Supplementary Materials for

High-quality sika deer omics data and integrative analysis reveal genic and cellular regulation of antler regeneration.

Zihe Li[1,11], Ziyu Xu[2,3,11], Lei Zhu[6,7,11], Tao Qin[1,11], Jinrui Ma[1,11], Zhanying Feng[2,10,11], Huishan Yue[1], Qing Guan[9], Botong Zhou[1], Ge Han[1], Guokun Zhang[8], Chunyi Li[8], Shuaijun Jia[6,7], Qiang Qiu[1]*, Dingjun Hao[6,7]*, Yong Wang[2,3,4,5]*, Wen Wang[1,9]*

[1] New Cornerstone Science Laboratory, Shaanxi Key Laboratory of Qinling Ecological Intelligent Monitoring and Protection, School of Ecology and Environment, Northwestern Polytechnical University, Xi'an 710072, China.

[2]CEMS, NCMIS, HCMS, MADIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China.

[3]School of Mathematics, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing 100049, China.

[4]Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou, 310024, China.

[5]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China.

[6]Department of Spine Surgery, Honghui Hospital, Xi’an Jiaotong University, Xi’an, Shaanxi 710054, China

[7]Shaanxi Key Laboratory of Spine Bionic Treatment, Xi’an, Shaanxi 710054, China

[8] Institute of Antler Science and Product Technology, Changchun Sci-Tech University; 130600 Changchun, China

[9]Key Laboratory of Genetic Evolution & Animal Models, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China.

[10]Department of Statistics, Department of Biomedical Data Science, Bio-X Program, Stanford University, Stanford, CA 94305, USA

[11]These authors contributed equally: Zihe Li, Ziyu Xu, Lei Zhu Tao Qin, Jinrui Ma, Zhanying Feng

* Correspondent authors: wenwang@nwpu.edu.cn (W.W.), ywang@amss.ac.cn (Y.W.), haodingjun@mail.xjtu.edu.cn (D.J.H.), qiuqiang@nwpu.edu.cn (Q.Q.)

This Supplementary Information includes:

Supplementary Methods

Supplementary Figures 1 – 11

Supplementary Methods

Animal Ethics.

Four 2-year-old male sika deers (*Cervus nippon*) were used for sampling regenerating antler Tissues. The deer were farmed in an enclosure, provided water and fed ad libitum, with feed including dry leaves (mainly oak tree leaves) and grass and some cooked grains (mainly soya bean cakes). All experimental protocols and procedures were approved by the Animal Ethics Committee of Changchun Sci-Tech University (Approval No. CKARI202007). On days 0, 2, 5 and 10 after casting, deer were anesthetized. The entire pedicles were thoroughly shaved, then after routine sterilization with 1% iodophor and 75% ethanol, the skin was cut open, pedicle periosteum or regenerating antler tissue (on the distal end of pedicle) was exposed and collected, then immediately placed in liquid nitrogen to froze. After sample collection, all animals were raised until natural death.

Genome size estimation.

We used chicken blood cells as an internal standard. 1ml of fresh heparin anticoagulant blood samples were collected from the veins beneath the wings of four chickens. The samples were stored on ice and transported back to the laboratory. Each 1 ml of anticoagulant blood was diluted with 10 ml of physiological saline, and then centrifuged at 4 °C and 1500 rpm for 10 minutes. The supernatant was discarded, and this step was repeated twice. Finally, 1 ml of physiological saline was used to suspend

the red blood cells, and chicken red blood cells were added to Galbraith's buffer at a ratio of 1:100 to prepare the chicken cell nuclear suspension.

Four sika deers were anesthetized, and 1 ml of fresh heparin anticoagulant blood samples were collected from their neck arteries. The samples were stored on ice and transported back to the laboratory. Each 1 ml of anticoagulant blood was diluted with 10 ml of physiological saline, and then centrifuged at 4 °C and 1500 rpm for 10 minutes. The supernatant was discarded, and this step was repeated twice. Finally, 1 ml of physiological saline was used to suspend the cells (after centrifugation, mainly white blood cells were obtained, along with a small number of other cells), and deer cells were added to Galbraith's buffer at a ratio of 1:100 to prepare the deer cell nuclear suspension. As a control sample, 1 ml of fresh heparin anticoagulant blood samples were collected from the tail veins of four rats. Each 1 ml of anticoagulant blood was diluted with 10 ml of physiological saline, and then centrifuged at 4 °C and 1500 rpm for 10 minutes. The supernatant was discarded, and this step was repeated twice. Finally, 1 ml of physiological saline was used to suspend the red blood cells, and rat cells were added to Galbraith's buffer at a ratio of 1:100 to prepare the rat cell nuclear suspension.

After preparing the cell nuclear suspension samples for deer, chicken, and rat, the relative fluorescence positions of the two samples were determined. The deer and rat cell nuclear suspensions were mixed in a ratio of 10:1, and the rat and chicken cell nuclear suspensions and the deer and chicken cell nuclear suspensions were mixed in

a ratio of 500:3. Then, 50 μl of propidium iodide (PI) staining solution was added, and the samples were stained in a dark environment at 4 °C for 20 minutes.

The cell nuclear suspensions were filtered through a 20 μm nylon membrane, transferred to flow cytometry tubes, and analyzed using a BD FACSCelesta 3 laser. The stained samples were detected under excitation light at 488 nm, and the relative fluorescence signals of the 2C nuclei of the internal standard samples and the test samples were measured to determine the average fluorescence peak. At least 10,000 nuclear fluorescence signals were collected for each sample, and the coefficient of variation (CV) was controlled within 5%.

