

Article

Differential regulation of alternate promoter regions in *Sox17* during endodermal and vascular endothelial development

Linh T. Trinh,^{1,3,4} Anna B. Osipovich,^{2,3} Leesa Sampson,³ Jonathan Wong,⁵ Chris V.E. Wright,^{1,3,4} and Mark A. Magnuson^{1,2,3,4,6,*}

SUMMARY

***Sox17* gene expression is essential for both endothelial and endodermal cell differentiation. To better understand the genetic basis for the expression of multiple *Sox17* mRNA forms, we identified and performed CRISPR/Cas9 mutagenesis of two evolutionarily conserved promoter regions (CRs). The deletion of the upstream and endothelial cell-specific CR1 caused only a modest increase in lymphovascuogenesis likely via reduced Notch signaling downstream of SOX17. In contrast, the deletion of the downstream CR2 region, which functions in both endothelial and endodermal cells, impairs both vascular and endodermal development causing death by embryonic day 12.5. Analyses of 3D chromatin looping, transcription factor binding, histone modification, and chromatin accessibility data at the *Sox17* locus and surrounding region further support differential regulation of the two promoters during the development.**

INTRODUCTION

The *Sox* gene family of transcription factors (TFs), as defined by the presence of a DNA binding SRY-related high-mobility group (HMG) box, are present in unicellular choanoflagellates, invertebrate and vertebrate species (King et al., 2008; Lefebvre et al., 2007; Sarkar and Hochedlinger, 2013). *Sox* family members broadly regulate cell fate determination, with the 20 *Sox* genes in vertebrate genomes performing highly divergent functions often involving lineage specification and the regulation of developmental potency (Lefebvre et al., 2007; Sarkar and Hochedlinger, 2013). Gene duplication during evolution has resulted in both functional redundancy and structural similarities between different *Sox* genes (Lefebvre et al., 2007; Schepers et al., 2002). The F subgroup contains *Sox7*, *Sox17*, *Sox18*, all playing crucial roles in endodermal, blood, and lymphatic vascular development (Kamachi and Kondoh, 2013).

Sox17 gene expression begins around embryonic day (E) 4.5 in the primitive endoderm layer of the mouse blastocyst, continues throughout subsequent gastrulation, and is often used as a marker for definitive endoderm (DE) (Burtscher et al., 2012; Kanai-Azuma et al., 2002; Nakan et al., 2010). Loss-of-function studies in mice have shown that *Sox17* is necessary for the survival of the foregut and expansion of prospective mid- and hindgut, although it is redundant to *Sox7* in extra-embryonic endodermal tissues (Kanai-Azuma et al., 2002). At E9.5, *Sox17* expression in DE becomes restricted to the hindgut and posterior foregut, and in the latter is critical for the formation of the liver, pancreas, and biliary system (Spence et al., 2009). Dysregulation of *Sox17* in DE is associated with biliary atresia, acute hepatitis (Uemura et al., 2013), gall bladder agenesis, and ectopic pancreatic development (Spence et al., 2009).

Sox17 is also critical for cardiovascular and hematopoietic development from mesodermal tissues, and conditional deletion of *Sox17* in endothelial progenitors causes embryonic lethality owing to disrupted hematopoiesis (Kim et al., 2007). *Sox17* expression in endothelial cells begins around E9 and is essential for the specification of arterial fate (Corada et al., 2013), maintenance of hemogenic endothelium, and emergence and maturation of fetal hematopoietic stem cells (Kim et al., 2007; Lizama et al., 2015). *Sox17* shares overlapping functions with *Sox18* in cardiac looping and vascular remodeling during both pre- and post-natal stages (Matsui et al., 2006; Sakamoto et al., 2007). Recent studies showed that *Sox17*-expressing mesodermal cells are precursors of the endocardium (Saba et al., 2019), and that *Sox17* facilitates the formation of coronary arteries (Gonzalez-Hernandez et al., 2020).

¹Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN 37232, USA

²Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN 37232, USA

³Center for Stem Cell Biology, Vanderbilt University, Nashville, TN 37232, USA

⁴Program in Developmental Biology, Vanderbilt University, Nashville, TN 37232, USA

⁵College of Arts and Science, Vanderbilt University, Nashville, TN 37232, USA

⁶Lead contact

*Correspondence: mark.magnuson@vanderbilt.edu

<https://doi.org/10.1016/j.isci.2022.104905>



Previous investigations into the regulation of *Sox17* gene expression have largely utilized non-mammalian vertebrate models. In zebrafish, *sox17* expression in DE is induced by *casanova* (*cas*; also known as *sox32*), a fish-specific *soxF* factor (Reim et al., 2004), plausibly regulated by the formation of an Eomes/Gata5/Bon TF complex (Zorn and Wells, 2007). In *Xenopus*, the initiation and maintenance of *sox17* expression in DE was found to be directly regulated by the binding of maternal VegT - a frog-specific T-box TF, and Smad2 - a Nodal signaling effector, at the *sox17 α* promoter (Clements and Woodland, 2003; Engleka et al., 2001; Howard et al., 2007). In mice, whereas combinatorial Nodal and Wnt signaling are required for DE formation (Liu et al., 1999; Huelsken et al., 2000; Lowe et al., 2001; Tremblay et al., 2000; Vincent et al., 2003), the expression of *Sox17* exhibit differences during gastrulation and embryo morphogenesis when compared with fish and amphibians (Zorn and Wells, 2007). Indeed, the TCF4/ β -CATENIN complex is among the few factors known to directly regulate *Sox17* expression in mouse endoderm (Engert et al., 2013).

The existence of at least two promoters in the *Sox17* gene was previously inferred from identifying long and short mRNA forms in endothelial cells and endoderm, respectively (Choi et al., 2012; Liao et al., 2009). However, the presence of multiple predicted transcriptional start sites (TSSs) and inconsistencies in the annotation of *Sox17* mRNA forms in mice (Choi et al., 2012; Liao et al., 2009) and humans (Katoh, 2002) across different online databases (i.e., between Ensembl and UCSC genome browser/Refseq), have hindered understanding of how the two promoters regulate *Sox17* gene expression in endoderm and endothelium. Herein we describe the use of CRISPR/Cas9-induced gene editing in mice and a *Luciferase* reporter assay in mouse embryonic stem cell-derived endothelial and endodermal cells to identify and functionally characterize two evolutionarily conserved regions (CRs) in the murine *Sox17* locus. Besides observing differential activities of the two promoters, we analyzed 3D chromatin looping to identify a distal element that appears to interact specifically with one promoter but not the other. We propose a model whereby alternate promoters in *Sox17* function as hubs for the binding of various cell type-specific TFs and/or for interacting with a distal element to transcriptionally regulate *Sox17* expression in different cell lineages during development.

RESULTS

Two evolutionarily conserved non-coding regions within the *Sox17*

To better understand how *Sox17* is regulated we began by searching for promoter-proximal *cis*-regulatory elements. As the annotations of murine *Sox17* gene are inconsistent across databases, and the location of the TSS for a short form of mRNA (TSS2; Figure 1A) has not been firmly established, we performed 5' Rapid Amplification of cDNA Ends (5' RACE) using mRNA extracted from mouse embryonic stem cell-derived definitive endoderm and vascular endothelial cells (Figure S1A). For both samples, we identified a major band corresponding to the predicted short form of *Sox17* mRNA. Cloning and Sanger-sequencing of the 5' RACE products confirmed that TSS2 lies between exon 3 and 4 (Figure S1B) and coincides precisely with the location predicted by the FANTOM5 CAGE database (Figure 1A). Our inability to detect the well-established long form of *Sox17* mRNA with 5' RACE assay is likely owing to the intrinsic bias of PCR amplification for shorter amplicons. Nevertheless, we gained confidence in using FANTOM5 CAGE database as an important starting point for further analysis of *Sox17* locus.

Having firmly established the location of TSS2 and the existence of a short *Sox17* mRNA form, we next used Vista-Point (Frazer et al., 2004; Mayor et al., 2000) to compare the murine *Sox17* gene locus with its human, chimpanzee, rhesus, rat, dog, cow and chicken homologs and identified two non-coding CRs, hereafter referred to as CR1 and CR2, located upstream of two putative TSSs, designated as either TSS1 or TSS2. We then extended our analysis using the conservation track in the UCSC browser (Kent et al., 2002) that reflects the sequence similarities of over 60 species. CR1 is located directly upstream of putative TSSs within exon 1 (TSS1) whereas CR2 lies between exons 3 and 4, upstream of an extended exon 4 where a TSS was confirmed based on our 5' RACE results and the FANTOM5 CAGE database (TSS2) (Lizio et al., 2015) (Figures 1A and 1B). RepeatMasker track (UCSC browser) indicated that CR1 and CR2 do not contain transposable elements (Figure S1C), and visual inspection did not reveal recognizable mRNA splice donor/acceptor/branch sites (Table S7). Using the RNAcentral database (v20), we observed a long non-coding RNA that spans the CR1 region (NONCODE: NONMMUT000027.2 - Figure S1D). Whereas this long non-coding RNA is expressed at a very low level in murine adult lung tissue (approximately 0.025 FPKM/TPM according to the source database NONCODE v6), we did not detect expression at early embryonic stages using RT-qPCR with two different primer pairs (Table S4), suggesting that CR1 function during developmental stages is unrelated to this long non-coding RNA. Thus, considering that both regions

