

The pioneer factor SOX9 competes for epigenetic factors to switch stem cell fates

Received: 8 January 2023

Accepted: 8 June 2023

Published online: 24 July 2023

 Check for updates

Yihao Yang^{1,7}, Nicholas Gomez^{1,2,7}, Nicole Infarinato^{1,3}, Rene C. Adam^{1,4}, Megan Sribour¹, Inwha Baek^{1,5}, Mélanie Laurin^{1,6} & Elaine Fuchs¹  

During development, progenitors simultaneously activate one lineage while silencing another, a feature highly regulated in adult stem cells but derailed in cancers. Equipped to bind cognate motifs in closed chromatin, pioneer factors operate at these crossroads, but how they perform fate switching remains elusive. Here we tackle this question with SOX9, a master regulator that diverts embryonic epidermal stem cells (EpdSCs) into becoming hair follicle stem cells. By engineering mice to re-activate SOX9 in adult EpdSCs, we trigger fate switching. Combining epigenetic, proteomic and functional analyses, we interrogate the ensuing chromatin and transcriptional dynamics, slowed temporally by the mature EpdSC niche microenvironment. We show that as SOX9 binds and opens key hair follicle enhancers *de novo* in EpdSCs, it simultaneously recruits co-factors away from epidermal enhancers, which are silenced. Unhinged from its normal regulation, sustained SOX9 subsequently activates oncogenic transcriptional regulators that chart the path to cancers typified by constitutive SOX9 expression.

From development to malignancy, cells face decisions of fate determination. Governing the reprogramming from one fate to another, pioneer factors are transcription factors that can recognize and access their cognate binding motifs in compacted and repressed chromatin¹. In vitro studies have shown that when a pioneer factor binds, it displaces the nucleosome, permitting the opening and remodelling of the chromatin landscape to change gene expression^{2,3}. Recent studies have begun to uncover interactions of various pioneer factors with histone-modifying enzymes and members of the SWI/SNF chromatin remodelling complex². However, the order of events in chromatin remodelling has remained elusive due to the rapid time frame of reprogramming *in vitro* where cells are outside local restraints of their tissue microenvironments. Even less clear is the role of pioneer factors in accomplishing the other side of fate switching, namely the silencing of a cell's previous identity².


In this Article, seeking the answers to these enigmas, we focused on the SOX superfamily of context-specific pioneer factors, whose

members are at the nexus of critical cell fate choices in embryonic development, tissue homeostasis and transition to malignancy⁴⁻⁷. In skin, SOX9 is first expressed when multipotent embryonic epidermal progenitors bifurcate to become SOX9⁺ hair follicle stem cells (HFSCs) and SOX9^{neg} epidermal stem cells (EpdSCs)⁸⁻¹⁰. In the next step of hair follicle morphogenesis, SOX9⁺ HFSCs bifurcate again to form SOX9^{neg} transit amplifying hair shaft progenitors. Basal cell carcinoma (BCC) formation from EpdSCs resembles the initial steps of embryonic hair follicle morphogenesis, but once re-activated, SOX9 is sustained, leading to invaginating follicle-like tumour masses that lack hair lineages¹¹⁻¹⁴. Here we recapitulated these reprogramming events by generating mice in which we could inducibly re-activate and sustain SOX9 expression in adult EpdSCs.

Not encountered *in vitro* or in embryogenesis, the mature tissue stem cell niche imposed physiological constraints that slowed SOX-mediated chromatin reprogramming. This enabled the unravelling

¹Howard Hughes Medical Institute, Robin Chemers Neustein Laboratory of Mammalian Cell Biology and Development, The Rockefeller University, New York, NY, USA. ²Present address: Allen Institute for Cell Sciences, Seattle, WA, USA. ³Present address: PRECISIONscientia, Yardley, PA, USA.

⁴Present address: Regeneron Pharmaceuticals, Tarrytown, NY, USA. ⁵Present address: Kyung Hee University, Seoul, South Korea. ⁶Present address: CHU de Québec-Université Laval Research Center, Québec City, Québec, Canada. ⁷These authors contributed equally: Yihao Yang, Nicholas Gomez.

 e-mail: fuchslb@rockefeller.edu

of sequential events that happen as SOX9 achieves a cell fate switch that when dysregulated later progresses to a tumourigenic state. By dissecting the temporal steps of epigenetic reprogramming, we show that SOX9 binds to closed chromatin at HFSC enhancers, where it recruits histone and chromatin modifiers to remodel and subsequently open chromatin for transcription. In doing so, SOX9 redistributes co-factors away from EpdSC enhancers, thereby silencing these genes indirectly but efficiently. Moreover, when the ability of SOX9 to bind DNA is abrogated, it still silences, but when it cannot bind chromatin remodellers, the switch fails altogether. Together, our findings illuminate how fate switching can be achieved through the direct activating functions of a pioneer factor, which then unleashes transcriptional repression through indirect competition for epigenetic co-factors. We further show that SOX9 regulates downstream transcription factors to drive tumourigenesis, which explains the delay in subsequent reprogramming events.

Results

SOX9 launches a transcriptional cascade towards BCC in EpdSC

To interrogate SOX9 reprogramming in adult tissue stem cells, we engineered mice harbouring a MYC-epitope-tagged *Sox9* transgene controlled by a tetracycline responsive enhancer and a minimal promoter (*TRE-Sox9*) (Extended Data Fig. 1a). After validating the specificity of transgene induction (Extended Data Fig. 1b), we bred selected mice to lines expressing the requisite tetracycline-inducible transcriptional activator (rtTA) driven by an epidermal (*Krt14*) promoter (*Krt14-rtTA*)¹⁵, and selected mice that induced MYC-SOX9 in EpdSCs at levels comparable to SOX9 in adult HFSCs (Extended Data Fig. 1c,d).

Upon doxycycline (DOX) administration (D0), adult mice were monitored weekly thereafter (Fig. 1a). Within the first 2 weeks, morphology and differentiation seemed unaffected (Extended Data Fig. 1e). However, by week (W)1, nuclear SOX9 was detected in the EpdSCs of the innermost (basal) epidermal layer (Fig. 1b). By W2, a rise in proliferation was detected, similar to that seen when SOX9 is naturally induced in embryonic epidermis¹⁰ (Extended Data Fig. 1f).

Between W2 and W6, de novo invaginations began to grow between native HFSCs (Fig. 1b and Extended Data Fig. 1e). As differentiation defects necessitated killing mice by W6, we monitored later events in SOX9 reprogramming by engrafting neonatal *Krt14-rtTA;TRE-Sox9* skin patches onto immunocompromised mice. Once normal skin pathology was restored (21 days after grafting), we induced SOX9 (Fig. 1a). By W12 post-induction, invaginations were dysplastic, resembling morphological and molecular (SOX9, EpCAM and KRT6) features of human BCCs (Extended Data Fig. 1g).

To gain further insights, we profiled the transcriptomic changes occurring in EpdSCs during SOX9-driven reprogramming. At each timepoint, two biological replicates of RNA sequencing (RNA-seq) were performed on fluorescence-activated cell sorting (FACS)-purified EpdSCs from *Krt14-rtTA;TRE-Sox9* skins (Extended Data Fig. 2a,b). By comparing transcriptomes across time, we identified the significantly variable genes ($P < 0.05$) along the reprogramming cascade (Fig. 1c and Supplementary Table 1). As expected, the D0 population displayed the hallmark signature of EpdSCs, replete with mRNAs encoding epidermal master regulator transcription factors, TRP63 and GATA3, key signalling pathways (NOTCH and EGFR), and epidermal structural proteins.

Despite few morphological changes within 2 weeks after induction, SOX9⁺ EpdSCs displayed dramatic transcriptional changes, mimicking transcriptional changes that occur when embryonic skin progenitors naturally induce SOX9 and divert from an epidermal to hair follicle fate¹⁰. Thus, epidermal genes were markedly suppressed, while classical markers of the embryonic hair bud and adult hair follicle outer root sheath (ORS) were upregulated, as supported by gene set enrichment analysis (GSEA) (Fig. 1c,d). The kinetics of these reprogramming events in adult EpdSCs, however, was markedly slower in the adult than in embryonic skin or in cultured cells, suggestive of the need to override the constraints of the mature epidermal niche.

As in BCC development, progression to mature HFSCs did not happen, in agreement with the need for *Sox9* downregulation for HFSCs to generate the hair and its channel^{10,16}. However, with sustained SOX9 expression, the transcriptional changes continued, and by W6–12, cancer-associated features appeared. At W12, GSEA revealed a strong correlation, both up and down, with the molecular signature of BCC compared with normal skin^{14,17} (Fig. 1c,d). Although the similarities in gene expression were strongest at late stages, they surfaced as early as W2, that is, before overt phenotypic changes, and clearly favoured a BCC versus squamous cell carcinoma (SCC) signature (Extended Data Fig. 2c,d).

SOX9 is a bona fide pioneer factor

To understand how SOX9 acts as a master regulator to induce these transcriptional dynamics, we began by performing CUT&RUN (cleavage under targets and release using nuclease; hereafter termed CNR) sequencing^{18,19} to temporally assay the binding of SOX9 to chromatin, and assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq)^{20,21} to interrogate chromatin accessibility during reprogramming (Fig. 2a,b). Biological replicates were concordant, and the SOX motif was most enriched in our SOX9 CNR peak sets (Extended Data Fig. 3a–c).

SOX9 binding to chromatin occurred rapidly within W1 and before the rise in proliferation. In contrast, the increase in accessibility at SOX9-binding sites occurred between W1 and W2, indicating that SOX9 can bind to closed chromatin (Fig. 2a). In fact, of all the SOX9 CNR peaks pooled from W1 to W12, nearly 30% were situated within closed chromatin at D0 (Fig. 2c). Moreover, by W2, nucleosome occupancy was lost at these sites as measured by histone H3 (Fig. 2c). Additionally, these SOX9-bound opening peaks displayed a time-dependent decrease in CNR fragment length. These features are hallmarks of nucleosome displacement and pioneer factor activity²², providing compelling evidence that SOX9 in skin EpdSCs binds to its cognate motifs within closed chromatin, and subsequently perturbs nucleosomes.

SOX9 induces global chromatin changes at distal enhancers

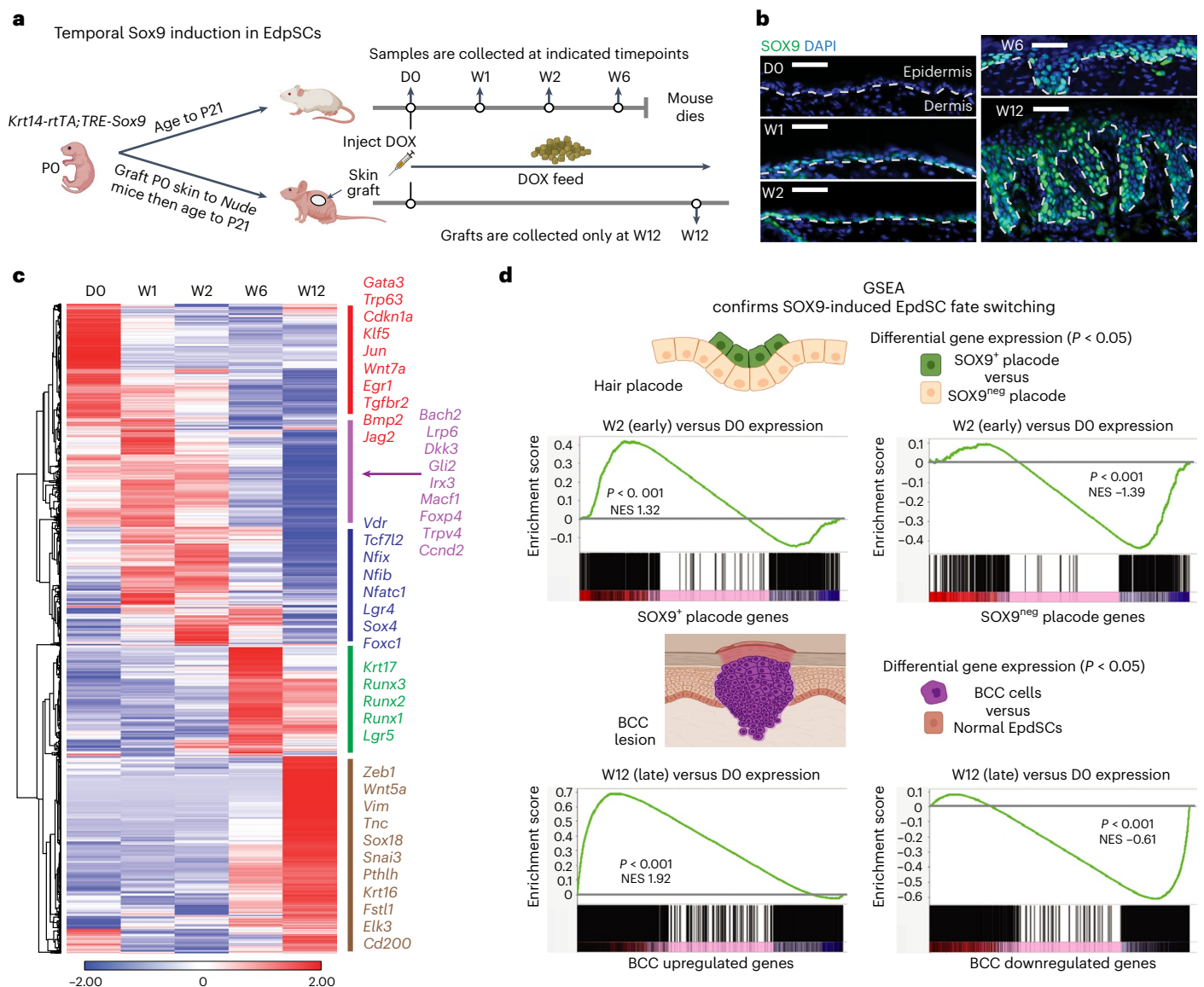
Since SOX9 bound to closed chromatin at W1, and presumptive nucleosome loss occurred soon thereafter, these events seemed unlikely to account fully for the tumourigenic transcriptional dynamics (Fig. 1c). Probing deeper, we examined how ATAC peaks and their associated genes changed over time.

Principal component analysis (PCA) of chromatin accessibility showed clustering according to times post-SOX9 induction (Fig. 2d). D0 and W1 samples clustered closely, W2 constituted an intermediary, and later timepoints (W6 and W12) made a second cluster. Comparative analyses across all timepoints revealed that many ATAC peaks were shared across samples, reflective of housekeeping genes and/or genes common to both EpdSCs and HFSCs (for example, *Krt5*) (Fig. 2e). By contrast, other ATAC peaks exhibited strikingly dynamic behaviour (for example, WNT-target *Cttnb1*), indicative of SOX9-induced temporal chromatin remodelling. These dynamic changes were particularly striking at W2 after SOX9 induction (Fig. 2e). Of the peaks that opened by W2, many persisted thereafter.

Upon binning peaks as either static (present at all timepoints, $n = 38,079$) or dynamic (absent in at least one timepoint, $n = 62,626$), it was clear that dynamic peaks were substantially more enriched in distal intergenic regions than static peaks (Fig. 2f and Extended Data Fig. 3e), suggesting a special role for SOX9 in eliciting chromatin changes at enhancers.

Direct and indirect chromatin remodelling induced by SOX9

K-means clustering of the dynamic ATAC peaks resolved the temporal changes in chromatin accessibility following SOX9 induction. Although more than 10,000 peaks (cluster C4) opened at later timepoints (Fig. 3a),



revealed five distinct patterns, shown as coloured bars on the right with representative transcripts indicative of specific fates. **d**, GSEA of W2 versus D0 SOX9-induced expression changes in adult EdpSCs compared with SOX9⁺ and SOX9^{neg} wild-type placode gene signatures from embryonic day E15.5 (top) (Kolmogorov–Smirnov test, $P < 0.001$ for both gene sets), and W12 versus D0 SOX9-induced EdpSC expression changes compared to BCC upregulated and downregulated signatures (bottom) (Kolmogorov–Smirnov test, $P < 0.001$ for both gene sets). NES, normalized enrichment score.

the most substantial changes occurred between W1 and W2. Using Genomic Regions Enrichment of Annotations Tool (GREAT), we assessed the biological pathways associated with each of the six clusters.

C1 and C6 closed within the first 2 weeks and were enriched for pathways with direct importance to EdpSCs (Fig. 3b and Supplementary Table 2). By contrast, C2 and C5 markedly increased their chromatin accessibility during this time and were enriched for hair follicle development and SHH signalling, key not only in stimulating ORS/HFSC lineage proliferation^{23–25} but also in driving BCCs^{11,12,14} (Fig. 3b). Also notable was a downregulation of API, EGFR and TGFβ signalling pathways, which are known to be elevated in BCCs that develop resistance to SHH inhibitors²⁶. Many of the pathways enriched in the C2/C5 clusters were also implicated in other cancers previously associated with SOX9 expression^{27–29}.

The role of SOX9 in activating the ORS/HFSC fate appeared to be direct, as chromatin that opened by W2 was associated with hair follicle development and displayed both SOX motifs and SOX9 binding (Fig. 3b,c). These peaks also persisted in an accessible state and were still prominent from W6 to W12 (Fig. 3a). By contrast, late-opening C4-associated genes were not related to HFSC fate. Many of their peaks only entered a more accessible state sometime after W2, coincident with the late BCC gene induction per our transcriptome analysis (Fig. 1c,d). Intriguingly, these peaks were not enriched for SOX but rather RUNX, API and NF-κB motifs (Fig. 3c). Moreover, whereas the *Sox9* transgene was induced by W1, *Runx1–Runx3* in particular were highest at W2–W6 (Fig. 1b,c). Given that RUNX1 suppresses basosquamous features in therapeutic-resistant human BCCs³⁰, the sustained *Runx* expression underscored a BCC-like rather than SCC-like phenotype.

