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.)

Abstract

Antler is the only organ that can fully regenerate annually in mammals. However, the regulatory pattern and mechanism of gene expression and cell differentiation during this process remain largely unknown. Here, we obtain comprehensive assembly and gene annotation of the sika deer (*Cervus nippon*) genome. Together with large-scale chromatin accessibility and gene expression data, we construct gene regulatory networks involved in antler regeneration, identifying four transcription factors, *MYC*, *KLF4*, *NFE2L2,* and *JDP2* with high regulatory activity across whole regeneration process. Comparative studies and luciferase reporter assay suggest the *MYC* expression driven by a cervid-specific regulatory element might be important for antler regenerative ability. We further develop a model called cTOP which integrates single-cell data with bulk regulatory networks and find *PRDM1*, *FOSL1*, *BACH1*, and *NFATC1* as potential pivotal factors in antler stem cell activation and osteogenic differentiation. Additionally, we uncover interactions within and between cell programs and pathways during the regeneration process. These findings provide insights into the gene and cell regulatory mechanisms of antler regeneration, particularly in stem cell activation and differentiation.

Introduction

The cyclic renewal of deer antlers, the only known mammal organ that can fully regenerate annually, has long captivated numerous biologists(Li et al. 2014). Each spring, deer shed their hard antlers, and the pedicle scar heals as new bone and cartilage regenerate from the pedicle periosteum. During late spring and early summer, antlers grow and calcify over 3 to 4 months at a rate of 2.75 cm per day, potentially reaching a weight of up to 15 kg by the end of the growth season (Price et al. 2005; Landete-Castillejos et al. 2019). In autumn, antlers rapidly calcify and shed their enveloping velvet skin, leaving an exposed bony surface. The cycle concludes with the casting of the antlers in spring, initiating a new regeneration phase. Researchers have discovered that antler regeneration is driven by stem cells located in the antler pedicle periosteum(Wang et al. 2019a). Casting of the pedicle periosteum results failure of regeneration. Several somatic stem cell markers such as *NT5E*, *THY1* and *ENG* have been found expressed in antler stem cells(Dong et al. 2020). Some researcher found embryonic stem cell factors such as *MYC, KLF4* and *POU5F1* expressed in the regenerative antler stem cell (Dąbrowska et al. 2016), while a single-cell transcriptome study suggested that these factors were specific to antlerogenic stem cells(Ba et al. 2022). Our recent work further unveiled the cell atlas of antler regeneration and elucidated the vital role of antler blastema progenitor cells (ABPCs) differentiated from *PRRX1*$^{+}$ mesenchymal stem cells (PMCs) in the bone regeneration of sika deer antler(Qin et al. 2023). This study also demonstrated the involvement of early developmental pathways, including the WNT, TGF-β,

and FGF pathways, during antler regeneration. Despite these advancements, a comprehensive understanding of the complex gene and cell regulatory dynamics in deer antler regeneration remains elusive.

Gene regulatory networks (GRNs) constructed from transcriptomic and epigenomic data have been widely used to explain complex biological processes such as development and regeneration(Johnston et al. 2021; Jimenez et al. 2022). Recent studies on retinal regeneration have identified key GRNs and regulatory factors, such as *NFIA/B*, which constrain mammalian retinal regeneration(Hoang et al. 2020). However, most of these studies are limited to model species due to the lack of necessary data resources, including high-quality genomic data in non-model organisms. The quality of genomic data significantly impacts multiomic studies(Chisanga et al. 2022). Among Cervidae, the most comprehensive genomic data so far is for red deer (*Cervus elaphus*)(Pemberton et al. 2021), sequenced and assembled by the Darwin Tree of Life Project and annotated using the latest RefSeq pipelines(Blaxter 2022; Pruitt et al. 2014). However, antler regeneration related data are limited and difficult to obtain for red deer in contrast to sika deer (*Cervus nippon*), as many herds of sika deer are reared for antler medicine production in China(Mccullough et al. 2009). Therefore, the sika deer could serve as a valuable model for mammalian regeneration research due to the ease of sampling. Several sika deer genome assemblies derived from short-read or long-read sequencing are available(Han et al. 2022; Xing et al. 2023; Qin et al. 2023; Chen et al. 2019). Most of these studies focused on evolutionary analysis, revealing the genetic basis of unique traits such as high-tannin diet adaptation, rapid antler growth and cancer resistance in sika

deer. These resources have been utilized in transcriptomic research on deer antler(Zhang et al. 2022). However, comprehensive genomic data, including high-quality genome assembly and annotation for sika deer, remain lacking. This gap, combined with the challenges of determining regulatory relationships between regulatory elements (REs) and target genes (TGs) using standard ATAC-seq and histone ChIP-seq methods, have greatly hindered efforts to unravel the gene and cell regulatory mechanisms underlying antler regeneration.

Recently, we developed PECA2 (paired gene expression and chromatin accessibility) to more accurately infer regulatory relationships by integrating paired chromatin accessibility data and gene expression data in mouse and human(Duren et al. 2017). However, this approach relies on extensive paired chromatin accessibility and gene expression data, which are currently unavailable for deer species. In this work, we aim to generate comprehensive omics data for sika deer, including genomic, transcriptomic, and chromatin accessibility data. By leveraging high-quality data and newly developed methods, we seek to gain insights into the multi-scaled regulatory dynamics of antler regeneration.

Results

Genome sequencing, assembly and annotation for sika deer

We first conducted flow cytometry to estimate the sika deer genome size as 3.5 Gb (Fig. S1A-C). Next, we employed multiple sequencing technologies to achieve a high-quality and highly contiguous genome assembly (Fig. 1A). Initially, we obtained 91.52 Gb (31x) of

PacBio HiFi reads and 360 Gb (120x) of Hi-C BGI reads (Supplementary Table1) from a male sika deer. We utilized Hifiasm(Cheng et al. 2022) with these HiFi and Hi-C data, as well as ONT long reads (N50 > 38k) from another individual in a previous study(Qin et al. 2023),to generate well-phased contig assemblies. Finally, we obtained two haplotype assemblies with sizes of 3.1 Gb and 3.0 Gb, respectively. We then employed a two-step haplotype-aware scaffolding strategy to finely phase and scaffold chromosomes for each haplotype using the Hi-C data (see details in Methods). Finally, we anchored the contigs to 66 phased pseudochromosomes (32 × 2 + X, Y) (Fig. 1B) (Supplementary Table 2). Compared to previous assemblies for sika deer (Han et al. 2022; Xing et al. 2023), our assembled haplotype sizes were much larger and closer to the estimated size. The majority of the additional assembly originated from satellite-rich (47.48%) heterochromatin sequences (Supplementary Table 2). The euchromatic contigs scaffolded to pseudochromosomes have a total size of about 2.6 Gb, similar to other ruminant assemblies. Additionally, we found that the X and Y Chromosomes in the previous genome assembly lacked the attachment of the PAR (pseudoautosomal region) (Fig. S1 D-E). In our assembly, the PAR region was fully attached to the X and Y Chromosomes and showed good collinearity with those of other mammals (Fig. S1 F). Compared with cow, the distal end of PAR in sika deer extended to *SHROOM2*, which is more similar to pig(Liu et al. 2019a). Seventeen chromosomes contain no more than three contigs, indicating high continuity. Other quality metrics of our assembly further demonstrated the high quality of our data (Supplementary Table 2).

We annotated 42.80% of repetitive sequences in the genome of sika deer. Notably, the

proportion of satellite sequences (9.96%), was much higher compared to the previous (0.07% - 1.53%) (Supplementary Table 3) sika deer genome assemblies(Han et al. 2022; Xing et al. 2023). Utilizing a range of methods, including RNA-seq and Iso-Seq transcriptome data (Supplementary Table 4), gene projection based on genome alignment, protein homology annotation, and *de novo* annotation, we annotated a total of 22,119 protein-coding genes, including 13 mitochondrial genes. The number of genes annotated was comparable to the RefSeq gene annotation of red deer. Functional annotations were obtained for 20,715 of these genes. The BUSCO completeness score of 99.5%, surpasses all previous sika deer annotations(Han et al. 2022; Xing et al. 2023; Qin et al. 2023) and annotations from published Cetartiodactyla reference genomes such as those of cattle, vaquita, and pig(Talenti et al. 2022; Morin et al. 2021; Warr et al. 2020) (Fig. 1C). Moreover, we identified 35 genes in our annotation that spanned more than 1 Mb, a rarity in previous annotations due to the splitting strategy. Finally, by integrating short-read and long-read transcriptomic data, we annotated 5' UTR for 14,979 genes and 3' UTR for 15,760 genes.

By utilizing our assembly and annotation to reanalyze previous single-cell data, we found a substantial increase in the number of detected cells and genes compared to our earlier version (Qin et al. 2023) (Fig. S2) (Supplementary Table 5). Furthermore, the integration of single-cell quality control using mitochondrial gene data resulted in a more precise cell atlas (Fig. 1D, Fig. 1E). During our quality control process, fibroblast populations with unknown functions identified in our previous study(Qin et al. 2023) were excluded due to their association with low-quality cells characterized by high mitochondrial gene content and low

read counts. We also successfully annotated T cells and mast cells for the first time in the context of antler regeneration. Furthermore, we discovered a proliferative subgroup of PMC (pPMC) that abundantly appeared at dac5 (days after casting), expressing both *TNN* (ABPCs marker) and *ACTA2* (pericytes marker). This suggests that these cells are activated stem cells with the potential for both osteochondrogenesis and angiogenesis. In other words, the pPMCs may represent the intermediate stem cells from the resting cell to the osteochondrogenesis APBCs and angiogenesis pericytes.