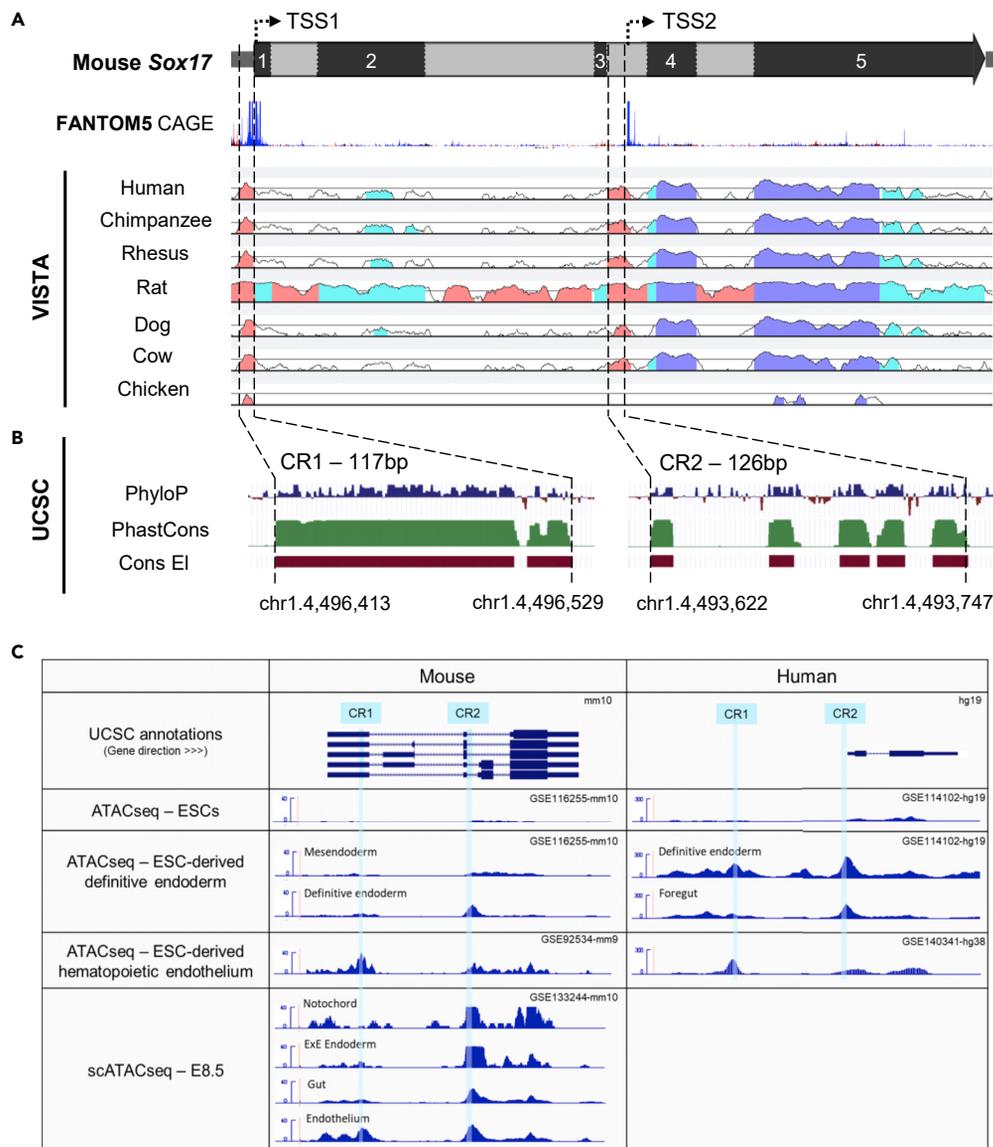


Figure 1. Identification of candidate *Sox17* proximal *cis*-regulatory elements in mice

(A) Mouse *Sox17* gene structure and location of two proximal *cis*-regulatory conserved regions (CRs). Exon 1–5, black; intron, gray. Dashed lines indicate boundaries of CR1 and CR2 positioned close to putative transcriptional start sites according to FANTOM5 CAGE (Cap Analysis Gene Expression), TSS1, and TSS2, respectively. VISTA-Point conservation analysis: introns, pink; untranslated exon sequence, blue; protein-coding sequence, purple.

(B) CR1 and CR2 conservation analysis by UCSC (University of California Santa Cruz) genome browser mouse assembly mm10: phylogenetic computation of *p* values (PhyloP) for conservation of individual nucleotides; phylogenetic analysis with space/time models for identification of conserved sites (PhastCons) and prediction of conserved elements (Cons EI). Chromosome 1 (chr1.) positions indicate CR1 (117 bp) and CR2 (126 bp) deletion breakpoints produced in mice via CRISPR-Cas9.

(C) *Sox17* locus chromatin accessibility data from published Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) data sets for mouse or human embryonic stem cells (ESCs), ESC-derived endoderm, ESC-derived endothelium, and E8.5 mouse endodermal and endothelial cells from single-cell (sc) ATAC-seq. CR1 and CR2 in both species are highlighted in light blue. Gene Expression Omnibus accession number (GSE) for each published data set is followed by the reference genome (mm, mouse, or hg, human).

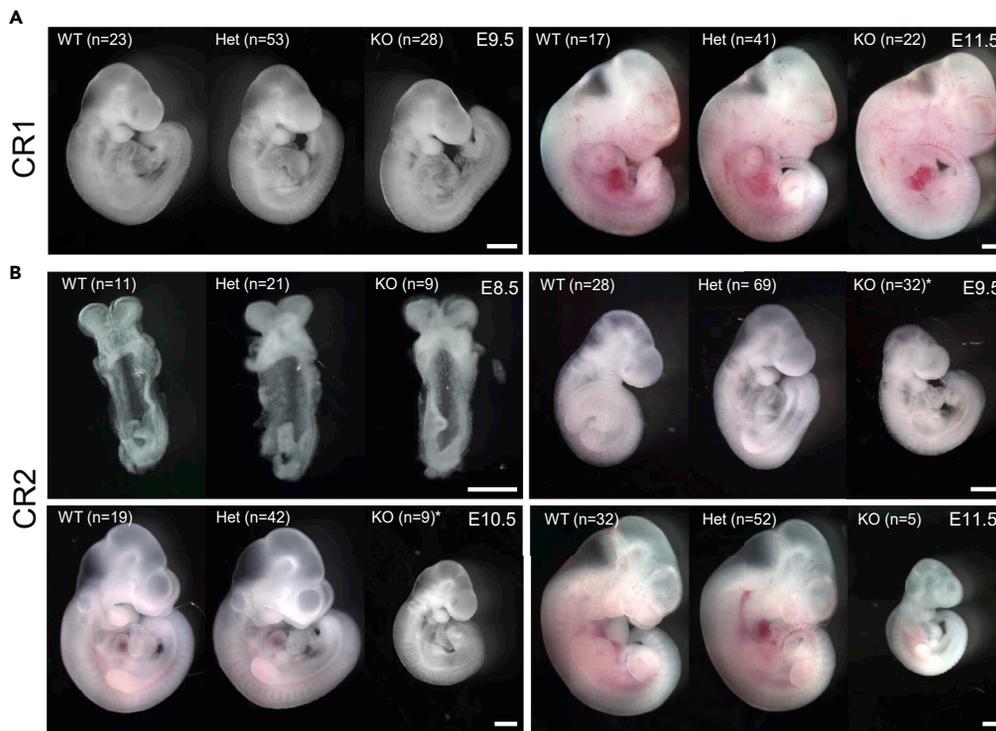


Figure 2. Deletion of *Sox17* CR2, but not CR1, results in developmental growth retardation and embryonic lethality

(A) Representative images of CR1 deletion line E9.5 ($n \geq 23$) and E11.5 ($n \geq 17$) embryos.

(B) Representative images of CR2 deletion line E8.5 ($n \geq 9$), E9.5 ($n \geq 28$), E10.5 ($n \geq 9$), and E11.5 ($n \geq 5$) embryos.

*phenotypic variation observed (Figure S2). WT, wild type; Het, heterozygous, KO, knockout embryos. Scale bars, 0.5 mm.

precede predicted TSSs in *Sox17*, and that they also contain sequence motifs for endodermal and/or endothelial transcription-factor binding by JASPAR analysis (Fornes et al., 2020) (Figure S1E), we hypothesized that CR1 and CR2 contain proximal cis-regulatory elements necessary for *Sox17* gene expression.

To expand our understanding of TSS1 and TSS2 expression, we also examined the chromatin accessibility around CR1 and CR2. To do so, we compiled a series of previously published ATAC-seq (Assay for Transposase-Accessible Chromatin sequencing) data sets from mouse and human embryonic stem cell (ESC)-derived hematopoietic endothelial and endodermal cells (Cernilogar et al., 2019; Jung et al., 2021; Lee et al., 2019), and from mouse embryonic tissues at E8.5 (Pijuan-Sala et al., 2020) (Figure 1C). Inspection of these data sets revealed differences in the accessibility of chromatin around CR1 and CR2 that varied both by cell lineage and stage of differentiation. Whereas chromatin accessibility at CR1 is open in mesoderm to hematopoietic endothelial lineages, CR2 is preferentially accessible in the endoderm.

Mice lacking CR1 are viable, whereas those lacking CR2 show embryonic lethality

To determine the functional importance of CR1 and CR2, we used CRISPR/Cas9 gene editing to derive alleles containing precise deletions of either CR1 (117 bp) or CR2 (126 bp). Mice heterozygous for either allele were grossly normal and fertile. However, whereas we could readily generate CR1 homozygous null (Δ CR1) mice from pairwise heterozygous matings, we were unable to identify any live-born pups that were homozygous for the CR2 deletion allele (Δ CR2) (Table S1). On account of this, we proceeded to examine the gross morphology of Δ CR1 and Δ CR2 embryos at different developmental stages using Theiler's criteria (Theiler, 1989).