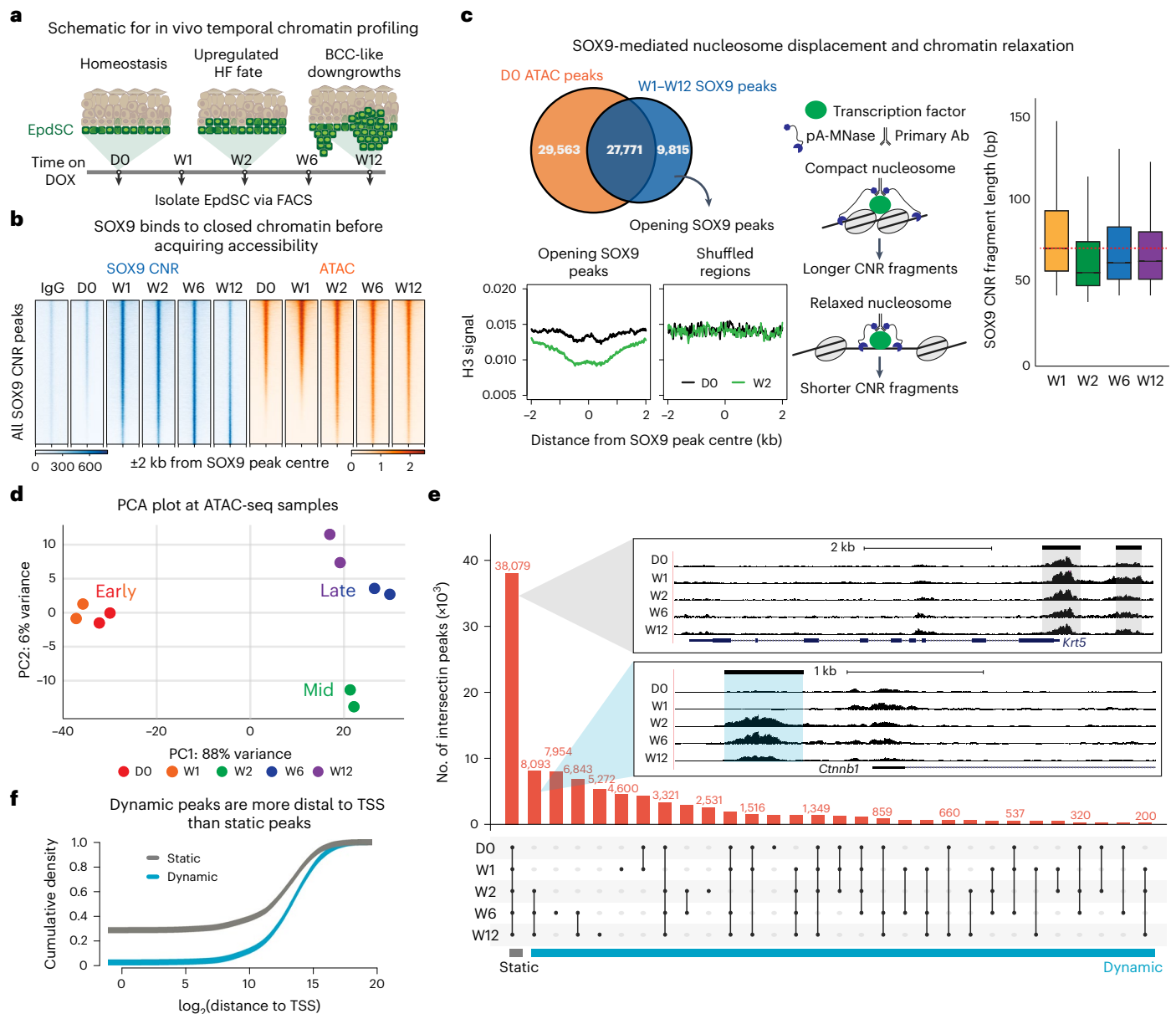


Fig. 2 | Upon induction, SOX9 opens chromatin at enhancers by evicting the nucleosome at its binding site and remodelling the flanking chromatin.

a, Schematic of morphological changes that occur temporally after DOX. Back skins were collected at indicated timepoints and subjected to EpdSC FACS purification followed by SOX9 CNR landscaping and ATAC-seq. **b**, Heat map of IgG control CNR (blue), SOX9 CNR (blue) and ATAC-seq (orange) signals at all SOX9-bound peaks across indicated timepoints. Peaks are arranged along the vertical axis on the basis of their accessibility at W1. Note that, within the cohort of peaks in the bottom half along this axis, SOX9 binding occurred by W1, while their ATAC-seq landscape did not change until the following week. **c**, Top left: Venn diagrams show that, although SOX9 binds to many pre-existing accessible chromatin peaks, 9,815 SOX9 peaks open de novo. MINT-ChIP for histone H3 (H3) coupled with SOX9 CNR reveals that, by W2 post-induction, SOX9 binds to these previously closed peaks, 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. $n = 2$ biological replicates. Box plot is centred at median and bound by first and the third quartile, and whiskers extend to 1.5 times interquartile range (IQR) on both ends. **d**, PCA of ATAC-seq duplicate samples over the five timepoints. **e**, Upset plot of ATAC-seq peaks for each timepoint following SOX9 induction. Shared regions between timepoints are indicated by the dots and connecting lines. 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 timepoints. **f**, 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.

We also performed temporal motif analysis with ChromVAR³¹, which considers both enrichment and chromatin accessibility variability at each motif. In addition to SOX, AP1 (FOS/JUN), GATA and RUNX were top variable motifs. To learn how motif accessibility varied over time, we plotted accessibility deviation scores for each timepoint and compared them with a motif (TBX) that showed no temporal variability.

Agreeing with motif enrichments in C4, the RUNX motifs continued to gain accessibility from W2 to W6 (Fig. 3c and Extended Data Fig. 3f).

Delving deeper, the *Runx1* gene locus was closed at DO, but within W1 after induction, the locus revealed SOX9 binding at multiple sites (Fig. 3d). Since the dynamic peaks were enriched at distal intergenic regions (Fig. 2f), we performed multiplexed T7-indexed chromatin

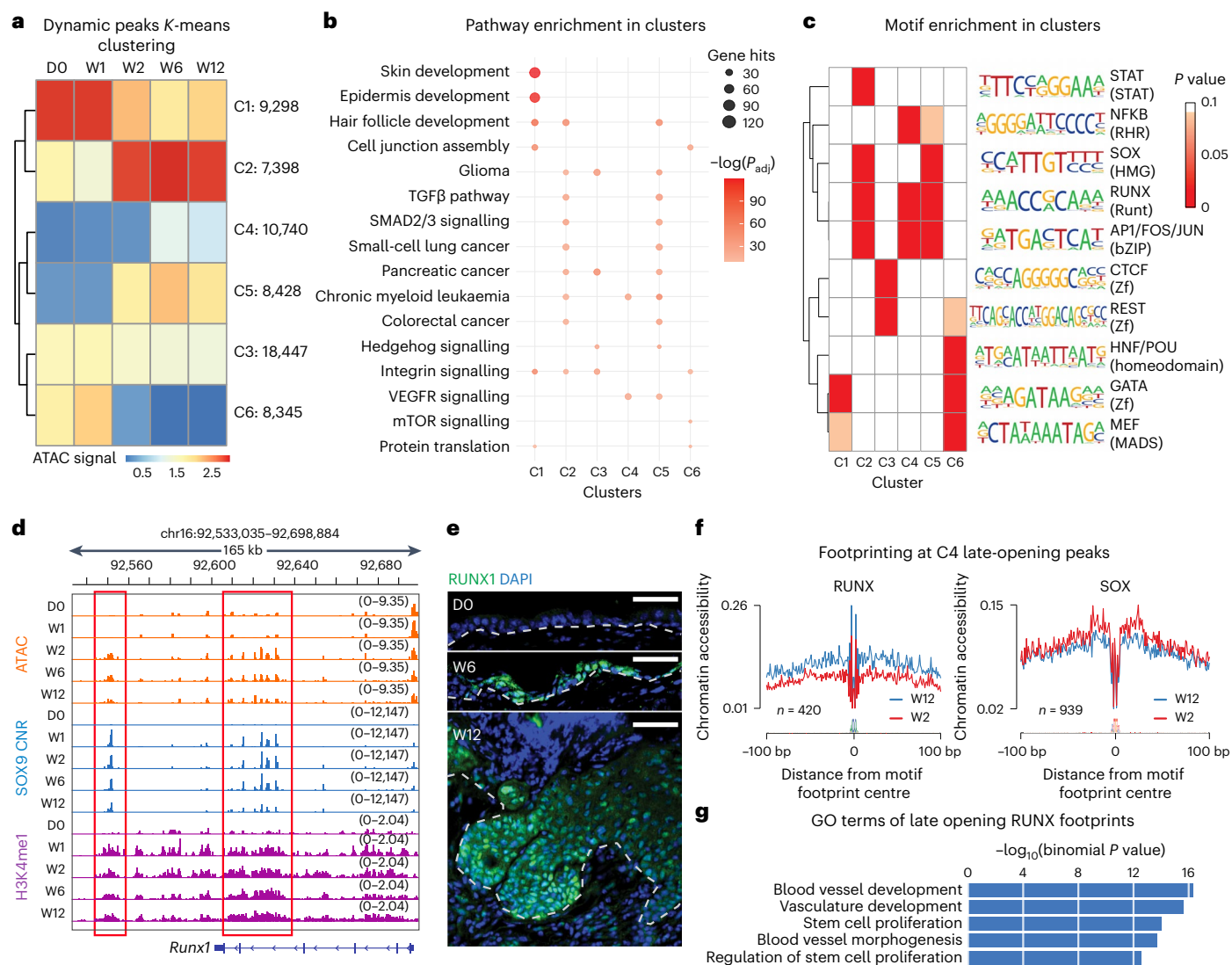


Fig. 3 | SOX9 triggers an activated HFSC fate, while its target gene transcription factors trigger BCC transformation. a, Heat map of K-means clustering of ATAC peaks based on signal across time. Number of peaks in each cluster are indicated on the right. **b**, GO-term analysis of the genes associated with each peak of a cluster. Size of circle reflects the number of gene hits for each pathway, while the shade of red indicates adjusted P value (P_{adj}) from binomial test. Note that cluster C1, whose chromatin is suppressed by W2, is enriched for epidermal genes, while C2, whose chromatin is accessible by W2, is enriched for hair follicle genes. C4, whose chromatin is accessible at later times, is the cluster most enriched for BCC genes. **c**, Motif enrichment analysis of each cluster with HOMER. Shade of red indicates P value with white representing ≥ 0.01 (binomial is used to calculate the significance). Note enrichment for: GATA motif in epidermal genes whose chromatin closes after SOX9 induction; SOX motif in hair follicle genes whose chromatin opens early; and the SOX9 target RUNX1 motif in genes

whose chromatin opens weeks after SOX9 has bound. **d**, 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. **e**, RUNX1 immunofluorescence reveals its absence in epidermal homeostasis, but its presence at late timepoints (W6 and W12) following SOX9 induction. Scale bars, 50 μ m. **f**, RUNX and SOX ATAC footprint analyses at W12 and W2, showing a late increase in chromatin accessibility at the RUNX footprint at a stage when phenotypic BCC-like invaginations are prevalent. By comparison, SOX footprints appear by W2 and are retained thereafter. **g**, GO terms of genes whose putative enhancers have RUNX footprints and open at W6 and W12 (binomial is used to calculate the significance).

immunoprecipitation (MINT-ChIP)³² on enhancer histone modification, H3K4me1, which also showed binding to this locus within W1. By contrast, accessibility did not occur until a week later (Fig. 3d). Immunofluorescence corroborated the delay in activating the *Runx1* locus, and underscored its prominence at later stages of reprogramming (Fig. 3e). Finally, footprint analyses exposed an increase in chromatin accessibility at RUNX footprints over the late-opening C4 ATAC peaks (Fig. 3f and Extended Data Fig. 3g). This contrasted with SOX footprints, which appeared early and then remained constant from W2 to W12 (Fig. 3f).

Gene Ontology (GO)-term analyses of the genes associated with these late-opening RUNX footprints reflected stem cell proliferation, and angiogenesis, hallmarks of cancers (Fig. 3g). Together, these data imply that later changes involved not only SOX9 but also transcription factors that were directly targeted by SOX9, notably of the RUNX family, whose motifs were also enriched as noted above. Although we did not address whether RUNX factors operate as pioneer factors, the enrichment of RUNX motifs coincident with the rise in proliferation during BCC-like downgrowth raised the possibility that proliferation may enhance if not allow accessibility of these factors to chromatin.

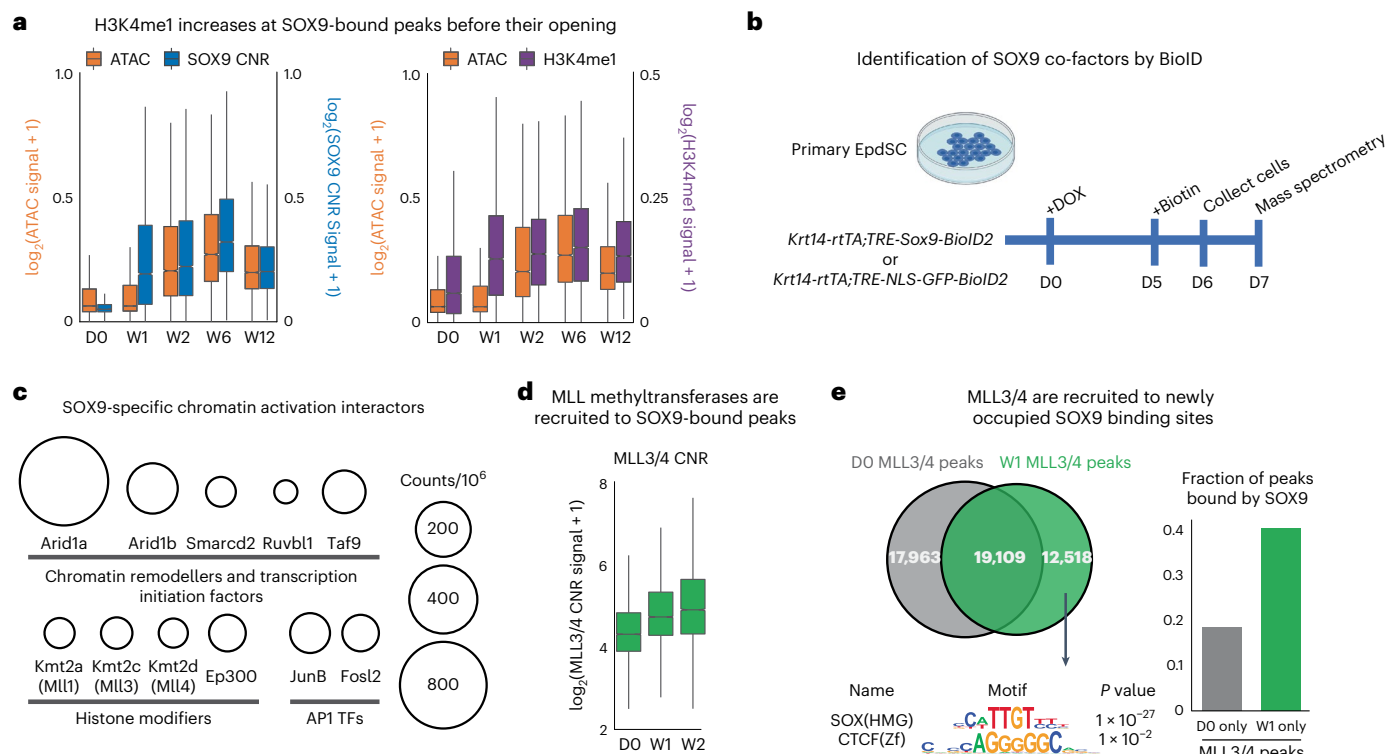


Fig. 4 | SOX9 recruits co-factors to epigenetically prime and then remodel chromatin to an open, transcriptionally accessible state.

a, Left: box plot 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: box plot 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. Box plots are centred at median and bound by first and the third quartile, and whiskers extend to 1.5 times interquartile range (IQR) on both ends. **b**, Schematic of BioID2 proximity labelling of proteins that interact with SOX9 induced in cultured EpdSCs. **c**, Selected SOX9-interacting proteins detected with mass spectrometry and that fall into the top GO terms of

chromatin remodellers of the SWI/SNF family; transcription initiation factors (TAF9), AP1 transcription factors; and enzymes that modify histones (MLL1, MLL3, MLL4 and p300). For full list, see Supplementary Table 3. Circle size corresponds to the strength of the hit as delineated at right. **d**, Box plot of MLL3/4 CNR signals reveals the appearance of MLL3/4 at SOX9-bound peaks following DOX. $n = 2$ biological replicates. Box plot is centred at median and bound by the first and the third quartile, and whiskers extend to 1.5 times IQR on both ends. **e**, 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 $P = 1 \times 10^{-27}$), 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.

SOX9 induces epigenetic remodelling before opening chromatin

The substantial delay between H3K4me1 and SOX9 versus chromatin accessibility and transcription of the *Runx1* gene led us to wonder whether this might be a general phenomenon of SOX9 reprogramming. To address this, we compared ATAC and histone modification signals over time at all opening SOX9 peaks (Fig. 2c). Correlating with SOX9 binding, H3K4me1 deposition occurred within W1 and levelled off thereafter (Fig. 4a), preceding chromatin accessibility changes at W2. By contrast, H3K27ac changes, while appearing by W1, were less robust and, in further contrast, continued to rise over time relative to SOX9 and H3K4me1 (Extended Data Fig. 4b).

Notably, although the nucleosomes directly over SOX9-binding sites appear to have been evicted, H3K4me1 was strongly enhanced on nucleosomes flanking SOX9 (Extended Data Fig. 4c). Moreover, the domain size of H3K4me1 gradually increased from D0 to W2. This did not occur at static peaks, but rather specifically at SOX9-bound opening peaks (Extended Data Fig. 4d).

Activating HFSC enhancers

To understand how SOX9 directly activates HFSC enhancers, we began by identifying SOX9-interacting co-factors. To this end, we transduced *Krt14-rtTA* primary EpdSCs in vitro with *TRE-Sox9-BioID2* and control *TRE-GFP-NLS-BioID2* and then induced expression of each transgene

using DOX (Extended Data Fig. 5a). One week later, biotinylated SOX9-interacting proteins were purified and analysed by mass spectrometry (Fig. 4b).