Chromatin accessibility landscapes and gene regulation networks for organs of sika deer

To construct regulatory networks involved in antler regeneration, we collected diverse samples to train the PECA2 model. In total, we performed RNA-seq and ATAC-seq sequencing for a total of 32 samples including two replicates for the lung, spleen, liver, adipose tissue, muscle, skin, rumen, and antler pedicle periosteum (PP) at four distinct stages of antler regeneration (Fig. 2A) (Supplementary Table 4,6). The PP samples were the same as those used in previously published single-cell studies on antler regeneration, which encompassed critical stages of antler regeneration. The ATAC-seq samples achieved an average of 47,686,163 unique read pairs, with enrichment observed at transcription start site (TSS) (Fig. S3). Overall, we identified 178,269 non-redundant open chromatin regions (OCRs) in these tissues. Among these, 41,142 OCRs were activated in the PP tissue, with 15,398 OCRs activated specifically in PP (Fig. S4A). Compared to all OCRs, PP-specific OCRs were less abundant in promoter regions, with a larger proportion found in distal

intergenic regions (Fig. S4B). This suggests that PP-specific OCRs are more likely to function as distal elements during regeneration. When projected onto the red deer genome, 98.1% OCRs were fully mapped (Fig. S4C), which is much higher than the ratio of synteny region (89.1%) among the global genomes between the two species, suggesting these identified OCRs are highly conserved and thus may have functional constraint. The 1,808 OCRs that were absent or fragmented in red deer were considered specific to sika deer. Additionally, 4,250 potential target genes of these 1,808 sika deer-specific OCRs identified by Pearson's correlation analysis were enriched in immune-related functions (Fig. S4D). Through hierarchical clustering and PCA analysis, we discovered notable distinctions between the PP and other tissue types, for both the ATAC-seq and RNA-seq datasets (Fig. 2B-C, Fig. S5). Meanwhile, the skin and rumen, both of which have a similar stratified squamous epithelium composition (Pan et al. 2021), clustered together, indicating that our data accurately capture the similarities between these tissues in the sika deer.

To identify regulators driving the variance of chromatin accessibility, we conducted TF (transcription factor) motif enrichment analysis on the peaks from each tissue (Fig. 2D). We identified the TF motifs with the most significant enrichment for each organ and found that these TFs played important roles in the corresponding organs of model organisms, such as HNF4A in the liver(Radi et al. 2023) and MEF2s in the muscle(Taylor and Hughes 2017). The regulatory pattern of PP tissues was similar to that of skin and rumen, contrasting with the results of ATAC-seq and RNA-seq analysis. This similarity was mainly attributed to the binding of AP-1s, KLFs and CNCs, which are TF families involved in cell proliferation and

stimulus response. Meanwhile, the specific motif enrichment of the RUNX family in PP tissue indicates a strong osteogenic potential of PP tissues, which differs from epithelial tissues.

To establish accurate TF-regulatory element (RE) and RE-target gene (TG) regulatory relationships, we used PECA2 to model the regulatory network. These tissue-specific networks contain an average of 6190 TGs and 318 TFs. The TGs of PP-specific REs were enriched in antler-specific genes identified in a previous study(Wang et al. 2019b) (Fisher's exact test, p-value=1.64×10^{-16}), suggesting our model can identify important REs function in gene co-option for antler regeneration. We performed Kleinberg's HITS analysis(Kleinberg 2000) to calculate hub scores which representing significance of TFs in network. We used Enrichr with 'ARCHS4 tissue' database (Xie et al. 2021; Lachmann et al. 2018)to perform enrichment analysis with the top 50 TFs with highest hub scores for each organ. The results for most major organs matched with corresponding human organs (Supplementary Table 7), indicating that our network could accurately identify the important TFs that contribute to tissue specificity. Tissue enrichment of PP tissues in four stages matched with omentum, an organ with strong regenerative potential(Di Nicola 2019) , suggesting that tissues with strong regenerative ability share similar TF regulatory patterns.

Hub TF dynamics in antler regeneration

Hierarchical clustering and TF-binding motif analysis of multi-omics data revealed two main stages of deer antler regeneration (Fig. 2C, D, Fig. 3A). The first stage, encompassing

dac0 and dac2, involved stem cell activity or early development (NFYs, SPs) and stimulus response (CREBs, CNCs, ZNF189, MITF, AP-1s, and sMAFs) as the main biological processes. From dac5 onwards, TF families related to development (RUNXs, NFATs, TEADs, MEF2s) predominated, suggesting that stem cells had activated and initiated developmental processes such as chondrogenesis. These findings align with a previous study on dac5 PP ectopic ossification and confirm that dac5 is a critical turning point for antler regeneration(Qin et al. 2023). We then selected the hub TFs (those with the top 10 highest hub scores) of GRNs at each stage to identify the key TFs in regeneration (Fig3 B). Similar to the TF motif patterns, some hub TFs are shared with other tissues, such as *KLF4* in lung and skin (Supplementary Table 8). Notably, *KLF4*, *JDP2*, *MYC*, and *NFE2L2* are hub TFs across all four stages, suggesting they serve as core TFs throughout the entire regeneration process. The hub TFs identified in the early regeneration stages were primarily associated with stimulus response (*RXRA*, *RXRB*, *NR3C1*), while those appearing later were mostly related to specific developmental processes such as chondrogenesis (*RUNX2*, *SP7*), angiogenesis (*SP3*). We also identified neural crest-related TFs (*TFAP2A*, *TEAD2*), which supports the earlier hypothesis that antler originate from the neural crest (Wang et al. 2019b).

We observed a high degree of overlap among the regulons of hub TFs across all stages (Fig. S6). Similarly, the expression profiles of hub TFs in the cellular atlas of antler regeneration reveal that most hub TFs are broadly expressed across various cell types, with only a few exceptions (Fig. S7). For example, *SP7* is specifically expressed in activated stem cells and osteoblasts. These findings suggest that the identified hub TFs may function

throughout the entire antler regeneration process through combinatorial coordination.

To further elucidate the stage dynamics of the antler regeneration process, we conducted functional enrichment analysis on the shared TGs of these hub TFs (Fig. 3C). The PI3K-AKT signaling pathway and collagen expression-related terms such as Extracellular Matrix Organization were enriched across all four periods. During the dac0 period, there was notable enrichment in Mesenchymal Cell Differentiation, indicating an early stem cell response to antler casting. Both the dac0 and dac2 periods exhibited similarities, with enrichment in terms associated with immune and stress responses, including the TNF signaling pathway, Antigen Processing and Presentation of Exogenous Peptide Antigen, and Wound Healing, indicating a response to wound exposure following antler casting. Significantly, both periods also showed enrichment in Osteoclast Differentiation, aligning with previous studies that suggest osteoclasts drive antler casting(Goss et al. 1992). In the dac5 period, when regenerative APBCs appeared, enrichment was observed in developmental functions such as Skeletal System Development and Embryonic Organ Development. Additionally, the emergence of the Relaxin signaling pathway, known for its role in inhibiting inflammation and fibrosis(Valkovic et al. 2019), marked the transition from an immune response to the initiation of the regenerative process during this period. The subsequent dac10 period showed enrichment in collagen and skeleton development-related terms. In summary, our networks provide a detailed explanation of the antler regeneration process, illustrating the development from antler stem cells to different tissues activating at dac5, driven by the immune response to antler casting.

As regeneration developmental process activated at dac5, we extracted the subnetwork of hub TFs at dac5 (Fig. 3D). This subnetwork has a clear hierarchical structure and consists of the time course pattern. The hub TFs related to stem cell and stimulus response in the initial stages are in the upstream of the subnetwork, while the hub TFs related to developmental processes are in the downstream of the subnetwork. Among the 4 core factors, *MYC* is located at the top of the subnetwork and *KLF4* has the most regulatory connections in the subnetwork, indicating a crucial role in generation activation of these two Yamanaka stem cell factors. Sika deer antler stem cells have shown higher stemness and proliferation stability (over 20 passages) than regular mesenchymal stem cells(Seo et al. 2014; Binato et al. 2013). Moreover, compared to the mouse digit tip regeneration, deer antler can regenerate many more times(Dolan et al. 2022). By comparing the stem cells from deer antler PP, mouse regenerative digit tip (P3), and non-regenerative digit tip (P2) (Fig. 3E), we found different quiescent stem cells from antler and P3 differentiated into similar activated stem cell to initiate regeneration (Fig. S8A, B). Specifically, *KLF4* shows differential expression between antler and P3 (p-value $< 1\times10^{-300}$) by Wilcoxon analysis, while *MYC* is specifically expressed in deer antler (Fig. 3F). Comparative analysis between human long bone and calvaria bone gene expression data(He et al. 2021) shows the different expression is not dominated by the difference between endochondral ossification (P2 and antler) and intramembranous ossification (P3) (Fig. S8C), suggesting that the high expression levels of these two factors may contribute to the strong stemness of deer antler stem cells. Additionally, we examined the evolutionary status of their network-related REs and found a deer-specific open element

(Chr14:13194714-13195323) with a deer-specific insertion upstream of *MYC* (Fig. 3G). This element contains a 7 bp deer-specific insertion in a conserved region, and our ATAC-seq data for different organs showed that it was exclusively open in PP tissue. Our dual luciferase reporter experiment demonstrates that following the cervid-specific insertion, this element exhibits a significantly enhancer effect than the wild-type RE with the pGL4.23 minimal promoter as a control (Fig. 3G). These findings suggest that the emergence of this stronger regulatory element has led to the recruitment of *MYC* in the deer antler stem cells, contributing to their strong stemness.

cTOP modeling uncovers regulation of cellular programs in antler regeneration

Although the hub TFs and their regulons have revealed the dynamics of gene regulatory profiles during antler regeneration, the regulation of cellular program dynamics remains unclear. To address this, we coupled our TF combinatorial regulatory modules identified from bulk GRNs with single cell expression profiles to explore the cell subpopulation heterogeneity. We developed a model called cTOP (combinatorial TF Oriented Program) (see details in Methods), which extracts cellular programs including combinatorial TF module and their specific TGs from single-cell RNA-seq data (Fig. 4A).

We used the cTOP model to analyze the single-cell data at PP dac5, a stage when regeneration has been activated. In total, we identified six Cellular Programs mediated by TF Combinatorial Regulations (CPCRs). Uniform manifold approximation and projection (UMAP) dimensionality reduction revealed that the cell types assigned by the traditional

clustering method (Fig. 4B) closely matched with the cellular program assignments based on the highest CPCR score (Fig. 4C), except for CPCR3 which was consistent with proliferating cells across different cell types. The strong correlation between functional enrichment (Fig. 4D) and cell classification demonstrated the accuracy and sensitivity of our method in extracting cellular programs. Furthermore, our findings revealed that cell types were not always defined by individual CPCRs, but rather by different compositions of multiple CPCRs. For instance, CPCR2 was detected in nearly all cell types (Fig. S9), indicating the involvement of immune responses triggered by antler casting in various biological processes during antler regeneration. Therefore, in our model, cell types are determined by the composition of CPCRs, and cell differentiation can be described as the dynamic interplay of CPCRs. We also applied the cTOP model to liver data. Our model successfully identified metabolism, immune, and angiogenesis programs in the relevant cell types (Fig. S10 A-E). Notably, the programs associated with endothelial cells in both PP and liver tissues shared 9 TFs (*AR*, *ID2*, *GATA2, KLF4, MAFK, MAFG,* and *PPARA*) and 55 TGs tightly related to angiogenesis(Torres-Estay et al. 2015; Sangwung et al. 2017; Lasorella et al. 2005; Dong et al. 2020) (Fig. S10F), suggesting that the cTOP model is robust in identifying similar regulatory programs across different tissues.