At E9.5–E11.5, Δ CR1 embryos were visibly indistinguishable from littermates (Figure 2A). In contrast, Δ CR2 embryos, which showed no noticeable difference from wide-type embryos at late head-fold and early somite stages (E8.5), displayed growth retardation from E9.5 onward (Figure 2B), becoming smaller at later stages until, at E12.5, rarely any Δ CR2 embryos were recovered. Some Δ CR2 embryos exhibited more

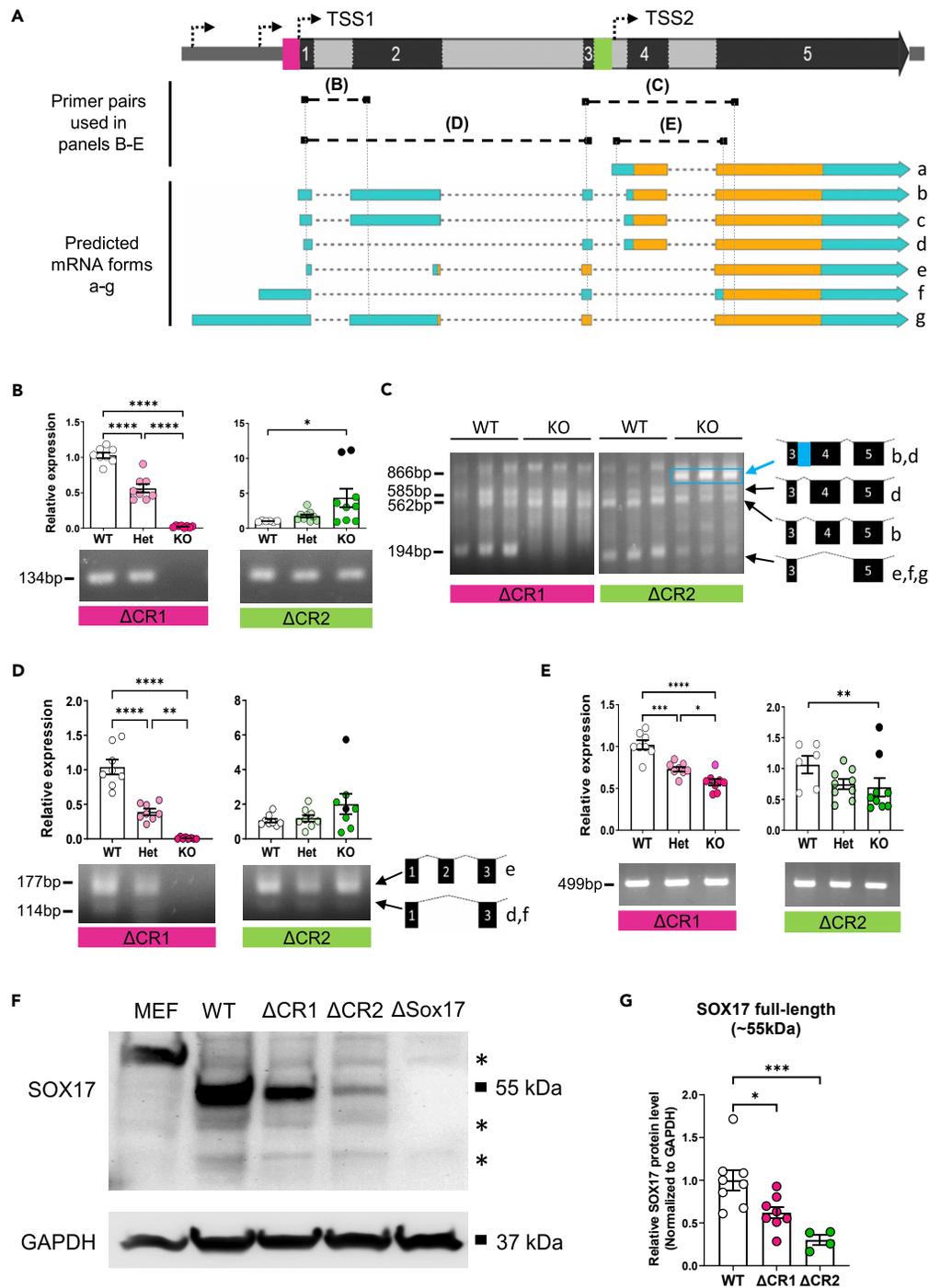


Figure 3. CR1 and CR2 are necessary for expression and processing of *Sox17* mRNA isoforms

(A) Mouse *Sox17* gene structure and predicted mRNA isoforms a-g schematic. Exons 1–5, black; introns, gray; CR1, magenta; CR2, green; dashed arrows, putative TSSs. Predicted *Sox17* mRNA isoforms a-g. Untranslated region, teal; protein-coding sequence, yellow. Black boxes connected by dashed lines indicate the location of primers and amplicons used for RT-PCR and RT-qPCR analyses presented in panels (B)–(E).

(B) Relative *Sox17* mRNA expression and representative electrophoresis gel of RT-qPCR products in CR1 and CR2 E9.5 embryos for primer pair B.

(C) *Sox17* mRNA RT-PCR analysis of CR1 and CR2 E9.5 embryos for primer pair C.

(D) Relative *Sox17* mRNA expression and representative electrophoresis gel of RT-qPCR products in CR1 and CR2 E9.5 embryos for primer pair D.

Figure 3. Continued

(E) Relative Sox17 mRNA expression and representative electrophoresis gel of RT-qPCR products in CR1 and CR2 E9.5 embryos for primer pair E. (B, D, E) WT, wild type; Het, heterozygous; KO, knockout embryos. CR1 ($n = 8$) and CR2 ($n \geq 6$). CR2 KO embryos with substantial tail effusion, marked with black dots, are outliers excluded from statistical analysis. (C) Arrows, Sox17 mRNA exon junctions confirmed by Sanger sequencing of PCR products. $n = 3$. (F) Representative western blot analysis for SOX17 and GAPDH in MEF (mouse embryonic fibroblast cells in culture) – negative control, E9.5 WT, CR1 KO (Δ CR1), CR2 KO (Δ CR2), and Sox17 KO (Δ Sox17) embryos – negative control. GAPDH, loading control. *non-specific or not fully validated bands. (G) Quantification of relative SOX17 protein levels. WT, $n = 8$; CR1 KO, $n = 8$; CR2 KO, $n = 4$. Loading control, GAPDH. Each dot represents one embryo. Bars and error bars represent the mean \pm SEM. Results were analyzed by a one-way ANOVA. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

growth retardation than others and exhibited either a more severely degraded posterior body trunk, resembling in some respects the Sox17-null phenotype (Kanai-Azuma et al., 2002), or effusions near the tail tip characterized by a network of disorganized blood vessels (Figure S2A). The folding and closure of neural plate, the initial formation of optic and otic vesicles, heart tube, bronchial arches, and limb buds appeared generally normal at the stages examined (Figure S2C). Vessels and blood islands were still observed in yolk sacs and embryo-proper tissues of Δ CR2 mice, although the vasculature was less developed and showed an unresolved primary capillary plexus-like morphology as compared with the remodeled, well-branched and hierarchical structure in wild type and heterozygous littermates (Figure S2B). In addition, Δ CR2 embryos show normal axis rotation (the lordosis-kyphosis transition) that is absent in whole-body Sox17 knock-out mice (Kanai-Azuma et al., 2002), indicating the loss of CR2 is developmentally less deleterious than the total ablation of Sox17 expression.

CR1 and CR2 are necessary for the expression and processing of different Sox17 mRNA forms

Given the profoundly different phenotypes of CR1- and CR2-null animals, we next determined how removing each region affects Sox17 mRNA expression. Using mRNA from whole embryos at E9.5 and a selective-primer strategy for discriminating exon junctions, we performed reverse transcription – polymerase chain reaction and qPCR (RT-PCR/qPCR) analyses to assess the impact of CR1 and CR2 deletions on the expression of all predicted Sox17 mRNA forms (Figures 3A–3E). As the long and short forms of Sox17 mRNA differ significantly in the exon content, we designed a series of different primer pairs that would recognize all predicted exon combinations present in the long mRNA forms (e.g., primer pairs B, C, D – Figure 3A). Similarly, as an extended exon 4 region is only present in the short mRNA and not in any of the previously described long mRNA forms, we designed an additional primer pair for this region and used it to assess the expression of the short mRNA form (primer pair E – Figure 3A).

For Δ CR1 embryos, we observed an allele-dosage-dependent decrease of exon junction B (mRNA forms b, c, and g; Figure 3B) in both heterozygous and null embryos, suggesting that CR1 contains required regulatory elements for TSS1. In addition, CR1 is also critical for the expression of the other long Sox17 mRNAs (forms d, e, f, and g; Figure 3A) that start in exon 1 and that encode truncated SOX17 proteins (Figures 3C–194 bp band), suggesting the reduction of forms e, f, g and Figure 3D indicating a reduction of forms d, e and f). Embryos lacking CR1 also showed an allele-dosage-dependent partial decrease in the expression of Sox17 short mRNA (form a; Figure 3E).

For Δ CR2 embryos, there was a modest increase in the long Sox17 mRNAs (forms b, c, and g; Figure 3B), perhaps as a compensatory response to the loss of CR2. Although our inspection of CR2 had revealed no recognizable consensus mRNA splicing sites, our PCR analysis showed an additional band of 866 bps indicating that the intron between exons 3 and 4 was retained in the Δ CR2 embryos, whereas the 194 bp product (Exon 3–5 junction skipping exon 4) was reduced in the absence of CR2 (Figure 3C). Even though primer pair E was designed to be specific for a single product, in Δ CR2 samples the primers not only amplified the short mRNA form, but also detected the long mRNA owing to retention of this intron. On account of this, we were unable to specifically assess the expression of the short form of Sox17 mRNA with primer pair E. However, even in the presence of the incompletely processed mRNA (long mRNA with intron retention), there was a notable reduction of amplicon E (Figure 3E), consistent with TSS2 activity being dependent on CR2 and with CR2 being necessary for the expression of the short Sox17 mRNA form.