Biological replicates correlated highly and formed distinct clusters by PCA (Extended Data Fig. 5b–d). Fifty-eight proteins interacted with SOX9 relative to NLS-GFP EpdSCs (Supplementary Table 3). On the basis of protein function and GO-term analysis, SOX9-interacting proteins were mainly DNA and chromatin binders enriched in chromatin modifications and nuclear activity. Among the strongest SOX9 interactions were with core members of the SWI/SNF chromatin remodelling complex (ARID1a/b and SMARCD2), TATA box binding protein TAF9 (TFIID) required for RNA polymerase II-mediated induction of transcription, and AP1 (FOSL2 and JUNB) (Fig. 4c and Extended Data Fig. 5e, f). Histone modifiers typifying key active enhancers in developmental contexts were also featured. As SOX9-induced opening peaks were more enriched at enhancers over promoters (Extended Data Fig. 3e), we were intrigued to find modifiers of two histone marks enriched at active enhancers: Ep300, the acetyltransferase for H3K27ac, and MLL3/MLL4, histone methyltransferases that not only can deposit H3K4me1 but possibly play additional emerging roles in enhancer activation^{33–35}.

Since we observed an increase in H3K4me1 at SOX9 targeted enhancers before H3K27ac or chromatin opening, we first focused on whether, as predicted, MLL3/4 are recruited by SOX9 to closed

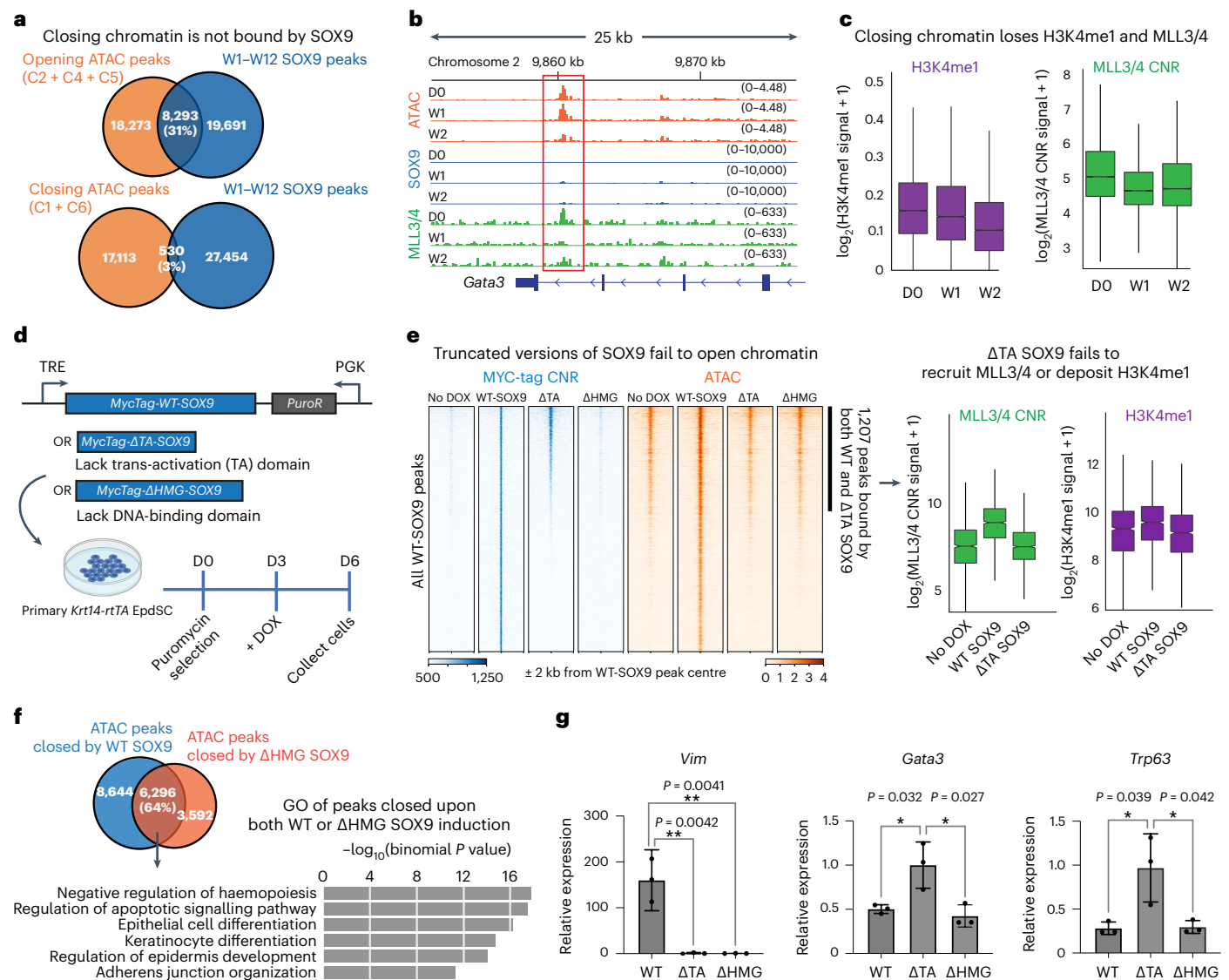


Fig. 5 | SOX9 achieves EpdSC fate silencing independent from DNA binding.

a, Top: Venn diagram shows robust overlap between opening peak clusters (C2 + C4 + C5) and SOX9 peaks. Bottom: Venn diagram shows only 3% overlap between closing peaks (C1 + C6) and SOX9 peaks. **b**, ATAC, SOX9 CNR and MLL3/4 CNR tracks at the *Gata3* locus, showing that, by W1 after SOX9 induction, MLL3/4 CNR peaks were diminished, and by W2, ATAC peaks closed, even though CNR showed no SOX9 binding in this region (red box). **c**, Box plots showing loss of MLL3/4 and H3K4me1 signal beginning at W1 post SOX9-induction and specifically at ATAC peaks that close by W2 (C1, C6). $n = 2$ biological replicates. **d**, Schematic of inducing of MYC-tagged wild-type (WT) or mutant versions of SOX9 in transduced *Krt14-rtTA* cultured EpdSCs. **e**, Profile plot and heat map showing MYC-tagged wild-type or variant SOX9 binding (blue) and accessibility (orange) before and after DOX. Peaks are sorted the same way across samples. Note that Δ HMG-SOX9 fails to bind DNA, and Δ TA-SOX9 binds only to the subset

of SOX9 peaks that were already accessible before DOX. Both mutants of SOX9 failed to open chromatin de novo. Right: box plot comparing MLL3/4 CNR and H3K4me1 signals at the peaks that are bound by wild-type SOX9 and Δ TA-SOX9. Note that only wild-type SOX9 brought additional MLL3/4 and deposited more H3K4me1 to these peaks. $n = 2$ biological replicates. **f**, Venn diagram shows overlap between the ATAC peaks that are closed by wild-type SOX9 or Δ HMG-SOX9 induction. GO terms reveal that epidermal enhancers close upon wild-type or Δ HMG-SOX9 induction (binomial is used to calculate the significance). **g**, Quantitative PCR 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 \pm s.d. * $P < 0.05$ and ** $P < 0.01$, two-tailed *t*-test. $n = 3$ biological replicates. All box plots are centred at median and bound by the first and third quartile, and whiskers extend to 1.5 times interquartile range (IQR) on both ends.

chromatin in vivo. To validate the physical interaction, we exploited the MYC tag of SOX9 and performed co-immunoprecipitations on cultured EpdSC lysates with or without SOX9 induction, and then probed for MLL4. Given the large size of MLL4 (>500 kDa) and the likelihood of degradation, we used CRISPR-Cas9 to ablate *Mll4* in these EpdSCs to ensure correct band identification (Extended Data Fig. 5g).

After validation, we performed MLL3/4 CNR, reasoning that, if MLL3/4 recruitment to chromatin is regulated by SOX9, de novo MLL3/4 targeted sites should be enriched with SOX9 binding. A marked increase

in MLL3/4 association with chromatin occurred between D0 and W2 at opening SOX9-bound enhancers (Fig. 4d). Moreover, upon analysing de novo MLL3/4 recruitment sites on chromatin at W1, we found that SOX motifs were significantly enriched (Fig. 4e). These data began to provide a clearer picture of how SOX9 functions as a pioneer factor, as it not only binds to closed chromatin but also recruits co-factors to epigenetically modify flanking histones. The data from Fig. 4c further hinted that SOX9 recruits the SWI/SNF complex to make the chromatin accessible for transcription.

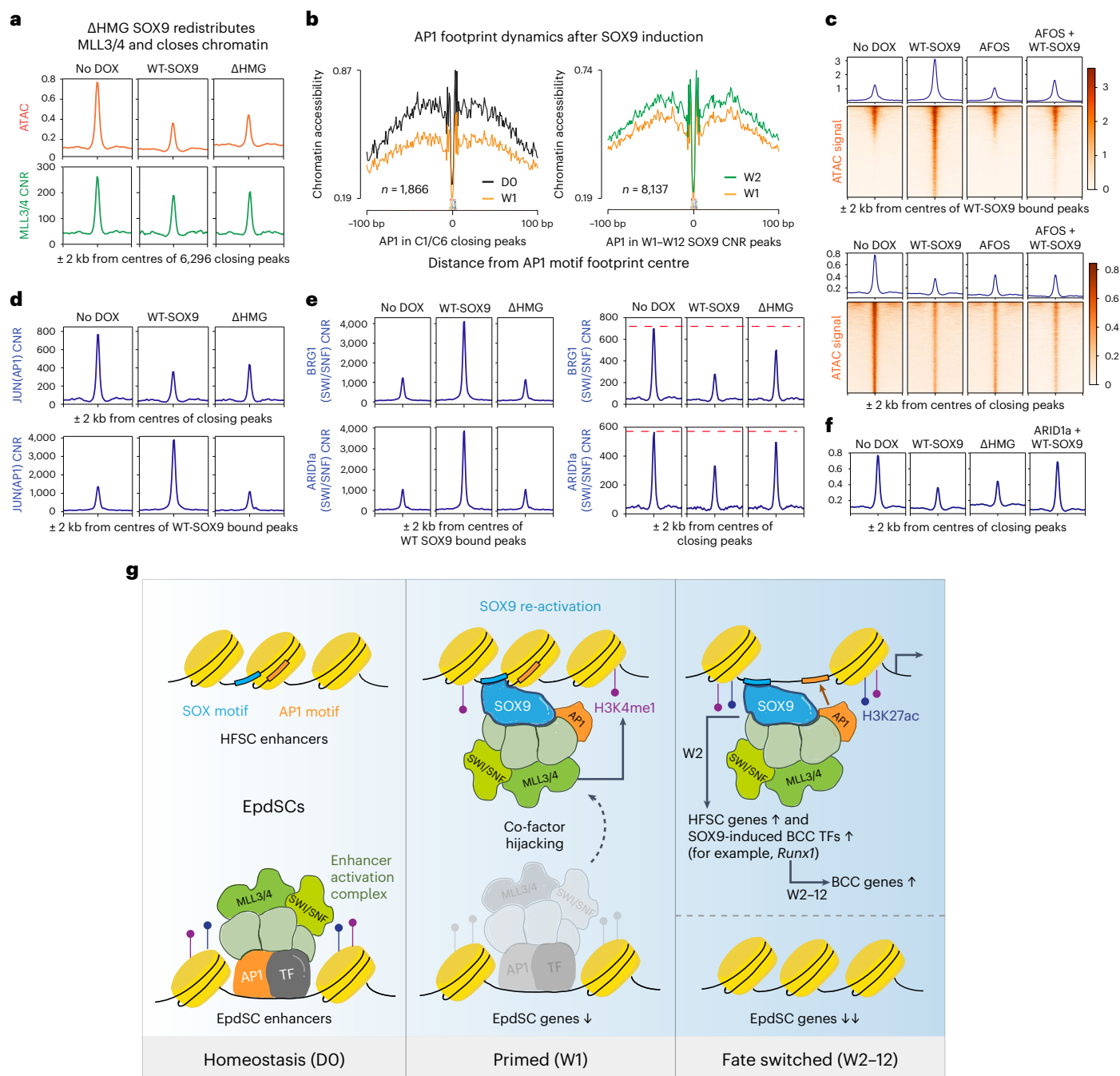


Fig. 6 | SOX9 redistributes chromatin remodelling co-factors to activate HFSC enhancers and silence EpdSC enhancers. **a**, Profile plots showing that ATAC and MLL3/4 CNR signals wane at EpdSC enhancers upon wild-type SOX9 and Δ HMG-SOX9 induction. Note that, when SOX9 cannot bind to DNA (Δ HMG), it still diminishes MLL3/4 at endogenous EpdSC enhancers and closes their chromatin. **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 (see also Fig. 5a). **c**, Top: profile plots and heat maps comparing ATAC signals at wild-type SOX9 (WT-SOX9) bound peaks in vitro. While WT-SOX9 can open chromatin, dominant negative FOS (AFOS) inhibits the opening as shown in the last column. Bottom: profile plots and heat maps comparing ATAC signals at closing peaks in vitro. AFOS phenocopies the indirect closing effect of SOX9 on API-associated epidermal enhancers. **d**, Top: profile plots showing that JUN CNR signals are reduced at EpdSC enhancers upon wild-type SOX9 and Δ HMG-SOX9 induction. Bottom: supportive of competition, wild-type SOX9 recruits API to HFSC enhancers, while Δ HMG-SOX9, lacking the DNA binding domain, fails

to do so. **e**, Profile plots showing that both BRG1 and ARID1a of the SWI/SNF complex behave similarly to API and MLL3/4: they are recruited to the opening HFSC enhancers by only wild-type SOX9 (left), while both wild-type SOX9 and Δ HMG-SOX9 reduce SWI/SNF association with epidermal enhancers (right). The red dotted line denotes CNR levels of indicated target at closing peaks before DOX. **f**, Profile plots comparing ATAC signals at closing epidermal enhancers with wild-type SOX9, Δ HMG-SOX9 or ARID1a overexpression together with wild-type SOX9. Note that, when ARID1a is overexpressed, SOX9 fails to close epidermal enhancers. **g**, Working model for how SOX9 achieves cell fate switching: pioneer factor SOX9 indirectly silences the epidermal fate by competing away co-factors and other transcription factor (TFs) including API from active EpdSC enhancers. Concomitantly, SOX9 binds directly to key hair follicle enhancers, bringing with it the hijacked chromatin remodelling machinery and activating the hair follicle fate. Also among the SOX9 target genes are transcription factors such as 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-to-HFSC fate transition.

Silencing the epidermal fate

Interestingly, while SOX9 binding was highly enriched within peaks that opened over time, it accounted for only 3% of peaks that closed over time (Fig. 5a). The differences were even more striking when we restricted our analysis to ATAC peaks changing over the first two weeks (Extended Data Fig. 6a). These findings were consistent with our SOX9 interactome, which was dominated by chromatin-activating remodellers. Moreover, in contrast to the hair-follicle-associated GO terms prominent in W2 SOX9-bound opening peaks, Epd-associated GO-terms were featured among closing peaks that were not bound by SOX9 (Extended Data Fig. 6b). We therefore hypothesized that SOX9 silences epidermal fate indirectly.

To further understand how epidermal fate is silenced, we were intrigued by GATA factors, whose motif was markedly enriched in ATAC peaks (C1, C6) that closed within the first 2 weeks after SOX9 induction and whose transcription factor footprint declined upon SOX9 induction (Extended Data Fig. 6c). GATAs surfaced upon analysing the transcription factors expressed by EpdSCs and whose motifs are highly enriched in closing chromatin (Extended Data Fig. 6d). GATA3 transcript and protein expression also declined concomitantly with the closure of GATA motifs (Extended Data Fig. 6e).

The *Gata3* gene locus also lost chromatin accessibility by W2, but the decline happened only at non-SOX9-bound peaks. The nearest SOX9-bound enhancer was >30 kb from the *Gata3* gene body, and like several other weaker peaks, this site was already open and MLL3/4-bound before SOX9 was induced. Notably, subsequent SOX9 binding had little or no effect on its status (Extended Data Fig. 6f). These findings suggest that the role of SOX9 in silencing epidermal fate is at least in part indirect. Moreover, the result appeared to be physiologically relevant as genes downregulated in SOX9⁺ embryonic skin progenitors were also silenced when SOX9 was induced in adult EpdSCs (Fig. 1d)³⁶.

MLL3/4 presence over opening SOX9-dependent enhancer peaks was robust by W2, as was H3K4me1 modification (Fig. 4a,d). By contrast, the >6,000 SOX9-independent enhancer peaks that closed during this time displayed plummeting MLL3/4 association and a more gradual loss of H3K4me1 (Extended Data Fig. 7a). These findings raised the tantalizing possibility that, in binding to nucleosomes at HFSC-enhancers, SOX9 might be recruiting co-factors including MLL3/4 away from active EpdSC enhancers.

To test this hypothesis, we engineered DOX-inducible MYC-tagged wild-type and mutant forms of SOX9 that lacked either the transactivation (Δ TA) domain or the DNA binding (Δ HMG) domain (Fig. 5d and Extended Data Fig. 7b). In the transduced primary EpdSCs, immunofluorescence levels of three versions of SOX9 were comparable, and the ectopically expressed proteins were of the expected size (Extended Data Fig. 7c,d). Additionally, as judged by co-immunoprecipitation, only wild-type SOX9 and Δ HMG-SOX9, but not Δ TA-SOX9, associated with MLL4, consistent with the inability of Δ TA to interact with chromatin remodellers (Extended Data Fig. 7e).

By using CNR with a MYC-tag antibody recognizing all three SOX9 variants equivalently, we verified that wild-type SOX9 and the Δ TA-SOX9 mutant, but not Δ HMG-SOX9, bound to DNA (Fig. 5e). Interestingly, without the TA domain to interact with co-factors, Δ TA-SOX9 only bound to chromatin that was already accessible in EpdSCs. Consistent with this result, the 1,207 peaks that were open before DOX and bound by Δ TA-SOX9 did not show MLL3/4 recruitment nor did they show H3K4me1 modification (shown at right). Additionally, and in contrast to wild-type SOX9, Δ TA-SOX9 failed to stably bind to closed chromatin of HFSC enhancers, indicating that, without binding to co-factors, SOX9 lost the defining feature of pioneer factors.

Although Δ HMG-SOX9 did not bind DNA, it had a striking effect on chromatin accessibility. Nearly 10,000 ATAC peaks closed and >8,000 peaks opened upon induction (Extended Data Fig. 7f). As this mutant was unable to bind DNA, it was not surprising to see that the GO-term

profile of the opening peaks was dramatically different than that of wild-type SOX9 (Extended Data Fig. 7g). Rather than HFSC features, the changes were more reflective of a stressed state. By contrast, in the ATAC peaks that closed in response to Δ HMG-SOX9, 64% of them were also closed by wild-type SOX9, and the GO terms corresponded to the same EpdSC genes indirectly silenced by wild-type SOX9 (Fig. 5f,g).

