lid under the
134 following conditions: 72 $^{\circ}$ C for 3 minutes (gap filling), 95 $^{\circ}$ C for 30 seconds, followed
135 by 12 cycles of 98 $^{\circ}$ C for 10 seconds and 60 $^{\circ}$ C for 5 seconds, with a final extension at
136 72 $^{\circ}$ C for 1 minute and hold at 4 $^{\circ}$ C. Library clean-up was performed using SPRI
137 Magnetic Beads (Beckman Coulter, Beverly, MA, USA), and DNA was purified using
138 phenol-chloroform-isoamyl alcohol extraction followed by ethanol precipitation. The
139 DNA concentration was assessed using a Qubit 4.0 fluorometer (Thermo Fisher
140 Scientific, USA), and the library quality was evaluated on BiOptic QSep400 (China).
141 Finally, the library was sequenced in 150 bp paired-end mode on a NovaSeq 6000
142 platform (USA).

143

144 **Quantitative real-time PCR**

145 Tissue samples including brain, skeletal muscle, stomach, liver and heart were collected
146 and ground in liquid nitrogen prior to RNA extraction. Total RNA was extracted using
147 a commercial RNA extraction kit according to the manufacturer's instructions. RNA
148 concentration and purity were assessed using a NanoDropTM spectrophotometer, and
149 RNA integrity was confirmed by agarose gel electrophoresis. Only high-quality RNA
150 samples with clear 28S and 18S rRNA bands were used for subsequent experiments.



Gene-specific primers for *PCSK2* and *GAPDH* were designed using Primer5.0 software and validated using NCBI Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). Primer specificity was further confirmed by agarose gel electrophoresis prior to qPCR analysis. Primer sequences were as follows: PCSK2-F:GGGACGACGACTCCAAGG; PCSK2-R:TGCCCAGGATGCTCTTCA; GAPDH-F:GCACAGTCAAGGCGGAGA; GAPDH-R:GGCAGAAGGGGCAGAGAT.

The first strand cDNA synthesis was performed using BeyoRT™ II cDNA Kit (Beyotime Biotechnology, Shanghai, China) according to the manufacturer's instructions. Quantitative real-time PCR was performed using the CFX96™ Real-Time System (Bio-Rad Laboratories, Hercules, CA, USA) with SYBR Green qPCR Master Mix Kit (Sangon Biotech (Shanghai) Co., Ltd., Shanghai, China). Thermal cycling conditions were: 95°C for 5 min, followed by 40 cycles of 95°C for 10 s and 60°C for 30 s. Three biological replicates and three technical replicates were set for each tissue. Relative expression levels were calculated using the $2^{-\Delta\Delta C_t}$ method (Livak & Schmittgen, 2001) with *GAPDH* as the internal reference gene and brain tissue as the calibrator. Data were expressed as mean $\pm$ SD of technical replicates.

Dual-Luciferase reporter assay

HEK293T cells were cultured to 70–80% confluency, then digested with trypsin. Digestion was terminated using complete medium, and the cell suspension was gently



173 pipetted to obtain a single-cell suspension. Cells were counted using a hemocytometer
174 and diluted to a final concentration of 1×10^5 cells/mL in complete medium. A total of
175 1 mL of the cell suspension was seeded into each well of a 24-well plate. The plate was
176 gently rocked in both directions to ensure even distribution and incubated at 37°C in a
177 5% CO₂ incubator for 16 hours, reaching 40–50% confluency prior to transfection.
178 Plasmid transfection was performed using Lipofectamine® 2000 (Thermo Fisher
179 Scientific, Cat# 11668027) following the manufacturer's protocol. For each well, the
180 culture medium was replaced with 1 mL of Opti-MEM, and cells were returned to the
181 incubator. A total of 1 µg of plasmid DNA (pGL3 firefly luciferase reporter and pRL-
182 TK Renilla control plasmid at a 5:1 ratio) was diluted in 50 µL Opti-MEM. Separately,
183 2 µL of Lipofectamine 2000 was diluted in another 50 µL Opti-MEM. The two
184 solutions were combined, incubated at room temperature for 20 minutes, and added
185 dropwise to the corresponding wells. After 4–6 hours of incubation, the transfection
186 medium was replaced with 1 mL of complete medium, and cells were cultured for an
187 additional 48 hours.

188
189 Luciferase activity was measured using the Dual-Luciferase® Reporter Assay System
190 (Promega, Cat# E1910) according to the manufacturer's instructions. The 5× Passive
191 Lysis Buffer (PLB) was diluted to 1× with nuclease-free water. Culture medium was
192 removed, and cells were washed once with PBS. Then, 100 µL of 1× PLB was added
193 to each well, and the plate was gently agitated at room temperature for 15 minutes.
194 Lysates were mixed thoroughly, and 20 µL of each lysate was transferred to a 96-well



plate in triplicate. Firefly luciferase activity was measured by adding 100 μ L of Luciferase Assay Reagent II (LAR II), followed by measurement of Renilla luciferase activity after adding 100 μ L of Stop & Glo® Reagent. Firefly luciferase signals were normalized to Renilla luciferase signals for data analysis.

In situ Hi-C library preparation and sequencing

Brain and liver tissues were selected for in situ Hi-C experiments following the protocol described by Rao et al (Rao et al., 2014). with minor modifications. Briefly, sample was crosslinked in 2% formaldehyde for 15 minutes at room temperature, and the reaction was quenched with glycine. Crosslinked samples were subsequently ground in liquid nitrogen, and nuclei were isolated through a series of centrifugation steps: first by filtration at 4,000 rpm for 20 minutes at 4°C, followed by two rounds of centrifugation at 14,000 rpm for 10 minutes at 4°C. The extracted chromatin was solubilized using dilute SDS and incubated at 65°C for 10 minutes. Triton X-100 was added to quench the SDS before digesting the chromatin overnight at 37°C with 400 units of MboI restriction enzyme (NEB). The resulting DNA fragment ends were labeled with biotin-14-dCTP (Invitrogen, 19518-018) and then proximity-ligated to form chimeric junctions. Crosslinks were reversed by incubation with proteinase K (Invitrogen, 25530-031) at 65°C, followed by DNA extraction using phenol-chloroform. To remove biotin from non-ligated ends, T4 DNA polymerase (NEB, M0203L) was applied. The DNA was then fragmented to 100–500 bp using sonication. Fragment end repair was performed using T4 DNA polymerase, T4 polynucleotide kinase (NEB,



M0201), and Klenow DNA polymerase. Biotinylated ligation products were selectively enriched using streptavidin magnetic beads. End repair and A-tailing were carried out with Klenow fragment (NEB, M0212L), and sequencing adapters (Illumina, PN 1001782 and 1001783) were ligated to the DNA. Hi-C libraries were constructed following Illumina protocols and sequenced using paired-end 150 bp reads on the Illumina HiSeq platform.

scRNA-seq and scATAC-seq library preparation and sequencing

scRNA-seq libraries were prepared following established protocols with minor modifications (Rao et al., 2023). Briefly, approximately 100 mg of pig thymic tissue was rinsed with phosphate buffered saline (PBS). Tissue dissociation was performed using the dissociation kit (mouse, Miltenyi, Germany) at 37 °C. The resulting cell suspension was filtered through a 70 µm cell strainer (Miltenyi, Germany) and centrifuged at $300 \times g$ for 5 minutes at 4 °C. After red blood cell lysis (Miltenyi, Germany), cell concentration and viability were assessed using a fluorescence-based cell analyzer (Countstar Rigel S2, China) with acridine orange/propidium iodide staining. Cells were then washed twice with RPMI 1640, resuspended at a final concentration of 1×10^6 cells/mL in RPMI 1640 supplemented with 0.04% bovine serum albumin (BSA). ScRNA-seq libraries were prepared using SeekOne® MM Single Cell 3' library preparation kit (SeekGene, China) and sequenced on illumina Nova Seq 6000 platform (USA).



For scATAC-seq, nuclei isolation and library preparation were performed following the 10x Genomics protocol (<https://support.10xgenomics.com/single-cell-atac>). Nuclei quantified and diluted to a concentration of 2,000-4,000 nuclei/ μ L. Approximately 10,000 nuclei were used for transposition and capture on the Chromium Controller (10x Genomics, USA). Library were prepared using the Chromium Single Cell ATAC Reagent Kit according to the manufacturer's instructions. Final scATAC-seq libraries were sequenced on an Illumina HiSeq 2000 platform (USA) using 150 bp paired-end reads.

CUT&Tag data analysis

The raw data were analyzed utilizing pipelines developed by the Henikoff laboratory (Henikoff et al., 2020), with minor modifications. In brief, paired-end reads containing adapters, ploy-N, and low-quality reads were trimmed using Trim Galore v0.6.10 (available at <https://github.com/FelixKrueger/TrimGalore>) with default parameters. High quality clean reads were then aligned to the porcine reference genome (self-assembled by the lab) using Bowtie2 v2.4.1 (Langmead & Salzberg, 2012). To removed potential PCR duplicates, the Picard toolkit v1.119 (<http://broadinstitute.github.io/picard/>) was employed. Reads with low mapping quality (mapq < 30) and mitochondrial DNA were filtered out using SAMtools v1.9 (Li et al., 2009). Bigwig files were generated with deepTools bamCoverage function (Ramírez et al., 2016). For peak calling, stringent parameters (--qvalue 0.01 --nomodel --shift 0 -B -SPMR --keep-dup all) were applied using MACS2 v2.12 (Zhang et al., 2008), with



high-quality peak regions defined by p-values < 10⁻⁵.

Identification of enhancers and promoters

A list of overlapping peaks across all H3K27ac CUT&tag data was generated using the BEDTools merge command (Quinlan & Hall, 2010). Merged peaks located more than 1 kb from transcript start sites (TSSs) of transcripts were defined as candidate enhancers. We defined putative promoters as merged peaks within 1 kb of TSSs, identified from the combined H3K27ac and H3K4me3 CUT&Tag datasets. Read coverage for peaks was computed using deepTools multiBamSummary (Ramírez et al., 2016), while peak intensity was measured as Peaks Per Kilobase Million using R program v3.5.1 (<https://www.r-project.org/>). Sample clustering based on enhancer activity profiles was performed using the hclust function with the ward.D method and visualized using the circlize (Gu et al., 2014) and dendextend packages (Galili, 2015).

To assess tissue specificity, the Tau score was calculated using the following formula:

$$\tau \text{ score} = \frac{\sum_{i=1}^N (1 - \hat{x}_i)}{N-1}, \text{ where } \hat{x}_i = \frac{x_i}{\max(x_i)}$$

Here, x_i is the average normalized enhancer intensity in tissue i , and N is the number of tissues.

Hi-C map construction and loop identification

Hi-C reads were mapped to the pig reference genome using the juicer v1.6 pipeline (Durand et al., 2016) with modified parameters “-s MboI -S early” and other setting



were kept at their defaults. Mapped chimeric reads were filtered to retain only high-confidence interactions with MAPQ > 30. Hi-C contact matrices were generated in .hic format using juicer tools v1.22.01 at a resolution of 5 kb. Chromatin loops were identified using the HiCCUPS algorithm within Juicer Tools with the resolution parameter set to -r 5000; all other parameters remained at default settings.

Data reproducibility

To assess reproducibility, we collected liver-matched tissue samples from six anatomical sites in three additional biological replicates from F12-generation individuals (see Supplementary Table S3 for details). The same experimental protocols and analytical procedures used in the primary study were applied. Read coverage across genomic regions in the filtered BAM files was calculated to using multiBamSummary bins function in deepTools v2.0 (Ramírez et al., 2016) to assess genome-wide similarity among replicates.

Build enhancer regulatory networks

To explore the underlying mechanism of enhancer clusters in regulating gene expression, we performed eNet algorithm (Hong et al., 2023) with minor modifications to build enhancer networks based on H3K27ac and H3K4me3 CUT&Tag datasets. Briefly, Spearman correlations were computed between enhancer intensity and promoter intensity. Promoters within 150 kb of each enhancer were included, and correlations with a coefficient ≥ 0.3 , and q-value < 0.01, where the enhancer and

promoter existed in the same sample, were considered significant. For each gene, each putative enhancer pair within the enhancer cluster were performed co-accessibility. Then significantly correlated enhancers were predicted enhancer-enhancer interactions (EEIs). Next, we built enhancer networks for each gene, where nodes represent enhancers in the enhancer network and edges represent EEIs. Because the resulting network was relatively dense, we further retained only the top 5% strongest EEIs (95th percentile cutoff) to reduce noise from weaker associations and improve the stringency of downstream network analysis. This cutoff was applied during edge filtering and was not used to directly define hub enhancers. We then quantified the complexity of the enhancer networks by the network size and connectivity. Network size was calculated as the number of enhancers. Network connectivity was calculated as two-fold of the number of edges divided by the number of nodes. We classified the enhancer networks into three modes based on their complexity: Simple ($<$ median of nodes), Multiple ($>$ median of nodes and network connectivity < 1), Complex ($>$ median of nodes and network connectivity > 1).