Deer antler regeneration is a stem cell-based epimorphic process(Wang et al. 2019a). Within the stromal cell lineage which includes PMCs, pericytes, APBCs, and osteoblasts, we identified five CPCR cell programs: CPCR1 (osteochondrogenesis), CPCR2 (stimulus response), CPCR3 (cell proliferation), CPCR4 (mesenchymal stem cell phenotype), and

CPCR5 (pericyte-related angiogenesis). Pseudotime trajectories demonstrate that mesenchymal stem cells differentiate into pericytes and APBCs through proliferative PMCs (Fig. 4E, Fig. S11 A-D). Throughout this process, the activity of CPCR4 gradually decreases, while CPCR3 is activated. Finally, the dominant program transitions into CPCR1 and CPCR5, respectively (Fig. 4F, Fig. S11 E-F).

At the beginning of differentiation, we found that CPCR4 functioned more as a PMC activation program rather than a PMC resting program, as inferred from the cell annotation. The TFs in this subnetwork include TF families (NFATCs and TEADs) that emerge at dac5 by motif enrichment analysis and regulate the osteochondrogenic TFs (*RUNX2*, *SOX9* and *EN1*) of the dominant program in ABPCs: CPCR1 (Fig. 4G). Core TFs in the CPCR4, including *FOSL1*, *FOS*, *PRDM1*, *BACH1* and *NFATC1* express higher at the first two days after casting, suggesting their early roles in regulating stem cell activation (Fig. 4H). TFs in the AP-1family including *FOS* and *FOSL1* have been found activating stem cells in muscle regeneration, and NFACT1 is crucial for ensuring bone repair and regeneration in skeletal stem cells(Yu et al. 2022) . These TFs may be key factors for stem cell activation by regulating genes in WNT and Hedgehog signaling pathway.

We further found regulatory identified between programs following stem cell activation (Fig. 4J). The proliferation program CPCR3 was regulated by *KLF4*, *ID2/E2Fs*, *CREBs*, and *SMADs* as well as *SP7* from CPCR1 and *SIX2* from CPCR5. Expression of *SIX2* and *SP7* specifically increased along differentiation of angiogenesis and osteochondrogenesis, respectively (Fig. 4I). *SP7* is a well-known master regulator for skeleton development while

SIX2 is important for cranial skeleton development from cranial neural crest stem cell(Liu et al. 2019b). Additionally, CPCR5, the pericyte-dominant program, also regulated the CPCR1, suggesting multifunctional roles for pericytes during antler regeneration. These findings indicate that TFs expressed by the differentiation programs can, in turn, regulate the stem cell proliferation program, and that *SIX2* and *SP7* may determine the fate of stem cell differentiation through this regulatory mechanism.

We also found significant effects of TGF-β pathway on different cellular programs involved in antler regeneration. In a mouse heterotopic ossification model, high levels of *TGFB1* induces the fibroblasts differentiating to chondrogenic progenitor cells(Sorkin et al. 2020). We observed a similar role of *TGFB1* in antler osteochondrogenesis in contrast to *TGFB2/3*. The expression of *TGFB2* and *TGFB3* peaks at the onset of osteochondrogenic differentiation and then decline, while the expression of *TGFB1* increases during this process (Fig. 4F). *TGFB2* and *TGFB3* are crucial for the maintenance of the valvular interstitial cell phenotype, a multipotent cell type in heart valves(Wang et al. 2021). Additionally, *TGFB2* has been reported to inhibit osteogenesis of mesenchymal stem cells, contrasting with the role of *TGFB1*(Li et al. 2022). These results suggest a critical role of *TGFB2/3* in PMCs stemness maintenance by suppressing differentiation. Similar interplays were also found between TGF-β pathway related TFs like *ID2* and *SMADs* in CPCR3. In airway basal stem cells, the *TGF-β-ID2* axis was reutilized to promote tissue regeneration, with overexpression of *ID2* leading to a tumorigenic phenotype(Kiyokawa et al. 2021). SMADs are canonical downstream factors of TGF-β pathway. *SMAD4*, in combination with TGF-β activated *SMAD2* regulates DNA

repairing and cell cycle to inhibit tumorigenesis. We found these factors regulating proliferative genes like *UBE2C* and *TOP2A* together. The coordination between *ID2/E2Fs* and SMADs may keep the balance between regenerative ability and cancer suppression in antler generation.

In summary, our cTOP approach effectively modeled the cellular dynamics of the stem cell differentiation process in antler regeneration. By constructing the subnetwork of CPCRs, we have found key factors operating within specific cellular programs and the interactions between these programs.

Discussion

The multiomic approach has become an efficient method for resolving complex biological processes. However, the application of such methods in non-model species has been hindered by the lack of high-quality omics data. This study generated the high-quality and comprehensive genomic data of sika deer, making it suitable for subsequent multiomic studies. We observed that the genome size estimated for flow cytometry is larger than those from *k*-mer estimation as well as previous assembly based on technologies like ONT and high-throughput sequencing, a discrepancy also noted in other ruminant species such as cows and reindeer(Kent et al. 1988; Goss et al. 1992). Our analysis indicates that this gap is mainly due to the failure of anchoring massive centromeric sequences, a characteristic feature of ruminant genomes. Although we have nearly completed assembling all the euchromatic

regions of the sika deer genome, overcoming the large centromeric regions and achieving a complete telomere-to-telomere genome assembly for ruminant species remains a significant challenge. Fortunately, the large gaps in the ruminant centromeric regions contain few or no genes because almost all mammalian genes have been annotated for the sika deer, and therefore they have limited impact on the downstream functional omics analyses.

Based on the high-quality sika deer genome, we were able to integrate large amounts of omics data including bulk RNA-seq, single cell RNA-seq, and ATAC-seq data of sika deer to reconstruct the tissue-specific gene regulatory networks and resolve cellular dynamics during antler regeneration. The results reveal that antler PP shared some key regulator, such as AP-1 and *KLF4*, with epithelial tissues. AP-1 complexes are prevalent in the regenerative elements of various species, including fruit flies, acoel worms, and bony fish, and are crucial for activating regenerative response enhancers in bony fish(Wang et al. 2020). *KLF4* is essential for the homeostasis and self-renew of epithelial cells in various tissues(Angel et al. 2001; Segre et al. 1999). The recruitment and combination of these TFs may underlie the regenerative potential of antler PP. Several hub TFs, including *MYC*, *KLF4*, *NFE2L2,* and *JDP2*, coordinated the regeneration process from wound healing towards skeletal development. The uncovered cellular regulation dynamics indicated that *FOSL1*, *PRDM1*, and *NFATC1* might drive the stem cell activation and balancing between oncogenic factors *ID2/E2Fs* and anticancer factors *SMADs* in the TGF-β pathway might contribute to the well-organized stem cell proliferation during antler regeneration (Fig. 5e). The proliferation program might also receive potential feedback regulation by *SP7* and *SIX2* from

differentiation programs. These findings for the first time revealed the gene and cell regulatory mechanism of deer antler regeneration.

Antler progenitor stem cells have been characterized as *PRRX1*$^{+}$ mesenchymal stem cells with certain embryonic stem cell characteristics. The involvement of embryonic stem cell-related TFs like *KLF4* and *MYC* in antler regeneration has been a subject of debate. Our study suggested the hub role of *KLF4* and *MYC* in the regulatory networks of antler regeneration, which could not be easily identified through gene expression analysis alone. Cross-species comparisons highlight their high expression might contribute to the high regenerative capacity of antler stem cells. Along with the identification of cervid-specific regulatory elements and the discovery of stem cell phenotype maintenance by *TGFB2/3*, we have preliminarily elucidated the evolutionary and molecular mechanisms underlying stemness in deer antler stem cells, which may provide valuable insights for regenerative medicine studies.

In addition, we proposed the new method, cTOP, in this study to combine our scRNA-seq and bulk ATAC-seq and decode cellular GRNs. Existing gene regulatory network inference methods are not ideally suited for sika deer data. Most methods, such as SCENIC+, GLUE, scReg(Bravo González-Blas et al. 2023; Duren et al. 2022; Cao and Gao 2022) require paired scRNA-seq and scATAC-seq, which is expensive and very sensitive to quality of library preparation. Other methods, such as SCENIC and GRNBoost(Aibar et al. 2017; P assemiers et al. 2022) construct a coarse network using only scRNA-seq. Compared to cTOP, scRNA-seq-only methods, which infer regulatory relationships primarily from co-expression at the single-cell level, may identify many false-positive regulations among TFs and TGs

422 within a co-expression module. The approach of cTOP, decomposing TF modules from bulk
423 GRNs and then coupling with single cell expression profile, offers a cost-effective alternative
424 to obtain highly confident cellular GRNs. Importantly, cTOP fully utilized all data generated
425 in a non-model animal. In future, we will generalize cTOP method to scRNA-seq data only
426 cases and compare with existing methods. Furthermore, cTOP can easily be used in model
427 animal-based development and disease research with even better performance, given the fact
428 that the original PECA2 method was developed with larger and more diverse paired
429 expression and chromatin accessibility data from mouse and human.

430 In conclusion, we have identified the key factors and pathways in stem cell activation,
431 proliferation and differentiation of antler regeneration through cellular GRN modeling based
432 on our high-quality omics data. The gene regulatory mechanism underlying the strong
433 regenerative capacity and delicate balance between regeneration and cancer suppression in
434 cervids will provide new clues for both regeneration and cancer medicine studies.