Single cell RNA-seq library preparation and sequencing

Liver sample was pooled, minced with a razor blade, and digested with collagenase I (100 μg/ml) + collagenase II (100 μg/ml) + collagenase IV (30 μg/ml) mixed solution (Gibco, 17100017, 17101015, and 17104019) at 37°C for 60 min (or until the tissues were fully digested), with intermittent shaking. Single cell suspensions were sequentially filtered through 70 μm and 40 μm cell strainers (Corning, 352350 and 352340), and centrifuged at 500 × g for 5 min. The cell pellets were resuspended in PBS and treated with the Red Blood Cell Lysis Buffer (Solarbio, R1010) to remove red blood cells. Live cells were then washed with PBS twice, resuspended in 0.04% UltraPure BSA solution (Thermo, 23209), and counted using a Countess automated cell counter (Bio-Rad TC20, US). Live cells (confirmed by staining with 7-aminoactinomycin D [BD Biosciences, 51- 68981E]; 90–95% viability after sorting)

were collected from samples collected at each stage, resuspended at 1 × 103 cells/ml and single cells were obtained using a Chromium Controller (10× Genomics, US). We followed a standard protocol for scRNA-seq library construction (10× Genomics, US). The libraries were sequenced with an Illumina NovaSeq6000 platform (Novogene, China).

ATAC-seq library preparation and sequencing.

ATAC-seq was performed as previously reported(Buenrostro et al. 2013). Briefly, nuclei were extracted from samples, and the nuclei pellet was resuspended in the Tn5 transposase reaction mix. The transposition reaction was incubated at 37°C for 30 min. Equimolar Adapter 1 and Adapter 2 were added after transposition, PCR was then performed to amplify the library. After the PCR reaction, libraries were purified with the AMPure beads and library quality was assessed with Qubit. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina Hiseq platform and 150 bp paired-end reads were generated.

RNA-seq library preparation and sequencing.

A total amount of 1 μg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext® UltraTM

RNA Library Prep Kit for Illumina® (NEB, USA) following manufacturer's recommendations and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly(T) oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer(5X). First-strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H). Second-strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEBNext Adaptor with hairpin loop structure were ligated to prepare for hybridization. In order to select cDNA fragments of preferentially 250~300 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then 3 μl USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95 °C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumia) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina NovaSeq platform and 150 bp paired-end

reads were generated.

HiFi library preparation and sequencing.

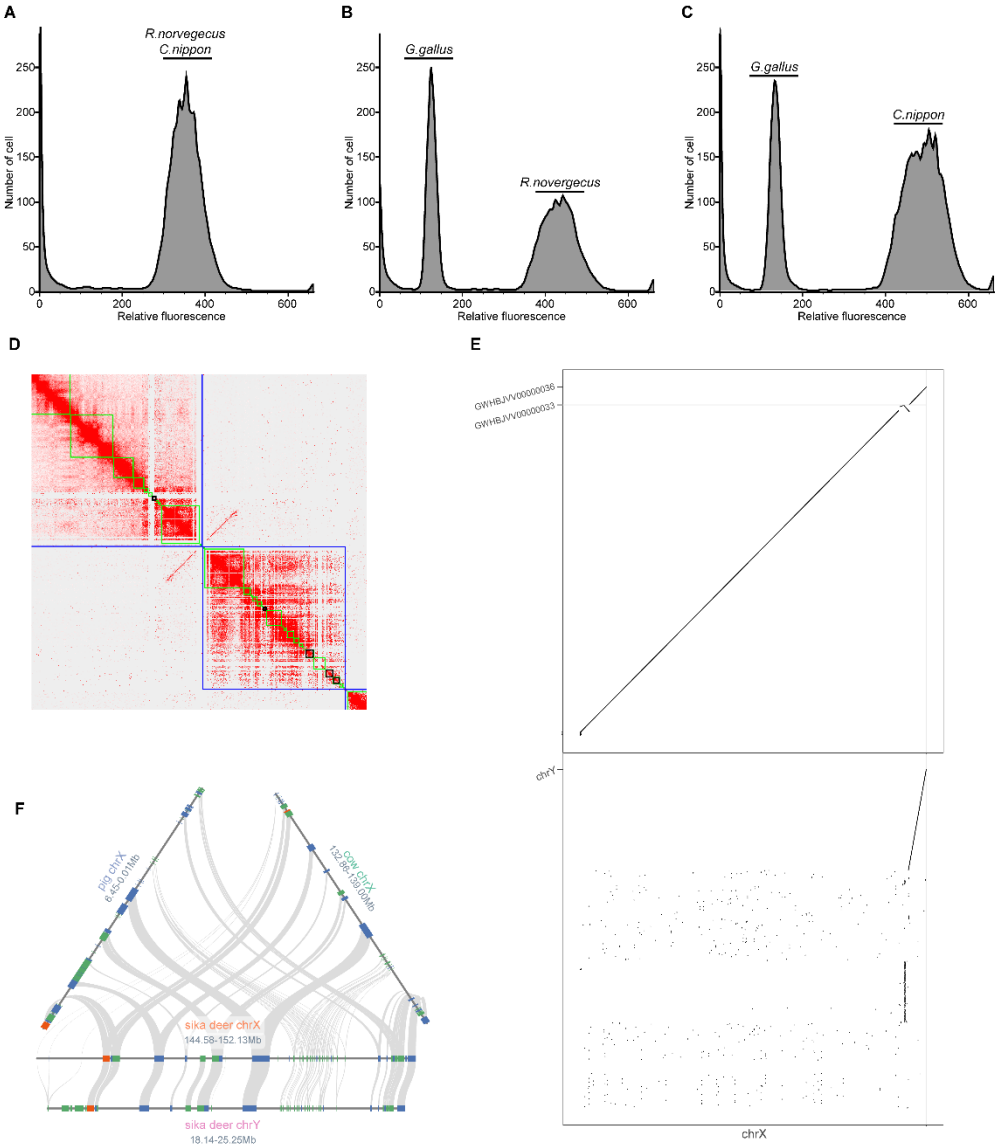
Genomic DNA was extracted from liver using the standard cetyltrimethylammonium bromide method and sequenced using the PacBio in HiFi mode. SMRTbell library construction and sequencing were performed at Novogene (Tianjin, China) following the official protocols of PacBio for preparing ~20-544kb SMRTbell libraries. After obtaining the sequence data, we processed the raw information with SMRT Link (version 8.0546 https://github.com/PacificBiosciences/pbcommand), using the CCS method with default parameters.