The differential effects of deleting CR1 or CR2 on the expression of long and short Sox17 mRNA forms were also indicated by the decrease in SOX17 full-length protein in both mouse models (Figures 3F and 3G).

Importantly, SOX17 protein was still found in endodermal and endothelial cells of E9.5 Δ CR1 embryos (Figure S3B), indicating that long *Sox17* mRNAs are not necessary for animal survival, and suggesting that endothelial SOX17 protein is instead produced from the short *Sox17* mRNAs. On the other hand, we analyzed predicted open reading frames (ORFs) and confirmed that the unexpected intron retention in Δ CR2 embryos did not result in any modification of the main ORF coding for full-length SOX17 from long mRNA forms (Figure S3A). The nearly undetectable level of SOX17 in both endodermal and endothelial cells in Δ CR2 mice (Figures 3F and 3G; Figure S3B) further suggests that in the absence of the short *Sox17* mRNA, the long mRNAs are unable to produce SOX17 protein efficiently, likely related to changes in two upstream ORFs (μ ORFs – Figure S3A) as the consequences of the intron retention presented above. As we were unable to firmly verify the identity of minor bands in the western blot analysis, the effects on the production and function of different SOX17 isoforms remain unclear. We also observed that several Δ CR2 embryos that exhibited tail effusions with pronounced vascular overgrowth and abnormalities (Figure S2) exhibited modestly higher expression of *Sox17* long mRNA and greater intron retention than most of the other Δ CR2 embryos (Figures 3B and 3E; black dots), consistent with long *Sox17* mRNAs being preferentially-expressed in endothelium (Choi et al., 2012; Liao et al., 2009).

Δ CR1 mice exhibit increased lympho-vasculogenesis associated with decreased *Notch1* expression

Our observation that deletion of CR1 prevents expression of long *Sox17* mRNAs without impairing embryonic survival prompted us to explore whether Δ CR1 mice have a non-lethal phenotype. Analysis of endodermal and mesodermal endothelial marker genes using mRNA from whole embryos at E9.5 revealed increased expression of endothelial markers (*Pecam1*, also known as *Cd31*, and *Tek*, also known as *Tie2*; Figure 4A) but not endodermal markers (*Foxa2*, *Sox2*, *Cdx2*, *Hnf4a*, *Sox9*, and *Pdx1*) in Δ CR1 embryos (Figure S4A). Prior studies have shown that *Sox17* is necessary for the generation of hematopoietic stem cells (Kim et al., 2007; Lizama et al., 2015), specification of arterial fate (Corada et al., 2013), and blood-vessel stabilization (Heinke et al., 2012; Lee et al., 2015). However, expression of the tissue-specific markers *Runx1* (hematopoietic progenitors), *Efnb2*, *Kdr* (also known as *Vegfr2*, an arterial cell marker), *Ephb4*, *Flt4* (also known as *Vegfr3*, a venous cell marker), *Pdgfb*, and *Tgfb1* (which is critical for the establishment of vasculature) were all unchanged across genotypes (Figure 4A; Figure S4A). In contrast, *Lyve1*, *Prox1*, *Pdpn*, and *Nr2f2* (also known as *CoupTFII*, a lymphatic endothelial cell marker) were all increased in Δ CR1 embryos (Figure 4A; Figure S4A).

Increased lympho-vasculogenesis but not arterial or venous development was further confirmed with flow cytometry by an increase in the number of LYVE1-expressing cells within the PECAM1-positive endothelial population (Figure 4B). Although *Sox18* is a key regulator of lympho-vasculogenesis (Hosking et al., 2009), no significant change was observed in RNA levels of either *Sox18* or *Sox7*, the other two *SoxF* members (Figure S4A). Instead, endothelial cells from Δ CR1 embryos had reduced *Notch1* and *Dll4* expression by almost 2-fold (Figure 4C). This finding is consistent with the regulation of *Notch1* by *Sox17* and the *Notch1*-*Dll4* signaling axis being necessary for the specification of arterial/venous/lymphatic cell fates, as previously reported (Chiang et al., 2017; Murtomaki et al., 2013).

Impaired formation of pancreatic buds and outgrowth of the hepato-pancreato-biliary system in Δ CR2 embryos

To further characterize the lethal phenotype of Δ CR2 embryos, we analyzed various endodermal and mesodermal lineage marker genes. At E9.5, there were no differences for *Foxa2* (pan-endoderm), *Sox2* and *Cdx2* (rostral versus caudal endoderm, respectively), *Nkx2-1* and *Foxe1* (lung and thyroid buds), *Hnf4a* (liver primordium), *Sox9* (biliary and pancreatic ductal progenitors), or *Pecam1* (endothelium) (Figures 5A and S5), suggesting that the initial formation of these tissues is normal. However, *Pdx1* (marking the anlagen of the pancreas, antral stomach, and rostral duodenum) was notably down-regulated (\sim 10-fold) (Figure 5A). Similar suppressions of *Hnf4a* (to a nearly undetectable level) and *Sox9* (down \sim 5-fold), as well as reductions to half or less of *Sox2* (foregut) and *Cdx2* (posterior fore- and mid/hindgut) were observed at later ages (E11.5) (Figures 5A and S5), suggesting an overall failure in survival and/or expansion of the entire endodermal tube and particularly the hepato-pancreato-biliary system that arises from it.

We also determined the expression of several lymphatic (*Prox1*, *Pdpn*, *Lyve1*, *Nr2f2*), arterial (*Efnb2*, *Kdr*), venous (*Ephb4*, *Flt4*), and hematopoietic (*Runx1*) marker genes in Δ CR2 embryos (Figure S5). Among the four tested lymphatic vascular genes, only *Lyve1* was significantly upregulated in the KO compared with

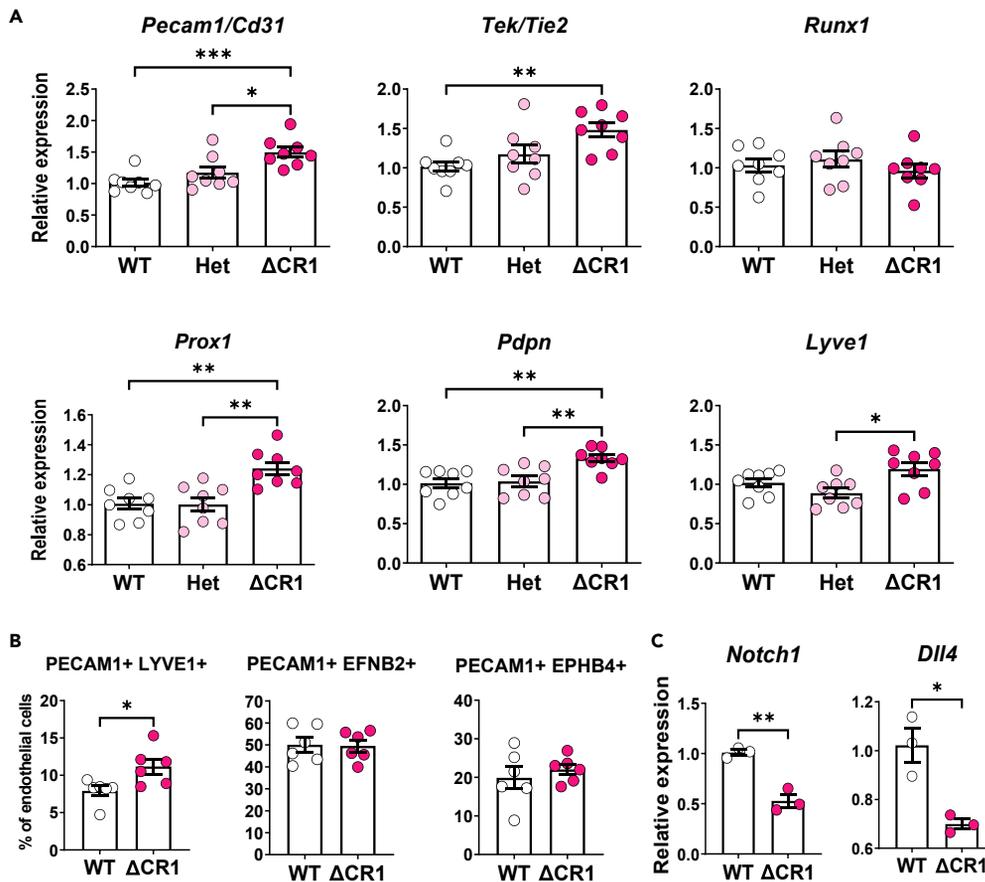


Figure 4. CR1 knockout mice exhibit increased lympho-vasculogenesis and reduced Notch receptor/ligand pair
(A) RT-qPCR analysis of relative mRNA expression in CR1 deletion line E9.5 embryos for endothelial, *Pecam1* (also known as *Cd31*) and *Tek* (also known as *Tie2*); hematopoietic, *Runx1*; and lymphatic endothelial, *Prox1*, *Pdpn* and *Lyve1*, markers. WT, wild type; Het, heterozygous; Δ CR1, CR1-null embryos. $n = 8$. Data are normalized to *Gapdh*. Each dot represents one embryo. Bars and error bars represent the mean \pm SEM analyzed by a one-way ANOVA. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

(B) Quantification of flow cytometry analysis for number of PECAM1-LYVE1, PECAM1-EFNB2, PECAM1-EPHB4-double positive cells in E10.5 WT and CR1-null embryo. $n = 6$.

(C) RT-qPCR analysis of relative mRNA expression for *Notch1* and *Dll4* in sorted PECAM1-positive cells of E10.5 WT and CR1-null embryos. $n = 3$. (B, C) Each dot represents one sorting experiment pooled from at least three embryos each genotype. Bars and error bars represent the mean \pm SEM analyzed by Student's *t* test. * $p < 0.05$; ** $p < 0.01$.