435

436 **Methods**

437 **Sample collection and genome sequencing.**

438 Four 2-year-old male sika deers (*Cervus nippon*) were used for sampling regenerating
439 antler Tissues, on days 0, 2, 5 and 10 after casting. Another 2-year-old male sika deer was
440 sacrificed for sampling normal organs. Blood from 4 chickens, sika deers and rats was

collected to conduct flow cytometry for genome size estimation (detail in Supplemental Methods).

Genomic DNA was extracted from liver using the standard cetyltrimethylammonium bromide method. For HiFi sequencing, SMRTbell library construction and sequencing were performed at Novogene (Tianjin, China) or BerryGenomics (Beijing, China) following the official protocols of PacBio for preparing ~20-544kb SMRTbell libraries. For Hi-C sequencing, we followed the standard protocol described previously with minor modification, using the same sample with HiFi sequencing(Belton et al. 2012) at Novogene (Tianjin, China). ATAC-seq was performed by standard protocol as previously reported(Buenrostro et al. 2013) at Novogene (Tianjin, China). Details of DNA and RNA library preparation are described in Supplementary Methods, and statistics of all data collected for each bat are provided in Supplementary Table 1 and 3.

Genome assembly.

Hifiasm (v0.19.5) was used to assemble the HiFi reads with ONT reads we generated previously, and Hi-C reads from same individual to generate phased contigs. Then the contigs of each haplotype were merged and scaffolded with Juicer(Durand et al. 2016) and 3DDNA(Dudchenko et al. 2017) to check switch error of phasing (most in sex chromosomes) (Fig.S1 D). Then each haplotig was scaffolded separately to increase scaffolding accuracy with higher Hi-C contact resolution.

Mitochondrial genome was assembled and annotated using MitoHiFi (v3.2.1)(Uliano-

Silva et al. 2023) with HiFi reads.

We assessed genome completeness and consensus quality value (QV) using Merqury(Rhie et al. 2020). Besides, we performed a BUSCO(Waterhouse et al. 2018) assessment of the genome sequences using the certa_odb10 database.

Genome annotation.

RepeatMasker(Tarailo-Graovac and Chen 2009) was first used to detect and mask the repetitive region. Then we integrated different evidence to predict genes. First, we use GeMoMa(v1.9) to do homology-based annotation with annotation of human, cattle, red deer and yarkand deer from NCBI or Ensembl as reference. Second, we processed whole genome alignment to these related species genomes with UCSC chain/net pipeline(Kent et al. 2003) and projected gene annotations using TOGA (v1.1.7)(Kirilenko et al. 2023). Third, we used transcriptome workflow in REAT (v0.6.1 https://github.com/EI-CoreBioinformatics/reat) to integrate short and long transcriptomic data and generate a highly confident gene model. De novo prediction was applied with BRAKER (v2.1.5)(Hoff et al. 2016) in with transcriptomic model as hint. All evidence was integrated using MAKER (v3.01.03)(Campbell et al. 2014).

ATAC-seq data process.

ATAC-seq reads were cleaned by fastp (v0.23.1)(Chen et al. 2018) and then were aligned to the reference genome using Bowtie 2(Langmead and Salzberg 2012). These reads were then filtered for high quality (MAPQ ≥ 13), we also removed reads that were not

properly paired and with PCR duplicates by Picard (version 2.25.7 https://broadinstitute.github.io/picard/). All peak calling was performed with MACS2 (version 2.1.0)(Zhang et al. 2008) using "–call-summits nomodel –shift -100 –extsize 200". Motif enrichment was performed by HOMER(Heinz et al. 2010).

RNA-seq data process.

Short RNA-seq reads were cleaned by fastp and then were aligned to the reference genome with HISAT2(Kim et al. 2019). The mapped reads of each sample were assembled by StringTie (v1.3.3b)(Kovaka et al. 2019).

Iso-Seq reads were preprocessed by ISOSEQ3 (v3.8.0 https://github.com/PacificBiosciences/IsoSeq). The FLNC reads from ISOSEQ3 were mapped to genome with minimap2 and processed with TAMA pipeline.

Single-cell RNA data process.

10x single-cell RNA-seq data were obtained from our previous research(Qin et al. 2023). Cell Ranger (v6.1.1) was used to make prepare reference and count the gene expression profile. Then we used SCANPY (v1.1.10)(Wolf et al. 2018) to process the analysis. First, we use Scrublet (v0.2.3)(Wolock et al. 2019) to mark and remove doublets. Then we further filtered out cells less than 500 counts or have more than 15% mitochondrial gene expression. After preprocessing, we used harmony (v0.0.9) to integrate data of all stages. We used the Leiden method to cluster cells and the Wilcoxon method to identify marker genes.

Pseudotime trajectory inference of dac5 data was applied with Monocle3 (v.0.5.0)(Setty et al. 2019) and Palantir (v1.3.0)(Setty et al. 2019) separately.

Optimization model to identify TF combinatorial programs by cTOP model.

cTOP is a method to infer TF combinatorial program from scRNA-seq data. In brief, a guidance TF-TG regulatory network is first constructed for a given context. Here we use PECA2 model from paired bulk gene expression and chromatin accessibility to construct TF-TG regulatory network. Second, a guidance TF combinatorial network is built from TF-TG network by connection specificity index (CSI) to represent the similarity of regulons between TFs. Then, we apply non-negative matrix factorization (NMF) based optimization model to both the CSI matrix and the single-cell expression matrix for identifying TF combinatorial program and cell embeddings on TF programs. Three essential steps of cTOP are detailed below.

1. Constructing TF-TG regulatory network and estimating TF combinatorial effect

We use the PECA2[12] model to build a TF-TG regulatory network. PECA2 takes paired gene expression (bulk RNA-seq) and chromatin accessibility data (bulk ATAC-seq) as input, with replicant merged for each tissue. The prior data of PECA2, including TF-TG correlation and RE-TG interaction, is calculated from data of multiple deer organs (Supplementary Table S5) for deer specific regulatory network. The output of PECA2 is a TF-TG regulatory

strength matrix, denotated by R matrix, for M TFs and N TGs. Specifically, R_{ij}, which is the i-th row and j-th column of TRS matrix, is the regulatory strength score of the i-th TF on j-th TG.

Then we use the connection specificity index (CSI)(Bass et al. 2013) to assess the combinatorial effect of two TFs. For the i-th and j-th TFs, we use R_i and R_j to represent their regulatory strength scores across all the TGs. Then we calculate the Pearson correlation of their regulatory strength $PCC_{ij} = PCC(R_i, R_j)$. Then CSI score considers the specificity of Pearson correlation to evaluate the combinatorial effect of i-th and j-th TFs:

$$CSI_{ij} = \frac{\#\{l: PCC_{il} \leq PCC_{ij} - \varepsilon, PCC_{jl} \leq PCC_{ij} - \varepsilon\}}{M} \quad (1)$$

Here M was the total number of TFs. ε was a constant with a default value of 0.05. A high CSI score indicated that two TFs specifically regulated a group of TGs.

2. Identifying TF combinatorial program from scRNA-seq data

We use cTOP model to identify TF combinatorial program from single cell RNA-seq data. There are three inputs of the cTOP model: 1) TF combinatorial network represented by above CSI matrix, 2) TF-TG regulatory network represented by above trans-regulatory score matrix R, and 3) single cell gene expression matrix E. We expected the gene expression matrix E to be the coupled with TF combinatorial programs through regulatory network R. Formally, the cTOP optimization model is formulated as follows.

$$\min_{X,H} \|C - XX^T\|_F^2 - \mu_1 C \circ (XX^T) + \mu_2 \|E - WH\|_F^2 - \mu_3 tr(X^T RW)$$

$$s.t. X \geq 0, H \geq 0, \sum_i x_{ik}^2 = 1, k = 1,2,\cdots K; \sum_k x_{ik} = 1, i = 1,2,\cdots M \quad (1.)$$

The cTop model has three components:

1) $\|C - XX^T\|_F^2 - \mu_1 C \circ (XX^T)$: The first two terms are designed to detect TF modules from TF-TF combinatorial effect matrix. The first term decomposes TF-TF combinatorial matrix for detecting TF combinatorial modules and the second term constrains the detected TF modules to be TF combinations with large CSI scores. This component will output variable X, which is a M by K matrix to reveal the combinatorial effect of M TFs in K TF combinatorial programs. We used X to obtain TF modules of TF programs. Given the TFs' combinatorial effect $X_k = (X_{1k}, X_{2k}, \cdots, X_{Mk})^T$ in the k-th combinatorial regulon, we computed the combinatorial effect of i-th TF and j-th TF in k-th TF programs:

$$CE_{ij}^k = X_{ik} \times X_{jk} \tag{2.}$$

We assumed the combinatorial effect of TF pairs in each TF programs followed Gamma distribution. We used threshold P-value≤ 0.01 to select TF pairs for k-th TF programs and the significant TF pairs formed the TF module of k-th TF programs.

2) $\|E - WH\|_F^2$: the third term was to cluster and obtained gene expression programs for scRNA-seq. This component will output matrix W and H. W was a N by K matrix to represent the gene expression program and each column of W indicated the mean expression of TGs regulated by the corresponding TF module. And W was used to obtain the TGs of each TF program by gene expression. H was a K by c matrix to reveal assignment weights of c cells for K TF programs. We assigned each cell to a TF program with the largest assignment weight.

3) $tr(X^T RW)$: The last two terms exerted specificity on TF program by coupling

the TF modules with the gene expression programs through the regulatory network. This component gave constraints to TF modules: the TF modules should regulate TGs that have specific expression in certain cell types/states. This constraint enabled TF programs to utilize not only the specificity of TFs but also the specificity of Res-TGs and TF combinatorial effect to identify core TF combinations for cell type/states.

These three outputs (X, W, H) would be used for describing TF combinatorial programs. We modeled the k-th TF combinatorial program as follows.

$$(X_k(W_{1k} + W_{2k})^T) \circ R \tag{3.}$$

Here X_k and W_k were the k-th column of X and W, respectively. The k-th TF combinatorial program was represented by the TF module defined by X_k with equation (6). The TGs of k-th TF combinatorial program were given by W. The k-th cell state, which was cells governed by k-th TF combinatorial program, was defined by H.

