SOX9 is sufficient to drive endothelial cells towards a mesenchymal fate by altering the chromatin landscape.

Bettina M. Fuglerud[1,2,3], Sibyl Drissler[1,4], Tabea L. Stephan[1,4], Avinash Thakur[1,2], Rebecca Cullum[1], Pamela A. Hoodless[1,2,5,*]

[1]Terry Fox Laboratory, BC Cancer, Vancouver, British Columbia, Canada.

[2]Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada.

[3]Department of Biosciences, University of Oslo, Oslo, Norway.

[4]Cell and Developmental Biology Program, University of British Columbia, Vancouver, British Columbia, Canada.

[5]School of Biomedical Engineering, University of British Columbia, Vancouver, British Columbia, Canada.

*Corresponding author

Email: hoodless@bccrc.ca

Abstract

The transcription factor SOX9 is expressed in multiple tissues during embryogenesis and directs developmental processes. SOX9 is activated upon epithelial-to-mesenchymal transition (EMT) and the closely related process, endothelial-to-mesenchymal transition (EndMT), but its role in regulating these processes is less clear. Both EMT and EndMT are fundamental processes in normal development and cancer progression. Here, we show that SOX9 expression alone is sufficient to activate mesenchymal enhancers and steer endothelial cells towards a mesenchymal fate. By genome-wide mapping of the chromatin landscape, we show that SOX9 acts as a pioneer transcription factor, having the ability to open the chromatin structure and increase enrichment of active histone marks at a specific subset of target enhancers with an associated SOX motif. SOX9 also displays widespread chromatin pausing that is not associated with SOX motifs. This leads to a switch in enhancer activity states resulting in activation of mesenchymal genes and concurrent suppression of endothelial genes to drive EndMT. Moreover, SOX9 induced widespread changes in chromatin states throughout the genome without SOX9 binding. Our study highlights the crucial developmental role of SOX9 and provides new insight into key molecular functions of SOX9 in regulating the chromatin landscape and mechanisms of EndMT.

Introduction

The SOX family of transcription factors (TFs) contains a highly conserved high mobility group (HMG) domain binding to ACAAa/tG-like DNA sequences. The HMG domain was first identified in *SRY*, a crucial factor for mammalian sex determination (Sinclair et al. 1990; Harley, Lovell-Badge, and Goodfellow 1994). Around 20 SOX proteins with divergent developmental functions have been identified, of which many act as master regulators that control critical transcriptional programs to determine cell fate and differentiation. SOX9 and other SOX factors bind DNA as either a monomer or a homo- or heterodimer and this can alter functionality (Hou, Srivastava, and Jauch 2017). During embryogenesis, SOX9 directs developmental processes including chondrogenesis, gliogenesis, sex determination, as well as development of the heart valves, hair follicles, pancreas, and liver (summarized in (Jo et al. 2014)).

Most knowledge of biological functions of SOX9 has come from developmental studies involving chondrogenesis, which has revealed its role in regulating extracellular matrix (ECM) production and organization (Lincoln et al. 2007; Akiyama 2008). During normal development, SOX9 is crucial for processes involving epithelial-to-mesenchymal transition (EMT) and the closely related process endothelial-to-mesenchymal transition (EndMT) (Lincoln et al. 2007; Akiyama et al. 2004; M. Cheung et al. 2005). Given the significance of EMT as a disease mechanism in cancer progression, SOX9 can drive cancer invasion and metastasis in multiple cancers, including prostate, breast, colon, gastric, and lung (summarized in (Aguilar-Medina et al. 2019)). SOX9 is also involved in promoting liver and cardiac fibrosis by stimulating inappropriate ECM deposition (Lee and Friedman 2011; Lacraz et al. 2017).

At the molecular level, SOX9 has been suggested to alter epigenetic signatures, but most of the work has been done on a single locus in the *COL2A1* enhancer. SOX9 remodeled chromatin and

activated transcription at this site using an *in vitro* assembled chromatin template (Coustry et al. 2010). Using the same enhancer, the histone acetyltransferase, p300, and SOX9 synergistically acetylated histones on the chromatin template, while SOX9 was not required to assist p300 in acetylating free histones (Furumatsu et al. 2005). The latter strongly suggests that SOX9 recruits the histone modifier to the chromatinized template. In chondrogenic systems, SOX9 can activate enhancers (Cheung et al. 2020), but may not be required for the initiation of chromatin changes (C.-F. Liu et al. 2018). In hair follicle stem cells, SOX9 has been put forward as a pioneer TF (Adam et al. 2015); however, this study only showed increased activation of enhancers and did not identify whether SOX9 opens chromatin, a hallmark of pioneer TFs. SOX9 acts as a master regulator in heart valve development, binding regulatory regions that control TFs important for proper valve development (Garside et al. 2015). Thus, the epigenetic requirements of SOX9 to alter chromatin and drive cell fate decisions remains uncertain and may vary between cell types.

Some developmental TFs have the remarkable ability to program one cell type into another by engaging genes that are developmentally silenced and in closed chromatin, thus acting as pioneer TFs to initiate transcriptional events (Iwafuchi-Doi and Zaret 2016). In the present work, we have used Human Umbilical Vein Endothelial Cells (HUVECs) to study cell programming initiated by SOX9 and its role in restructuring the chromatin landscape. Transcriptome sequencing revealed that ectopic expression of SOX9 in endothelial cells is sufficient to induce expression of mesenchymal marker genes. In addition, the cells acquired increased invasive properties. We used the chromatin profiling methods CUT&RUN and CUT&Tag to map SOX9 binding and histone marks (H3K4me1, H3K27ac, and H3K27me3). In addition, we mapped chromatin accessibility by ATAC-seq. Altogether, we find that at a subset of its binding sites in silent chromatin, SOX9 occupancy increases chromatin accessibility and enrichment of active histone

marks to induce gene expression, demonstrating SOX9's ability to act as a pioneer TFs to program cell fate. This pioneer function seems to be motif encoded and occurs predominantly in distal regulatory regions. Furthermore, we show global changes in the chromatin landscape downstream of SOX9, including a switch in enhancers, which may be mediated through other SOX proteins to drive endothelial cells towards a mesenchymal phenotype. Together, our work demonstrates how a single TF can have widespread effects on chromatin structure and downstream cellular phenotypes.

Results

Ectopic expression of SOX9 alters morphology and migratory properties of endothelial cells.

To determine if ectopic SOX9 expression in endothelial cells would force cells to adopt a mesenchymal phenotype, we transduced HUVECs with a construct encoding human SOX9 or an empty vector control, both containing a green fluorescent protein (GFP) gene (Fig. 1A). HUVECs, a recognized EndMT model, lack endogenous SOX9 expression. Expression of SOX9 was confirmed (Fig. 1B) and cellular GFP levels were monitored during growth. Six days after transduction, a change in morphology occurred, from the cobblestone-like shape of endothelial cells to the elongated spindle-shape characteristic of mesenchymal cells (Fig. 1C). This change in morphology is consistent with other studies' observations of EndMT in HUVECs (Yu et al. 2017; Liguori et al. 2019; Noseda et al. 2004). Furthermore, we observed enhanced migratory properties of the HUVECs after SOX9 expression, as shown by transwell assays (Fig. 1D). The above results show that ectopic expression of SOX9 alone promotes a mesenchymal phenotype in endothelial cells.

SOX9 alters expression of genes associated with EndMT.

To investigate global transcriptional changes caused by SOX9, we employed RNA-seq using HUVECs harvested four days after transduction with SOX9 or empty vector. 4806 genes were differentially expressed between the two conditions (log2 ratio >0.5 or <-0.5, p-value <0.05). Of these genes, 2587 genes were upregulated and 2219 genes were downregulated in cells transduced with SOX9 compared to empty vector (Fig. 2A). Among the most upregulated genes were components of ECM, including collagens, as well as mesenchymal markers like Periostin (*POSTN*), N-cadherin (*CDH2*), and Cadherin-11 (*CDH11*). The most downregulated genes

included endothelial markers like von Willebrand factor (*VWF*), VE-cadherin (*CDH5*), and CD31 (*PECAM1*), as well as some key endothelial-associated TFs, such as SOX7, SOX18, ERG and LMO2 (Fig. 2A, right panel). Gene Set Enrichment Analysis (GSEA) revealed that genes upregulated by SOX9 were significantly enriched for hallmark gene sets for EMT (Fig. 2B). Downregulated genes were enriched for several inflammatory signaling pathways, reflecting the role of endothelial cells as regulators of inflammatory responses. Although EMT and EndMT are similar processes, both leading to a mesenchymal phenotype, gene sets for EndMT currently do not exist in gene ontology (GO) databases. Therefore, we created a list of endothelial and mesenchymal genes, including genes shown to be down- and up-regulated upon EndMT (Supplemental Table 1). SOX9 expression resulted in enhanced expression of mesenchymal genes, with concurrent downregulation of endothelial gene expression (Fig. 2C). Together, these results show that SOX9 expression is sufficient to induce EndMT. Of note, we observed a downregulation of some TFs known to have a role in EMT, like SNAI1 and ZEB1, indicating that the upregulation of mesenchymal markers does not depend on these TFs in this context.

CUT&RUN identifies SOX9 direct target genes.

To examine the regulatory role of SOX9 in directing endothelial cells to a mesenchymal fate, we profiled SOX9 chromatin binding genome-wide by CUT&RUN in HUVECs four days after transduction with SOX9 or empty vector. We identified 3381 SOX9 bound regions (Fig. 3A).

Searching for enriched DNA binding motifs in the SOX9 peaks, we identified a valid SOX motif (Fig. 3B). Surprisingly, the SOX motif was listed as the sixth most significantly enriched motif and was only present in 649 (19%) of the SOX9 bound regions. Approximately 6% of all SOX9 bound regions were either directly over a transcriptional start site (TSS) or in the 5 kb proximal promoter regions (Fig. 3C). About 45% of bound regions were located within gene bodies,

whereas 49% were distal (Fig. 3C, lower panel). These data indicate that in HUVECs, SOX9 primarily binds distal regulatory elements inside or outside gene bodies rather than promoter regions of genes.

Of genes associated with SOX9 bound regions (3094 genes total), 867 were differentially expressed (Fig. 3D), 843 were stably expressed, while 1384 genes were not expressed. Genes with both a SOX9 bound region and differential gene expression were considered to be direct transcriptional targets of SOX9. Of these, 543 genes (63%) were upregulated by SOX9 expression, while 324 genes (37%) were downregulated. Among the upregulated genes are previously characterized SOX9 target genes, such as ECM genes encoding collagens IX and XI (Oh et al. 2014), as well as other TFs like SOX5, SOX6, CREB3L2 and NFIA, highlighting the role of SOX9 as a master regulator (Akiyama et al. 2002; Hino et al. 2014; Garside et al. 2015). Of note, we did not observe SOX9 bound to the enhancer of a well-characterized SOX9 target gene, *COL2A1*, although this gene was highly upregulated upon SOX9 expression. Other interesting genes, previously not characterized as SOX9 target genes, encode the Homeobox TFs PRRX1 and PKNOX2, both of which have roles in mesenchyme development (Lu et al. 1999; W. Zhou et al. 2013). PRRX1 can also promote EMT in cancers (Reichert et al. 2013; Takahashi et al. 2013). Genes repressed by SOX9 include endothelial markers, such as von Willebrand factor and VE-Cadherin, as well as TFs important for endothelial cell functions, including ERG (Kalna et al. 2019) and ETS2 (Wei et al. 2009). Taken together, identification of SOX9 direct target genes supports a proximate role for SOX9 in the regulation genes important for EndMT

SOX9 acts as a pioneer TF to program mesenchymal cell fate.