WT embryos. Of note, embryos that have the highest level of *Prox1* and *Lyve1* also expressed the highest level of hepatic marker *Hnf4a* (highlighted by orange dots in Figure S5), consistent with the fact that *Prox1* and *Lyve1* are known to be expressed in the developing liver (Burke and Oliver, 2002; Nonaka et al., 2007). Arterial, venous, and hematopoietic gene expression were also significantly increased in the Δ CR2 embryos compared with either heterozygous null or WT embryos, suggesting vascular abnormalities in the absence of CR2. Consistent with the previously discussed gross morphology analysis, two CR2-null embryos that exhibit tail effusion with pronounced vascular overgrowth (marked by black dots in Figure S5) expressed the highest level of tested vascular genes.

To better visualize morphological changes in the posterior foregut region at E9.5 in Δ CR2 embryos, we combined lineage tracing genetics and light-sheet microscopy to obtain 3D images of the posterior foregut region as the outgrowth of the hepato-pancreato-biliary primordia begins to occur. The lineage tracing strategy we employed used both of a previously described null *Sox17*^{GFP^{Cre}} allele (Choi et al., 2012) and a *R26*^{LSL:TdTomato} reporter allele, enabling us to identify cells that are actively expressing *Sox17* by their green fluorescence and to identify cells that previously expressed *Sox17* by red fluorescence. This dual fluorescence strategy allowed us to simultaneously visualize the whole gut tube together with the emerging vasculature of intact embryos and to identify the

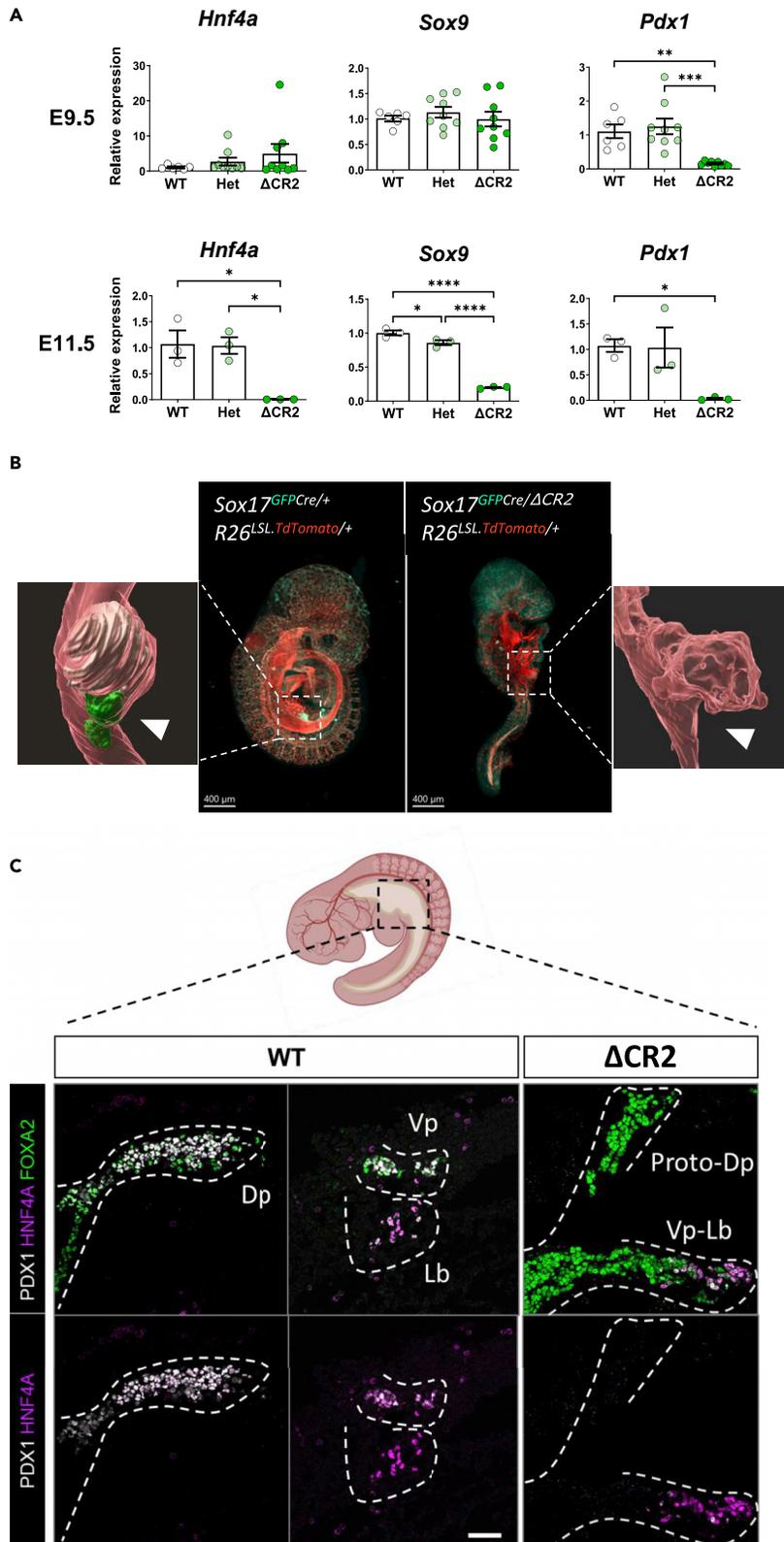


Figure 5. CR2 KO embryos exhibit impaired pancreatic bud formation and failure to establish the hepato-pancreato-biliary system

(A) RT-qPCR analysis of relative mRNA expression in CR2 deletion line embryos for hepatic, *Hnf4*; ductal, *Sox9*; and pancreatic, *Pdx1* markers. WT, wild type; Het, heterozygous; Δ CR2, CR2-null embryos. E9.5, $n \geq 6$; E11.5, $n = 3$. Data are normalized to *Gapdh*. Counts represent the mean \pm SEM analyzed by a one-way ANOVA. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

(B) Representative light sheet microscopy images of E9.5 control (*Sox17*^{GFP*Cre/+*}; *R26*^{LSL*TdTomato/+*} – center left) and CR2-deficient (*Sox17*^{GFP*Cre/ΔCR2*}; *R26*^{LSL*TdTomato/+*} – center right) embryos. Reconstruction of posterior foregut surface visualizes hepatopancreatobiliary bud formation in the presence (hepatic bud, white; pancreatobiliary bud, green) and the absence of CR2 (lack of visible pancreatobiliary bud, arrow head). Scale bar, 400 μ m.

(C) Representative immunofluorescence staining of E9.5 WT and CR2-null embryonic sections from the dashed square region for FOXA2 (endoderm, green), HNF4A (liver bud, LB, magenta), and PDX1 (dorsal bud, DP/Proto-DP, and ventral pancreatic bud, VP, white). $n = 3$. Scale bar, 50 μ m.

outgrowing hepatic domain (TdTomato single positive bud) from the ventral pancreato-biliary primordia (GFP and TdTomato double-positive bud) in E9.5 control embryos (*Sox17*^{GFP*Cre/+*}; *R26*^{LSL*TdTomato/+*}, Figure 5B – left panel). To determine the morphological effects of removing CR2, we generated *Sox17* compound heterozygous embryos (*Sox17*^{GFP*Cre/ΔCR2*}; *R26*^{LSL*TdTomato/+*}). Analysis of these embryos at E9.5 showed growth-retardation with failure in axis rotation and more severely degraded posterior trunk compared with Δ CR2 (*Sox17* ^{Δ CR2/ Δ CR2}) embryos, further indicating the importance of *Sox17* gene dosage in early morphogenesis. Moreover, by using the readily observable red fluorescence from these embryos, we derived a 3D reconstruction of the endodermal foregut region. Inspection of the 3D reconstructions clearly indicated the absence of the ventral, smaller bud domain (corresponding to the pancreato-biliary bud) posterior to the hepatic-like bud in compound heterozygous embryos (arrow heads, Figure 5B insets).

The posterior foregut region at E9.5 was further examined with immunofluorescence staining of both wild-type and Δ CR2 embryonic serial sections (Figure 5C). Consistent with previous findings (Spence et al., 2009), there was a scarcity of PDX1 production in the proto-dorsal pancreatic region at E9.5 in the absence of SOX17, suggesting failure or delay of the dorsal pancreas specification program. Few observed PDX1-expressing cells in E9.5 Δ CR2 embryos in the ventral domain were abnormally intermingled with HNF4A-positive hepatic progenitors, indicating that Δ CR2 embryos exhibit a strongly impaired segregation of the hepatic bud from ventral pancreatic and biliary progenitors. This finding suggests that in the absence of compensatory upregulation from TSS1, the appropriate level and/or timing of SOX17 protein production associated with CR2-directed transcription could be critical for the proper development of the hepato-pancreato-biliary system.

Both CRs exhibit directional promoter activity in cell culture

To better understand the cell-type-specific functions of CR1 and CR2, we tested their ability to drive *Luciferase* expression in reporter constructs delivered into DE or hematopoietic vascular endothelial (HVE) cells at various stages of their being derived from mouse embryonic stem cells (mESCs) using previously published directed differentiation protocols (Borowiak et al., 2009; Chiang and Wong, 2011) (Figure 6A). At day 6 of each differentiation protocol, approximately 70% of the cell population exhibited the expected morphologies (Figure S6A), and key marker genes were expressed as expected throughout each differentiation protocol (Figure S6B). In addition, the long and short *Sox17* mRNAs showed preferential expression in HVE and DE, respectively (Figure S6C), further supporting the presence of alternate cell type-specific promoters. From day 6–8 of differentiation, CR1 exhibited promoter activity only in HVE cells, whereas CR2 was active in both DE and HVE. CR1 and CR2 were unidirectional in their ability to drive reporter gene expression (Figure 6B), supporting the argument that CR1 does not function as a promoter for the predicted nearby long non-coding RNA (Figure S1D – NONCODE: NONMMUT000028.2). CR2 activity was gradually upregulated throughout the time-course of DE differentiation. In contrast, CR1 lacked transcriptional activity at any stage of DE-directed differentiation (Figure 6C).