3. Annotating cell clusters with linear combinations of TF combinatorial programs

A cell cluster is a group of cells in scRNA-seq data. For a cell cluster, we suppose this cell cluster has n cells $G = \{g_1, g_2, \cdots g_n\}$. We represent this cell cluster as the TF combinatorial program with the averaged coefficients of all the cells in G, respectively:

$$D = \frac{1}{n}\sum_{i=1}^{n} H_{g_i} \tag{4.}$$

where H_{g_i} is the columns corresponding to the cell g_i. Then the TF combinatorial program combination coefficients of the given cell cluster will be D.

4. Model initiation, parameter selection, and optimization algorithm

To initiate our model, we first solved the component (1) and (2) of our model independently. These three components gave us the initiation of five variable: X^0, W^0, H^0.

The initiation matrix of three variables enabled us to determine the hyper-parameters in our model. There were three hyper-parameters in our model: μ_1, μ_2, μ_3, and K. μ_1, μ_2, μ_3, were parameters to balance the scale of five terms in our model, which could be determined by the initiation matrix:

$$\mu_1 = \|C - X^0 \cdot X^{0T}\|_F^2 / \left(C \circ (X^0 \cdot X^{0T})\right) \quad (5.)$$

$$\mu_2 = \|C - X^0 \cdot X^{0T}\|_F^2 / \|E - W^0 \cdot H^0\|_F^2 \quad (6.)$$

$$\mu_3 = \|C - X^0 \cdot X^{0T}\|_F^2 / tr(X^{0T} \cdot R \cdot W^0) \quad (7.)$$

The hyper-parameter K was the number of TF combinatorial programs, which was consistent with number of TF modules in C and number of cell type/states for single cell data. K could be determined in two ways. First, if we had prior knowledge about the number of TF modules or cell types, we could direct assign this number to K. Second, if we don't have biological insights about the data beforehand, we could try different K to find modules and compute modularity in the TF combinatorial effect matrix to select the best one. And the number of K could also be determined by a method similar to that in Brunet et al.

Starting from the initiation matrices and hyper-parameters, we proposed a multiplicative update algorithm to solve the optimization problem of the cTOP model. We used X_{ij} to represent the element of the i-th row and the j-th column in matrix X and W_{ij} and H_{ij} be the corresponding elements in W and H. We adopted the following update roles and stopped

the iteration when the relative error is less than 0.0001.

$$X_{ij} = X_{ij} \cdot \frac{(4 + 2\mu_1)C \cdot X + \mu_4 R \cdot W}{4X \cdot X^T \cdot X} \quad (8.)$$

$$W_{ij} = W_{ij} \cdot \frac{E \cdot H^T + \mu_3/\mu_2 R^T \cdot X}{W \cdot H \cdot H^T} \quad (9.)$$

$$H_{ij} = H_{ij} \cdot \frac{W^T E}{H^T W H} \quad (10.)$$

Cervid specific structure variation identification.

To check the structure variants in elements, we first used UCSC chain/net pipeline to generate whole genome alignment to rein deer (HlranTar1). Then we used liftOver to map elements from sika deer to rein deer. Then we extract maf file with rein deer as reference from the hal file generated by cactus in Zoonomia project(Genereux et al. 2020). An in-house script called SV_caller was used to find structure variants follow these criterions: 1. Identity of sequence for ingroup species in the SVs is not below 90%. 2. Identity of sequence for all species 50bp flank SVs is not below 50%. 3. Length of SV is over 5bp.

Comparative analysis of OCRs.

To check the evolutionary conservation of OCRs, we first used UCSC chain/net pipeline to generate whole genome alignment to red deer (mCerEla1.1). Then we used liftOver to map elements from sika deer to red deer. Potential target genes were identified with Pearson's correlation test. Functional enrichment was conducted with Enrichr.

Luciferase reporter assay.

The dual luciferase reporter constructs engineered in this study were developed on the pGL4.23[luc2/minP] Vector (E8411，Promega)，pGL4.74[hRluc/TK] Vector(E6911，Promega) which were served as the base plasmid. The Dual Glo Luciferase luminescent assay (E1960，Promega) was carried out in accordance with the manufacturer's protocol with slight modifications. Detailed protocols for transient transfection and stable measurements are described here. After transient transfection in HEK-293T adherent cells, discard the cell culture medium and rinse the cells twice with PBS. Add 120ul 1×passive lysis buffer (E194A, Promega) to each well to lyse the cells. After incubation at room temperature for 10 min, lysates were transferred to 96-well flat-bottomed white polystyrene plates (3912, Corning). Using the automatic loading function of a multifunctional microplate reader (Synergy Neo2, BioTek), add 50ul of firefly luminescence detection solution and 50ul of sea cucumber luminescence detection solution according to 2.5ul of the sample to be tested in each well, and load the sample and detect the final luminescence value.

Data access

All raw sequencing data generated in this study have been submitted to the Genome Sequence Archive of China National Center for Bioinformation (GSA; https://ngdc.cncb.ac.cn/gsa) under accession number CRA018294 (single cell RNA-seq data), CRA018238 (genomic data), and CRA015420 (RNA-seq and ATAC-seq data). Genome

assembly was submitted to NCBI GenBank database with accession number GCA_038088365.1. Genome assembly, gene annotation and network files were also updated to Zenodo database with https://zenodo.org/records/13298156. The Source code is provided with this paper in Supplementary Codes. The source codes and sample data for cTOP are available at https://github.com/AMSSwanglab/cTOP. The codes for structure variation are available at https://github.com/lizihe21/SV_caller.

Competing interest statements

The authors declare no competing interests.

Acknowledgements

This work was supported by two grants of NSFC (Nos. 32030016 and 32220103005) and the New Cornerstone Investigator Program to WW. National Key Research and Development Program of China (2022YFA1004800), CAS Project for Young Scientists in Basic Research, and the National Natural Science Foundation of China (12025107) to YW.

Author contributions

W.W., Y.W., Q.Q. and D.J.H conceived and designed the project. Z.H.L., Z.Y.F., W.W., and Y.W. drafted the manuscripts. Z.H.L., Q.T., B.T.Z., W.W., S.J.J. and L.Z. revised the manuscript. T.Q., G.K.Z. and C.Y.L. collected and prepared the samples, help in assistance

with the experiment. J.R.M., G.Q. and H.S.Y. designed and performed all experiments. Z.H.L. Z.Y.X. Z.Y.F. L.Z. and G.H. performed the data analysis.

References

Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine JC, Geurts P, Aerts J, et al. 2017. SCENIC: single-cell regulatory network inference and clustering. *Nature Methods 2017 14:11* **14**: 1083–1086.

Angel P, Szabowski A, Schorpp-Kistner M. 2001. Function and regulation of AP-1 subunits in skin physiology and pathology. *Oncogene 2001 20:19* **20**: 2413–2423.

Ba H, Wang X, Wang D, Ren J, Wang Z, Sun HX, Hu P, Zhang G, Wang S, Ma C, et al. 2022. Single-cell transcriptome reveals core cell populations and androgen-RXFP2 axis involved in deer antler full regeneration. *Cell Regeneration* **11**.

Bass JIF, Diallo A, Nelson J, Soto JM, Myers CL, Walhout AJM. 2013. Using networks to measure similarity between genes: association index selection. *Nat Methods* **10**: 1169–1176.

Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. 2012. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**: 268–276.

Binato R, de Souza Fernandez T, Lazzarotto-Silva C, Du Rocher B, Mencalha A, Pizzatti L, Bouzas LF, Abdelhay E. 2013. Stability of human mesenchymal stem cells

during in vitro culture: considerations for cell therapy. *Cell Prolif* **46**: 10.

Blaxter ML. 2022. Sequence locally, think globally: The Darwin Tree of Life Project. *Proc Natl Acad Sci U S A* **119**.

Bravo González-Blas C, De Winter S, Hulselmans G, Hecker N, Matetovici I, Christiaens V, Poovathingal S, Wouters J, Aibar S, Aerts S. 2023. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nature Methods 2023 20:9* **20**: 1355–1367.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods 2013 10:12* **10**: 1213–1218.

Campbell MS, Holt C, Moore B, Yandell M. 2014. Genome Annotation and Curation Using MAKER and MAKER-P. *Curr Protoc Bioinformatics* **48**: 4.11.1-4.11.39.

Cao ZJ, Gao G. 2022. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat Biotechnol* **40**: 1458–1466.

Chen L, Qiu Q, Jiang Y, Wang K, Lin Z, Li Z, Bibi F, Yang Y, Wang J, Nie W, et al. 2019. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science* **364**.

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–i890.

Cheng H, Jarvis ED, Fedrigo O, Koepfli KP, Urban L, Gemmell NJ, Li H. 2022. Haplotype-resolved assembly of diploid genomes without parental data. *Nature*

Biotechnology 2022 40:9 **40**: 1332–1335.

Chisanga D, Liao Y, Shi W. 2022. Impact of gene annotation choice on the quantification of RNA-seq data. *BMC Bioinformatics* **23**: 1–21.

Dąbrowska N, Kiełbowicz Z, Nowacki W, Bajzert J, Reichert P, Bieżyński J, Zebrowski J, Haczkiewicz K, Cegielski M. 2016. Antlerogenic stem cells: molecular features and potential in rabbit bone regeneration. *Connect Tissue Res* **57**: 539–554.

Di Nicola V. 2019. Omentum a powerful biological source in regenerative surgery. *Regen Ther* **11**: 182.

Dolan CP, Yang TJ, Zimmel K, Imholt F, Qureshi O, Falck A, Gregory J, Mayes M, Ritchie K, Koester H, et al. 2022. Epimorphic regeneration of the mouse digit tip is finite. *Stem Cell Res Ther* **13**: 1–12.

Dong Z, Haines S, Coates D. 2020. Proteomic Profiling of Stem Cell Tissues during Regeneration of Deer Antler: A Model of Mammalian Organ Regeneration. *J Proteome Res* **19**: 1760–1775.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science (1979)* **356**: 92–95.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**: 95–98.

Duren Z, Chang F, Naqing F, Xin J, Liu Q, Wong WH. 2022. Regulatory analysis of

single cell multiome gene expression and chromatin accessibility data with scREG. *Genome Biol* **23**.

Duren Z, Chen X, Jiang R, Wang Y, Wong WH. 2017. Modeling gene regulation from paired expression and chromatin accessibility data. *Proc Natl Acad Sci U S A* **114**: E4914–E4923.