Since SOX9 primarily bound to distal regulatory elements (enhancers), we questioned which chromatin states allow occupancy by SOX9 and if SOX9 itself is able to alter the chromatin

landscape and act as a pioneer TF to program cell fate. To determine the epigenomic impact of SOX9 in HUVECs, we performed ATAC-seq and CUT&Tag for the enhancer marks H3K4me1, H3K27ac, and H3K27me3 on cells four days after transduction with SOX9 or empty vector. We observed SOX9 bound in both open and closed chromatin regions and in regions with and without pre-existing enhancer marks (Fig. 4A). We classified SOX9 bound regions into four distinct clusters, termed c1-c4, determined by chromatin state upon SOX9 expression and in the empty vector control cells. The first cluster, c1, are regions with high chromatin accessibility regardless of SOX9 presence, as in the enhancer upstream of the gene encoding ECM-associated protein CTGF. We also observed high levels of H3K4me1 and H3K27ac in these regions, indicative of an active enhancer, which increased slightly with SOX9 occupancy. c2 regions displayed closed chromatin and had no enrichment of the assayed histone marks prior to SOX9 expression, but with SOX9 expression gained high chromatin accessibility and enrichment of H3K4me1 and H3K27ac. At these newly opened chromatin regions, the changes to the chromatin confirm pioneer TF activity by SOX9. As shown at the *PRRX1* locus, these regions cleanly gained opening of chromatin and addition of active histone marks with SOX9 binding. The c2 cluster included regulatory regions for several other TF genes (Supplemental Fig. 1B). c3 consisted of regions with closed chromatin and no active histone marks but high levels of the repressive mark H3K27me3, as shown in the *FOXR1* locus. Neither chromatin opening, active histone marks nor H3K27me3 levels were affected by SOX9 binding in this cluster. The final c4 cluster, composed of over 50% of SOX9 bound regions, did not show changes to the chromatin with SOX9 binding and was in quiescent chromatin: closed regions with low levels of histone marks. The locus encoding the ECM enzyme MMP2 is an interesting example as it had a c4 type region in addition to a c2 region and this gene was highly upregulated upon SOX9 expression.

As SOX9 binding only resulted in increased chromatin accessibility at a subset of its bound regions, we questioned if DNA sequence might intrinsically function to dictate how SOX9 binding affects chromatin accessibility or histone marks. For c1 and c2, the primary motifs recovered were SOX motifs, but for c3 and c4 no SOX motifs were significantly enriched (Fig. 4B and Supplemental Fig. 1A). The SOX dimer motif was enriched in c1 but was more frequently observed in c2 (9% vs 29%, respectively). Additionally, 47 % of c1 regions and 79% of c2 regions contained one or more SOX monomer motifs. Thus, this data suggests that SOX9 binding to closed chromatin, as in c2-c4, was not dictated by a SOX motif in the DNA sequence, but its ability to increase chromatin accessibility, as in c2, was dependent on the presence of a SOX motif. While all regions were primarily distal, c1 contained more regions closer to TSS/promoter regions than the other clusters (Fig. 4C).

Chromatin changes induced by SOX9 binding in c1 and c2 lead to an overall upregulation of the associated genes (Fig. 4D and Supplemental Fig. 1B). There was also upregulation of the genes associated with regions in c4, while expression of genes associated with c3 regions were not significantly changed. Of note, a large number of the genes associated with c4 regions also had SOX9 binding in c1 or c2 regions, which may explain why these genes were upregulated (Supplemental Fig. 1C). While over 70% of genes with c1 or c2 type SOX9 binding were expressed, only 47% of the genes with c3 type SOX9 binding were expressed in either condition (Fig. 4D).

Comparing the average ATAC and CUT&Tag profiles at SOX9 bound regions associated with upregulated genes, we saw a massive increase of ATAC signal as well as H3K4me1 and H3K27ac. At the regions associated with downregulated genes, there was also an overall increase but less so than the sites associated with upregulated genes. Additionally, upregulated genes with

regions in c1 and c2 were overrepresented compared to downregulated genes (Supplemental Fig. 2A and 2B).

The regions in c1 and c2 were associated with genes specifically important for gene ontology processes connected with mesenchymal cells (Fig. 4E). Notably, we did not identify biological processes associated with the regions in c4, although this is the largest of the four clusters. The latter may indicate that this cluster represents non-specific binding or scanning by SOX9 (Zaret, Lerner, and Iwafuchi-Doi 2016), although SOX9 must be pausing at these sites to permit detection. Taken together, our profiling of the chromatin landscape demonstrate that SOX9 binds closed chromatin and increases chromatin accessibility at a subset of enhancers, demonstrating its role as a pioneer TF. Moreover, these enhancers are crucial for EndMT initiated by SOX9 and their activation leads to upregulation of associated genes.

Major global changes in the epigenetic landscape are induced by SOX9.

To investigate whether there was a global impact on the chromatin landscape regardless of SOX9 binding at a region, we examined the global dynamics of histone marks at enhancers after SOX9 expression (Fig. 5A). We defined all enhancers as being more than 5000 bp upstream or 500 bp downstream from a TSS and classified enhancer states as active (H3K4me1 + H3K27ac), bivalent (also referred to as poised) (H3K4me1 + H3K27me3), primed (H3K4me1 only), quiescent (no marks), or repressed (H3K27me3 only). Based on histone marks alone, SOX9 expression caused major changes in enhancer states; in total 53012 enhancers changed from one state to another upon SOX9 expression. 6592 enhancers that were primed, bivalent, quiescent, or repressed in cells transduced with empty vector gained activation marks in cells with SOX9 expression (H3K4me1 and H3K27ac). By incorporating ATAC-seq, we observed that 2218 (34%) of the newly activated enhancer regions showed increased chromatin accessibility after

SOX9 expression (Fig. 5B). Genes associated with newly activated and newly opened regions were linked with mesenchymal phenotypes, such as collagen-fibril organization and ECM organization, as well as EMT (Fig. 5C).

To analyze the downstream effects of SOX9 on chromatin opening, we subtracted SOX9 bound regions from the entire set that showed a gain in chromatin accessibility. We subdivided the remaining 5315 regions into three clusters, named C1-C3 (Fig. 5D). Regions in C1, for example the *MMP12* locus, included the newly activated enhancers mentioned above with gained H3K4me1 and H3K27ac. Regions in C2 only showed increased H3K4me1 in addition to increased chromatin accessibility and included the known chondrogenic SOX9 target gene *COL2A1*. A region downstream of the proto-oncogene *PIM1* is an example of a region in C3, which only showed increased chromatin accessibility. We also did clustering on sites of decreased chromatin accessibility (Supplemental Fig. 3A). As expected, with a loss of chromatin accessibility, we observed loss of active histone marks (C1' and C2'). Both the regions that gained chromatin accessibility and lost it exclusively occurred in distal regions, which highlights the role of enhancers in defining cell-type specificity (Fig. 5E, 5F, and Supplemental Fig. 3B).

Genes associated with regions of increased chromatin accessibility were significantly upregulated for all three clusters, not only regions with increased active histone marks (Fig. 5G). Genes with decreased chromatin accessibility regions did not show a significant decrease in gene expression (Supplemental Fig. 3C). Interestingly, at regulatory regions for 1835 genes we observed both an increase and a decrease in chromatin accessibility, for instance *PRRX1* (Fig. 4A), and these include both upregulated and downregulated genes (Supplemental Fig. 4). We found enriched GO terms related to mesenchymal cells in all three clusters and EMT for the genes associated with C3, suggesting that all types of increased chromatin accessibility are driving the EndMT

process (Fig. 5H). Some endothelial processes were enriched for the regions with decreased chromatin accessibility (Supplemental Fig. 3D).

We did not observe significant changes in enrichment of the repressive mark H3K27me3 in SOX9 bound regions or regions with changes in chromatin accessibility. However, by focusing on regions with differential H3K27me3 enrichment following SOX9 expression, we noted that the majority (74%) lost H3K27me3 while only 26% gained the repressive mark (Supplemental Figure 5A). Regions with differential H3K27me3 in open chromatin were more likely to be located around TSS/promoter sites, while those in closed chromatin were almost exclusively distal (Supplemental Figure 5B). Of note, neither the active marks nor chromatin accessibility were affected by changes in H3K27me3. With loss of H3K27me3, expression of associated genes increased, however with gain in H3K27me3, expression of associated genes did not change significantly. Notably, only about 50% of all genes marked with H3K27me3 were expressed (Supplemental Figure 5C) and the genes associated with several developmental processes, but none specifically relevant to our studied system (Supplemental Figure 5D).

In summary, although SOX9 directly alters chromatin at a subset of enhancers, we found major changes in the chromatin landscape downstream of SOX9, including a switch in enhancers and opening of chromatin at regions associated with genes involved in promoting EndMT. This highlights the crucial developmental role of SOX9, being able to act as a pioneer TF at a subset of critical sites, frequently targeting other TF genes, which then further results in a global transition of the chromatin landscape to steer cells to their fate.

Other SOX factors may be responsible for chromatin opening downstream of SOX9.

SOX9 was directly responsible for less than 10% of all the chromatin opening events we identified, suggesting other TFs targeted by SOX9 could be responsible for additional chromatin

changes and enhancer activation. After subtraction of all identified SOX9 bound regions, SOX motifs were nonetheless found to be the most significantly enriched in the newly opened chromatin regions (Fig. 6A, upper panel). The core SOX motif is the preferred binding sequence for all 20 SOX TFs (Badis et al. 2009) and many SOX factors can bind DNA as homo- and heterodimers (Lefebvre 2010; Huang et al. 2015). Of the top five significantly enriched motifs in sites with increased chromatin accessibility, two were monomeric SOX sites and three were palindromic SOX dimer motifs. In fact, 87% of all newly opened chromatin without detected SOX9 binding contained a SOX motif.

We also analyzed all the activated enhancers in already open chromatin regions (Fig. 6A, lower panel) and found enrichment for motifs for the AP-1 family of TFs, including JUNB, JUND, and FOSL1. Interestingly, these motifs were also enriched in sites with decreased chromatin accessibility and deactivated enhancers, together with motifs for the endothelial TFs ETS and ERG (Supplemental Fig. 3E). This suggests that SOX factors may specifically be required to open chromatin at mesenchymal enhancers, while other TFs, such as within the AP-1 family, act in already open chromatin regions.

Although several SOX factors are expressed in HUVECs, SOX2, SOX5, and SOX6 were not expressed prior to SOX9 transduction (Fig. 6B). Of interest, the *SOX6* locus had eight SOX9 bound regions, consistent with the high upregulation of this gene. Three regions showed chromatin opening upon SOX9 binding (Fig. 6C, upper right panel). We also observed SOX9 chromatin opening and enhancer activation upstream of the *SOX2* locus and within the *SOX5* locus. The combination of SOX5, SOX6, and SOX9 is known to cooperate to drive chondrogenic differentiation, and SOX9 is essential for SOX5 and SOX6 expression in chondrogenesis (Ikeda et al. 2004; Akiyama et al. 2002). SOX5 and SOX6 have not been reported to open chromatin,

while SOX2 is a known pioneer TF (Iwafuchi-Doi 2019). SOX2 can also promote EMT and cancer metastasis in cancers (X. Liu et al. 2018; Li et al. 2013). Thus, SOX9 acts as a master regulator and pioneer TF upstream of these factors to activate their enhancers, and these may in turn act downstream of SOX9 to open chromatin and activate enhancers, which conjunctly drives the transition from an endothelial to a mesenchymal cell fate.