Hub enhancers for tissue-specific promoters were identified only from Complex networks. For each complex network, the degree of each enhancer node was calculated and normalized by the total number of edges in the network. Network hub enhancers were defined as enhancers showing high normalized interaction degree within the corresponding complex network.



Validation of enhancer–promoter interactions within the same TADs

Based on the Hi-C data generated from brain, liver, and muscle tissues (Li et al., 2025), we identified TADs for each tissue. To evaluate the spatial validity of EPIs, both enhancers and promoters involved in EPIs were intersected with TAD coordinates using bedtools intersect (Quinlan & Hall, 2010). An interaction was considered TAD-supported when both the enhancer and promoter were assigned to the same TAD. The proportion of TAD-supported interactions was calculated as the number of EPIs with both anchors located within the same TAD divided by the total number of EPIs evaluated in the corresponding tissue.

Function annotation and visualization

GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analysis were conducted using DAVID (<https://david.ncifcrf.gov/tools.jsp>). To visualize continuous CUT&Tag signal, BigWig files normalized to 1× genome coverage in 20 bp bins were created using deepTools bamCoverage (Ramírez et al., 2016), and visualized with pyGenomeTracks (Lopez-Delisle et al., 2021).

Enrichment analysis of housekeeping and disease genes

For a given gene set, the enrichment score was defined as the fold change relative to the genomic baseline, calculated by the formula:

$$(m/n) / (M/N),$$

where m and M indicate the number of target genes within the gene set and across the

entire genome, respectively, while n and N represent the total number of genes in the set and genome-wide, respectively. All housekeeping genes were downloaded from <https://housekeeping.unicamp.br/?homePageGlobal>. Enhancers from Vista Enhancer browser database <https://enhancer.lbl.gov/vista/downloads> were converted to pig reference genome using the LiftOver tool (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>). Brain disease genes were collected from <https://download.big.ac.cn/brainbase/Brain%20Diseases/>.

To evaluate the performance of enhancer networks in predicting the housekeeping genes, we ranked all genes by two scoring methods, including network connectivity and network size. We then calculated the fold-enrichment of housekeeping genes in top-ranked genes with a moving window of 50.

TF motif enrichment analysis

Motif occurrence within enhancers from three modes was predicted using HOMER (Heinz et al., 2010) with default settings to predict motif occurrences within enhancers from the three regulatory modes. TFs with motif enrichment P -values < 0.01 were extracted to generate a list of enriched TFs. To characterize the distinctive features of enhancers across the three modes, TFs with $-\log_{10}(P\text{-value}) > 2$ were visualized.

CTCF enrichment analysis in hub enhancers

To assess the enrichment of CTCF binding within hub enhancers, motif scanning was



conducted against the CTCF motifs from the non-redundant JASPAR2022 CORE vertebrates motif database using FIMO (Grant et al., 2011) with the default settings. To evaluate CTCF enrichment, we conducted a Fisher's exact test comparing the frequency of CTCF-bound enhancers in hub versus non-hub enhancer sets.

Identification of super-enhancers and construction of core regulatory networks

SEs were identified using H3K27ac data with ROSE algorithm (https://bitbucket.org/young_computation/rose). SEs from all samples were then merged to obtain consensus super-enhancers. The normalization process followed the same steps as those used for typical enhancers.

To identify the CRCs, *CRC mapper* (Saint-André et al., 2016) was used with sequence alignment data. SEs were identified using H3K27ac data with ROSE algorithm (https://bitbucket.org/young_computation/rose). Then, screening the closest active genes that are TFs and find the TFs regulated by SEs. TF motifs were identified with FIMO (Grant et al., 2011) using default parameters. Finally, a TF was defined as a self-regulatory TF when it could bind to an SE element that regulates the TF itself. Within the CRC, each TF binds its own SE as well as the SEs of the other TFs in the circuitry, collectively driving an extended catalog of SEs.

ScRNA-seq data processing

Raw FASTQ files were aligned to the porcine reference genome (self-assembled by the



lab) using cellranger-7.0.0 (<https://github.com/10XGenomics/cellranger>). The resulting output matrices were imported into Seurat (v 4.3.0.1) (Hao et al., 2021) for quality control. Overall, 10,515 cells passed the filtering criteria, including all genes expressed in less than three cells, the number of genes expressed per cell was greater than 200 and less than 7,500, the UMI counts were less than 50,000, the percent of mitochondrial-DNA derived gene expression was less than 5%. To visualize cellular heterogeneity, we regressed out the effects of total expression and mitochondrial content using Scanpy version 1.7.1 (Wolf et al., 2018) with Python version 3.7.4. Dimensionality reduction was performed using principal component analysis (PCA), followed by uniform manifold approximation and projection (UMAP) for embedding into two-dimensional space. Cell-type annotation was achieved by identifying marker genes via the Wilcoxon rank-sum test and selecting the top 500 differentially expressed genes, which were cross-referenced with canonical markers of thymic cell types.

ScATAC-seq processing

Raw scATAC-seq reads were processed using the Cell Ranger ATAC pipeline (v2.0.0, 10x Genomics). Accessibility profiles were generated for each library using the cellranger-atac count command. Downstream analyses were performed with the ArchR package (v1.0.1) (Granja et al., 2021). During quality control, nuclei with TSS enrichment scores < 4 or fewer than 1,000 unique fragments were excluded. Additionally, cells with elevated doublet scores were filtered out. Dimensionality reduction was performed using latent semantic indexing (LSI) via the addIterativeLSI



function, followed by the construction of a nearest-neighbor graph for clustering and UMAP visualization. Gene activity scores were computed using the addGeneScoreMatrix function, and cluster-specific marker genes were identified with getMarkerFeatures. To evaluate peak-calling reproducibility, pseudobulk replicates were generated, and peaks were identified using MACS2 v2.1.2 (Zhang et al., 2008).

Integration of scRNA-seq data and scATAC-seq data

To annotate scATAC-seq profiles, we integrated the scATAC-seq dataset with the matched scRNA-seq dataset using the addGeneIntegrationMatrix() function in ArchR (v1.0.1) (Granja et al., 2021). Briefly, gene activity scores from scATAC-seq cells, represented by the GeneScoreMatrix, were used as the query features, and the annotated scRNA-seq Seurat object was used as the reference. Cell-type labels in the scRNA-seq dataset were obtained from the metadata field cell_type, which had been assigned based on canonical marker genes.

Peak-to-gene linkage analysis was performed using addPeak2GeneLinks() in ArchR (v1.0.1) (Granja et al., 2021) with reducedDims = "IterativeLSI" and useMatrix = "GeneScoreMatrix". A permissive correlation cutoff of corCutOff = 0.2 was used to detect candidate regulatory associations. Peak-to-gene links were extracted with getPeak2GeneLinks() and used for downstream visualization and locus-level interpretation.

For visual validation of the integrated data, browser-style tracks were generated with `plotBrowserTrack()` using either user-defined genomic regions or marker genes of interest. Pseudobulk accessibility tracks grouped by `predictedGroup`, together with inferred peak-to-gene links, were used to visually examine whether known marker loci showed expected cell-type-specific accessibility patterns.

Integration analysis between single-cell and bulk data

To compare single-cell regulatory elements with hub enhancers identified from bulk thymus data, accessible chromatin regions derived from scATAC-seq were intersected with bulk-defined hub enhancers using `bedtools intersect` command (Quinlan & Hall, 2010). Overlap was defined on the basis of genomic coordinate intersection between the two sets of regions.

To further assess cell-type specificity, Tau scores were calculated for bulk-defined hub enhancers based on their accessibility across scATAC-seq cell types. Track visualization was used to compare the accessibility landscape surrounding representative loci and to examine the relationship between bulk-defined hub enhancers and cell-type-resolved accessible elements in single-cell data using the ‘`plotBrowserTrack`’ function (Granja et al., 2021).

GWAS enrichment analysis

To investigate the enrichment of GWAS signals within hub enhancers and super-



enhancers, we retrieved QTL data for 133 traits available in the PigBiobank database (<https://pigbiobank.ipiginc.com/home>). The enrichment calculation was structured as follows: $\text{Odd Ratio} = A/B$, where A represents the number of lead SNPs within target regions / total number of lead SNPs and B represents the number of background SNPs within target regions / Total number of background SNPs. A fold change value greater than 1 indicated that trait-associated SNPs were specifically enriched in H3K27ac-marked regions, suggesting a significant association between these regions and trait-linked regulatory activity. For each odds ratio (OR), statistical significance was assessed using Fisher's exact test.

RESULTS

Data Summary

To systematically characterize enhancer regulatory architecture across pig tissues, we generated a comprehensive CUT&Tag dataset profiling both H3K27ac and H3K4me3 in 151 anatomical sites from 58 distinct tissues. The samples were collected from a castrated male pig together with seven reproductive sites from its full-sibling female, yielding broad anatomical coverage across the pig body. Across all samples, we identified an average of 34,812 H3K4me3 peaks and 72,983 H3K27ac peaks per sample. After merging peaks across samples, we obtained 222,290 consensus H3K4me3 peaks (mean length = 1,013 bp) and 486,064 consensus H3K27ac peaks (mean length = 1,088 bp). All samples passed stringent quality-control criteria, including FRiP, NSC, and RSC, supporting the overall robustness of the chromatin profiles (Supplementary

Tables S1, S2).

To further assess reproducibility, we generated additional H3K27ac profiles from liver samples collected from three independent pigs across six anatomical sites (Supplementary Table S3). These validation samples clustered closely with the corresponding liver samples generated in this study, showed high pairwise correlations, and remained clearly separated from other tissue types (Supplementary Figure S1A). In parallel, clustering based on peak activity across all samples grouped tissues according to their biological identities (Supplementary Figure S1B), further supporting the reliability and tissue specificity of the regulatory profiles.

By integrating H3K4me3 and H3K27ac signals across all 151 samples, we annotated 36,857 promoters in transcription start site (TSS)-proximal regions and 448,648 distal enhancers located more than 1 kb from annotated TSSs. This 1 kb threshold was used as an operational criterion to distinguish promoter-proximal from distal regulatory elements, consistent with previous epigenomic studies (Jiang et al., 2022; Zhu et al., 2022). In our dataset, both H3K4me3 and H3K27ac signals were strongly enriched within 1 kb of annotated TSSs (Supplementary Figure S1C), supporting the suitability of this cutoff for separating promoter-associated regions from distal putative enhancers.

Complexity of the whole-genome enhancer networks across 58 pig tissues

We next applied the eNet framework (Hong et al., 2023) to infer enhancer-centered



regulatory networks across tissues. To define an appropriate search space for enhancer–promoter interactions (EPIs), we generated Hi-C data from brain and liver and integrated publicly available muscle Hi-C data (Li et al., 2025). Topologically associating domain (TAD) analysis revealed a median size of approximately 280 kb (Supplementary Figure S1D), suggesting that a 150-kb window around each enhancer represents a biologically plausible range for local cis-regulatory interactions. Sensitivity analyses across different enhancer–promoter distance thresholds showed that the proportion of positively correlated enhancer–promoter pairs located within the same TAD increased with decreasing genomic distance but plateaued between 100 and 150 kb (Supplementary Figure S1E). Consistently, positive enhancer–promoter correlations were enriched at shorter distances, whereas negative correlations were depleted (Supplementary Figure S1F). Together, these results supported the use of a 150-kb window for subsequent analyses.

Using this window, we inferred EPIs based on activity correlations (Spearman's $r \geq 0.3$, $q < 0.01$) (Ren et al., 2026), identifying 741,616 interactions, including 697,368 positive EPIs (94.03%), linking 273,855 enhancers to 34,158 promoters. Approximately 80% of positive EPIs were located within the same TAD, providing independent support from Hi-C data (Supplementary Figure S1E). For each promoter, linked enhancers were grouped into enhancer clusters, and enhancer–enhancer interactions (EEIs) were inferred to construct gene-centered enhancer networks, where nodes represent enhancers and edges represent EEIs (Figure 1). Network complexity



was quantified by network size (number of enhancers) and connectivity (twice the number of edges divided by the number of nodes). Following the eNet framework, we retained networks larger than the median size with connectivity > 1 and selected the top 5% strongest EEIs to construct high-confidence enhancer networks for downstream analyses (Figure 2A).