Having observed CR2 activity in both DE and HVE cells, we next tested whether specific conserved motifs within CR2 (Figure 6D) contribute to promoter activity in a cell-type-specific manner. To do so, we generated six different fusion-gene constructs in which 9 bps were mutated by transversion (G to A and C to T, and vice versa – Figure 6D). Except for mutation 2 (m2), located in a non-CR of CR2 - thereby providing an internal control - mutations within all five of the conserved motifs (m1, m3–m6) reduced promoter activity in

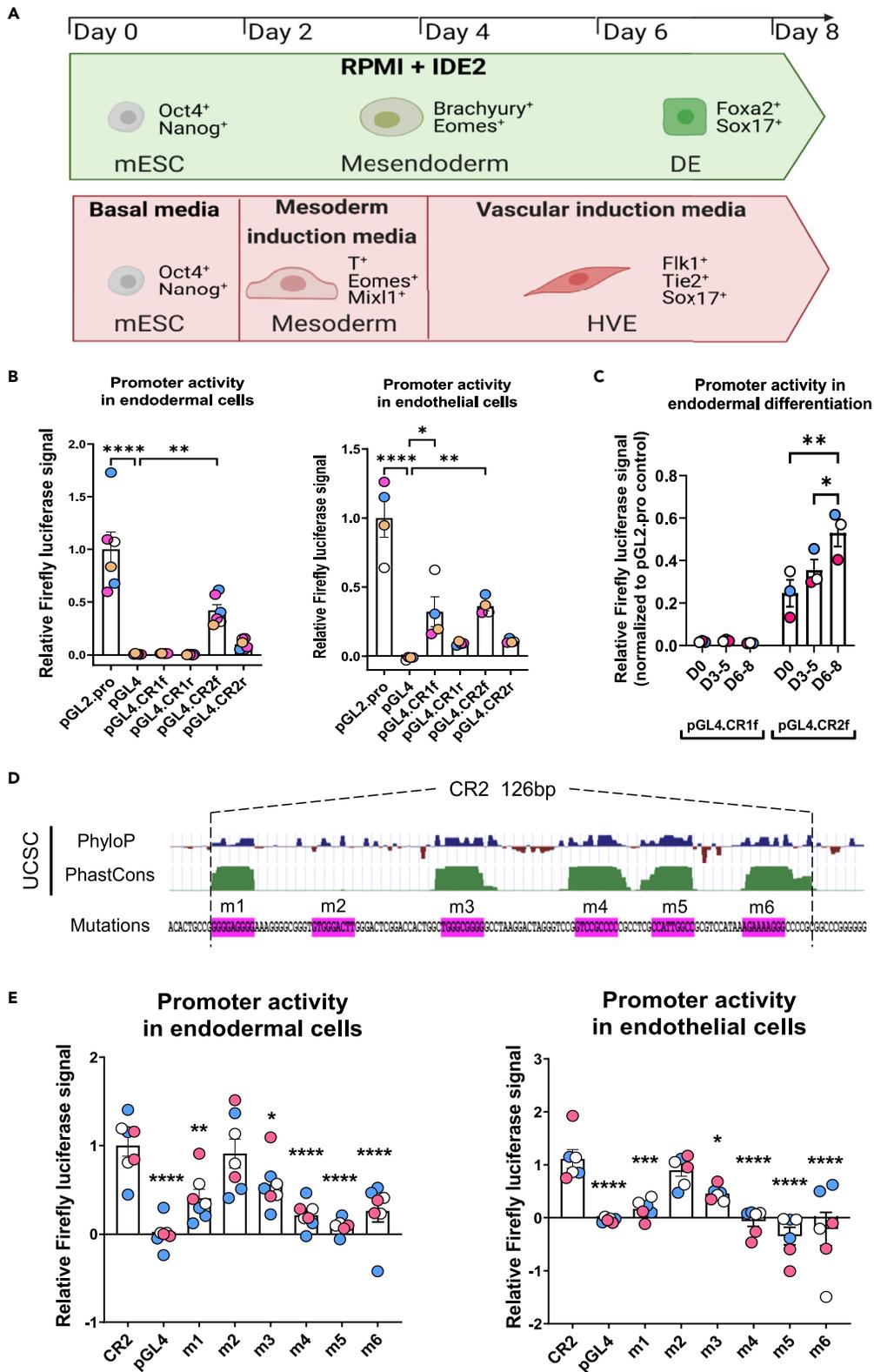


Figure 6. Sox17 CR1 and CR2 exhibit promoter activity in vitro

(A) The 8-day mESC differentiation protocol used to derive definitive endoderm-like and hematopoietic vascular endothelial-like cells. Terminal cellular identity was confirmed by cellular morphology and lineage-specific markers as indicated (see Figure S6).

(B) Luciferase-activity readout at days 6–8 of endodermal and endothelial differentiation. CR1 and CR2 DNA sequences were inserted in sense (CR1f, CR2f) or antisense (CR1r, CR2r) orientation. pGL4 backbone vector without a promoter, negative control. pGL2.pro vector with an SV40 promoter, positive control. Endoderm protocol, $n = 6$. Endothelial protocol, $n = 4$.

(C) Luciferase activity from CR1f and CR2f from three time points during the endodermal differentiation protocol. $n = 3$. Data are normalized to pGL2.pro.

(D) Five conserved motifs within CR2 identified with UCSC conservation track. Six 9-basepair blocks (magenta on wild-type CR2 sequence) were mutated by base transversion (G–A and C–T, or vice versa).

(E) Luciferase activity readout at days 6–8 of the endodermal or endothelial differentiation protocol for WT and six block-mutations within the CR2 sequence (m1–m6). pGL4, negative control. pGL4.CR2f, positive control. Endoderm protocol, $n = 7$. Endothelial protocol, $n = 6$.

(B, C, E) Dot color represents samples differentiated from the same date (same color) or different dates (different colors). Each dot represents one well of differentiated cells. Bars and error bars represent the mean \pm SEM. Results were analyzed by a one-way ANOVA. * $p < 0.05$; ** $p < 0.01$. *** $p < 0.001$; **** $p < 0.0001$.

both DE and HVE cells (Figure 6E). Thus, we infer that all five motifs contribute to CR2 transcriptional activity, but do not register as DE or HVE selective in the context of this *Luciferase* reporter assay.

Histone-modification landscape and transcription-factor binding to CR1 and CR2

Lastly, we performed a meta-analysis of previously reported Chromatin immuno-precipitation analyzed by sequencing (ChIP-seq) data of histone modifications and transcription-factor binding (Aksoy et al., 2014; Cernilogar et al., 2019; Goode et al., 2016; Lie et al., 2018; Liu et al., 2015; Tosic et al., 2019) in and around CR1 and CR2 (Figure S7) to assess if our experimental results could be correlated with activation-repression marks or DE versus HVE discriminatory transcription-factor binding. In mESCs and mesendoderm cells, *Sox17* is characterized by suppressive H3K27me3 marks with either low or moderate-level H3K4me3 (promoter-activation marking) at and around both CRs, suggesting poised chromatin states. However, significant changes in the epigenetic landscape occur throughout *Sox17* locus during differentiation of mESCs to DE or HVE cells, with repressive H3K27me3 being replaced by activating H3K27ac in both lineages. There is an increase of H3K4me3 marks preferentially around CR2 in the endoderm but around both CRs in the mesoderm. In both DE and HVE lineages, CR1 and CR2 locations are coincident with valleys of the H3K27ac signal, suggesting nucleosome depletion at both sites (Calo and Wysocka, 2013; Pundhir et al., 2016; Shlyueva et al., 2014), in agreement with ATAC-seq peaks described previously (Figure 1C). A potentially important exception may be CR1 in DE cells where despite being nucleosome-free (inferred from the depleted H3K27ac signal), only limited chromatin accessibility is detected by ATAC-seq.

We note that CR1 and CR2 contain the binding motifs and are coincident with cell culture ChIP-seq determined bindings of several stem-ness and lineage-regulating TFs. For instance, binding of KLF4 and KLF5 (pluripotency factors), or BRACHYURY and EOMES (mesendodermal TFs), was detected at both CRs in mESCs (Aksoy et al., 2014) and mesendoderm (Tosic et al., 2019), respectively. Whereas CR2 contains a potential GATA binding motif, ChIP-seq data suggest that GATA4, together with FOXA2, binds only weakly to CR1 and not at all to CR2 (Cernilogar et al., 2019). On the other hand, in endothelial lineages, ETV2, GATA2, LMO2, FLI1, TAL1, GFI1/GFI1B, and RUNX1, all appear to bind at or near CR1 (Goode et al., 2016), consistent with CR1 being an endothelial cell-preferential promoter-activating region.