Discussion

Cell programming in embryonic development is crucial for cell fate and often involves combinations of TFs. Certain TFs, termed pioneer TFs, initiate cell programming events by targeting silent chromatin and TF genes and are thus among the master regulators of cell fate (Iwafuchi-Doi and Zaret 2016). Previous work has shown that SOX9 is essential during cushion formation in cardiac valve development, which involves EndMT, and where SOX9 activates transcription of key regulators of heart valve development (Akiyama et al. 2004; Garside et al. 2015). However, SOX9's role in inducing EndMT has not been evident. In this study, we used human endothelial cells, HUVECs, to examine whether ectopic expression of SOX9 can program a mesenchymal cell fate by altering the chromatin landscape. By using genome-wide techniques, we show that at a subset of enhancers in HUVECs, SOX9 functions as a pioneer TF by opening chromatin and increasing active histone marks. The activation of these enhancers was sufficient for the upregulation of mesenchymal genes and for cells to undergo EndMT, highlighting the ability of one single TF to induce major chromatin restructuring and cell fate decisions.

Collectively, our analyses provide evidence that ectopic SOX9 expression in HUVECs can drive EndMT and transition cells to a mesenchyme state. SOX9 inducted expression of a plethora of TFs associated with mesenchyme, including PRRX1, PKNOX2, FOXC1, and other SOX TFs. Of interest, some TFs with known functions in EndMT/EMT, such as SNAI1 and ZEB1, were downregulated in our dataset. SNAI1 and ZEB1 are upregulated by SOX9 upon EMT in gastric carcinoma cells and mammary stem cells, respectively (C.-J. Zhou et al. 2011; Kogata et al. 2018). However, these TFs were not directly targeted by SOX9 DNA binding in the HUVECs. Another factor promoting EMT, SNAI2, was not expressed in our datasets. Interestingly, both SNAI1 and SNAI2 are required for EndMT in cardiac cushion morphogenesis in mice (Niessen

et al. 2008). Thus, our data suggests that these factors function upstream of SOX9 and that SOX9 can initiate EndMT independently of SNAI1 and SNAI2.

Mapping of SOX9 binding and the chromatin landscape identified different modes of SOX9 binding (Fig. 3 and 4, summarized in Fig. 7). Surprisingly, very few (less than 10%) SOX9 bound regions localized to open chromatin regions already present in HUVECs, and were rarely found in promoter regions. Previous studies of SOX9 genome-wide chromatin binding have found that about 20-30% of SOX9 bound regions occurred within TSS/proximal promoter sites in developing heart valves and limbs in mice and in mouse chondrocytes (Garside et al. 2015; Ohba et al. 2015). In mouse chondrocytes, a SOX motif was enriched in distal regions while promoter regions were enriched for motifs that suggested binding of SOX9 was mediated through the basal transcriptional machinery (Ohba et al. 2015). Similarly, in the c1 cluster, where chromatin was open prior to SOX9 binding, the SOX motif was less enriched, suggesting that at these sites, SOX9 may be binding to other proteins and not requiring its own motif to bind DNA. Furthermore, as previous studies of SOX9 genomic binding assayed endogenous SOX9, SOX9 recruitment to promoters may be dependent on factors which are absent in HUVECs.

Pioneer TFs have the unique ability to engage their target sites in closed chromatin, render them accessible and allow other factors to bind (Iwafuchi-Doi 2019). We show that SOX9 harbours this ability but, as observed with other pioneer TFs (Fuglerud et al. 2017; Mayran et al. 2018), not all SOX9 binding events result in opening of chromatin. The SOX motif seems to work as an intrinsic signal to guide SOX9 pioneer activity, as pioneering events occurred in regions with SOX dimer and/or monomer motifs. The pluripotency factors OCT4, SOX2 and KLF4,when acting as pioneer TFs, bind nucleosomes *in vitro* by targeting partial DNA motifs displayed on the nucleosome surface (Soufi et al. 2015). According to our data, when SOX9 is acting as a

pioneer TF, it requires a full motif, supporting studies showing that SOX9 chromatin remodeling of nucleosomes assembled *in vitro* was abolished when the SOX motif was disrupted (Coustry et al. 2010). At sites of SOX9 binding and chromatin opening, we also observed enrichment of the active enhancer marks H3K4me1 and H3K27ac. SOX9 can recruit the histone acetyltransferase p300 to chromatin (Furumatsu et al. 2005), but whether SOX9 also recruits histone methyltransferases to target sites remains unknown. H3K27ac is acquired almost exclusively in the context of pre-existing H3K4me1 at enhancers (Bonn et al. 2012), which suggests that SOX9 may indeed recruit a H3K4 monomethyltransferase to its pioneering target sites. The precise molecular mechanisms underlying pioneer function remain undetermined for most pioneer TFs. Although binding to closed chromatin is a unique property of pioneer TFs, whether pioneer TFs are able to perform chromatin opening fully independently of other factors is still a matter of debate, and the possibility that chromatin remodelers are recruited to sites of pioneer TF binding cannot be ruled out.

Pioneer TF chromatin binding exhibits the potential for low-affinity genome-wide interactions, which may indicate a scanning mechanism (Zaret, Lerner, and Iwafuchi-Doi 2016). The SOX9 bound regions observed in quiescent chromatin that lack changes in the chromatin landscape upon SOX9 binding may represent scanning. The lower average peak signal in the SOX9 data at these regions suggests that these are low-affinity binding events (Supplemental Fig. 6). SOX9 binding in c3, with high H3K27me3, also had a lower signal compared to c1 and c2, suggesting these events may also represent chromatin scanning rather than specific binding by SOX9. Since our analysis was performed on a pool of cells, enrichment of SOX9 at the same site suggests there must be other qualities to sites in c3 and c4 that made SOX9 pause at these regions, such as the presence of a histone mark not assayed or other chromatin bound factors. Together, the

different modes of SOX9 chromatin binding and impact on chromatin resulted in activation of genes driving a mesenchymal cell fate and downregulation of genes important for the endothelial cell identity (Fig. 7).

Although we only observed a direct impact of SOX9 on the chromatin landscape in ~10% of bound regions, the downstream effects on chromatin were profound. After removing SOX9 bound regions, our motif analysis suggested other SOX factors may act as pioneer TFs downstream of SOX9. SOX2, which was upregulated by SOX9 in HUVECs, functions as a pioneer TF in induced pluripotent stem cell reprogramming (Soufi, Donahue, and Zaret 2012), where it forms a heterodimer with OCT4. OCT4 was not expressed in our dataset but SOX2 can also direct chromatin opening independently of OCT4 (Malik et al. 2019). SOX6 was one of the most highly upregulated TFs and we uncovered multiple enhancers activated by SOX9 in and around the *SOX6* locus (Fig. 6C). In chondrogenesis, SOX9 activates transcription of SOX5 and SOX6, which are suggested to secure efficient chondrogenesis by stabilizing SOX9 binding to DNA (Han and Lefebvre 2008). SOX9 and SOX6 have also been found to target super-enhancers in chondrocytes (C.-F. Liu and Lefebvre 2015). In the latter study, SOX6 was shown to target tandem SOX dimer motifs, whereas SOX9 preferred inverted SOX dimer motifs, like the ones we identified in newly accessible sites upon SOX9 expression (Fig. 6A). Motif preference may depend on context, but based on the inverted dimer presence, SOX9 may be opening the previously silent chromatin sites prior to cell harvesting and then evacuating. Pioneering may be viewed as a one-time event, meaning that SOX9 does not remain bound at all sites opened by SOX9.

We observed almost an equal number of enhancers with decreased accessibility as increased accessibility in HUVECs after SOX9 expression, demonstrating a switch in enhancers initiated

by SOX9. The enhancers that are switched off are highly enriched for motifs targeted by AP-1 family TFs and ETS family TFs, many of which are highly expressed in HUVECs (Supplemental Fig. 3E). Some of these, including *FOSL1*, *FOSL2*, *JUN*, *JUNB*, *ETS1*, *ETS2*, and *ERG*, were downregulated upon SOX9 expression and we found SOX9 bound regions associated with the *JUN, ETS1, ETS2,* and *ERG* loci, suggesting these were direct repressive targets of SOX9. ERG has been found to regulate super-enhancers to promote endothelial homeostasis, while ETS1 has been shown to remodel nucleosomes during T-cell differentiation (Kalna et al. 2019; Cauchy et al. 2016). It is conceivable that ETS family TFs maintain active enhancers in HUVECs and the downregulation of these factors leads to subsequent loss of chromatin accessibility at their target enhancers. The presence of AP-1 TF motifs in the activated enhancers with stable chromatin accessibility also suggested that these may function as transcriptional regulators without impacting the chromatin accessibility in HUVECs (Fig. 6A).

The requirement for SOX9 in regulating the chromatin landscape may differ between cellular systems (Adam et al. 2015; C.-F. Liu et al. 2018). For instance, SOX9 may not be crucial for epigenetic changes during chondrogenesis, but our data shows that expression of SOX9 alone is sufficient to remodel the chromatin landscape for induction of EndMT. In mouse embryonic heart valve development, endocardial cells undergo EndMT at embryonic day 9.5, coinciding with SOX9 expression (Akiyama et al. 2004). Although the requirement of SOX9 for inducing EndMT in heart valve development is not clear, our results indicate that it could be sufficient for this transition. In adults, tissue damage may stimulate EndMT to give rise to fibroblasts during wound healing or fibrotic diseases. Furthermore, EndMT forms cancer-associated fibroblasts in the tumor microenvironment, which regulates disease progression (Lin, Wang, and Zhang 2012).

Thus, SOX9 may be a potential driver of disease mechanisms when aberrantly expressed in endothelial cells.

Materials and Methods

Cell culture

Primary Human Umbilical Vein Endothelial Cells (HUVECs, Thermo Fisher Scientific) were maintained in Medium 200 supplemented with 2% low serum growth supplement (LSGS, Thermo Fisher Scientific), with penicillin and streptomycin. HUVECs were plated in 6-well plates and lentiviral supernatants were titrated to transduce >95% of the cells (indicated by GFP expression). The media was replaced with fresh Medium 200 24 hours after transduction and cells were collected three days later or seeded for migration assays. SOX9 overexpression was confirmed by western blotting (described in Supplemental Materials and Methods).

Migration assays

Transwell invasion assays were carried out according to manufacturer's recommendation. Briefly, 15,000 HUVECs were seeded in the upper chamber of 5 µm 24-well polycarbonate transwell inserts (Corning). HUVECs were maintained for 24 hours under normal culture conditions. After 24 hours, the non-migratory cells were removed from the upper chamber with a cotton swab. The membrane was washed with PBS, fixed with 4% formaldehyde for 10 minutes and stained with DAPI nuclear stain for 10 minutes. The lower side of the membrane was photographed with a Zeiss AxioImager fluorescence microscope. Cells that had migrated through the membrane were counted using Fiji/ImageJ software (http://fiji.sc/). To control for cell death and proliferation, SOX9 and empty vector transduced HUVECs of equal density were seeded in 24-wells and monitored for differences after 24h.

RNA-seq and analysis

RNA was extracted from cells using TRIzol (Thermo Fisher Scientific). Three biological

replicates were sequenced by the Biomedical Research Centre (BRC) Sequencing Core at the University of British Columbia (UBC). Reads were aligned to Hg19/ GRCh37 using STAR (Dobin et al. 2013). Fragments per kb of exon per million reads (FPKMs) were calculated using Cufflinks (Trapnell et al. 2012). Differential expression was determined by log2 ratio between SOX9 overexpression and empty vector FPKM values. For downstream analyses, only genes with log2 ratio > 0.5 or < -0.5 were included (p-value < 0.05). To determine biological functions and pathways we used Hallmark gene sets (Liberzon et al. 2015).