In total, we built 20,468 enhancer networks across pig tissues (Figure 2B). Among these, we observed several brain function-related genes, including *APBA2* (Da Costa et al., 2017), *OPCML* (Rougon & Hobert, 2003) and *SLAMF1* (Chen et al., 2020b), which exhibited notably large network size and high network connectivity (Figure 2B). The finding suggested that complex enhancer networks are more likely to regulate critical genes for tissue-specific functions, particularly in the brain, where intricate regulatory mechanisms may be required to support highly specialized neural processes. We then classified the enhancer networks into three distinct modes based on their regulatory complexity: Simple mode, multiple mode, and complex mode. This classification yielded 10,308 simple, 5,337 multiple and 4,751 complex enhancer networks (Figure 2B). For example, the *FGLI* gene (Chen et al., 2024), which was highly expressed in the liver and may be involved in liver regeneration and metabolic regulation, was regulated by a complex enhancer network consisting 53 EEIs among 25 enhancers. In contrast, the *LDHA* gene (Gray et al., 2014), which was widely expressed and involved in regulation of pyruvate metabolism, was controlled by a network with the same number of enhancers but only a single EEI (Figure 2C). This association was further



reproduced in an independent liver dataset. Specifically, using matched H3K27ac and RNA-seq data from 259 samples from our previous study (Ling et al., 2024). we reconstructed an enhancer–gene regulatory network and confirmed that genes embedded in more complex enhancer networks were expressed at significantly higher levels than genes associated with simple or less connected regulatory architectures (Supplementary Figure S1G). This comparison highlighted the diversity in enhancer network architectures, where genes with distinct biological functions may exhibit varying levels of regulatory complexity.

Enhancer network complexity is associated with tissue-specific gene regulation

We next examined the relationship between enhancer network complexity and tissue specificity. Using H3K27ac activity for enhancers, H3K4me3 activity for promoters, and matched RNA-seq data from pigGTEx (Fang et al., 2025) for genes, we calculated Tau scores across tissues. Tau scores of enhancers were positively correlated with those of their associated promoters and target genes (Supplementary Figures S2A, S2B), indicating coordinated tissue-specific regulation across multiple layers of the regulatory hierarchy. Overall, enhancers exhibited higher Tau scores than promoters and genes, consistent with distal enhancers carrying substantial tissue-specific regulatory information. However, among highly tissue-specific genes, promoter and gene Tau scores frequently approached or exceeded those of their associated enhancers, suggesting that tissue specificity may be further reinforced during signal transmission from enhancers to promoters and gene expression outputs (Supplementary Figures S2A,



S2B). Importantly, this relationship followed a clear hierarchical trend across regulatory architectures. As regulatory mode progressed from simple to multiple and ultimately to complex, Tau scores increased significantly for enhancers, promoters, and target genes (Supplementary Figures S2C, S2D). In simple networks, enhancers were generally more tissue-specific than their downstream genes. In contrast, under multiple and especially complex regulatory architectures, promoters and target genes often achieved tissue specificity comparable to or greater than that of their associated enhancers (Supplementary Figures S2C, S2D). These observations suggest that increasing enhancer network complexity is associated with progressively stronger tissue specificity and support a hierarchical regulatory model in which complex enhancer networks reinforce and concentrate tissue-specific regulatory signals at the promoter and transcriptional levels.

To further characterize the regulatory properties of the three enhancer network modes, we compared their associated promoters in terms of tissue specificity, housekeeping gene enrichment, and TF-binding features. Promoters in the complex mode exhibited significantly higher Tau scores (a tissue specificity index) (Yanai et al., 2005) (Figure 2D), indicating that genes associated with the complex mode were more tissue-specific than those in the other two modes. Consistent with this, housekeeping genes were depleted in the complex mode but enriched in the simple mode when assessed against the whole genome as background (Figure 2E). Furthermore, housekeeping genes showed significantly lower enrichment when genes were ranked by network

connectivity than when ranked by network size (Figure 2F), suggesting that the complexity of enhancer–promoter connectivity, rather than the mere number of network nodes, more strongly reflects tissue-specific regulatory programs. We next compared TF-binding motifs among promoters across the three modes and found that promoters in the simple mode were enriched for motifs of general TFs, including YY1 and SP1 (Figure 2G). Collectively, these results suggest that tissue-specific gene regulation is preferentially mediated by complex enhancer networks.

Tissue-specific enhancers enriched in complex enhancer networks

To further explore the tissue specificity of complex enhancer networks, we focused on promoters with tissue-specific activity (Smith et al., 2007). Using a Tau score threshold of 0.85, we identified 6,902 tissue-specific promoters across 58 tissues and linked them to 1,535 complex enhancer networks (Supplementary Figure S3A). Notably, nervous system-related tissues harbored a greater number of regulatory networks, consistent with their high transcriptional and regulatory complexity (Supplementary Figures S3B, S3C). We then examined whether these networks were associated with tissue-specific functions across tissues. Using brain regions (34/151 samples) as an example, we identified 942 promoters with highly specific activity, including 12 promoters corresponding to five brain function-related genes (*APBA2*, *DPP6*, *FAM135B*, *NTM* and *OPCML*) represented among the top 10 complex enhancer networks (Supplementary Figures S4A, S4B). Consistent with this, genes associated with the complex mode in brain regions showed the strongest enrichment for brain disease-



related genes, with a 2.59-fold enrichment compared to the whole genome as the background (Supplementary Figure S4C). We next examined the constituent enhancers within these networks and found that they also exhibited marked tissue specificity, with approximately 30% showing Tau scores greater than 0.8 (Supplementary Figure S4D). Applying a stricter threshold ($\text{Tau} > 0.85$), we identified 3,132 tissue-specific enhancers, which were enriched for tissue-specific TFs and associated with highly tissue-specific functions (Figure 3A; Supplementary Tables S4, S5). Collectively, these results suggest that tissue-specific enhancers play important roles in shaping complex enhancer networks.

Identification and characterization of hub enhancers within complex networks

Enhancer clusters are organized hierarchically, comprising both necessary and dispensable components that coordinate gene expression regulation (Huang et al., 2018a). To analyze hierarchically organized enhancers in complex enhancer networks associated with tissue-specific promoters, we defined network hub enhancers as those with high-frequency EEIs, as previously described (Hong et al., 2023). In total, we identified 1,511 hub enhancers linked to 629 core genes (Figure 3B; Supplementary Table S6). To investigate the functional properties of hub enhancers, we first compared the enrichment of CTCF binding motifs between hub and non-hub enhancers. Hub enhancers exhibited significantly higher CTCF motif enrichment than non-hub enhancers (2.21-fold, $P = 1.42\text{E-}23$, binomial test; Figure 3C). Additionally, hub enhancers were more strongly enriched in experimentally validated VISTA enhancers

635 than non-hub enhancers (7.45-fold, $P = 8.16\text{E-}10$, binomial test; Figure 3D).
636 Permutation analyses (1,000 iterations) confirmed that hub enhancers exhibited
637 significantly higher Tau score than expected by chance (empirical $P < 0.001$).
638 Consistently, genes linked to hub enhancers also displayed elevated Tau score relative
639 to randomly sampled genes (empirical $P < 0.001$; Supplementary Figures S4E, S4F). A
640 representative example was observed at the *PCSK2* locus, a neuronal marker gene
641 (Ferguson et al., 2019) (Figure 3A). We constructed *PCSK2*-associated enhancer
642 networks across multiple brain regions, including the brain, brain stem, cerebellum,
643 pituitary, thalamus and spinal cord. Comparative analysis revealed dynamic changes in
644 enhancer composition and network connectivity across these regions (Figure 3E).
645 Notably, a 16-enhancer network was specifically established in the cerebellum and
646 brain, consistent with elevated enhancer activity in these regions (Figure 3F). Within
647 this network, enhancer 16 acted as a hub enhancer that coordinated network interactions
648 and contributed to network formation (Figure 3E). We used dual-luciferase reporter
649 assays to test the regulatory activity of the predicted hub enhancer and performed qPCR
650 analysis across five tissues to examine the expression pattern of its target gene. These
651 experiments showed that *PCSK2* is specifically highly expressed in the brain and that
652 the hub enhancer identified in this study significantly increased reporter activity
653 (Figures 3G, 3H; Supplementary Figures S4G, S4H; Supplementary Table S7),
654 supporting its regulatory effect on *PCSK2* expression. Collectively, these findings
655 indicate that hub enhancers represent central regulatory elements within complex
656 enhancer networks and are enriched for CTCF-binding motifs, experimentally validated

enhancers, and elevated tissue-specific activity.

Genome-wide identification and characterization of super-enhancers

SEs, a distinct category of enhancer clusters, are characterized by high-density H3K27ac signals and can rapidly regulate target genes in response to external stimuli. Using the ROSE algorithm (Whyte et al., 2013), we stitched nearby enhancers (<12.5 kb) and ranked regions by average H3K27ac signal, defining SEs as regions with higher signal (slope >1) and the remainder as typical enhancers (Supplementary Figure S5A). Across samples, each contained ~1,500 SEs (median = 1,563), with a median size of 43.6 kb and typically encompassing eight constituent enhancers (Supplementary Figures S5B–D). Compared to typical enhancers, SEs exhibited stronger H3K27ac enrichment (Supplementary Figure S5E). Merging SEs from all 151 samples yielded 7,448 non-redundant SEs, most of which were present in only a few or single samples (Supplementary Figure S5F), reflecting their high tissue specificity (Pott & Lieb, 2015). To further explore specificity of SEs, we performed principal component analysis (PCA) using the Jaccard distance to assess SE location dissimilarity across all 151 samples. The PCA demonstrated clear separation among tissue types (Supplementary Figure S5G), which was corroborated by hierarchical clustering of SE occurrence across samples (Supplementary Figure S5H). Collectively, these results highlight the highly tissue-specific distribution of SEs.

Characterization of core regulatory circuitries based on super-enhancers

SEs preferentially recruit master TFs and play a central role in maintaining tissue-specific transcriptional programs. Therefore, constructing SE-driven CRCs enables systematic identification of key TFs and their cooperative interactions that define tissue identity in pigs (Almeida et al., 2021; Jung et al., 2024; Li et al., 2021). Using CRC *mapper* (Saint-André et al., 2016), we systematically constructed SE-driven CRCs across 151 pig tissue samples. In total, we identified 477 SEs (ranging from 14 to 173, mean = 74) associated with 368 core TFs (ranging from 13 to 156, mean = 58) (Figures 4A, 4B; Supplementary Table S8). Analysis of core TF occurrence patterns revealed marked tissue specificity. Approximately 36% of core TFs were detected in fewer than five samples, nearly 17% in only one sample, and ~11% in more than half of all samples (Figure 4C; Supplementary Figure S6). Representative examples include pituitary-specific POU1F1 and multi-tissue FOS (Figures 4D, 4E). At the tissue level, ~18% of core TFs were restricted to a single tissue (Figure 4C). Thus, tissue-restricted TFs constitute the predominant component of the CRC-associated TF repertoire. Notably, TFs with wider sample distribution tend to occupy more super-enhancers overall (Figure 4F). These results suggested that common TFs tended to perform similar regulatory functions in different samples by binding to more common SEs, while specific TFs play specific roles in maintaining tissue identity, which can be achieved by binding to the relatively few tissue-specific SEs (Feng et al., 2023).

Core regulatory circuitries for tissue-specific functions

Previous study has revealed that core TFs exhibited high total degrees of connectivity



within their respective tissues (Boyer et al., 2005). Consistent with this observation, we found that brain and lymph node ranked among the top two tissues in terms of total network connectivity (Figure 5A), suggesting highly interconnected regulatory architectures in these tissues. This may reflect the elevated regulatory complexity required for neural information processing and immune cell coordination, both of which rely on finely tuned transcriptional control. Motivated by this pattern, we next focused on core transcription factors in the pituitary. We observed that PITX1, PITX2 and GLI2 as core TFs in the pituitary (Figures 5B, 5C). Functionally, *PITX1* was known to play a crucial role in pituitary development (Suh et al., 2002), *PITX2* was required at multiple stages of pituitary organogenesis (Suh et al., 2002), and truncating and Zn-finger variants in *GLI2* were associated with a high risk of hypopituitarism (Arnhold et al., 2015; Corder et al., 2022). Additionally, their motifs were significantly enriched in SEs (Figure 5D). Further analysis revealed that these three TFs formed a feed-forward auto-regulatory circuitry and co-regulated a subset of other SEs, thereby establishing an extended regulatory network (Figure 5B). Intriguingly, pituitary-related TFs such as HES1 (Monahan et al., 2009), POU1F1 (Bando et al., 2023), ISL2 (Yan et al., 2023) and INSM1 (Welcker et al., 2013) were also integrated into this network. Moreover, the binding motifs of PITX1, PITX2 and GLI2 were significantly enriched near their respective gene loci (Figures 5E-G), further supporting the possibility of their roles as core regulators in the pituitary. Collectively, these results highlighted the importance of TF-mediated CRCs in tissue identity and underscored the role of core TFs in CRCs for tissue-specific functions.



