2023

Pioneer Factors Compete for Epigenetic Factors in Switching Stem Cell Fates

Yihao Yang

PIONEER FACTORS COMPETE FOR EPIGENETIC FACTORS IN SWITCHING STEM CELL FATES

A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfillment of the Requirements for
the degree of Doctor of Philosophy

by
Yihao Yang
June 2023

PIONEER FACTORS COMPETE FOR EPIGENETIC FACTORS IN SWITCHING STEM CELL FATES

Yihao Yang
The Rockefeller University 2023

During development, progenitor cells can activate one cell fate while simultaneously silencing another, a process that is tightly regulated in adult tissues. However, this process is often derailed in diseases such as cancers, leading to uncontrolled growth and malignancy. At the crossroads of fate-switching, pioneer factors are a class of transcription factors equipped to bind cognate motifs in closed chromatin. Once they access the closed chromatin, pioneer factors can either act as transcriptional activators or repressors of cell fates by recruiting co-activators or co-repressors. Nevertheless, whether and how a single factor can simultaneously silence the old fate while activating the new cell fate remains largely unknown.

Here, I tackled this question with SOX9, a master regulator that diverts embryonic epidermal stem cells (EpdSCs) into becoming hair follicle stem cells (HFSCs). I triggered a temporal fate-switching by engineering mice to re-activate SOX9 in adult EpdSCs. Combining epigenetic, proteomic, and functional analyses, I interrogated the ensuing transcriptional and epigenetic dynamics, slowed temporally by the mature EpdSC niche microenvironment. My findings demonstrate that SOX9 plays a dual role in the fate-switching process. As it binds and opens key HFSC enhancers, it also redistributes co-factors away from EpdSC enhancers, leading to their silencing. Furthermore, in the absence of its normal regulation, prolonged expression of SOX9 in EpdSCs leads to the activation of downstream transcriptional regulators associated with basal cell carcinoma.

To C. David Allis, Ph.D.
for illuminating the path to epigenetic understanding through his inspirational mentorship

ACKNOWLEDGMENTS

As I conclude my PhD journey, I am overwhelmed with deep gratitude and emotion for all who have provided me with mentorship and guidance, and those who embraced me with encouragement. The journey to completing my PhD has been a challenging one as a first-generation immigrant in the midst of a global pandemic, but it has also been an immensely rewarding experience thanks to the individuals who have supported me scientifically and in life at large. I could only have achieved this far with you.

First and foremost, I want to express my heartfelt gratitude to my advisor, Elaine Fuchs, who has given me the freedom to pursue my interests in epigenetics and stem cells and provided unparalleled support and guidance to ensure my success as an independent scientist. Elaine has inspired my scientific career beyond direct mentorship, but also as a role model through her diligent dedication, boundless curiosity for science, and the spirit of understanding every biology "more than skin deep."

Science, any science, would not be possible without collaboration. I want to acknowledge my closest collaborator and a great mentor, Nick Gomez. All the results during my graduate career were only possible because of him. From the beginning of my graduate school, I worked closely with Nick to learn from the basic experimental designs and analysis to how to think about epigenetics. He nurtured me to become the independent and collaborative scientist I am today through his encouragement and countless discussions.

Elaine has always urged her trainees to reflect on the "big picture questions" related to our research. I reflect on my graduate school journey, and the biggest picture solution I acquired is my friends and collogues that have supported me boundlessly throughout my time at Rockefeller. I

am forever grateful to the members of Fuchs Lab, Deniz, Anita, Martina, Leo, Nicole, Irina, Inwha, Hanseul, Chris and Raj, for your friendship and many inspiring scientific conversations.

I would also like to thank the core facilities and the dean's office at Rockefeller University. Stanka, Samer, Songyan and Svetlana from FCRC provided tremendous help and support with our FACS experiments. Christine, Jackie, Bin and Connie from GRC attended our sequencing requests closely. Henrik from the proteomics core assisted our mass-spec experiments. And finally, everyone from the dean's office offered constant support throughout my graduate program years.

Beyond Rockefeller, my interest in science was first spurred by Li Diao and Wen He, excellent teachers who directed me to become a scientist. Furthermore, my mentors in college, Anna Sahakyan, Kathrin Plath, and Ira Clark, gave me the confidence and knowledge to pursue graduate school.

Finally, I thank my family and friends outside academia. To my parents Hao and Beilei, who always believed in me and provided everything they could to support any of my life decisions. To my friends Simon Marin and Felix Xie, who gave me the confidence to be who I am. And to my chosen family Michael McKeever, you are behind everything I have achieved and will pursue. I hope I made all of you proud.

Scientific contributions

The work presented in this thesis report involves my original works but also contributions from my close collaborators. Here, I would like to address the efforts of Nick Gomez (N.G.), Nicole Infarinato (N.I.), Rene C. Adam (R.C.A.), Megan Sribour (M.S.), Inwha Baek (I.B.) and Mélanie Laurin (M.L.) in details. N.G., myself and Elaine, designed the experiments and interpreted the data. R.C.A. generated the SOX9-inducible transgenic mice with help from M.S. N.G. performed

the *in vivo* ATAC-seq experiments and initial data processing. N.G. performed sample preparation and initial data processing of the proteomic experiments with help from M.L. The *in vitro* sequencing studies were performed with help from I.B. M.S. also participated in SOX9 mouse experiments and tumor cell engraftments. N.I. assisted in high-throughput data generation, immunofluorescence microscopy and quantifications. In addition, I thank Maria Nikolova, Ellen Wong, June Racelis, Dr. Tatiana Omenchenko and John Levorse for other experimental assistance.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ATAC	Assay for Transposase-Accessible Chromatin
BCC	Basal Cell Carcinoma
BioID	Proximity-dependent Biotin Identification
ChIP-seq	Chromatin Immunoprecipitation Sequencing
CNR	Cleavage Under Targets and Release Using Nuclease, or Cut-and-Run
co-IP	Co-Immunoprecipitation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DOX	Doxycycline
EdU	5-Ethynyl-2'-Deoxyuridine
EMT	Epithelial-to-Mesenchymal Transition
EpdSC	Epidermal Stem Cell
FACS	Fluorescence Activated Cell Sorting
GFP	Green Fluorescent Protein
GO	Gene Ontology
GREAT	Genomic Regions Enrichment of Annotations Tool
GSEA	Gene Set Enrichment Analysis
H3K27ac	Histone 3 Lysine 27 acetylation
H3K4me1	Histone 3 Lysine 4 mono-methylation
H3K9	Histone 3 Lysine 9
hPSCs	human pluripotent stem cells
HFSC	Hair Follicle Stem Cell
IF	Immunofluorescence
iPSC	induced pluripotent stem cell
MINT-ChIP	Multiplexed T7-indexed chromatin immunoprecipitation
ORS	Outer Root Sheath
PCA	Principal Component Analysis
PPIs	Protein-protein Interactions
PROTACs	Proteolysis-Targeting Chimeras
qPCR	Quantitative Polymerase Chain Reaction
rtTA	Reverse Tetracycline-Controlled Trans-activator
TRE	Tetracycline Response Element
ΔHMG	SOX9 with High Mobility Group domain truncation
ΔTA	SOX9 Trans-Activation domain truncation

CHAPTER 1. Introduction

1.1 Epigenetics regulates gene expression in cell fate switching

From the early stages of development to the progression toward malignancy, cells undergo a series of crucial fate-determination decisions that dictate their function and identity within an organism. These decisions are critical to establishing and maintaining specific gene expression patterns in diverse cell types. Epigenetics study the heritable changes in gene expression that do not involve alterations to the underlying DNA sequence during cell fate switching. Instead, epigenetic regulations, such as chromatin accessibility, histone modifications, and non-coding RNA molecules, contribute to the complex interplay between the cellular environment and the genome, ultimately influencing cell fate decisions (Wang et al., 2021).

One well-known example of epigenetic regulation in cell fate switching is the conversion of somatic cells to induced pluripotent stem cells (iPSCs) by ectopically activating a combination of transcription factors, or Yamanaka factors, OCT4, SOX2, KLF4, and c-MYC (Chronis et al., 2017; Takahashi et al., 2007; Takahashi and Yamanaka, 2006). The ectopic expression of these factors in differentiated cells triggers a cascade of epigenetic changes, including the gaining of chromatin accessibility at pluripotency genes and the establishment of a pluripotent chromatin state with histone modifications that mark active chromatin, such as H3K4me1 and H3K27ac (Chronis *et al.*, 2017). Together, these epigenetic changes elicited by Yamanaka factors enable the somatic cells to regain pluripotency (Wang *et al.*, 2021).

1.2 Pioneer factors orchestrate epigenetic reprogramming

The specificity of epigenetic regulation on gene expression is achieved through the interactions between specific epigenetic enzymes and transcription factors. In various situations where multiple transcription factors are expressed at the junction of cell fate decision, a particular group of them

consistently exhibits the most pronounced impact on cell fate conversion. Structural and genomic evidence suggests that these transcription factors, known as "pioneer factors," can be identified by their unique ability to interact with target DNA sites in closed chromatin before other factors bind, open, or modify surrounding nucleosomes (Iwafuchi-Doi and Zaret, 2014; Soufi et al., 2015; Zaret, 2020).

Pioneer factors can bind to their target sites on the chromatin even when these regions are in a compacted and repressed chromatin state to facilitate the subsequent binding of other transcription factors and co-factors, such as chromatin remodeling enzymes, ultimately leading to the activation of genes associated with a new cell fate (Balsalobre and Drouin, 2022; Iwafuchi-Doi and Zaret, 2014). This ability allows them to access and modify the chromatin landscape during cell fate transitions. It is also worth noting that these newly activated genes, in turn, reinforce the new cell fate by maintaining the expression of the new set of transcription factors and suppressing old fate-related genes (Iwafuchi-Doi and Zaret, 2014). For instance, in the case of iPSCs, OCT4, SOX2, and KLF4 act as pioneer factors and bind to closed chromatin regions. They recruit various epigenetic modifiers, including histone-modifying enzymes and DNA methyltransferases, which help to erase the repressive epigenetic marks and establish a pluripotent chromatin state (Chronis *et al.*, 2017). This process involves the removal of repressive marks, such as DNA methylation and histone H3K9 methylation, and the addition of activating marks, such as histone H3K4 methylation, at pluripotency-associated genes (Balsalobre and Drouin, 2022; Wang *et al.*, 2021).

In addition to iPSCs, numerous *in vitro* studies have highlighted the role of pioneer factors in governing diverse cell fate transitions through epigenetic mechanisms. For example, since the initial discovery that FOXA (also known as HNF3) can directly and stably interact with nucleosomes through biochemical assays, several studies have demonstrated the role of FOXA1

and FOXA2 as pioneer factors in facilitating liver-specific enhancer activation in collaboration with GATA transcription factors (Gualdi et al., 1996; Iwafuchi-Doi and Zaret, 2014). More recently, mechanistic investigations with next-generation sequencing using *in vitro* pancreatic differentiation from embryonic stem cells and MCF-7 breast cancer cells have further illuminated the way FOXA proteins can recruit co-factors, such as KMT2C, to deposit H3K4me1 marks and prime as well as activate de novo enhancers (Dorighi et al., 2017; Jozwik et al., 2016; Lee et al., 2019; Rao and Dou, 2015; Sze and Shilatifard, 2016).

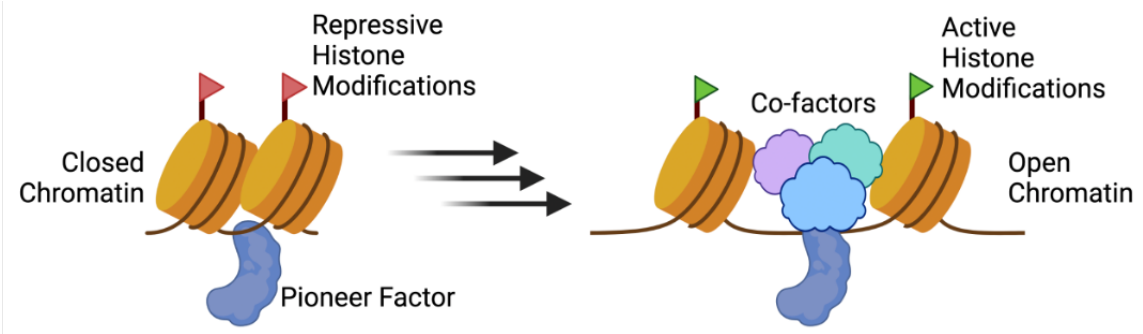

Fig. 1 Schematic of the pioneer factor action in epigenetically switching to a new cell fate Pioneer factor can access closed chromatin in a repressive environment marked by histone modifications such as H3K9 and H3K27 methylation. By doing so, the pioneer factor can recruit co-factors to modify the chromatin with active histone modifications such as H3K27ac and H3K4 methylation, making the chromatin more accessible for new gene expression. However, the sequence of pioneering events, such as pioneer factor binding, co-factor recruitment, and changes in histone modifications, has yet to be discovered.

While these studies frequently emphasize the mechanism of pioneer factors' influence on chromatin and gene expression, most of them concentrate on the initial and final stages of cell fate switching. As a result, the sequence of events in chromatin remodeling has been difficult to decipher, mainly due to the accelerated timeframe of reprogramming *in vitro*, where cells are not subject to the local constraints of their native tissue microenvironments (Fig. 1). Moreover, while

pioneer factors could function as either transcriptional activators or repressors by recruiting co-activators or co-repressors, the mechanisms through which a single pioneer factor can concurrently suppress the previous cell fate while promoting the new one remains unexplored.

1.3 SOX9 regulates hair follicle fate and basal cell carcinoma progression

In light of these limitations, it becomes essential to delve deeper into the intermediate steps of cell fate transitions orchestrated by pioneer factors and to uncover the molecular mechanisms that enable them to simultaneously suppress a cell's previous identity while promoting a new fate. The skin epidermis emerges as an ideal system to address these questions.

The skin epidermis consists of two main stem cell compartments: the hair follicle stem cells (HFSCs) and the interfollicular epidermal stem cells (EpdSCs) (Gonzales and Fuchs, 2017). These compartments become specified during prenatal development and are maintained by distinct signature transcription factors that help regulate their identities and functions. HFSCs are regulated by SOX9, LHX2, and TCF3/4, whereas EpdSCs express TP63, GATA3 and KLF5 to maintain the epidermal gene regulatory network (Adams et al., 2015; Cangkrama et al., 2013; Nguyen et al., 2006; Yi, 2017).

The transition between these two cell fates is first observed during epidermal development when the multipotent embryonic epidermal progenitors bifurcate to become $SOX9^{+}$ HFSCs and $SOX9^{neg}$ EpdSCs (Nowak et al., 2008; Ouspenskaia et al., 2016). In the next step of HF morphogenesis, $SOX9^{+}$ HFSCs bifurcate again to form $SOX9^{neg}$ transit-amplifying hair shaft progenitors. Without SOX9, the differentiation of EpdSCs in embryonic skin continues as usual, given their inherent absence of this factor. However, HFSCs don't develop, leading to halted hair formation (Nowak *et al.*, 2008; Vidal et al., 2005). Furthermore, when SOX9 is conditionally

knocked out in adult HFSCs, they fail to suppress the epidermal fate, leading to a transformation of HFs into epidermal cysts (Kadaja et al., 2014).

These roles of SOX9 parallel the pioneering function of FOXA transcription factors in liver development, where the FOXA footprints on DNA precede the expression of liver specification genes such as albumin (*Alb*) (Gualdi *et al.*, 1996; Lee et al., 2005). Moreover, given the known pioneer factor properties of SOX2 in reprogramming, another member of the SOX family, it is reasonable to hypothesize that SOX9 may also act as a pioneer factor, influencing chromatin landscape to guide cell fate decisions in skin stem cells (Chronis *et al.*, 2017; Julian et al., 2017; Kamachi and Kondoh, 2013).

In the context of disease, when EpdSCs switch towards a tumorigenesis fate, SOX9 is highly activated in basal cell carcinoma (BCC) and is essential for BCC progression (Larsimont et al., 2015). In the mouse SmoM2 model of BCC, where the SHH signaling pathway is constitutively active and drives tumor progression, the loss of function of SOX9 leads to the regression of tumor progression *in vivo* (Larsimont *et al.*, 2015). Furthermore, the genes with reduced expression in BCC following SOX9 loss of function are associated with embryonic hair follicle progenitors.

In adult HFs, SOX9 binds to epicenters of HFSC genes, which are clusters of H3K27ac enhancer peaks that act as hubs for gene regulation (Adam et al., 2015). When SOX9 is ectopically induced in adult EpdSCs, it induces the expression of HF genes such as *Lhx2*. A more mechanistic study involving the reprogramming of endothelial cells to a mesenchymal fate by inducing SOX9 in cultured human umbilical vein endothelial cells has further demonstrated that the chromatin activated by SOX9 is associated with H3K4me1 and H3K27ac (Fuglerud et al., 2022). However, it has not been definitively demonstrated whether SOX9 can interact with closed or repressed chromatin as a pioneer factor to induce a new cell fate. Yet collectively, these findings highlight

the potential of SOX9 as a pioneer factor in various cellular contexts and underscore the need for further investigation into its molecular mechanisms and functions in pioneering a cell fate switching.

CHAPTER 2. SOX9 re-activation in EpdSCs initiates fate-switching cascade

2.1 SOX9 launches phenotypic changes in the epidermis towards BCC-like lesions

To investigate the role of SOX9 in cell fate switching in adult tissue stem cells, Rene Adam engineered mice carrying a myc-epitope-tagged *Sox9* transgene regulated by a tetracycline-responsive enhancer and a minimal promoter (*TRE-Sox9*). After confirming the specificity of myc-tag expression, Nicholas Gomez and Lisa Polak bred selected mice with lines expressing the tetracycline-inducible transcriptional activator (rtTA) driven by an epidermal (*Krt14*) promoter (*Krt14rtTA*) so that Nicholas and I could induce SOX9 expression in basal EpdSCs upon doxycycline (DOX) treatment (Fig. 2.1a) (Nguyen *et al.*, 2006). In addition, I selected mice that induced SOX9 in EpdSCs at levels comparable to SOX9 level in adult HFSCs (Fig. 2.1c, d).

At postnatal day 21, the mice were treated with DOX through Intraperitoneal administration, and their dorsal skin was monitored for 12 weeks (Fig. 2.1a). By week 1 (W1), SOX9 was observed in the nuclei of EpdSCs, and its nuclear expression remained steady throughout the inspection period (Fig. 2.1b). In first two weeks, dorsal skin with SOX9 induction showed normal morphology and differentiation, but by week 2 (W2), an increase in EpdSC proliferation was observed with S-phase labeling with nucleotide analogue 5-ethynyl-2'-deoxyuridine (EdU), administered just before analysis, similar to that observed in embryonic epidermis and BCCs where SOX9 is ectopically induced (Fig. 2.2a-c) (Ouspenskaia *et al.*, 2016). From W2 and beyond, *de novo* invaginations began to form between native hair follicles, accompanied by an elevated proliferation and disrupted EpdSC differentiation (Fig. 2.1b and Fig. 2.2a-c).

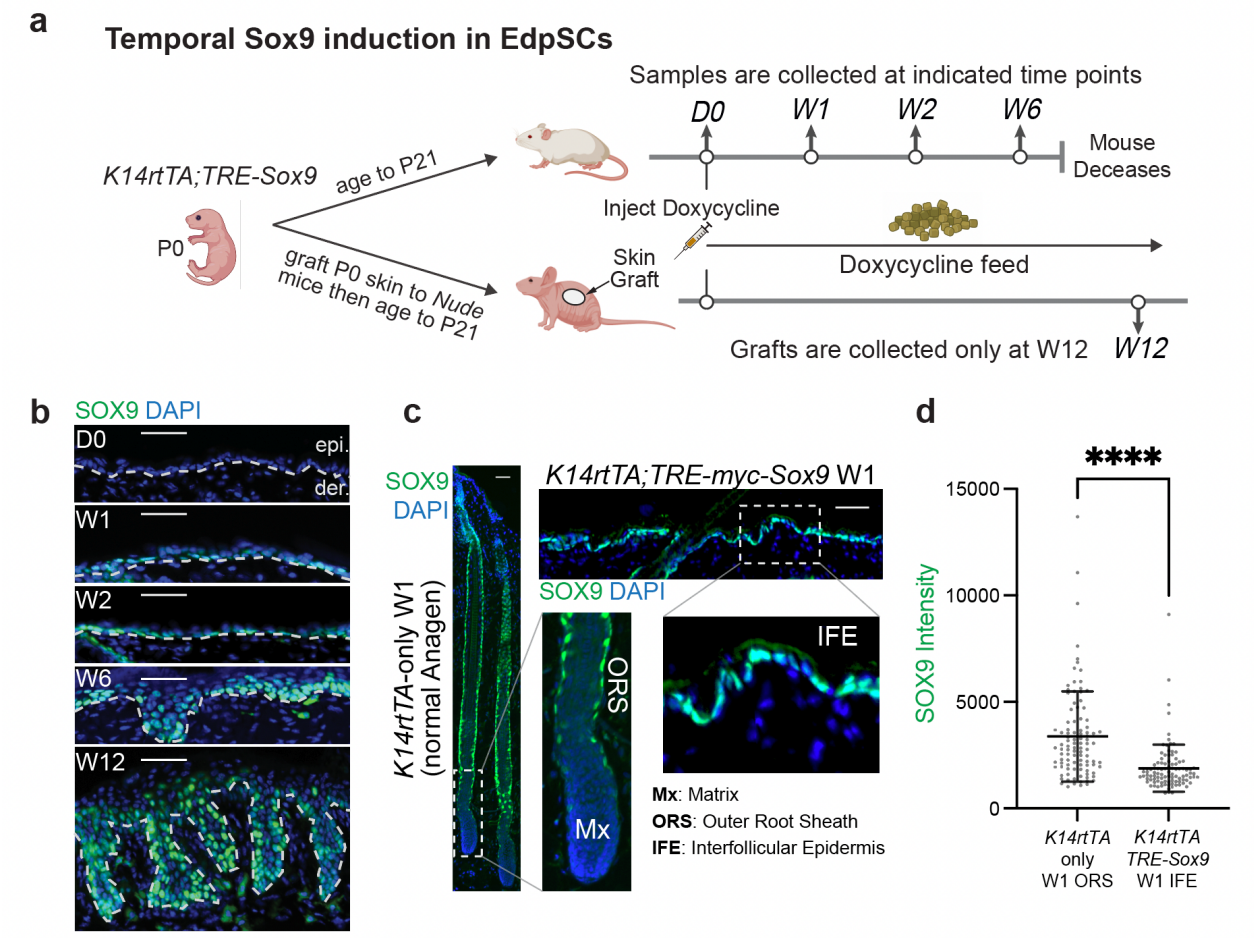

Fig. 2.1: Inducing SOX9 in EpdSCs with transgenic mice

a, Schematic of SOX9 induction in EpdSCs of adult skin and in post-engrafted skin.

b, Immunofluorescence reveals SOX9 protein in nuclei of EpdSCs after doxycycline induction. Dotted lines denote epidermal-dermal borders. Scale bars, 50μm.

c, Immunofluorescence comparing SOX9 expression in the W1 tissue after SOX9 induction in adult EpdSCs and the age-matched sample with normal anagen ORS (containing activated HFSCs). All scale bars are 50μm.

d, Quantifications of SOX9 immunofluorescence intensity in the two states in (c) shows that SOX9 levels in ectopically induced adult EpdSCs are not higher than in native ORS HFSCs.

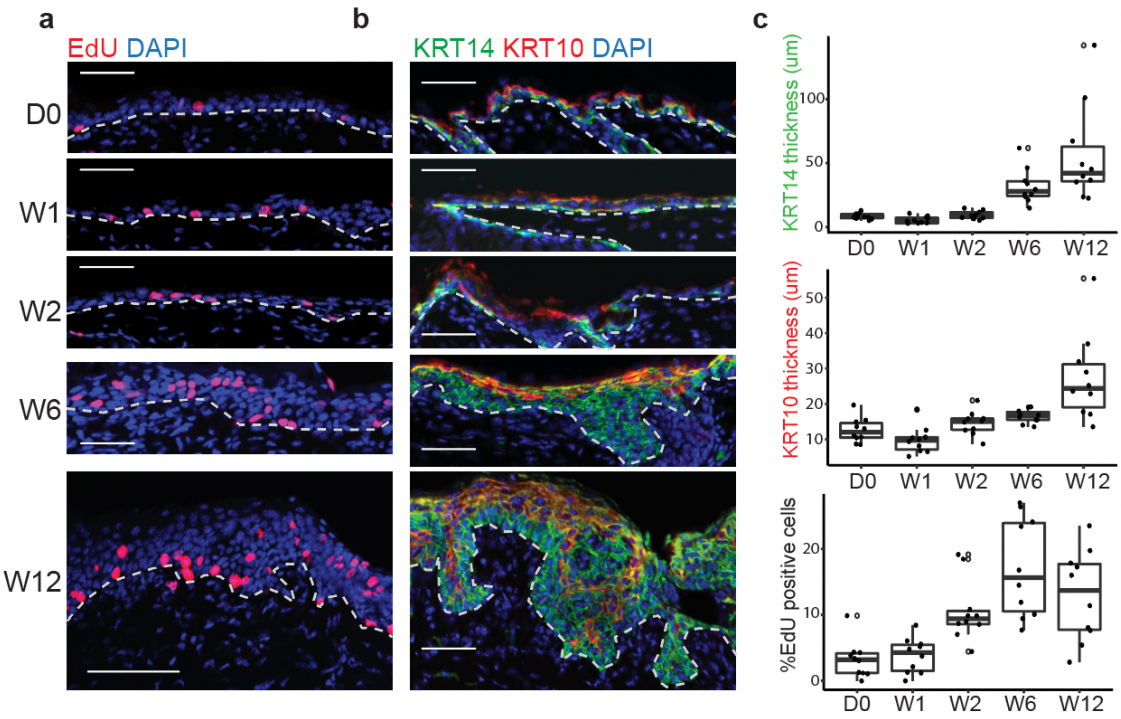

Fig. 2.2 Ectopic re-activation of SOX9 in adult EpdSCs elicit phenotypic abnormality
a, EdU staining after SOX9-induction over time to assess cell proliferation. Scale bars are 50μm. **b,** KRT14 and KRT10 immunofluorescence in the epidermis after SOX9-induction in EpdSCs. Note that the KRT14 skin progenitor layer significantly expanded over time. Scale bars are 50μm. **c,** (top) Quantification of the thickness of KRT14 and KRT10 epidermal layers over 12 weeks. (bottom) Quantification of % proliferating cells in the basal layer of EpdSCs after SOX9 induction.

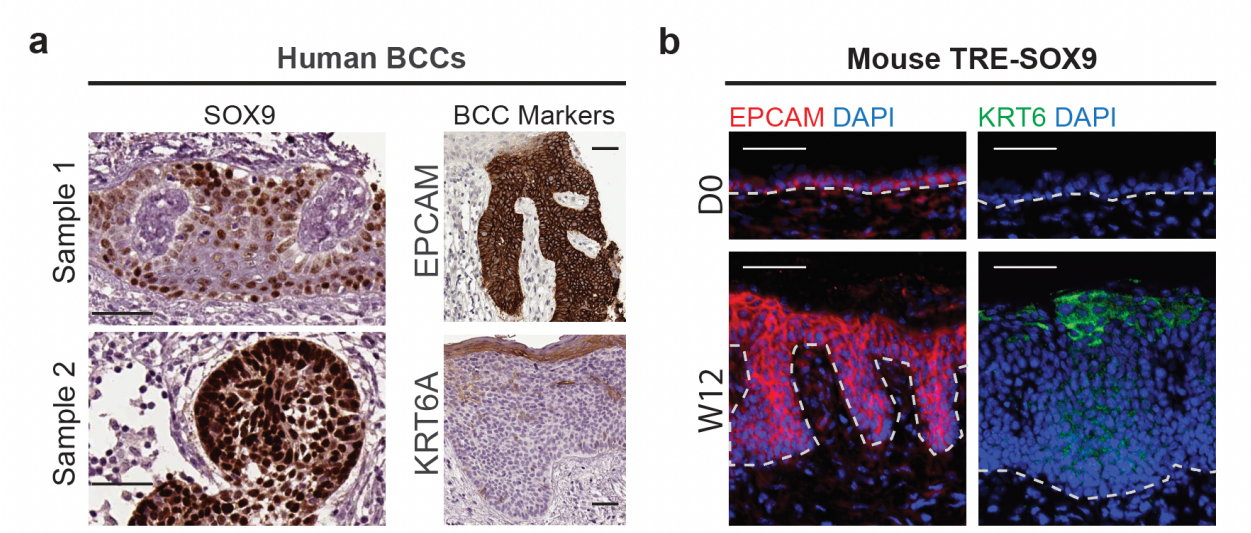

Fig. 2.3 SOX9 induced lesions resemble human BCC

a, Immunohistochemistry staining of human BCC samples from The Human Protein Atlas. SOX9 (Antibody: CAB068240) and EpCAM (Antibody: CAB030012) show strong staining in the body of the BCC lesion, whereas KRT6A (Antibody: HPA061168) is restricted to the apical side of the tumor tissue.

b, Immunofluorescence shows the lesions in W12 samples have similar EpCAM and KRT6A staining pattern as human BCC. Dotted lines denote the dermo-epidermal border. All scale bars are 50μm.