CUT&RUN

CUT&RUN was performed as described (Skene, Henikoff, and Henikoff 2018). Briefly, 500,000 cells per experiment were harvested and washed with Wash Buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine and protease inhibitors), bound to Concanavalin A-coated magnetic beads and incubated with SOX9 antibody (AB5535, Millipore) diluted in wash buffer containing 0.05% digitonin (Dig-Wash) overnight at 4 °C. Cells were washed and incubated with Guinea Pig anti-Rabbit secondary antibody (Novus Biologicals) diluted in Dig-Wash for 1 hour at room temperature. Cells were washed again and incubated with CUTANA™ pAG-MNase (Epicypher) for 10 minutes at room temperature (RT). Slurry was washed again and placed on ice and incubated with Dig-Wash containing 2 mM $CaCl_2$ for 30 min to activate digestion. Stop buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.05% Digitonin, 0.05 mg/mL glycogen, 5 µg/mL RNase A) was added to stop the reaction, and fragments were released by 30 min incubation at 37 °C. DNA was extracted with phenol-chloroform and ethanol precipitation. Libraries were constructed and sequenced by the BRC Sequencing Core at UBC. See Supplemental Material for detailed information about data analysis.

CUT&Tag

CUT&Tag was performed as described (Kaya-Okur et al. 2019) and pA-Tn5 adapter complex was kindly provided by Steven Henikoff (Fred Hutchinson Cancer Research Center). Cells were harvested, washed, and incubated with primary and secondary antibodies as described for CUT&RUN. Cells were washed again and incubated with pA-Tn5 adapter complex diluted in Dig-300 buffer (0.01% Digitonin, 20 mM HEPES, pH 7.5, 300 mM NaCl, 0.5 mM Spermidine, protease inhibitors) for 1 hour at RT. Cells were washed and incubated with Dig-300 buffer containing 10 mM $MgCl_2$ for 1 hour at 37°C. To stop tagmentation, 16.6 mM EDTA, 0.1% SDS, and 50 µg Proteinase K was added and incubated at 50 °C for 1 hour. DNA was extracted with phenol-chloroform and ethanol precipitation. Libraries were generated using Ad1_noMX and Ad2.3–2.8 barcoded primers from (Buenrostro et al. 2013) and amplified for 14 cycles. DNA purification was carried out using sparQ PureMag Beads (Quantabio). Libraries were sequenced by the BRC Sequencing Core at UBC. Primary antibodies against the following proteins were applied: H3K4me1 (C15410037, Diagenode), H3K27ac (ab4729, Abcam), and H3K27me3 (C15410069, Diagenode). See Supplemental Material for detailed information about data analysis.

ATAC-seq

ATAC-seq was performed as described (Buenrostro et al. 2013) with small modifications. 100,000 cells were used per experiment. Libraries were generated using Ad1_noMX and Ad2.1–2.2 barcoded primers from (Buenrostro et al. 2013) and amplified for 10 cycles. DNA was purified using sparQ PureMag Beads (Quantabio) following the manufacturer's protocol for size selection and library preparation. Libraries were sequenced by the BRC Sequencing Core at UBC. See Supplemental Material for detailed information about data analysis.

Data availability

High-throughput sequencing data is publicly available through GEO (accession numbers: GSE155282 and GSE155290).

Acknowledgements

We are grateful to Steven Henikoff for providing the pA-Tn5 adapter complex for CUT&Tag experiments. B.M.F. is financed by the Research Council of Norway FRIPRO Mobility Research Grant (FRIPRO 275783). P.A.H. receives funding from the Canadian Institutes for Health Research (Funding reference number: 159512).

Author contributions

B.M.F. and P.A.H. planned the experiments. T.L.S. performed transwell migration assays, all other experiments were performed by B.M.F. S.D. performed alignments, processing, and peak-calling on CUT&RUN, CUT&Tag and ATAC-seq data. B.M.F. analysed the data. All authors contributed to the writing of the manuscript. B.M.F. and A.T. made the figures.

References

Adam RC, H Yang, S Rockowitz, SB Larsen, M Nikolova, DS Oristian, L Polak, M Kadaja, A Asare, D Zheng, et al. 2015. Pioneer Factors Govern Super-Enhancer Dynamics in Stem Cell Plasticity and Lineage Choice. *Nature* **521**: 366–70. doi:10.1038/nature14289.

Aguilar-Medina M, M Avendaño-Félix, E Lizárraga-Verdugo, M Bermúdez, JG Romero-Quintana, R Ramos-Payan, E Ruíz-García, and C López-Camarillo. 2019. SOX9 Stem-Cell Factor: Clinical and Functional Relevance in Cancer. *Journal of Oncology* **2019**: 1–16. doi:10.1155/2019/6754040.

Akiyama H. 2008. Control of Chondrogenesis by the Transcription Factor Sox9. *Modern Rheumatology* **18**: 213–19. doi:10.1007/s10165-008-0048-x.

Akiyama H, M-C Chaboissier, RR Behringer, DH Rowitch, A Schedl, JA Epstein, and B de Crombrugghe. 2004. Essential Role of Sox9 in the Pathway That Controls Formation of Cardiac Valves and Septa. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 6502–7. doi:10.1073/pnas.0401711101.

Akiyama H, M-C Chaboissier, JF Martin, A Schedl, and B de Crombrugghe. 2002. The Transcription Factor Sox9 Has Essential Roles in Successive Steps of the Chondrocyte Differentiation Pathway and Is Required for Expression of Sox5 and Sox6. *Genes & Development* **16**: 2813–28. doi:10.1101/gad.1017802.

Badis G, MF Berger, AA Philippakis, S Talukder, AR Gehrke, SA Jaeger, ET Chan, G Metzler, A Vedenko, X Chen, et al. 2009. Diversity and Complexity in DNA Recognition by Transcription Factors. *Science (New York, N.Y.)* **324**: 1720–23.

doi:10.1126/science.1162327.

Bonn S, RP Zinzen, C Girardot, EH Gustafson, A Perez-Gonzalez, N Delhomme, Y Ghavi-Helm, B Wilczyński, A Riddell, and EEM Furlong. 2012. Tissue-Specific Analysis of Chromatin State Identifies Temporal Signatures of Enhancer Activity during Embryonic Development. *Nature Genetics* **44**: 148–56. doi:10.1038/ng.1064.

Buenrostro JD, PG Giresi, LC Zaba, HY Chang, and WJ Greenleaf. 2013. Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position. *Nature Methods* **10**: 1213–18. doi:10.1038/nmeth.2688.

Cauchy P, MA Maqbool, J Zacarias-Cabeza, L Vanhille, F Koch, R Fenouil, M Gut, I Gut, MA Santana, A Griffon, et al. 2016. Dynamic Recruitment of Ets1 to Both Nucleosome-Occupied and -Depleted Enhancer Regions Mediates a Transcriptional Program Switch during Early T-Cell Differentiation. *Nucleic Acids Research* **44**: 3567–85. doi:10.1093/nar/gkv1475.

Cheung K, MJ Barter, J Falk, CJ Proctor, LN Reynard, and DA Young. 2020. Histone ChIP-Seq Identifies Differential Enhancer Usage during Chondrogenesis as Critical for Defining Cell-Type Specificity. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology* **34**: 5317–31. doi:10.1096/fj.201902061RR.

Cheung M, M-C Chaboissier, A Mynett, E Hirst, A Schedl, and J Briscoe. 2005. The Transcriptional Control of Trunk Neural Crest Induction, Survival, and Delamination. *Developmental Cell* **8**: 179–92. doi:10.1016/j.devcel.2004.12.010.

Coustry F, C Oh, T Hattori, SN Maity, B de Crombrugghe, and H Yasuda. 2010. The

Dimerization Domain of SOX9 Is Required for Transcription Activation of a Chondrocyte-Specific Chromatin DNA Template. *Nucleic Acids Research* **38**: 6018–28. doi:10.1093/nar/gkq417.

Dobin A, CA Davis, F Schlesinger, J Drenkow, C Zaleski, S Jha, P Batut, M Chaisson, and TR Gingeras. 2013. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics (Oxford, England)* **29**: 15–21. doi:10.1093/bioinformatics/bts635.

Fuglerud BM, RB Lemma, P Wanichawan, AYM Sundaram, R Eskeland, and OS Gabrielsen. 2017. A C-Myb Mutant Causes Deregulated Differentiation Due to Impaired Histone Binding and Abrogated Pioneer Factor Function. *Nucleic Acids Research* **45**: 7681–96. doi:10.1093/nar/gkx364.

Furumatsu T, M Tsuda, K Yoshida, N Taniguchi, T Ito, M Hashimoto, T Ito, and H Asahara. 2005. Sox9 and P300 Cooperatively Regulate Chromatin-Mediated Transcription. *The Journal of Biological Chemistry* **280**: 35203–8. doi:10.1074/jbc.M502409200.

Garside VC, R Cullum, O Alder, DY Lu, R Vander Werff, M Bilenky, Y Zhao, SJM Jones, MA Marra, TM Underhill, et al. 2015. SOX9 Modulates the Expression of Key Transcription Factors Required for Heart Valve Development. *Development (Cambridge, England)* **142**: 4340–50. doi:10.1242/dev.125252.

Han Y, and V Lefebvre. 2008. L-Sox5 and Sox6 Drive Expression of the Aggrecan Gene in Cartilage by Securing Binding of Sox9 to a Far-Upstream Enhancer. *Molecular and Cellular Biology* **28**: 4999–5013. doi:10.1128/MCB.00695-08.

Harley VR, R Lovell-Badge, and PN Goodfellow. 1994. Definition of a Consensus DNA Binding Site for SRY. *Nucleic Acids Research* **22**: 1500–1501. doi:10.1093/nar/22.8.1500.

Hino K, A Saito, M Kido, S Kanemoto, R Asada, T Takai, M Cui, X Cui, and K Imaizumi. 2014. Master Regulator for Chondrogenesis, Sox9, Regulates Transcriptional Activation of the Endoplasmic Reticulum Stress Transducer BBF2H7/CREB3L2 in Chondrocytes. *The Journal of Biological Chemistry* **289**: 13810–20. doi:10.1074/jbc.M113.543322.

Hou L, Y Srivastava, and R Jauch. 2017. Molecular Basis for the Genome Engagement by Sox Proteins. *Seminars in Cell & Developmental Biology* **63**: 2–12. doi:10.1016/j.semcdb.2016.08.005.

Huang Y-H, A Jankowski, KSE Cheah, S Prabhakar, and R Jauch. 2015. SOXE Transcription Factors Form Selective Dimers on Non-Compact DNA Motifs through Multifaceted Interactions between Dimerization and High-Mobility Group Domains. *Scientific Reports* **5**: 10398. doi:10.1038/srep10398.

Ikeda T, S Kamekura, A Mabuchi, I Kou, S Seki, T Takato, K Nakamura, H Kawaguchi, S Ikegawa, and U Chung. 2004. The Combination of SOX5, SOX6, and SOX9 (the SOX Trio) Provides Signals Sufficient for Induction of Permanent Cartilage. *Arthritis and Rheumatism* **50**: 3561–73. doi:10.1002/art.20611.

Iwafuchi-Doi M. 2019. The Mechanistic Basis for Chromatin Regulation by Pioneer Transcription Factors. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine* **11**: e1427. doi:10.1002/wsbm.1427.

Iwafuchi-Doi M, and KS Zaret. 2016. Cell Fate Control by Pioneer Transcription Factors. *Development (Cambridge, England)* **143**: 1833–37. doi:10.1242/dev.133900.

Jo A, S Denduluri, B Zhang, Z Wang, L Yin, Z Yan, R Kang, LL Shi, J Mok, MJ Lee, et al. 2014. The Versatile Functions of Sox9 in Development, Stem Cells, and Human Diseases.

Genes & Diseases **1**: 149–61. doi:10.1016/j.gendis.2014.09.004.

Kalna V, Y Yang, CR Peghaire, K Frudd, R Hannah, A V Shah, L Osuna Almagro, JJ Boyle, B Göttgens, J Ferrer, et al. 2019. The Transcription Factor ERG Regulates Super-Enhancers Associated With an Endothelial-Specific Gene Expression Program. *Circulation Research* **124**: 1337–49. doi:10.1161/CIRCRESAHA.118.313788.