Hi-C library preparation and sequencing.

For Hi-C sequencing, we followed the standard protocol described previously with minor modifications. In brief, liver of sika deer was used to isolate cells. The cells were cross-linked and then homogenized by tissue lysis and digested with the restriction enzyme MboI overnight. The proximal chromatin DNA was re-ligated using a ligation enzyme. The nuclear complexes were reversely cross-linked by incubation with Proteinase K at 65°C. DNA was purified using phenol-chloroform extraction, and biotin was removed from non-ligated fragment ends using T4 DNA polymerase. The ends of sheared fragments (300–500 base pairs) were repaired using a mixture of T4 DNA polymerase, T4 polynucleotide kinase, and Klenow DNA

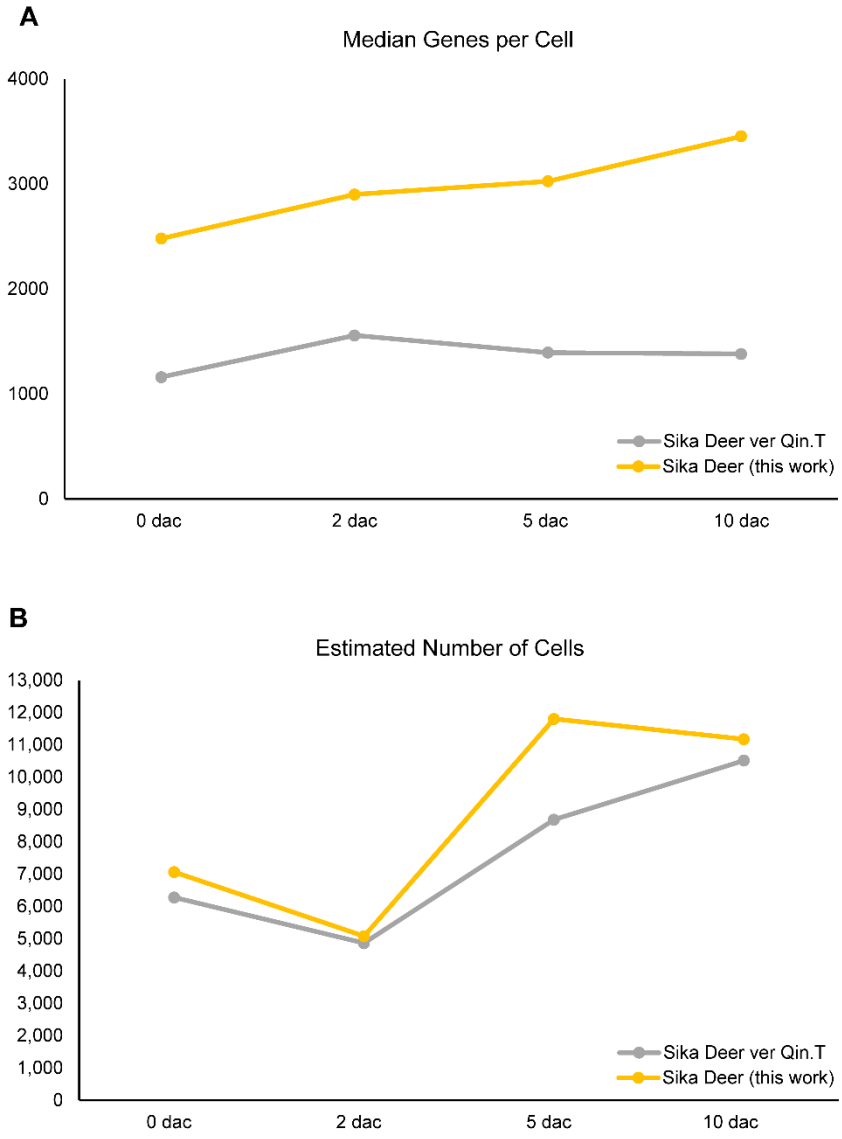
polymerase. Biotin-labeled Hi-C samples were specifically enriched using streptavidin C1 magnetic beads. After adding A-tails to the fragment ends and ligating Illumina paired-end (PE) sequencing adapters, Hi-C sequencing libraries were amplified by PCR (12–14 cycles) and sequenced on the Illumina NovaSeq sequencing platform at Novogene (Tianjin, China).

Supplementary Figures



Supplementary Figure 1. Genome estimate, assembly and evaluation.

The erythrocyte of chicken (*Gallus gallus*) as an internal reference standard and rat (*Rattus norvegicus*) as a positive control were selected to determine the genome size of sika deer. This figure shows the flow cytometric histogram of relative DNA content of nuclei from (**A**) mixture of rat and sika deer, (**B**) mixture of chicken and rat and (**C**) mixture of sika deer and chicken. (**D**) Hi-C heatmap for X chromosome and Y chromosome of sika deer. Green blocks referred to contigs and black blocks referred to contigs with switch error phased by hifiasm. (**E**) Dot plot of genome alignment from previous assembly (upper) and Y chromosome in our assembly to X chromosome in our assembly suggesting that our assembly located PAR regions finely in X and Y chromosome which is lost in previous assembly. (**F**) Synteny of PAR regions between X chromosomes from cow, pig and sika deer and Y chromosomes of sika deer firstly revealed the structure of PAR in Cervidae. Red blocks referred to starting point of PAR.