Competition for SOX9-interacting chromatin remodellers

Consistent with the hypothesis that SOX9 closes chromatin by competing for and redistributing co-factors, MLL3/4 CNR signal diminished over EpdSC enhancers upon Δ HMG-SOX9 induction (Fig. 6a). Probing deeper, we turned to AP1 transcription factors, which surfaced in our SOX9 interactome. In agreement with the dynamics observed for MLL3/4, footprint analysis in vivo revealed that AP1 binding decreased in closing non-SOX9-bound epidermal enhancers and increased in SOX9-bound chromatin (Fig. 6b). Moreover, in these SOX9-bound opening peaks, SOX and AP1 motifs were mostly found within one-nucleosome distance, supporting a role for SOX9 in targeting AP1 transcription factors to their canonical binding sites upon opening hair follicle enhancers (Extended Data Fig. 7h).

Notably, motif analyses revealed the presence of AP1-binding sites in both closing and opening enhancers (Extended Data Fig. 3f), suggesting that the interaction between SOX9 and AP1 may be functionally important for both opening SOX9⁺ HFSC enhancers and closing SOX9^{neg} EpdSC enhancers. To test the possibility that enhancers might be competing for AP1 binding, we used the strategy delineated in Fig. 5d to induce AFOS, a dominant negative version of c-FOS that can heterodimerize with AP1 transcription factors and block their binding to DNA^{37,38}. We performed these experiments in the presence and absence of wild-type SOX9, and then carried out ATAC-seq. In the peaks that were bound by wild-type SOX9, AFOS clearly interfered with the opening of the HFSC enhancers, while also phenocopying the closing effects of SOX9 at EpdSC enhancers when expressed alone (Fig. 6c).

Our data thus far suggested that, like MLL3/4, AP1 transcription factors function on both sides of the fate coin. To test whether other members of the interactome are targets for this putative competition for SOX9-binding partners, we focused on AP1 transcription factors and the SWI/SNF complex. After first validating their association with SOX9 (Extended Data Fig. 7i,j), we performed CNR analysis. We found that induction of wild-type SOX9 resulted in increased JUN(AP1) binding at SOX9-bound peaks, and decreased JUN binding at closing epidermal peaks. Notably, while Δ HMG-SOX9 failed to recruit JUN and open hair follicle enhancers, it still diminished JUN binding at closing epidermal peaks (Fig. 6d). Similarly, when we performed CNR on both structural (ARID1a) and enzymatic (BRG1) members of the SWI/SNF complex, we observed a decline in their association with epidermal enhancers when either wild-type SOX9 or Δ HMG-SOX9 were induced, but an increased association with SOX9-bound peaks only after wild-type SOX9 and not Δ HMG-SOX9 induction (Fig. 6e).

Together, these data suggest that SOX9 orchestrates the redistribution of transcription factors and epigenetic co-factors that are shared by the enhancers of both cell fates. Moreover, this competition appeared to be predicated in part on limiting levels of chromatin remodelling factors, as when we overexpressed ARID1a in the presence of SOX9, epidermal enhancers were rescued from closing (Fig. 6f and Extended Data Fig. 7i,j).

Discussion

Elegant studies by the Zaret lab launched the field of pioneer factors, now examined in various fate-switching scenarios and distinguished by their ability to bind their sequence motifs within closed chromatin^{1,2}. However, the precise sequence of nucleosome eviction, opening of surrounding chromatin, and reprogramming fate choices has been difficult to unravel, particularly in in vitro settings, where fate choices lack constraints imposed by native tissue microenvironments. By exploiting the slowed kinetics of our in vivo reprogramming system, we discovered

that SOX9 not only perturbs its target nucleosome but also recruits enzymes that modify the flanking enhancer nucleosomes. Like SOX9 binding itself, these features precede the subsequent chromatin opening to the transcriptional machinery. As our Δ TA-SOX9 studies imply, these dynamics appear to be achieved by SOX9 recruiting of chromatin remodelling factors such as AP1 TFs and the SWI/SNF complex^{39,40}.

It has generally been viewed that a pioneer factor can act either as a transcriptional activator or as a repressor through recruiting different cohorts of co-activators or co-repressors^{1,2}. At first glance, this notion seems well suited to nodes of lineage switching, where one fate is silenced while another is chosen. However, increasing evidence suggests that pioneer factors may bind and directly regulate the enhancers of only one lineage at the crossroads, leaving a conundrum as to how the other lineage becomes silenced to achieve the switch.

Our findings showed clearly that EpdSC gene silencing occurs shortly after SOX9 induction, a timing that is at odds with the notion that SOX9 might induce transcriptional repressors that then subsequently silence epidermal genes. Moreover, in contrast to HFSC enhancers, many of which bind SOX9 and are opened de novo, EpdSC enhancers show a paucity of SOX9 binding and yet close rapidly upon SOX9 induction.

Rather to prevailing notions, our findings favour a dual function model whereby a pioneer factor actively hijacks and redistributes shared co-factors to achieve cost-effective and coordinated fate switching from one lineage to another (Fig. 6g). Thus, following SOX9 induction in EpdSCs, MLL3/4 binding increased at SOX9-bound opening HFSC enhancers, while diminished at closing non-SOX9-bound EpdSC enhancers. Our studies with wild-type SOX9 and Δ HMG-SOX9 revealed that not only does SOX9 interact with MLL3/4, but also with a compendium of co-factors essential to activate enhancers, which include not only MLL3/4 but also AP1 and SWI/SNF complex.

In closing, although direct repressive mechanisms independent from chromatin accessibility are still formally possible, our data suggest that at least some chromatin remodellers that are generally required for enhancer activity are in short supply, thereby setting up the competition to achieve fate switching once a pioneer factor such as SOX9 is activated. By utilizing such a mechanism, cellular fate plasticity is minimized, while simultaneously expediting the shift in density of shared transcriptional regulators to genomic loci of new fate determinants.

Finally, it is noteworthy that, to make tissue, stem cells must undergo a fate choice, which for SOX9⁺HFSCs, is achieved by down-regulating SOX9 (ref. 16). In our model as in BCC, SOX9 was constitutive and hence the choice to make hair was never made. Moreover, when left outside the instructive microenvironment of the quiescent hair follicle bulge niche, the proliferating cells with sustained SOX9 activated SOX9 downstream target transcription factor genes, such as those encoding the RUNX family, that secondarily drove further dynamic changes in the chromatin landscape. These findings begin to explain how and why in adult tissue stem cells, sustained re-activation of a pioneer factor involved in embryonic fate decisions frequently leads to cancer^{5,6,41}.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41556-023-01184-y>.

References

- Zaret, K. S. Pioneer transcription factors initiating gene network changes. *Annu. Rev. Genet.* **54**, 367–385 (2020).
- Balsalobre, A. & Drouin, J. Pioneer factors as master regulators of the epigenome and cell fate. *Nat. Rev. Mol. Cell Biol.* **23**, 449–464 (2022).
- Soufi, A. et al. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555–568 (2015).
- Fuglerud, B. M. et al. SOX9 reprograms endothelial cells by altering the chromatin landscape. *Nucleic Acids Res.* **50**, 8547–8565 (2022).
- Grimm, D. et al. The role of SOX family members in solid tumours and metastasis. *Semin. Cancer Biol.* **67**, 122–153 (2020).
- Julian, L. M., McDonald, A. C. & Stanford, W. L. Direct reprogramming with SOX factors: masters of cell fate. *Curr. Opin. Genet. Dev.* **46**, 24–36 (2017).
- Kamachi, Y. & Kondoh, H. Sox proteins: regulators of cell fate specification and differentiation. *Development* **140**, 4129–4144 (2013).
- Nowak, J. A., Polak, L., Pasolli, H. A. & Fuchs, E. Hair follicle stem cells are specified and function in early skin morphogenesis. *Cell Stem Cell* **3**, 33–43 (2008).
- Kadaja, M. et al. SOX9: a stem cell transcriptional regulator of secreted niche signaling factors. *Genes Dev.* **28**, 328–341 (2014).
- Ouspenskaia, T., Matos, I., Mertz, A. F., Fiore, V. F. & Fuchs, E. WNT-SHH antagonism specifies and expands stem cells prior to niche formation. *Cell* **164**, 156–169 (2016).
- Oro, A. E. et al. Basal cell carcinomas in mice overexpressing sonic hedgehog. *Science* **276**, 817–821 (1997).
- Vidal, V. P. et al. Sox9 is essential for outer root sheath differentiation and the formation of the hair stem cell compartment. *Curr. Biol.* **15**, 1340–1351 (2005).
- Youssef, K. K. et al. Identification of the cell lineage at the origin of basal cell carcinoma. *Nat. Cell Biol.* **12**, 299–305 (2010).
- Larsimont, J. C. et al. Sox9 controls self-renewal of oncogene targeted cells and links tumor initiation and invasion. *Cell Stem Cell* **17**, 60–73 (2015).
- Nguyen, H., Rendl, M. & Fuchs, E. Tcf3 governs stem cell features and represses cell fate determination in skin. *Cell* **127**, 171–183 (2006).
- Yang, H., Adam, R. C., Ge, Y., Hua, Z. L. & Fuchs, E. Epithelial–mesenchymal micro-niches govern stem cell lineage choices. *Cell* **169**, 483–496 e413 (2017).
- Fiore, V. F. et al. Mechanics of a multilayer epithelium instruct tumour architecture and function. *Nature* **585**, 433–439 (2020).
- Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006–1019 (2018).
- Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
- Corces, M. R. et al. The chromatin accessibility landscape of primary human cancers. *Science* <https://doi.org/10.1126/science.aav1898> (2018).
- Meers, M. P., Janssens, D. H. & Henikoff, S. Pioneer factor–nucleosome binding events during differentiation are motif encoded. *Mol. Cell* **75**, 562–575 e565 (2019).
- Oro, A. E. & Scott, M. P. Splitting hairs: dissecting roles of signaling systems in epidermal development. *Cell* **95**, 575–578 (1998).
- Hsu, Y. C., Li, L. & Fuchs, E. Transit-amplifying cells orchestrate stem cell activity and tissue regeneration. *Cell* **157**, 935–949 (2014).
- Oshimori, N. & Fuchs, E. Paracrine TGF- β signaling counterbalances BMP-mediated repression in hair follicle stem cell activation. *Cell Stem Cell* **10**, 63–75 (2012).

26. Kuonen, F. et al. c-FOS drives reversible basal to squamous cell carcinoma transition. *Cell Rep.* **37**, 109774 (2021).
27. Wang, L. et al. Oncogenic role of SOX9 expression in human malignant glioma. *Med. Oncol.* **29**, 3484–3490 (2012).
28. Zhou, C. H. et al. Clinical significance of SOX9 in human non-small cell lung cancer progression and overall patient survival. *J. Exp. Clin. Cancer Res.* **31**, 18 (2012).
29. Matheu, A. et al. Oncogenicity of the developmental transcription factor Sox9. *Cancer Res.* **72**, 1301–1315 (2012).
30. Haensel, D. et al. LY6D marks pre-existing resistant basosquamous tumor subpopulations. *Nat. Commun.* **13**, 7520 (2022).
31. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
32. van Galen, P. et al. A multiplexed system for quantitative comparisons of chromatin landscapes. *Mol. Cell* **61**, 170–180 (2016).
33. Sze, C. C. & Shilatifard, A. MLL3/MLL4/COMPASS family on epigenetic regulation of enhancer function and cancer. *Cold Spring Harb. Perspect. Med.* <https://doi.org/10.1101/cshperspect.a026427> (2016).
34. Dorigi, K. M. et al. Mll3 and Mll4 facilitate enhancer RNA synthesis and transcription from promoters independently of H3K4 monomethylation. *Mol. Cell* **66**, 568–576 e564 (2017).
35. Rao, R. C. & Dou, Y. L. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat. Rev. Cancer* **15**, 334–346 (2015).
36. Nowak, J. A., Polak, L., Pasolli, H. A. & Fuchs, E. Hair follicle stem cells are specified and function in early skin morphogenesis. *Cell Stem Cell* **3**, 33–43 (2008).
37. Ahn, S. et al. 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 (1998).
38. Olive, M. et al. A dominant negative to activation protein-1 (AP1) that abolishes DNA binding and inhibits oncogenesis. *J. Biol. Chem.* **272**, 18586–18594 (1997).
39. Vierbuchen, T. et al. AP-1 transcription factors and the BAF complex mediate signal-dependent enhancer selection. *Mol. Cell* **68**, 1067–1082 e1012 (2017).
40. Wolf, B. K. et al. Cooperation of chromatin remodeling SWI/SNF complex and pioneer factor AP-1 shapes 3D enhancer landscapes. *Nat. Struct. Mol. Biol.* <https://doi.org/10.1038/s41594-022-00880-x> (2022).
41. Kumar, P. & Mistri, T. K. Transcription factors in SOX family: potent regulators for cancer initiation and development in the human body. *Semin. Cancer Biol.* **67**, 105–113 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Methods

Ethical regulation compliance

All animals used in this study were maintained and bred under specific-pathogen-free conditions at the Comparative Bioscience Center at The Rockefeller University, which is an Association for Assessment and Accreditation of Laboratory Animal Care-accredited facility. All procedures were performed with the Institutional Animal Care and Use Committee-approved protocols (20012-H and 20066-H).

Generating and handling *TRE-Sox9* mice

To generate the conditional SOX9 transgenic mice, the *Sox9* coding sequence was cloned into the pTRE2 vector harbouring a DOX-inducible, minimal CMV2 promoter. A MYC tag was added to the N terminus of SOX9. Transgenic mice were generated as described previously⁴². The resulting *TRE-Sox9* mice were then genotyped and crossed to *Krt14-rtTA* transgenic mice¹⁵ to allow for DOX-inducible expression of MYC–SOX9 specifically in skin epithelium.

Primary cell isolation

Primary *Krt14-rtTA*; *TRE-Sox9* EpdSCs were isolated from newborn male pups (postnatal day 0, or P0) as described previously^{16,43}. Briefly, mouse back skin was collected 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 medium⁴⁴ (0.05 mM CaCl₂) at 37 °C with 7.5% CO₂.

DOX treatment

A total of 0.1 mg of DOX (Sigma) in 100 µl phosphate-buffered saline (PBS) was administered by intraperitoneal injection to *Krt14-rtTA*; *TRE-Sox9* and *Krt14-rtTA*-only mice at postnatal day P21, and the mice were thereafter maintained on mouse chow containing 2 mg g⁻¹ DOX throughout the experimental time course. Phenotypic mice were housed with at least one control littermate for adequate grooming. To maintain proper body fluid, 100 µl PBS was administered through intraperitoneal injection every other day after 4 weeks of SOX9 induction. For W12 samples, epidermis from the back skin of P0 *Krt14-rtTA*; *TRE-Sox9* or *Krt14-rtTA*-only pups were grafted onto 6–8-week-old immunocompromised (*Nude*) female mice. Grafts were allowed to heal for 21 days, and DOX was administered as above. For induction of SOX9 and its variants, AFOS and ARID1a in cultured cells, DOX was added to a final concentration of 1 µg ml⁻¹ in E-low medium for BioID or SOX9 variant experiments.

Immunofluorescence

Mouse back skin was fixed in 4% paraformaldehyde at room temperature for 15 min, and then washed three times with PBS for 15 min at 4 °C. Following PBS washes, samples were dehydrated in 30% sucrose in PBS 4 °C overnight. The dehydrated samples were then embedded in optimal cutting temperature (OCT) medium (VWR) and frozen on dry ice. Cryosections (16 µm) were blocked in immunofluorescence buffer containing 0.3% Triton X-100, 2.5% normal donkey serum, 2.5% normal goat serum, 1% bovine serum albumin and 1% gelatin in PBS for 1 h at room temperature. After blocking, the sections were stained with primary antibodies in immunofluorescence buffer at 4 °C overnight: MYC-tag (rabbit, 1:1,000, Cell Signaling), SOX9 (rabbit, 1:5,000, Millipore), ITGA6 (rat, 1:1,000, BD), KRT14 (chicken, 1:1,000, BioLegend), KRT10 (rabbit, 1:250, Fuchs Lab), EpCAM (rabbit, 1:100, Abcam), KRT6 (guinea pig, 1:1,000, Fuchs Lab), RUNX1 (rabbit, 1:100, Abcam), and GATA3 (rat, 1:100, Invitrogen). After primary antibody staining, all sections were washed three times with immunofluorescence buffer containing 0.1% Triton X-100 in PBS for 5 min at room temperature. Sections were then stained with Alexa 488, 546 or 647 conjugated secondary donkey antibodies (1:500, Thermo Fisher), mounted with Prolong Diamond anti-fade mounting medium with 4',6'-diamidino-2-phenylindole (DAPI, Thermo Fisher) and imaged with Zeiss Axio Observer Z1 with Apotome 2 microscope.

Images were collected and analysed with Fiji (ImageJ v.2.3.0). For the Human Atlas immunostaining, the following antibodies were used: SOX9 (CAB068240), EpCAM (CAB030012) and KRT6A (HPA061168).

For cultured cells, cells were plated onto chamber slides (Thermo Fisher). At collection, cells were fixed with 4% paraformaldehyde for 10 min, and then washed three 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, 1:1,000, Cell Signaling), GFP (chicken, 1:2,000, Fuchs Lab), RFP (rat, 1:1,000, ChromoTek), and MYC-tag (rabbit, 1:1,000, Cell Signaling).

For 5'-ethynyl-2'-deoxyuridine (EdU) experiments, mice were injected IP with EdU (50 µg g⁻¹ body weight) 2 h 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 (1:5,000), 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

Krt14-rtTA; *TRE-Sox9* and *Krt14-rtTA*-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/ethylenediaminetetraacetic acid (EDTA) (Gibco) for 45–60 min at 37 °C. After quenching the trypsin with cold FACS buffer (5% foetal bovine serum, 10 mM EDTA and 1 mM HEPES in PBS), the epidermal layer and HF were scraped off the epidermal side of the skin. The tissues were mechanically separated and filtered through a 70 µm cell strainer (BD) into a single-cell suspension for immunolabelling. Single-cell suspensions were immunolabelled with antibodies: Ly6A/E-APCCy7 (1:500, BioLegend), CD49f-PECy7 (1:1,000, 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), TruStain FcX for blocking (1:1,000, BioLegend) and streptavidin-FITC (1:1,000, BioLegend) in 300 µl of FACS buffer. Stained cells were washed and resuspended with FACS buffer with 100 ng ml⁻¹ DAPI before analysis or sorting. EpdSCs were collected using an Aria Cell Sorters (BD Biosciences) with BD FACSDiva (v. 8.0) 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 transcripts per million (TPM) 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 (DEG) gene list was generated with DESeq2 (v.1.16.1).