Dissecting hub enhancers within super-enhancers mediating core regulatory circuitries

Inspired by the discovery that hub enhancers within enhancer networks play a pivotal role in gene regulation, we systematically characterized the structural organization of SEs within CRCs. To this end, we analyzed CTCF binding sites (CBSs) within enhancers contained in SEs of CRCs, retaining only the enhancer with the highest number of CBSs for each SE. This approach identified 535 hub enhancers corresponding to 473 SEs ([Supplementary Table S9](#)). Interestingly, the hub enhancers identified via CRCs showed limited overlap with those identified through enhancer network analyses: only eight hub enhancers and 21 core genes were shared between the two approaches (Figure 6A; [Supplementary Table S10](#)). [This discrepancy likely reflects the distinct principles underlying the two methods.](#) We further examined whether hub enhancers identified by enhancer networks were components of super-enhancers, and found that most of them were indeed SE constituents ([Supplementary Figure S7A](#)), highlighting the complementary nature of the two approaches. Specifically, the CRC-based method captures hub enhancers within regulatory circuitries driven by core TFs, whereas enhancer networks identify hub enhancers based on inter-enhancer interactions. Together, these strategies substantially expand the catalog of functional enhancer clusters and deepen our understanding of their organization and function.

[To evaluate evolutionary constraint, we next assessed sequence conservation of both](#)

hub enhancers and their associated genes. Using a permutation framework, we compared the median phastCons scores of the overlapping hub enhancers against a null distribution generated from 1,000 random samplings of other hub enhancers matched in number. The overlapping hub enhancers exhibited significantly higher conservation than expected by chance (empirical $P < 0.05$) (Supplementary Figure S7B), indicating strong evolutionary constraint. We then applied the same permutation-based approach to the corresponding core genes. The median conservation of the 21 shared genes was compared against randomly sampled sets of other core genes, and we observed a consistent pattern of elevated conservation relative to expectation (empirical $P < 1 \times 10^{-4}$) (Supplementary Figure S7C). These results suggest that both regulatory elements and their target genes within the overlapping hub modules are under strong purifying selection. Consistent with these findings, the overlapping hub enhancers also displayed strong functional activity signatures, including significant enrichment of H3K27ac and CTCF signals across all eight loci (Figure 6B; Supplementary Figures S7D–S7J; Supplementary Table S11), suggesting robust enhancer activity and potential involvement in chromatin loop anchor formation.

To further characterize the structural and regulatory characteristics of overlapping hub enhancers, we selected a representative *SATB1*-associated locus for in-depth analysis. At this locus, a hub enhancer located at chr13:20,331,411–20,333,161 was identified as a putative regulatory element of *SATB1* (Figure 6C), a master regulator of thymus development (Naito et al., 2023). We then examined local histone modification patterns



and binding motifs. The promoter of *SATB1* was marked by both H3K27ac and H3K4me3, and co-localized with RAD21 and CTCF binding motifs (Figure 6D). In addition, CTCF and RAD21 binding motifs were also present within the hub enhancer, suggesting that it may function as an architectural element facilitating enhancer–promoter communication at the *SATB1* locus. Notably, these observations were further supported by Hi-C data generated from the same thymus sample, in which a clear chromatin loop structure connecting the hub enhancer and the *SATB1* promoter was detected (Figure 6D). We further compared the distribution of SATB1 binding profiles (Kitagawa et al., 2017), H3K27ac, H3K4me1, H3K4me3 (<https://www.encodeproject.org/>), and RAD21 (Seitan et al., 2013) at the orthologous mouse locus. Cross-species comparison revealed conserved RAD21 motifs at the *SATB1* promoter and strong enrichment of active histone modifications across the SE containing the hub enhancer (Figure 6E), supporting the functional conservation of this regulatory architecture. Collectively, these results suggested that hub enhancers were characterized by the presence of active enhancer-associated histone modifications and the binding of architectural proteins such as CTCF and RAD21, consistent with a potential role in organizing enhancer–promoter communication within tissue-specific regulatory circuits.

Single-cell chromatin accessibility reveals limited cell-type specificity of hub enhancers in the thymus

To investigate whether hub enhancers exhibit cell-type-specific accessibility, we



performed single-cell ATAC-seq (scATAC-seq) on pig thymus using the Chromium platform (10x Genomics). After stringent quality control, 10,126 high-quality cells were retained for downstream analysis (Supplementary Figures S8A, S8B). We identified a total of 173,994 peaks mapped to the nuclear genome, with a median of 16,416.5 fragments per cell. To infer cell-type identities, we calculated gene activity scores by summing the fragments within gene promoter and gene bodies, and then transferred cell-type labels from a reference scRNA-seq dataset. Guided by canonical markers and reference annotations from Gu et al (Gu et al., 2022), we annotated the major cell clusters into double-positive (DP) thymocytes, T regulatory (Treg) cells, proliferating cells, double-negative (DN) cells, single-positive (SP) cells, $\gamma\delta$ T cells and B cells (Supplementary Figures S8C-S8E).

We next sought to identify cell-type-specific cis-regulatory elements (cCREs). Using a Tau score threshold > 0.85 , we identified 11,938 cCREs with strong cell-type specificity (Figure 7A). Notably, none of these cCREs overlapped with the hub enhancers previously defined in bulk thymus. This lack of overlap suggests that hub enhancers identified from bulk tissue may not correspond to highly cell-type-restricted regulatory elements. Consistent with this interpretation, the Tau scores of bulk-defined hub enhancers were generally centered around 0.5 (Figures 7B, 7C), indicating broad accessibility across cell types rather than strong cell-type specificity. Chromatin accessibility profiles further supported this observation, as most hub enhancers exhibited robust signals across multiple thymic cell populations (Supplementary



Figures S9A, S9B). Together, these results suggest that bulk-derived hub enhancers may preferentially capture regulatory elements shared across major thymic cell populations, rather than cCREs restricted to a single cell type. The lack of overlap may also reflect the sparse nature of scATAC-seq data (Preissl et al., 2023), as well as the difference between hub enhancers defined by network interaction properties in bulk tissue and cCREs defined by cell-type-specific chromatin accessibility.

To further examine this pattern at a representative locus, we focused on *SATB1*, a gene associated with a shared hub enhancer identified by both bulk-based approaches in thymus. Although the bulk-defined *SATB1* hub enhancer was not recovered through peak-to-gene linkage in the single-cell analysis, accessible regulatory elements at the *SATB1* locus remained detectable and showed concordant regulatory associations (Figure 7D). This discrepancy may reflect differences in signal aggregation, linkage sensitivity, and cell-type composition between bulk and single-cell datasets. It may also indicate that bulk-defined hub enhancers function as shared regulatory elements across multiple thymic cell types, rather than as highly cell-type-specific enhancers, consistent with their broadly distributed accessibility profiles and intermediate Tau scores. In addition, the inherent sparsity and limited sensitivity of single-cell ATAC-seq may hinder the detection of, or peak-to-gene linkage for, some functionally important enhancers with low or transient accessibility.

Taken together, these findings suggest that hub enhancers identified from bulk-tissue

enhancer networks are more likely to represent broadly accessible regulatory elements with shared functions across thymic cell lineages than strongly cell-type-specific enhancers.

Enrichment of GWAS variants in hub enhancers and super-enhancers

Hub enhancers and super-enhancers have been proposed as key regulatory elements that coordinate cell identity and complex transcriptional programs (Boeva et al., 2017; Hong et al., 2023). Given their central positions within enhancer regulatory networks, we hypothesized that these elements would preferentially harbor genetic variants underlying complex traits.

To test this hypothesis, we integrated publicly available pig GWAS signals spanning 133 complex traits (<https://pigbiobank.ipiginc.com/home>) with the hub enhancers and super-enhancers identified in this study (Supplementary Table S12). Both hub enhancers and super-enhancers exhibited significant enrichment of GWAS variants (Supplementary Figures S10A, S10B), indicating that trait-associated variants are preferentially concentrated within highly connected and regulatory-dense enhancer elements.

Notably, phenotypic categories related to health, meat and carcass traits, as well as production and reproduction, were all significantly enriched, with reproductive and production phenotypes showing the highest enrichment in super-enhancers (SEs), such as the number of stillborn pigs (S_NBS_DRP), number born alive (D_NBA), and



average daily gain from 30 to 115 kg (S_ADG_30T115). Similarly, hub enhancers were significantly enriched for the number of healthy piglets born (Y_NBH), sperm count (S_SPCOUNT_DRP), and average daily gain (Y_ADG). These observations raise the possibility that SEs and hub enhancers might act as important regulatory hubs underlying complex economic traits in pigs. The preferential enrichment of reproductive and growth traits in SEs is consistent with a model in which these phenotypes are modulated by coordinated enhancer clusters, potentially contributing to canalization against environmental fluctuations. In contrast, the selective enrichment of traits such as healthy litter size and sperm count in hub enhancers suggests a potentially more concentrated regulatory architecture, wherein a limited set of enhancers may integrate multiple signaling pathways. From a breeding perspective, these results nominate specific non-coding genomic regions as priority candidates for marker-assisted selection or gene-editing strategies, particularly for traits with low heritability or that are challenging to measure directly. Functionally, the identified SEs and hub enhancers provide a starting point for dissecting the molecular basis of pleiotropic trade-offs between reproduction and growth, as well as for investigating the genetic determinants of neonatal survival and male fertility.

DISCUSSION

Enhancers frequently function in clusters to ensure robust and precise control of gene expression during tissue specification, often through additive, synergistic, or redundant mechanisms (Blayney et al., 2023; Boeva et al., 2017; Hong et al., 2008; Parker et al.,



2013). However, the hierarchical organization and regulatory logic of enhancer clusters remain poorly understood, particularly in pigs. By systematically reconstructing enhancer-centered regulatory architectures across diverse pig tissues, this study provides a broad framework for understanding how coordinated enhancer activity contributes to tissue-specific gene regulation.

A major strength of this work lies in its genetically matched sampling design, which minimized confounding effects introduced by inter-individual genetic variation and enabled cleaner resolution of tissue-specific regulatory hierarchies. Unlike population-based cross-tissue studies, in which differences in regulatory landscapes may reflect both tissue identity and genetic heterogeneity (Whitehead & Crawford, 2005), our design allowed enhancer network variation to be interpreted primarily in the context of tissue-specific regulatory programs. Together with broad tissue coverage, this strategy provided a robust foundation for comparative dissection of enhancer organization across the pig body.

Within this framework, two complementary layers of enhancer-mediated regulation emerged. First, enhancer regulatory networks reconstructed by the eNet algorithm (Hong et al., 2023), captured widespread tissue-specific regulatory architectures, in which clustered enhancers were linked to promoter selectivity and lineage-relevant biological functions. The marked variation in network size and connectivity across tissues further suggested that regulatory complexity is not uniformly distributed, but



899 instead reflects differences in tissue specialization and transcriptional demand (Lin et
900 al., 2025). Second, SE-derived CRCs highlighted a distinct layer of regulation centered
901 on tissue-restricted TF–TF interaction modules (Arnhold et al., 2015; Suh et al., 2002)..
902 Compared with typical enhancers, SEs displayed stronger H3K27ac enrichment and
903 more pronounced tissue specificity, consistent with their established role in reinforcing
904 lineage identity (Adam et al., 2015). Together, these findings indicate that tissue-
905 specific gene regulation is shaped not simply by the presence of enhancers, but by their
906 coordinated organization into distinct regulatory architectures with different
907 topological and functional properties.

908
909 A particularly notable finding is that hierarchical dissection from global networks to
910 central nodes revealed hub enhancers as prominent candidates for regulatory control
911 (Hong et al., 2023; Shu et al., 2022). Hub enhancers identified from enhancer regulatory
912 networks and CRCs likely represent functionally important nodes, yet the limited
913 overlap between these two strategies suggests that they capture distinct but partially
914 nested regulatory layers. Enhancer-regulatory-network–derived hubs appear to reflect
915 broader enhancer cooperativity linked to complex tissue functions, whereas CRC-
916 associated hubs are more closely connected to TF-centered autoregulatory circuitry.
917 Thus, rather than representing redundant approaches, these two frameworks provide
918 complementary views of enhancer organization and jointly refine the interpretation of
919 tissue-specific regulatory logic.

920



921 Importantly, hub enhancers displayed stronger CTCF enrichment than non-hub
922 enhancers, suggesting that enhancer centrality may be coupled not only to
923 transcriptional activity but also to higher-order chromatin organization (Uyehara &
924 Apostolou, 2023). Given the established roles of CTCF and cohesin in chromatin
925 looping and topological domain architecture (Jeong et al., 2024), this enrichment
926 supports a model in which highly connected enhancers act as structural-regulatory
927 nodes that facilitate frequent enhancer–promoter communication. In this view, [hub](#)
928 [enhancers may function beyond strong regulatory activity, potentially serving to](#)
929 [integrate transcriptional control with the spatial organization of the genome](#) (Uyehara
930 & Apostolou, 2023).