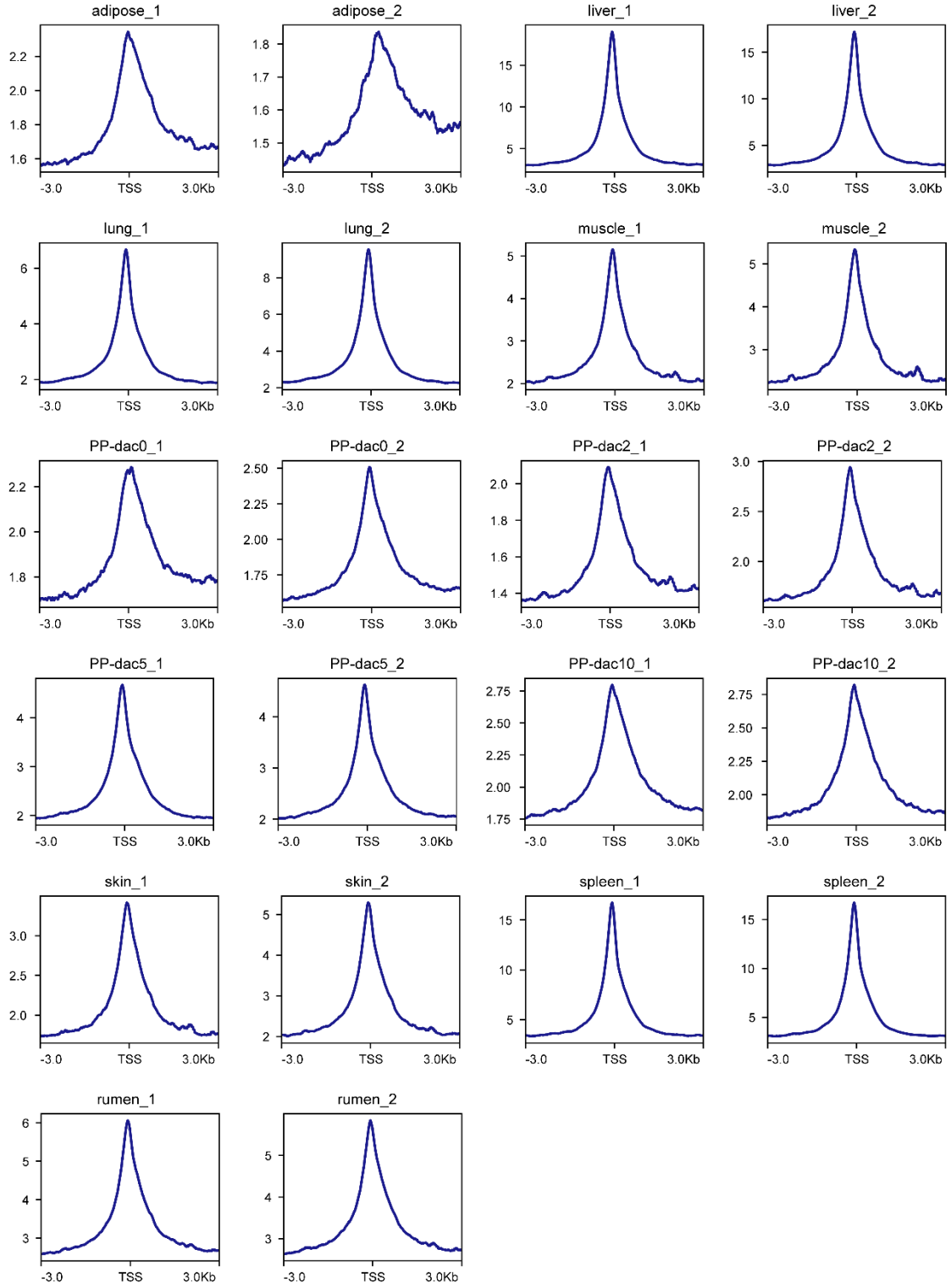
186

187 **Supplementary Figure 2. Single cell RNA-seq quality metrics in different genome**

188 **versions.**

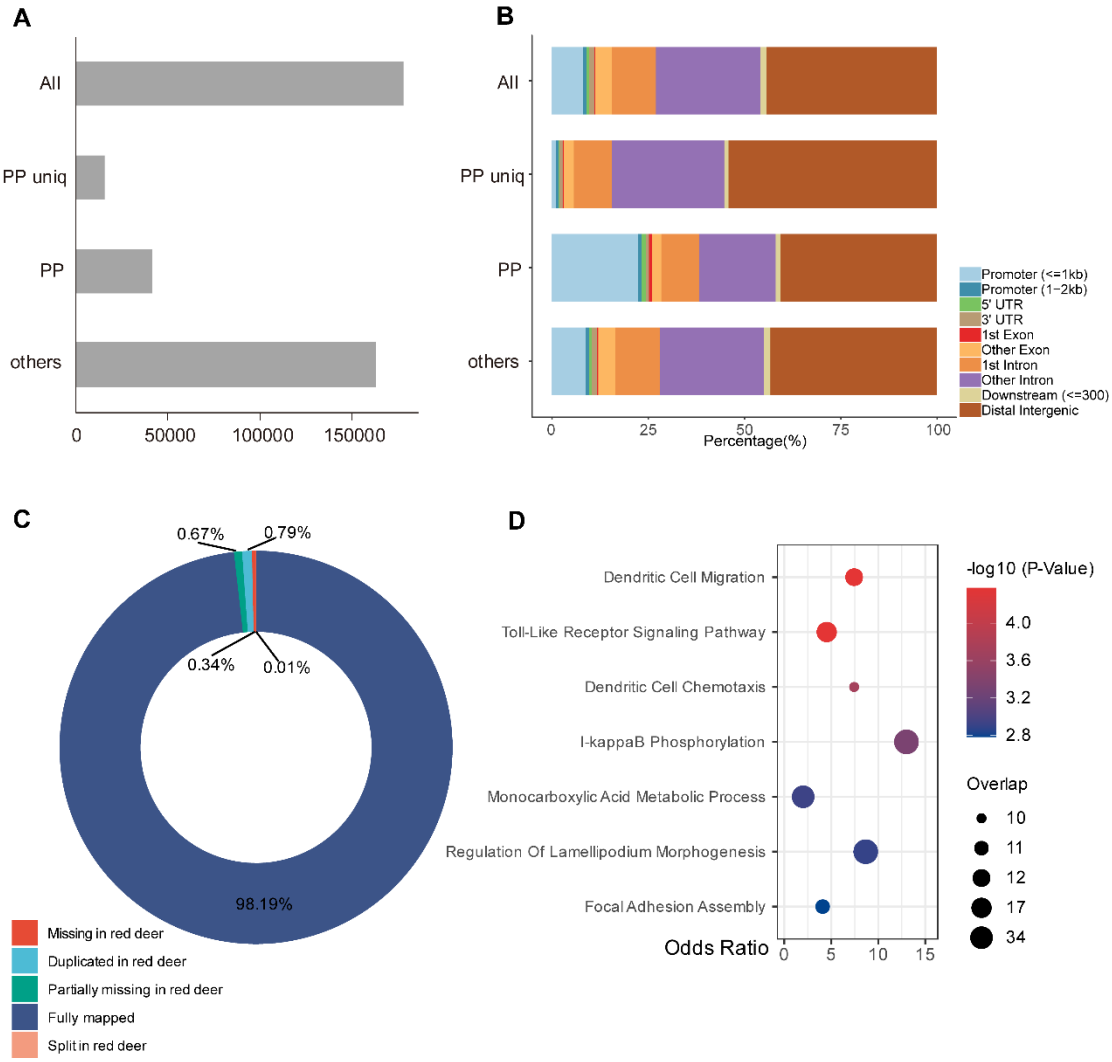
189 Line plots to show median genes per cell (A) and estimated number of cells (B)