As the differentiation defects occurred throughout the entire dorsal skin, mice had to be euthanized at week 6 (W6). To track later events in SOX9 fate switching, Megan Sribour engrafted neonatal *Krt14rtTA;TRE-Sox9* transgenic skin patches onto immunodeficient mice. Once normal skin pathology was restored (21 days after grafting), Nicholas and I induced SOX9 (Fig. 2.1a). By week 12 (W12) post-induction, numerous invaginations were observed that resembled the morphological and molecular features of human basal cell carcinomas (BCCs), including SOX9, EpCAM, and KRT6 (Fig. 2.3a,b) (Crowson, 2006; Gaiser et al., 2018).

2.2 SOX9 induces temporal EpdSC fate switching transcriptionally

To gain a better understanding of the transcriptomic changes occurring in EpdSCs during the SOX9-driven reprogramming process, I performed RNA-seq on fluorescence activated cell sorting (FACS)-purified EpdSCs from *Krt14rtTA;TRE-Sox9* epidermis at each time point after SOX9 induction (Fig. 2.4a). The two biological replicates performed at each time point show a high degree of correlation at all time points (Fig. 2.4b).

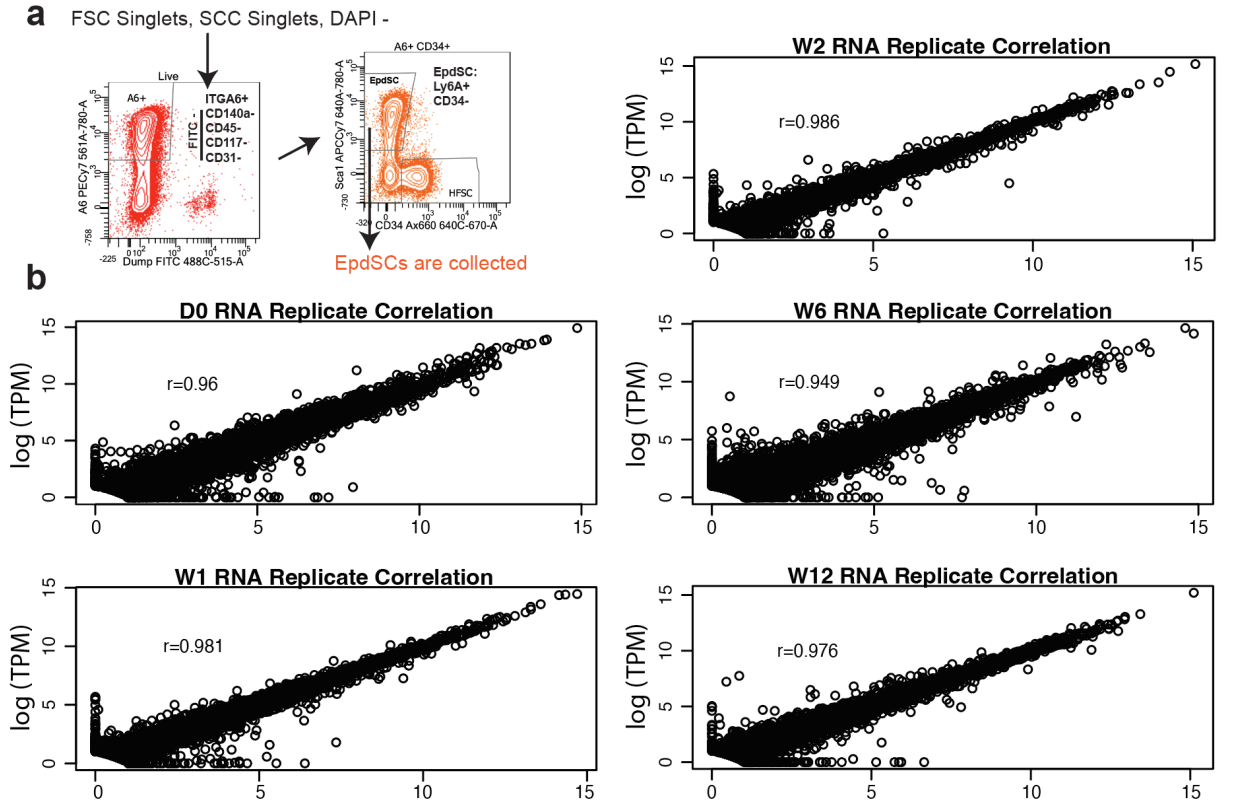

Fig. 2.4 Sample preparation and quality control of SOX9-induced transcriptome dynamics
a, FACS gating strategies for isolating EpdSCs. DAPI-negative singlets were first gated on ITGA6 (epithelial progenitors) and immune cell, fibroblast/adipocyte, endothelial and melanocyte lineage markers (CD45, CD140a, CD31 and CD117). Lineage-negative and ITGA6^{+} cells were further gated on CD34 and Ly6A/E (SCA-1) to distinguish HFSCs and EpdSCs, respectively.
b, Replicate correlation analyses of RNA-seq show strong correlation ($r > 0.94$) between samples across time points.

To analyze the transcriptional dynamics across time, I selected the significantly variable genes ($p < 0.05$) along the SOX9-induced reprogramming cascade (Fig. 2.5a and Table 1). As expected, at D0, the FACS sorted cell population displayed the hallmark molecular signature of EpdSCs, with mRNAs encoding the epidermal master regulator transcription factors (TP63, KLF5, GATA3), key signaling factors (NOTCH ligand JAG2 and WNT10a) and epidermal structural proteins (keratins KRT5, KRT15, KRT77). These epidermal transcripts were markedly downregulated at W1 SOX9 induction. Conversely, transcripts encoding classical markers of proliferative HFSC progenitors of the embryonic hair bud and/or adult ORS/HFSCs (KRT17, LRIG1, FZD1, FZD7), including HFSC transcription factors (NFIB, NFIX, FOXC1, TCF3), were all transcriptionally elevated within the first two weeks after SOX9 induction. Notably, markers of the hair shaft and inner root sheath were missing, as were their matrix progenitors, all of which requires SOX9 downregulation in the HF differentiation and relies upon other transcription factors (Yang et al., 2017).

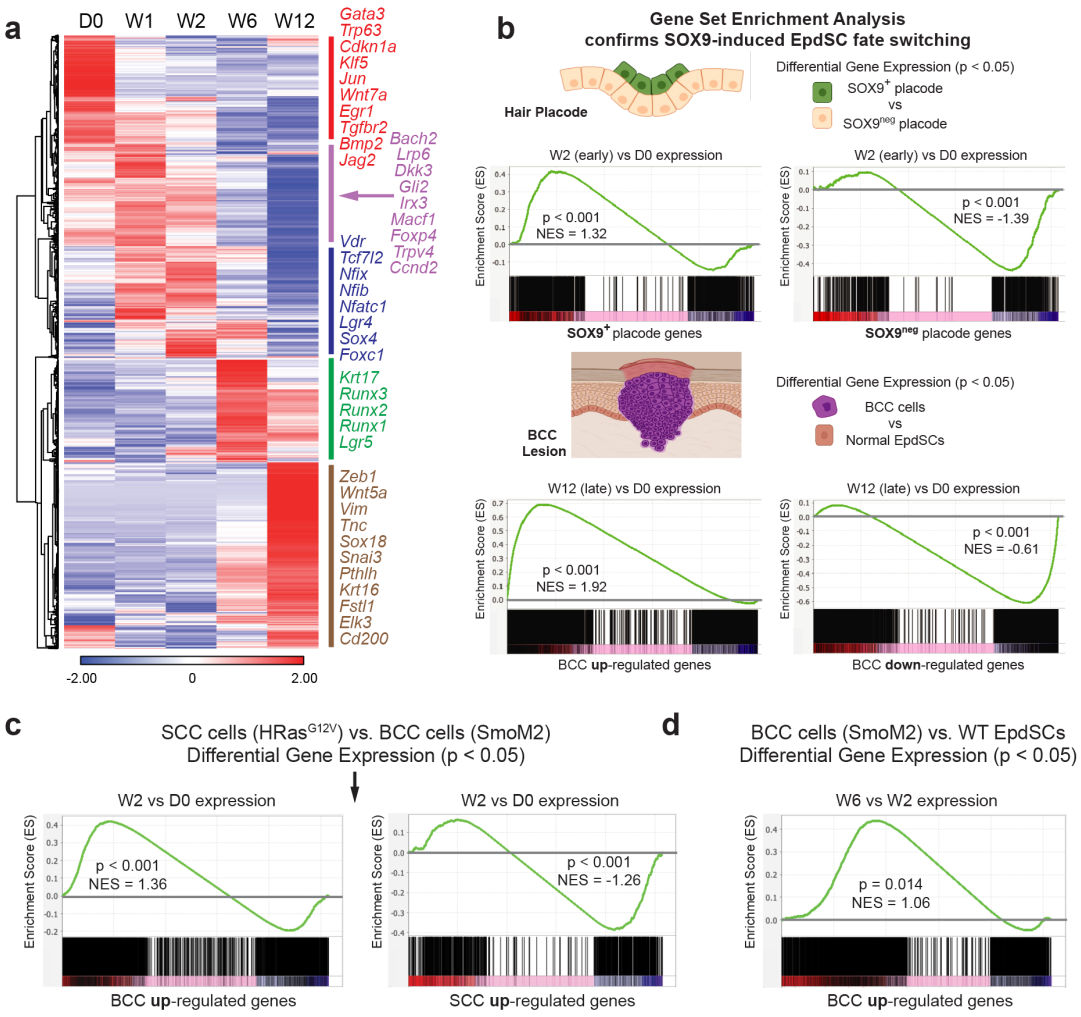

Fig. 2.5: SOX9 re-activation silences the epidermal while progressively activating the hair follicle and then BCC fates on a transcriptome level

a, Heatmap showing expression of significantly variable genes (adjusted p-value < 0.05) across combined independent replicates (r > 0.94) of each of 5 indicated time points. Hierarchy clustering revealed 5 distinct patterns, shown as colored bars on the right with representative transcripts indicative of specific fates.

b, Gene Set Enrichment Analyses (GSEA) of W2 vs D0 SOX9-induced expression changes in adult EpdSCs compared to $SOX9^{+}$ and $SOX9^{neg}$ placode gene signatures from embryonic day E15.5 (top), and W12 vs D0 SOX9-induced EpdSC expression changes compared to BCC up-regulated and down-regulated signatures (bottom).

c, At W2 following SOX9 induction, similarities to BCC were already apparent. By contrast, a negative correlation to SCC was observed.

d, In comparing all BCC genes to W6 and W2 data, the shift towards a BCC signature continued to rise, concomitant with the phenotypic changes described in Fig. 2.1 and Fig. 2.2.

2.2.1 Early transcriptionalchanges demonstrate an EpdSC to HFSC fate

While there were limited morphological changes observed in the epidermis within the initial two-week period after SOX9 induction, EpdSCs underwent significant transcriptional changes. To formally test if SOX9 induction reprogram the epidermal transcriptome to a hair follicle fate as observed in hair follicle development, I employed Gene Set Enrichment Analysis (GSEA). GSEA with gene sets significantly ($p < 0.05$) upregulated or downregulated in $SOX9^+$ cells in embryonic hair bud showed a strong resemblance between W2 sample and hair bud $SOX9^+$ cells (Fig. 2.5b). These results provide compelling evidence that the early transcriptional changes elicited by SOX9 in adult EpdSCs closely mimics the events that occur when embryonic skin progenitors naturally induce SOX9, leading to a switch from an epidermal to a hair follicle fate (Larsimont *et al.*, 2015).

2.2.2 Late transcriptionalchanges demonstrate an activation of BCC fate

In a manner similar to BCC development, progression towards mature hair follicles did not occur following SOX9 induction. This observation is consistent with the need for SOX9 downregulation for HFSCs to generate transit amplifying, short-lived progeny that produce hair and its channel (Ouspenskaia *et al.*, 2016; Yang *et al.*, 2017). Nevertheless, sustained SOX9 expression led to continued dynamic transcriptional reprogramming beyond W2, and by W6-W12, cancer-associated features began to emerge. At W12, GSEA revealed a strong correlation, both up and down from normal EpdSCs, between the molecular signature of BCC and the EpdSCs with SOX9 induction compared to normal skin (Fiore et al., 2020; Larsimont *et al.*, 2015)(Fig. 2.5b).

Interestingly, although the similarities in gene expression to BCC were strongest at the later stages, they surfaced as early as W2, prior to phenotypic changes. Moreover, they clearly favored a BCC signature over a squamous cell carcinoma (SCC) signature which are two of the most common types of skin cancer (Fig. 2.5c,d) (Fijalkowska et al., 2021; Haensel et al., 2022). These

findings underscore the critical role of SOX9 in driving the transcriptional reprogramming of adult EpdSCs towards a hair follicle fate and when left unchecked, contribute to the development of BCC.

CHAPTER 3. SOX9 induces temporal epigenetic changes in EpdSCs

3.1 SOX9 induces chromatin remodeling mainly at intergenic regions

Given the marked temporal dynamics in transcription following SOX9 induction, my findings suggested that SOX9 is necessary and sufficient to drive the epidermal to HF fate transition not only in embryonic epidermal progenitors and BCC but also in adult EpdSCs. Therefore, I turned to addressing how SOX9 remodels chromatin to exert its master regulatory powers to induce these transcriptional dynamics. To this end, Nicholas and I began by performing Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) on FACS-purified EpdSCs over 12 weeks following SOX9 induction to illustrate the chromatin accessibility dynamics associated with the cellular fate reprogramming from an EpdSC to HFSC to a BCC-like state (Buenrostro et al., 2013; Corces et al., 2018).

Principal component analysis (PCA) displayed that the biological replicates were highly concordant and had a clustering pattern associated with the duration after SOX9 induction (Fig. 3.1a). The D0 and W1 samples were closely clustered, while W2 acted as an intermediary, and later time points (W6 and W12) formed a separate cluster.

To compare chromatin accessibility across all time points, I analyzed ATAC peaks in a comprehensive upset plot, revealing that some were shared among samples. These shared peaks are in proximity to representative housekeeping genes or genes common to EpdSC, HFSC and BCC fate (such as *Krt5*) (Fig. 3.1b). Other ATAC peaks, however, demonstrated highly dynamic behavior, such as one in the gene locus of WNT-target *Ctnnb1*, suggesting SOX9-induced chromatin remodeling over time. While some peaks showed increased accessibility, others became less accessible (Fig. 3.1b). Notably, the most significant changes occurred at the W2 time point after SOX9 induction, with many of the newly accessible peaks at W2 remaining open in subsequent time points.

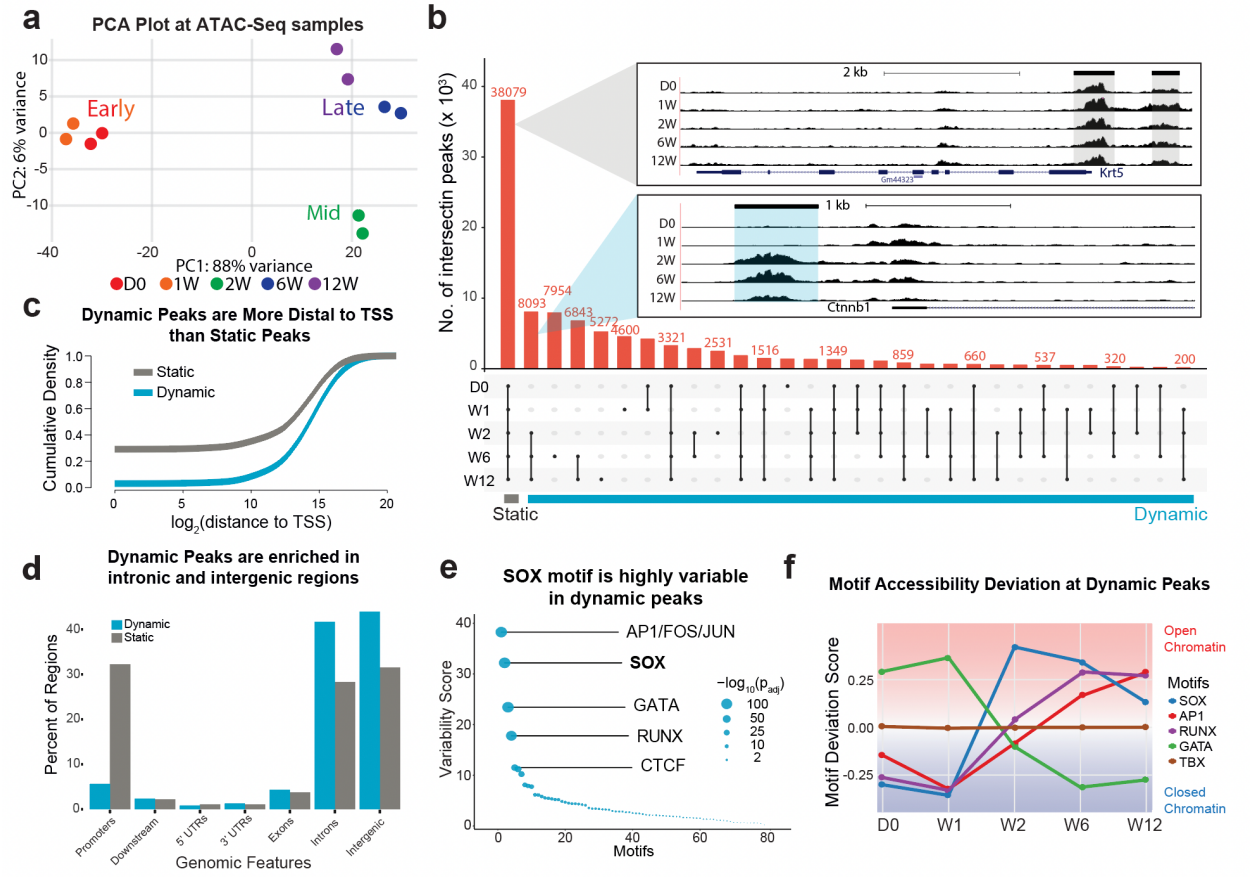

Fig. 3.1: Upon induction, SOX9 provokes chromatin changes at intergenic regions

a, Principal Component Analysis (PCA) of ATAC-seq duplicate samples over the 5 time points.

b, Upset plot of called ATAC-seq peaks for each time point following SOX9-induction. Shared regions between time points are indicated by the dots and connecting lines while regions that are unique to that time point are indicated by a lone dot. Grey inset shows a genome browser track of the generic skin stem cell *Krt5* locus as an example of a static gene region. Blue inset depicts a dynamic region of the WNT-target HFSC gene *Ctnnb1* that is open by W2 and remains open across W6, and W12 time points.

c, Empirical cumulative distribution plot of dynamic (blue) and static peaks (grey) and their density relative to the transcription start site (TSS) of the nearest gene. Note that dynamic peaks are primarily associated with distal regions, typically encompassing enhancers.

Probing more deeply, I binned ATAC-peaks as either static (present at all timepoints, n = 38,079), or dynamic (absent in at least one timepoint, n = 62,626) (Fig. 3.1b). While both static and dynamic peaks were enriched near transcription start sites (TSS), dynamic peaks situated further from TSS and were significantly more enriched in intergenic regions instead of promoters than static peaks (Fig. 3.1c, d), suggesting that SOX9 was primarily acting at enhancers rather than directly at promoters. This observation further supports the idea that SOX9 functions as a pioneer factor, as its activity primarily focuses on activating chromatin at enhancer regions.

To infer if SOX9 directly targets the dynamic chromatin, I performed temporal motif analysis with ChromVAR, which considers both enrichment and chromatin accessibility variability at each motif (Schep et al., 2017). As expected, SOX motif is in the top variable motifs in dynamic ATAC peaks (Fig. 3.1e). In addition to SOX, AP1(FOS/JUN), GATA and RUNX were top variable motifs. To learn the directionality of how motif accessibility varied over time, I plotted accessibility deviation scores for each time point and compared them to a motif (TBX) which showed no temporal variability (Fig. 3.1f).

3.2 Direct and early chromatin remodeling induced by SOX9

K-means clustering of the 62,626 dynamic ATAC peaks offered a more detailed map of the temporal changes in chromatin accessibility following SOX9 induction (Fig. 3.2a). While the most significant changes were observed between the W1 and W2 timepoints, the clustering additionally identified more than 10,000 peaks that opened at later time points (cluster C4 in Fig. 3.2a). Further analysis of the biological pathways associated with each of the six clusters was performed using the Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010).

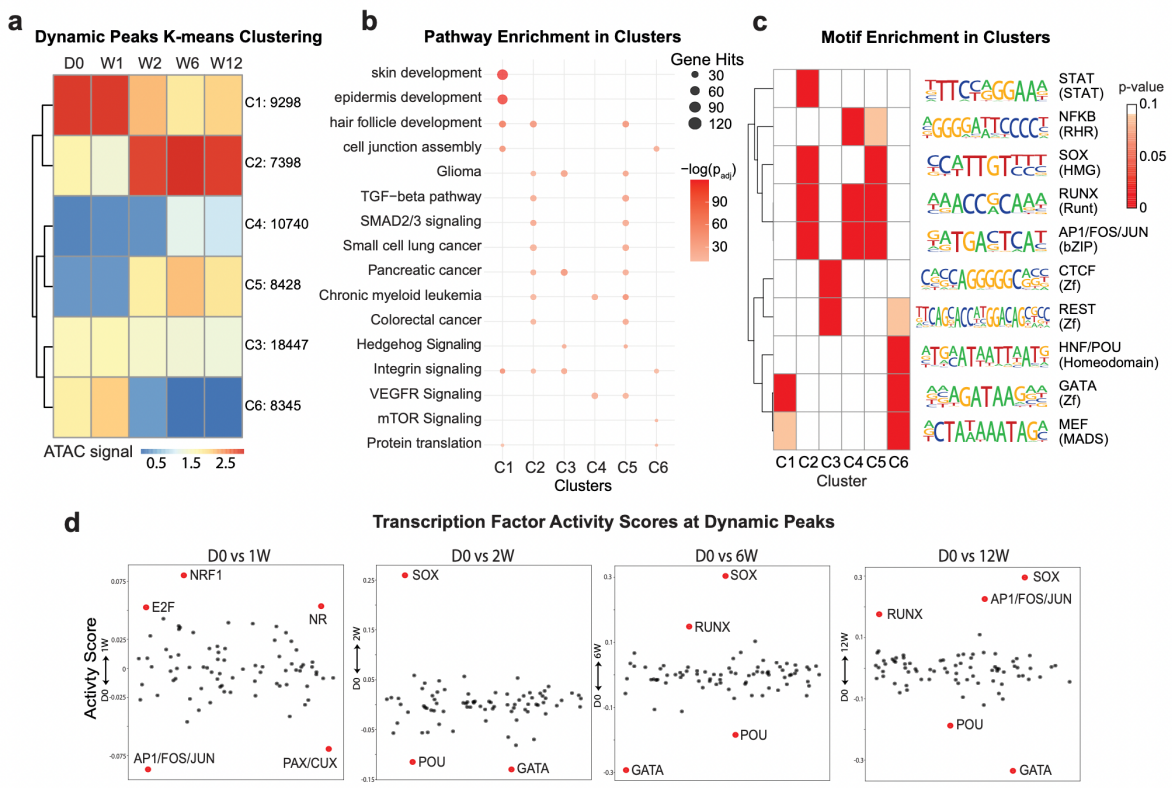

Fig. 3.2: Temporal chromatin changes elicited by SOX9

a, K-means clustering of ATAC peaks based on signal across time. Number of peaks in each cluster are indicated on the right.

b, GO-term analysis of the genes associated with each cluster of peaks. Size of circle reflects the number of gene hits for each pathway, while the shade of red indicates adjusted p-value (P_{adj}).

c, Motif enrichment analysis of each cluster. Shade of red indicates p-value with white representing ≥ 0.01. Note enrichment for: GATA motif in epidermal genes whose chromatin closes after SOX9 induction; SOX motif in HF genes whose chromatin opens early and RUNX motif in genes whose chromatin opens weeks after SOX9 has bound.

d, Footprint analysis shows how TF footprints differ in activity at indicated time points compared to D0. Motifs indicated with red dot show significant changes in activity score over D0.

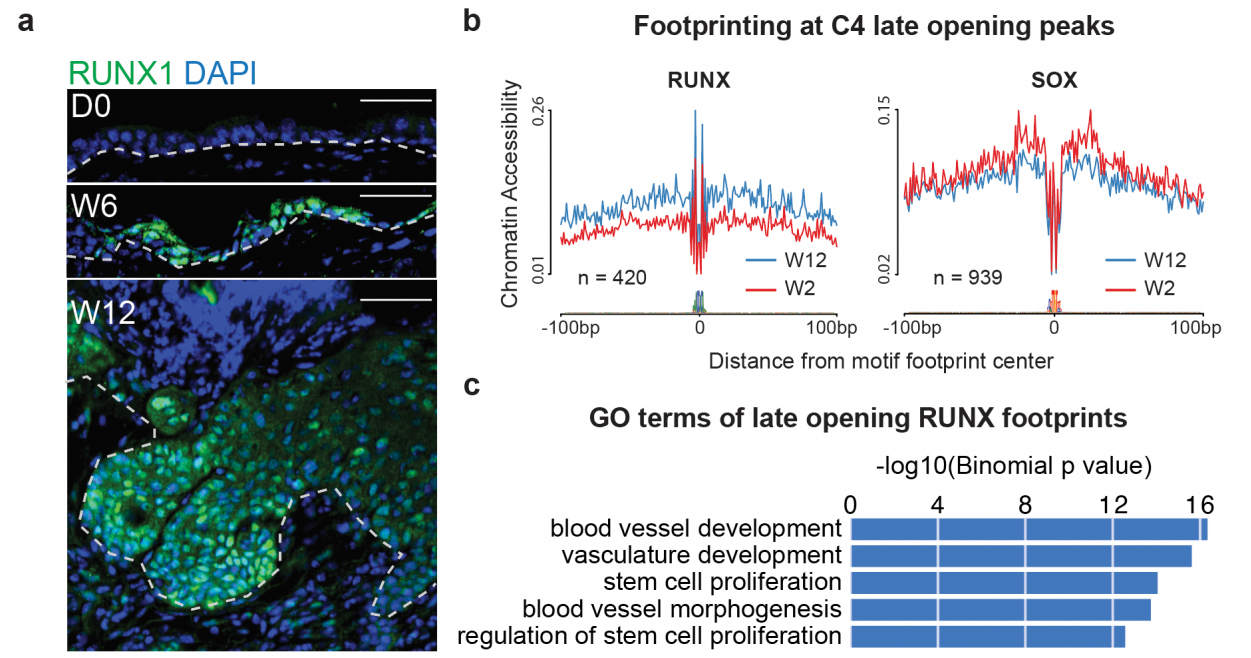

Fig. 3.3: Late chromatin changes elicited by RUNX transcription factors
a, Immunofluorescence staining for RUNX1 shows its absence during epidermal homeostasis, but it is detectable at later time points (W6 and W12) after SOX9 induction. Scale bars, 50μm.
b, RUNX and SOX motif footprint analyses at W2 and W12, showing a late increase in chromatin accessibility at the RUNX footprints when phenotypic BCC-like invaginations are prevalent. In contrast, the accessibility of SOX footprints remained consistent between W2 and W12.
c, GO terms of genes whose putative regulatory regions have RUNX footprints and open at W6.