Kaya-Okur HS, SJ Wu, CA Codomo, ES Pledger, TD Bryson, JG Henikoff, K Ahmad, and S Henikoff. 2019. CUT&Tag for Efficient Epigenomic Profiling of Small Samples and Single Cells. *Nature Communications* **10**: 1930. doi:10.1038/s41467-019-09982-5.

Kogata N, P Bland, M Tsang, E Oliemuller, A Lowe, and BA Howard. 2018. Sox9 Regulates Cell State and Activity of Embryonic Mouse Mammary Progenitor Cells. *Communications Biology* **1**: 228. doi:10.1038/s42003-018-0215-3.

Lacraz GPA, JP Junker, MM Gladka, B Molenaar, KT Scholman, M Vigil-Garcia, D Versteeg, H de Ruiter, MW Vermunt, MP Creyghton, et al. 2017. Tomo-Seq Identifies SOX9 as a Key Regulator of Cardiac Fibrosis During Ischemic Injury. *Circulation* **136**: 1396–1409. doi:10.1161/CIRCULATIONAHA.117.027832.

Lee UE, and SL Friedman. 2011. Mechanisms of Hepatic Fibrogenesis. *Best Practice & Research Clinical Gastroenterology* **25**: 195–206. doi:https://doi.org/10.1016/j.bpg.2011.02.005.

Lefebvre V. 2010. The SoxD Transcription Factors--Sox5, Sox6, and Sox13--Are Key Cell Fate Modulators. *The International Journal of Biochemistry & Cell Biology* **42**: 429–32. doi:10.1016/j.biocel.2009.07.016.

Li X, Y Xu, Y Chen, S Chen, X Jia, T Sun, Y Liu, X Li, R Xiang, and N Li. 2013. SOX2

Promotes Tumor Metastasis by Stimulating Epithelial-to-Mesenchymal Transition via Regulation of WNT/β-Catenin Signal Network. *Cancer Letters* **336**: 379–89. doi:https://doi.org/10.1016/j.canlet.2013.03.027.

Liberzon A, C Birger, H Thorvaldsdóttir, M Ghandi, JP Mesirov, and P Tamayo. 2015. The Molecular Signatures Database (MSigDB) Hallmark Gene Set Collection. *Cell Systems* **1**: 417–25. doi:10.1016/j.cels.2015.12.004.

Liguori TTA, GR Liguori, LFP Moreira, and MC Harmsen. 2019. Adipose Tissue-Derived Stromal Cells' Conditioned Medium Modulates Endothelial-Mesenchymal Transition Induced by IL-1β/TGF-B2 but Does Not Restore Endothelial Function. *Cell Proliferation* **52**: e12629. doi:10.1111/cpr.12629.

Lin F, N Wang, and T-C Zhang. 2012. The Role of Endothelial-Mesenchymal Transition in Development and Pathological Process. *IUBMB Life* **64**: 717–23. doi:10.1002/iub.1059.

Lincoln J, R Kist, G Scherer, and KE Yutzey. 2007. Sox9 Is Required for Precursor Cell Expansion and Extracellular Matrix Organization during Mouse Heart Valve Development. *Developmental Biology* **305**: 120–32. doi:10.1016/j.ydbio.2007.02.002.

Liu C-F, M Angelozzi, A Haseeb, and V Lefebvre. 2018. SOX9 Is Dispensable for the Initiation of Epigenetic Remodeling and the Activation of Marker Genes at the Onset of Chondrogenesis. *Development (Cambridge, England)* **145**. doi:10.1242/dev.164459.

Liu C-F, and V Lefebvre. 2015. The Transcription Factors SOX9 and SOX5/SOX6 Cooperate Genome-Wide through Super-Enhancers to Drive Chondrogenesis. *Nucleic Acids Research* **43**: 8183–8203. doi:10.1093/nar/gkv688.

Liu X, B Qiao, T Zhao, F Hu, AK-Y Lam, and Q Tao. 2018. Sox2 Promotes Tumor

Aggressiveness and Epithelial-mesenchymal Transition in Tongue Squamous Cell Carcinoma. *International Journal of Molecular Medicine* **42**: 1418–26. doi:10.3892/ijmm.2018.3742.

Lu MF, HT Cheng, MJ Kern, SS Potter, B Tran, TG Diekwisch, and JF Martin. 1999. Prx-1 Functions Cooperatively with Another Paired-Related Homeobox Gene, Prx-2, to Maintain Cell Fates within the Craniofacial Mesenchyme. *Development (Cambridge, England)* **126**: 495–504.

Malik V, L V Glaser, D Zimmer, S Velychko, M Weng, M Holzner, M Arend, Y Chen, Y Srivastava, V Veerapandian, et al. 2019. Pluripotency Reprogramming by Competent and Incompetent POU Factors Uncovers Temporal Dependency for Oct4 and Sox2. *Nature Communications* **10**: 3477. doi:10.1038/s41467-019-11054-7.

Mayran A, K Khetchoumian, F Hariri, T Pastinen, Y Gauthier, A Balsalobre, and J Drouin. 2018. Pioneer Factor Pax7 Deploys a Stable Enhancer Repertoire for Specification of Cell Fate. *Nature Genetics* **50**: 259–69. doi:10.1038/s41588-017-0035-2.

Niessen K, Y Fu, L Chang, PA Hoodless, D McFadden, and A Karsan. 2008. Slug Is a Direct Notch Target Required for Initiation of Cardiac Cushion Cellularization. *The Journal of Cell Biology* **182**: 315–25. doi:10.1083/jcb.200710067.

Noseda M, G McLean, K Niessen, L Chang, I Pollet, R Montpetit, R Shahidi, K Dorovini-Zis, L Li, B Beckstead, et al. 2004. Notch Activation Results in Phenotypic and Functional Changes Consistent with Endothelial-to-Mesenchymal Transformation. *Circulation Research* **94**: 910–17. doi:10.1161/01.RES.0000124300.76171.C9.

Oh C, Y Lu, S Liang, Y Mori-Akiyama, D Chen, B de Crombrugghe, and H Yasuda. 2014.

SOX9 Regulates Multiple Genes in Chondrocytes, Including Genes Encoding ECM Proteins, ECM Modification Enzymes, Receptors, and Transporters. *PloS One* **9**: e107577. doi:10.1371/journal.pone.0107577.

Ohba S, X He, H Hojo, and AP McMahon. 2015. Distinct Transcriptional Programs Underlie Sox9 Regulation of the Mammalian Chondrocyte. *Cell Reports* **12**: 229–43. doi:10.1016/j.celrep.2015.06.013.

Reichert M, S Takano, J von Burstin, S-B Kim, J-S Lee, K Ihida-Stansbury, C Hahn, S Heeg, G Schneider, AD Rhim, et al. 2013. The Prrx1 Homeodomain Transcription Factor Plays a Central Role in Pancreatic Regeneration and Carcinogenesis. *Genes & Development* **27**: 288–300. doi:10.1101/gad.204453.112.

Sinclair AH, P Berta, MS Palmer, JR Hawkins, BL Griffiths, MJ Smith, JW Foster, AM Frischauf, R Lovell-Badge, and PN Goodfellow. 1990. A Gene from the Human Sex-Determining Region Encodes a Protein with Homology to a Conserved DNA-Binding Motif. *Nature* **346**: 240–44. doi:10.1038/346240a0.

Skene PJ, JG Henikoff, and S Henikoff. 2018. Targeted in Situ Genome-Wide Profiling with High Efficiency for Low Cell Numbers. *Nature Protocols* **13**: 1006–19. doi:10.1038/nprot.2018.015.

Soufi A, G Donahue, and KS Zaret. 2012. Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome. *Cell* **151**: 994–1004. doi:10.1016/j.cell.2012.09.045.

Soufi A, MF Garcia, A Jaroszewicz, N Osman, M Pellegrini, and KS Zaret. 2015. Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate

Reprogramming. *Cell* **161**: 555–68. doi:10.1016/j.cell.2015.03.017.

Takahashi Y, G Sawada, J Kurashige, R Uchi, T Matsumura, H Ueo, Y Takano, S Akiyoshi, H Eguchi, T Sudo, et al. 2013. Paired Related Homoeobox 1, a New EMT Inducer, Is Involved in Metastasis and Poor Prognosis in Colorectal Cancer. *British Journal of Cancer* **109**: 307–11. doi:10.1038/bjc.2013.339.

Trapnell C, A Roberts, L Goff, G Pertea, D Kim, DR Kelley, H Pimentel, SL Salzberg, JL Rinn, and L Pachter. 2012. Differential Gene and Transcript Expression Analysis of RNA-Seq Experiments with TopHat and Cufflinks. *Nature Protocols* **7**: 562–78. doi:10.1038/nprot.2012.016.

Wei G, R Srinivasan, CZ Cantemir-Stone, SM Sharma, R Santhanam, M Weinstein, N Muthusamy, AK Man, RG Oshima, G Leone, et al. 2009. Ets1 and Ets2 Are Required for Endothelial Cell Survival during Embryonic Angiogenesis. *Blood* **114**: 1123–30. doi:10.1182/blood-2009-03-211391.

Yu C-H, Suriguga, M Gong, W-J Liu, N-X Cui, Y Wang, X Du, and Z-C Yi. 2017. High Glucose Induced Endothelial to Mesenchymal Transition in Human Umbilical Vein Endothelial Cell. *Experimental and Molecular Pathology* **102**: 377–83. doi:10.1016/j.yexmp.2017.03.007.

Zaret KS, J Lerner, and M Iwafuchi-Doi. 2016. Chromatin Scanning by Dynamic Binding of Pioneer Factors. *Molecular Cell* **62**: 665–67. doi:10.1016/j.molcel.2016.05.024.

Zhou C-J, J-Q Guo, K-X Zhu, Q-H Zhang, C-R Pan, W-H Xu, H-J Wang, and B Liu. 2011. Elevated Expression of SOX9 Is Related with the Progression of Gastric Carcinoma. *Diagnostic Cytopathology* **39**: 105–9. doi:10.1002/dc.21348.

Zhou W, H Zhu, J Zhao, H Li, Y Wan, J Cao, H Zhao, J Yu, R Zhou, Y Yao, et al. 2013. Misexpression of Pknox2 in Mouse Limb Bud Mesenchyme Perturbs Zeugopod Development and Deltoid Crest Formation. *PloS One* **8**: e64237. doi:10.1371/journal.pone.0064237.