ATAC-seq and raw file processing

ATAC-seq²⁰ was performed on FACS-purified EpdSCs (two to four male mice per replicate) at indicated timepoints (D0, W1, W2, W6 and W12) and cultured keratinocytes. Briefly, cells were lysed in ATAC lysis buffer for 5 min and then transposed with Tn5 transposase (Illumina) for 30 min. Samples were barcoded and sequencing libraries were prepared according to the manufacturer's guidelines (Illumina) and sequenced on an Illumina NextSeq. For sequencing analysis, 50 bp paired-end FASTQs were aligned to the mouse genome (GRCm38/mm10) using the PEPATAC (v.0.10.3) pipeline⁴⁵. 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, we normalized peak calling as performed as described²¹ with a threshold of 3, and we quantified reads in filtered peaks (RIP) for generating normalized bigwig files. To do so, 1,000,000/RIP was used as input for Deeptools ‘bam-coverage’ with the ‘-scaleFactor’ option. Shared peaks were defined as regions that had ≥ 1 base pair overlap between two timepoints as shown in Fig. 2e. Dynamic peaks were defined as those accessible chromatin regions that were absent from at least one timepoint. 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.

CNR and raw file processing

EpdSCs were FACS purified, and the CNR sequencing was performed as previously described^{19,46} with minor modifications indicated below. Briefly, 500,000–1,000,000 EpdSCs were washed with ice-cold PBS, resuspended in crosslinking buffer (10 mM HEPES–NaOH pH 7.5, 100 mM NaCl, 1 mM egtazic acid (EGTA), 1 mM EDTA and 1% formaldehyde) and rotated at room temperature for 10 min. Crosslinked cells were quenched with glycine at a final concentration of 0.125 M for 5 min at room temperature. Cells were washed with cold $1\times$ PBS and resuspended in NE1 buffer (20 mM HEPES–KOH pH 7.9, 10 mM KCl, 1 mM MgCl₂, 1 mM dithiothreitol, 0.1% Triton X-100 supplemented with Roche complete protease inhibitor EDTA-free) and rotated for 10 min at 4 °C. Nuclei were washed twice with CNR wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5% bovine serum albumin and 0.5 mM spermidine supplemented with protease inhibitor) and incubated with concanavalin-A (ConA) beads washed with CNR binding buffer (20 mM HEPES–KOH pH 7.9, 10 mM KCl, 1 mM CaCl₂ and 1 mM MnCl₂) for 10 min at 4 °C. ConA-bead-bound nuclei were incubated overnight at 4 °C in CNR antibody buffer (CNR wash buffer supplemented with 0.1% Triton X-100 and 2 mM EDTA) and antibody. After antibody incubation, ConA-bead-bound nuclei were washed once with CNR Triton wash buffer (CUT&RUN wash buffer supplemented with 0.1% Triton X-100) then resuspended and incubated at 4 °C for 1 h in CUT&RUN antibody buffer and 2.5 μ l pAG-MNase (EpiCypher). ConA-bound-nuclei were then washed twice with CUT&RUN Triton wash buffer and resuspended in 100 μ l of Triton wash buffer and incubated on ice for 5 min. Then, 2 μ l 100 mM CaCl₂ was added and mixed gently to each 100 μ l ConA-bound-nuclei. The reaction was then incubated at 0 °C for 30 min. The reaction was stopped by addition of 100 μ l $2\times$ stop buffer (340 mM NaCl, 20 mM EDTA, 4 mM egtazic acid, 0.1% Triton X-100 and 50 μ g ml⁻¹ RNaseA) and incubated at 37 °C for 10 min. All buffers mentioned above were filtered with 0.22 μ m filter before use. After incubation, ConA-bound-nuclei were captured using a magnet and supernatant containing CNR DNA fragments were collected. Supernatant was incubated at 70 °C for 4 h with 2 μ l 10% sodium dodecyl sulfate and 2.5 μ l 20 mg ml⁻¹ proteinase K. DNA was purified using PCI reagent (phenol:chloroform:isoamyl alcohol, Millipore) and overnight ethanol precipitation with glycogen at –20 °C. DNA was resuspended in elution buffer (1 mM Tris–HCl pH 8.0 and 0.1 mM EDTA).

CNR 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 $1\times$ ratio of SPRI beads (Beckman) and eluted in 15 μ l EB buffer (Qiagen). All CNR 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 ≤ 120 bp 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 reads per kilobase of transcript per million mapped reads (RPKM) normalization and presented with Integrative Genomics Viewer software. CNR peaks were called using SEACR⁴⁷ from bedGraph files generated from RPKM-normalized Bigwig files (bigWigToBedGraph, UCSC Tools) using stringent setting and a numeric threshold of 0.01. Peaks were further filtered to have peak scores $>1,800$ for a set of high-confidence 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 (ref. 48). Samples were independently normalized to the number of 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.

BioID and mass spectrometry

For identification of SOX9-interacting partners we transduced primary *Krt14-rtTA* 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. We induced expression of recombinant proteins using 1 μ g ml⁻¹ DOX. Cells were allowed to expand for 5 days and were pulsed with 50 μ M biotin (Sigma) for 16 h before reaching confluence. Cells were purified and proteins isolated as previously described⁴⁹ with minor modifications mentioned below. Immediately after sonication, lysates were washed using Zeba desalting columns (7K molecular weight cut-off, ThermoFisher cat. no. 89894) with 50 mM Tris pH 7.4 to remove excess biotin. Beads were also washed three times with 2 M urea and a final two times with PBS before being resuspended with 500 μ l 50 mM 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 μ l 8 M urea, 50 mM ammonium bicarbonate and 10 mM dithiothreitol for 1 h and replaced with 100 μ l 40 mM iodoacetamide and incubated in the dark for 30 min. Alkylation solution was replaced with 1 μ g trypsin (Promega) dissolved in 100 μ l 50 mM ammonium bicarbonate and incubated for 4 h. Supernatant was then removed and re-digested overnight using 0.5 μ g trypsin and 0.5 μ g Endopeptidase Lys-C (Wako). Peptides were desalted and concentrated using C18-based Stage tips⁵⁰ 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.

Data were queried with UniProts Complete Proteome mouse database and concatenated with known common contaminants. Proteome Discover and Mascot was used to analyse the resulting data produced. Data were further filtered using a percolator⁵¹ to calculate peptide false discovery rates and set a threshold of 1%. Proteins were specific to SOX9’s proximity if they were identified in two of the three MYC-BioID2-SOX9 replicates and absent from all the MYC-BioID2-GFP-NLS samples. For the full list of SOX9-specific interactors and raw counts, see Supplementary Table 3.

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

Three versions of MYC-tagged SOX9 (WT, Δ TA and Δ HMG as indicated in Extended Data Fig. 7b) 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 described⁵². *Krt14-rtTA* EpdSCs were cultured and transduced with 1 μ l concentrated lentivirus in 10 ml E-low medium with 8 μ g ml⁻¹ polybrene (hexadimethrine bromide, Sigma 107689-100MG) overnight. Transduced cells were then selected with 2 μ g ml⁻¹ puromycin for 5 days before DOX treatment. For AFOS and ARID1a experiments, Flag-tagged *AFOS* or *Arid1a* CDS were cloned into the described pKO vector for lentiviral production. *Krt14-rtTA* or *Krt14-rtTA;TRE-mycSOX9* EpdSCs were cultured and transduced with 1 μ l concentrated lentivirus as described above. Transduced cells were also selected with puromycin for 5 days before DOX treatment.

CRISPR-mediated *Mll4* knockout

To generate *Mll4* (also known as *Kmt2d*) null lines, we cultured keratinocytes from the EpdSCs of our *Krt14-rtTA*, *TRE-Sox9* mice. Lines were generated with the Alt-R CRISPR-Cas9 system (Integrated DNA Technologies). Briefly, a recombinant Cas9 protein, a validated single guide RNA (TGCTCGGCAACAGACGTGAC) targeting *Mll4* or a negative control single guide RNA (Integrated DNA Technologies), and an ATTO-550 conjugated tracer RNA were used to form a ribonucleoprotein were 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 1 \times PBS, lysed in RIPA buffer (Millipore) supplemented with protease and phosphatase inhibitors (Roche), and collected by scraping. Cells were lysed for 15 min on ice and then centrifuged to collect the supernatant. Co-immunoprecipitation was performed as previously described⁵³ with the modification where protein-A/G-conjugated magnetic beads (Pierce) were used to bind antibodies instead, and proteins were eluted from beads with 1 \times NuPAGE LDS Sample Buffer (Invitrogen) with 2.5% 2-mercaptoethanol at 70 $^{\circ}$ C for 10 min. Protein concentration was determined by BCA Assay (Pierce) against a bovine serum albumin standard curve. Then 15 μ g protein of each sample was run on NuPAGE 4–12% Bis-Tris Gels (Invitrogen) for 2 h at 110 V in NuPAGE MOPS SDS Running Buffer (Invitrogen). Protein was transferred onto nitrocellulose membrane (Cytiva) in NuPAGE Transfer Buffer (Invitrogen) at 15 V overnight at 4 $^{\circ}$ C. 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, we often cut the blots on the basis of size 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 h at room temperature before incubating with primary antibodies. The following primary antibodies were diluted in blocking buffer: MYC-tag (mouse, 1:1,000, Cell Signaling), MLL4 (mouse, 1:200, Santa Cruz Biotechnology), cJUN (rabbit, 1:1,000, Cell Signaling), ARID1a (rabbit, 1:1,000, Abcam) and β -actin (mouse, 1:10,000, Cell Signaling). The membranes were incubated in primary antibodies overnight at 4 $^{\circ}$ C. Membranes were then washed three times in 0.1% Tween-20 in TBS before incubating with HRP secondary (1:10,000) antibody for 1 h at room temperature. After secondary antibody incubation, membranes were then washed four times in 0.1% Tween-20 in TBS and incubated in ECL Prime reagents (Cytiva) for 5 min before chemiluminescence detection. Membranes were imaged with an GE Amsham AI600 Imager. For clarity, we show the bands of the correct sizes. However, all full blots (cut before processing as delineated above) are shown in corresponding source data.

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. Complementary DNAs from each sample were normalized using primers against *Rps16*. All primers used are provided in Supplementary Table 4.

Bioinformatic analyses

GSEA. For comparing with both hair placodes and BCC, TPM matrices in D0, W2 and W12 were used as GSEA (v. 4.1.0) input. The DEG lists as illustrated in Fig. 1d were used as gene set inputs. For the BCC sample, DEG list of genes with $P < 0.05$ was curated from GSE152487 in the Gene Expression Omnibus depository¹⁷. 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.

Heat maps and box plots. All heat maps 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 and MLL3/4 CNR box plots in R with matrix output from Deeptools compute-matrix as input. The histone H3 profile plot was also generated with Profileplyr in R.

GO analysis. We 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)⁵⁴ with default gene association settings and the whole mouse genome (GRCm38/mm10) as the background.

Transcription factor motif and footprint analyses. For motif enrichment analysis on peak sets, HOMER⁵⁵ (v. 4.10) findMotifGenome.pl was used with a customized motif database from JASPAR2018 (ref. 56). The motif input for HOMER was generated from the 79 clusters of JASPAR2018 vertebrates CORE central transcription factor 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 our ATAC-seq clusters we ran HOMER for each cluster using the union set of dynamic peaks as our background (-bg) set with the options -size given -h. The resulting heat map was generated by combining the significant ($P < 0.05$) motifs for each cluster and plotting the associated P value. For motif distance measuring, we overlapped SOX9-bound opening peaks with known API and SOX motifs curated by HOMER (mm10-191020) and measured the distance from SOX motifs to the closest API motifs with Bedtools. For footprint analysis, we used HINT-ATAC⁵⁷ with our 79 motif clusters as the input as well. For transcription factor motif variability score analysis, we ran ChromVAR³¹ (1.18.0) on the dynamic peaks for differential chromatin accessibility across our 79 motif clusters to find the top variable motifs in dynamic peaks. We further used ChromVAR to calculate the motif deviation scores over time at the top variable motifs.

Illustrations. Schematics were prepared using BioRender and Adobe Illustrator (v. 26.0.1).

Statistics and reproducibility. No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those reported in previous publications^{14,16,46}. No data points were excluded. Upon collection, 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-immunoprecipitation and immunoblot experiments were repeated twice with samples collected on different days. The statistics in Fig. 5g

and Extended Data Fig. 1d were analysed 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 \pm s.d. **P* < 0.05, ***P* < 0.01, ****P* < 0.001 and *****P* < 0.0001.

Resource availability

Lead contact. Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, E.F. (fuchslb@rockefeller.edu).

Materials availability. Will be provided upon request and available upon publication.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data that support the findings of this study are available within the paper and its supplementary files. Sequencing data that support the findings of this study have been deposited in the Gene Expression Omnibus under accession code [GSE208072](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE208072). Previously published RNA-seq data from BCC, SCC and normal EpdSCs that were re-analysed here are available under accession code [GSE152487](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152487). Source data are provided with this paper. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Code availability

All bioinformatic analysis tools and pipelines used in this study are documented in the method section. Codes are available from the corresponding author upon reasonable request.

References

42. Vasioukhin, V., Degenstein, L., Wise, B. & Fuchs, E. The magical touch: genome targeting in epidermal stem cells induced by tamoxifen application to mouse skin. *Proc. Natl Acad. Sci. USA* **96**, 8551–8556 (1999).
43. Blanpain, C., Lowry, W. E., Geoghegan, A., Polak, L. & Fuchs, E. Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. *Cell* **118**, 635–648 (2004).
44. Rheinwald, J. G. & Green, H. Epidermal growth factor and the multiplication of cultured human epidermal keratinocytes. *Nature* **265**, 421–424 (1977).
45. Smith, J. P. et al. PEPATAC: an optimized pipeline for ATAC-seq data analysis with serial alignments. *NAR Genom. Bioinform.* **3**, lqab101 (2021).
46. Larsen, S. B. et al. Establishment, maintenance, and recall of inflammatory memory. *Cell Stem Cell* **28**, 1758–1774 e1758 (2021).
47. Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42 (2019).
48. Stovner, E. B. & Saetrom, P. epic2 efficiently finds diffuse domains in ChIP-seq data. *Bioinformatics* **35**, 4392–4393 (2019).
49. Kim, D. I. & Roux, K. J. Filling the void: proximity-based labeling of proteins in living cells. *Trends Cell Biol.* **26**, 804–817 (2016).
50. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).
51. Kall, L., Canterbury, J. D., Weston, J., Noble, W. S. & MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **4**, 923–925 (2007).
52. Beronja, S., Livshits, G., Williams, S. & Fuchs, E. Rapid functional dissection of genetic networks via tissue specific transduction and RNAi in mouse embryos. *Nat. Med.* **16**, 821–827 (2010).

53. Guzzi, N. et al. Pseudouridine-modified tRNA fragments repress aberrant protein synthesis and predict leukaemic progression in myelodysplastic syndrome. *Nat. Cell Biol.* **24**, 299–306 (2022).
54. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
55. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
56. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018).
57. Li, Z. et al. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* **20**, 45 (2019).

Acknowledgements

We thank M. Nikolova, E. Wong, J. Racelis, T. Omenchenko and J. Levorse for experimental assistance; M. Parigi, N. Guzzi, M. D. Abdusselamoglu, S. Yuan, A. Gola, K. Gonzales and C. Cowley for discussions; S. Mazel, S. Semova, S. Han and S. Shalaby for conducting FACS sorting; C. Lai (for high-throughput sequencing and raw data analyses; H. Molina for conducting mass spectrometry. E.F. is a Howard Hughes Medical Investigator. N.G. was the recipient of Burroughs Wellcome Diversity fellowship (1017355), and an F32 postdoctoral fellowship from the National Cancer Institute (5F32CA221353). N.I. was the recipient of an F31 from the National Institutes of Health (5F31AR073110). M.L. was the recipient CIHR postdoctoral fellowship. This study was supported by grants to E.F. from the National Institutes of Health (R01-AR31737 and R01-AR050452).

Author contributions

Y.Y., N.G. and E.F. conceptualized the study, designed the experiments, interpreted the data and wrote the manuscript. Y.Y. and N.G. performed and analysed *in vivo* high throughput data. M.L. assisted with proteomic experiments. R.C.A. generated the SOX9-inducible transgenic mice. Y.Y. performed *in vitro* studies with help from I.B., immunofluorescence microscopy and quantifications. M.S. participated in SOX9 mouse experiments and tumour cell engraftments. N.I. participated in high-throughput data generation, immunofluorescence microscopy and quantifications. All authors provided input on the final manuscript.

Competing interests

The authors declare no competing financial interests in this research, but E.F. was on the Scientific Advisory Boards of L’Oreal and Arsenal Biosciences during a period while these studies were ongoing.