The results from GREAT analysis revealed that C1 and C6, which were closed within the first 2 weeks, were enriched for biological pathways that are directly related to EpdSC fate (Fig. 3.2b). In contrast, C2 and C5 exhibited a marked increase in chromatin accessibility during this time and were enriched for pathways associated with HF development and SHH-signaling, which are important for stimulating ORS/HFSC lineage proliferation and BCC progression (Fig. 3.2b) (Hsu et al., 2014; Larsimont *et al.*, 2015; Oro and Scott, 1998; Vidal *et al.*, 2005). It is worth noting that AP1, EGFR, and TGFβ signaling pathways, which are known to be elevated in BCCs that develop resistance to SHH inhibitors, were downregulated (Kuonen et al., 2021; Oshimori and Fuchs, 2012; Yao et al., 2020). Many of the pathways enriched in the C2 and C5 clusters were also found to be

implicated in other cancers previously associated with SOX9 expression (Matheu et al., 2012; Wang et al., 2012; Zhou et al., 2012).

SOX9's role in activating the ORS/HFSC fate appeared to be direct, as chromatin that showed opening by W2 following SOX9 induction were associated with HF development genes and displayed SOX motif enrichment, SOX footprint enrichment, as well as drastic increase in SOX motif deviation score by ChromVAR (Fig 3.1e,f and Fig. 3.2c,d). These peaks also persisted in a highly accessible state from W6 to W12 (Fig. 3.2a).

3.3 Indirect and late chromatin remodeling induced by SOX9

While pathways shared by HF fate and BCC progression enriched in C2 and C5 were directly targeted by SOX9, by contrast, cluster C4 is associated with genes irrelevant to HFSC fate. Instead, the accessibility of many peaks associated with BCC fate increased between W2 and W6, consistent with the transcriptome data (Fig3.2b and Fig. 2.5a). Interestingly, these C4 peaks were not enriched for SOX motifs, but rather for RUNX, AP1, and NFkB motifs (Fig. 3.2c). Moreover, while nuclear SOX9 present in EpdSCs from W1, transcripts for *Runx1-3* showed a delayed upregulation (Fig. 2.5a). Upon SOX9 induction, *Runx1* transcript levels increased by 8.5-fold at W2, and RUNX1 immunofluorescence confirmed its presence in the later stages of reprogramming at W6 (Fig. 3.3a). The sustained expression of *Runx*1 further highlighted a BCC-like rather than SCC-like phenotype, given that *Runx1* loss of function favors basosquamous tumors found in treatment-resistant human BCCs (Haensel *et al.*, 2022).

Consistent with motif enrichments in Fig. 3.2c, the RUNX motif continued to become more accessible from W2 to W6 (Fig. 3.1f). Finally, footprint analyses revealed that RUNX footprints in the late-opening C4 ATAC peaks increase in chromatin accessibility from, whereas SOX footprints appeared early and remained constant from W2 to W12 (Fig. 3.3b). These late-opening

RUNX footprints are enriched in proximity to stem cell proliferation genes based on GO-term analysis, as well as vasculature and angiogenesis genes, which are all hallmarks of cancers (Fig. 3.3c). These observations suggest that the early transcriptional changes were primarily mediated by SOX9, while the later changes involved not only SOX9, but also other transcription factors induced by SOX9, including those of the RUNX family whose motifs were enriched, as noted earlier. Although it is unclear whether RUNX factors function as pioneer factors, the enrichment of RUNX motifs during the rise in proliferation during BCC-like downgrowth raises the possibility that proliferation may facilitate the chromatin accessibility for these factors.

CHAPTER 4. Temporal epigenetic mechanisms behind HF fate activation by SOX9

4.1 SOX9 is a *bona fide* pioneer factor

I initiated the investigation of the mechanisms underlying SOX9's regulatory influence on transcriptional and chromatin dynamics by conducting temporal assessments of SOX9's binding to chromatin using a low-input ChIP-seq approach, namely Cleavage-Under-Targets-and-Release-Using-Nuclease (Cut-and-Run) sequencing (Skene et al., 2018; Skene and Henikoff, 2017). The biological replicates were highly correlated, and the SOX motif exhibited the highest enrichment in the SOX9 Cut-and-Run peaks (Fig. 4.1a,b).

SOX9 rapidly bound to chromatin in a week following induction in EpdSCs before proliferation increases. Additional binding at later time points was not observed, nor any of the binding sites lost SOX9 binding through 12 weeks. On the other hand, the increase in chromatin accessibility at SOX9 binding sites was observed between W1 and W2, denoting SOX9's ability to bind to closed chromatin (Fig 4.1c). In fact, almost 30% of the SOX9 binding sites in all SOX9 Cut-and-Run peaks pooled from W1 to W12 were positioned within closed chromatin at D0 (Fig. 4.1d). Furthermore, at these sites, nucleosome occupancy was lost by W2, as measured by histone H3 as a proxy of nucleosome occupancy (Fig. 4.1d). Additionally, these SOX9-bound opening peaks exhibited a progressive reduction in SOX9 Cut-and-Run fragment length over time, indicative of nucleosome displacement and pioneer factor activity (Meers et al., 2019b). These findings provide strong support that in the chromatin environment of skin EpdSCs, SOX9 can function as a genuine pioneer factor by binding to its corresponding motifs within closed chromatin and subsequently displacing nucleosomes.

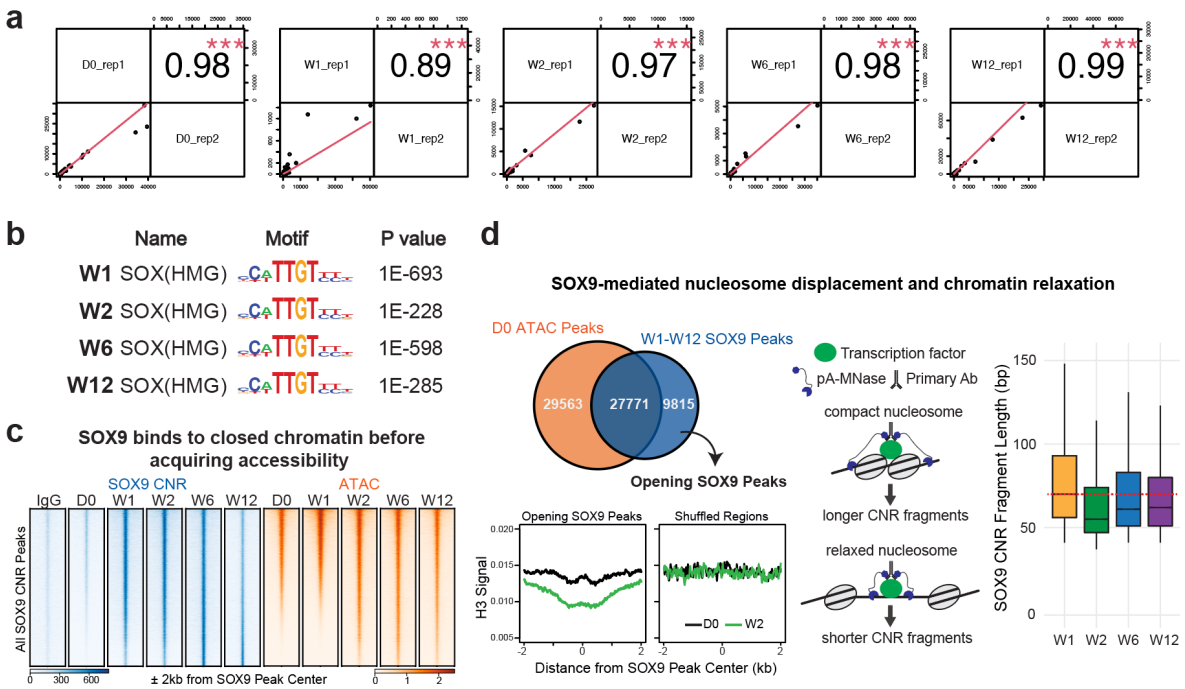

Fig. 4.1: SOX9 is a *bona fide* pioneer factor, binds closed chromatin and displaces nucleosome

a, The correlation analyses of SOX9 Cut-and-Run replicates demonstrate a robust correlation between the duplicate samples at every time point.

b, HOMER motif analysis shows that SOX(HMG) motifs are most strongly enriched in SOX9 CNR peaks in W1 to W12 samples.

c, Heatmap of IgG CNR (blue), SOX9 CNR (blue) and ATAC-seq (orange) signals at all SOX9-bound peaks over time. Peaks are arranged along the vertical axis based upon their accessibility at W1. Note that within the cohort of peaks in the bottom half along this axis, SOX9 binding has occurred by W1, while their ATAC-seq landscape does not change until the following week.

d, (top left) Venn diagrams show that although SOX9 binds to many pre-existing accessible chromatin peaks, 9815 SOX9 peaks open *de novo.* MINT-ChIP for histone H3 (H3) reveals that by W2 post-induction, SOX9 binds closed chromatin, concomitant with displacement of histone H3 at the SOX9-bound site. (right) Schematic and data showing that CNR fragment length shortens between W1 and W2, correlating with SOX9 binding and nucleosome eviction.

4.2 SOX9 triggers epigenetic modifications prior to chromatin opening and remodeling

The dynamics at SOX9 binding sites match the analysis of SOX motif enrichment and footprint where SOX9 is directly responsible for chromatin changes from W1 to W2 related to HF fate. Further investigating SOX9 targets that may regulate W2-W12 chromatin changes related to BCC fate, I noticed SOX9 binding at multiple sites within the yet closed Runx1 gene locus at W1. These sites became accessible a week later (Fig. 4.2b). Since the dynamic peaks were enriched at distal intergenic regions (Fig. 3.1c,d), it prompted me to perform multiplexed T7-indexed chromatin immunoprecipitation (MINT-ChIP) analysis on H3K4me1, a histone modification enriched at enhancers (van Galen et al., 2016). The results showed high correlation between replicates and that H3K4me1 was concomitant with SOX9 binding at the Runx1 intergenic region (Fig. 4.2a,b). Interestingly, both SOX9 binding and enhancer histone modification H3K4me1 appeared at the Runx1 locus prior to peak accessibility (Fig. 4.2b).

This substantial delay from H3K4me1 deposition at SOX9-bound peaks in the regulatory regions of *Runx1* locus to the chromatin becoming accessible and *Runx1* transcription made me hypothesize that this is a general phenomenon of SOX9-induced fate switching. To test this, I compared chromatin accessibility and enhancer H3K4me1 signals over 12 weeks at all peaks that become open de novo upon SOX9 binding (Fig. 4.3a). H3K4me1 deposition, which temporally correlated with SOX9 binding, occurred within one week after DOX and then leveled off. Both SOX9 binding and H3K4me1 preceded chromatin accessibility changes at W2 (Fig. 4.3a). Conversely, activation of these enhancers, measured by H3K27ac changes at W1, was less robust and continued to increase over time relative to SOX9 and H3K4me1 (Fig. 4.3b). Remarkably, H3K4me1 deposition was markedly amplified in nucleosomes adjacent to, but not directly over, SOX9 binding sites (Fig. 4.3c). Additionally, the domain sizes of H3K4me1 modifications

gradually increased from D0 to W2 only at SOX9-bound opening peaks, but not at static peaks that were already open (Fig. 4.3d).

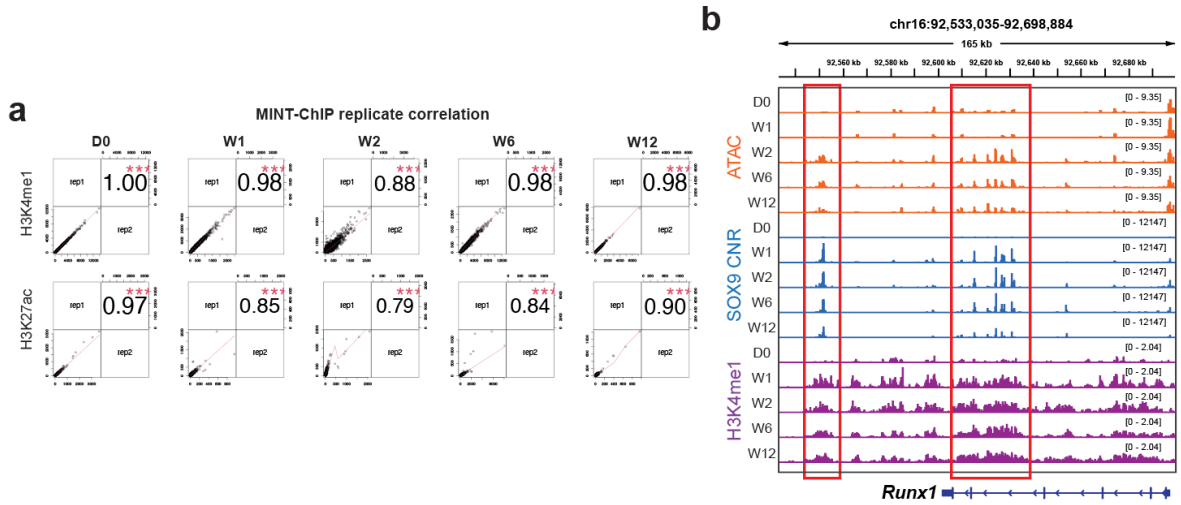

Fig. 4.2: Temporal epigenetic changes induced upon SOX9 binding in *Runx1* locus

a, Replicate correlation analyses of the H3K4me1 and H3K27ac MINT-ChIP show strong correlation between each duplicate sample across all time points.

b, Chromatin landscape of the *Runx1* gene locus, showing that pioneer factor SOX9 binds to this gene (blue), concomitant with early increases in H3K4me1 modifications across the locus (purple), while chromatin accessibility (orange) comes afterwards. Red boxes indicate regions that are bound by SOX9, primed at W1, and opened at W2.

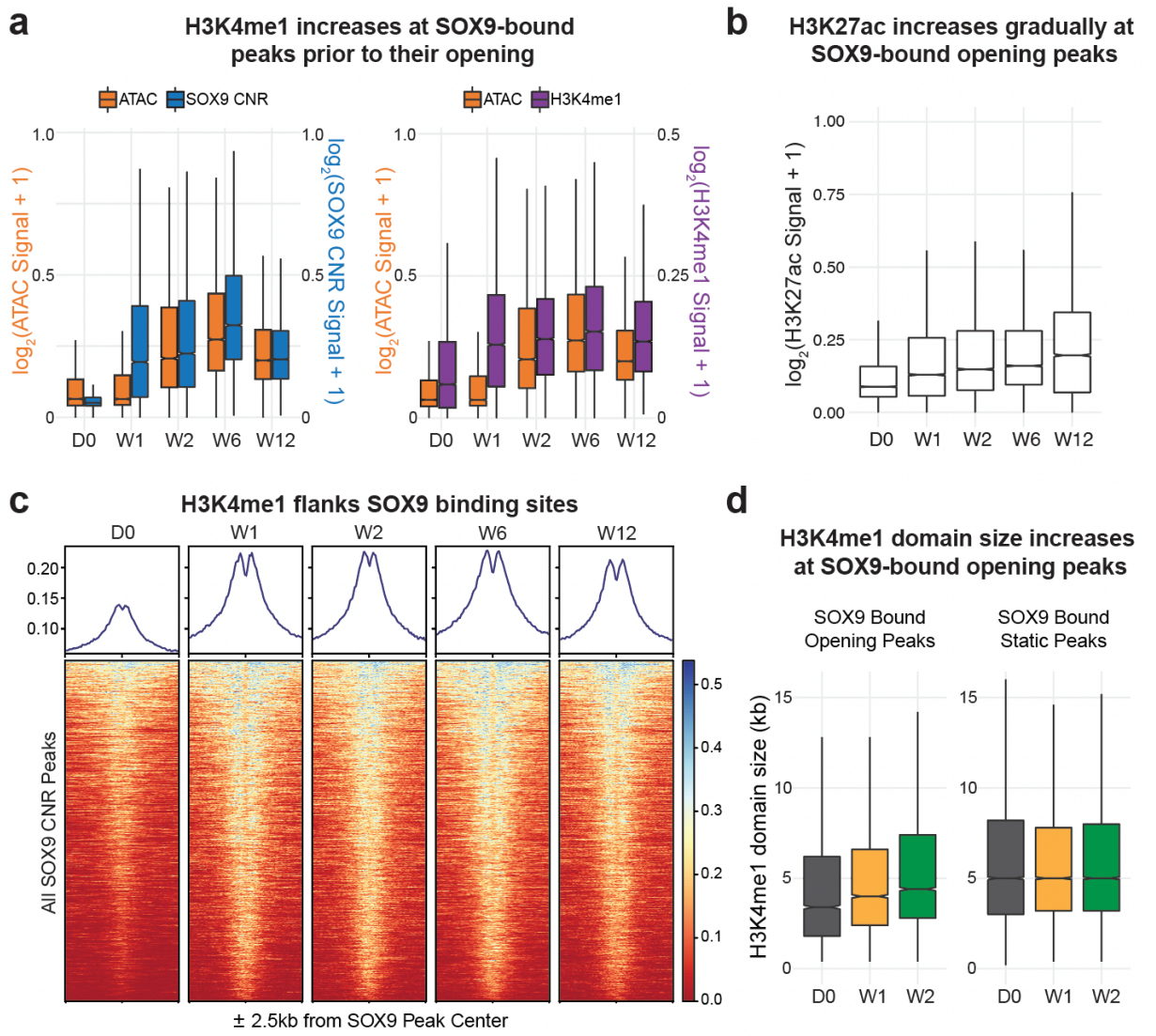

Fig. 4.3: Temporal epigenetic changes induced upon SOX9 binding across genome

a, (left) Boxplot comparing ATAC (orange) and SOX9 CNR (blue) signals at SOX9 peaks that transition from closed to open chromatin over time. Note that SOX9-binding increases markedly by W1, preceding chromatin accessibility at these sites by nearly a week. (right) Boxplot comparing ATAC (orange) and MINT-ChIP H3K4me1 (purple) signals at SOX9 peaks that open over time. Note that H3K4me1 follows the time course of SOX9 binding, again preceding chromatin accessibility. n=2 biological replicates. Boxplots are centered at median and bounds by 1st and the 3rd quartile, and whiskers extends to 1.5 times IQR on both ends

b, Boxplot of H3K27ac signals at opening SOX9 peaks overtime. Note that the signal increases gradually from W1 to W12. n=2 biological replicates. Boxplot is centered at median and bounds by 1st and the 3rd quartile, and whiskers extends to 1.5 times IQR on both ends.

c, Profile plot (top) and heatmap (bottom) showing the H3K4me1 signals at all SOX9 Cut-and-Run peaks across time points. Note the strong flanking pattern of the H3K4me1 signals adjacent to each SOX9 binding site (center dip) from W1 to W12.

d, Boxplot of H3K4me1 domain sizes at opening SOX9 peaks and SOX9 bound static peaks shows gradual increase of H3K4me1 domain size from W1 to W2 after SOX9 induction. n=2 biological replicates. Boxplots are centered at median and bounds by 1st and the 3rd quartile, and whiskers extends to 1.5 times IQR on both ends.

4.3 SOX9 directly interacts with epigenetic co-factors

To understand how SOX9 directly induced epigenetic changes at HF enhancers prior to chromatin accessibility, Nicholas and I started by identifying proteins that interact with SOX9 in primary EpdSCs. To accomplish this, we induced expression of *TRE* promoter driven fusion proteins SOX9-BioID2 and control GFP-NLS-BioID2 in *Krt14rtTA* primary EpdSCs *in vitro* using DOX (Fig 4.3a,b) (Kim et al., 2016). One week later, we purified and analyzed biotinylated SOX9-interacting proteins with mass spectrometry (Fig. 4.4a).

PCA analysis revealed that biological replicates were highly correlated and formed distinct clusters (Fig. 4.4c). Fifty-eight proteins were identified to interact specifically with SOX9 compared to NLS-GFP in EpdSCs (Table 2). GO-term analysis and protein function showed that SOX9-interacting proteins were mainly chromatin and DNA-binding proteins enriched with functions in nuclear activity and chromatin modifications (Fig. 4.4d). Among the most robust SOX9 interactions were core members of the SWI/SNF chromatin remodeling complex (ARID1a/b and SMARCD2), TAF9 (TFIID) that binds to the TATA box, required for RNA polymerase II-mediated transcriptional induction, and AP1 transcription factors (FOSL2 and JUNB) (Fig. 4.5 and Table 2).

Histone modifiers, key interactors of pioneer factors in other developmental contexts, were also prominently featured (Balsalobre and Drouin, 2022). Given that SOX9-induced opening peaks were significantly more enriched at enhancers than promoters (Fig. 3.1c), I was intrigued to

discover modifiers of two histone marks that are enriched at active enhancers, namely Ep300, the acetyltransferase for H3K27ac, and MLL3 and MLL4, histone methylases that not only generate H3K4me1 but may also have other emerging roles in enhancer activation (Dorighi *et al.*, 2017; Rao and Dou, 2015; Sze and Shilatifard, 2016).

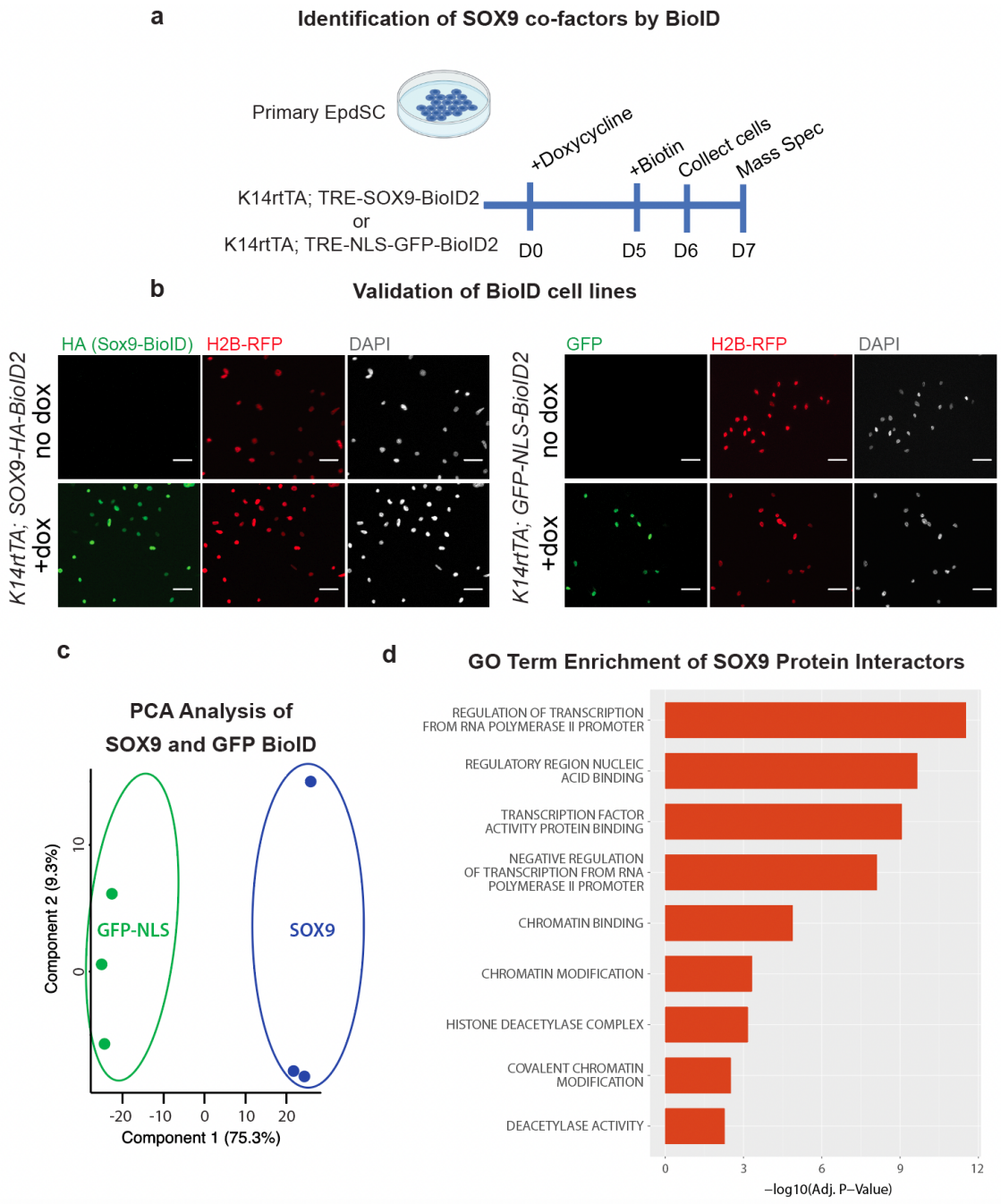

Fig. 4.4: Identify SOX9 interacting proteins with proximity labeling

a, Schematic of BioID2 proximity labeling of proteins that interact with SOX9.

b, Immunofluorescence validation of *Krt14rtTA* primary keratinocytes transduced with TRE-HA-SOX9-BioID2; H2B-RFP (left) or TRE-GFP-NLS-BioID2; H2B-RFP (right). Transgenes were induced with doxycycline and immunolabeled for HA (left) or GFP (right). All scale bars are 50μm.

c, PCA analysis of GFP and SOX9 BioID protein interactors demonstrate sample-specific clusters.

d, Gene ontology enrichment of proteins specific to SOX9 compared to GFP-NLS (binomial is used to calculate the significance).

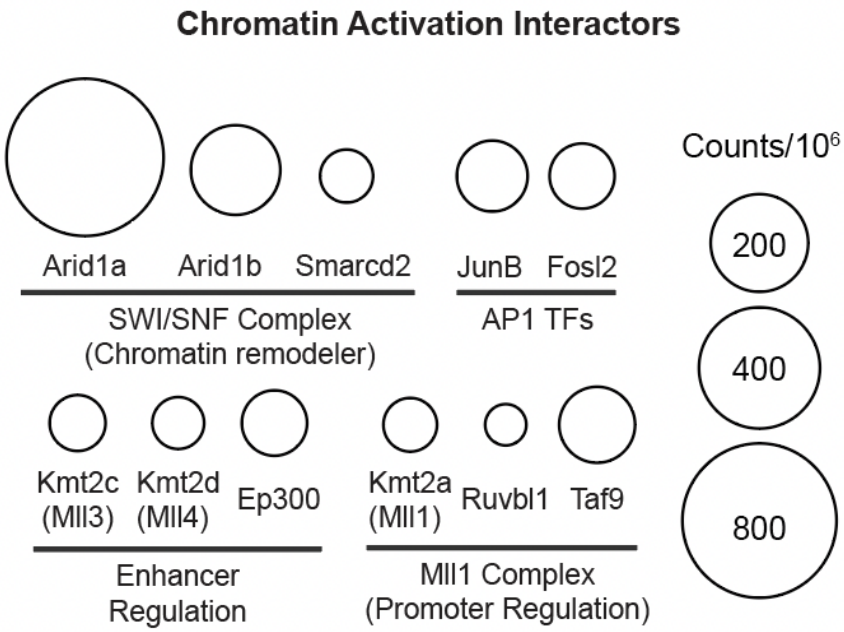

Fig. 4.5: Top SOX9 interacting proteins identified by BioID
Selected SOX9-interacting proteins detected with mass spectrometry and which fall into the top GO-terms of chromatin remodelers of the SWI/SNF family; transcription initiation factors (TAF9), AP1 transcription factors (TFs); and enzymes that modify histones (MLL1, MLL3, MLL4, p300). See Table 2 for full list. Circle size corresponds to the strength of the hit as delineated at right.

While I observed SOX9's interacting with epigenetic co-factors in vitro, ultimately, I want to test if *in vivo* SOX9 directly recruits MLL3/4 to closed chromatin, cooperating with the observation where H3K4me1 increases at SOX9 targeted enhancers before chromatin opening (Fig 4.2c). To investigate this further, I conducted MLL3/4 Cut-and-Run experiments, hypothesizing that if SOX9 indeed directly regulates MLL3/4 recruitment to chromatin in vivo, SOX9 binding should be enriched in newly targeted MLL3/4 sites. I observed a gradual increase in MLL3/4 CNR signal from D0 and W2 at SOX9-bound enhancers, preceding chromatin opening (Fig. 4.6a). In

addition, I observed a significant enrichment of SOX motifs in W1 unique MLL3/4 CNR peaks when compared to D0, whereas no motif enrichment was observed in D0 unique peaks (Fig. 4.6b).