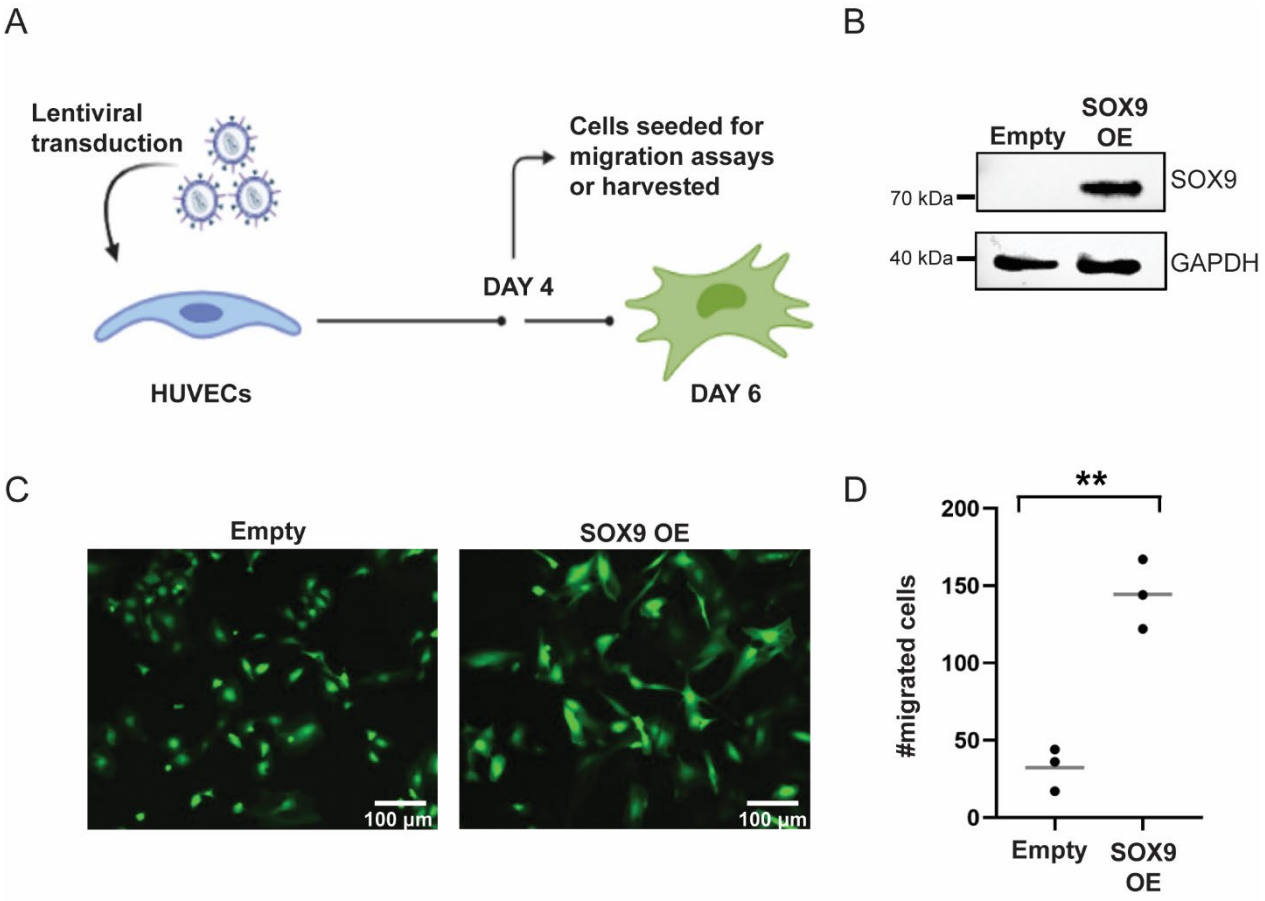


Figure 1. SOX9 causes morphological changes and increased migratory properties of HUVECs.

A) *Schematic overview of experimental setup.* ***B)*** *Expression of SOX9 (SOX9 OE) confirmed by western blotting.* ***C)*** *Fluorescence microscopy of the HUVECs six days after transduction with SOX9 or empty vector plasmid, both containing a GFP gene, demonstrating the change in morphology upon SOX9 expression. Scale bars, 100 µm.* ***D)*** *Indicated number of HUVECs migrated after transduction with SOX9 or empty vector. Three transwells were used per condition. Significance was evaluated by unpaired, two-tailed t-tests with P-values (*P<0.05; **P<0.01; ***P<0.001; ****P<0.0001; ns P>0.05). Mean is indicated by gray line.*

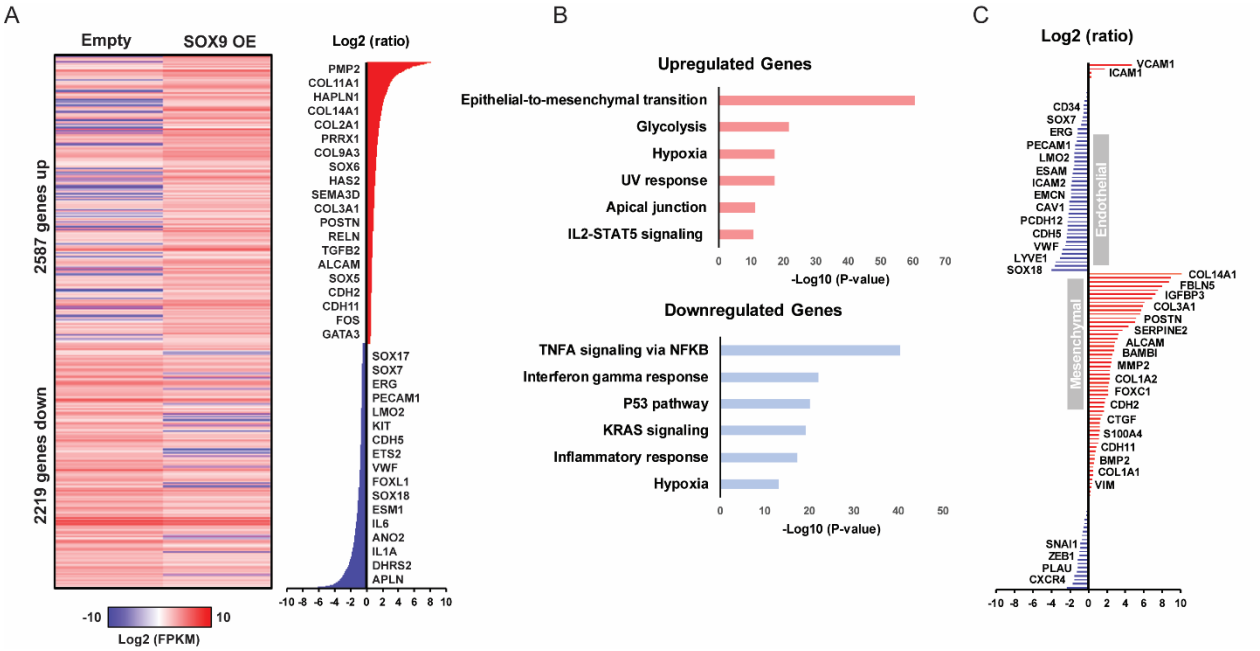


Figure 2. SOX9 alters expression of EndMT genes.

A) *Heatmap showing differentially expressed genes between cells transduced with SOX9 or empty vector. Heatmap shows log2 FPKM values while log2 ratios are displayed in the right panel with selected genes highlighted. Mean values of three biological replicates are shown.* ***B)*** *Top enriched Hallmark gene sets (based on P-values) for upregulated and downregulated genes upon SOX9 expression.* ***C)*** *Log2 ratios of SOX9 OE over empty vector for "endothelial" and "mesenchymal" genes with selected genes highlighted.*

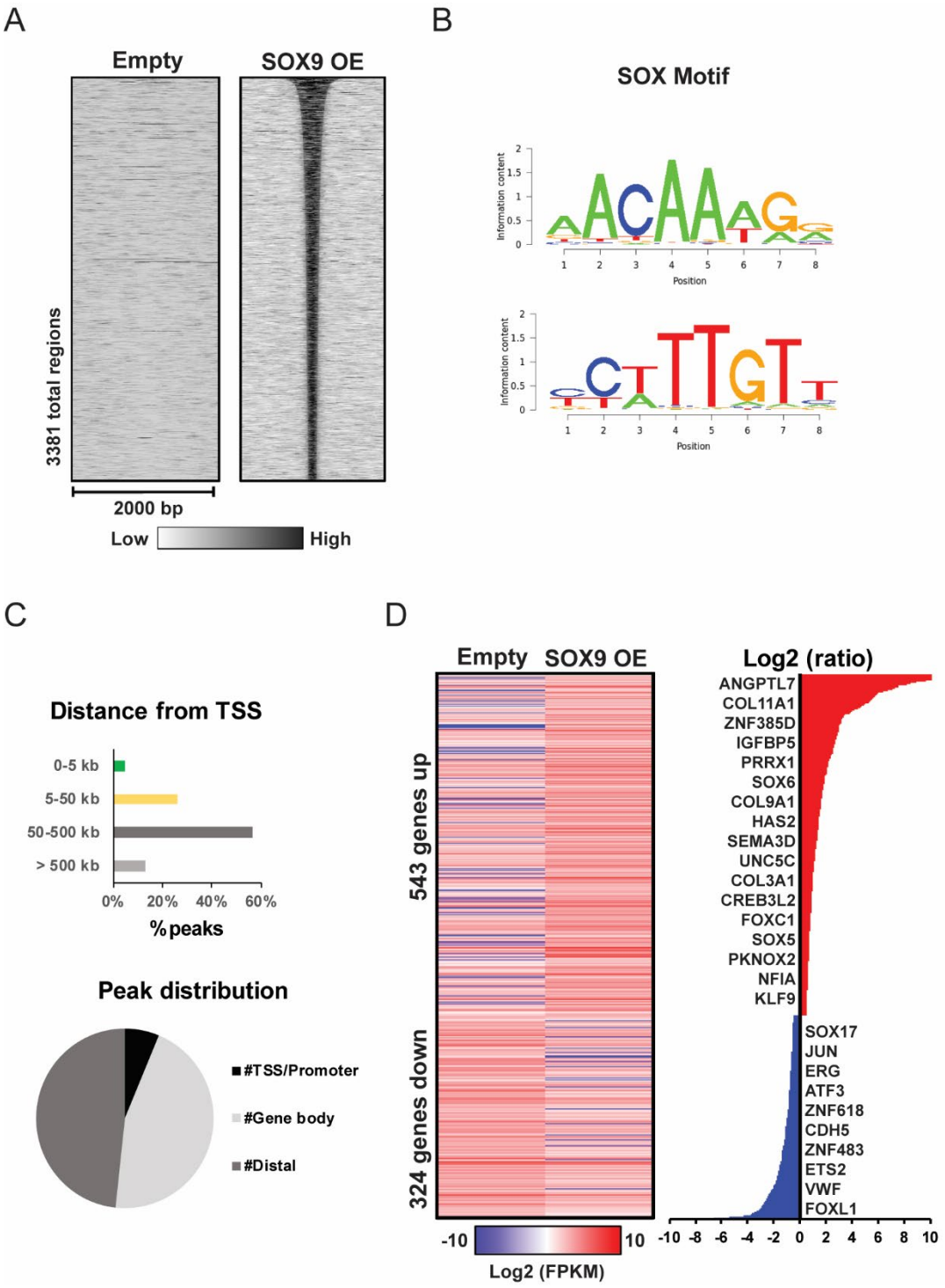


Figure 3. CUT&RUN profiling identifies SOX9 direct target genes.