931
932 Single-cell analyses in the thymus further refined this interpretation by extending
933 tissue-level network features to cellular resolution. A representative shared hub
934 enhancer associated with *SATB1* showed strong CTCF and RAD21 occupancy and
935 overlapped conserved regulatory signals in mouse datasets, supporting its potential
936 architectural and evolutionary importance. [Furthermore, Hi-C data generated from the](#)
937 [same thymus sample revealed a clear chromatin loop connecting the hub enhancer and](#)
938 [the *SATB1* promoter, providing additional structural evidence for its role in enhancer–](#)
939 [promoter interaction](#). Notably, accessibility at this locus remained broadly distributed
940 across multiple thymic cell populations, consistent with a shared regulatory role rather
941 than a lineage-restricted enhancer state. More broadly, many hub enhancers defined at
942 the tissue level were accessible across diverse thymic cell types (Elmentaite et al., 2022),



suggesting that tissue-level enhancer networks preferentially capture shared regulatory backbones with pleiotropic roles. Such organization may provide a stable regulatory scaffold that supports overall tissue identity while permitting diversification of cell states within the same tissue context.

Consistent with this interpretation, none of the 11,938 cell-type-specific cCREs overlapped with the bulk-defined hub enhancers. Rather than challenging the biological relevance of hub enhancers, this observation suggests the distinct regulatory information captured by bulk and single-cell epigenomic approaches. Hub enhancers were defined according to enhancer interaction density and coordinated regulatory activity within the tissue context and therefore likely represent core regulatory nodes that sustain transcriptional programs shared across dominant thymic cell populations (Perlman et al., 2024). In contrast, cell-type-specific cCREs were identified based on accessibility patterns that distinguish individual cellular states. Consequently, tissue-level hub enhancers may preferentially reflect regulatory backbones underlying common thymic functions, whereas cCREs capture regulatory features associated with cellular specialization. This distinction may be particularly relevant in the thymus, where closely related immune cell populations share substantial portions of their chromatin accessibility landscapes (Ashby & Hogquist, 2024). In addition, methodological differences may further contribute to the limited overlap. Compared with bulk H3K27ac profiling, scATAC-seq data are substantially sparser and primarily measure chromatin accessibility rather than enhancer activity itself (Baek & Lee, 2020).



Moreover, accessibility-based cCRE identification and interaction-based hub enhancer inference rely on fundamentally different criteria. Together, these findings support a hierarchical view of cis-regulatory organization in which tissue-level hub enhancers define shared regulatory frameworks, whereas cell-type-specific cCREs contribute to the diversification of cellular identities within those frameworks. Future integration of perturbation-based validation and single-cell multi-omics data will help clarify how these complementary regulatory layers cooperate to shape gene expression programs across cellular hierarchies.

Several limitations should be acknowledged. Although the present dataset provides broad anatomical coverage, it does not fully capture population-level regulatory variation across breeds, developmental stages, or physiological conditions. Accordingly, the primary objective of this study was to establish a high-resolution whole-body regulatory atlas under a relatively controlled genetic background, rather than to define the full spectrum of regulatory diversity in pig populations. Future investigations including larger and more diverse cohorts will be important for evaluating the robustness, conservation, and generalizability of the regulatory patterns identified in this work. The inferred enhancer–promoter relationships and hub enhancers were derived primarily from correlative epigenomic features rather than direct perturbation or chromatin conformation measurements. It should be noted that the absolute numbers of inferred enhancer networks, hub enhancers, and shared hubs depend on the parameter settings used for network construction and interaction filtering.



The thresholds applied in this study were chosen to provide a relatively stringent framework that prioritizes high-confidence enhancer-enhancer interactions and improves interpretability of network structure. Accordingly, the reported numbers should be interpreted in the context of these analytical choices rather than as fixed estimates independent of parameter selection. In addition, single-cell validation was performed in the thymus, and whether similar principles apply broadly across other tissues remains to be determined. Future integration of chromatin interaction assays and perturbation-based functional experiments will be essential to establish the causal mechanisms underlying enhancer centrality and tissue-specific network organization.

CONCLUSION

In conclusion, this study establishes a comprehensive atlas of enhancer-centered regulatory architectures across pig tissues and reveals that tissue-specific gene regulation is organized through multiple, complementary layers of enhancer coordination. By integrating enhancer-regulatory-network analysis and SE-driven CRC inference, we show that classical enhancer clusters and super-enhancer-associated TF circuitries represent distinct but partially nested regulatory frameworks underlying tissue identity. We further identify hub enhancers as key regulatory nodes, with evidence suggesting that their centrality may reflect both transcriptional importance and architectural integration through CTCF-associated chromatin organization. Finally, single-cell analyses indicate that many tissue-level hub enhancers function as broadly shared regulatory backbones rather than lineage-restricted switches. Together, these

findings provide both a conceptual framework and a valuable resource for dissecting tissue function, developmental regulation, and complex trait biology in pigs.

DATA AVAILABILITY

The raw H3K27ac CUT&Tag datasets in this study have been deposited in the GSA database with the accession number CRA042065. To complement our data, we obtained publicly available ChIP-seq datasets from the ENCODE portal (<https://www.encodeproject.org/>) with the following identifiers: H3K4me1 ChIP-seq, ENCFF695SKH and ENCFF726SHR; H3K4me3 ChIP-seq, ENCFF708XKV and ENCFF331TER; H3K27ac ChIP-seq, ENCFF653GDA and ENCFF019NFC, and ChIP-seq for RAD21 and SATB1 was downloaded from the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE48763 and GSE90635. Any additional information required to reanalyze the data reported in this study is available from the lead contact upon request.

CODE AVAILABILITY

The data analysis pipeline in our study was described on the 10× Genomics (<https://www.10xgenomics.com>), Seurat websites (<https://github.com/satijalab/seurat>), and ArchR websites (<https://www.archrproject.com>). Customized code used in this study has been posted on the GitHub repository at <https://github.com/TaoHJiang/A-high-resolution-H3K27ac-bodymap-reveals-tissue-specific-and-haplotype-resolved-enhancer-in-pigs> and <https://github.com/TaoHJiang/A-whole-body-atlas-of-enhancer>



and-super-enhancer-regulatory-networks-in-pigs. Publicly available software used in this study are listed in the Methods.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

T.J. was responsible for formal analysis, analysis and wrote the original draft. S.Y.L. and J.H.Z. contributed to methodology, investigation, and data curation. Q.Y., S.L., and N.L. contributed to investigation and data curation. J.Z.H. contributed to conceptualization, methodology, data curation and wrote the original draft. All authors have read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We sincerely thank Professor Lusheng Huang for initiating this experiment and for his valuable guidance on the design of experiments, data analysis, and manuscript revision.

We thank member of the National Key Laboratory for Swine Genetic Improvement and Germplasm Innovation, Jiangxi Agricultural University for their assistance in experiments and data analysis. We thank High-Performance Computing Centre (HPC)



1053 of the National Key Laboratory for Swine Genetic Improvement and Germplasm
1054 Innovation, Jiangxi Agricultural University for providing computing resources. This
1055 work was supported by the National Key Research and Development Program of China
1056 (2024YFF1000201) and the Postdoctoral Fellowship Program of CPSF under Grant
1057 Number GZC20240619.

uncorrected proof



Reference

- Adam RC, Yang H, Rockowitz S, et al. 2015. Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature*, **521**(7552): 366-370.
- Almeida N, Chung MWH, Drudi EM, et al. 2021. Employing core regulatory circuits to define cell identity. *Embo j*, **40**(10): e106785.
- Arnhold IJ, França MM, Carvalho LR, et al. 2015. Role of GLI2 in hypopituitarism phenotype. *J Mol Endocrinol*, **54**(3): R141-150.
- Ashby KM & Hogquist KA. 2024. A guide to thymic selection of T cells. *Nat Rev Immunol*, **24**(2): 103-117.
- Baek S & Lee I. 2020. Single-cell ATAC sequencing analysis: From data preprocessing to hypothesis generation. *Comput Struct Biotechnol J*, **18**: 1429-1439.
- Bando H, Brinkmeier ML, Castinetti F, et al. 2023. Heterozygous variants in SIX3 and POU1F1 cause pituitary hormone deficiency in mouse and man. *Hum Mol Genet*, **32**(3): 367-385.
- Blayney JW, Francis H, Rampasekova A, et al. 2023. Super-enhancers include classical enhancers and facilitators to fully activate gene expression. *Cell*, **186**(26): 5826-5839.e5818.
- Boeva V, Louis-Brennetot C, Peltier A, et al. 2017. Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat Genet*, **49**(9): 1408-1413.
- Boyer LA, Lee TI, Cole MF, et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, **122**(6): 947-956.
- Chen J, Wu L, Li Y. 2024. FGL1 and FGL2: emerging regulators of liver health and disease. *Biomark Res*, **12**(1): 53.
- Chen L, Huang M, Plummer J, et al. 2020a. Master transcription factors form interconnected circuitry and orchestrate transcriptional networks in oesophageal adenocarcinoma. *Gut*, **69**(4): 630-640.
- Chen SD, Li HQ, Shen XN, et al. 2020b. Genome-Wide Association Study Identifies SLAMF1 Affecting the Rate of Memory Decline. *J Alzheimers Dis*, **74**(1): 139-149.
- Corder ML, Berland S, Førsvoll JA, et al. 2022. Truncating and zinc-finger variants in GLI2 are associated with hypopituitarism. *Am J Med Genet A*, **188**(4): 1065-1074.
- Da Costa IB, De Labio RW, Rasmussen LT, et al. 2017. Change in INSR, APBA2 and IDE Gene Expressions in Brains of Alzheimer's Disease Patients. *Curr Alzheimer Res*, **14**(7): 760-765.
- Durand NC, Shamim MS, Machol I, et al. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst*, **3**(1): 95-98.
- Elmentaite R, Domínguez Conde C, Yang L, et al. 2022. Single-cell atlases: shared and tissue-specific cell types across human organs. *Nat Rev Genet*, **23**(7): 395-410.
- Fang LandTeng JandLin Q, et al. 2025. The Farm Animal Genotype-Tissue Expression (FarmGTEx) Project. *Nat Genet*, **57**(4): 786-796.
- Feng C, Song C, Jiang Y, et al. 2023. Landscape and significance of human super enhancer-driven core transcription regulatory circuitry. *Mol Ther Nucleic Acids*, **32**: 385-401.
- Ferguson LB, Zhang L, Kircher D, et al. 2019. Dissecting Brain Networks Underlying Alcohol Binge Drinking Using a Systems Genomics Approach. *Mol Neurobiol*, **56**(4): 2791-2810.
- Galili T. 2015. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, **31**(22): 3718-3720.
- Granja JM, Corces MR, Pierce SE, et al. 2021. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet*, **53**(3): 403-411.



1101 Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics*,
 1102 **27**(7): 1017-1018.

1103 Gray LR, Tompkins SC, Taylor EB. 2014. Regulation of pyruvate metabolism and human disease. *Cell*
 1104 *Mol Life Sci*, **71**(14): 2577-2604.

1105 Gu W, Madrid DMC, Joyce S, et al. 2022. A single-cell analysis of thymopoiesis and thymic iNKT cell
 1106 development in pigs. *Cell Rep*, **40**(1): 111050.

1107 Gu Z, Gu L, Eils R, et al. 2014. circlize Implements and enhances circular visualization in R.
 1108 *Bioinformatics*, **30**(19): 2811-2812.

1109 Hao Y, Hao S, Andersen-Nissen E, et al. 2021. Integrated analysis of multimodal single-cell data. *Cell*,
 1110 **184**(13): 3573-3587.e3529.

1111 Heinz S, Benner C, Spann N, et al. 2010. Simple combinations of lineage-determining transcription
 1112 factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, **38**(4):
 1113 576-589.

1114 Henikoff S, Henikoff JG, Kaya-Okur HS, et al. 2020. Efficient chromatin accessibility mapping in situ
 1115 by nucleosome-tethered tagmentation. *Elife*, **9**.

1116 Hnisz D, Abraham BJ, Lee TI, et al. 2013. Super-enhancers in the control of cell identity and disease.
 1117 *Cell*, **155**(4): 934-947.

1118 Hong D, Lin H, Liu L, et al. 2023. Complexity of enhancer networks predicts cell identity and disease
 1119 genes revealed by single-cell multi-omics analysis. *Brief Bioinform*, **24**(1).

1120 Hong JW, Hendrix DA, Levine MS. 2008. Shadow enhancers as a source of evolutionary novelty.
 1121 *Science*, **321**(5894): 1314.

1122 Huang J, Li K, Cai W, et al. 2018a. Dissecting super-enhancer hierarchy based on chromatin interactions.
 1123 *Nat Commun*, **9**(1): 943.

1124 Huang M, Chen Y, Yang M, et al. 2018b. dbCoRC: a database of core transcriptional regulatory
 1125 circuitries modeled by H3K27ac ChIP-seq signals. *Nucleic Acids Res*, **46**(D1): D71-d77.