To strengthen the observation of physical interaction between SOX9 and MLL3/4, I performed co-immunoprecipitations (co-IP). I used CRISPR/Cas9 to ablate MLL4 in these EpdSCs to ensure accurate band identification, given MLL4's substantial size (>500kDa) and high probability of degradation (Fig. 4.6c,d). The co-IP results on primary EpdSC lysates with or without SOX9 induction utilizing antibody against SOX9's myc-tag, followed by probing for MLL4 with immunoblot, showed specific detection of MLL4 in whole cell lysates and pulldown elution only in the sample with SOX9 induction (Fig. 4.6e). These findings provide a better understanding of how SOX9 operates as a pioneer factor in activating HF enhancers, as it not only binds to closed chromatin but also recruits epigenetic co-factors to modify surrounding histones, preceding chromatin opening.

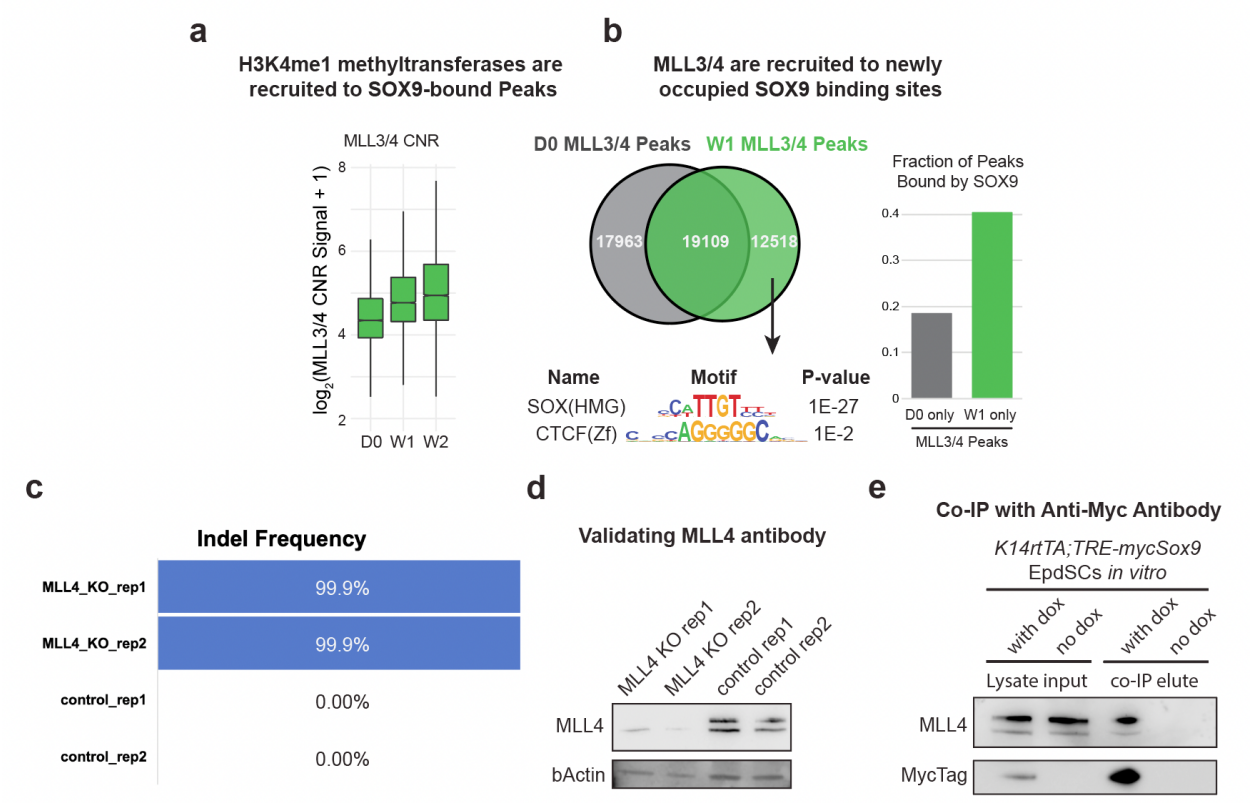

Fig. 4.6: SOX9 directly recruits epigenetic co-factors to closed chromatin prior to opening

a, Boxplot of MLL3/4 CNR signals reveals the appearance of MLL3/4 at SOX9-bound peaks following DOX. n=2 biological replicates. Boxplot is centered at median and bounds by 1st and the 3rd quartile, and whiskers extends to 1.5 times IQR on both ends.

b, Venn diagrams providing further evidence that the *de novo* MLL3/4 peaks that appear between D0 and W1 are highly enriched for bound SOX9 (binomial p = 1E-27), whereas the 17963 MLL3/4 peaks lost at W1 do not show any significant motif enrichment over all D0 MLL3/4 peaks and are not associated with SOX9 bound sites, inconsistent with a repressor role for SOX9.

c, Two *Mll4* null keratinocyte cell lines (*Krt14rtTA; TRE-mycSox9*) were generated with CRIPSR/Cas9 with 99.9% indel frequency.

d, The *Mll4* null keratinocyte cell lines validated the efficacy of the MLL4 antibody by immunoblot, which detected a ~500 kDa protein in the control but not the knockout (KO) cell lines. MW, molecular weight.

e, Immunoblot showing that the ~500 kDa protein, identified in (g) as MLL4, is pulled down with mycSOX9 in an anti-myc tag antibody immunoprecipitation. See also in associated source data.

CHAPTER 5. SOX9 competes for epigenetic co-factors to simultaneously activate HF fate and silence EpdSC fate

5.1 SOX9 indirectly silences EpdSC genes before chromatin remodeling

Pioneer factors situate at the junction of stem cell fate switching, which requires both new fate activation and efficient old fate repression. In line with this concept, I observed efficient suppression of epidermal fate-related genes upon SOX9 induction prior to the induction of HF-related genes. A simple hypothesis is that SOX9 directly repressed epidermal enhancers by recruiting co-repressors such as polycomb repressive complexes (Balsalobre and Drouin, 2022).

However, while SOX9 was strongly associated with ATAC peaks that opened over time, it was rarely found in closing peaks. In fact, only 3% of peaks that closed over time were bound by SOX9 (Fig. 5.1a). The differences were even more noticeable when I restricted the analysis to peaks that opened or closed in the first two weeks (Fig. 5.1b). SOX9 binding was detected in almost 50% of ATAC peaks that opened in the first two weeks, whereas SOX9 bound less than 5% of closing peaks. Additionally, while the closing peaks that SOX9 did not bind featured epidermal-associated GO-terms, the SOX9-bound opening peaks were enriched with HF-fate related GO-terms (Fig. 5.1c,d). This strong association between SOX9 and opening but closing peaks was con consistent with SOX9 interacting proteins identified by BioID, which were mainly chromatin-activating enzymes (Fig. 4.5). Therefore, I hypothesized that the silencing of epidermal enhancers and genes were regulated by SOX9 indirectly.

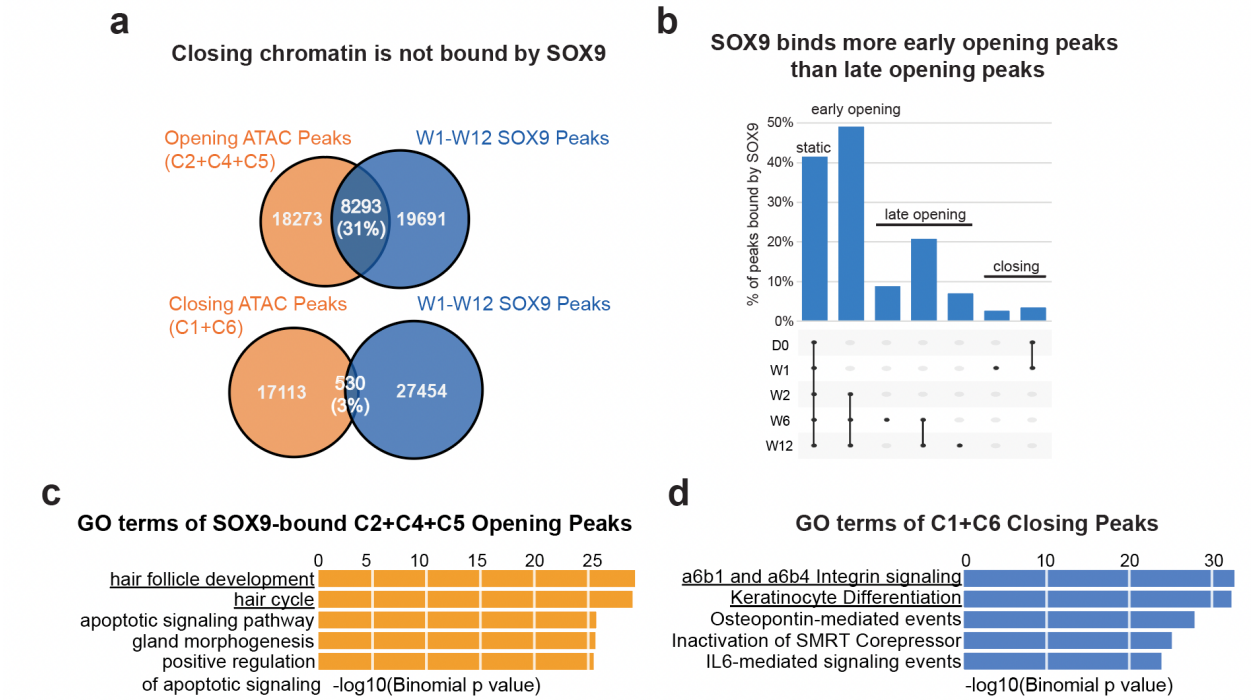

Fig. 5.1: SOX9 silences epidermal enhancers through indirect mechanisms

a, (top) Venn diagram shows robust overlap between opening peaks (C2+C4+C5) and SOX9 peaks. (bottom) Venn diagram shows only 3% overlap between closing peaks (C1+C6) and SOX9 peaks.

b, Upset plot shows % of peaks bound by SOX9 in the top peak sets from Fig. 2D. Note that peaks opened by W2 are more often bound by SOX9 than later opening or closing peaks.

c, GO terms enriched in SOX9-bound opening peaks (C2, C4, C5)

d, GO terms enriched in SOX9-bound all closing peaks (C1, C6)

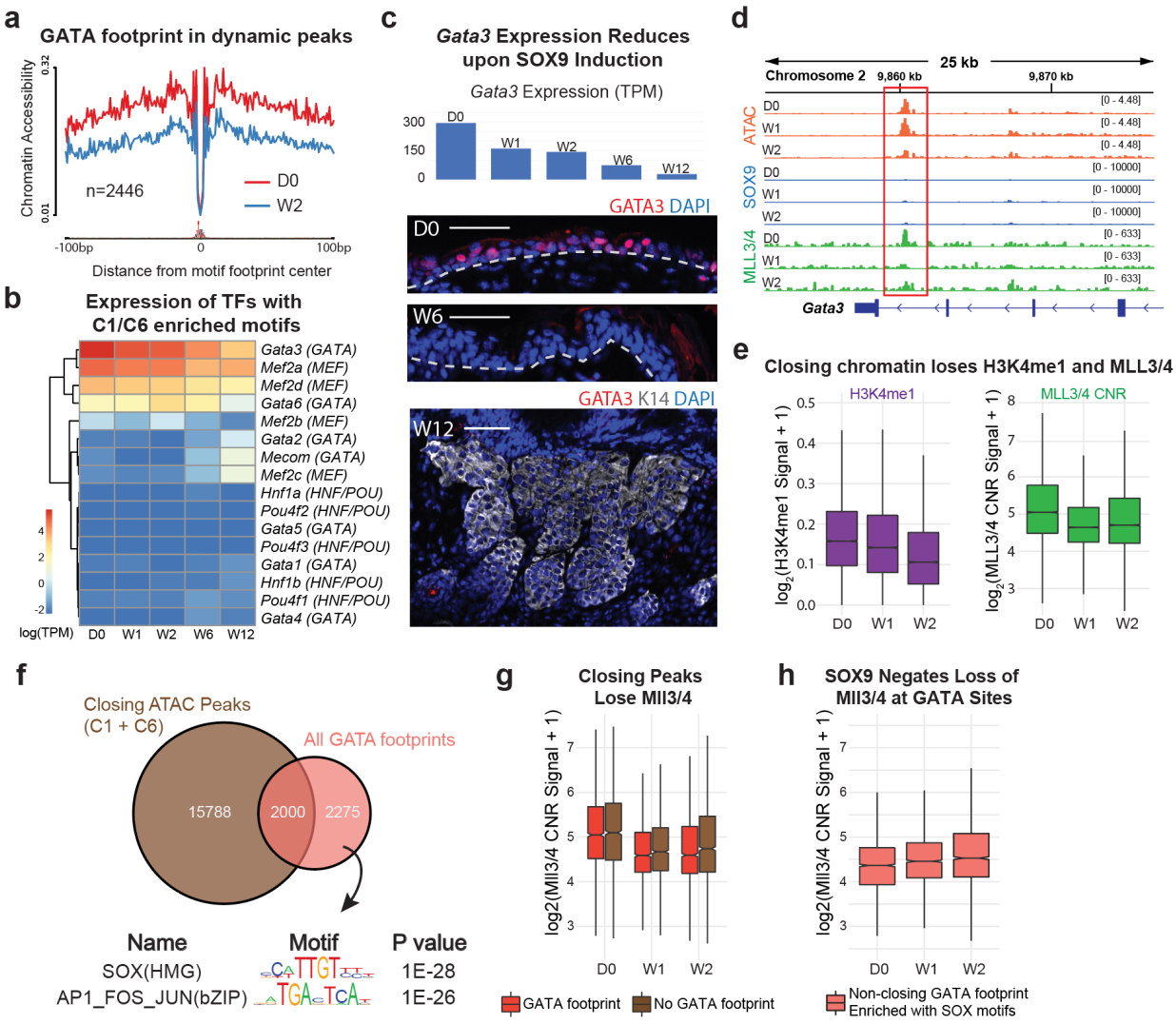

Fig. 5.2: SOX9 silences epidermal enhancers through indirect mechanisms

a, ATAC footprint analysis at D0 and W2 shows a decrease in chromatin accessibility at GATA footprint in dynamic ATAC peaks following SOX9 induction.

b, EpdSCs expression of transcription factors that belong to TF families whose motifs are enriched in closing ATAC peaks (C1 and C6 in Fig. 3C) (binomial is used to calculate the significance).

c, (top) Transcript levels of *Gata3* over time following SOX9 induction. (bottom) GATA3 immunofluorescence of epidermis at D0, W6 and W12. Scale bars, 50µm.

d, ATAC, SOX9 CNR and MLL3/4 CNR tracks at the *Gata3* locus, showing that by W1 after SOX9 induction, MLL3/4 CNR peaks have diminished markedly, and by W2, ATAC peaks are closed, even though CNR shows no SOX9 binding in this region (red box).

e, Boxplots showing loss of MLL3/4 and H3K4me1 signal beginning at W1 post SOX9-induction and specifically at ATAC peaks that close within W2 (C1, C6). n=2 biological replicates.

f, (top) Venn diagram shows overlap between closing peaks (C1,C6) and GATA footprints. (bottom) Interestingly, peaks with GATA footprints that were not closed post SOX9-induction were enriched with SOX and AP1 motifs.
g, Boxplots showing loss of MLL3/4 signal at closing epidermal enhancers with or without GATA footprints post SOX9-induction. n=2 biological replicates.
h, Boxplots showing MLL3/4 signal was not reduced at non-closing peaks with both GATA and SOX motifs. n=2 biological replicates.
All boxplots are centered at median and bounds by 1st and the 3rd quartile, and whiskers extends to 1.5 times IQR on both ends.

To further understand how epidermal fate is silenced, I focused on the GATA factors, which are master transcription factors of the epidermal fate. HOMER analyses demonstrated that GATA motifs were markedly enriched in the closing ATAC peaks in C1 and C6 (Fig. 3.2a). In addition, the GATA TF footprint was strongly accessible at D0 and closed at W2 upon SOX9 induction (Fig. 5.2a). Agreeing with the motif dynamics, GATA factors, especially GATA3's transcript and protein expression level also dropped concurrently with the closing of GATA motifs (Fig. 5.2b,c).

GATA3 has been recognized as a pioneer factor in multiple studies (Takaku et al., 2016; Tanaka et al., 2020). Even a minor reduction of *Gata3* expression at W1 could potentially result in the closure of epidermal enhancers, given enhancers can be sensitive to changes in pioneer factor levels (Naqvi et al., 2023). However, enhancers closed by SOX9 did not appear to depend on GATA binding, as they display consistent levels and dynamics of MLL3/4 loss, regardless of the presence or absence of GATA footprints (Fig 5.2g). Moreover, enhancers with GATA footprints that remain open after SOX9 induction were notably enriched with SOX motifs (Fig 5.2f). These enhancers, characterized by the presence of both GATA and SOX motifs, exhibit a slight increase in MLL3/4 binding *in vivo* (Fig 5.2h). These observations challenge the notion that the closure of epidermal enhancers is solely due to GATA3 loss. Instead, they further highlights the role of SOX9 in recruiting co-factors such as MLL3/4.

At the *Gata3* gene locus, a putative enhancer markedly declined in chromatin accessibility over the first 2 weeks of SOX9 induction (Fig. 5.2d). In the meantime, I observed MLL3/4 binding declined preceding the closure of this enhancer at W1. On a global level, MLL3/4 binding at all opening SOX9-bound HF enhancers was robust by W2, as was H3K4me1 modification (Fig. 4.3a and Fig. 4.6a). In contrast, the over 6,000 SOX9-independent enhancers including the GATA3 enhancer that closed during the first 2 weeks of SOX9 induction showed reduced MLL3/4 association starting from W1 and a more gradual loss of H3K4me1 (Fig. 5.2e). These results suggest an attractive possibility that SOX9, prior to binding to HFSC enhancers, could be capturing co-factors, such as MLL3/4, away from active EpdSC-enhancers to achieve their fast silencing.

5.2 Competition for SOX9-interacting epigenetic co-factors

To test this possibility, I started by examining the functional domains of pioneer factor SOX9. SOX9 binds DNA through its high-mobility group (HMG) domain and interacts with co-factors with the trans-activation (TA) domain (Fig. 5.3a) (Ming et al., 2022). By engineering DOX-inducible myc-tagged wildtype (WT) and mutant forms of SOX9 that either lacked either the co-factor binding (ΔTA) or the DNA binding (ΔHMG) domain, I could more carefully study the interaction among SOX9, chromatin and co-factors (Fig. 5.3b). I transduced primary EpdSCs with one of the three SOX9 constructs and validated that the immunofluorescence levels were comparable (Fig. 5.3c). As assessed by immunoblot, the ectopically induced proteins all had the expected size (Fig. 5.3d). Additionally, co-IP followed by immunoblot of MLL4 confirmed that only WT and ΔHMG-SOX9, but not ΔTA-SOX9, had the ability to bind MLL4, substantiating that the TA domain of SOX9 was necessary for interacting with co-factors (Fig. 5.3e).

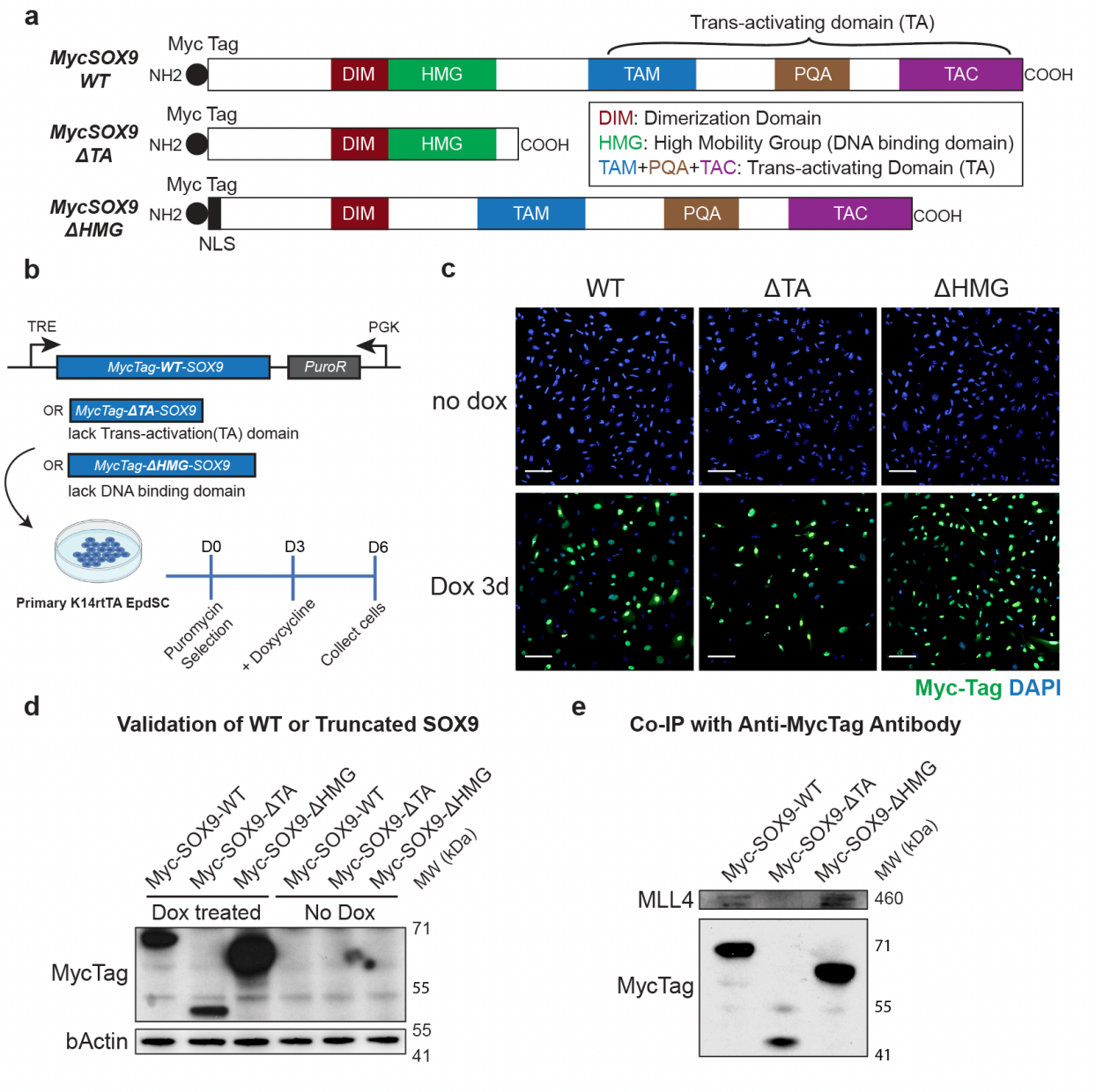

Fig. 5.3: Truncated SOX9 displays impaired DNA binding or co-factor recruitment

a, Schematic illustrating constructs engineered to express WT SOX9 and two variants of SOX9.

b, Schematic of inducing of myc-tagged WT or mutant versions of SOX9 in transduced *Krt14rtTA* cultured EpdSCs.

c, Immunofluorescence reveals similar intensities of WT and mutant SOX9 in induced EpdSCs. Scale bars, 100μm.

d, Immunoblot validating the sizes of different versions of SOX9. MW, molecular weight.

e, Immunoblot showing that the transactivating (TA) domain of SOX9 is sufficient in binding MLL4.

I next used a myc-tag antibody recognizing all three SOX9 variants equivalently and performed Cut-and-Run to confirm that while the WT and ΔTA-SOX9 variants can bind SOX9 binding sites, ΔHMG-SOX9 failed to bind to DNA compared to primary EpdSCs without induction (Fig. 5.4a). When compared in parallel to ATAC-seq signals with different variants of SOX9, it was interesting that while no induction or ΔHMG-SOX9 had no impact on peaks otherwise opened by WT-SOX9 as expected, ΔTA-SOX9 only bound to chromatin that was already accessible in EpdSCs but not opening HFSC enhancers (Fig. 5.4a). In line with these findings, the 1207 peaks that were accessible prior to the induction of SOX9 variants and bound by ΔTA-SOX9 did not exhibit MLL3/4 recruitment or increase in H3K4me1 modification (Fig. 5.4b). These results suggested that when SOX9 failed to bind co-factors, it lacked the characteristic feature of pioneer factors.

It was also striking that even though ΔHMG-SOX9 did not bind DNA directly, it indirectly but profoundly remodeled chromatin accessibility landscape. Upon induction, nearly 10,000 ATAC peaks were closed, while more than 8,000 peaks exhibited opening (Fig. 5.4c). Given that ΔHMG-SOX9 lacked DNA binding ability, the profile of opening peaks would be expected to differ from that of WT-SOX9 significantly. GO-term profile of the opening peaks in ΔHMG-SOX9 induced EpdSCs reflected a stressed state rather than the HFSC enhancers typically opened by WT-SOX9 (Fig. 5.4d). However, the closing peaks in WT and ΔHMG-SOX9 induced EpdSCs displayed a 64% overlap, and the GO-terms are composed of the same EpdSC genes indirectly silenced by WT-SOX9 (Fig. 5.5a). Additionally, the closure of these enhancers by WT and ΔHMG-SOX9 reduced the expression of epidermal master transcription factors such as *Gata3* and *Trp63,* whereas ΔTA-SOX9 failed to do so (Fig. 5.5b). To test the mechanism behind ΔHMG-SOX9's transcriptional and chromatin impact, I performed MLL3/4 Cut-and-Run since ΔHMG-SOX9 can still interact with co-factors. MLL3/4 Cut-and-Run signal diminished over closing EpdSC

enhancers by ΔHMG-SOX9, which is consistent with the hypothesis that SOX9 closes chromatin by competing for and redistributing co-factors in dependent from DNA binding (Fig. 5.5c).

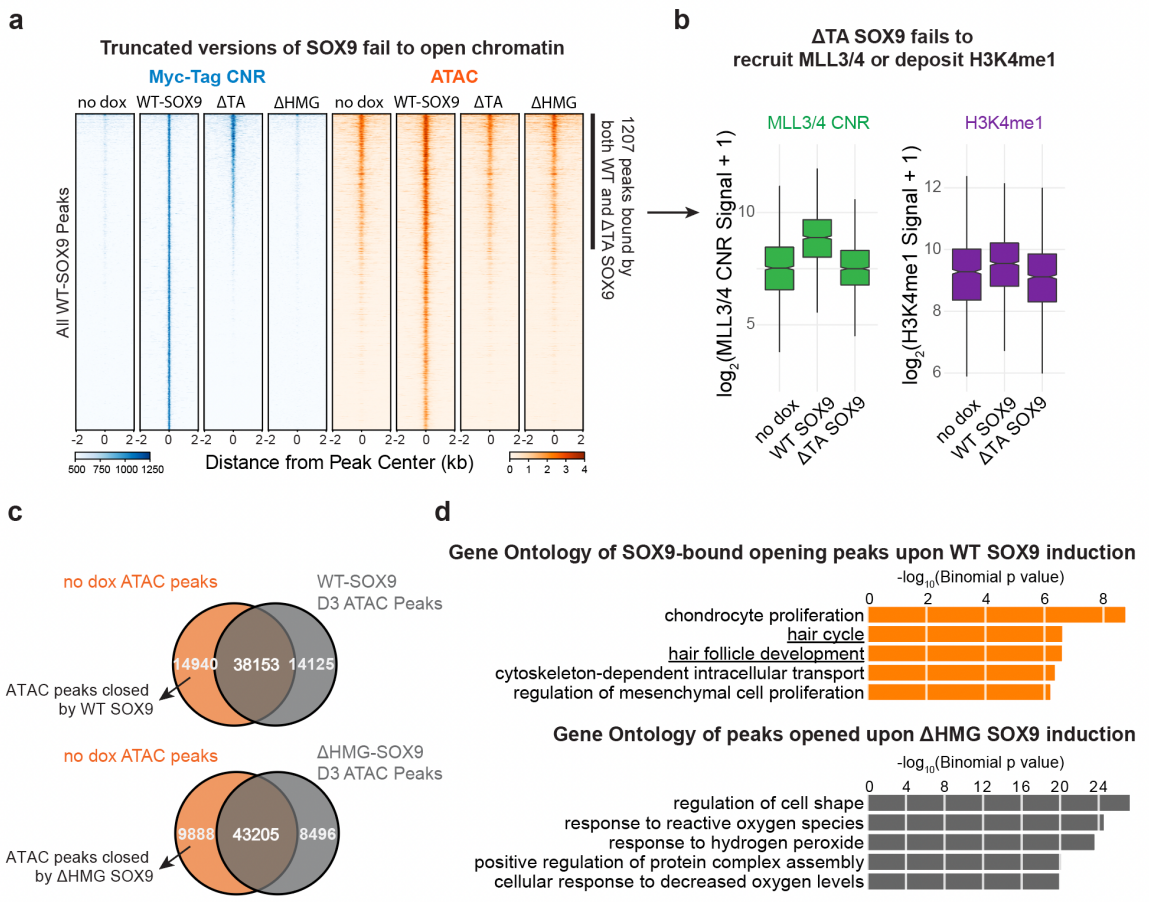

Fig. 5.4: Truncated SOX9 displays impaired pioneering functions

a, Profile plot and heatmap showing myc-tagged WT or variant SOX9 binding (blue) and accessibility (orange) prior to and after DOX. Peaks are sorted the same way across samples. Note that ΔHMG-SOX9 fails to bind DNA, and ΔTA-SOX9 binds only to the subset of SOX9 peaks which were already accessible prior to DOX. Both mutants of SOX9 failed to open chromatin *de novo*.

b, Boxplot comparing MLL3/4 CNR and H3K4me1 signals at the peaks that are bound by WT and ΔTA-SOX9. Note that only WT-SOX9 brought additional MLL3/4 and deposited more H3K4me1 to these peaks. n=2 biological replicates.

c, Venn diagrams show that the peak sets closed by WT and ΔHMG-SOX9 are comparable in size.

d, (top) Top 5 biological process gene ontologies of genes associated with SOX9-bound opening peaks upon WT SOX9 induction *in vitro* from GREAT. (bottom) Top 5 biological process gene ontologies of genes associated with peaks opened upon ΔHMG SOX9 induction *in vitro* from GREAT. Binomial is used to calculate the significance.