190 identified using the same scRNA-seq data but different reference genomes.



Supplementary Figure 3. Quality control of ATAC-seq samples. Line plot to show TSS enrichment analysis of ATAC-seq for each sample suggested high quality of our data.

196



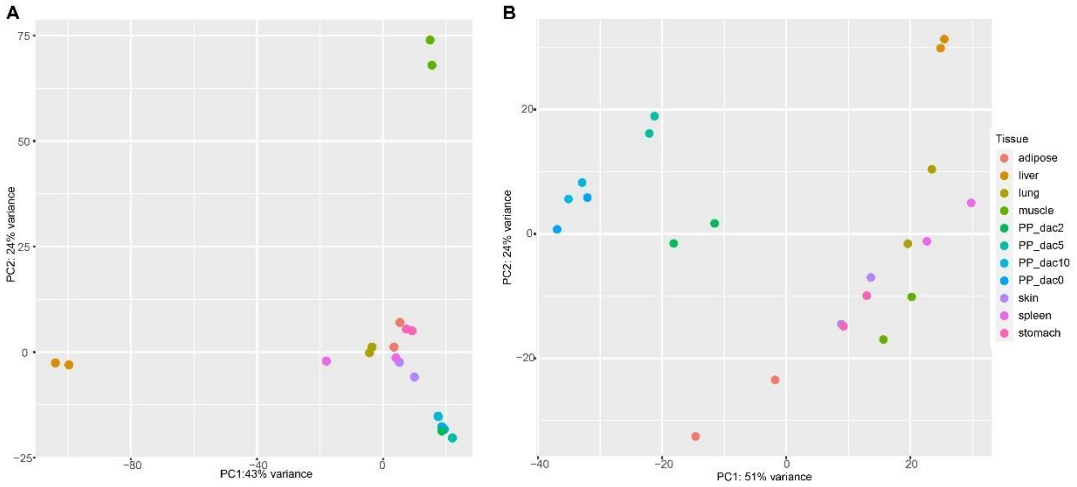
197 **Supplemental Figure 4. Statistic of ATAC-seq peaks.**

198 **A** Bar plot for the number of peaks for all samples and PPs. **B** Bar plot of fraction of

199 different type of peaks for all samples and PPs tissue. **C** Pie plot to show the statistic

200 of the OCRs mapping from sika deer genome to red deer genome. **D** Bubble plot to

201 show functional enrichment of genes related to sika deer-specific OCRs.



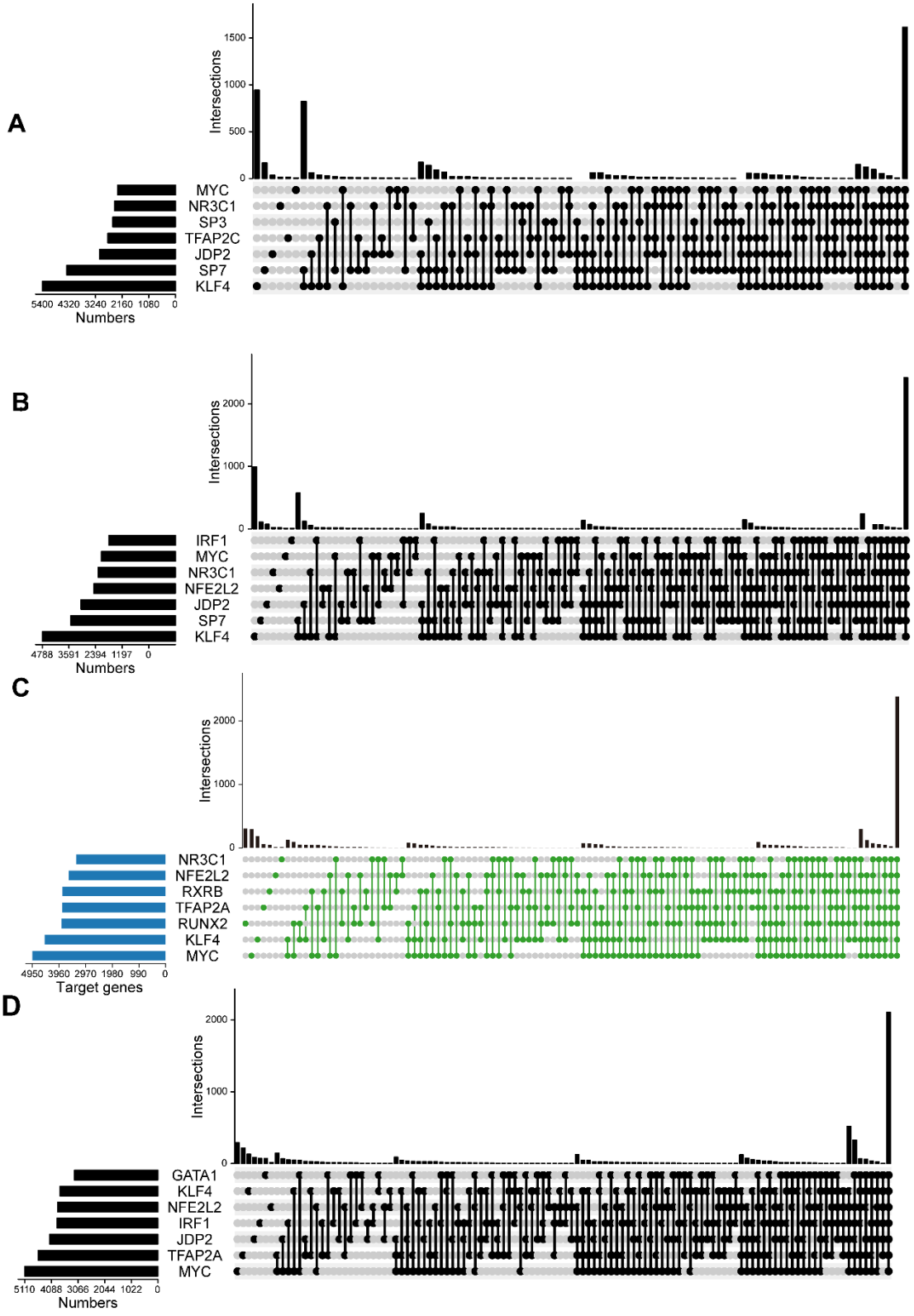
202 **Supplemental Figure 5. PCA analysis for RNA-seq and ATAC-seq of sika deer**