Additional information

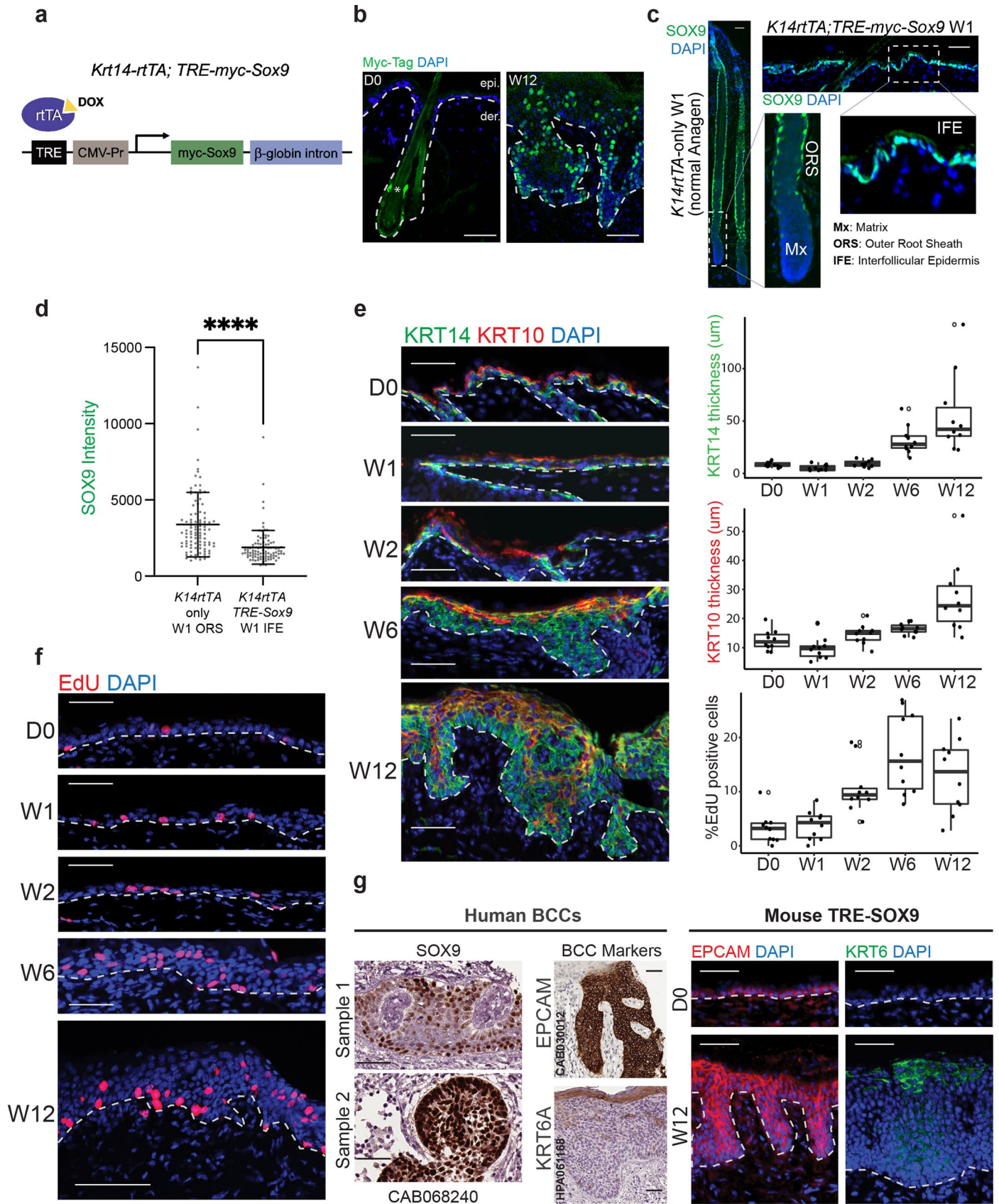
Extended data is available for this paper at <https://doi.org/10.1038/s41556-023-01184-y>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41556-023-01184-y>.

Correspondence and requests for materials should be addressed to Elaine Fuchs.

Peer review information *Nature Cell Biology* thanks Yali Dou, Anthony Oro and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

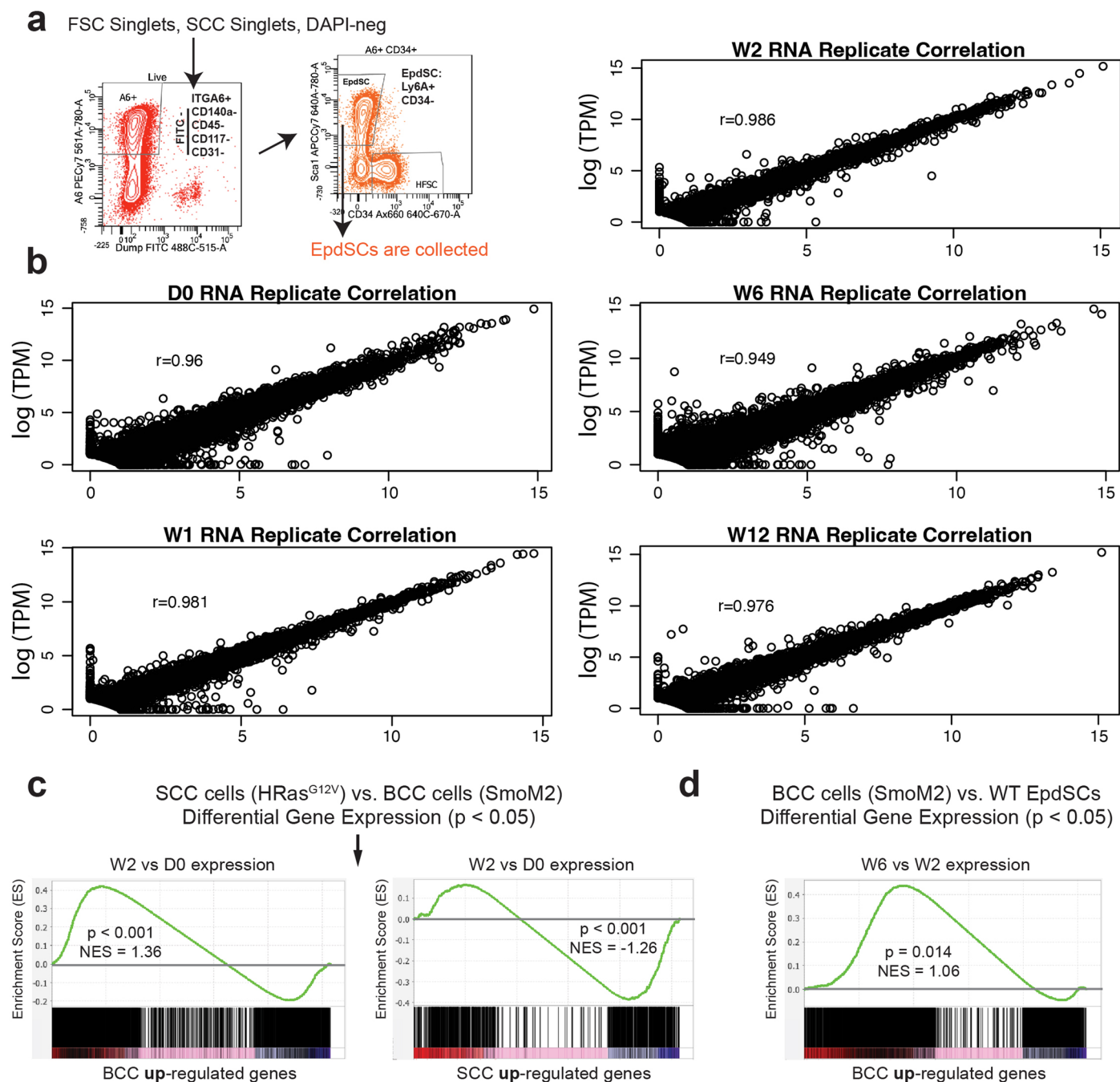
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | See next page for caption.

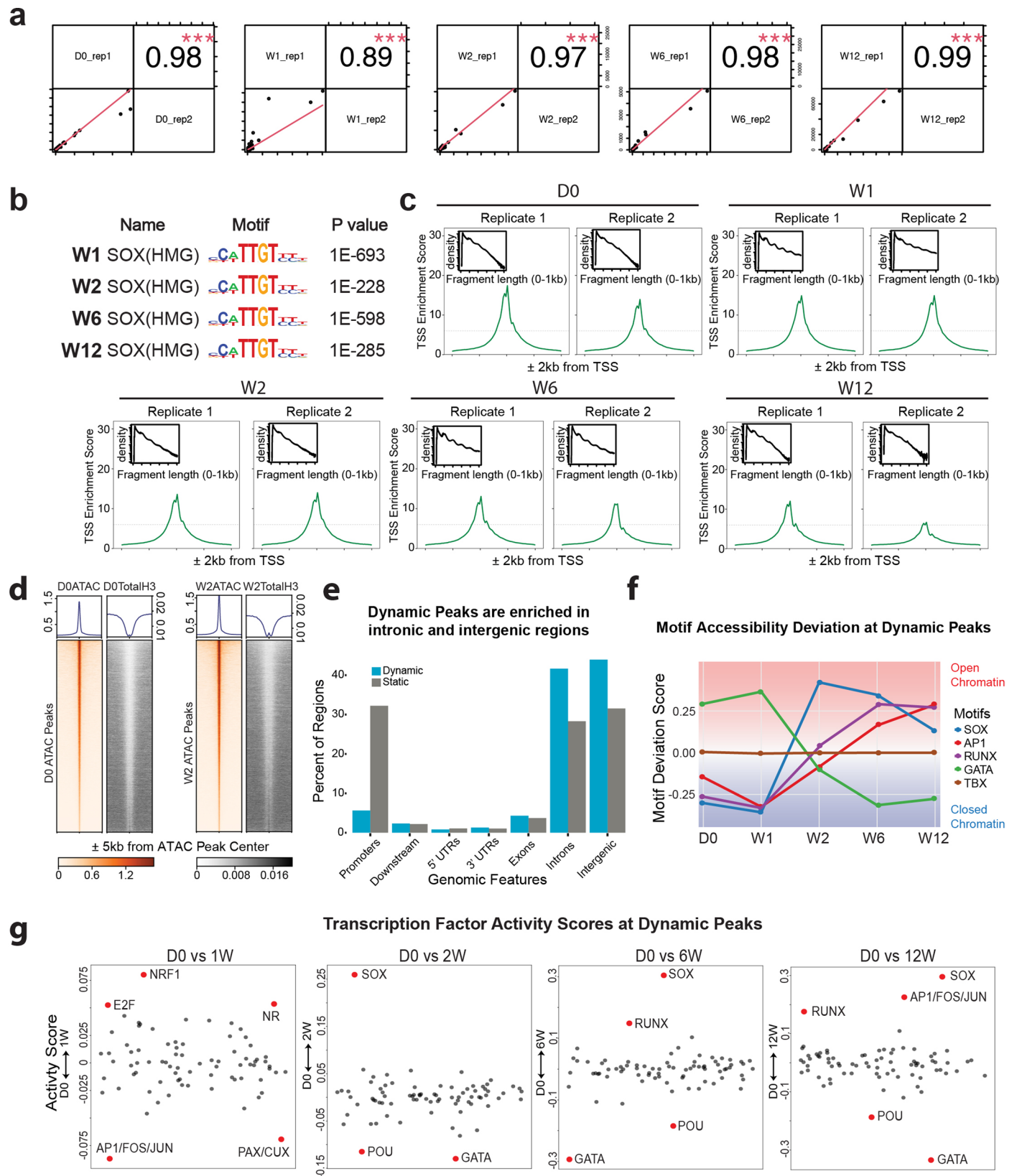
Extended Data Fig. 1 | Ectopic reactivation of SOX9 in adult EpdSCs silences epidermal fate and activates hair follicle fate within 2W and then progresses to BCC-like lesions. **a**, Schematic of the *TRE-Sox9* construct for generating SOX9-inducible transgenic mice. **b**, Immunofluorescence showing myc-tag staining at D0 and W12 SOX9-induction. This validates myc-tag expression is only induced and specifically induced in EpdSCs. All scale bars are 50µm. **c**, Immunofluorescence comparing SOX9 expression in the normal anagen ORS (containing activated HFSCs) and induced SOX9 expression in the adult EpdSCs. All scale bars are 50µm. **d**, Quantifications of SOX9 immunofluorescence intensity in the two states in (b) shows that SOX9 levels in ectopically induced adult EpdSCs are not higher than in native ORS HFSCs. n = 100 cells measured over 5 biological replicates. All the error bars are mean ± SD. Statistical significance from two-tailed t-test is denoted by ****(p < 0.0001). **e**, (left) KRT14 and KRT10 immunofluorescence of the skin after SOX9-induction in EpdSCs.

Note that the KRT14 skin progenitor is markedly expanded over time. All scale bars are 50µm. (right) Quantification of the thickness of KRT14 and KRT10 layers over time. n = 5 biological replicates with 2 measurement per sample. Boxplots are centered at median and bound by 1st and the 3rd quartile, and whiskers extend to 1.5 times IQR on both ends. The solid dots are data points, and the empty circles are outliers beyond 1.5 times IQR. **f**, (top) cell proliferation as assessed by EdU immunofluorescence after SOX9-induction. All scale bars are 50µm. (bottom left) quantification of % proliferating cells in the basal layer of EpdSCs. **g**, (left) Immunostaining of human BCC samples from The Human Protein Atlas. SOX9 and EpCAM show strong staining in the body of the lesion, whereas KRT6A is restricted to the apical epidermis. (right) Immunofluorescence images show the lesions in 12-week samples have similar EpCAM and KRT6 staining pattern as human BCC. Dotted lines denote the dermo-epidermal border. All scale bars are 50µm.



Extended Data Fig. 2 | Quality control for transcriptome analysis and temporal comparisons between SOX9-induced transcriptome analyses and cutaneous cancers. **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

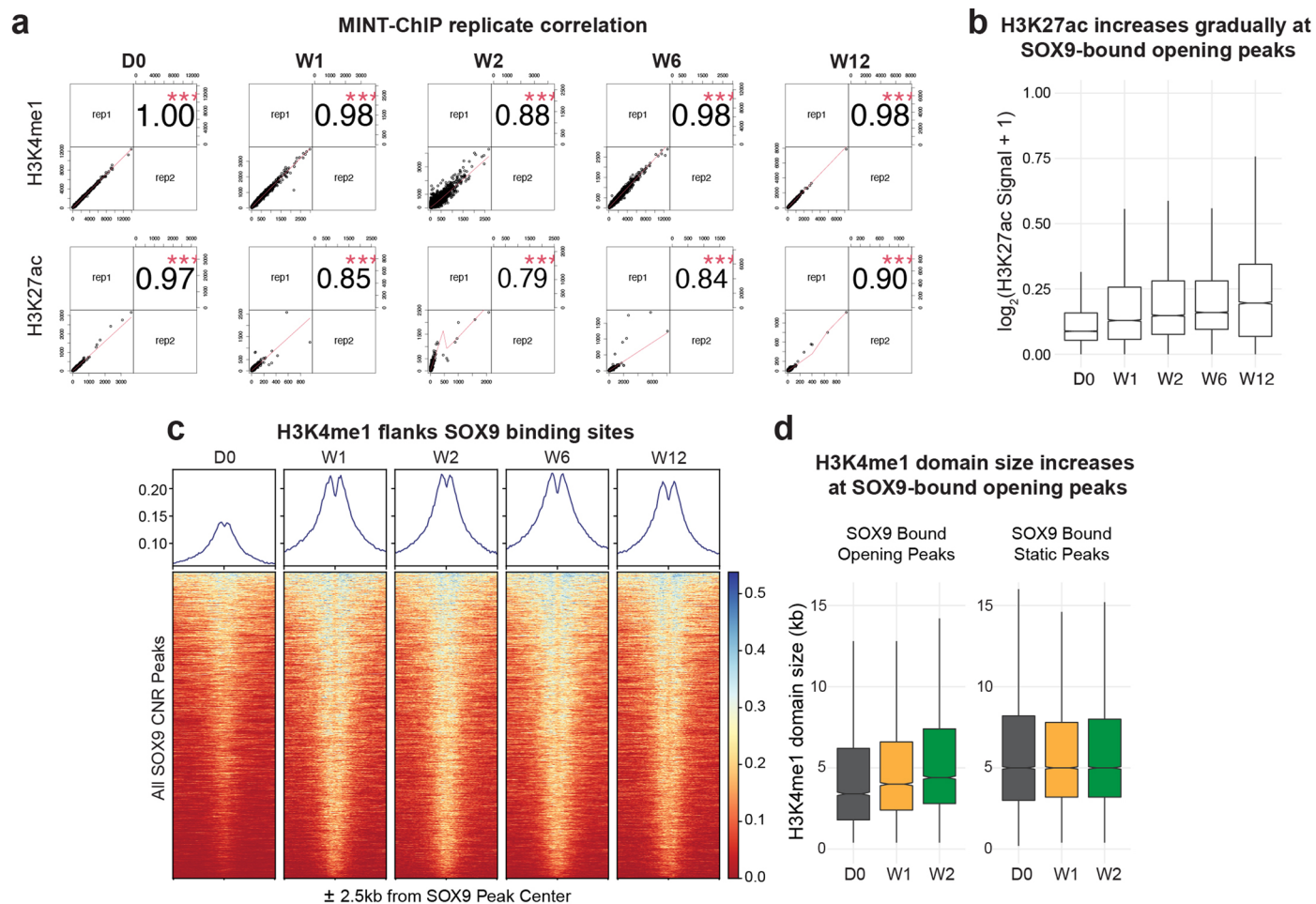
EpdSCs, respectively. **b**, Replicate correlation analyses of RNA-seq show strong correlation ($r > 0.94$) between samples across time points. **c**, At W2 following SOX9 induction, similarities to BCC were already apparent (K-S test, $p < 0.001$). By contrast, a negative correlation to SCC was observed (K-S test, $p < 0.001$). **d**, In comparing all BCC genes to our W6 and W2 data, the shift towards a BCC signature continued to rise, concomitant with the phenotypic changes described in Fig. 1 (K-S test, $p = 0.014$).



Extended Data Fig. 3 | See next page for caption.