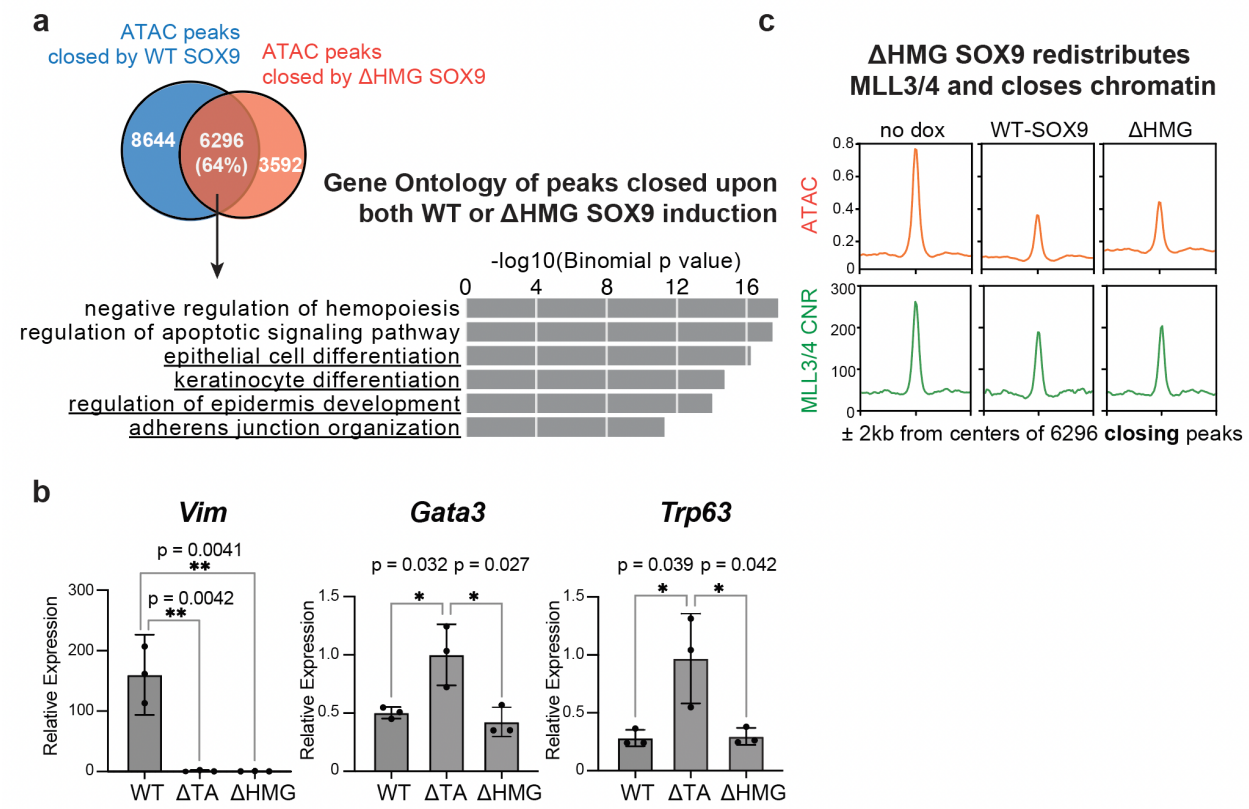

Fig. 5.5: DNA-binding impaired SOX9 can still silence epidermal fate

a, Venn diagram shows overlap between the ATAC peaks that are closes by WT-SOX9 or ΔHMG-SOX9 induction. GO terms reveal that epidermal enhancers close upon WT or ΔHMG-SOX9 induction (binomial is used to calculate the significance).

b, qPCR analysis of genes that are directly induced (*Vim*) or indirectly repressed (*Gata3* and *Trp63*) by SOX9 in EpdSC cells. All the error bars are mean ± SD. Statistical significance from two-tailed t-test is denoted by *($p<0.05$) **($p<0.01$). n=3 biological replicates.

c, Profile plots showing that ATAC and MLL3/4 CNR signals wane at EpdSC enhancers upon WT and ΔHMG-SOX9 induction. Note that when SOX9 cannot bind to DNA (ΔHMG), it still diminishes MLL3/4 at endogenous EpdSC enhancers and closes their chromatin.

All boxplots are centered at median and bounds by 1st and the 3rd quartile, and whiskers extends to 1.5 times IQR on both ends.

It has been proposed that the cooperative binding of pioneer factors and other transcription factors is essential for chromatin activation. Since I observed epidermal AP1 transcription factors in the SOX9 interactome, I sought to examine the cooperativity between SOX9 and AP1 factors. In the opening peaks bound by SOX9, the presence of SOX and AP1 motifs was predominantly observed in close proximity to each other, often within one nucleosome distance (Fig. 5.6a). Moreover, consistent with the observed dynamics of MLL3/4, *in vivo* footprint analysis demonstrated a decrease in AP1 binding in closing epidermal enhancers that were not bound by SOX9, while an increase in AP1 binding was observed in SOX9-bound chromatin (Fig. 5.6b). Additionally, co-IP experiments in WT and ΔHMG-SOX9 induced EpdSCs showed a physical interaction between SOX9 and c-JUN, an AP1 TF (Fig. 5.6c). These finding provides support for the hypothesis that SOX9 plays a role in guiding AP1 factors to their canonical binding sites upon the opening of HF enhancers.

Notably, ChromVAR analysis uncovered the presence of AP1 binding sites in both closing and opening enhancers (Fig. 3.1e,f). This suggests that the interaction between SOX9 and AP1 may play a crucial functional role in both the opening of HFSC enhancers and the closing of epidermal enhancers. To investigate the possibility of enhancers competing for AP1 binding, I created a EpdSCs where AFOS, a dominant negative version of c-FOS could be induced with DOX in the presence or absence of SOX9. AFOS has the ability to form heterodimers with other AP1 factors at their DNA binding domains, thereby obstructing their binding to DNA (Ahn et al., 1998; Olive et al., 1997). I carried out ATAC-seq following AFOS induction. Interestingly, when AFOS was expressed alone, it mimicked the closing effects of WT and ΔHMG-SOX9 on EpdSC enhancers. In the peaks bound by WT-SOX9, it was evident that AFOS had a clear inhibitory effect on the opening of HFSC enhancers (Fig. 5.6d).

Based on the findings thus far, it appears that AP1 factors, similar to MLL3/4, play a dual role in regulating stem cell fate. Additionally, considering the observed interactions with members of the SWI/SNF complex (Fig. 4.5 and 5.5c), the data suggests that SOX9 may recruit the SWI/SNF complex to facilitate chromatin accessibility for transcription as well. Therefore, I decided to focus the investigation on AP1 factors and the SWI/SNF complex to examine whether other members of the interactome are also involved in the potential enhancer competition for SOX9 binding partners.

To do so, I performed Cut-and-Run analysis for members of the AP1 factors and SWI/SNF complex. The results revealed that WT-SOX9 led to an increase in JUN (AP1) binding at the SOX9-bound opening peaks, whereas JUN binding decreased at the closing epidermal peaks. Remarkably, even though ΔHMG-SOX9 failed open HF enhancers, it still reduced JUN binding at the closing epidermal peaks (Fig. 5.6e). Likewise, in the Cut-and-Run analysis of the structural component (ARID1a) and enzymatic component (BRG1) of the SWI/SNF complex, I observed a decrease in their association with epidermal enhancers following the induction of both WT and ΔHMG-SOX9 (He et al., 2020). However, the enhanced association of the SWI/SNF complex with SOX9-bound peaks was specific to the induction of WT SOX9 but not ΔHMG-SOX9 (Fig. 5.6f).

Collectively, these findings indicate that SOX9 plays a crucial role in redistributing transcription factors and epigenetic co-factors that are common to enhancers involved in different cell fates, namely epidermal and hair follicle. This competition among enhancers is influenced, at least in part, by the availability of chromatin remodeling factors. Supporting this notion, I observed that overexpression of ARID1a in the presence of SOX9, rescued the closure of epidermal enhancers (Fig. 5.6g,h).

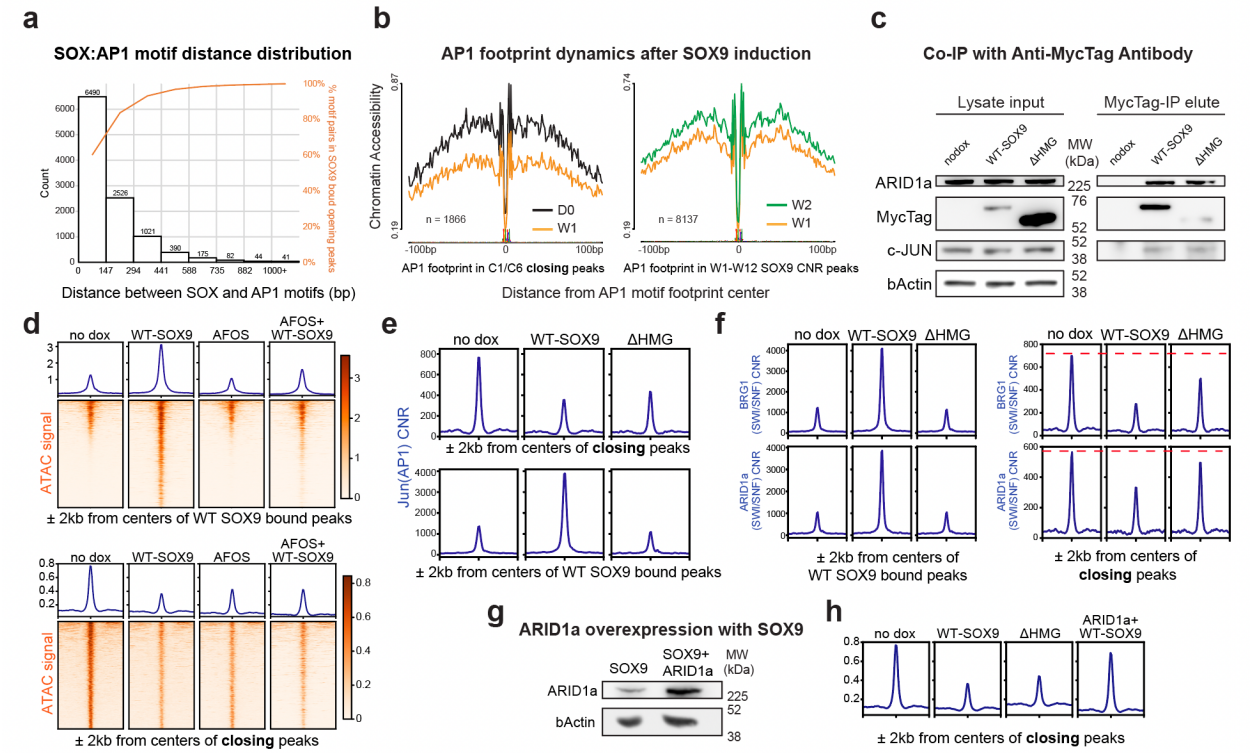

Fig. 5.6: SOX9 redistributes epigenetic co-factors to achieve efficient fate switching

a, Distribution of the distance between a SOX motif and its closest AP1 motif in the SOX9 bound opening peaks. Note that the x-axis is binned by multiplies of one nucleosome size (147bp). The cumulation plot of the distribution is shown in orange on the secondary y-axis.

b, AP1 footprint analyses in closing ATAC peaks and SOX9 CNR peaks. Note that chromatin accessibility at AP1 footprints decreases over closing epidermal peaks by W1 and increases over opening SOX9 bound peaks by W2.

c, Immunoblot showing that both WT-SOX9 and ΔHMG-SOX9 are capable of binding c-JUN and ARID1a.

d, (top) Profile plots and heatmaps comparing ATAC signals at WT-SOX9 bound peaks *in vitro*. While WT-SOX9 can open chromatin, AFOS inhibits the opening as shown in the last column. (bottom) Profile plots and heatmaps comparing ATAC signals at closing peaks *in vitro*. AFOS phenocopies the indirect closing effect of SOX9 on AP1-associated epidermal enhancers.

e, Profile plots showing that JUN CNR signals are reduced at EpdSC enhancers upon WT-SOX9 and ΔHMG-SOX9 induction (top). Supportive of competition, WT-SOX9 recruits AP1 to HFSC enhancers, while ΔHMG-SOX9, lacking the DNA binding domain, fails to do so (bottom).

f, Profile plots showing that both BRG1 and ARID1a of the SWI/SNF complex behave similarly to AP1 and MLL3/4: they are recruited to the opening HFSC enhancers by only WT-SOX9 (left), while both WT and ΔHMG-SOX9 reduce SWI/SNF association with epidermal enhancers (right). The red dotted line denotes CNR levels of indicated target at closing peaks before DOX.

g, Immunoblot validating ARID1a can be induced 3x higher in the SOX9 expressing keratinocytes. For full blots, see associated source data.

CHAPTER 6. Discussion and future directions

6.1 Temporal regulation of cell fate switching by pioneer factors

6.1.1 Significance

As first coined in liver development, pioneer factors have the ability to bind closed chromatin, which renders other transcription factors and transcriptional machinery inaccessible, and their transcriptional dynamics are often correlated with fate-switching processes (Balsalobre and Drouin, 2022; Gualdi *et al.*, 1996; Iwafuchi-Doi and Zaret, 2014). Various pieces of literature have documented that this correlation is functionally relevant and the fact that pioneer factors epigenetically open and activate the enhancers of the new fate that was once buried in closed chromatin (Balsalobre and Drouin, 2022). However, studies aiming to address the molecular mechanisms behind fate-switching often took a reductionist approach where cells were removed from their native niche, and reprogramming events were examined *in vitro*. While studies with cultured cells provided important, valuable and detailed insights into important pioneering epigenetic events with their kinetics of reprogramming accelerated, I still needed a clear picture of the temporal sequence of events involved in this process *in vivo*. Additionally, given the frequent dysregulation of pioneer factors in cancers, dissecting the sequential steps that lead to the establishment of a new tumorigenic chromatin state while concurrently suppressing the normal cell fate in a healthy tissue becomes crucial (Iwafuchi-Doi and Zaret, 2014).

6.1.2 Choosing SOX9 re-activation as a model for pioneer factor induced fate-switching

Considering these questions, my study tackles the mechanisms underlying stem cell fate switching *in vivo*. I focused on SOX9, which directs fate decisions in many different tissues and states. SOX9 was first studied to have a critical role in male sex determination with its function and sequence similarity to SRY, Sex-determining Region Y (Wright et al., 1993). It regulates the

development of Sertoli cells downstream of SRY, which is necessary for testis formation in the mammalian embryo. In the absence of SOX9, the gonad develops into an ovary. As the sex reversal phenotype and SOX9 mutations were often found to be associated with a severe skeletal malformation syndrome, campomelic dysplasia, SOX9's role in chondrocyte differentiation has been illustrated in multiple studies (Foster et al., 1994; Kwok et al., 1995; Wagner et al., 1994; Wirth et al., 1996). SOX9 is involved in the early differentiation of mesenchymal cells into chondrocytes and directly targets essential cartilage matrix genes, such as types II collagen (Lefebvre et al., 1997). In addition, SOX9 displays cooperativity with SOX5 and SOX6 in fully activating the expression of cartilage-specific genes (Akiyama et al., 2002).

During skin development, SOX9 is expressed during embryogenesis when multipotent skin progenitors activate SOX9 to induce the formation of hair placodes and repress the epidermal fate (Nowak *et al.*, 2008; Ouspenskaia *et al.*, 2016). The antagonism between WNT and SHH signaling pathways is critical for the specification and expansion of stem cells during HF development, even before the formation of their niche. SHH signaling is crucial for expanding the population of $SOX9^{+}$ HF progenitor cells and enhancing their fate commitment, likely through the suppression of WNT signaling. SOX9, being the downstream target of this process, in turn specifies HFSCs from $SOX9^{neg}$ epidermal progenitors (Ouspenskaia *et al.*, 2016).

In adult skin, SOX9 loss of function leads to HF malformations and deficiencies in the HFSC compartment (Vidal *et al.*, 2005). Besides being essential for maintaining HFSCs, SOX9 regulates the expression of several important signaling factors that help maintain the stem cell niche (Kadaja *et al.*, 2014). And finally, SOX9 is repressed when HFSCs commit to becoming short-lived transit-amplifying cells that give rise to the differentiated layers in a mature HF (Adam *et al.*, 2015; Yang *et al.*, 2017).

SOX9 is not only essential in regulating HFSC fate during skin development and homeostasis, but also important in cancers originating from EpdSCs, affecting their initiation and progression. SHH, a key regulator of HF progenitor growth and differentiation as aforementioned, is frequently activated in BCC originated from EpdSCs, and this activation has been linked to the increased SOX9 expression. Subsequently, SOX9 contributes to the proliferation and survival of BCC cells. The oncogenic activation of SOX9 in the adult epidermal cells with constitutive SHH leads to hair placode-like invaginations that further progress to malignancy (Larsimont *et al.*, 2015).

6.1.3 Findings and open questions

Regarding the order of pioneering events

By developing a genetically modified mouse model, I re-activated SOX9 in adult EpdSCs at levels similar to those observed in natural physiological contexts, i.e., HFSCs. This mouse model provided a unique system for investigating the *in vivo* temporal events of SOX9-induced fate-switching over a 12-week period. Induction of SOX9 in EpdSCs resulted in the suppression of the epidermal fate by W1 and the activation of HF fate by W2 on a transcription level. This experimental setup also mirrored the conditions in BCC's progression, where SOX9 expression remains constitutively active for 12 weeks (Fig. 6.1).

On a molecular level, Adam et al. had predicted the pioneer function of SOX9 by observing that SOX9 binds in HF super-enhancers, clusters of H3K27ac-rich enhancers that served as putative regulatory hubs and contained multiple DNA motifs for transcription factor binding including SOX9 (Adam *et al.*, 2015; Parker et al., 2013; Whyte et al., 2013). It was also clear that SOX9 could induce a handful of crucial HF proteins when induced in EpdSCs. However, owing to the technical difficulty of collecting millions to billions of EpdSCs from mice for ChIP-seq, if

and how SOX9 could bind closed chromatin were never directly answered *in vivo* (Skene and Henikoff, 2017). Therefore, I adapted a low-input ChIP-seq protocol, Cut-and-Run, to be able to test if SOX9 is a genuine pioneer factor directly (Meers et al., 2019a; Skene *et al.*, 2018; Skene and Henikoff, 2017).

With ATAC-seq and SOX9 Cut-and-Run over the course of SOX9-induced fate-switching, it became clear that SOX9 directly binds its cognate motifs in both open and closed chromatin, especially at intergenic regions rather than promoters related to HF fate. These results not only concur with pioneer factors' defining feature of accessing compacted chromatin but also provide additional examples of their preference for binding at intergenic regions (Chronis *et al.*, 2017).

By profiling histone modification over time, I illustrated that these SOX9 bound intergenic regions were enhancers and became activated according to their enrichment of H3K4me1 and H3K27ac, respectively. Moreover, the priming of these HF enhancers with H4K4me1 was observed a week before they became open, a drastic delay rarely documented in cultured reprogramming (Mayran et al., 2018). The detection of an increased MLL3/4 binding and the physical interaction between SOX9 and MLL3/4 at these enhancers before opening further cooperated the notion where SOX9 recruited epigenetic co-factors to prime closed chromatin prior to enhancer opening and activation.

Studies on pioneer factors during fate switching progresses often deduce the event of pioneer factors' binding to closed chromatin through observing a set of prior closed chromatin become open and bound by the pioneer factor of interest (Balsalobre and Drouin, 2022). While this observation is strong evidence of pioneering function, it is not direct and only captures the initial and end stage of fate switching. Nevertheless, one study carefully documented PAX7's binding and melanotrope enhancers' opening dynamics in minute-resolution with an inducible system *in*

vitro (Mayran *et al.*, 2018). Only with this high temporal resolution, Mayran et al. were able to observe a state where PAX7 is directly associated with closed chromatin. Alternatively, the stepwise enhancer activation can be achieved through stepwise *in vitro* differentiation. In studying the role of a well-known pioneer factor FOXA2, Lee et al. induced the differentiation of human pluripotent stem cells (hPSCs) into comparatively uniform populations of definitive endoderm (DE), posterior foregut (FG), and primary pancreatic progenitor (PP1) step by step (Lee *et al.*, 2019). With this system, they demonstrated that at both FG and PP1 stages, the PP1-specific enhancers exhibited H3K4me1 signals. However, ATAC-seq and H3K27ac signals became evident only at the PP1 stage, indicating a sequential process of enhancer priming and subsequent activation. While an increasing number of studies have started to acknowledge the temporal aspect of pioneer factor directed fate switching *in vitro*, I was delighted to corroborate it in an *in vivo* setting for the first time. However, I could not definitively conclude if MLL3/4 binding to pioneer factor and H3K4me1 deposition were functionally required merely because they preceded enhancer opening and activation, especially when the role of MLL3/4 in activating and maintaining gene expression is still a question of ongoing debate (Henikoff and Shilatifard, 2011).

Regarding the specificity of fate activation

The transcriptomic and chromatin data suggested that SOX9 directly induced HF fate, yet it is curious why other cell fates, such as chondrocyte differentiation, associated with SOX9 were not prominently induced. Similar to the ectopic re-activation system, Fuglerud et al. expressed SOX9 in cultured human endothelial cells, and its induction alone is enough to trigger the activation of a mesenchymal fate rather than HF genes (Fuglerud *et al.*, 2022). This discrepancy could potentially be attributed to several factors.

Firstly, the functional specificity of SOX9 in different tissue contexts might be determined by its interaction with distinct sets of non-pioneer transcription factors, and their availability of them can vary in different cell types and developmental stages. Specifically, I observed that SOX9 recruits AP1 as a cooperative transcription factor, which was also observed to interact with another putative pioneer factor, STAT3, in reaction to inflammation in EpdSCs (Larsen et al., 2021). On the other hand, transcription factors such as SOX5 and SOX6 were identified to aid SOX9 in chondrocyte differentiation (Akiyama *et al.*, 2002). However, SOX5 was absent from EpdSCs per my RNA-seq data, and SOX6 expression was significantly downregulated by W1 in response to SOX9 induction.

Additionally, closed chromatin consists of different subtypes of heterochromatin whose modulation of cell fate should not be overlooked (Balsalobre and Drouin, 2022). Constitutive heterochromatin refers to extensive regions of densely packed chromatin with H3K9me3 marks. These regions are generally unaffected by most pioneer factors investigated, including Yamanaka factors (Mayran *et al.*, 2018; Soufi et al., 2012). In contrast, pioneer factors open chromatin sites mostly within facultative heterochromatin, which is defined by its relatively higher abundance of H3K9me2 (Mayran *et al.*, 2018). Therefore, further ChIP-seq experiments on H3K9me3 and H3K9me2 in EpdSCs are necessary to address why SOX9 only opened HF enhancers rather than enhancers of other associated fates.

Finally, in contexts where SOX9 induces different fates, the level of SOX9 expression might need to reach a certain threshold, suggesting a possible dose dependency of SOX9's actions. It is also known that the levels of pioneer factors like SOX9 are tightly regulated with additional enhancers, further emphasizing the importance of its expression level in dictating cell fate (Arnold et al., 2013; Naqvi *et al.*, 2023). Insufficient expression of *Sox9* due to haploinsufficiency has been

shown to contribute to impaired formation of cartilage and premature skeletal mineralization (Bi et al., 2001). Recently, Naqvi et al. more precisely showed how cranial neural crest genes can have sensitized or buffered responses to SOX9 dosage changes with a degrader titration system (Naqvi *et al.*, 2023). Besides SOX9, another study induced different levels of SOX2 with DOX and demonstrated level-dependent effects of SOX2 during caudal epiblast development (Blassberg et al., 2022). Besides the dose, time also can be in play as I demonstrated the effect of prolonged SOX9 expression at a physiologically relevant level in EpdSCs. Interestingly, rather than giving rise to fully differentiated hair follicles, EpdSCs with prolonged SOX9 expression formed placode-like and BCC-like lesions, consistent with the required downregulation of SOX9 and upregulation of other transcription factors during HF maturation. The sustained SOX9 instead induced RUNX transcription factors in EpdSCs, which further regulated the progression to a BCC fate (Fig. 6.1).

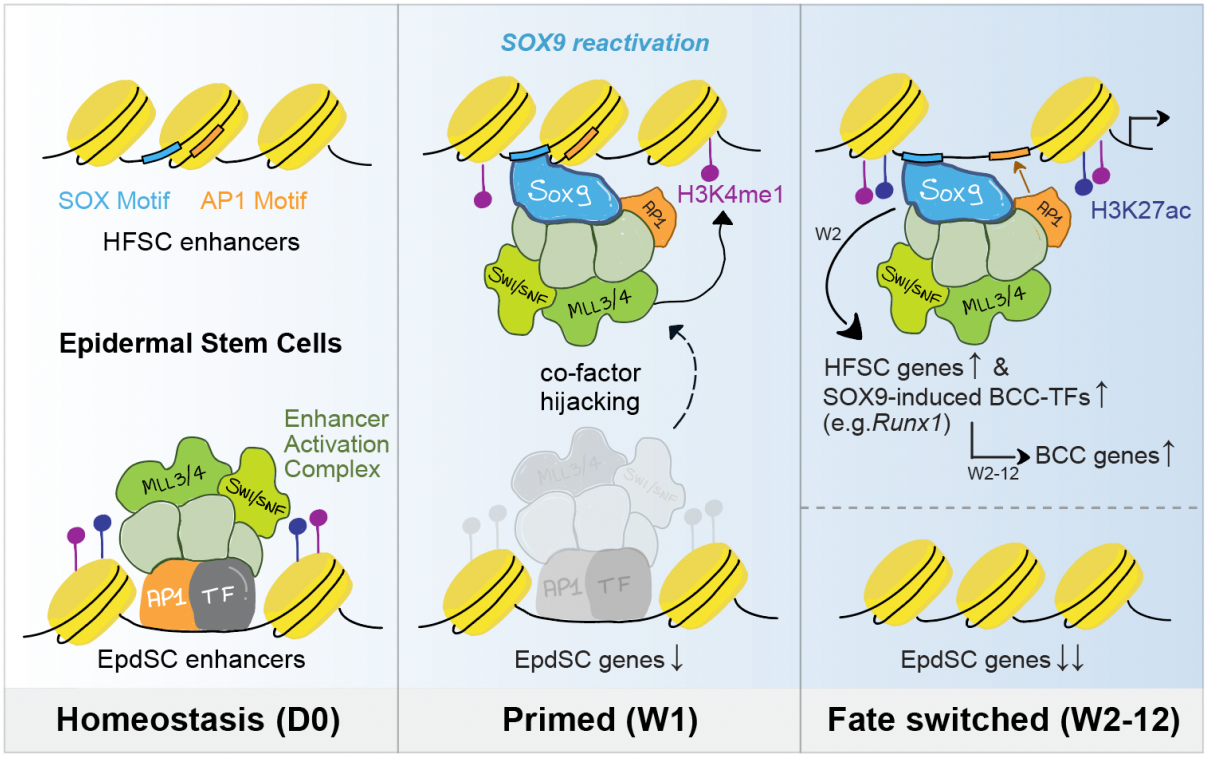

Fig. 6.1: Working model for how SOX9 achieves cell fate switching

Pioneer factor SOX9 competes for co-factors and other transcription factors to indirectly silence EpdSC enhancers, while directly activating HFSC enhancers. AP1 motifs appear in both EpdSC and HFSC enhancers, but SOX9 can competitively recruit AP1 away from EpdSC and to HFSC enhancers. SOX9's binding opens HFSC enhancers and exposes the AP1 sites (prevalent within 147 bp of SOX9 sites). Also among SOX9's target genes are transcription factors like RUNX1, whose footprints appear to participate in the delayed activation of BCC cancer genes, leading to further fate switching to BCC, downstream of the EpdSC:HFSC fate transition.