Identification of a distal element that binds FOXA2 and GATA4, and that interacts specifically with CR2 but not CR1

To better understand why *Sox17* may require the use of alternate promoters during development, we hypothesized the existence of distal regulatory elements that might differentially interact with each CR. Considering that CR2 is highly accessible in definitive endoderm (Figure 1C) but that the binding of FOXA2 and GATA4, two well-known endodermal TFs, are not detected in this region (Figure S7), we expanded our analysis of published ATAC-seq and ChIP-seq results (Cernilogar et al., 2019) to a 100-kb genomic region surrounding *Sox17* (± 50 kb). By doing so, we identified a third evolutionarily CR located approximately 10-kb upstream of *Sox17* that is characterized both by an open chromatin state and strong binding of both FOXA2 and GATA4 in the definitive endoderm (5 \times stronger than at CR1) (Figure 7A). Given that this CR is located within a gene desert and that *Sox17* lies nearby, we hypothesized that the FOXA2 and

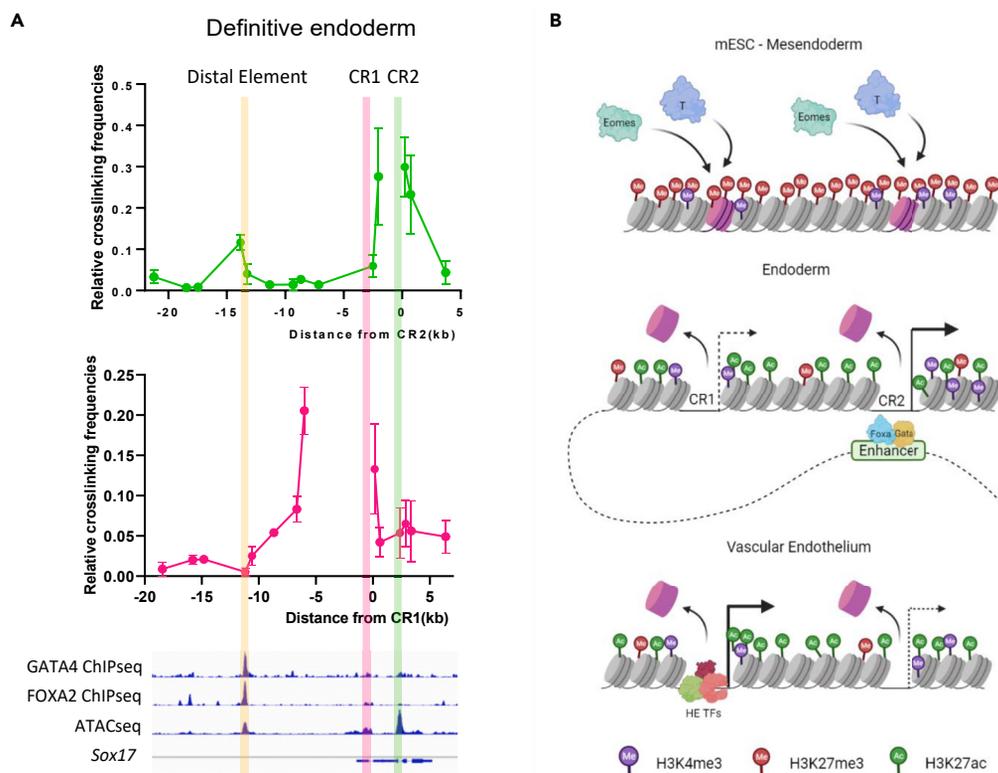


Figure 7. Multifactorial promoter functions for Sox17 CR1 and CR2 in endoderm versus endothelium lineages
(A) Chromatin conformation capture assay analyzed by qPCR (3C-qPCR) in mESC-derived definitive endoderm assesses 3D distance of proximal and distal DNA fragments to CR2- (green line) or CR1- (magenta line) containing fragment. Dots and error bars represent the mean \pm SEM. $n = 3$. Approximate position of CR1, CR2 and the identified distal element are highlighted in pink, green, and yellow columns, respectively. GATA4, FOXA2 ChIPseq, and ATACseq analyses of the corresponding DNA region to the 3C-qPCR assay are from Gene Expression Omnibus GSE116262 and GSE116255.
(B) Proposed model of alternate promoter usage to regulate Sox17 expression in endoderm and vascular endothelium. Pioneer TFs, EOMES and T, can recognize and bind to two poised CRs. Repressive H3K27me3 marking (red) is replaced by active H3K27ac (green), whereas histone cores are displaced from two CRs when Sox17 is activated in the definitive endoderm or hematopoietic endothelial (HE) cells. In endoderm, FOXA2 and GATA4 binding at distal regulatory element that spatially interacts with CR2 facilitates CR2-driven TSS2 activity usage. Meanwhile, in HE cells, CR1 is the preferentially used TSS1-driver under the regulation of several HE TFs.

GATA4 binding sequence at -10 kb may interact with CR2 to drive the expression of TSS2 in the endoderm. To test this hypothesis, we performed chromatin conformation capture (3C) analysis to assess interaction frequency between upstream genomic sequences with either CR1 or CR2. The 3C-qPCR results revealed a higher probability of the -10 -kb distal element lying in close spatial proximity to CR2 but not to CR1 (Figure 7A).

DISCUSSION

Using multi-species sequence alignment, we identified two evolutionarily conserved non-coding regions in the Sox17 gene, CR1 and CR2, located near putative TSSs for the long and short forms of Sox17 mRNA, respectively. We then determined the functional importance of these regions by performing independent deletion in mice and mutational analysis in differentiated mouse embryonic stem cells. Using 3C assay and meta-analysis of available ChIP-seq data sets, we elaborated on the multifactorial mechanism of two promoter usage to regulate Sox17 expression in different lineages.

CR1 enables transcription from TSS1

Our studies indicate that CR1 functions preferentially in endothelial cells and is necessary for the generation of long Sox17 mRNAs. As Sox17 expression in endothelial and hematopoietic lineages is vital (Kim et al., 2007), it is noteworthy that eliminating transcription from TSS1 did not adversely affect embryonic

survival. However, this is likely explained by the continuing expression of short *Sox17* mRNAs from TSS2, which remains functional in the absence of CR1. On account of this, and from our finding that there were no differences in the expression of either *Sox7* or *Sox18*, two closely related *SoxF* family members that are co-expressed with *Sox17* in endothelial cells (Lilly et al., 2017), we were unable to assess whether there was any functional redundancy exerted by these other *SoxF* family members.

Whereas *Sox17* has been linked previously to the commitment of endothelial cells to arterial fate and generation of hematopoietic lineages from endothelium (Corada et al., 2013; Kim et al., 2007), these processes were unaffected by loss of CR1. Instead, the decreased SOX17 protein level in Δ CR1 mice (Figure 3G) resulted in a modest increase in the number of lymphatic vascular cells. This finding suggests that different concentrations of *Sox17* may be required for different endothelial-related functions. Specifically, expression of *Sox17* from TSS2 alone, together with the presence of the other two *SoxF* family members, may be sufficient for the development of endothelial and hematopoietic cells but insufficient to keep lympho-vasculogenesis in check.

Consistent with *Notch1* being a central negative regulator of lymphatic vasculature development and being transcriptionally regulated by *SoxF* members (Chiang et al., 2017; Murtomaki et al., 2013), we observed decreased *Notch1* and *Dll4* expression in the Δ CR1 endothelial cells. A significant reduction of *Dll4* expression in Δ CR1 non-endothelial cells was also seen (Figure S4D), likely from pericytes and smooth muscle cells, which are known to contribute to blood vessel formation and maintenance (Bergers and Song, 2005; Hungerford and Little, 1999).

We also observed that CR1 deletion causes a modest decrease in expression of short *Sox17* mRNAs, suggesting that CR1 or TSS1-based transcription somehow affects transcription initiating from TSS2. We note that whereas no gross morphological abnormalities were seen in endoderm-derived tissues in Δ CR1 mice at either embryonic or adult stages, we cannot rule out subtle abnormalities that may manifest themselves during adult aging.

CR2 enables transcription from TSS2

In contrast to the mild effects of removing CR1, removing CR2 produced a more dramatic and complex phenotype arising from disrupted transcriptional initiation from TSS2 and defective splicing of transcripts originating from TSS1. First, the loss of CR2 results in embryonic lethality before E12.5 from a spectrum of endodermal and endothelial-related deficiencies. Variations in the stage of death and phenotype are likely owing to our use of an outbred mouse strain as a non-homogenous genetic background that is known to influence phenotypic expressivity and penetrance (Doetschman, 2009). In particular, *Sox17* gene deletion in a C57BL/6 genetic background was reported to exhibit a more severe phenotype than in a mixed background (Uemura et al., 2013). In general, however, the major phenotypes of Δ CR2 mice are notably milder than those reported for global (Kanai-Azuma et al., 2002) or endothelium-specific (e.g., *Tie2-cre*; *Sox17^{F/F}*) knockouts of *Sox17* (Kim et al., 2007) perhaps reflecting the incomplete elimination of SOX17 protein (Figure 3G). Conversely, the Δ CR2 phenotype is more severe than previous *Sox17* endoderm-specific knockout under either *Pdx1-cre* (regional endoderm) or *FoxA3-cre* (pan-endoderm) (Spence et al., 2009), possibly related to additional abnormal vasculature as the result of deficient SOX17 protein expression from the intron-retaining long mRNAs in the endothelium of Δ CR2 embryos.

The unexpectedly defective mRNA processing in Δ CR2 embryos complicated our effort to determine unequivocally the role of CR2 in regulating *Sox17* transcription. Although CR2 contains no consensus splice donor/acceptor or branch sites, deleting this short region revealed it is necessary for efficient exon 3–exon 4 splicing. Intron retention in long-form *Sox17* mRNA in the Δ CR2 context is predicted to obstruct protein translation via nuclear sequestration or inhibitory transcript secondary structure (Jacob and Smith, 2017), possibly linked to changes in the μ ORF (Figure S3A). Thus, despite increased levels of long *Sox17* mRNAs in the absence of CR2, intron retention prevents CR1-driven TSS1-derived transcripts from compensating.