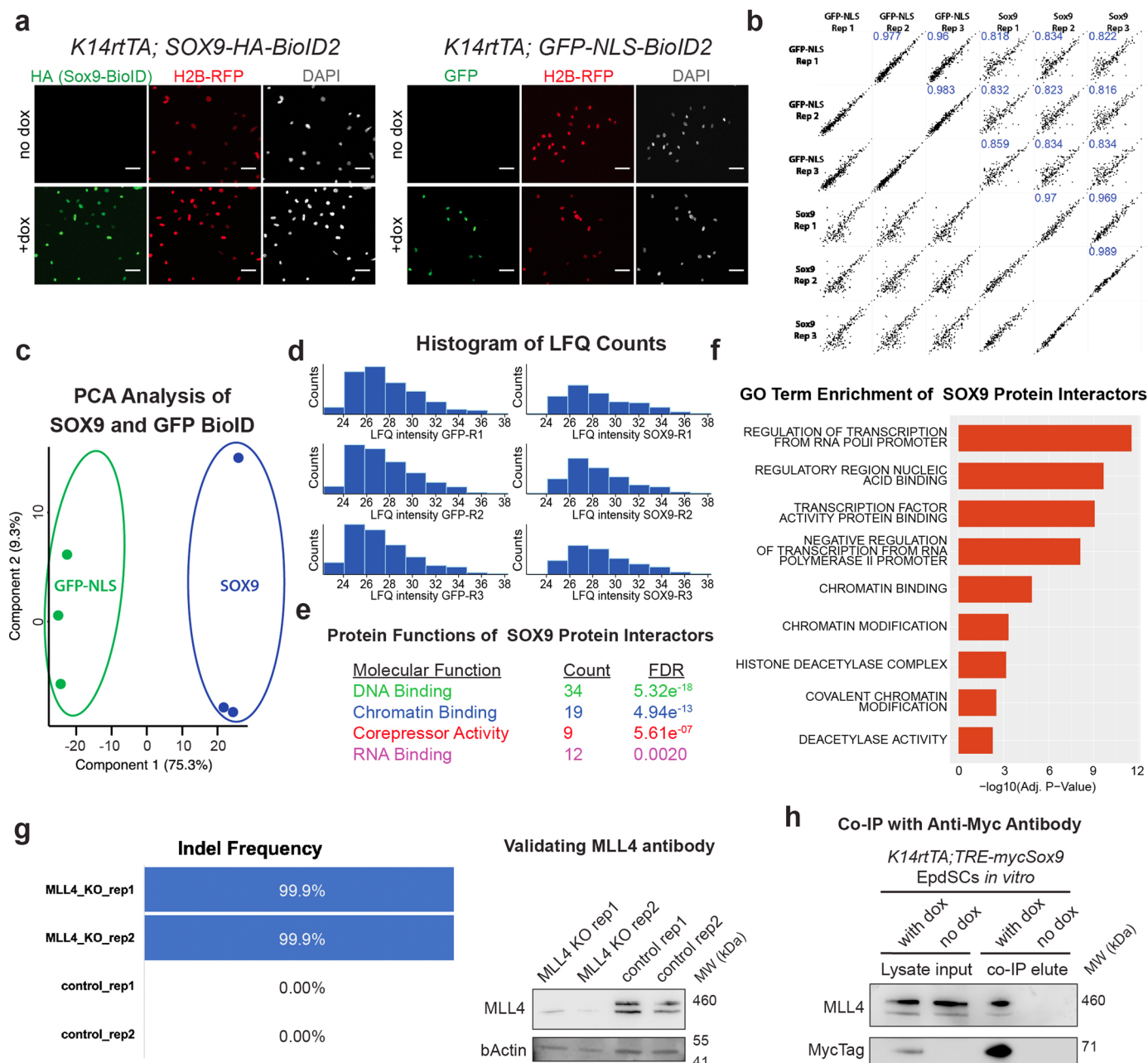
Extended Data Fig. 3 | Analyses and dynamics of ATAC-seq and Cut-and-Run and peak distributions. **a**, Replicate correlation analyses of SOX9 Cut-and-Run show strong correlation between duplicate samples at each time point. **b**, HOMER motif analysis shows that SOX(HMG) motifs are strongly enriched in SOX9 CNR peaks in W1 to W12 samples. **c**, ATAC-seq samples show clear nucleosome patterning and TSS enrichment. **d**, Total histone H3 signals are mutually exclusive from ATAC signals at D0 and W2 time points. **e**, Distribution of dynamic and static peaks at different genomic features. Dynamic peaks are more enriched in intronic and intergenic regions, while static peaks are more

enriched at promoters. **f**, ChromVAR analysis of motif deviation scores of SOX (blue), API/FOS/JUN (red), RUNX (purple), GATA (green), and TBX (brown) motifs at indicated time points. Note that SOX motif accessibility rises markedly in dynamic peaks that open within the first two weeks post SOX9 induction while GATA motif accessibility declines. API and RUNX motif accessibility rise between W2 and W6. TBX motif is shown as a control which does not change accessibility overtime. **g**, HINT-ATAC footprint analysis shows how transcription factor (TF) motif footprints differ in activity score at indicated time points compared to D0. Motifs indicated with red dot show significant changes in activity score over D0.



Extended Data Fig. 4 | MINT-ChIP experiments and analyses of H3K4me1 relative to SOX9 binding. **a**, Replicate correlation analyses of the H3K4me1 and H3K27ac MINT-ChIP show strong correlation between each duplicate sample across all time points. **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 bound by 1st and the 3rd quartile, and whiskers extend to 1.5 times IQR on both ends. **c**, Profile plot (top) and

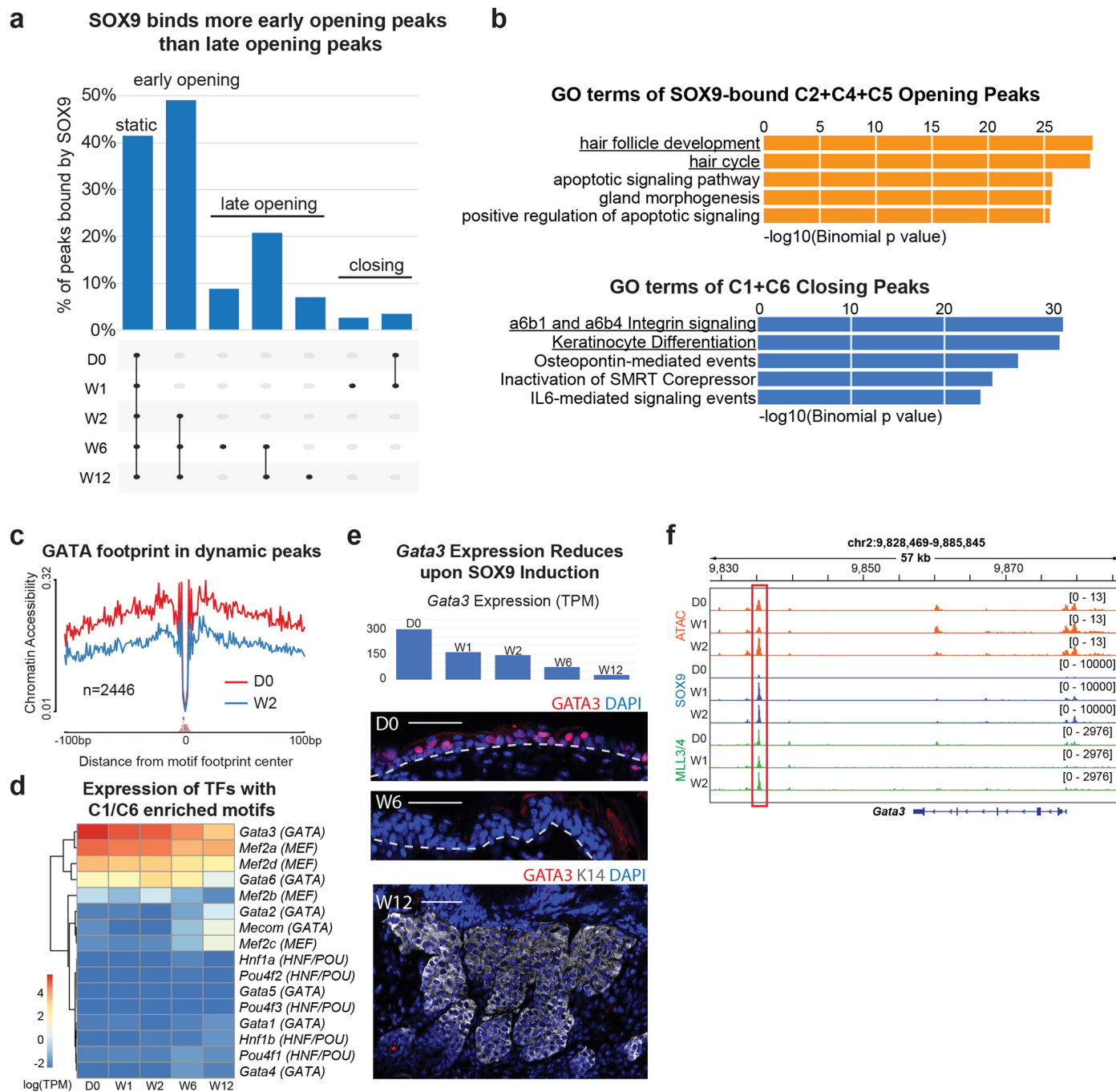
heatmap (bottom) showing the H3K4me1 signals at all SOX9 Cut-and-Run peaks across time points. Note the strong flanking pattern of the H3K4me1 signals adjacent to each SOX9 binding site (center dip) from W1 to W12. **d**, Boxplot of H3K4me1 domain sizes at opening SOX9 peaks and SOX9 bound static peaks shows gradual increase of H3K4me1 domain size from W1 to W2 after SOX9 induction. $n = 2$ biological replicates. Boxplots are centered at median and bound by 1st and the 3rd quartile, and whiskers extend to 1.5 times IQR on both ends.



Extended Data Fig. 5 | BioID experiments and identification of SOX9

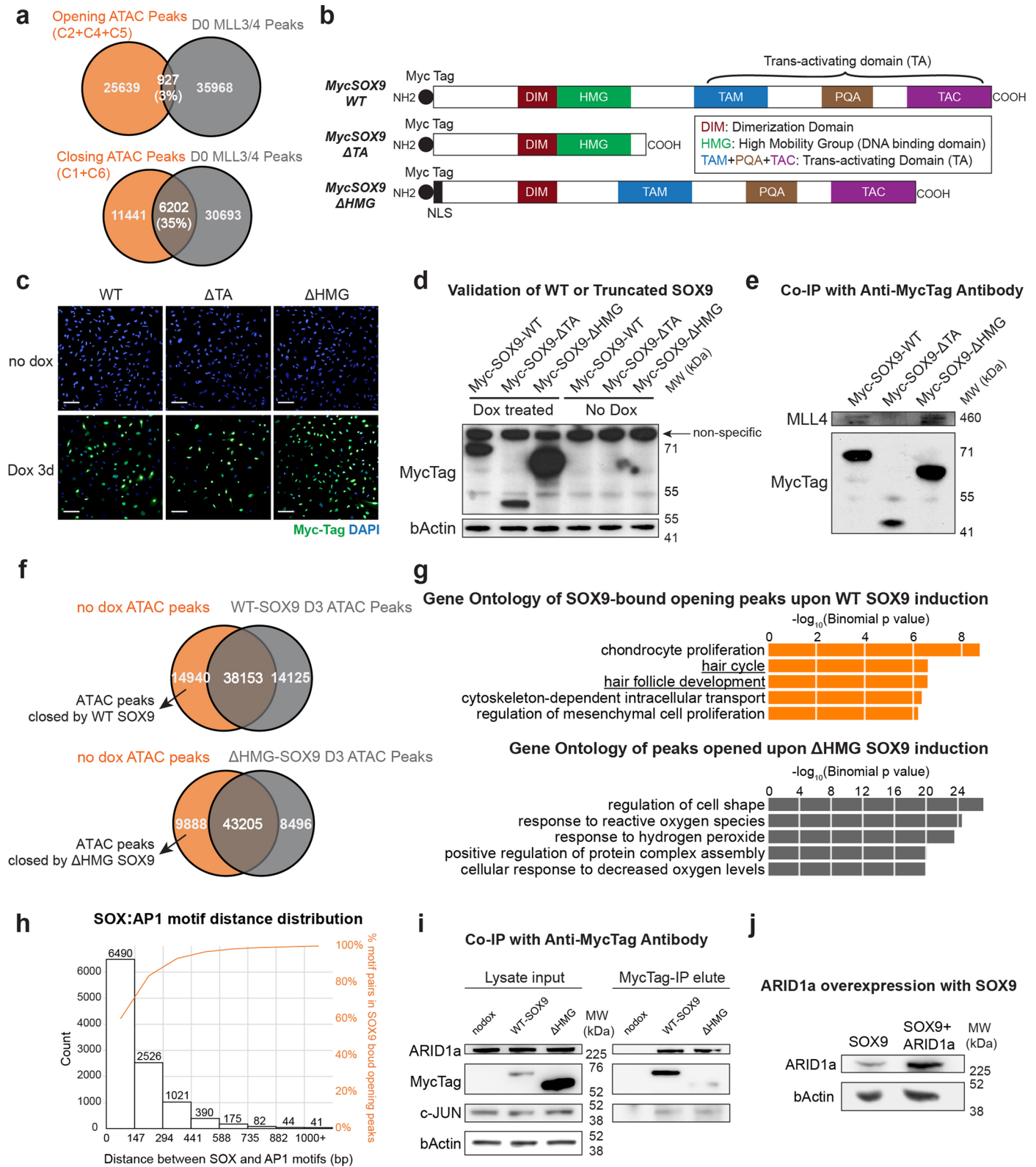
co-factors. **a**, Immunofluorescence validation of *Krt14rtTA* primary keratinocytes transduced with TRE-HA-SOX9-BioID2; H2B-RFP (left) or TRE-GFP-NLS-BioID2; H2B-RFP (right). Transgenes were induced with doxycycline and immunolabeled for HA (left) or GFP (right). All scale bars are 50µm. **b**, Correlation plots of Label Free Quantification (LFQ) values identified by BioID experiments across replicates and samples. SOX9-BioID and GFP-NLS-BioID samples share stronger correlations between replicates than each other. Blue number represents r^2 value. **c**, PCA analysis of GFP and SOX9 BioID protein interactors demonstrate sample-specific clusters. **d**, Histogram of LFQ intensity values for

the GFP (left) and SOX (right) BioID replicates. **e**, Molecular function enrichment of proteins specifically interacting with SOX9. **f**, Gene ontology enrichment of proteins specific to SOX9 (binomial is used to calculate the significance). **g**, Two *Mll4* null keratinocyte cell lines (*Krt14rtTA; TRE-mycSox9*) were generated with CRISPR/Cas9 with 99.9% indel frequency. These 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. **h**, 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.



Extended Data Fig. 6 | Epidermal TFs diminish rapidly upon SOX9 induction in EpcSCs. **a**, 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. **b**, GO terms enriched in SOX9-bound opening peaks (C2, C4, C5) and all closing peaks (C1, C6). **c**, ATAC footprint analysis at D0 and W2 shows a decrease in chromatin accessibility at GATA footprint in dynamic ATAC peaks following SOX9 induction. **d**, EpcSCs expression of TFs 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). **e**, (top) Transcript levels of *Gata3* over time following SOX9 induction. (bottom) GATA3 immunofluorescence of epidermis at D0, W6 and W12. Scale bars, 50µm. **f**, Integrative Genomics Viewer (IGV) snapshot of SOX9, ATAC and H3K4me1 tracks within the *Gata3* locus. The red box indicates a peak >30kb downstream from the gene that is bound by SOX9 and MLL3/4, and stays open at W2 even though the gene body of *Gata3* closes its chromatin (see Fig. 5b).



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Truncated SOX9 displays impaired DNA binding or co-factor recruitment. **a**, (top) Venn diagram reveals almost no overlap between ATAC peaks (C2,C4,C5) that mostly open between W2-W12 and D0 MLL3/4 Cut-and-Run peaks. (bottom) Venn diagram shows substantial overlap between ATAC peaks (C1,C6) that mostly close between W1-W2 and D0 MLL3/4 Cut-and-Run peaks. **b**, Schematic illustrating constructs engineered to express WT SOX9 and two variants of SOX9. **c**, Immunofluorescence reveals similar intensities of WT and mutant SOX9 in induced EpdSCs. Scale bars, 100 μ m. **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. **f**, Venn diagrams show that the peak sets closed by WT and Δ HMG-SOX9 are comparable in size. **g**, (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 Δ HMG SOX9 induction *in vitro* from GREAT. Binomial is used to calculate the significance. **h**, 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 multiples of one nucleosome size (147bp). The cumulation plot of the distribution is shown in orange on the secondary y-axis. **i**, Immunoblot showing that both WT-SOX9 and Δ HMG-SOX9 are capable of binding c-JUN and ARID1a. **j**, Immunoblot validating ARID1a can be induced 3x higher in the SOX9 expressing keratinocytes. For full blots, see associated source data.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	<input type="text" value="No human research participants were used."/>
Population characteristics	<input type="text" value="See above"/>
Recruitment	<input type="text" value="See above"/>
Ethics oversight	<input type="text" value="See above"/>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	<input type="text" value="Sample sizes are similar to those reported in previous publications using a similar approach (Naik and Larsen et al. Nature, 2017, Adam et al. Nature, 2015)"/>
Data exclusions	<input type="text" value="No data was excluded for the analysis."/>
Replication	<input type="text" value="For our genomic experiments, each sample consisted of 3-4 mice per time point. Replicative analysis showed strong correlations for every one of our data types."/>
Randomization	<input type="text" value="K14rtTA+;TRE-SOX9+ mice were confirmed via early postnatal PCR genotyping. K14rtTA+;TRE-SOX9+ mice were housed with at least 2 K14rtTA+ only males. Upon harvest, samples were separated into experimental and control samples and randomly allocated to genomic and immunofluorescence experiments."/>
Blinding	<input type="text" value="Blinding was not possible or relevant in our study. Our mice were phenotypic upon harvesting. Additionally, we conducted a time-course study with D0 as control."/>

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	<input type="text" value="Ly6A/E-APCCy7,BioLegend,cat#108126
CD49f-PECy7,BioLegend,cat#313622
CD34-Alexa660,Invitrogen,cat#50-0341-82"/>
-----------------	--