6.2 The hijack of epigenetic co-factors by pioneer factors

6.2.1 Significance

To complete a cell fate switch, a new gene regulatory network arises while an old one fades out. On the chromatin level, fate-switching manifests as one set of enhancers is silenced while another is activated. An increasing number of studies are concentrating on the role of pioneer factors in activating or poising a new cellular fate after opening enhancers, depending on the activating or repressive nature of the recruited co-factors. However, since an efficient fate transition also requires the closing of enhancers related to the old fate, it has remained largely speculated if and how a single pioneer factor can simultaneously silence the old fate while it activates the new fate.

6.2.2 Findings and open questions

In the case of SOX9, if it were to directly silence the epidermal fate, I would observe SOX9 binding at the closing enhancers. While not being absolutely zero, only 3% of the closing epidermal enhancers contained binding sites for SOX9 (Fig. 5.1a). Furthermore, the closure of epidermal enhancers was independent of having GATA motifs and hence could not be attributed to the reduction of *Gata3* expression (Fig. 5.2e). Intriguingly, even the ΔHMG-SOX9 that was unable to bind DNA at its target sites was still able to effectively silence these EpdSC enhancers (Fig. 5.5a,b). These findings suggested that the repression mechanism involved an indirect pathway that relied on the presence of SOX9 rather than its direct binding to chromatin.

Alternatively, SOX9 could induce the expression of downstream repressors that directly repressed epidermal enhancers. Nevertheless, the transcriptional kinetics revealed that the repression of epidermal genes occurred before the activation of HF genes and any possible

repressors (Fig. 2.5). Inspired by the interactome of SOX9 and the fact that the dynamics of MLL3/4 loss at epidermal enhancers coincided with the fast repression of epidermal genes at W1, I tested the possibility that SOX9 silenced epidermal genes by hijacking co-factors to HFSC gene enhancers.

I solidified the competition hypothesis by showing that ΔHMG-SOX9, although lacking its DNA binding domain, can still recruit MLL3/4 away from epidermal enhancers. In addition, AP1 factors were also competed away by SOX9 from epidermal enhancers resulting in their closure, and this repression mechanism could be recapitulated by specifically preventing AP1 binding by overexpressing AFOS, a dominant negative AP1 factor, alone (Fig. 5.5d). Curiously, when SOX9 was induced in the presence of AFOS, SOX9's ability to open closed chromatin was inhibited but not entirely abolished (Fig. 5.5d), indicating that AP1 is required for maintaining HF enhancer opening with SOX9 rather than for the initial pioneering in closed chromatin by SOX9. Similarly, when AFOS was induced at the onset of epidermal inflammation, it partially reduced enhancers opening at inflammatory genes (Larsen *et al.*, 2021).

I further extended the competition model to SWI/SNF complex whose core components were found in the SOX9 interactome. While it was not surprising to find SWI/SNF interacting with a pioneer factor, as documented before, it was reassuring to find SWI/SNF being recruited away from closing enhancers related to epidermal fate (Balsalobre and Drouin, 2022). Compared to other chromatin remodelers such as Iswi and Ino80, SWI/SNF preferably binds at cell-fate-related enhancers, and targeted degradation of SWI/SNF causes rapid closure of developmental enhancers (Hendy et al., 2022; Iurlaro et al., 2021; Park et al., 2021; Schick et al., 2021). Since I did not find chromatin remodelers other than SWI/SNF in SOX9 interactome (Table 2), and most dynamic peaks were enriched at enhancers (Fig. 3.1c,d), these results started to ascertain a common

pioneering mechanism at play where a pioneer factor plays a dual function in cell identify enhancers' direct activation and indirect repression rather than at promoters or housekeeping genes (Dejosez et al., 2023).

The biggest assumption behind the competition model was that co-factors are limiting so that a new cell fate can be activated at the expense of an old fate. I verified this assumption by supplying additional ARID1a, a SOX9 interacting SWI/SNF component, and observed that the closing of epidermal enhancers by SOX9 was attenuated. The concept of limiting co-factors is important when considering cellular plasticity. When multiple lineage-specific transcription factors are present at the junction of cell fate decision, limiting interacting co-factors can ensure efficient and complete fate switching, especially for differentiation of stem cells (Fig. 6.2). On the other hand, when co-factor levels are dysregulated, unwanted lineage genes may become activated. This notion also aligns with the fact that copy number variation of epigenetic co-factors is often associated with cancer progression where lineage infidelity is often observed (Kadoch et al., 2013; Sima et al., 2019).

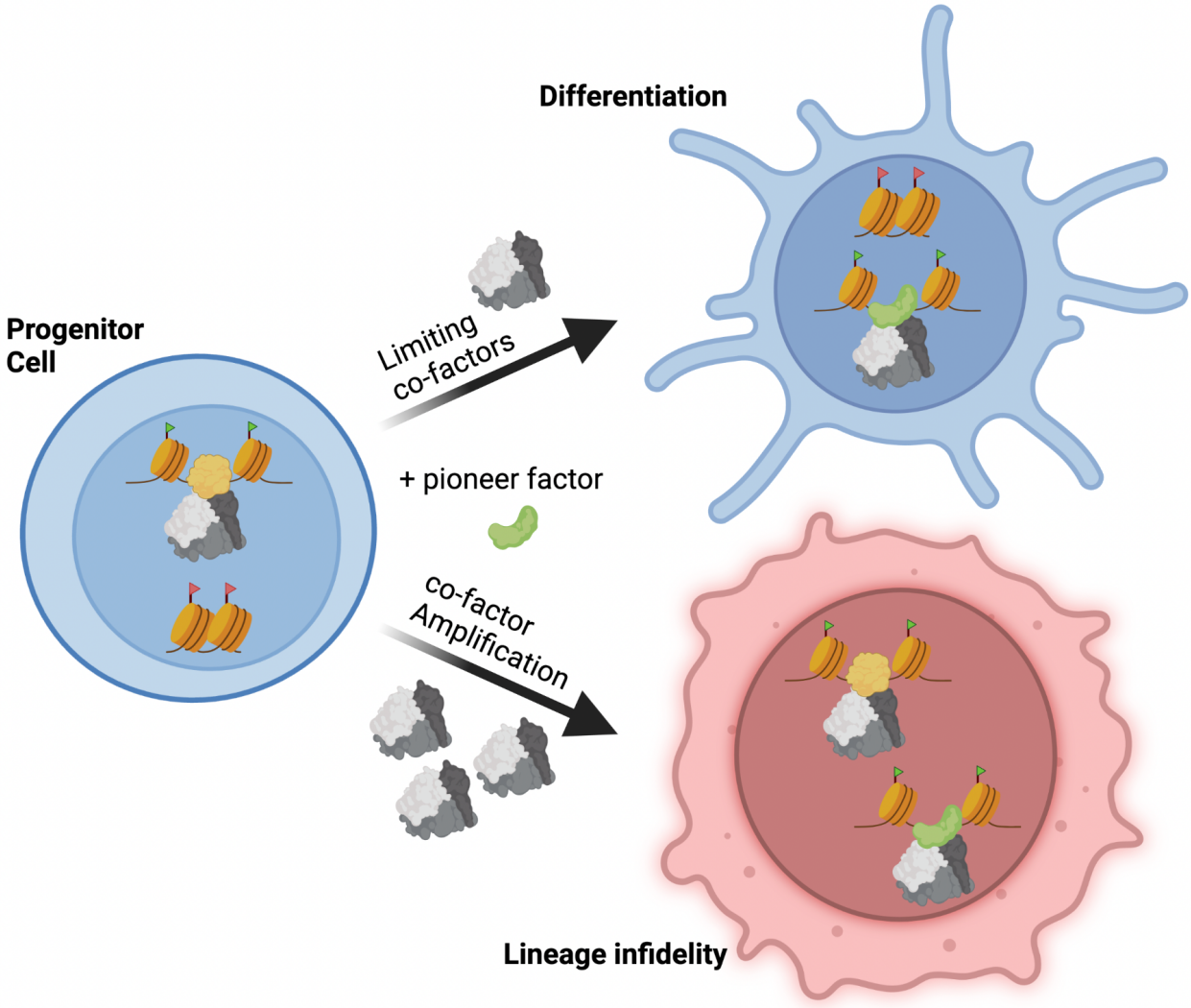

Fig. 6.2: Limiting co-factor is essential in avoiding cell lineage infidelity
When pioneer factor (green bean-shaped protein) is induced in normal progenitor cells, it competes for limiting co-factors (grey protein complex) to close progenitor enhancers and open lineage enhancers. However, if the co-factors are in excess, new enhancers are opened while the progenitor enhancers fail to be silenced, resulting lineage infidelity.

6.3 Future Directions

6.3.1 Other pioneering mechanisms

With SOX9-mediated epidermal to HF fate switching as an *in vivo* model, I thus far illustrated how SOX9 hijacks co-factors to silence epidermal fate, prime HF enhancers with H3K4me1 and recruit SWI/SNF and AP1 to open and maintain HF enhancers. Future experiments are necessary to address other potential mechanisms behind SOX9-directed fate switching.

In my ChromVAR and motif analysis, CTCF surfaced as a top candidate for additional regulation related to enhancer-promoter interactions. CTCF is known to shape the 3D chromatin organization of the genome through its interactions with the cohesin complex, which extrudes DNA to form chromatin loops (Tang et al., 2015). These loops can facilitate enhancer-promoter interactions to activate gene expression. While CTCF was not detected in SOX9 interactome, cohesin release factor (WAPL) and mediator (MED15) were within the list of SOX9-specific interacting proteins. Since CTCF motifs were enriched in dynamic ATAC peaks, additional experiments could illustrate whether loop formation and/or resolution were essential for fate switching.

Chromatin looping is also closely associated with phase separation and condensate formation, which can serve as a regulatory hub for transcription (Hnisz et al., 2017). Increasing studies have illustrated how TA domains of transcription factors can be intrinsically disordered regions (IDRs) that facilitate phase separation (Lyons et al., 2023). My experiments demonstrated SOX9's TA domain is crucial for its interaction with co-factors and pioneering function. More detailed studies on SOX9's TA domain could answer if it acts as IDR and if SOX9 involves in phase separation.

6.3.2 Potential application in therapeutics

Elevated SOX9 expression has been associated with the progression and poor prognoses of tumors in the prostate, central nervous system, pancreas, ovary and esophagus (Aguilar-Medina et al., 2019; Jo et al., 2014). Traditionally, transcription factors have been deemed "undruggable" because they often lack well-defined active sites, which are typically used as the target sites for small-molecule drugs. However, this paradigm is shifting with new technologies and strategies. One such strategy is targeting protein-protein interactions (PPIs) (Henley and Koehler, 2021). As observed, PPI through its TA domain is crucial for SOX9's function in both activating new fate and repressing old fate.

Another emerging approach is the use of Proteolysis-Targeting Chimeras (PROTACs). PROTACs are bifunctional molecules that bind to the target protein on one end and to an E3 ubiquitin ligase on the other, leading to the ubiquitination and subsequent degradation of the target protein (Henley and Koehler, 2021; Nalawansha and Crews, 2020). However, to effectively target SOX9 using these strategies, it is critical to understand the structural basis of SOX9-cofactor interactions. High-resolution structures of SOX9 bound to its co-factors would be highly valuable, as they could reveal potential druggable sites.

Single-cell technologies have been revolutionizing our understanding of the heterogeneity of biological systems. I have shown that H3K4me1 can be present at enhancers before the chromatin becomes accessible and before the initiation of RNA transcription. Single-cell multiomics such as Paired-Tag would be essential for us to confirm the order of pioneering events at a single-cell resolution (Vandereyken et al., 2023; Zhu et al., 2021). Furthermore, in the context of the epithelial-to-mesenchymal transition (EMT), a highly heterogenous process implicated in cancer metastasis, the early detection of changes in the enhancer landscape on a single-cell level prior to

accessibility and transcription could be particularly valuable as an early diagnostic marker for EMT (Gerstberger et al., 2023; Yang et al., 2022).

CHAPTER 7. Other Documentations

7.1 Materials and Methods

Generating and handling TRE-SOX9 mice

To generate the conditional SOX9 transgenic mice, the *Sox9* coding sequence was cloned into the pTRE2 vector harboring a doxycycline inducible, minimal CMV2 promoter. A MYC-tag was added to the N-terminus of SOX9. Transgenic mice were generated as described previously (Vasioukhin et al., 1999). The resulting *TRE-Sox9* mice were then genotyped and crossed to *Krt14-rtTA* transgenic mice to allow for doxycycline-inducible expression of mycSOX9 specifically in skin epithelium (Nguyen *et al.*, 2006). Animals were maintained and bred under specific-pathogen-free conditions at the Comparative Bioscience Center (CBC) at The Rockefeller University, which is an Association for Assessment and Accreditation of Laboratory Animal Care (AALAC) – accredited facility. All procedures were performed with the Institutional Animal Care and Use Committee (IACUC) – approved protocols.

Primary Cell Isolation

Primary *Krt14rtTA*;*TRE-Sox9* EpdSCs were isolated from newborn male pups (postnatal day 0, or P0) as described previously (Blanpain et al., 2004; Yang *et al.*, 2017). Briefly, mouse back skin was harvested from P0 pups and treated with dispase (Gibco) overnight at 4°C. Epidermis was manually separated from dermis and disassociated into a single cell suspension. Epidermal cells were passaged and maintained in E-low calcium media (0.05 mM CaCl2) at 37°C with 7.5% CO_2 (Rheinwald and Green, 1977).

Doxycycline Treatment

0.1mg of doxycycline (Sigma) in 100µl phosphate buffered saline (PBS) was administered by intraperitoneal injection (IP) to *Krt14rtTA*;*TRE-Sox9* and *Krt14rtTA*-only mice at postnatal day P21 and thereafter were maintained on mouse chow containing 2mg/g doxycycline throughout the experimental time course. Phenotypic mice were housed with at least one control littermate for adequate grooming. To maintain proper body fluid, 100µl PBS was administered through IP injection every other day after 4-weeks of SOX9 induction. For W12 samples, epidermis from the back skin of P0 *Krt14rtTA*;*TRE-Sox9*, or *Krt14rtTA*-only pups were grafted onto 6 to 8-week-old immunocompromised (*Nude*) female mice. Grafts were allowed to heal for 21 days, and doxycycline was administered as above. For induction of BioID, SOX9 and its variants, AFOS and ARID1a in cultured cells, doxycycline was added to a final concentration of 1µg/ml in E-low media for BioID or SOX9 variant experiments.

Immunofluorescence

Mouse back skin was fixed in 4%PFA at room temperature for 15min, and then washed 3 times with PBS for 15min at 4°C. Following PBS washes, samples were dehydrated in 30% sucrose in PBS 4°C overnight. The dehydrated samples were then embedded in OCT medium (VWR) and frozen on dry ice. Cryosections (16µm) were blocked in immunofluorescence (IF) buffer containing 0.3% Triton X-100, 2.5% normal donkey serum, 2.5% normal goat serum, 1% BSA, and 1% gelatin in PBS for 1hr at room temperature. After blocking, the sections were stained with primary antibodies in IF buffer at 4°C overnight: Myc-tag (Rabbit, 1:1000, Cell Signaling), SOX9 (Rabbit, 1:5000, Millipore), ITGA6 (Rat, 1:1000, BD), KRT14 (Chicken, 1:1000, Biolegend), KRT10 (Rabbit, 1:250, Fuchs Lab), EpCAM (Rabbit, 1:100, Abcam), KRT6 (Guinea pig, 1:1000,

Fuchs Lab), RUNX1 (Rabbit, 1:100, Abcam), and GATA3 (Rat, 1:100, Invitrogen). After primary antibody staining, all sections were washed for 3 times with IF buffer containing 0.1% Triton X-100 in PBS for 5min at room temperature. Sections were then stained with Alexa 488, 546 or 647 conjugated secondary antibodies (1:500, Thermo Fisher), mounted with Prolong Diamond anti-fade mounting media with DAPI (Thermo Fisher), and imaged with Zeiss Axio Observer Z1 with Apotome 2 microscope. Images were collected and analyzed with Fiji (ImageJ v.2.3.0). For the Human Atlas immunostaining, the following antibodies were used: SOX9 (CAB068240), EpCAM (CAB030012), KRT6A (HPA061168).

For cultured cells, cells were plated onto chamber slides (Thermo Fisher). At collection, cells were fixed with 4%PFA for 10min, and then washed 3 times with PBS at room temperature. After washing, the cells were blocked and stained with primary antibodies the same way as described above for sections with the following primary antibodies: HA-tag (Rabbit, 1:1000, Cell Signaling), GFP (Chicken, 1:2000, Fuchs Lab), RFP (Rat ,1:1000, ChromoTek), and MYC-tag (Rabbit, 1:1000, Cell Signaling).

For 5'-ethynyl-2'deoxyuridine (EdU) experiments, mice were injected IP with EdU (50μg/g body weight) 2 hours before analysis. Quantifications were performed by counting the number of EdU^+ EpdSCs within the basal layer. For quantifying the SOX9 signal in the native ORS and the SOX9-induced epidermis, sections were stained with same SOX9 antibody concentration (1:5000), and same laser intensity and exposure time were used to acquire images. From each sample, 100 cells were quantified with the multi-point tool in Fiji.

Flow Cytometry and Cell Sorting

Krt14rtTA;*TRE-Sox9* and *Krt14rtTA*-only male mice were used for FACS experiments to obtain maximal cell numbers and to control for variation due to sex. Briefly, the whole back skins were first dissected from the mouse. After scraping off the fat tissues from the dermal side, the tissues were incubated in 0.25% Trypsin/EDTA (Gibco) for 45 to 60 min at 37ºC. After quenching the Trypsin with cold FACS buffer (5%FBS, 10mM EDTA, 1mM HEPES in PBS), the epidermal layer and hair follicles were scraped off the epidermal side of the skin. The tissues were mechanically separated and filtered through a 70μm cell strainer (BD) into a single-cell suspension for immunolabeling. Single cell suspensions were immunolabeled with antibodies: Ly6A/E-APCCy7 (1:500, BioLegend), CD49f-PECy7 1:1000, BioLegend), CD34-Alexa660 (1:50 Invitrogen), CD45-biotin (1:200, BioLegend), CD31-biotin (1:200, BioLegend), CD140a-biotin (1:200, BioLegend), CD117-biotin (1:200, BioLegend), and Streptavidin-FITC (1:1000, BioLegend) in 300 μl of FACS buffer. Stained cells were washed and resuspended with FACS buffer with 100ng/ml DAPI prior to analysis or sorting. EpdSCs were collected using an Aria Cell Sorters (BD Biosciences) into either FACS buffer for genomic experiments or Trizol LS (Invitrogen) for RNA extraction.

RNA-Seq and raw file processing

EpdSCs were collected by FACS as described above directly into Trizol LS (Invitrogen). RNA libraries were generated using SMARTer RNA kit for low input RNA-seq. Libraries were sequenced on Illumina NovaSeq SP. Raw FASTQ files were trimmed of barcodes using Skewer (v.0.2.2) and transcript abundance quantified using Salmon (v.1.4.0) with a modified GENCODE transcript index (version GRCm38 release M24) to include *TRE-Sox9*. Gene level counts and

TPMs were calculated using the Tximport (v.1.12.3) package in R (v.3.6.1). For hair placode RNA-seq data, after generating the raw counts, differentially expressed gene list were generated with DESeq2 (v.1.16.1).

ATAC-Seq and raw file processing

ATAC-seq was performed on FACS-purified EpdSCs (2-4 male mice per replicate) at indicated time points (D0, W1, W2, W6 and W12) and cultured keratinocytes (Buenrostro *et al.*, 2013). Briefly, cells were lysed in ATAC lysis buffer for 5 minutes and then transposed with Tn5 transposase (Illumina) for 30 minutes. Samples were barcoded and sequencing libraries were prepared according to manufacturer's guidelines (Illumina) and sequenced on an Illumina NextSeq. For sequencing analysis, 50bp paired end FASTQs were aligned to the mouse genome (GRCm38/mm10) using the PEPATAC pipeline (Smith et al., 2021). Replicate BAM files were merged, and peak calling was performed using Model-based Analysis of ChIP-Seq 2 (MACS2) with the option of "--keep-dup all" to keep duplicates generated during the combining of experimental replicates. Because peak calling is greatly influenced by number of reads and sequencing depth, I normalized peak calling as performed as described with a threshold of 3, and I quantified reads in filtered peaks (RIP) for generating normalized bigwig files (Corces *et al.*, 2018). To do so, 1,000,000/RIP was used as input for Deeptools "bamcoverage" with the "--scaleFactor" option. Shared peaks were defined as regions that had ≥1 base pair overlap between two time points. Dynamic peaks were defined as those accessible chromatin regions that were absent from at least one time point. For PCA analysis, peaks called from combined replicates were merged to create a union set of peaks across the samples. Read counts under the union peaks were

summed for each individual replicate and used as input for PCA analysis or generating K-means clusters in R.

Cut-and-Run and raw file processing

EpdSCs were FACS-purified, and the Cut-and-Run sequencing was performed as previously described with minor modifications indicated below (Larsen *et al.*, 2021; Skene and Henikoff, 2017). Briefly, 500,000 to 1,000,000 EpdSCs were washed with ice cold PBS, resuspended in crosslinking buffer (10mM HEPES-NaOH pH 7.5, 100mM NaCl, 1mM EGTA, 1mM EDTA, 1% formaldehyde) and rotated at room temperature for 10 min. Crosslinked cells were quenched with glycine at a final concentration of 0.125M for 5 min at room temperature. Cells were washed with cold 1X PBS and resuspended in NE1 buffer (20mM HEPES-KOH pH7.9, 10mM KCl, 1mM MgCl2, 1 mM DTT, 0.1% triton X-100 supplemented with Roche complete protease inhibitor EDTA-free) and rotated for 10 min at 4^0C. Nuclei were washed twice with Cut-and-Run wash buffer (20mM HEPES pH7.5, 150mM NaCl, 0.5% BSA, 0.5mM spermidine supplemented with protease inhibitor) and incubated with concanavalin-A (ConA) beads washed with Cut-and-Run binding buffer (20mM HEPES-KOH pH 7.9, 10mM KCl, 1mM CaCl2, 1mM MnCl2) for 10 min at 4^0C. ConA-bead-bound nuclei were incubated overnight at 4^{o}C in Cut-and-Run antibody buffer (Cut-and-Run wash buffer supplemented with 0.1% Triton X-100 and 2mM EDTA) and antibody. After antibody incubation, ConA-bead-bound nuclei were washed once with Cut-and-Run Triton wash buffer (CUT&RUN wash buffer supplemented with 0.1% triton X-100) then resuspended and incubated at 4^0C for 1hr in CUT&RUN antibody buffer and 2.5µL pAG-MNase (EpiCypher). ConA-bound-nuclei were then washed twice with CUT&RUN triton wash buffer and resuspended in 100µL of triton wash buffer and incubated on ice for 5 min. Then, add 2µl 100mM CaCl2 and

mix gently to each 100ul ConA-bound-nuclei. The reaction was then incubated at 0ºC for 30 min. The reaction was stopped by addition of 100µl 2x stop buffer (340mM NaCl, 20mM EDTA, 4mM EGTA, 0.1% Triton X-100, 50µg/ml RNaseA) and incubated at 37ºC for 10 min. All buffers mentioned above were filtered with 0.22 µm filter before use. After incubation, ConA-bound-nuclei were captured using a magnet and supernatant containing Cut-and-Run DNA fragments were collected. Supernatant was incubated at 70ºC for 4hrs with 2µL 10% SDS and 2.5µL 20mg/mL proteinase-K. DNA was purified using PCI reagent (Phenol:Chloroform:Isoamyl Alcohol, Millipore) and overnight ethanol precipitation with glycogen at -20ºC. DNA was resuspended in elution buffer (1mM Tris-HCl pH8.0, 0.1mM EDTA). A full list of antibodies used is provided in Table 4.

Cut-and-Run sequencing libraries were generated using NEBNext Ultra II DNA Library Prep Kit for Illumina and NEBNext Multiplex Oligos for Illumina. PCR-amplified libraries were purified using 1X ratio of SPRI beads (Beckman) and eluted in 15µL EB buffer (Qiagen). All Cut-and-Run libraries were sequenced on Illumina NextSeq using 40-bp paired-end reads. Reads were trimmed with Skewer and aligned to reference genome (mm10) using Bowtie2 (v.2.2.9) and deduplicated with Java (v.2.3.0) Picard tools (http://broadinstitute.github.io/picard). Reads were filtered to ≤120bp using Samtools (v.1.3.1). BAM files for each replicate were combined using Samtools. Bigwig files were generated using Deeptools (v.3.1.2) with RPKM normalization and presented with Integrative Genomics Viewer (IGV) software. Cut-and-Run peaks were called using SEACR from bedGraph files generated from RPKM normalized Bigwig files (bigWigToBedGraph, UCSC Tools) using stringent setting and a numeric threshold of 0.01 (Meers et al., 2019c). Peaks were further filtered to have peaks scores > 1800 for a set of high confident peaks.

MINT-ChIP-Seq and raw file processing

EpdSCs were FACS-purified and subjected to histone ChIP-seq (MINT-ChIP) with antibodies recognizing H3K4me1 (Rabbit, Cell Signaling), H3K27ac (Rabbit, Active Motif), and Total H3 (Mouse, Active Motif). Pooled samples were then sequenced using 50-bp paired-end Illumina NextSeq. Resulting FASTQ files were demultiplexed for specific histone antibodies by using the unique barcode present in sequenced read2. Resulting paired reads were then trimmed for adapters using Skewer and aligned to mouse genome (GRCm38/mm10) using Bowtie2. Duplicated reads were marked and removed using Picard and replicates were merged with Samtools. Peak calling for H3K27ac was performed using MACS2 while broad domains of H3K4me1 were called using epic2 (Stovner and Saetrom, 2019). Samples were independently normalized to the number of reads in peaks (RIP). For visualization, Bigwig files were generated on the combined BAM files using Deeptools "bamcoverage" with (1,000,000/RIP) as input for the "--scalefactor" option. For total H3, RPKM was used for normalization. A full list of antibodies used is provided in Table 4.