203 **major organ and regenerative antler**.

204 Dot plot to show the PCA analysis of gene expression (**A**) and chromatin accessibility

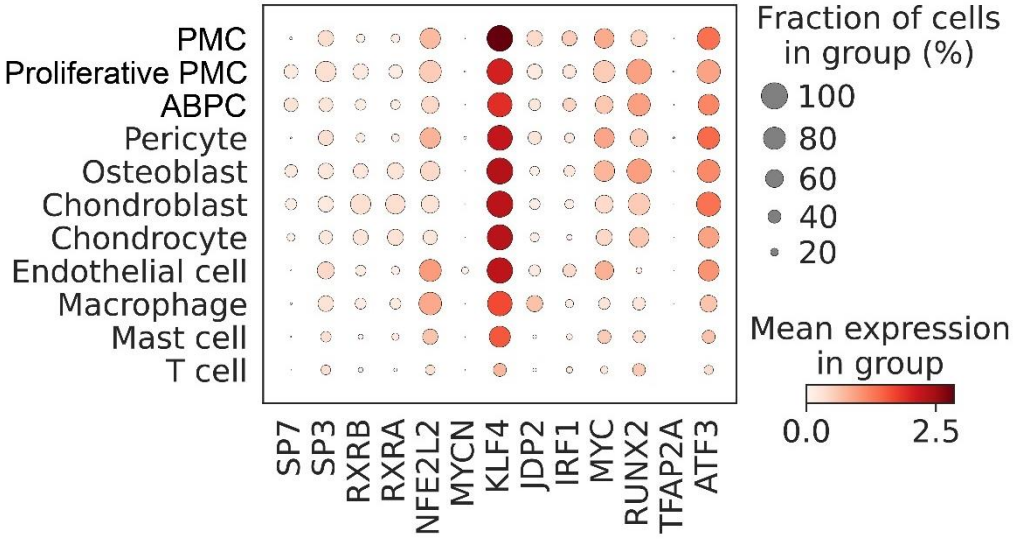
205 (**B**) of sika deer major organ.

206



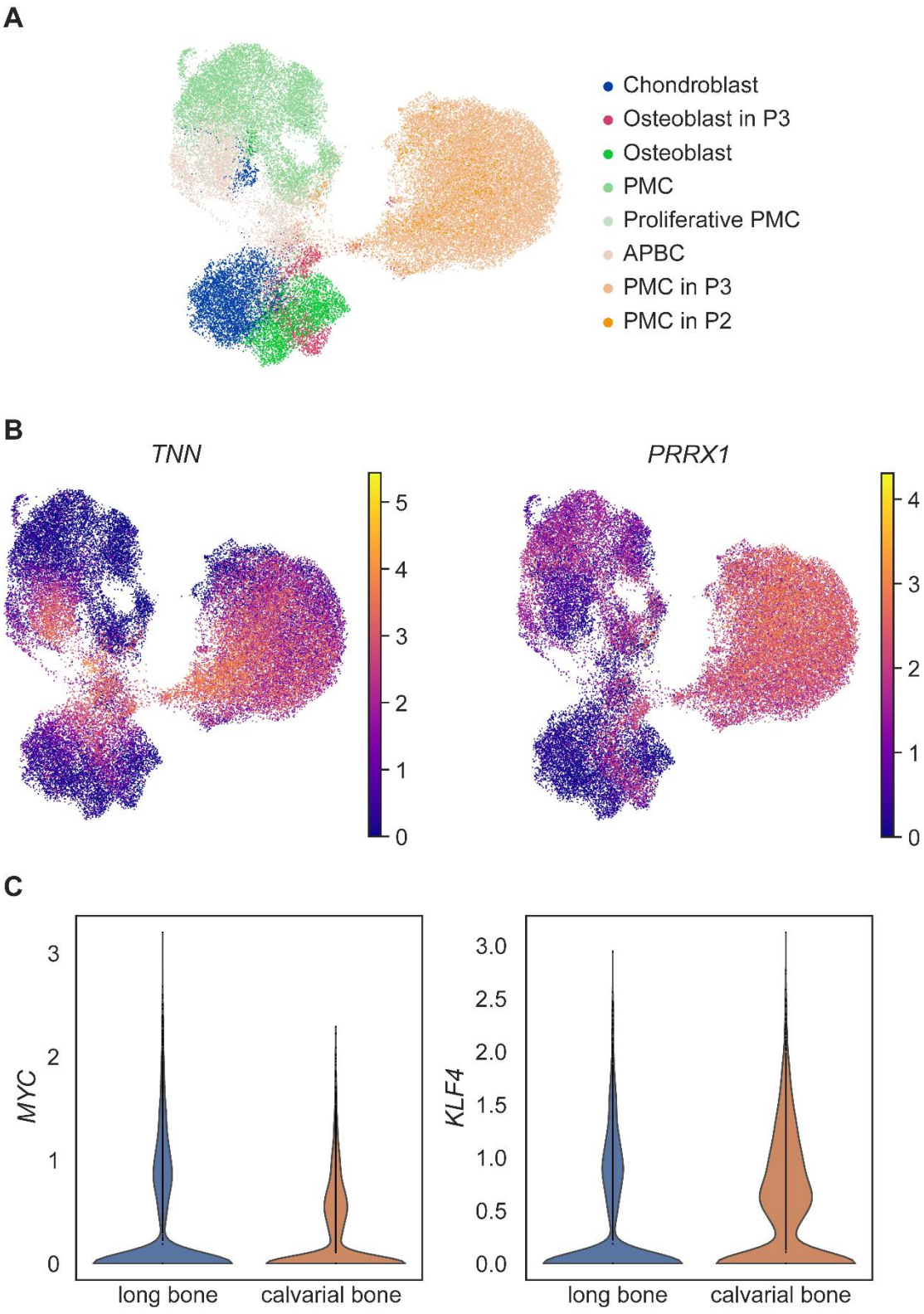
Supplementary Figure 6. Highly overlap between hub TFs during antler regeneration.