Genereux DP, Serres A, Armstrong J, Johnson J, Marinescu VD, Murén E, Juan D, Bejerano G, Casewell NR, Chemnick LG, et al. 2020. A comparative genomics multitool for scientific discovery and conservation. *Nature 2020 587:7833* **587**: 240–245.

Goss RJ, Van Praagh A, Brewer P. 1992. The mechanism of antler casting in the fallow deer. *Journal of Experimental Zoology* **264**: 429–436.

Han R, Han L, Zhao X, Wang Q, Xia Y, Li H. 2022. Haplotype-resolved Genome of Sika Deer Reveals Allele-specific Gene Expression and Chromosome Evolution. *Genomics Proteomics Bioinformatics*.

He J, Yan J, Wang J, Zhao L, Xin Q, Zeng Y, Sun Y, Zhang H, Bai Z, Li Z, et al. 2021. Dissecting human embryonic skeletal stem cell ontogeny by single-cell transcriptomic and functional analyses. *Cell Res* **31**: 742–757.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**: 576–589.

Hoang T, Wang J, Boyd P, Wang F, Santiago C, Jiang L, Yoo S, Lahne M, Todd LJ, Jia M, et al. 2020. Gene regulatory networks controlling vertebrate retinal regeneration. *Science* **370**.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**: 767–769.

Jimenez E, Slevin CC, Song W, Chen Z, Frederickson SC, Gildea D, Wu W, Elkahloun AG, Ovcharenko I, Burgess SM. 2022. A regulatory network of Sox and Six transcription factors initiate a cell fate transformation during hearing regeneration in adult zebrafish. *Cell Genomics* **2**: 100170.

Johnston H, Warner JF, Amiel AR, Nedoncelle K, Carvalho JE, Röttinger E. 2021. Whole body regeneration deploys a rewired embryonic gene regulatory network logic. *bioRxiv* 658930.

Kent M, Chandler R, Wachtel S. 1988. DNA analysis by flow cytometry. *Cytogenet Cell Genet* **47**: 88–89. https://pubmed.ncbi.nlm.nih.gov/3356174/ (Accessed February 27, 2024).

Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. 2003. Evolution's cauldron: Duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci U S A* **100**: 11484–11489. https://www.pnas.org/doi/abs/10.1073/pnas.1932072100 (Accessed February 20, 2024).

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome

alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology 2019 37:8* **37**: 907–915.

Kirilenko BM, Munegowda C, Osipova E, Jebb D, Sharma V, Blumer M, Morales AE, Ahmed AW, Kontopoulos DG, Hilgers L, et al. 2023. Integrating gene annotation with orthology inference at scale. *Science (1979)* **380**.

Kiyokawa H, Yamaoka A, Matsuoka C, Tokuhara T, Abe T, Morimoto M. 2021. Airway basal stem cells reutilize the embryonic proliferation regulator, Tgfβ-Id2 axis, for tissue regeneration. *Dev Cell* **56**: 1917-1929.e9.

Kleinberg JM. 2000. *Hubs, Authorities, and Communities*.

Kovaka S, Zimin A V., Pertea GM, Razaghi R, Salzberg SL, Pertea M. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol* **20**: 1–13.

Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, Silverstein MC, Ma'ayan A. 2018. Massive mining of publicly available RNA-seq data from human and mouse. *Nature Communications 2018 9:1* **9**: 1–10.

Landete-Castillejos T, Kierdorf H, Gomez S, Luna S, García AJ, Cappelli J, Pérez-Serrano M, Pérez-Barbería J, Gallego L, Kierdorf U. 2019. Antlers - Evolution, development, structure, composition, and biomechanics of an outstanding type of bone. *Bone* **128**: 115046.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Lasorella A, Rothschild G, Yokota Y, Russell RG, Iavarone A. 2005. Id2 Mediates Tumor Initiation, Proliferation, and Angiogenesis in Rb Mutant Mice. *Mol Cell Biol* **25**: 3563.

Li C, Zhao H, Liu Z, McMahon C. 2014. Deer antler – A novel model for studying organ regeneration in mammals. *Int J Biochem Cell Biol* **56**: 111–122.

Li J, Ge L, Zhao Y, Zhai Y, Rao N, Yuan X, Yang J, Li J, Yu S. 2022. TGF-β2 and TGF-β1 differentially regulate the odontogenic and osteogenic differentiation of mesenchymal stem cells. *Arch Oral Biol* **135**.

Liu R, Low WY, Tearle R, Koren S, Ghurye J, Rhie A, Phillippy AM, Rosen BD, Bickhart DM, Smith TPL, et al. 2019a. New insights into mammalian sex chromosome structure and evolution using high-quality sequences from bovine X and Y chromosomes. *BMC Genomics* **20**: 1–11.

Liu Z, Li C, Xu J, Lan Y, Liu H, Li X, Maire P, Wang X, Jiang R. 2019b. Crucial and Overlapping Roles of Six1 and Six2 in Craniofacial Development. *J Dent Res* **98**: 572–579.

Mccullough DR, Jiang Z-G, Li C-W. 2009. Sika Deer in Mainland China.

Morin PA, Archer FI, Avila CD, Balacco JR, Bukhman Y V., Chow W, Fedrigo O, Formenti G, Fronczek JA, Fungtammasan A, et al. 2021. Reference genome and demographic history of the most endangered marine mammal, the vaquita. *Mol Ecol Resour* **21**: 1008–1020.

Pan X, Cai Y, Li Z, Chen X, Heller R, Wang N, Wang Y, Zhao C, Wang Y, Xu H, et al.

2021. Modes of genetic adaptations underlying functional innovations in the rumen. *Sci China Life Sci* **64**: 1–21.

Passemiers A, Moreau Y, Raimondi D. 2022. Fast and accurate inference of gene regulatory networks through robust precision matrix estimation. *Bioinformatics* **38**: 2802.

Pemberton J, Johnston SE, Fletcher TJ. 2021. The genome sequence of the red deer, Cervus elaphus Linnaeus 1758. *Wellcome Open Res* **6**: 336.

Price JS, Allen S, Faucheux C, Althnaian T, Mount JG. 2005. Deer antlers: a zoological curiosity or the key to understanding organ regeneration in mammals? *J Anat* **207**: 603.

Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, Farrell CM, Hart J, Landrum MJ, McGarvey KM, et al. 2014. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res* **42**.

Qin T, Zhang G, Zheng Y, Li S, Yuan Y, Li Q, Hu M, Si H, Wei G, Gao X, et al. 2023. A population of stem cells with strong regenerative potential discovered in deer antlers. *Science* **379**: 840–847.

Radi SH, Vemuri K, Martinez-Lomeli J, Sladek FM. 2023. HNF4α isoforms: the fraternal twin master regulators of liver function. *Front Endocrinol (Lausanne)* **14**: 1226173.

Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol* **21**: 1–27.

Sangwung P, Zhou G, Nayak L, Chan ER, Kumar S, Kang DW, Zhang R, Liao X, Lu Y,

Sugi K, et al. 2017. KLF2 and KLF4 control endothelial identity and vascular integrity. *JCI Insight* **2**.

Segre JA, Bauer C, Fuchs E. 1999. Klf4 is a transcription factor required for establishing the barrier function of the skin. *Nat Genet* **22**: 356–360.

Seo MS, Park SB, Choi SW, Kim JJ, Kim HS, Kang KS. 2014. Isolation and characterization of antler-derived multipotent stem cells. *Cell Transplant* **23**: 831–843.

Setty M, Kiseliovas V, Levine J, Gayoso A, Mazutis L, Pe'er D. 2019. Characterization of cell fate probabilities in single-cell data with Palantir. *Nature Biotechnology 2019 37:4* **37**: 451–460.

Sorkin M, Huber AK, Hwang C, Carson WF, Menon R, Li J, Vasquez K, Pagani C, Patel N, Li S, et al. 2020. Regulation of heterotopic ossification by monocytes in a mouse model of aberrant wound healing. *Nat Commun* **11**. https://pubmed.ncbi.nlm.nih.gov/32024825/ (Accessed October 16, 2023).

Talenti A, Powell J, Hemmink JD, Cook EAJ, Wragg D, Jayaraman S, Paxton E, Ezeasor C, Obishakin ET, Agusi ER, et al. 2022. A cattle graph genome incorporating global breed diversity. *Nat Commun* **13**.

Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* **Chapter 4**.

Taylor M V., Hughes SM. 2017. Mef2 and the skeletal muscle differentiation program. *Semin Cell Dev Biol* **72**: 33–44.

Torres-Estay V, Carreño D V., San Francisco IF, Sotomayor P, Godoy AS, Smith GJ.

2015. Androgen receptor in human endothelial cells. *J Endocrinol* **224**: R131.

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, Blaxter M, Mieszkowska N, Hall N, Holland P, Durbin R, Richards T, Kersey P, et al. 2023. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics* **24**: 1–13.

Valkovic AL, Bathgate RA, Samuel CS, Kocan M. 2019. Understanding relaxin signalling at the cellular level. *Mol Cell Endocrinol* **487**: 24–33.

Wang D, Berg D, Ba H, Sun H, Wang Z, Li C. 2019a. Deer antler stem cells are a novel type of cells that sustain full regeneration of a mammalian organ—deer antler. *Cell Death Dis* **10**.

Wang F, Zhang C, Kwagh J, Strassle B, Li J, Huang M, Song Y, Lehman B, Westhouse R, Palanisamy K, et al. 2021. TGFβ2 and TGFβ3 mediate appropriate context-dependent phenotype of rat valvular interstitial cells. *iScience* **24**.

Wang W, Hu CK, Zeng A, Alegre D, Hu D, Gotting K, Granillo AO, Wang Y, Robb S, Schnittker R, et al. 2020. Changes in regeneration-responsive enhancers shape regenerative capacities in vertebrates. *Science (1979)* **369**.

Wang Y, Zhang C, Wang N, Li Z, Heller R, Liu R, Zhao Y, Han J, Pan X, Zheng Z, et al. 2019b. Genetic basis of ruminant headgear and rapid antler regeneration. *Science (1979)* **364**.

Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, Chow W, Eory L, Finlayson HA, Flicek P, et al. 2020. An improved pig reference genome sequence to enable pig

genetics and genomics research. *Gigascience* **9**.

Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva E V., Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* **35**: 543–548.