BioID and Mass Spectrometry

For identification of SOX9-interacting partners Nicholas transduced primary *Krt14rtTA* EpdSCs with LV-TRE-Myc-BioID2-GFP-NLS-H2B-RFP or LV-TRE-Myc-BioID2-SOX9-H2B-RFP. RFP^+ transduced cells were then isolated using FACS and stable EpdSC lines were established. Nicholas induced expression of recombinant proteins using 1 μg/mL doxycycline. Cells were allowed to expand for 5 days and were pulsed with 50 μM biotin (Sigma) for 16 hours before reaching confluence. Cells were purified and proteins isolated as previously described with minor modifications mentioned below (Kim and Roux, 2016). Immediately after sonication lysates were washed using Zeba desalting columns (7K MWCO, ThermoFisher cat# 89894) with 50mM

Tris pH 7.4 to remove excess biotin. Beads were also washed 3X with 2M urea and a final 2X with PBS before being resuspended with 500µL 50mM Tris, pH 8.0. All washes were performed using a magnetic stand. New tubes were used in between each urea and PBS washes. Wash buffer was removed from suspension of magnetic beads and replaced with 100 µl 8M Urea, 50 mM ammonium bicarbonate, 10 mM dithiothreitol for 1 hr and replaced with 100 µl 40 mM iodoacetamide and incubated in the dark for 30 min. Alkylation solution was replaced with 1 µg trypsin (Promega) dissolved in 100 µl 50 mM ammonium bicarbonate and incubated for 4 hrs. Supernatant was then removed and re-digested overnight using 0.5 µg trypsin and 0.5 µg Endopeptidase Lys-C (Wako). Peptides were desalted and concentrated using C18 based Stage tips and separated by nanoLC (gradient: 2%B/98%A to 38%B/62%A in 70 min, A: 0.1% Formic Acid, B: 90% acetonitrile/0.1% Formic acid) coupled to a Fusion Lumos (Thermo Scientific) operated in high/high mode (Rappsilber et al., 2007). Data were queried with UniProts Complete Proteome mouse database and concatenated with known common contaminants. Proteome Discover and Mascot was used to analyze the resulting data produced. Data was further filtered using a percolator to calculated peptide False Discovery Rates and set a threshold of 1%. Proteins were specific to SOX9's proximity if they were identified in 2 of the three Myc-BioID2-SOX9 replicates and absent of all the Myc-BioID2-GFP-NLS samples (Kall et al., 2007). See Supplementary Table 2 for full list of SOX9-specific interactors and raw counts.

Generation of EpdSC lines expressing SOX9 and variants, AFOS or ARID1a.

Three versions of MYC-tagged SOX9 (WT, ΔTA and ΔHMG as indicated in Fig. 5.3a) were cloned into plKO vectors with a TRE promoter and a puromycin-resistance gene (*puroR*) under the control of a constitutive promoter (PGK). Three lentiviruses were produced as described

(Beronja et al., 2010). *Krt14rtTA* EpdSCs were cultured and transduced with 1µl concentrated lentivirus in 10ml E-low media with 8µg/ml polybrene (hexadimethrine bromide, Sigma 107689-100MG) overnight. Transduced cells were then selected with 2µg/ml puromycin for 5 days before doxycycline treatment. For AFOS and ARID1a experiments, FLAG-tagged *AFOS* or *Arid1a* CDS were cloned into the described plKO vector for lentiviral production. *Krt14rtTA* or *Krt14rtTA;TRE-mycSOX9* EpdSCs were cultured and transduced with 1µl concentrated lentivirus as described above. Transduced cells were also selected with puromycin for 5 days before doxycycline treatment.

CRISPR mediated Mll4 Knockout

For generating the *Mll4* null lines, I cultured keratinocytes extracted from *Krt14-rtTA;TRE-Sox9* mice. Lines were generated with the Alt-R CRISPR-Cas9 system (IDT). Briefly, a recombinant Cas9 protein, a validated sgRNA (TGCTCGGCAACAGACGTGAC) targeting *Kmt2d* (*Mll4*) gene or a negative control sgRNA (IDT), and an ATTO-550 conjugated tracer RNA were used to form a ribonucleoprotein (RNP) are mixed with RNAiMax reagent (Thermo Fisher). Then, keratinocytes were transfected with the mixture overnight, and FACS purified into 96-well plates to produce clonal cell lines. The knockout cell lines were validated through sequencing of the target region for indel efficiency via MiSeq and used for the immunoblot of MLL4.

Immunoblotting and Co-Immunoprecipitation

Cultured EpdSCs were washed on the plate in cold 1X PBS, lysed in RIPA Buffer (Millipore) supplemented with protease and phosphatase inhibitors (Roche), and collected by scraping. Cells were lysed for 15 minutes on ice and then centrifuged to collect the supernatant. Co-

Immunoprecipitation was performed as previously described with the modification where protein-A/G-conjugated magnetic beads (Pierce) were used to bind antibodies instead, and proteins are eluted from beads with 1X NuPAGE LDS Sample Buffer (Invitrogen) with 2.5% 2-Mercaptoethanol at 70ºC for 10min (Guzzi et al., 2022). Protein concentration was determined by BCA Assay (Pierce) against a bovine serum albumin standard curve. 15µg protein of each sample was run on NuPAGE 4-12% Bis-Tris Gels (Invitrogen) for 2 hours at 110V in NuPAGE MOPS SDS Running Buffer (Invitrogen). Protein was transferred onto nitrocellulose membrane (Cytiva) in NuPAGE Transfer Buffer (Invitrogen) at 15V overnight at 4ºC. Given the marked differences in expected sizes of some of the proteins, overlapping host species of the antibodies raised, and the paucity of primary cell lysates for immunoprecipitates, the blots were cut based on sizes and performed immunoblotting on each piece with different antibodies. Membranes were then treated with blocking buffer with 5% Non-fat dry milk and 0.1% Tween-20 in TBS for 1 hour at room temperature before incubating with primary antibodies. The following primary antibodies were diluted in blocking buffer: MYC-tag (Mouse, 1:1000, Cell Signaling), MLL4 (Mouse, 1:200, Santa Cruz Biotechnology), cJUN (Rabbit, 1:1000, Cell Signaling), ARID1a (Rabbit, 1:1000, Abcam) and β-Actin (Mouse, 1:10000, Cell Signaling). The membranes were incubated in primary antibodies overnight at 4ºC. Membranes were then washed 3 times in 0.1% Tween-20 in TBS before incubating with HRP secondary (1:10000) antibody for 1 hour at room temperature. After secondary antibody incubation, membranes were then washed 4 times in 0.1% Tween-20 in TBS and incubated in ECL Prime reagents (Cytiva) for 5min before chemiluminescence detection. Membranes were imaged with an GE Amsham AI600 Imager. A full list of antibodies used is provided in Table 4.

Quantitative PCR

Equal amounts of RNA extracted from cultured cells were collected with AllPrep DNA/RNA Kits (Qiagen) and reverse transcribed using the superscript VILO cDNA synthesis kit (Invitrogen). For quantitative PCR, biological replicates represent the average of three technical replicates per individual sample. cDNAs from each sample were normalized using primers against *Rps16*. All primers used are provided in Supplementary Table 4.

Gene Set Enrichment Analysis (GSEA)

For comparing to both hair placodes and BCC, TPM matrices in D0, W2 and W12 were used as GSEA (v. 4.1.0) input. The differentially expressed gene (DEG) lists as illustrated in Fig.1D were used as gene set inputs. For the BCC and SCC samples, DEG list of genes with $p < 0.05$ was curated from GSE152487 in the GEO depository (Fiore *et al.*, 2020). GSEA was run with default settings, without collapsing, and with the gene set as the permutation type. The leading-edge analysis function was used to determine the significance of gene set enrichment.

Heatmaps and Boxplots

All heatmaps showing sequencing signals over binding sites are generated with Deeptools from RIP or RPKM normalized bigwig files. Profileplyr (v. 1.4.3) was used to generate ATAC, H3K4me1, H3K27ac, MLL3/4 CNR boxplots in R with matrix output from Deeptools compute-matrix as input. The histone H3 profile plot was also generated with Profileplyr in R.

Gene Ontology analysis

I performed GO analysis of each ATAC-Seq cluster by associating each region with genes and performing enrichment analysis using Genomic Regions Enrichment of Annotation Tool

(GREAT, version 3) with default gene association settings and the whole mouse genome (GRCm38/mm10) as the background (McLean *et al.*, 2010).

TF Motif and Footprint Analyses

For motif enrichment analysis on peak sets, HOMER (v. 4.10) findMotifGenome.pl was used with a customized motif database from JASPAR2018 (Heinz et al., 2010; Khan et al., 2018). The motif input for HOMER was generated from the 79 clusters of JASPAR2018 vertebrates CORE central TF motifs using 80% of the maximum log-odds expectation for each motif as the detection threshold for HOMER. To identify cluster specific motif enrichment in the ATAC-seq clusters. HOMER was used for each cluster using the union set of dynamic peaks as background (-bg) set with the options --size given –h. The resulting heatmap was generated by combining the significant (pval<0.05) motifs for each cluster and plotting the associated p value. For motif distance measuring, I overlapped SOX9 bound opening peaks with known AP1 and SOX motifs curated by HOMER (mm10-191020) and measured the distance from SOX motifs to the closest AP1 motifs with Bedtools. For footprint analysis, I used HINT-ATAC with the 79 motif clusters as the input as well. For TF motif variability score analysis, I ran ChromVAR (1.18.0) on the dynamic peaks for differential chromatin accessibility across 79 motif clusters to find the top variable motifs in dynamic peaks (Li et al., 2019). I further used ChromVAR to calculate the motif deviation scores over time at the top variable motifs.

Statistics

No statistical methods were used to pre-determine sample sizes, but my sample sizes are similar to those reported in previous publications (Larsen *et al.*, 2021; Larsimont *et al.*, 2015; Yang *et al.*, 2017). No data points were excluded. Upon harvest, mice with the same genetic background

were randomly allocated to genomic or immunofluorescence experiments. Data collection and analysis were not performed blind to the conditions of the experiments as the mice appears phenotypical after SOX9 induction. All immunofluorescence experiments were repeated three times with samples collected from different mice. All co-IP and immunoblot experiments were repeated twice with samples collected on different days. The statistics were analyzed with two-tailed t-test on the Graphpad Prism (9.0). Data distribution was assumed to be normal, but this was not formally tested. All the error bars are mean ± SD. Degree of statistical significance is denoted by ns ($p>=0.05$), * ($p<0.05$), ** ($p<0.01$), *** ($p<0.001$), and **** ($p<0.0001$).

Data Deposition

All sequencing data that support the findings have been deposited in the Gene Expression Omnibus under accession code GSE208072. Sequencing depths of samples are reported in Table 5 as quality control. Published RNA-seq data from BCC, SCC and normal EpdSCs that were re-analysed here are available under accession code GSE152487.

7.2 Tables

Table. 1 Selected genes and expression levels in SOX9-indeuced EpdSCs (z-score) over time

Gene ID	Symbol	D0	W1	W2	W6	W12	class
ENSMUSG00000024922.2	Ovol1	1.16	-0.37	-1.42	-0.08	0.71	D0
ENSMUSG00000014773.13	Dll1	1.67	-0.19	0.04	-0.75	-0.77	D0
ENSMUSG00000022510.14	Trp63	1.68	-0.24	0.02	-0.55	-0.9	D0
ENSMUSG00000052684.4	Jun	1.71	-0.41	0.02	-0.54	-0.78	D0
ENSMUSG00000023067.14	Cdkn1a	1.74	-0.08	-0.71	-0.51	-0.44	D0
ENSMUSG00000019880.10	Rspo3	1.77	-0.53	-0.53	-0.52	-0.19	D0
ENSMUSG00000000125.5	Wnt3	1.77	-0.28	-0.54	-0.59	-0.36	D0
ENSMUSG00000022528.8	Hes1	1.78	-0.46	-0.27	-0.55	-0.5	D0
ENSMUSG00000021250.13	Fos	1.78	-0.54	-0.29	-0.4	-0.55	D0
ENSMUSG00000026628.13	Atf3	1.78	-0.5	-0.42	-0.55	-0.3	D0
ENSMUSG00000002799.6	Jag2	0.9	0.75	0.48	-0.86	-1.28	D0
ENSMUSG00000001419.17	Mef2d	1.21	0.53	0.28	-0.76	-1.26	D0
ENSMUSG00000034910.5	Pygo1	1.27	0.45	0.29	-0.81	-1.2	D0
ENSMUSG00000024985.20	Tcf7l2	-0.76	0.77	0.89	0.45	-1.36	W1to6
ENSMUSG00000008575.17	Nfib	-0.38	0.41	1.28	0.11	-1.42	W1to6
ENSMUSG00000001911.16	Nfix	0.09	0.51	1.3	-0.6	-1.3	W1to6
ENSMUSG00000048402.14	Gli2	-0.94	0.99	1.14	-0.34	-0.85	W1to6
ENSMUSG00000022479.15	Vdr	-0.73	0.68	1.38	-0.33	-1	W1to6
ENSMUSG00000050199.13	Lgr4	-0.69	0.69	1.3	-0.13	-1.16	W1to6
ENSMUSG00000033016.16	Nfatc1	-0.7	0.87	1.15	-0.13	-1.19	W1to6
ENSMUSG00000050295.4	Foxc1	-1.21	-0.06	1	1	-0.72	W1to6
ENSMUSG00000039153.17	Runx2	-0.94	-1.03	0.35	1.36	0.27	W2to6
ENSMUSG00000022952.17	Runx1	-1.29	-0.22	1.33	0.59	-0.41	W2to6
ENSMUSG00000020140.15	Lgr5	-0.5	-0.83	-0.48	1.68	0.13	W2to6
ENSMUSG00000035557.9	Krt17	-0.95	-0.24	-0.46	1.68	-0.02	W2to6
ENSMUSG00000070691.10	Runx3	-0.84	-0.54	-0.07	1.71	-0.26	W2to6
ENSMUSG00000021994.15	Wnt5a	0.77	-0.79	-0.92	-0.4	1.34	W12
ENSMUSG00000028364.15	Tnc	-0.79	-0.81	-0.41	0.5	1.52	W12
ENSMUSG00000048776.8	Pthlh	-1.27	-0.39	0.13	0.03	1.49	W12
ENSMUSG00000024232.2	Bambi	0.03	-0.98	-0.29	-0.44	1.67	W12
ENSMUSG00000022150.16	Dab2	-0.62	-0.59	-0.43	-0.11	1.75	W12
ENSMUSG00000008398.15	Elk3	-0.74	-0.61	-0.55	0.25	1.65	W12
ENSMUSG00000022661.14	Cd200	-0.8	-0.72	-0.3	0.16	1.66	W12
ENSMUSG00000022816.11	Fstl1	-0.51	-0.56	-0.62	-0.06	1.74	W12
ENSMUSG00000053797.10	Krt16	-0.46	-0.55	-0.62	-0.13	1.76	W12
ENSMUSG00000016494.9	Cd34	-0.54	-0.55	-0.54	-0.13	1.76	W12
ENSMUSG00000006587.6	Snai3	-0.61	-0.58	-0.51	-0.04	1.74	W12
ENSMUSG00000043251.6	Exoc3l	-0.82	-0.27	-0.58	-0.05	1.71	W12
ENSMUSG00000024238.15	Zeb1	-0.51	-0.51	-0.51	-0.25	1.78	W12
ENSMUSG00000046470.5	Sox18	-0.45	-0.47	-0.47	-0.39	1.79	W12
ENSMUSG00000026728.9	Vim	-0.48	-0.43	-0.45	-0.43	1.79	W12

Table. 2 SOX9-specific interacting proteins

Gene	GFPNLS_1	GFPNLS_2	GFPNLS_3	SOX9_1	SOX9_2	SOX9_3	SOX9_mean
Vim	6.06E+08	0.00E+00	0.00E+00	2.05E+08	2.30E+08	0.00E+00	1.45E+08
Hnrnpa3	0.00E+00	7.09E+06	0.00E+00	9.36E+06	9.55E+06	1.38E+07	1.09E+07
Arid1a	0.00E+00	0.00E+00	0.00E+00	9.82E+08	7.04E+08	8.39E+08	8.42E+08
Gm382	0.00E+00	0.00E+00	0.00E+00	5.14E+07	4.48E+07	4.98E+07	4.87E+07
Wapl	0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.51E+06	1.90E+06	1.47E+06
Tcf20	0.00E+00	0.00E+00	0.00E+00	8.68E+06	7.61E+06	7.65E+06	7.98E+06
Foxp1	0.00E+00	0.00E+00	0.00E+00	2.13E+07	1.50E+07	1.62E+07	1.75E+07
Bicd2	0.00E+00	0.00E+00	0.00E+00	5.96E+06	1.68E+07	9.01E+06	1.06E+07
Ncoa2	0.00E+00	0.00E+00	0.00E+00	3.25E+07	2.61E+07	2.51E+07	2.79E+07
Ep300	0.00E+00	0.00E+00	0.00E+00	1.79E+07	5.27E+07	4.25E+07	3.77E+07
Prr12	0.00E+00	0.00E+00	0.00E+00	7.73E+06	5.13E+06	0.00E+00	4.29E+06
Irf2bp2	0.00E+00	0.00E+00	0.00E+00	7.12E+06	4.51E+06	4.73E+06	5.45E+06
Ncor1	0.00E+00	0.00E+00	0.00E+00	1.81E+07	1.39E+07	1.78E+07	1.66E+07
Elmsan1	0.00E+00	0.00E+00	0.00E+00	5.30E+06	9.19E+06	1.02E+07	8.22E+06
Arid1b	0.00E+00	0.00E+00	0.00E+00	1.69E+08	1.20E+08	1.53E+08	1.47E+08
Sf1	0.00E+00	0.00E+00	0.00E+00	3.96E+06	8.63E+06	1.95E+07	1.07E+07
Med15	0.00E+00	0.00E+00	0.00E+00	6.44E+06	9.26E+06	9.90E+06	8.53E+06
Cux1	0.00E+00	0.00E+00	0.00E+00	1.26E+07	9.41E+06	0.00E+00	7.33E+06
Grhl1	0.00E+00	0.00E+00	0.00E+00	9.50E+06	5.23E+06	1.23E+07	9.01E+06
Ubn2	1.93E+06	0.00E+00	0.00E+00	6.31E+06	7.56E+06	1.42E+07	9.37E+06
Phf21a	0.00E+00	0.00E+00	0.00E+00	2.15E+07	1.50E+07	1.41E+07	1.69E+07
Ncor2	0.00E+00	0.00E+00	0.00E+00	2.88E+07	3.31E+07	2.97E+07	3.06E+07
Pias2	0.00E+00	0.00E+00	0.00E+00	6.85E+07	4.27E+07	3.53E+07	4.88E+07
Kmt2c	0.00E+00	0.00E+00	0.00E+00	1.16E+07	1.49E+07	1.79E+07	1.48E+07
Kmt2d	0.00E+00	0.00E+00	0.00E+00	1.13E+07	6.06E+06	7.70E+06	8.35E+06
Purb	0.00E+00	0.00E+00	0.00E+00	6.22E+07	5.90E+07	6.74E+07	6.29E+07
Pias1	0.00E+00	0.00E+00	0.00E+00	6.17E+07	1.02E+08	8.38E+07	8.25E+07
Junb	0.00E+00	0.00E+00	0.00E+00	4.36E+07	5.74E+07	6.82E+07	5.64E+07
Msn	0.00E+00	0.00E+00	0.00E+00	5.69E+06	1.11E+06	0.00E+00	2.27E+06
Fosl2	0.00E+00	0.00E+00	0.00E+00	3.36E+07	3.53E+07	4.17E+07	3.69E+07
Kmt2a	0.00E+00	0.00E+00	0.00E+00	1.19E+07	6.85E+06	1.40E+07	1.09E+07
Ruvbl1	0.00E+00	0.00E+00	0.00E+00	0.00E+00	8.81E+05	3.04E+06	1.31E+06
Snrpd3	0.00E+00	0.00E+00	0.00E+00	3.36E+07	2.17E+07	2.37E+07	2.63E+07
Ncoa1	0.00E+00	0.00E+00	0.00E+00	3.89E+06	2.84E+06	4.56E+06	3.76E+06
Sox9	0.00E+00	0.00E+00	0.00E+00	4.39E+08	4.58E+08	5.17E+08	4.71E+08
Rreb1	0.00E+00	0.00E+00	0.00E+00	5.56E+06	6.25E+06	0.00E+00	3.94E+06
Hivep2	0.00E+00	0.00E+00	0.00E+00	7.94E+06	9.35E+06	6.05E+06	7.78E+06
Tet2	0.00E+00	0.00E+00	0.00E+00	2.20E+07	2.71E+07	2.10E+07	2.34E+07
Gpkow	0.00E+00	0.00E+00	0.00E+00	9.72E+06	0.00E+00	9.58E+06	6.44E+06
Coil	0.00E+00	0.00E+00	7.37E+06	6.30E+06	0.00E+00	6.49E+05	2.32E+06
Bcl9l	0.00E+00	0.00E+00	0.00E+00	1.71E+07	1.89E+07	1.03E+07	1.55E+07

Hdac4	0.00E+00	0.00E+00	0.00E+00	9.25E+06	1.43E+07	1.98E+07	1.44E+07
Tpr	0.00E+00	0.00E+00	0.00E+00	2.29E+06	2.71E+06	7.36E+06	4.12E+06
Tfap2a	0.00E+00	0.00E+00	0.00E+00	1.40E+07	1.22E+07	1.33E+07	1.31E+07
Hdac2	0.00E+00	9.06E+06	0.00E+00	6.01E+06	5.18E+06	0.00E+00	3.73E+06
Tox4	0.00E+00	0.00E+00	0.00E+00	3.47E+06	3.59E+06	4.68E+06	3.92E+06
Per1	0.00E+00	0.00E+00	0.00E+00	2.97E+07	2.83E+07	1.85E+07	2.55E+07
Nrip1	0.00E+00	0.00E+00	0.00E+00	3.09E+06	3.67E+06	4.75E+06	3.84E+06
Irf2bpl	0.00E+00	0.00E+00	0.00E+00	0.00E+00	7.57E+06	7.55E+06	5.04E+06
Fam114a2	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.53E+06	1.44E+06	9.90E+05
Taf9	0.00E+00	9.61E+06	0.00E+00	7.33E+07	7.64E+07	7.97E+07	7.65E+07
Cic	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.97E+07	1.64E+07	1.20E+07
Smarcd2	0.00E+00	0.00E+00	0.00E+00	1.72E+07	1.27E+07	0.00E+00	9.97E+06
Ints12	1.14E+07	0.00E+00	0.00E+00	1.65E+07	1.88E+07	5.99E+06	1.38E+07
Tbcb	0.00E+00	0.00E+00	0.00E+00	5.21E+06	4.86E+06	5.29E+06	5.12E+06
Trir	0.00E+00	0.00E+00	7.15E+06	1.77E+07	1.07E+07	1.43E+07	1.42E+07
Nup50	1.69E+06	0.00E+00	0.00E+00	1.08E+07	1.17E+07	2.04E+07	1.43E+07
Mta2	0.00E+00	0.00E+00	0.00E+00	7.91E+06	5.89E+06	0.00E+00	4.60E+06

Table. 3 qPCR and genotyping primers

qPCR_Vim_F	CGGAAAGTGGAATCCTTGCAGG
qPCR_Vim_R	AGCAGTGAGGTCAGGCTTGGAA
qPCR_Gata3_F	CCTCTGGAGGAGGAACGCTAAT
qPCR_Gata3_R	GTTTCGGGTCTGGATGCCTTCT
qPCR_Trp63_F	CAGTACCTCCCTCAGCACACG
qPCR_Trp63_R	CAGAGGTGGGGAACTGTTGCC
qPCR_Rps16_F	GACGTCCCCTGGAGATGAT
qPCR_Rps16_R	CGAATATCCACACCAGCAAA
geno_mycSox9_mycF	GGAACAAAAACTTATTTCTGAAGAAGATCTG
geno_mycSox9_Sox9R	CCCTGAGATTGCCCAGAGTG

Table. 4 List of antibodies

Ab	Host	Dilution	Application	Cat#	Source
Ly6A/E-APCCy7	Rat	1:500	FACS	108126	BioLegend
CD49f-PECy7	Rat	1:1000	FACS	313622	BioLegend
CD34-Alexa660	Rat	1:50	FACS	50-0341-82	Invitrogen
CD45-biotin	Rat	1:200	FACS	103104	BioLegend
CD31-biotin	Rat	1:200	FACS	102404	BioLegend
CD140a-biotin	Rat	1:200	FACS	135910	BioLegend
CD117-biotin	Rat	1:200	FACS	105804	BioLegend
Streptavidin-FITC	N/A	1:1000	FACS	405202	BioLegend
TruStain FcX	Rat	1:100	FACS	101320	BioLegend
SOX9	Rabbit	1:250	CNR	ab5535	Millipore
MLL3/4	Rabbit	1:50	CNR	N/A	Wysocka Lab
totalH3	Mouse	1:67	MINT-ChIP	39763	Active Motif
H3K27ac	Rabbit	1:100	MINT-ChIP	39133	Active Motif
H3K4me1	Rabbit	1:67	MINT-ChIP	D1A9	Cell Signaling
SOX9	Rabbit	1:5000	IF	ab185966	Abcam
ITGA6	Rat	1:1000	IF	555734	BD
KRT14	Chicken	1:2000	IF	906004	BioLegend
KRT10	Rabbit	1:250	IF	N/A	Fuchs Lab
EpCAM	Rabbit	1:100	IF	ab71916	Abcam
KRT6	Guinea pig	1:1000	IF	N/A	Fuchs Lab
RUNX1	Rabbit	1:100	IF	ab229482	Abcam
GATA3	Rat	1:100	IF	14-9966-82	Invitrogen
HA-tag	Rabbit	1:1000	IF	C29F4	Cell Signaling
GFP	Chicken	1:2000	IF	N/A	Fuchs Lab
RFP	Rat	1:1000	IF	5F8	ChromoTek
MYC-tag	Rabbit	1:1000	IF	71D10	Cell Signaling
MYC-tag	Mouse	1:1000	WB	9B11	Cell Signaling
beta-Actin	Mouse	1:10000	WB	8H10D10	Cell Signaling
MLL4	Rabbit	1:200	WB	sc-293217	Santa Cruz Biotechnology
MYC-tag	Rabbit	6ul/30ul beads	co-IP	71D10	Cell Signaling
IgG	Rabbit	8ul/30ul beads	co-IP	2729S	Cell Signaling
Arid1a	Rabbit	1:50	CNR	cat#12354S	Cell Signaling
ARID1a	Rabbit	1:200	WB	cat#ab182560	Abcam
c-Jun	Rabbit	1:50	CNR	cat#3753S	Cell Signaling
BRG1	Rabbit	1:20	CNR	cat#13-2002	Epicypher

Table. 5 Sequencing depth of samples

Sequencing experiment	Sample name	Aligned and deduplicated read depth
Cut-and-Run	Sox9_0D_mllF7_1to50_rep1_rep2_POOL_fragmentsAtOrUnder120bp	16,842,376
Cut-and-Run	Sox9_14D_mllF7_1to50_rep1_rep2_POOL_fragmentsAtOrUnder120bp	14,787,092
Cut-and-Run	Sox9_7D_mllF7_1to50_rep1_rep2_POOL_fragmentsAtOrUnder120bp	20,021,710
Cut-and-Run	WT_dox3d_mllCNR_POOL_sorted_fragmentunder120bp	3,253,510
Cut-and-Run	WT_dox3d_mycCNR_under120bp_POOL_sorted	6,091,558
Cut-and-Run	WT_nodox_mllCNR_POOL_sorted_fragmentunder120bp	16,657,444
Cut-and-Run	WT_nodox_mycCNR_under120bp_POOL_sorted	4,940,618
Cut-and-Run	noHMG_mllCNR_POOL_sorted_fragmentunder120bp	10,461,714
Cut-and-Run	noHMG_mycCNR_under120bp_POOL_sorted	4,035,336
Cut-and-Run	noTA_mllCNR_POOL_sorted_fragmentunder120bp	12,076,070
Cut-and-Run	noTA_mycCNR_under120bp_POOL_sorted	4,875,308
Cut-and-Run	D0_rep1_rep2_POOL_under120bp	32,529,086
Cut-and-Run	W1_rep1_rep2_POOL_under120bp	17,354,030
Cut-and-Run	W2_rep1_rep2_POOL_under120bp	13,089,458
Cut-and-Run	W6_rep1_rep2_POOL_under120bp	23,822,438
Cut-and-Run	W12_rep1_rep2_POOL_under120bp	21,758,868
Cut-and-Run	WT_dox_Brg1CNR_rep1_fragmentunder120bp	2,699,652
Cut-and-Run	noHMG_dox_Brg1CNR_rep1_fragmentunder120bp	3,256,860
Cut-and-Run	WT_nodox_Brg1CNR_rep1_fragmentunder120bp	7,242,716
Cut-and-Run	noHMG_dox_CSTArid1aCNR_rep2_fragmentunder120bp	898,082
Cut-and-Run	WT_NOdox_CSTArid1aCNR_rep2_fragmentunder120bp	1,102,352
Cut-and-Run	WT_dox_CSTArid1aCNR_rep2_fragmentunder120bp	443,050
Cut-and-Run	WT_dox_JunCNR_rep1_fragmentunder120bp	3,227,292
Cut-and-Run	WT_nodox_JunCNR_rep1_fragmentunder120bp	3,797,824
Cut-and-Run	noHMG_dox_JunCNR_rep1_fragmentunder120bp	1,561,838
ATAC	noHMG_rep1_rep2_POOL	92,690,912
ATAC	noTA_rep1_rep2_POOL	167,114,236
ATAC	WT_D0_rep1_rep2_POOL	83,369,464
ATAC	WT_D3_rep1_rep2_POOL	107,949,908
ATAC	K14SOX9_AFOS_rep1_rep2_readsPOOL	98,263,856
ATAC	K14only_AFOS_rep1_rep2_readsPOOL	99,744,508
ATAC	SOX9_W6_Rep1_Rep2_dedup_sorted_PEPATAC_normalized	49,829,490
ATAC	SOX9_M3_Rep1_Rep2_dedup_sorted_PEPATAC_normalized	92,332,298
ATAC	SOX9_D14_Rep1_Rep2_dedup_sorted_PEPATAC_normalized	32,562,188
ATAC	SOX9_D7_Rep1_Rep2_dedup_sorted_PEPATAC_normalized	72,107,920

ATAC	SOX9_D0_Rep1_Rep2_sort_dedup_sorted_PEPATAC_normalized	83,704,660
ATAC	K14SOX9_ARID1a_rep1_rep2_readsPOOL	78,031,988
MINT-ChIP	SOX9_12W_IFE_High_Low_POOL_TOTALH3_MERGE_sorted	20,007,732
MINT-ChIP	SOX9_6W_IFE_High_Low_POOL_TOTALH3_MERGE_sorted	67,713,364
MINT-ChIP	SOX9_D0_IFE_High_Low_POOL_H3K27ac_MERGE_sorted	6,962,366
MINT-ChIP	SOX9_D0_IFE_High_Low_POOL_H3K4me1_MERGE_sorted	46,655,899
MINT-ChIP	SOX9_D0_IFE_High_Low_POOL_TOTALH3_MERGE_sorted	70,350,434
MINT-ChIP	SOX9_D14_IFE_High_Low_POOL_TOTALH3_MERGE_sorted	67,982,223
MINT-ChIP	SOX9_D7_IFE_High_Low_POOL_H3K27ac_MERGE_sorted	3,785,286
MINT-ChIP	SOX9_D7_IFE_High_Low_POOL_H3K4me1_MERGE_sorted	16,137,643
MINT-ChIP	SOX9_D7_IFE_High_Low_POOL_TOTALH3_MERGE_sorted	26,301,826
MINT-ChIP	SOX9_W12_IFE_High_Low_POOL_H3K27ac_MERGE_sorted	1,342,114
MINT-ChIP	SOX9_W12_IFE_High_Low_POOL_H3K4me1_MERGE_sorted	15,304,804
MINT-ChIP	SOX9_W2_IFE_High_Low_POOL_H3K27ac_MERGE_sorted	4,959,398
MINT-ChIP	SOX9_W2_IFE_High_Low_POOL_H3K4me1_MERGE_sorted	16,694,998
MINT-ChIP	SOX9_W6_IFE_High_Low_POOL_H3K27ac_MERGE_sorted	9,412,015
MINT-ChIP	SOX9_W6_IFE_High_Low_POOL_H3K4me1_MERGE_sorted	35,913,180
MINT-ChIP	WT_dox_H3K4me1_MERGE_POOL	30,630,980
MINT-ChIP	WT_nodox_H3K4me1_MERGE_POOL	25,607,945
MINT-ChIP	noHMG_dox_H3K4me1_MERGE_POOL	24,534,294
MINT-ChIP	noTA_dox_H3K4me1_MERGE_POOL	26,643,114

7.3 Bibliography

Adam, R.C., Yang, H., Rockowitz, S., Larsen, S.B., Nikolova, M., Oristian, D.S., Polak, L., Kadaja, M., Asare, A., Zheng, D., and Fuchs, E. (2015). Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. Nature *521*, 366-370.