Upset plot to show the overlap in target genes between top 7 hub TFs in (**A**) dac0, (**B**) dac2, (**C**) dac5 and (**D**) dac10 during antler regeneration.



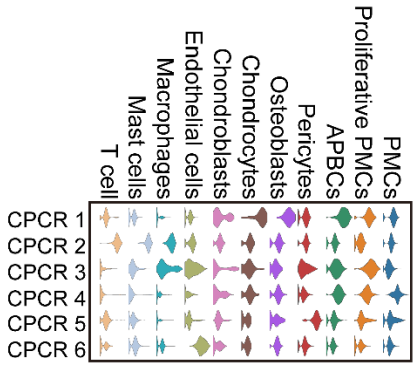
Supplementary Figure 7. Cellular expression of hub TFs

Dotplot of cellular expression profile of hub TFs in antler regeneration cell atlas, show most hub TFs don't show cellular specifity.



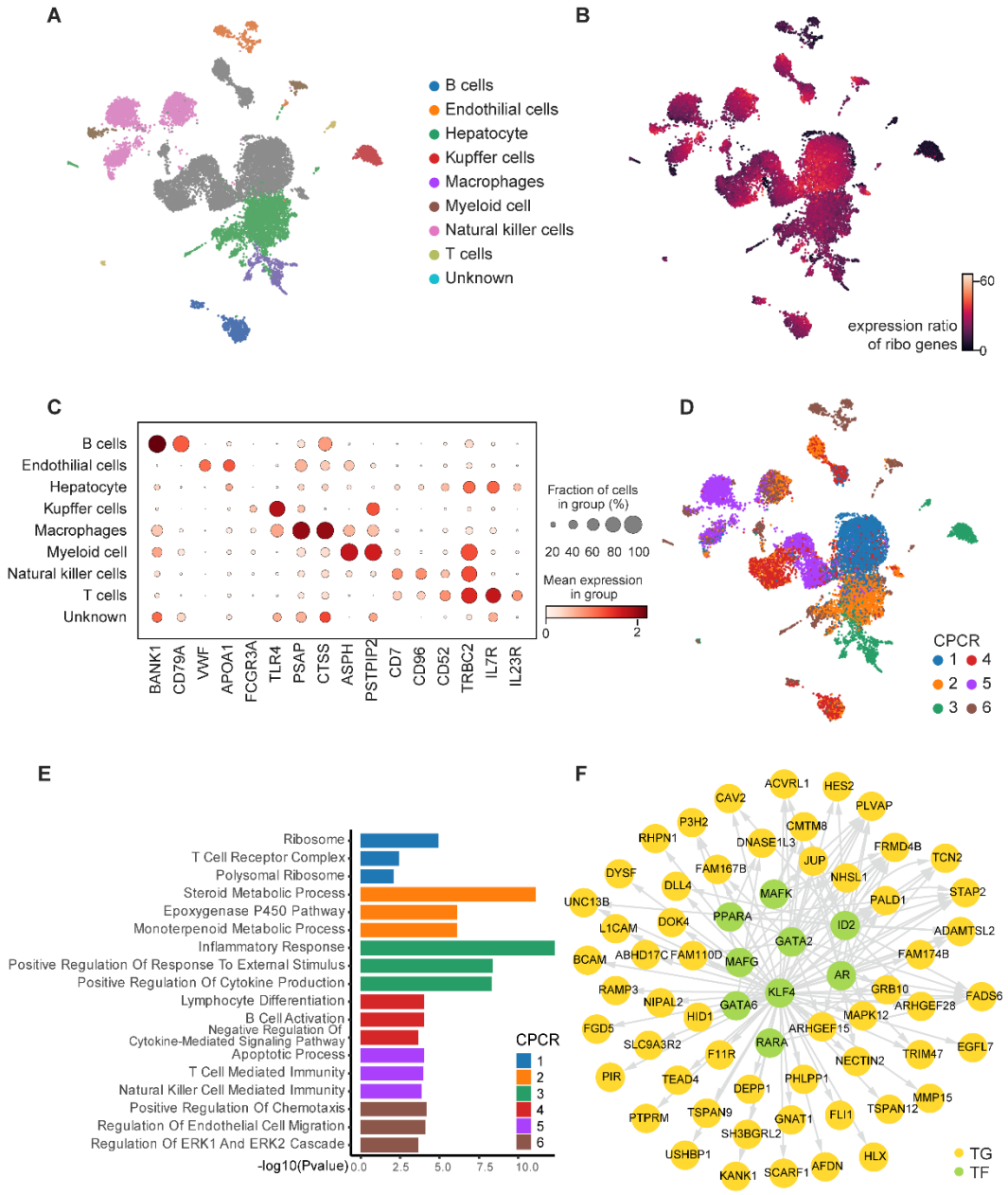
Supplementary Figure 8. Comparison between stem cells population from antler and mouse digit tip.