Wolf FA, Angerer P, Theis FJ. 2018. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol* **19**: 1–5.

Wolock SL, Lopez R, Klein AM. 2019. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**: 281-291.e9.

Xie Z, Bailey A, Kuleshov M V., Clarke DJB, Evangelista JE, Jenkins SL, Lachmann A, Wojciechowicz ML, Kropiwnicki E, Jagodnik KM, et al. 2021. Gene Set Knowledge Discovery with Enrichr. *Curr Protoc* **1**.

Xing X, Ai C, Wang T, Li Y, Liu H, Hu P, Wang G, Liu H, Wang H, Zhang R, et al. 2023. The First High-quality Reference Genome of Sika Deer Provides Insights into High-tannin Adaptation. *Genomics Proteomics Bioinformatics* **21**: 203–215.

Yu F, Li F, Yu P, Zhou B, Ye L. 2022. Identification and characterization of NFATc1+ skeletal stem cells in bone regeneration. *Cell Rep* **41**.

Zhang R, Dong Y, Xing X. 2022. Comprehensive transcriptome analysis of sika deer antler using PacBio and Illumina sequencing. *Sci Rep* **12**: 16161.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: 1–9.

885

886

Figures lengends

Figure 1 Assembly and annotation of the sika deer genome.

A Workflow of phased genome assembly of sika deer. **B** Hi-C interaction heatmap of phased pseudochromosomes of sika deer genome. **C** Bar plot of BUSCO evaluation of published genome annotations of sika deer and Cetartiodactyla species in RefSeq. **D** UMAP of cell atlas of antler regeneration based on our genome assembly and annotation. **E** Dot plot of cell marker expression profiles in each cell type. Dot size represent the proportion of cells expressing gene in a cluster.

Figure 2 Chromatin accessibility and gene expression landscape for different organs of sika deer.

A Schematic drawing of the study design covering primary organs and key stages of antler regeneration and collecting paired expression and chromatin accessibility data. **B**, **C** Hierarchical clustering heatmap of gene expression and chromatin accessibility of sika deer organs. **D** Hierarchical clustering heatmap of TF-binding motif match profile showing similar motif profile among self-renewable tissues.

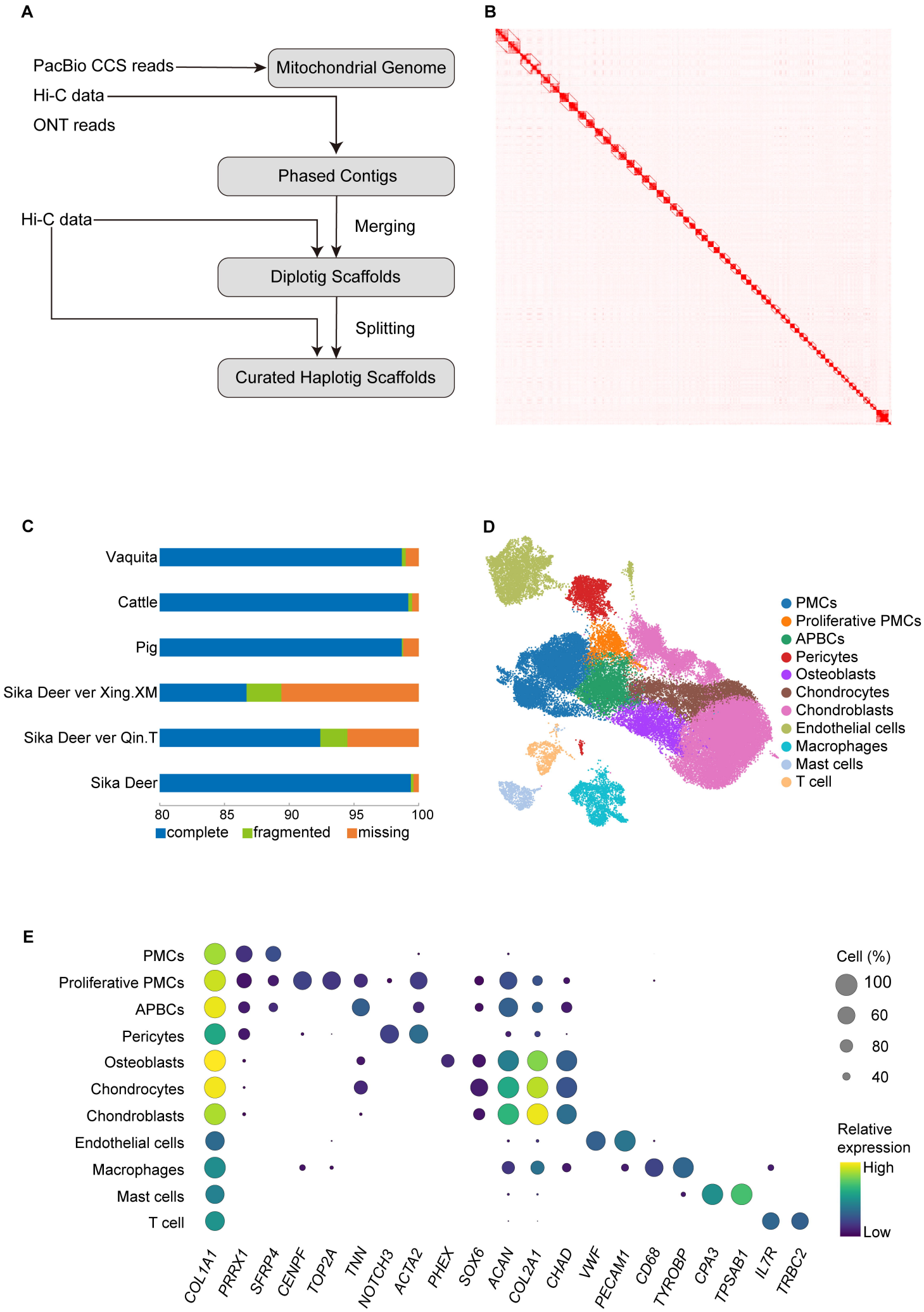
Figure 3 Hub TFs in antler regeneration.

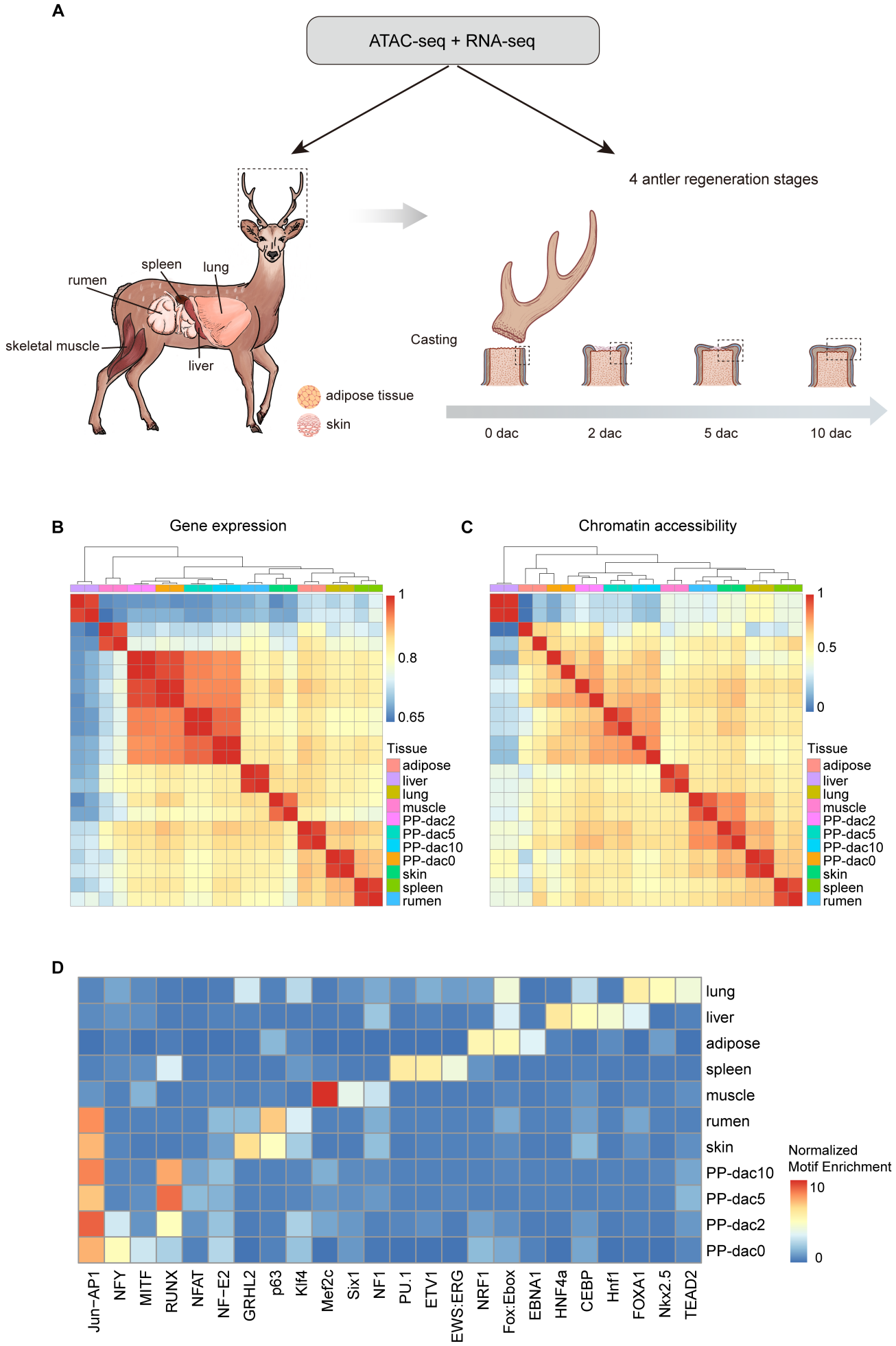
A Hierarchical cluster heatmap of the top 5 enriched motifs for each stage in antler regeneration. **B** Function and time course schema of top 10 TFs for each stage, suggesting 4 core TFs and the dynamic from cell stemness and stimulus response to development. **C** Functional enrichment of combinatory regulons of hub TFs in each antler regeneration stage. **D** Hierarchical structure of 10 hub TFs subnetwork in antler regeneration at dac5 with similar pattern with time course schema. **E** UMAP projection of PMC cell lineage from antler and mouse digit tip distinguishes PMC of antler and mouse. Dash circle highlighted the shared activated stem cell in antler and regenerative digit tip. **F** Higher *KLF4* and *MYC* expression in PMC of deer antler than mouse digit tips. **G** Genome track of antler specific element nearby *MYC* (up), sequence alignment (lower left) and luciferase assay (lower right) suggest cervid-specific insertion has significantly increased the expression of *MYC* in antler.