CD45-biotin, BioLegend, cat#103104
 CD31-biotin, BioLegend, cat#102404
 CD140a-biotin, BioLegend, cat#135910
 CD117-biotin, BioLegend, cat#105804
 Streptavidin-FITC, BioLegend, cat#405202
 TruStain FcX, BioLegend, cat#101320
 SOX9, Millipore, cat#ab5535
 MYC-tag, Cell Signaling, cat#71D10
 MLL3/4, Wysocka Lab, cat#N/A
 totalH3, Active Motif, cat#39763
 H3K27ac, Active Motif, cat#39133
 H3K4me1, Cell Signaling, cat#D1A9
 SOX9, Abcam, cat#ab185966
 ITGA6, BD, cat#555734
 KRT14, BioLegend, cat#906004
 KRT10, Fuchs Lab, cat#N/A
 EpCAM, Abcam, cat#ab71916
 KRT6, Fuchs Lab, cat#N/A
 RUNX1, Abcam, cat#ab229482
 GATA3, Invitrogen, cat#14-9966-82
 HA-tag, Cell Signaling, cat#C29F4
 GFP, Fuchs Lab, cat#N/A
 RFP, ChromoTek, cat#5F8
 MYC-tag, Cell Signaling, cat#71D10
 MYC-tag, Cell Signaling, cat#9B11
 beta-Actin, Cell Signaling, cat#8H10D10
 MLL4, Santa Cruz Biotechnology, cat#sc-293217
 MYC-tag, Cell Signaling, cat#2276S
 IgG, Cell Signaling, cat#2729S
 ARID1a, Cell Signaling, cat#12354S
 ARID1a, Abcam, cat#ab182560
 JUN, Cell Signaling, cat#3753S
 BRG1, EpiCypher, cat#13-2002
 Secondary (all with donkey as host):
 AF488-Rabbit, Thermo Fisher, A-21206
 AF488-Chicken, Thermo Fisher, A78948
 AF546-Rabbit, Thermo Fisher, A10040
 AF546-Rat, Thermo Fisher, A78947

Validation

The fluorophore conjugated antibodies from and BioLegend were validated for flow cytometry:

Ly6A/E-APCCy7, BioLegend, cat#108126
 CD49f-PECy7, BioLegend, cat#313622
 CD34-Alexa660, Invitrogen, cat#50-0341-82
 CD45-biotin, BioLegend, cat#103104
 CD31-biotin, BioLegend, cat#102404
 CD140a-biotin, BioLegend, cat#135910
 CD117-biotin, BioLegend, cat#105804
 Streptavidin-FITC, BioLegend, cat#405202
 TruStain FcX, BioLegend, cat#101320

The following antibodies are validated with western blot:

SOX9, Millipore, cat#ab5535
 MYC-tag, Cell Signaling, cat#71D10
 MYC-tag, Cell Signaling, cat#9B11
 beta-Actin, Cell Signaling, cat#8H10D10
 MLL4, Santa Cruz Biotechnology, cat#sc-293217 (also validated with KO in this study)

The following antibodies are validated with immunofluorescence:

SOX9, Abcam, cat#ab185966
 ITGA6, BD, cat#555734
 KRT14, BioLegend, cat#906004
 KRT10, Fuchs Lab, cat#N/A
 EpCAM, Abcam, cat#ab71916
 KRT6, Fuchs Lab, cat#N/A
 RUNX1, Abcam, cat#ab229482
 GATA3, Invitrogen, cat#14-9966-82
 HA-tag, Cell Signaling, cat#C29F4
 GFP, Fuchs Lab, cat#N/A
 RFP, ChromoTek, cat#5F8
 all secondary antibodies

The following antibodies are validated for CHIP:

MLL3/4, Wysocka Lab, cat#N/A
 totalH3, Active Motif, cat#39763
 H3K27ac, Active Motif, cat#39133
 H3K4me1, Cell Signaling, cat#D1A9

ARID1a, Cell Signaling, cat#12354S
 ARID1a, Abcam, cat#ab182560
 JUN, Cell Signaling, cat#3753S
 BRG1, EpiCypher, cat#13-2002

The following antibodies are validated for colP:
 MYC-tag, Cell Signaling, cat#2276S
 IgG, Cell Signaling, cat#2729S

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	Mouse keratinocyte cell lines from K14rtTA+ or K14rtTA+;TRE-SOX9+ mice. The sex is not available from the cell lines.
Authentication	The cell lines used for this study were generated within the Fuchs laboratory and confirmed by PCR genotyping.
Mycoplasma contamination	These specific Cell lines were not tested for mycoplasma but our lab routinely performs mycoplasma contamination checks of randomly selected lines throughout the year.
Commonly misidentified lines (See ICLAC register)	No cell lines in the ICLAC database were used.

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	Mice {Mus musculus} from 2 transgenic mouse lines: TRE-MYC-SOX9;K14rtTA and K14rtTA only. 3-4 male mice at age 3-9 weeks were pooled per time point analyzed. For skin grafts, 6-8 week old female Nude mice were used as recipients for P0 male mouse skin engraftment.
Wild animals	The study did not involve wild animals.
Reporting on sex	In order to maximize cell numbers and minimize variation due to sex, we used male mice for all experiments. Male mice were generally larger enabling more surface area of the skin to harvest EpdSCs.
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	All experimental procedures were conducted with accordance and approval of the Institutional Animal Care and Use Committee (IACUC) – approved protocols at the Rockefeller University (20012-H and 20066-H).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

May remain private before publication.

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE208072>

Files in database submission

ATAC:
 Sample 1-D0_ATAC_rep1
 Sample 2-D0_ATAC_rep2
 Sample 3-W1_ATAC_rep1
 Sample 4-W1_ATAC_rep2
 Sample 5-W2_ATAC_rep1
 Sample 6-W2_ATAC_rep2
 Sample 7-W6_ATAC_rep1
 Sample 8-W6_ATAC_rep2
 Sample 9-W12_ATAC_rep1
 Sample 10-W12_ATAC_rep2
 Sample 11-cultured_WT_dox_rep1
 Sample 12-cultured_WT_nodox_rep1
 Sample 13-cultured_noHMG_dox_rep1
 Sample 14-cultured_noTA_dox_rep1
 Sample 15-cultured_WT_dox_rep2
 Sample 16-cultured_WT_nodox_rep2

Sample 17-cultured_noHMG_dox_rep2
 Sample 18-cultured_noTA_dox_rep2
 Sample 19-cultured_K14AFOS_dox_rep1
 Sample 20-cultured_K14AFOS_dox_rep2
 Sample 21-cultured_K14SOX9AFOS_dox_rep1
 Sample 22-cultured_K14SOX9AFOS_dox_rep2
 Sample 23-cultured_K14SOX9ARID1a_dox_rep1
 Sample 24-cultured_K14SOX9ARID1a_dox_rep2
 MINT-ChIP:
 Sample 1-D0_H3K4me1_rep1_rep2
 Sample 2-D0_H3K27ac_rep1_rep2
 Sample 3-D0_totalH3_rep1_rep2
 Sample 4-W1_H3K4me1_rep1_rep2
 Sample 5-W1_H3K27ac_rep1_rep2
 Sample 6-W1_totalH3_rep1_rep2
 Sample 7-W2_H3K4me1_rep1_W12_H3K4me1_rep1_rep2
 Sample 8-W2_H3K4me1_rep2
 Sample 9-W2_H3K27ac_rep1_W12_H3K27ac_rep1_rep2
 Sample 10-W2_H3K27ac_rep2
 Sample 11-W2_totalH3_rep1_W12_totalH3_rep1_rep2
 Sample 12-W2_totalH3_rep2
 Sample 13-W6_H3K4me1_rep1
 Sample 14-W6_H3K4me1_rep2
 Sample 15-W6_H3K27ac_rep1
 Sample 16-W6_H3K27ac_rep2
 Sample 17-W6_totalH3_rep1
 Sample 18-W6_totalH3_rep2
 Bulk RNA-Seq:
 Sample 1-RNA_D0_rep1
 Sample 2-RNA_D0_rep2
 Sample 3-RNA_W1_rep1
 Sample 4-RNA_W1_rep2
 Sample 5-RNA_W2_rep1
 Sample 6-RNA_W2_rep2
 Sample 7-RNA_W6_rep1
 Sample 8-RNA_W6_rep2
 Sample 9-RNA_W12_rep1
 Sample 10-RNA_W12_rep2
 Sample 11-RNA_SOX9neg_rep1
 Sample 12-RNA_SOX9neg_rep2
 Sample 13-RNA_SOX9pos_rep1
 Sample 14-RNA_SOX9pos_rep2
 Cut-and-Run:
 Sample 1-SOX9CNR_D0_rep1
 Sample 2-SOX9CNR_D0_rep2
 Sample 3-SOX9CNR_W1_rep1
 Sample 4-SOX9CNR_W1_rep2
 Sample 5-SOX9CNR_W2_rep1
 Sample 6-SOX9CNR_W2_rep2
 Sample 7-SOX9CNR_W6_rep1
 Sample 8-SOX9CNR_W6_rep2
 Sample 9-SOX9CNR_W12_rep1
 Sample 10-SOX9CNR_W12_rep2
 Sample 11-mllCNR_D0_rep1
 Sample 12-mllCNR_D0_rep2
 Sample 13-mllCNR_W1_rep1
 Sample 14-mllCNR_W1_rep2
 Sample 15-mllCNR_W2_rep1
 Sample 16-mllCNR_W2_rep2
 Sample 17-mllCNR_WTSOX9_nodox_rep1
 Sample 18-mllCNR_noHMG_dox_rep1
 Sample 19-mllCNR_noTA_dox_rep1
 Sample 20-mycCNR_WTSOX9_dox_rep1
 Sample 21-mycCNR_WTSOX9_nodox_rep1
 Sample 22-mycCNR_noHMG_dox_rep1
 Sample 23-mycCNR_noTA_dox_rep1
 Sample 24-mllCNR_WTSOX9_nodox_rep2
 Sample 25-mllCNR_noHMG_dox_rep2
 Sample 26-mllCNR_noTA_dox_rep2
 Sample 27-mycCNR_WTSOX9_dox_rep2
 Sample 28-mycCNR_WTSOX9_nodox_rep2
 Sample 29-mycCNR_noHMG_dox_rep2
 Sample 30-mycCNR_noTA_dox_rep2
 Sample 31-mllCNR_WTSOX9_dox_rep1
 Sample 32-mllCNR_WTSOX9_dox_rep2
 Sample 33-JunCNR_WTSOX9_nodox_rep1

Sample 34-JunCNR_WTSOX9_dox_rep1
 Sample 35-JunCNR_noHMG_dox_rep1
 Sample 36-Arid1aCNR_WTSOX9_nodox_rep1
 Sample 37-Arid1aCNR_WTSOX9_dox_rep1
 Sample 38-Arid1aCNR_noHMG_dox_rep1
 Sample 39-BRG1CNR_WTSOX9_nodox_rep1
 Sample 40-BRG1CNR_WTSOX9_dox_rep1
 Sample 41-BRG1CNR_noHMG_dox_rep1

Genome browser session
 (e.g. [UCSC](#))

Available upon reasonable request. All processed bigwig files are provided in GEO for visualization.

Methodology

Replicates

Each data set was replicated twice at independent times with different cohorts of mice. Experimental replicates were highly concordant and displayed within the extended data.

Sequencing depth

Sample, Aligned Reads, Paired/Single End reads
 D0_Rep1_ATAC, 43858222, 50bp Paired-end
 D0_Rep2_ATAC, 39846438, 50bp Paired-end
 W1_Rep1_ATAC, 34846318, 50bp Paired-end
 W1_Rep2_ATAC, 37261602, 50bp Paired-end
 W2_Rep1_ATAC, 18838504, 50bp Paired-end
 W2_Rep2_ATAC, 13723684, 50bp Paired-end
 W6_Rep1_ATAC, 30156910, 50bp Paired-end
 W6_Rep2_ATAC, 19672580, 50bp Paired-end
 W12_Rep1_ATAC, 72898350, 50bp Paired-end
 W12_Rep2_ATAC, 19433948, 50bp Paired-end
 D0_Rep1_H3K27ac, 3841033, 50bp Paired-end
 D0_Rep2_H3K27ac, 3121333, 50bp Paired-end
 W1_Rep1_H3K27ac, 711249, 50bp Paired-end
 W1_Rep2_H3K27ac, 3074037, 50bp Paired-end
 W2_Rep1_H3K27ac, 1030516, 50bp Paired-end
 W2_Rep2_H3K27ac, 3928882, 50bp Paired-end
 W6_Rep1_H3K27ac, 7956088, 50bp Paired-end
 W6_Rep2_H3K27ac, 1455927, 50bp Paired-end
 W12_Rep1_H3K27ac, 1258550, 50bp Paired-end
 W12_Rep2_H3K27ac, 1403434, 50bp Paired-end
 D0_Rep1_H3K4me1, 24592016, 50bp Paired-end
 D0_Rep2_H3K4me1, 22063883, 50bp Paired-end
 W1_Rep1_H3K4me1, 6587117, 50bp Paired-end
 W1_Rep2_H3K4me1, 9550526, 50bp Paired-end
 W2_Rep1_H3K4me1, 6041330, 50bp Paired-end
 W2_Rep2_H3K4me1, 10653668, 50bp Paired-end
 W6_Rep1_H3K4me1, 27215545, 50bp Paired-end
 W6_Rep2_H3K4me1, 8697635, 50bp Paired-end
 W12_Rep1_H3K4me1, 6916330, 50bp Paired-end
 W12_Rep2_H3K4me1, 8388474, 50bp Paired-end
 D0_Rep1_Total H3, 37680307, 50bp Paired-end
 D0_Rep2_Total H3, 32670127, 50bp Paired-end
 W1_Rep1_Total H3, 12964069, 50bp Paired-end
 W1_Rep2_Total H3, 13337757, 50bp Paired-end
 W2_Rep1_Total H3, 9340777, 50bp Paired-end
 W2_Rep2_Total H3, 58641446, 50bp Paired-end
 W6_Rep1_Total H3, 40744095, 50bp Paired-end
 W6_Rep2_Total H3, 26969269, 50bp Paired-end
 W12_Rep1_Total H3, 8736431, 50bp Paired-end
 W12_Rep2_Total H3, 11271301, 50bp Paired-end
 D0_Rep1_SOX9_CNR, 20174088, 50bp Paired-end
 D0_Rep2_SOX9_CNR, 19645106, 50bp Paired-end
 W1_Rep1_SOX9_CNR, 23997258, 50bp Paired-end
 W1_Rep2_SOX9_CNR, 370168, 50bp Paired-end
 W2_Rep1_SOX9_CNR, 9768786, 50bp Paired-end
 W2_Rep2_SOX9_CNR, 7253702, 50bp Paired-end
 W6_Rep1_SOX9_CNR, 30653330, 50bp Paired-end
 W6_Rep2_SOX9_CNR, 2028772, 50bp Paired-end
 W12_Rep1_SOX9_CNR, 5090868, 50bp Paired-end
 W12_Rep2_SOX9_CNR, 20178366, 50bp Paired-end
 D0_Rep1_MII_CNR, 7024229, 50bp Paired-end
 D0_Rep2_MII_CNR, 3703596, 50bp Paired-end
 W1_Rep1_MII_CNR, 10820527, 50bp Paired-end
 W1_Rep2_MII_CNR, 6310082, 50bp Paired-end
 W2_Rep1_MII_CNR, 5048765, 50bp Paired-end
 W2_Rep2_MII_CNR, 5542309, 50bp Paired-end

	<p>WTSOX9_Nodox_Rep1_Mll_CNR,10946976,50bp Paired-end noHMG_Dox_Rep1_Mll_CNR,6801070,50bp Paired-end noTA_Dox_Rep1_Mll_CNR,7902664,50bp Paired-end WTSOX9_Nodox_Rep1_Myc_CNR,3199032,50bp Paired-end noHMG_Dox_Rep1_Myc_CNR,2733154,50bp Paired-end WTSOX9_Dox_Rep1_Myc_CNR,3983006,50bp Paired-end noTA_Dox_Rep1_Myc_CNR,3532914,50bp Paired-end</p>
Antibodies	<p>SOX9, Millipore, cat#ab5535 MLL3/4, Wysocka Lab, cat#N/A totalH3, Active Motif, cat#39763 H3K27ac, Active Motif, cat#39133 H3K4me1, Cell Signaling, cat#D1A9 MYC-tag, Cell Signaling, cat#2276S IgG, Cell Signaling, cat#2729S ARID1a, Cell Signaling, cat#12354S JUN, Cell Signaling, cat#3753S BRG1, EpiCypher, cat#13-2002</p>
Peak calling parameters	<p>ATAC-Seq: 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.</p> <p>H3K27ac: Replicate BAM files were merged with keep-duplicate option and MACS2 called with standard parameters and total H3 as input.</p> <p>H3K4me1: Replicate BAM files were merged and converted to BEDPE. --treatment was sample and -c was Total H3.</p> <p>Cut-and-Run: D0 to W12 SOX9 peaks were called using SEACR37 from bedGraph files generated from RPKM normalized Bigwig files (bigWigToBedGraph, UCSC Tools) using stringent setting and a numeric threshold of 0.01. Peaks were further filtered to have peak scores > 1800 for a set of high confident peaks.</p>
Data quality	<p>Data was checked for high correlations between replicate samples and enrichment for known regulatory regions such as TSS for ATAC, and H3K27ac. SOX9 CNR was assayed by unbiased motif enrichment at called peaks. MLL3/4 CNR was assayed by enrichment at enhancer elements relative to promoter regions.</p>
Software	<p>Skewer (v0.2.2), R (v 3.6.1), MACS2 (v2.2.7.1), Bowtie2 (v2.2.9), Picard (v2.3.0), Samtools (v1.3.1), Deeptools (v3.1.2), Integrative Genomics Viewer, SEACR, epic2 (v0.0.52), GSEA, profileplyr(v1.4.3) HOMER (v4.10), JASPAR (2018), HINT-ATAC, ChromVar(v1.18.0)</p>

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	<p>Backskin of K14rtTA+;TRE-SOX9; and K14rtTA+ only mice were harvested and subjected to 0.25% trypsin/EDTA dissociation for 1 hour at 37C. After neutralization with FACS buffer cell suspension was strained, centrifuged and resuspended prior to antibody staining. A cocktail of antibodies for surface markers was prepared at predetermined concentrations in FACS buffer followed by washing and resuspension in secondary antibody. The cells were washed and resuspended again in FACS buffer containing 100 ng/mL of DAPI and filtered again through 70 uM filter caps before FACS.</p>
Instrument	<p>BD Biosciences FACSria equipped with FACSDiva software for sorting, BD Biosciences FACS Fortessa with FACSDiva software for analysis.</p>
Software	<p>FACSDiva 8.0 for operating the sorter or analyzer.</p>
Cell population abundance	<p>Post sorting of the samples was routinely performed and consistently showed greater than 90% purity of the isolated populations.</p>
Gating strategy	<p>Single cell suspensions of harvested skin cells were first gated on ITGA6+ and Lineage negative (CD140a-, CD45-, CD117-, CD31-) followed by enrichment for EpdSC by gating on cells which were Ly6a+, CD34-, while HFSC were Ly6a-, CD34+. Please see extended data figure 2.</p>

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.