A) *Heatmap displaying SOX9 CUT&RUN signal within a 2 kb window around the summit of SOX9 bound regions.* ***B)*** *Positional weight models for the SOX motif identified.* ***C)*** *Distance from*

SOX9 bound region and the TSS of the associated genes (upper panel) and the distribution of SOX9 bound regions between TSS/promoter regions (here delimited from 5000 bp upstream to 500 bp downstream of TSS), within gene bodies, or distal (lower panel). ***D)*** *Heatmap of differentially expressed genes with an associated SOX9 bound region. Heatmap shows log2 FPKM values while log2 ratios are displayed in the right panel with selected genes highlighted.*

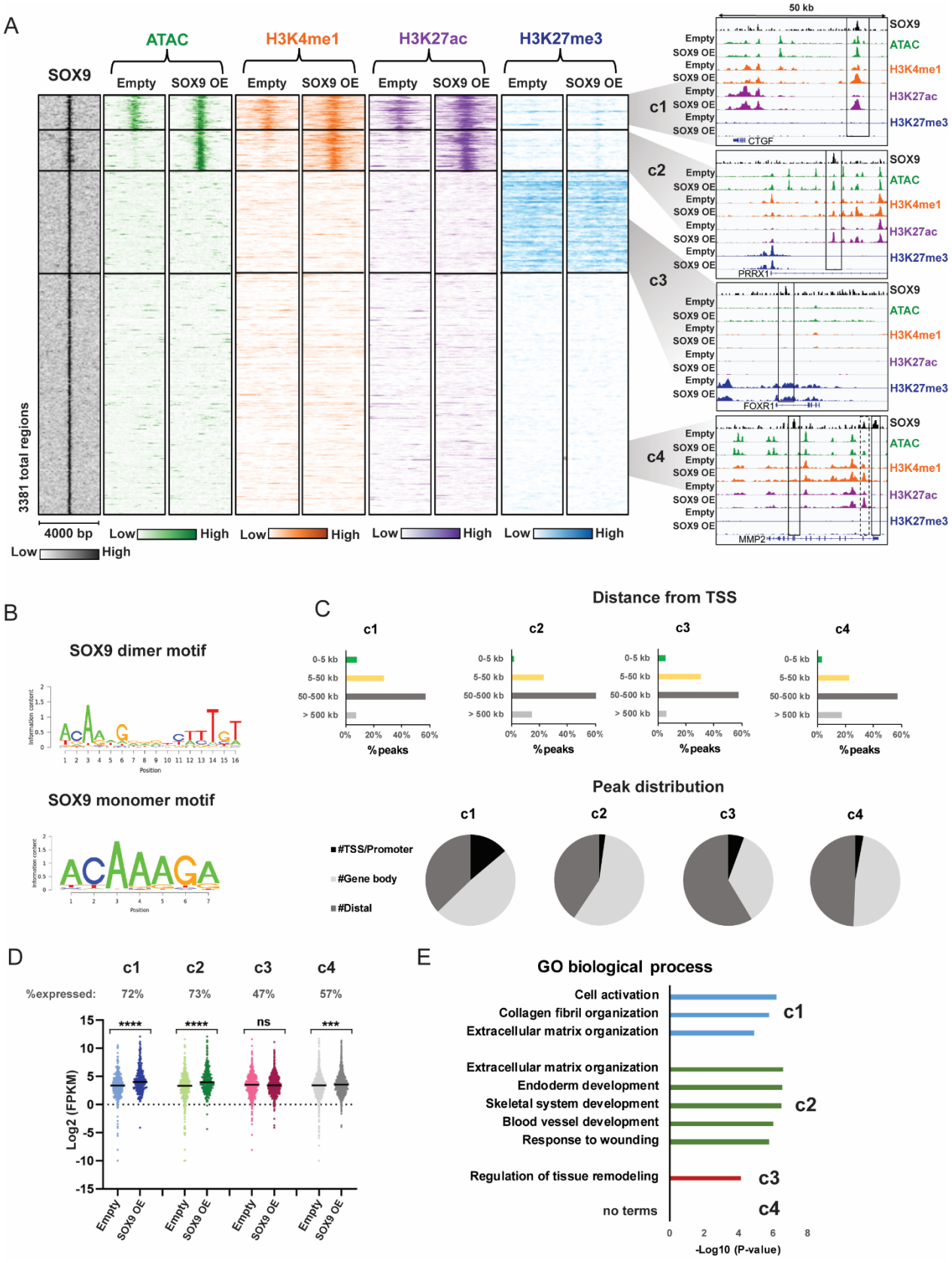
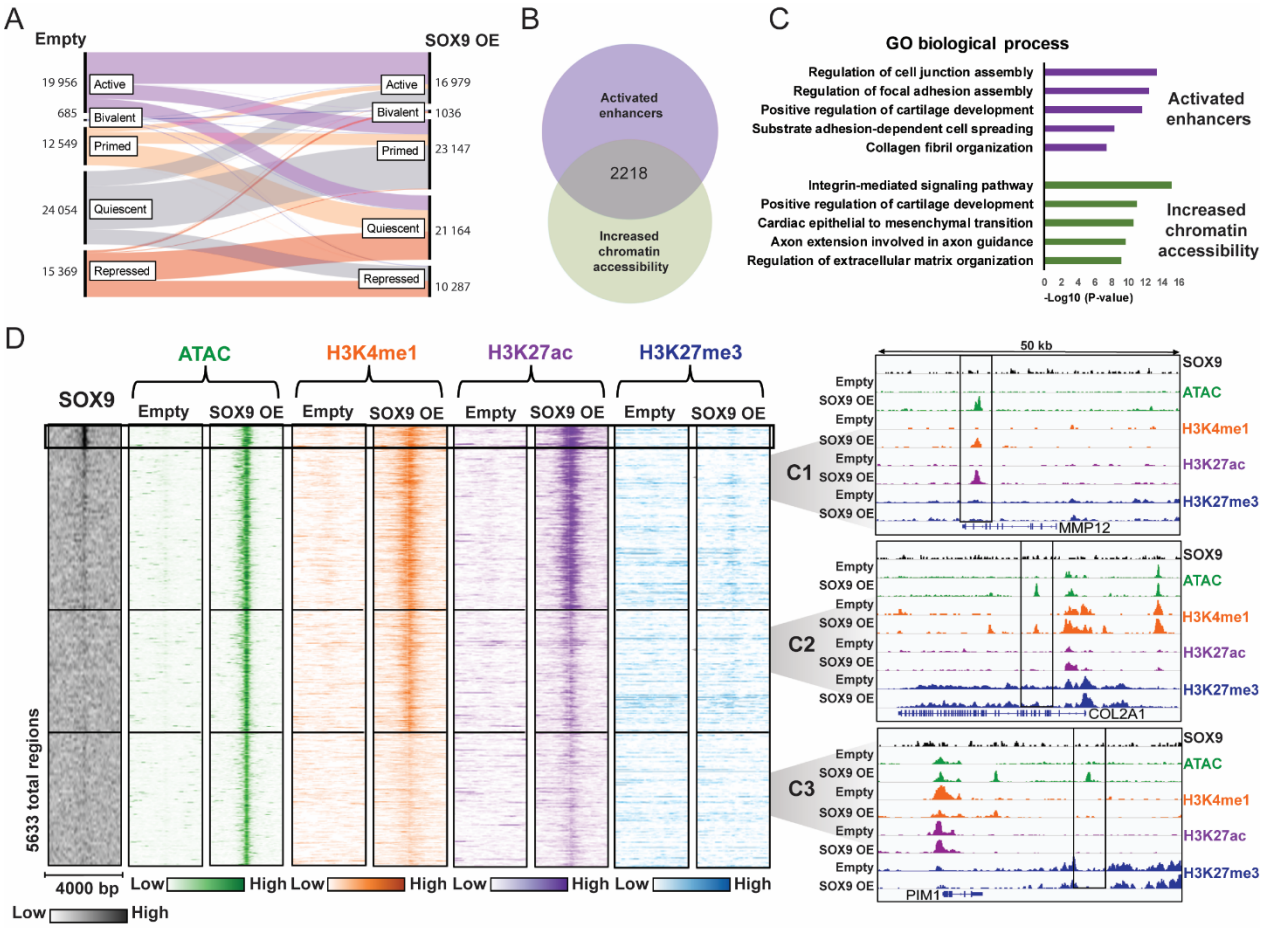
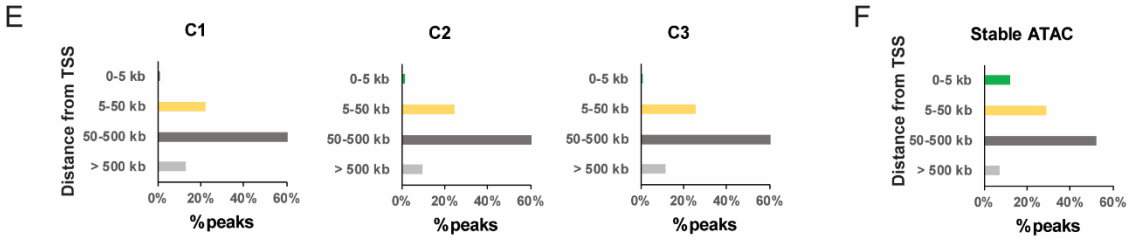

A
SOX9
ATAC
H3K4me1
H3K27ac
H3K27me3
Empty
SOX9 OE
3381 total regions
4000 bp
Low
High
50 kb
c1
c2
c3
c4
CTGF
PRRX1
FOXR1
MMP2
B
SOX9 dimer motif
SOX9 monomer motif
C
Distance from TSS
0-5 kb
5-50 kb
50-500 kb
> 500 kb
%peaks
Peak distribution
#TSS/Promoter
#Gene body
#Distal
D
%expressed:
72%
73%
47%
57%
Log2 (FPKM)
****
ns
***
E
GO biological process
Cell activation
Collagen fibril organization
Extracellular matrix organization
Endoderm development
Skeletal system development
Blood vessel development
Response to wounding
Regulation of tissue remodeling
no terms
-Log10 (P-value)

Figure 4. CUT&Tag and ATAC-seq identifies different modes of SOX9 chromatin binding. ***A)*** *Left panel: heatmap displaying SOX9, ATAC, and H3K4me1, H3K27ac, and H3K27me3 signal within a 4 kb window around the summit of SOX9 bound regions. We identified four clusters (c1-c4). The right panel displays representative loci in each cluster. Regions with a SOX9 peak are highlighted with black boxes, and the c2 region in the* MMP2 *locus is highlighted with a dotted black box.* ***B)*** *Positional weight models for the top enriched SOX dimer and monomer motifs in c1 and c2, respectively. An extended motif list is shown in Supplemental Fig. 1A.* ***C)*** *Distance from SOX9 bound region and the TSS of associated genes (upper panel) and the distribution of SOX9 bound regions between TSS/promoter regions, gene body positions, or distal (lower panel) for the SOX9 bound regions in each cluster.* ***D)*** *Scatterplots of log2 FPKM values of expressed genes (percentage is indicated) with an associated SOX9 bound region in each cluster. P-values are indicated as in Fig. 1D. Mean is indicated with black line.* ***E)*** *Top enriched (or the only enriched for c1 and c3) biological processes for genes associated with SOX9 bound regions in each cluster.*


A
Empty
SOX9 OE
19 956
Active
685
Bivalent
12 549
Primed
24 054
Quiescent
15 369
Repressed
16 979
1036
23 147
21 164
10 287
B
Activated enhancers
2218
Increased chromatin accessibility
C
GO biological process
Regulation of cell junction assembly
Regulation of focal adhesion assembly
Positive regulation of cartilage development
Substrate adhesion-dependent cell spreading
Collagen fibril organization
Activated enhancers
Integrin-mediated signaling pathway
Positive regulation of cartilage development
Cardiac epithelial to mesenchymal transition
Axon extension involved in axon guidance
Regulation of extracellular matrix organization
Increased chromatin accessibility
-Log10 (P-value)
D
ATAC
H3K4me1
H3K27ac
H3K27me3
SOX9
Empty
SOX9 OE
5633 total regions
4000 bp
Low
High
50 kb
C1
C2
C3
MMP12
COL2A1
PIM1


E
C1
C2
C3
Distance from TSS
0-5 kb
5-50 kb
50-500 kb
> 500 kb
0%
20%
40%
60%
%peaks
F
Stable ATAC

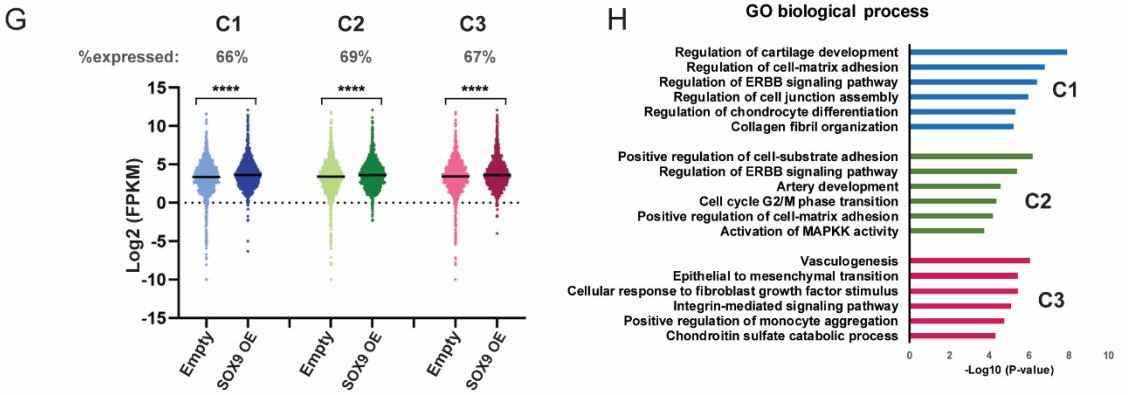

G
C1
C2
C3
%expressed:
66%
69%
67%
****
Log2 (FPKM)
Empty
SOX9 OE
H
GO biological process
Regulation of cartilage development
Regulation of cell-matrix adhesion
Regulation of ERBB signaling pathway
Regulation of cell junction assembly
Regulation of chondrocyte differentiation
Collagen fibril organization
Positive regulation of cell-substrate adhesion
Regulation of ERBB signaling pathway
Artery development
Cell cycle G2/M phase transition
Positive regulation of cell-matrix adhesion
Activation of MAPKK activity
Vasculogenesis
Epithelial to mesenchymal transition
Cellular response to fibroblast growth factor stimulus
Integrin-mediated signaling pathway
Positive regulation of monocyte aggregation
Chondroitin sulfate catabolic process
-Log10 (P-value)

Figure 5. Global chromatin changes upon SOX9 expression.