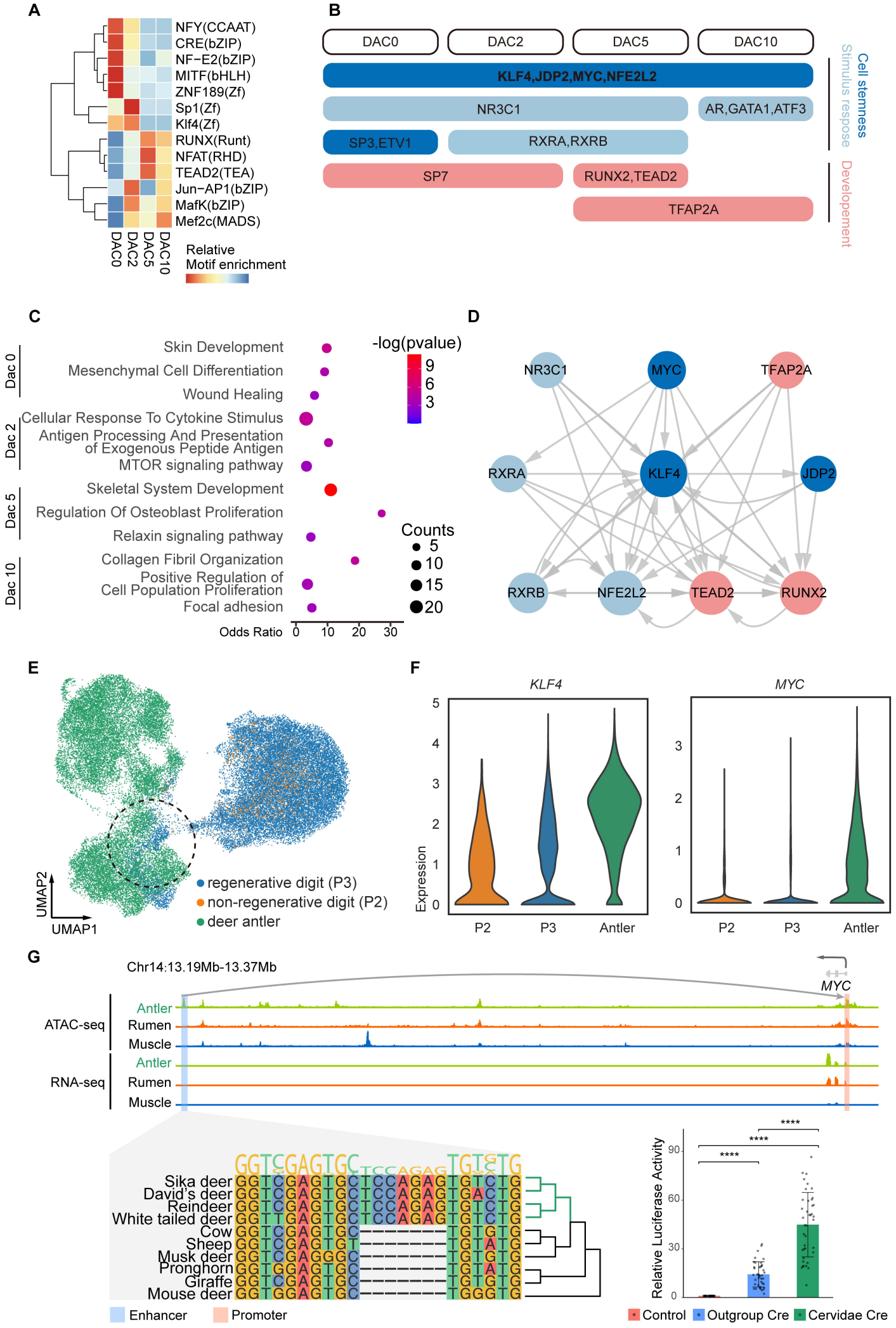
Figure 4 cTOP models cell programs in antler regeneration.

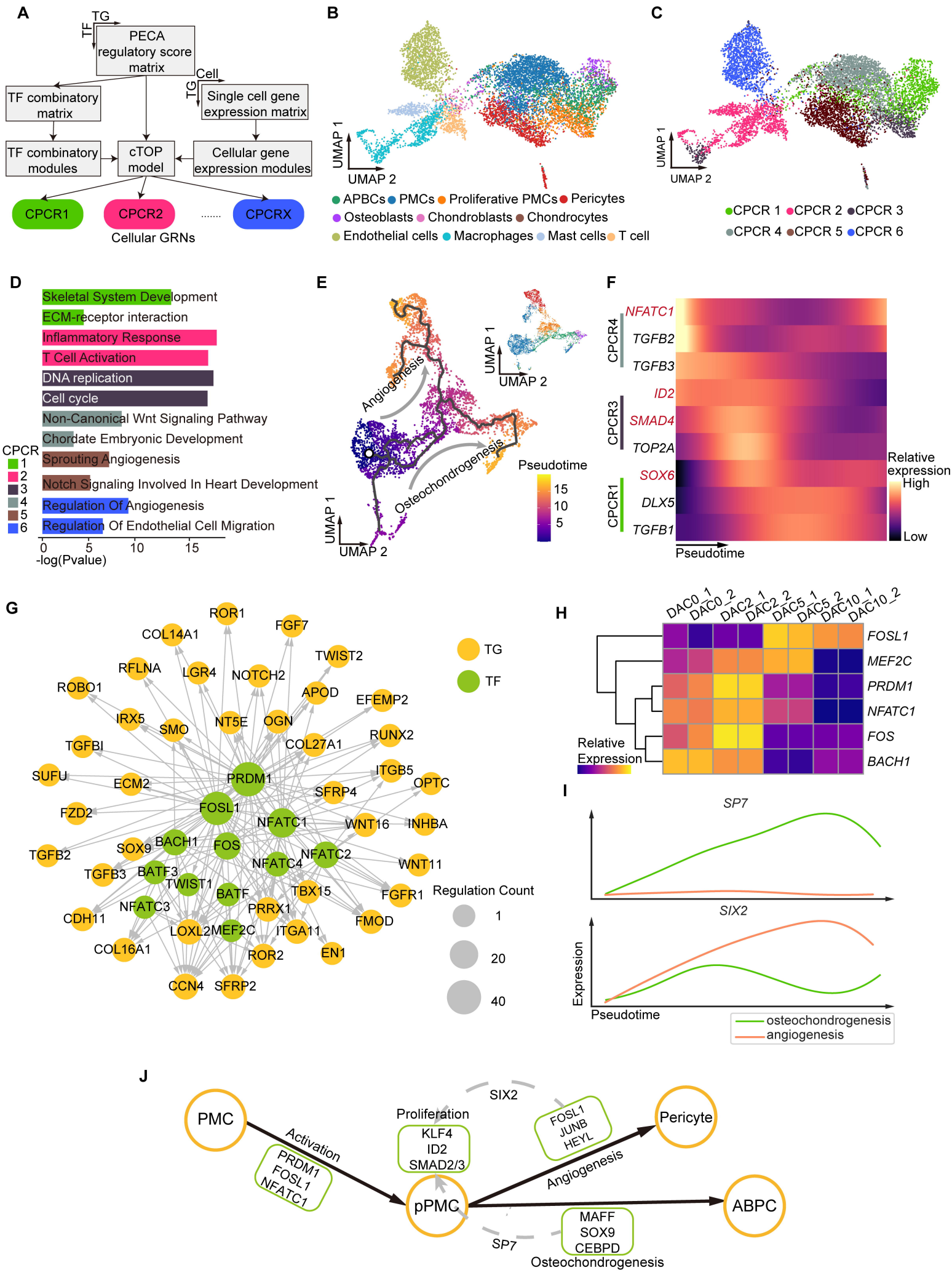
A cTOP couples cell expression modules and TF combinatory modules under the PECA regulatory network to model cellular regulatory programs. **B, C** UMAP of cell types and highest CPCR in cell atlas at dac5. **D** Functional enrichment for TGs of each CPCR. **E** UMAP of pseudotime trajectory of stromal cells at dac5 by Monocle3. **F** cTOP -relative TFs (red) and genes (black) expression dynamics across the osteochondrogenesis trajectory. **G** Network of CPCR4 suggests *PRDM1*, *FOSL1* and *NFATC1* as the key factors for stem cell activation in antler regeneration. **H** Expression dynamics of TFs in CPCR4 during antler regeneration

suggest their early function in antler regeneration. **I** Expression dynamic across differentiation of *SP7* and *SIX2* showing their program specific expression. **G** Schematic regulatory model of cellular programs in antler stem cell differentiation. Black arrows represent differentiation, and gray dashed arrows represent regulation.




A
ATAC-seq + RNA-seq
4 antler regeneration stages
rumen
spleen
lung
skeletal muscle
liver
adipose tissue
skin
Casting
0 dac
2 dac
5 dac
10 dac
B
Gene expression
1
0.8
0.65
C
Chromatin accessibility
1
0.5
0
Tissue
adipose
liver
lung
muscle
PP-dac2
PP-dac5
PP-dac10
PP-dac0
skin
spleen
rumen
D
lung
liver
adipose
spleen
muscle
rumen
skin
PP-dac10
PP-dac5
PP-dac2
PP-dac0
Normalized
Motif Enrichment
10
0
Jun−AP1
NFY
MITF
RUNX
NFAT
NF−E2
GRHL2
p63
Klf4
Mef2c
Six1
NF1
PU.1
ETV1
EWS:ERG
NRF1
Fox:Ebox
EBNA1
HNF4a
CEBP
Hnf1
FOXA1
Nkx2.5
TEAD2





GENOME RESEARCH

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

Zihe Li, Ziyu Xu, Lei Zhu, et al.

Genome Res. published online November 14, 2024
Access the most recent version at doi:10.1101/gr.279448.124