1126 Jeong D, Shi G, Li X, et al. 2024. Structural basis for the preservation of a subset of topologically
 1127 associating domains in interphase chromosomes upon cohesin depletion. *Elife*, **12**.

1128 Jiang T, Ling Z, Zhou Z, et al. 2022. Construction of a transposase accessible chromatin landscape
 1129 reveals chromatin state of repeat elements and potential causal variant for complex traits in pigs. *J Anim*
 1130 *Sci Biotechnol*, **13**(1): 112.

1131 Jung YL, Zhao W, Li I, et al. 2024. Epigenetic profiling reveals key genes and cis-regulatory networks
 1132 specific to human parathyroids. *Nat Commun*, **15**(1): 2106.

1133 Kaya-Okur HS, Wu SJ, Codomo CA, et al. 2019. CUT&Tag for efficient epigenomic profiling of small
 1134 samples and single cells. *Nat Commun*, **10**(1): 1930.

1135 Kitagawa Y, Ohkura N, Kidani Y, et al. 2017. Guidance of regulatory T cell development by Satb1-
 1136 dependent super-enhancer establishment. *Nat Immunol*, **18**(2): 173-183.

1137 Langmead B & Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**(4): 357-
 1138 359.

1139 Li C, Ge M, Long K, et al. 2025. Mechanism of parent-of-origin effects revealed by multi-omic data in
 1140 euro-chinese hybrid pigs. *Nat Commun*, **16**(1): 7542.

1141 Li H, Handsaker B, Wysoker A, et al. 2009. The Sequence Alignment/Map format and SAMtools.
 1142 *Bioinformatics*, **25**(16): 2078-2079.

1143 Li M, Huang H, Li L, et al. 2021. Core transcription regulatory circuitry orchestrates corneal epithelial
 1144 homeostasis. *Nat Commun*, **12**(1): 420.



1145 Lin H, Ye X, Chen W, et al. 2025. Modular organization of enhancer network provides transcriptional
 1146 robustness in mammalian development. *Nucleic Acids Res*, **53**(2).

1147 Livak KJ & Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative
 1148 PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**(4): 402-408.

1149 Lopez-Delisle L, Rabbani L, Wolff J, et al. 2021. pyGenomeTracks: reproducible plots for multivariate
 1150 genomic datasets. *Bioinformatics*, **37**(3): 422-423.

1151 Monahan P, Rybak S, Raetzman LT. 2009. The notch target gene HES1 regulates cell cycle inhibitor
 1152 expression in the developing pituitary. *Endocrinology*, **150**(9): 4386-4394.

1153 Naito T, Ise M, Tanaka Y, et al. 2023. Crucial Roles of SATB1 in Regulation of Thymocyte Migration
 1154 after Positive Selection. *J Immunol*, **211**(2): 209-218.

1155 Osterwalder M, Barozzi I, Tissières V, et al. 2018. Enhancer redundancy provides phenotypic robustness
 1156 in mammalian development. *Nature*, **554**(7691): 239-243.

1157 Parker SC, Stitzel ML, Taylor DL, et al. 2013. Chromatin stretch enhancer states drive cell-specific gene
 1158 regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A*, **110**(44): 17921-17926.

1159 Perlman BS, Burget N, Zhou Y, et al. 2024. Enhancer-promoter hubs organize transcriptional networks
 1160 promoting oncogenesis and drug resistance. *Nat Commun*, **15**(1): 8070.

1161 Pott S & Lieb JD. 2015. What are super-enhancers? *Nat Genet*, **47**(1): 8-12.

1162 Preissl S, Gaulton KJ, Ren B. 2023. Characterizing cis-regulatory elements using single-cell epigenomics.
 1163 *Nat Rev Genet*, **24**(1): 21-43.

1164 Quinlan AR & Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.
 1165 *Bioinformatics*, **26**(6): 841-842.

1166 Ramírez F, Ryan DP, Grüning B, et al. 2016. deepTools2: a next generation web server for deep-
 1167 sequencing data analysis. *Nucleic Acids Res*, **44**(W1): W160-165.

1168 Rao L, Cai L, Huang L. 2023. Single-cell dynamics of liver development in postnatal pigs. *Sci Bull*
 1169 *(Beijing)*, **68**(21): 2583-2597.

1170 Rao SS, Huntley MH, Durand NC, et al. 2014. A 3D map of the human genome at kilobase resolution
 1171 reveals principles of chromatin looping. *Cell*, **159**(7): 1665-1680.

1172 Ren L, Yang K, Zeng Y, et al. 2026. Chromatin-mediated mechanisms of hybrid fertility in F(1) male
 1173 hybrids of Culter alburnus and Megalobrama amblycephala. *BMC Biol*, **24**(1): 42.

1174 Rougon G & Hobert O. 2003. New insights into the diversity and function of neuronal immunoglobulin
 1175 superfamily molecules. *Annu Rev Neurosci*, **26**: 207-238.

1176 Saint-André V, Federation AJ, Lin CY, et al. 2016. Models of human core transcriptional regulatory
 1177 circuitries. *Genome Res*, **26**(3): 385-396.

1178 Seitan VC, Faure AJ, Zhan Y, et al. 2013. Cohesin-based chromatin interactions enable regulated gene
 1179 expression within preexisting architectural compartments. *Genome Res*, **23**(12): 2066-2077.

1180 Shu M, Hong D, Lin H, et al. 2022. Single-cell chromatin accessibility identifies enhancer networks
 1181 driving gene expression during spinal cord development in mouse. *Dev Cell*, **57**(24): 2761-2775.e2766.

1182 Smith AD, Sumazin P, Zhang MQ. 2007. Tissue-specific regulatory elements in mammalian promoters.
 1183 *Mol Syst Biol*, **3**: 73.

1184 Suh H, Gage PJ, Drouin J, et al. 2002. Pitx2 is required at multiple stages of pituitary organogenesis:
 1185 pituitary primordium formation and cell specification. *Development*, **129**(2): 329-337.

1186 Thomas HF, Kotova E, Jayaram S, et al. 2021. Temporal dissection of an enhancer cluster reveals distinct
 1187 temporal and functional contributions of individual elements. *Mol Cell*, **81**(5): 969-982.e913.

1188 Uyehara CM & Apostolou E. 2023. 3D enhancer-promoter interactions and multi-connected hubs:

Organizational principles and functional roles. *Cell Rep*, **42**(4): 112068.

Vermunt MW, Reinink P, Korving J, et al. 2014. Large-scale identification of coregulated enhancer networks in the adult human brain. *Cell Rep*, **9**(2): 767-779.

Welcker JE, Hernandez-Miranda LR, Paul FE, et al. 2013. Insm1 controls development of pituitary endocrine cells and requires a SNAG domain for function and for recruitment of histone-modifying factors. *Development*, **140**(24): 4947-4958.

Whitehead A & Crawford DL. 2005. Variation in tissue-specific gene expression among natural populations. *Genome Biol*, **6**(2): R13.

Whyte WA, Orlando DA, Hnisz D, et al. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**(2): 307-319.

Wolf FA, Angerer P, Theis FJ. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol*, **19**(1): 15.

Xu Q, Georgiou G, Frölich S, et al. 2021. ANANSE: an enhancer network-based computational approach for predicting key transcription factors in cell fate determination. *Nucleic Acids Res*, **49**(14): 7966-7985.

Yan CY, Wu FY, Sun F, et al. 2023. The isl2a transcription factor regulates pituitary development in zebrafish. *Front Endocrinol (Lausanne)*, **14**: 920548.

Yanai I, Benjamin H, Shmoish M, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, **21**(5): 650-659.

Zhang Y, Liu T, Meyer CA, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, **9**(9): R137.

Zhang YD, Guo C, Liu H, et al. 2024. Repressor elements provide insights into tissue development and phenotypes in pigs. *Zool Res*, **45**(6): 1313-1326.

Zhu Y, Zhou Z, Huang T, et al. 2022. Mapping and analysis of a spatiotemporal H3K27ac and gene expression spectrum in pigs. *Sci China Life Sci*, **65**(8): 1517-1534.

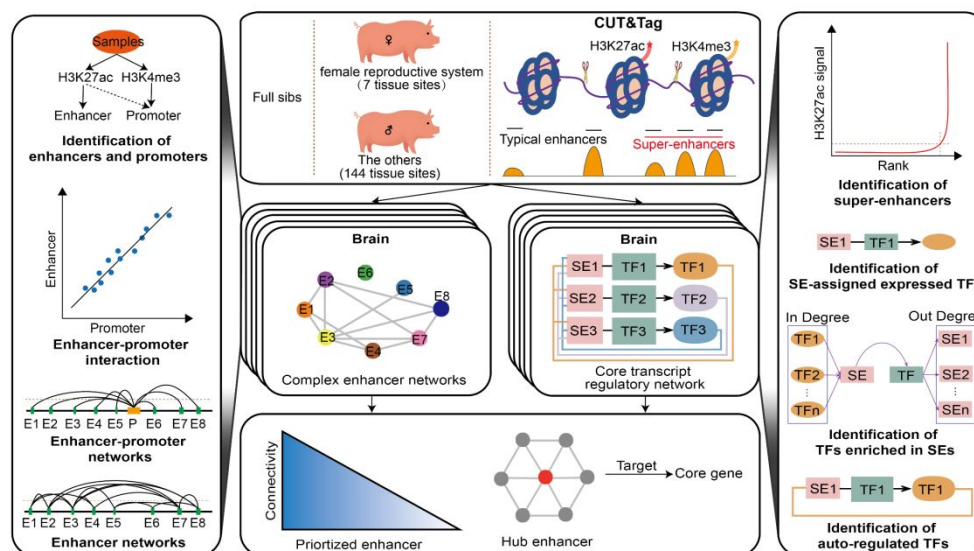


Figure 1 Identification enhancer networks across pig tissues. Left panel (from top to bottom): Identification of the enhancer and promoter using H3K27ac and H3K4me3 CUT&Tag data. Mapping enhancer clusters that regulate the promoters of the same gene based on co-accessibility. Constructing enhancer networks regulating a specific target gene. Middle panel (from top to bottom): Overview of tissue sampling and CUT&Tag workflow. Building enhancer networks using two computational

approaches. Detecting hub enhancers and core genes within the constructed networks. Right panel (from top to bottom): Defining super-enhancers (SEs) across pig tissues. Identification of SE-assigned expressed TFs and TFs enriched in SEs. Identification of auto regulatory TFs.

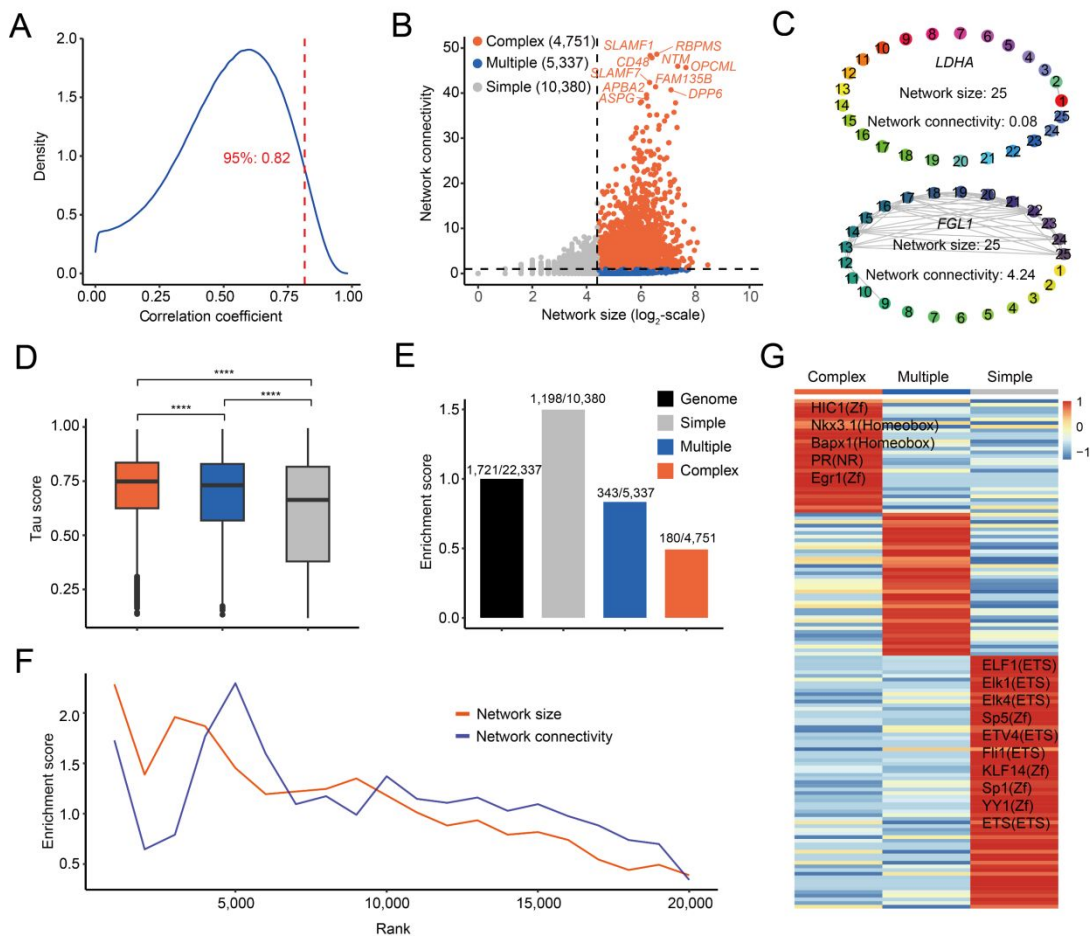


Figure 2 Characterization of enhancer networks.