A UMAP plot of cell type annotation in stem cell population from antler, regenerative mouse digit tip (P3) and unregenerative mouse digit tip (P2). **B** UMAP plot of gene expression of *PRRX1* (the marker of PRRX1+ mesenchymal stem cell) and *TNN* (the marker of antler blastema progenitor cell). **C** Violin plot to show expression of *MYC* and *KLF4* in long bone and calvarial bone data.



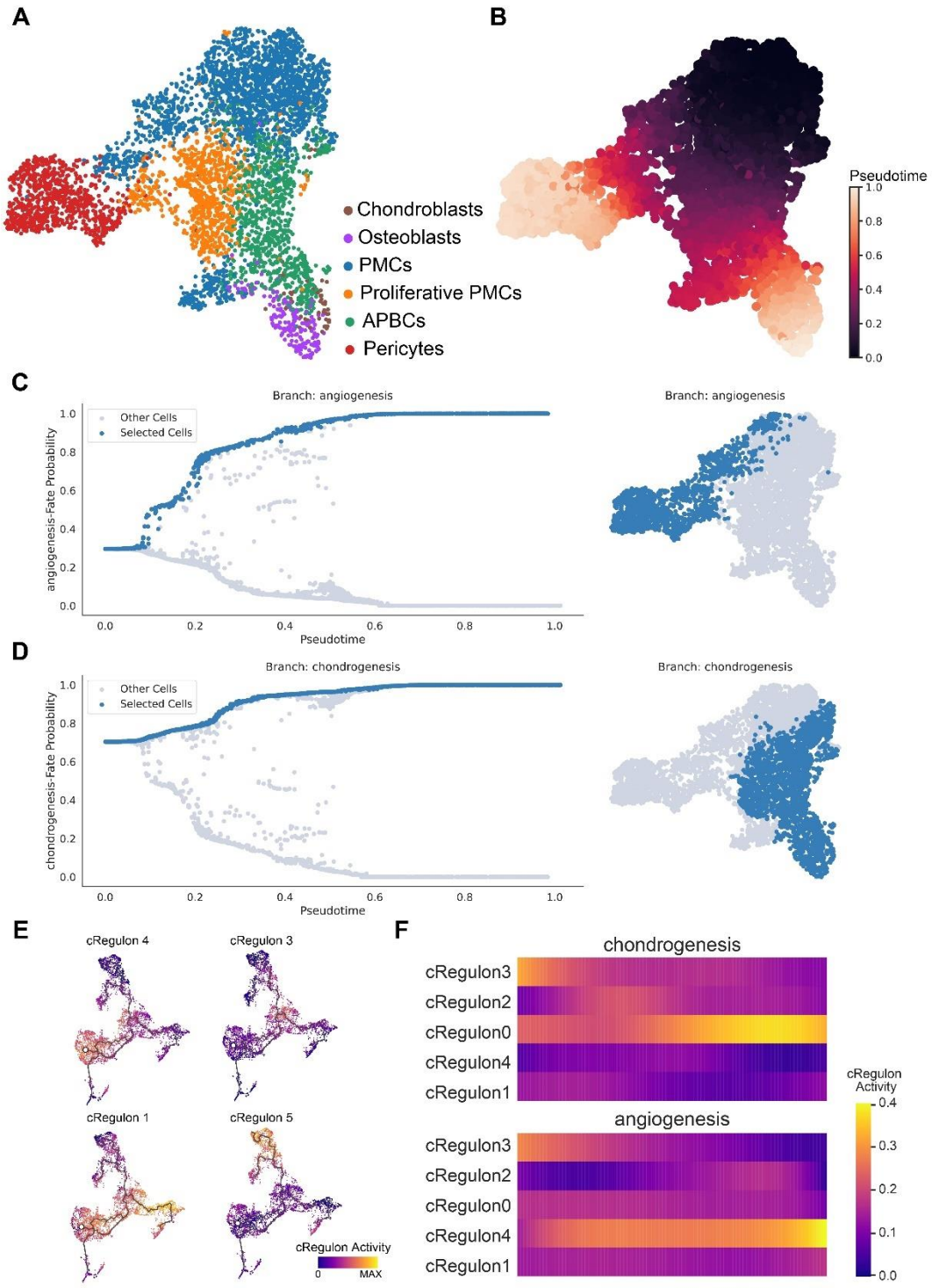
Supplementary Figure 9. CPCR activity in different cell types.

Violin plot to show each CPCR activity in different cell types.



Supplementary Figure 10. cTOP model in liver scRNA-seq data.

A UMAP plot of cell annotations in sika deer liver. **B** UMAP plot to show percentage of ribosome genes expression in a cell. **C** Dot plot to show marker genes expression in each cell type. **D** UMAP plot to show highest CPCR of cells. **E** Functional enrichment for TGs of each CPCR. **F** network of shared TFs and TGs between endothelial cell related programs in liver and antler PP.



Supplementary Figure 11. Pesudotime trajectory analysis of PMCs differentiation.

UMAP of (**A**) cell type and (**B**) pseudotime in dac5 cell atlas analyzed using Palantir. The branching fate to (**C**) angiogenesis and (**D**) chondrogenesis is consisting with the result from Monocle3 (Fig. 4E). The CPCR activity dynamics across trajectories from (**E**) Monocle3 and (**F**) Palantir are similar.