A) *Alluvial plot showing dynamics in enhancer states between cells transduced with SOX9 or empty vector. Alluvium colours are based on the state of the enhancers in the empty vector control cells. The number of sites in each alluvium is indicated on the sides.* ***B)*** *Venn diagram of activated enhancers (6592) and regions with increased chromatin accessibility (5633). Activated enhancers include enhancers that acquired H3K27ac upon SOX9 expression, while regions with increased chromatin accessibility include regions with an increase of more than 2 fold in ATAC-seq signal upon SOX9 expression.* ***C)*** *Top enriched biological processes for genes associated with activated enhancers (upper panel) or regions with increased chromatin accessibility (lower panel).* ***D)*** *Left panel: heatmap displaying SOX9, ATAC, and H3K4me1, H3K27ac, and H3K27me3 signal within a 4 kb window around the summit of increased ATAC signal upon SOX9 expression. Three clusters were identified (C1-C3) based on acquisition of H3K4me1 and/or H3K27ac in addition to an increased ATAC signal. SOX9 bound regions are highlighted with a black box (subtracted from further analyses). The right panel displays representative loci in each cluster. Sites of increased ATAC signal are highlighted with black boxes.* ***E)*** *Distance from increased ATAC regions and the TSS of the associated genes in each cluster.* ***F)*** *Distance from ATAC regions with no change in signal between cells transduced with SOX9 or empty vector and the TSS of associated genes.* ***G)*** *Scatterplots of log2 FPKM values of expressed genes with an associated increased ATAC peak in each cluster. P-values are indicated as in Fig. 1D. Mean is indicated with black line.* ***H)*** *Top enriched biological processes for genes associated with increased ATAC regions in each cluster.*

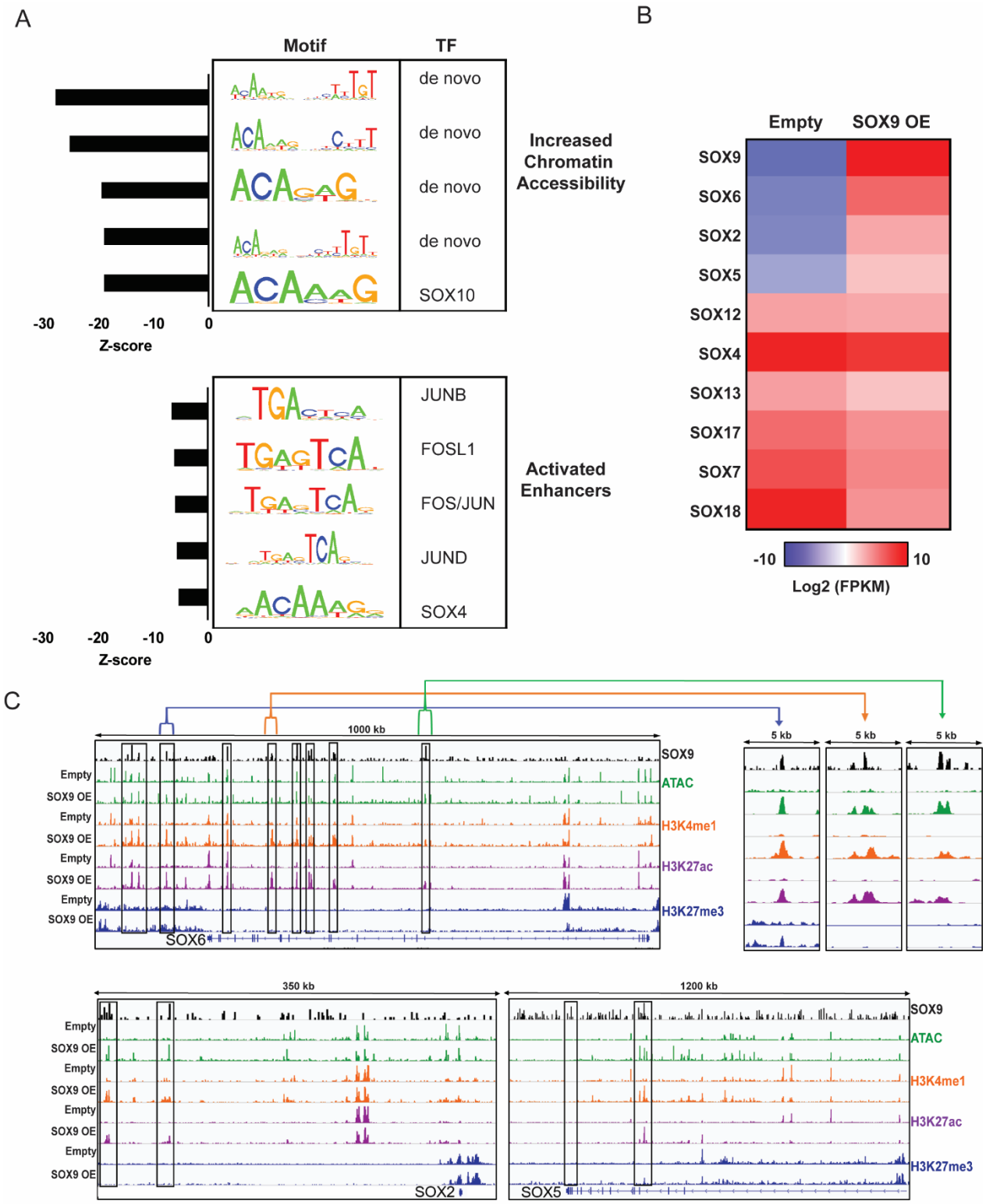


Figure 6. Potential pioneer TFs acting downstream of SOX9.

A) *The top five TF motifs identified in regions with increased chromatin accessibility (upper*

panel) and in activated enhancers without an increase in chromatin accessibility (lower panel). ***B)*** *Heatmap displaying the log2 FPKM values of expressed SOX factors in cells transduced with SOX9 or empty vector.* ***C)*** *The SOX6 locus (upper panel), SOX2 locus (lower left), and SOX5 locus (lower right). The three SOX9 bound enhancers displaying increased chromatin accessibility are highlighted in the upper right panel.*

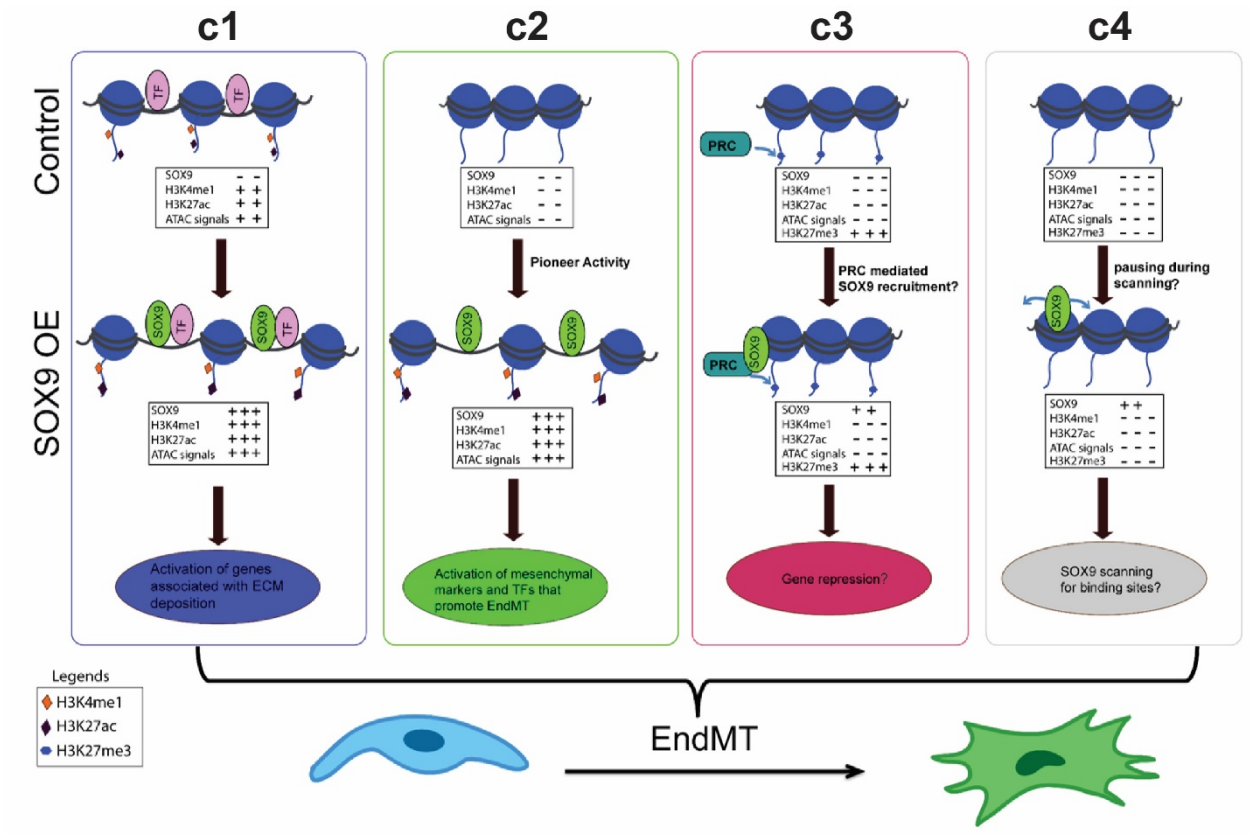


Figure 7. Modes of SOX9 binding and impact on chromatin.

SOX9 binding displays four different modes. In c1 regions, SOX9 binds to open chromatin where other TFs are likely already present, which leads to a slight increase in active histone marks and activation of ECM genes. The chromatin in c2 regions is closed and contains no active histone marks, but upon SOX9 binding the chromatin opens and acquires enrichment of active histone marks, which leads to activation of TF genes and markers for mesenchymal cells. The repressive mark H3K27me3 is highly enriched in c3 regions, both with and without SOX9 binding, and may involve SOX9 recruitment through other factors such as Polycomb Repressive Complexes (PRC) and might represent a mechanism of transcriptional repression by SOX9. SOX9 binding to c4 regions, which contains none of the histone marks assayed with or without SOX9 binding, represents low-affinity binding and potential pause sites during chromatin scanning by SOX9.