A: Density map showing the distribution of correlation coefficients of EEI, with the 95% quantile labeled. B: Distribution of enhancer networks, where the x axis ($\log_2$ -scaled) represents the network size (the number of enhancers per network) and the y axis indicates the network connectivity, quantified as the average number of edges per node in network. The top 10 complex enhancer networks are labeled with their target genes. C: Representative enhancer networks illustrating multiple (top) mode and complex (bottom) mode. D: Comparison of promoter specificity (Tau score) across complex, multiple, and simple modes. Statistical significance is indicated as $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$ (Student's t-test). E: Enrichment analysis of housekeeping genes among different modes, using the whole genome as the background. F: Enrichment of housekeeping genes (y axis) across gene subsets (x axis) ranked by network connectivity and size. G: Differences in TFs enriched in target promoters corresponding to the three modes of enhancer regulatory networks. Several key transcription factors with significant differences are labeled.

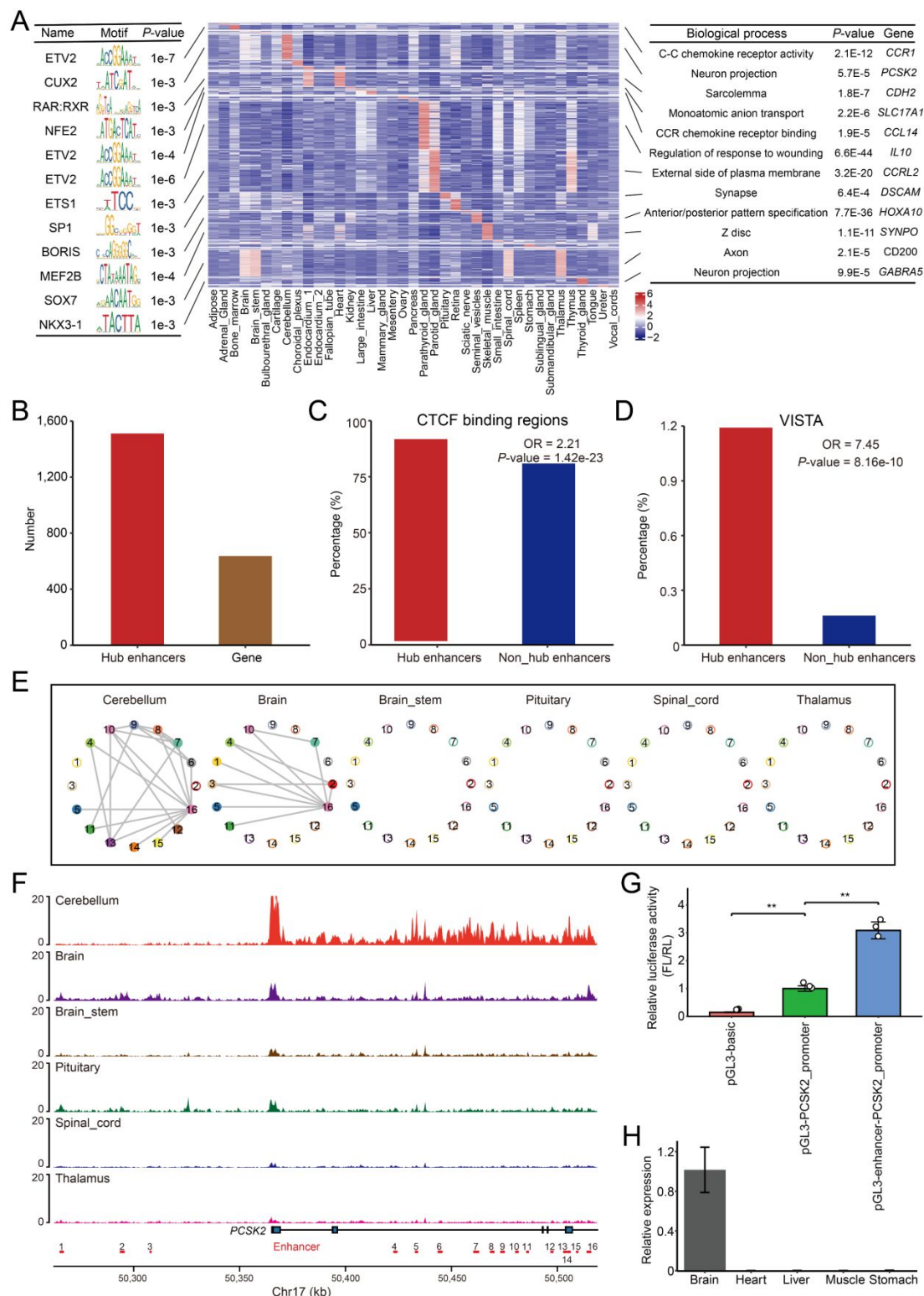


Figure 3 Complex enhancer networks for tissue specificity.

A: Heatmap showing the association of 58 tissues (columns) with tissue-specific enhancers (rows). The representative gene ontology (GO) terms (right) and TF binding motifs (left) enriched in each tissue were displayed. B: Bar plot illustrating the number of network hub enhancers and their targeted genes. C: Bar plot illustrating the percentage of hub enhancers enriched for CTCF binding motifs (relative to all hub

enhancers) with the percentage of non-hub enhancers enriched CTCF binding motifs (relative to all non-hub enhancers). Fold enrichment and *P*-values were calculated using the binomial test. D: Bar plot illustrating the percentage of hub enhancers enriched for VISTA validated enhancers (relative to all hub enhancers) with the percentage of non-hub enhancers enriched VISTA validated enhancers (relative to all non-hub enhancers). Fold enrichment and *P*-values were calculated using the binomial test. E: Enhancer networks for *PCSK2* across all brain sites. Nodes represent enhancers within the enhancer network, with blank nodes indicating removed enhancer. The edges represent the EEIs. F: Genome browser tracks showing aggregated H3K27ac signals for enhancer clusters in the enhancer networks for *PCSK2* across all brain sites. G: Dual-luciferase reporter assay showing the activating effect of the predicted hub enhancer on the *PCSK2* promoter. Three reporter constructs were examined: basic, basic + promoter, and basic + enhancer + promoter. Data are presented as mean ± SD from three independent experiments. Statistical significance was determined using two-tailed Student's *t*-tests. ns, not significant; **P* < 0.05; ***P* < 0.01. H: RT-qPCR analysis showing the expression levels of *PCSK2* across different tissues.



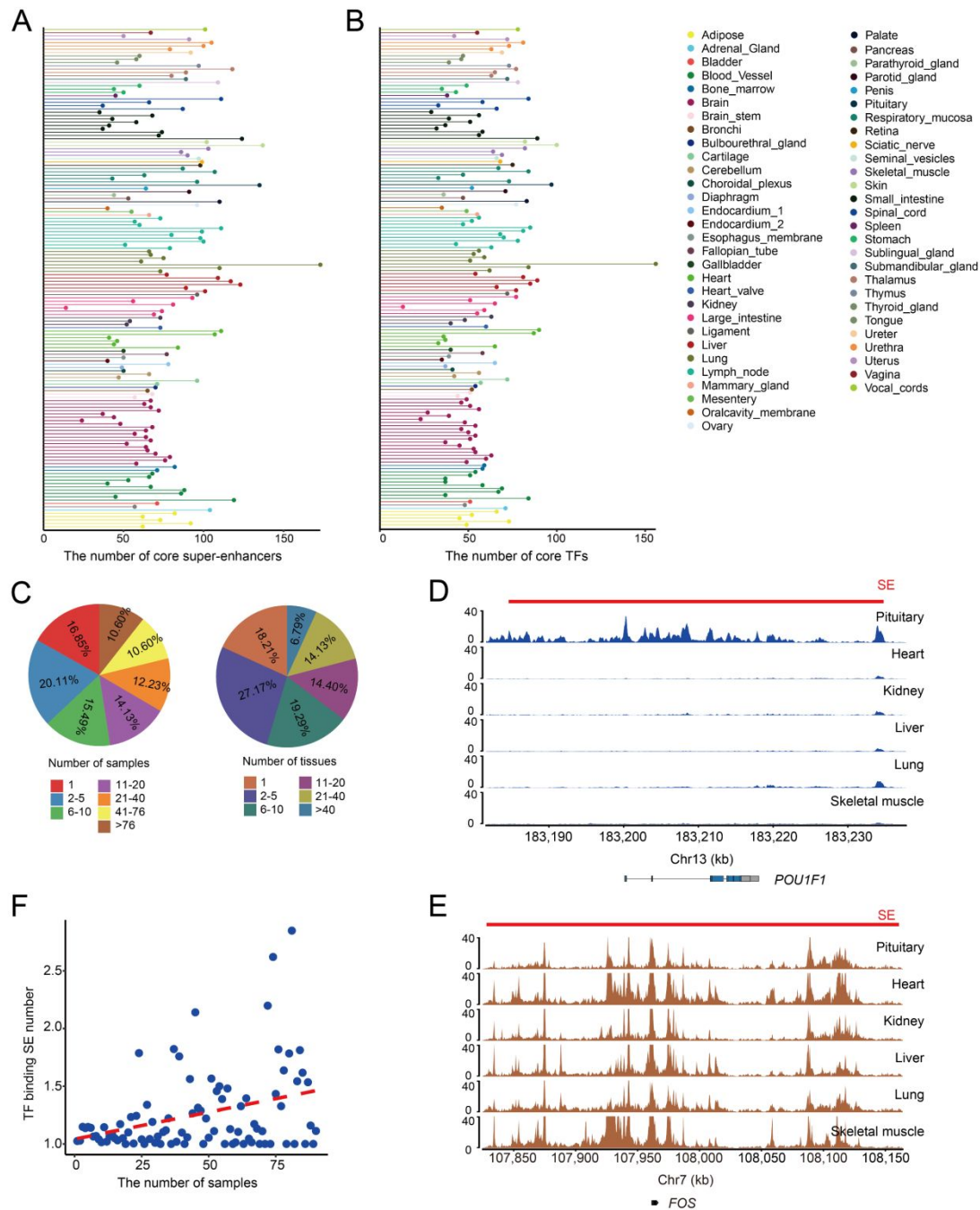


Figure 4 Identification of core regulatory circuitries in 151 samples.

A-B: Bar plot illustrating the distribution of the number of SEs (A) and TFs (B) identified across each sample. C: Left: The pie chart demonstrating the percentage of core TFs present in different numbers of samples. Right: The pie chart demonstrating the percentage of core TFs present in different numbers of tissues. D-E: Genome browser tracks depicting H3K27ac signals around the loci of *POU1F1* (D) and *FOS* (E) in six representative tissues. F: Scatter plot showing the relationship between the number of samples in which TFs were shared and the number of SEs they bound to. The x axis is the number of samples and the y axis is the number of SEs.