Adams, R., Yang, H., Rockowitz, S., Larsen, S., Nikolova, M., Oristian, D., Polak, L., Kadaja, M., Asare, A., Zheng, D., and Fuchs, E. (2015). Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. Nature *521*, 366-370.

Aguilar-Medina, M., Avendano-Felix, M., Lizarraga-Verdugo, E., Bermudez, M., Romero-Quintana, J.G., Ramos-Payan, R., Ruiz-Garcia, E., and Lopez-Camarillo, C. (2019). SOX9 Stem-Cell Factor: Clinical and Functional Relevance in Cancer. J Oncol *2019*, 6754040. 10.1155/2019/6754040.

Ahn, S., Olive, M., Aggarwal, S., Krylov, D., Ginty, D.D., and Vinson, C. (1998). A dominant-negative inhibitor of CREB reveals that it is a general mediator of stimulus-dependent transcription of c-fos. Mol Cell Biol *18*, 967-977. 10.1128/MCB.18.2.967.

Akiyama, H., Chaboissier, M.C., Martin, J.F., Schedl, A., and de Crombrugghe, B. (2002). The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. Genes Dev *16*, 2813-2828.

Arnold, C.D., Gerlach, D., Stelzer, C., Boryn, L.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science *339*, 1074-1077. 10.1126/science.1232542.

Balsalobre, A., and Drouin, J. (2022). Pioneer factors as master regulators of the epigenome and cell fate. Nature Reviews Molecular Cell Biology *23*, 449-464. 10.1038/s41580-022-00464-z.

Beronja, S., Livshits, G., Williams, S., and Fuchs, E. (2010). Rapid functional dissection of genetic fetworks via tissue specific transduction and RNAi in mouse embryos. Nature Medicine *16*, 821-827.

Bi, W., Huang, W., Whitworth, D.J., Deng, J.M., Zhang, Z., Behringer, R.R., and de Crombrugghe, B. (2001). Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. Proc Natl Acad Sci U S A *98*, 6698-6703. 10.1073/pnas.111092198.

Blanpain, C., Lowry, W.E., Geoghegan, A., Polak, L., and Fuchs, E. (2004). Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. Cell *118*, 635-648.

Blassberg, R., Patel, H., Watson, T., Gouti, M., Metzis, V., Delas, M.J., and Briscoe, J. (2022). Sox2 levels regulate the chromatin occupancy of WNT mediators in epiblast progenitors responsible for vertebrate body formation. Nat Cell Biol *24*, 633-644. 10.1038/s41556-022-00910-2.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218.

Cangkrama, M., Ting, S.B., and Darido, C. (2013). Stem Cells behind the Barrier. International Journal of Molecular Sciences *14*, 13670-13686. 10.3390/ijms140713670.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell *168*, 442-+. 10.1016/j.cell.2016.12.016.

Corces, M.R., Granja, J.M., Shams, S., Louie, B.H., Seoane, J.A., Zhou, W., Silva, T.C., Groeneveld, C., Wong, C.K., Cho, S.W., et al. (2018). The chromatin accessibility landscape of primary human cancers. Science *362*. 10.1126/science.aav1898.

Crowson, A.N. (2006). Basal cell carcinoma: biology, morphology and clinical implications. Modern Pathology *19*, S127-S147. 10.1038/modpathol.3800512.

Dejosez, M., Dall'Agnese, A., Ramamoorthy, M., Platt, J., Yin, X., Hogan, M., Brosh, R., Weintraub, A.S., Hnisz, D., Abraham, B.J., et al. (2023). Regulatory architecture of housekeeping genes is driven by promoter assemblies. Cell Rep *42*, 112505. 10.1016/j.celrep.2023.112505.

Dorighi, K.M., Swigut, T., Henriques, T., Bhanu, N.V., Scruggs, B.S., Nady, N., Still, C.D., 2nd, Garcia, B.A., Adelman, K., and Wysocka, J. (2017). Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. Mol Cell *66*, 568-576 e564. 10.1016/j.molcel.2017.04.018.

Fijalkowska, M., Koziej, M., and Antoszewski, B. (2021). Detailed head localization and incidence of skin cancers. Scientific Reports *11*. ARTN 12391 10.1038/s41598-021-91942-5.

Fiore, V.F., Krajnc, M., Quiroz, F.G., Levorse, J., Pasolli, H.A., Shvartsman, S.Y., and Fuchs, E. (2020). Mechanics of a multilayer epithelium instruct tumour architecture and function. Nature *585*, 433-439. 10.1038/s41586-020-2695-9.

Foster, J.W., Dominguez-Steglich, M.A., Guioli, S., Kwok, C., Weller, P.A., Stevanovic, M., Weissenbach, J., Mansour, S., Young, I.D., Goodfellow, P.N., and et al. (1994). Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. Nature *372*, 525-530. 10.1038/372525a0.

Fuglerud, B.M., Drissler, S., Lotto, J., Stephan, T.L., Thakur, A., Cullum, R., and Hoodless, P.A. (2022). SOX9 reprograms endothelial cells by altering the chromatin landscape. Nucleic Acids Research *50*, 8547-8565. 10.1093/nar/gkac652.

Gaiser, M.R., Hirsch, D., and Gaiser, T. (2018). Loss of epithelial cell adhesion molecule (EpCAM) in infiltrative basal cell carcinoma. International Journal of Clinical and Experimental Pathology *11*, 406-412.

Gerstberger, S., Jiang, Q., and Ganesh, K. (2023). Metastasis. Cell *186*, 1564-1579. 10.1016/j.cell.2023.03.003.

Gonzales, K.A.U., and Fuchs, E. (2017). Skin and Its Regenerative Powers: An Alliance between Stem Cells and Their Niche. Dev Cell *43*, 387-401. 10.1016/j.devcel.2017.10.001.

Gualdi, R., Bossard, P., Zheng, M.H., Hamada, Y., Coleman, J.R., and Zaret, K.S. (1996). Hepatic specification of the gut endoderm in vitro: Cell signaling and transcriptional control. Genes Dev *10*, 1670-1682. DOI 10.1101/gad.10.13.1670.

Guzzi, N., Muthukumar, S., Ciesla, M., Todisco, G., Ngoc, P.C.T., Madej, M., Munita, R., Fazio, S., Ekstrom, S., Mortera-Blanco, T., et al. (2022). Pseudouridine-modified tRNA fragments repress aberrant protein synthesis and predict leukaemic progression in myelodysplastic syndrome. Nat Cell Biol *24*, 299-306. 10.1038/s41556-022-00852-9.

Haensel, D., Gaddam, S., Li, N.Y., Gonzalez, F., Patel, T., Cloutier, J.M., Sarin, K.Y., Tang, J.Y., Rieger, K.E., Aasi, S.Z., and Oro, A.E. (2022). LY6D marks pre-existing resistant basosquamous tumor subpopulations. Nat Commun *13*, 7520. 10.1038/s41467-022-35020-y.

He, S., Wu, Z.H., Tian, Y., Yu, Z.S., Yu, J.L., Wang, X.X., Li, J., Liu, B.J., and Xu, Y.H. (2020). Structure of nucleosome-bound human BAF complex. Science *367*, 875-+. 10.1126/science.aaz9761.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589. 10.1016/j.molcel.2010.05.004.

Hendy, O., Serebreni, L., Bergauer, K., Muerdter, F., Huber, L., Nemcko, F., and Stark, A. (2022). Developmental and housekeeping transcriptional programs in Drosophila require distinct chromatin remodelers. Molecular Cell *82*, 3598-+. 10.1016/j.molcel.2022.08.019.

Henikoff, S., and Shilatifard, A. (2011). Histone modification: cause or cog? Trends Genet *27*, 389-396. 10.1016/j.tig.2011.06.006.

Henley, M.J., and Koehler, A.N. (2021). Advances in targeting 'undruggable' transcription factors with small molecules. Nat Rev Drug Discov *20*, 669-688. 10.1038/s41573-021-00199-0.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13-23. 10.1016/j.cell.2017.02.007.

Hsu, Y.C., Li, L., and Fuchs, E. (2014). Transit-amplifying cells orchestrate stem cell activity and tissue regeneration. Cell *157*, 935-949.

Iurlaro, M., Stadler, M.B., Masoni, F., Jagani, Z., Galli, G.G., and Schubeler, D. (2021). Mammalian SWI/SNF continuously restores local accessibility to chromatin. Nat Genet *53*, 279-287. 10.1038/s41588-020-00768-w.

Iwafuchi-Doi, M., and Zaret, K.S. (2014). Pioneer transcription factors in cell reprogramming. Genes Dev *28*, 2679-2692. 10.1101/gad.253443.114.

Jo, A., Denduluri, S., Zhang, B., Wang, Z., Yin, L., Yan, Z., Kang, R., Shi, L.L., Mok, J., Lee, M.J., and Haydon, R.C. (2014). The versatile functions of Sox9 in development, stem cells, and human diseases. Genes Dis *1*, 149-161. 10.1016/j.gendis.2014.09.004.

Jozwik, K.M., Chernukhin, I., Serandour, A.A., Nagarajan, S., and Carroll, J.S. (2016). FOXA1 Directs H3K4 Monomethylation at Enhancers via Recruitment of the Methyltransferase MLL3. Cell Reports *17*, 2715-2723. 10.1016/j.celrep.2016.11.028.

Julian, L.M., McDonald, A.C., and Stanford, W.L. (2017). Direct reprogramming with SOX factors: masters of cell fate. Curr Opin Genet Dev *46*, 24-36. 10.1016/j.gde.2017.06.005.

Kadaja, M., Keyes, B.E., Lin, M., Pasolli, H.A., Genander, M., Polak, L., Stokes, N., Zheng, D., and Fuchs, E. (2014). SOX9: a stem cell transcriptional regulator of secreted niche signaling factors. Genes Dev *28*, 328-341.

Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J., and Crabtree, G.R. (2013). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nat Genet *45*, 592-601. 10.1038/ng.2628.

Kall, L., Canterbury, J.D., Weston, J., Noble, W.S., and MacCoss, M.J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods *4*, 923-925. 10.1038/nmeth1113.

Kamachi, Y., and Kondoh, H. (2013). Sox proteins: regulators of cell fate specification and differentiation. Development *140*, 4129-4144. 10.1242/dev.091793.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Cheneby, J., Kulkarni, S.R., Tan, G., et al. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res *46*, D260-D266. 10.1093/nar/gkx1126.

Kim, D.I., Jensen, S.C., Noble, K.A., Kc, B., Roux, K.H., Motamedchaboki, K., and Roux, K.J. (2016). An improved smaller biotin ligase for BioID proximity labeling. Mol Biol Cell *27*, 1188-1196. 10.1091/mbc.E15-12-0844.

Kim, D.I., and Roux, K.J. (2016). Filling the Void: Proximity-Based Labeling of Proteins in Living Cells. Trends Cell Biol *26*, 804-817. 10.1016/j.tcb.2016.09.004.

Kuonen, F., Li, N.Y., Haensel, D., Patel, T., Gaddam, S., Yerly, L., Rieger, K., Aasi, S., and Oro, A.E. (2021). c-FOS drives reversible basal to squamous cell carcinoma transition. Cell Rep *37*, 109774. 10.1016/j.celrep.2021.109774.

Kwok, C., Weller, P.A., Guioli, S., Foster, J.W., Mansour, S., Zuffardi, O., Punnett, H.H., Dominguez-Steglich, M.A., Brook, J.D., Young, I.D., and et al. (1995). Mutations in SOX9, the gene responsible for Campomelic dysplasia and autosomal sex reversal. Am J Hum Genet *57*, 1028-1036.

Larsen, S.B., Cowley, C.J., Sajjath, S.M., Barrows, D., Yang, Y., Carroll, T.S., and Fuchs, E. (2021). Establishment, maintenance, and recall of inflammatory memory. Cell Stem Cell *28*, 1758-1774 e1758. 10.1016/j.stem.2021.07.001.

Larsimont, J.C., Youssef, K.K., Sanchez-Danes, A., Sukumaran, V., Defrance, M., Delatte, B., Liagre, M., Baatsen, P., Marine, J.C., Lippens, S., et al. (2015). Sox9 Controls Self-Renewal of Oncogene Targeted Cells and Links Tumor Initiation and Invasion. Cell Stem Cell *17*, 60-73. 10.1016/j.stem.2015.05.008

S1934-5909(15)00217-9 [pii].

Lee, C.S., Friedman, J.R., Fulmer, J.T., and Kaestner, K.H. (2005). The initiation of liver development is dependent on Foxa transcription factors. Nature *435*, 944-947. 10.1038/nature03649.

Lee, K., Cho, H., Rickert, R.W., Li, Q.V., Pulecio, J., Leslie, C.S., and Huangfu, D.W. (2019). FOXA2 Is Required for Enhancer Priming during Pancreatic Differentiation. Cell Reports *28*, 382-+. 10.1016/j.celrep.2019.06.034.

Lefebvre, V., Huang, W., Harley, V.R., Goodfellow, P.N., and de Crombrugghe, B. (1997). SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. Mol Cell Biol *17*, 2336-2346. 10.1128/MCB.17.4.2336.

Li, Z., Schulz, M.H., Look, T., Begemann, M., Zenke, M., and Costa, I.G. (2019). Identification of transcription factor binding sites using ATAC-seq. Genome Biol *20*, 45. 10.1186/s13059-019-1642-2.

Lyons, H., Veettil, R.T., Pradhan, P., Fornero, C., De La Cruz, N., Ito, K., Eppert, M., Roeder, R.G., and Sabari, B.R. (2023). Functional partitioning of transcriptional regulators by patterned charge blocks. Cell *186*, 327-345 e328. 10.1016/j.cell.2022.12.013.

Matheu, A., Collado, M., Wise, C., Manterola, L., Cekaite, L., Tye, A.J., Canamero, M., Bujanda, L., Schedl, A., Cheah, K.S., et al. (2012). Oncogenicity of the developmental transcription factor Sox9. Cancer Res *72*, 1301-1315. 10.1158/0008-5472.CAN-11-3660.

Mayran, A., Khetchoumian, K., Hariri, F., Pastinen, T., Gauthier, Y., Balsalobre, A., and Drouin, J. (2018). Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. Nature Genetics *50*, 259-+. 10.1038/s41588-017-0035-2.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501. 10.1038/nbt.1630.

Meers, M.P., Bryson, T.D., Henikoff, J.G., and Henikoff, S. (2019a). Improved CUT&RUN chromatin profiling tools. Elife *8*. 10.7554/eLife.46314.

Meers, M.P., Janssens, D.H., and Henikoff, S. (2019b). Pioneer Factor-Nucleosome Binding Events during Differentiation Are Motif Encoded. Mol Cell *75*, 562-575 e565. 10.1016/j.molcel.2019.05.025.

Meers, M.P., Tenenbaum, D., and Henikoff, S. (2019c). Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. Epigenetics Chromatin *12*, 42. 10.1186/s13072-019-0287-4.

Ming, Z.H., Vining, B., Bagheri-Fam, S., and Harley, V. (2022). SOX9 in organogenesis: shared and unique transcriptional functions. Cellular and Molecular Life Sciences *79*. ARTN 522 10.1007/s00018-022-04543-4.

Nalawansha, D.A., and Crews, C.M. (2020). PROTACs: An Emerging Therapeutic Modality in Precision Medicine. Cell Chem Biol *27*, 998-1014. 10.1016/j.chembiol.2020.07.020.

Naqvi, S., Kim, S., Hoskens, H., Matthews, H.S., Spritz, R.A., Klein, O.D., Hallgrimsson, B., Swigut, T., Claes, P., Pritchard, J.K., and Wysocka, J. (2023). Precise modulation of transcription factor levels identifies features underlying dosage sensitivity. Nature Genetics. 10.1038/s41588-023-01366-2.

Nguyen, H., Rendl, M., and Fuchs, E. (2006). Tcf3 governs stem cell features and represses cell fate determination in skin. Cell *127*, 171-183.

Nowak, J.A., Polak, L., Pasolli, H.A., and Fuchs, E. (2008). Hair follicle stem cells are specified and function in early skin morphogenesis. Cell Stem Cell *3*, 33-43.

Olive, M., Krylov, D., Echlin, D.R., Gardner, K., Taparowsky, E., and Vinson, C. (1997). A dominant negative to activation protein-1 (AP1) that abolishes DNA binding and inhibits oncogenesis. J Biol Chem *272*, 18586-18594. 10.1074/jbc.272.30.18586.

Oro, A.E., and Scott, M.P. (1998). Splitting hairs: dissecting roles of signaling systems in epidermal development. Cell *95*, 575-578. 10.1016/s0092-8674(00)81624-4.

Oshimori, N., and Fuchs, E. (2012). Paracrine TGF-beta signaling counterbalances BMP-mediated repression in hair follicle stem cell activation. Cell Stem Cell *10*, 63-75.

Ouspenskaia, T., Matos, I., Mertz, A.F., Fiore, V.F., and Fuchs, E. (2016). WNT-SHH Antagonism Specifies and Expands Stem Cells prior to Niche Formation. Cell *164*, 156-169.

Park, Y.K., Lee, J.E., Yan, Z., McKernan, K., O'Haren, T., Wang, W., Peng, W., and Ge, K. (2021). Interplay of BAF and MLL4 promotes cell type-specific enhancer activation. Nat Commun *12*, 1630. 10.1038/s41467-021-21893-y.

Parker, S.C.J., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., Black, B.L., et al. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc Natl Acad Sci U S A *110*, 17921-17926. 10.1073/pnas.1317023110.

Rao, R.C., and Dou, Y.L. (2015). Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. Nature Reviews Cancer *15*, 334-346. 10.1038/nrc3929.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc *2*, 1896-1906. 10.1038/nprot.2007.261.

Rheinwald, J.G., and Green, H. (1977). Epidermal growth factor and the multiplication of cultured human epidermal keratinocytes. Nature *265*, 421-424.

Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods *14*, 975-978. 10.1038/nmeth.4401.

Schick, S., Grosche, S., Kohl, K.E., Drpic, D., Jaeger, M.G., Marella, N.C., Imrichova, H., Lin, J.G., Hofstatter, G., Schuster, M., et al. (2021). Acute BAF perturbation causes immediate changes in chromatin accessibility. Nat Genet *53*, 269-278. 10.1038/s41588-021-00777-3.

Sima, X., He, J., Peng, J., Xu, Y., Zhang, F., and Deng, L. (2019). The genetic alteration spectrum of the SWI/SNF complex: The oncogenic roles of BRD9 and ACTL6A. PLoS One *14*, e0222305. 10.1371/journal.pone.0222305.

Skene, P.J., Henikoff, J.G., and Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. Nat Protoc *13*, 1006-1019. 10.1038/nprot.2018.015.

Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife *6*.

Smith, J.P., Corces, M.R., Xu, J., Reuter, V.P., Chang, H.Y., and Sheffield, N.C. (2021). PEPATAC: an optimized pipeline for ATAC-seq data analysis with serial alignments. NAR Genom Bioinform *3*, lqab101. 10.1093/nargab/lqab101.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994-1004. 10.1016/j.cell.2012.09.045.

Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell *161*, 555-568.

Stovner, E.B., and Saetrom, P. (2019). epic2 efficiently finds diffuse domains in ChIP-seq data. Bioinformatics *35*, 4392-4393. 10.1093/bioinformatics/btz232.

Sze, C.C., and Shilatifard, A. (2016). MLL3/MLL4/COMPASS Family on Epigenetic Regulation of Enhancer Function and Cancer. Cold Spring Harb Perspect Med *6*. 10.1101/cshperspect.a026427.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861-872.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Takaku, M., Grimm, S.A., Shimbo, T., Perera, L., Menafra, R., Stunnenberg, H.G., Archer, T.K., Machida, S., Kurumizaka, H., and Wade, P.A. (2016). GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler. Genome Biology *17*. ARTN 36
10.1186/s13059-016-0897-0.

Tanaka, H., Takizawa, Y., Takaku, M., Kato, D., Kumagawa, Y., Grimm, S.A., Wade, P.A., and Kurumizaka, H. (2020). Interaction of the pioneer transcription factor GATA3 with nucleosomes. Nat Commun *11*. ARTN 4136
10.1038/s41467-020-17959-y.

Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B., et al. (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell *163*, 1611-1627.
10.1016/j.cell.2015.11.024.

van Galen, P., Viny, A.D., Ram, O., Ryan, R.J., Cotton, M.J., Donohue, L., Sievers, C., Drier, Y., Liau, B.B., Gillespie, S.M., et al. (2016). A Multiplexed System for Quantitative Comparisons of Chromatin Landscapes. Mol Cell *61*, 170-180. 10.1016/j.molcel.2015.11.003.

Vandereyken, K., Sifrim, A., Thienpont, B., and Voet, T. (2023). Methods and applications for single-cell and spatial multi-omics. Nat Rev Genet, 1-22. 10.1038/s41576-023-00580-2.

Vasioukhin, V., Degenstein, L., Wise, B., and Fuchs, E. (1999). The magical touch: genome targeting in epidermal stem cells induced by tamoxifen application to mouse skin. Proc Natl Acad Sci U S A *96*, 8551-8556.

Vidal, V.P., Chaboissier, M.C., Lutzkendorf, S., Cotsarelis, G., Mill, P., Hui, C.C., Ortonne, N., Ortonne, J.P., and Schedl, A. (2005). Sox9 is essential for outer root sheath differentiation and the formation of the hair stem cell compartment. Curr Biol *15*, 1340-1351.

Wagner, T., Wirth, J., Meyer, J., Zabel, B., Held, M., Zimmer, J., Pasantes, J., Bricarelli, F.D., Keutel, J., Hustert, E., et al. (1994). Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. Cell *79*, 1111-1120. 10.1016/0092-8674(94)90041-8.

Wang, H., Yang, Y., Liu, J., and Qian, L. (2021). Direct cell reprogramming: approaches, mechanisms and progress. Nat Rev Mol Cell Biol *22*, 410-424. 10.1038/s41580-021-00335-z.

Wang, L., He, S., Yuan, J., Mao, X., Cao, Y., Zong, J., Tu, Y., and Zhang, Y. (2012). Oncogenic role of SOX9 expression in human malignant glioma. Med Oncol *29*, 3484-3490. 10.1007/s12032-012-0267-z.

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

Wirth, J., Wagner, T., Meyer, J., Pfeiffer, R.A., Tietze, H.U., Schempp, W., and Scherer, G. (1996). Translocation breakpoints in three patients with campomelic dysplasia and autosomal sex reversal map more than 130 kb from SOX9. Hum Genet *97*, 186-193. 10.1007/BF02265263.

Wright, E.M., Snopek, B., and Koopman, P. (1993). Seven new members of the Sox gene family expressed during mouse development. Nucleic Acids Res *21*, 744. 10.1093/nar/21.3.744.

Yang, D., Jones, M.G., Naranjo, S., Rideout, W.M., 3rd, Min, K.H.J., Ho, R., Wu, W., Replogle, J.M., Page, J.L., Quinn, J.J., et al. (2022). Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. Cell *185*, 1905-1923 e1925. 10.1016/j.cell.2022.04.015.

Yang, H., Adam, R.C., Ge, Y., Hua, Z.L., and Fuchs, E. (2017). Epithelial-Mesenchymal Micro-niches Govern Stem Cell Lineage Choices. Cell *169*, 483-496 e413.

Yao, C.D., Haensel, D., Gaddam, S., Patel, T., Atwood, S.X., Sarin, K.Y., Whitson, R.J., McKellar, S., Shankar, G., Aasi, S., et al. (2020). AP-1 and TGF ss cooperativity drives non-canonical Hedgehog signaling in resistant basal cell carcinoma. Nat Commun *11*. ARTN 5079 10.1038/s41467-020-18762-5.

Yi, R. (2017). Concise Review: Mechanisms of Quiescent Hair Follicle Stem Cell Regulation. Stem Cells *35*, 2323-2330. 10.1002/stem.2696.

Zaret, K.S. (2020). Pioneer Transcription Factors Initiating Gene Network Changes. Annu Rev Genet *54*, 367-385. 10.1146/annurev-genet-030220-015007.

Zhou, C.H., Ye, L.P., Ye, S.X., Li, Y., Zhang, X.Y., Xu, X.Y., and Gong, L.Y. (2012). Clinical significance of SOX9 in human non-small cell lung cancer progression and overall patient survival. J Exp Clin Cancer Res *31*, 18. 10.1186/1756-9966-31-18.

Zhu, C., Zhang, Y., Li, Y.E., Lucero, J., Behrens, M.M., and Ren, B. (2021). Joint profiling of histone modifications and transcriptome in single cells from mouse brain. Nat Methods *18*, 283-292. 10.1038/s41592-021-01060-3.