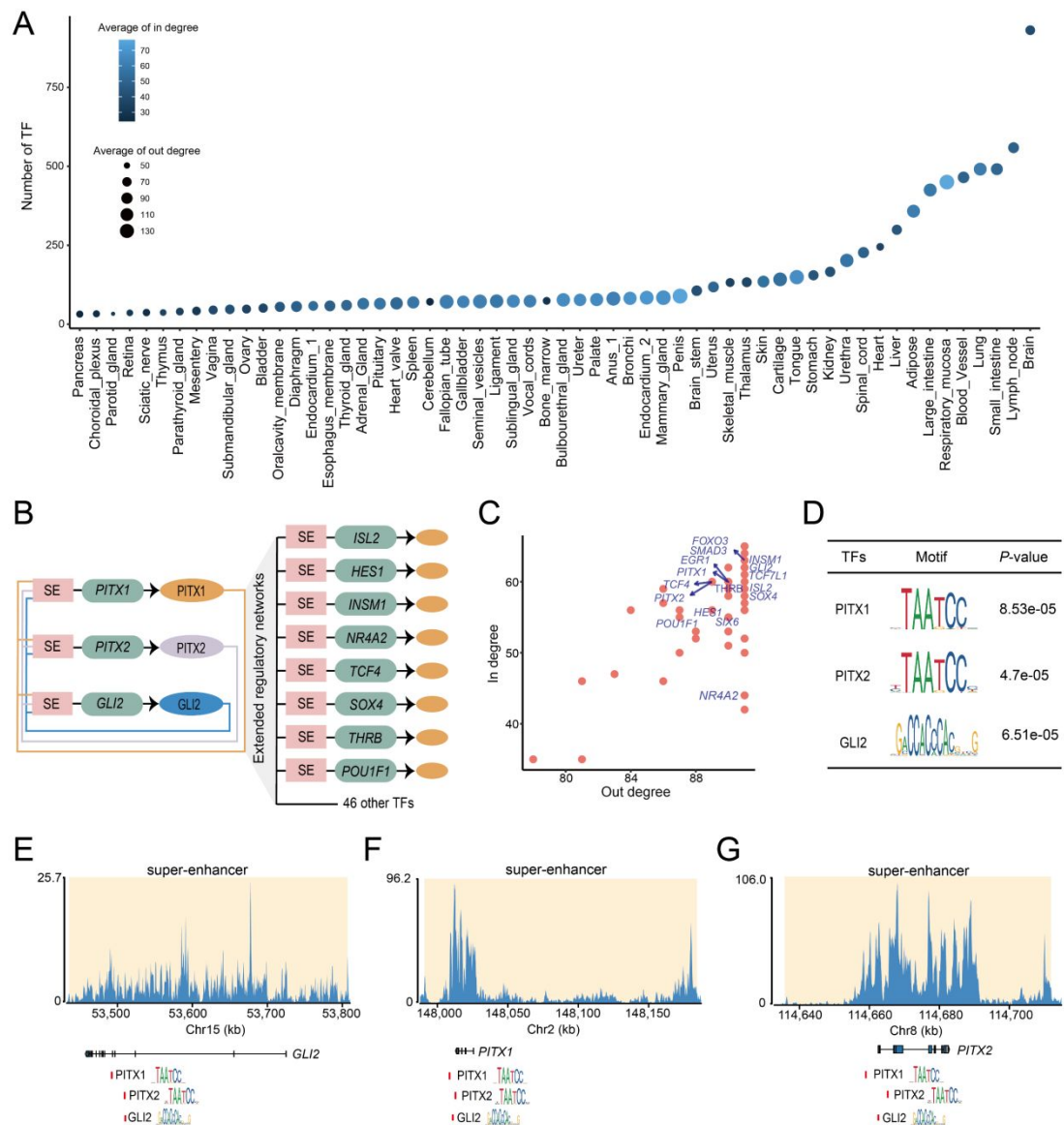


Figure 5 Core and extended regulatory circuitry in the pituitary.

A: Distribution of core TFs across tissues. Point size represents the average out-degree, and color intensity indicates the average in-degree, reflecting regulatory influence and regulatory input, respectively. **B:** Schematic model illustrating the core regulatory circuitry in the pituitary, established by PITX1, PITX2 and GLI2. **C:** Scatter plot of in-degree (number of TFs binding to the SE of a node TF) and out-degree (number of SEs bound by a node TF) for TFs in the SE-mediated regulatory network. **D:** Enrichment of PITX1, PITX2 and GLI2 motifs at SEs. **E-G:** Genome browser tracks depicting H3K27ac signals around the indicated SE loci in the pituitary, with logos and positions of the enriched TF motifs shown below.

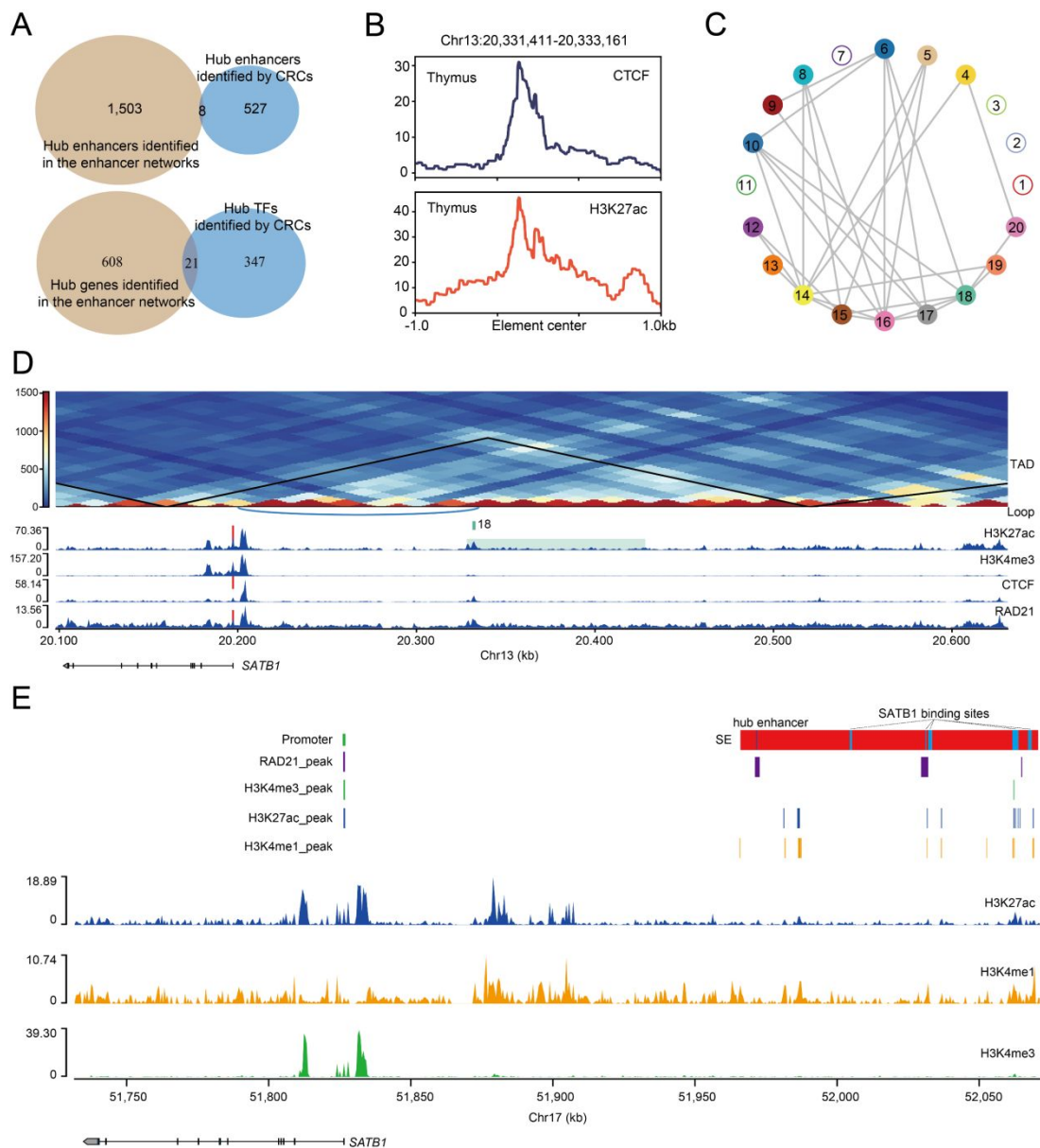


Figure 6 Hub enhancers identified by CRCs.

A: Venn diagrams illustrating the overlap of hub enhancers (upper panel) and core genes (bottom panel) identified by enhancer networks and CRCs. B: H3K27ac and CTCF signals within a 1 kb window centered on the hub enhancer. C: The enhancer network of SATB1, with code 18 representing the hub enhancer. D: Genome browser tracks depicting H3K27ac, H3K4me3, CTCF, RAD21, and Hi-C interaction signals around the *SATB1* locus in the thymus, highlighting enhancer-promoter contacts within the same TAD and supported chromatin loop. E: Genome browser tracks depicting H3K27ac, H3K4me3 and H3K4me1 signals around the loci of *SATB1* in the mouse. The hub enhancer, the SE containing the hub enhancer, and the promoter of *SATB1* were lifted over from the pig genome to the mouse genome. Histone modification peaks (H3K27ac, H3K4me3 and H3K4me1), SATB1 and RAD21 binding sites were showed.

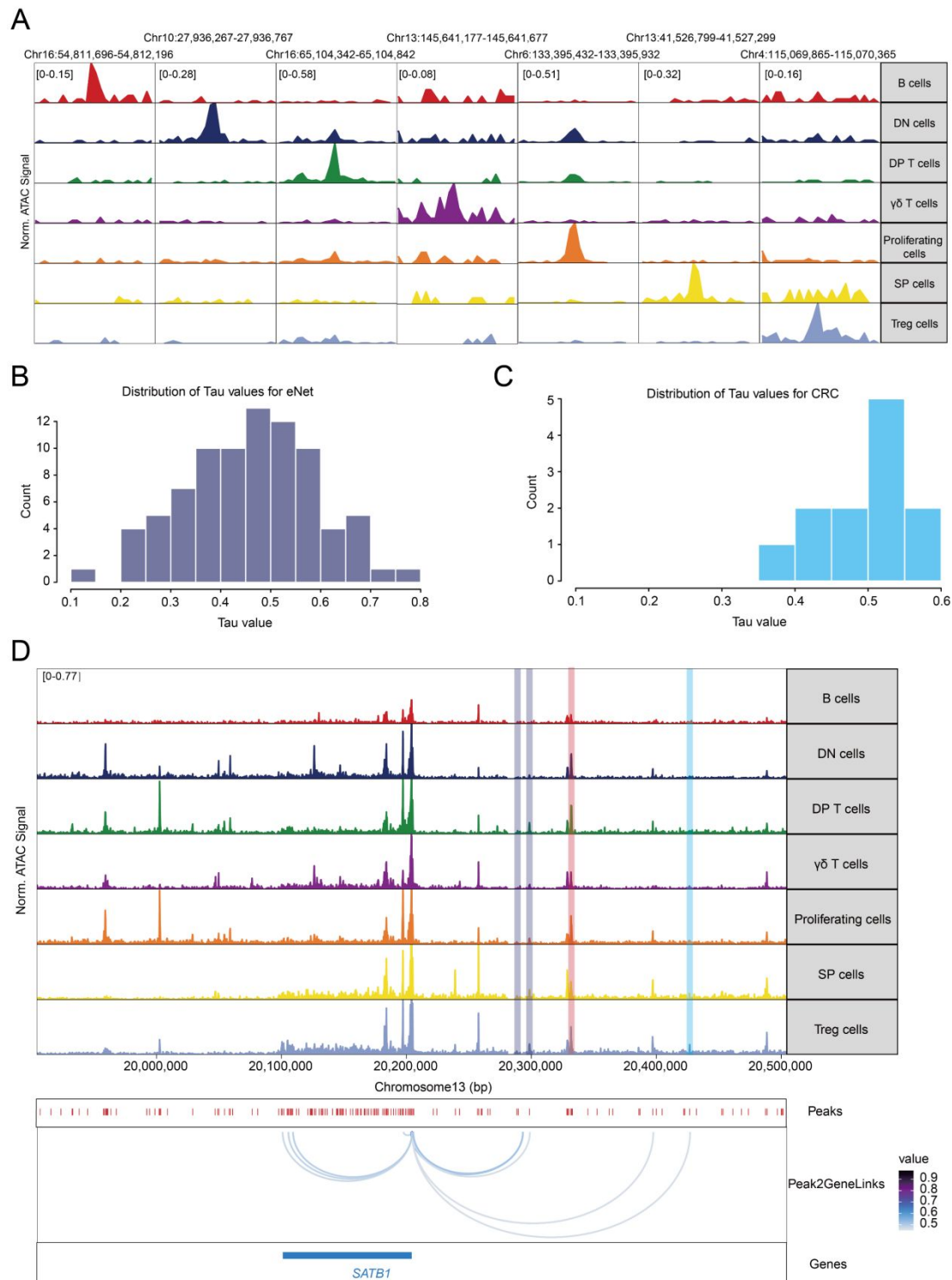


Figure 7 Identification of cell-type-specific cis-regulatory elements (cCREs) in the thymus.

A: Genome browser tracks depicting chromatin accessibility peaks of representative cCREs across thymic cell types identified by scATAC-seq. B-C: Bar plot showing the distribution of Tau scores for hub enhancers identified by the eNet algorithm (B) and CRC analysis (C), respectively. D: Genome browser tracks depicting chromatin accessibility signals around the *SATB1* locus across thymic cell types. The red shading

1306 denotes hub enhancer intervals identified by both the eNet and CRC approaches; blue
1307 shading marks hub enhancers identified only within CRCs, and grey shading indicates
1308 hub enhancers uniquely detected by the eNet algorithm.

1309

uncorrected proof