It is unclear why the deletion of CR2 showed such a regional effect on organogenesis from within the posterior foregut but apparently not from other parts of the endodermal gut tube. The scarcity of PDX1 in the proto-dorsal pancreatic bud and the presence of PDX1-positive cells mixed into the hepatic HNF4A-expressing domain strongly suggests that *Sox17* transcription from CR2 contributes to the formation of the pancreas and segregation of the hepato-pancreato-biliary system (Spence et al., 2009). The absence

of CR2 did not affect the development of the lung, thyroid, and thymus buds – at least over the stages that could be examined pre-lethality – in line with findings from others that the development of these organs is independent of *Sox17* (Kanai-Azuma et al., 2002), or only requires *Sox17* at later stages (Lange et al., 2014; Zhu et al., 2012). In either case, it may be informative to determine the effects of more discrete mutations in CR2 to determine whether the formation of the hepato-pancreato-biliary system is affected, whether embryonic lethality might be circumvented, and whether CR2 is also necessary for the development or function of other endodermal-derived organs.

Multilineage regulation of *Sox17* expression through two conserved promoter regions

Given the complex and potential compensatory functions of CR1 and CR2 in mice, we turned to reporter gene studies in cell culture and found that whereas both CR1 and CR2 exhibited promoter activity, CR1 activity was limited to endothelial cells, but CR2 was active in both endothelial and endodermal cells. We also aligned and examined the TF-binding and histone-modification data reported by others, and identified a potential distal regulatory element that binds endodermal TFs and interacts with CR2 but not CR1 in the endoderm. Taken together, we propose a stepwise model for how the alternate promoters in *Sox17* are differentially activated to discriminate two lineages during development (Figure 7B).

Consistent with previous reports (Aksoy et al., 2014; Tosic et al., 2019), we speculate that the binding of KLF4 and/or KLF5 at two poised conserved promoter regions (indicated by H3K27me3 and a low level of H3K4me3) maintains *Sox17* expression at a low level in mESCs, and that pioneer TFs such as EOMES and BRACHYURY (T) replace the pluripotent factors and open the chromatin of *Sox17* locus in mesodermal differentiation from the naive stem-cell state. This leads to swapping out of repressive H3K27me3 marks for active H3K27ac marks, expansion of active promoter H3K4me3 marks, and the potential subsequent displacement of histone octamers at two CRs.

In the endoderm, strong binding of the pioneer TFs FOXA2 and GATA4 was observed at the –10-kb distal element that is spatially close to the CR2 promoter, corresponding to high ATAC-seq accessibility at both regions in DE (Figures 7A and S1). The co-binding of FOXA2 and GATA4 at the distal element likely reflects the enhancer priming process to initiate general endoderm differentiation via activating *Sox17* expression from TSS2 and may help reinforce CR2 promoter usage by facilitating further recruitment of lineage-specific TFs to either the distal element or to CR2 at later stages (Geusz et al., 2021). In contrast, for the endothelial-vascular lineages, strong binding at CR1 of a combination of endothelial and hematopoietic TFs—ETV2 in mesoderm; GATA2 and LMO2 in hemangioblasts; FLI1, TAL1, GFI1/GFI1B, and RUNX1 in hematopoietic endothelial and progenitor cells - is likely to drive expression of long *Sox17* mRNAs.

In addition to each CR being controlled by several regulatory layers, our results suggest functional compensation between the two promoter regions. Specifically, in CR1-null embryos, the expression of short mRNA from the downstream promoter in endothelial cells rescues the absence of long mRNAs from the upstream promoter. Moreover, an increase in the expression of long mRNA forms is also seen in CR2-null embryos, suggesting an increase in gene expression from the upstream promoter in the absence of a functional downstream promoter. These observations suggest additional and yet-to-be-explored complexities in the regulation of *Sox17* during development.

Limitations of the study

Although the Δ CR2 mutation was designed based on all information available at the time, the disruption of RNA processing by the mutation significantly complicated our interpretations of the results. In the future, more discrete mutations, such as those corresponding to discrete conserved motifs within CR2 (Figure 6D), may be able to impair promoter function without also impairing RNA splicing. Finally, although our meta-analysis of published data sets and 3C-qPCR assay provided informative results, this approach does not exclude other possible regulatory mechanisms of the two CRs including unknown TFs binding directly to CR2 and/or distal elements interacting with each CR in endoderm and endothelial cells.

Concluding comments

We have identified two functionally distinct promoters in *Sox17* that are critical for the proper development of the endoderm and vascular endothelium. The precise identification of these two regions will facilitate both the identification and functional characterization of distal regulatory elements and potential

epigenetic modulator and cell type-specific TFs in regulating the expression of this essential SoxF family member during hematopoiesis and formation of the hepato-pancreato-biliary system.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Experimental mice
 - Cell lines
- METHOD DETAILS
 - Bioinformatics analysis
 - CRISPR/Cas9 mutagenesis
 - Mouse genotyping
 - Reverse transcription
 - PCR
 - Quantitative PCR
 - Immuno-fluorescence staining
 - Light-sheet microscopy
 - *Luciferase* reporter assay
 - Fluorescence-activated cell sorting
 - Western blot analysis
 - Chromatin conformation capture assay
 - Graphical illustration, quantification and statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.104905>.

ACKNOWLEDGMENTS

We thank the staff of the Vanderbilt Genome Editing Resource (VGER), the Vanderbilt Flow Cytometry Shared Resource (FCSR), and the Vanderbilt Cell Imaging Shared Resource (CISR) for their expert assistance. VGER is supported by CA68485 and DK020593, FCSR by CA068485 and DK058404, and CISR by CA68485, DK20593, DK58404, DK59637, and EY08126. We thank Eunyoung Choi for helpful discussions and reading the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization, L.T.T. and M.A.M.; methodology, L.T.T., L.S., and A.B.O.; formal analysis, L.T.T.; investigation, L.T.T., A.B.O., and J.W.; resources, C.V.E.W.; data curation, L.T.T.; writing – original draft, L.T.T. and M.A.M.; writing – review and editing, L.T.T., A.B.O., L.S., C.V.E.W., and M.A.M.; visualization, L.T.T.; supervision, C.V.E.W. and M.A.M.; funding acquisition, M.A.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 21, 2021

Revised: April 6, 2022

Accepted: August 5, 2022

Published: September 16, 2022

REFERENCES

- Aksoy, I., Giudice, V., Delahaye, E., Wianny, F., Aubry, M., Mure, M., Chen, J., Jauch, R., Bogu, G.K., Nolden, T., et al. (2014). Klf4 and Klf5 differentially inhibit mesoderm and endoderm differentiation in embryonic stem cells. *Nat. Commun.* 5, 3719.
- Bergers, G., and Song, S. (2005). The role of pericytes in blood-vessel formation and maintenance. *Neuro Oncol.* 7, 452–464.
- Borowiak, M., Maehr, R., Chen, S., Chen, A.E., Tang, W., Fox, J.L., Schreiber, S.L., and Melton, D.A. (2009). Small molecules efficiently direct endodermal differentiation of mouse and human embryonic stem cells. *Cell Stem Cell* 4, 348–358.
- Burke, Z., and Oliver, G. (2002). Prox1 is an early specific marker for the developing liver and pancreas in the mammalian foregut endoderm. *Mech. Dev.* 118, 147–155.
- Burtscher, I., Barkey, W., Schwarzfischer, M., Theis, F.J., and Lickert, H. (2012). The Sox17-mCherry fusion mouse line allows visualization of endoderm and vascular endothelial development. *Genesis* 50, 496–505.
- Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? *Mol. Cell.* 49, 825–837.
- Cernilogar, F.M., Hasenöder, S., Wang, Z., Scheibner, K., Burtscher, I., Sterr, M., Smialowski, P., Groh, S., Evenroed, I.M., Gilfillan, G.D., et al. (2019). Pre-marked chromatin and transcription factor co-binding shape the pioneering activity of Foxa2. *Nucleic. Acids. Res.* 47, 9069–9086.
- Chiang, I.K.N., Fritzsche, M., Pichol-Thieuvend, C., Neal, A., Holmes, K., Lagendijk, A., Overman, J., D'angelo, D., Omini, A., Hermkens, D., et al. (2017). SoxF factors induce Notch1 expression via direct transcriptional regulation during early arterial development. *Development* 144, 2629–2639.
- Chiang, P.M., and Wong, P.C. (2011). Differentiation of an embryonic stem cell to hemogenic endothelium by defined factors: essential role of bone morphogenetic protein 4. *Development* 138, 2833–2843.
- Choi, E., Kraus, M.R.C., Lemaire, L.A., Yoshimoto, M., Vemula, S., Potter, L.A., Manduchi, E., Stoeckert, C.J., Jr., Grapin-Botton, A., and Magnuson, M.A. (2012). Dual lineage-specific expression of Sox17 during mouse embryogenesis. *Stem. Cell.* 30, 2297–2308.
- Clements, D., and Woodland, H.R. (2003). VegT induces endoderm by a self-limiting mechanism and by changing the competence of cells to respond to TGF-beta signals. *Dev. Biol.* 258, 454–463.
- Corada, M., Orsenigo, F., Morini, M.F., Pitulescu, M.E., Bhat, G., Nyqvist, D., Breviaro, F., Conti, V., Briot, A., Iruela-Arispe, M.L., et al. (2013). Sox17 is indispensable for acquisition and maintenance of arterial identity. *Nat. Commun.* 4, 2609.
- Doetschman, T. (2009). Influence of genetic background on genetically engineered mouse phenotypes. *Methods. Mol. Biol.* 530, 423–433.
- Engert, S., Burtscher, I., Liao, W.P., Dulev, S., Schotta, G., and Lickert, H. (2013). Wnt/beta-catenin signalling regulates Sox17 expression and is essential for organizer and endoderm formation in the mouse. *Development* 140, 3128–3138.
- Engleka, M.J., Craig, E.J., and Kessler, D.S. (2001). VegT activation of Sox17 at the midblastula transition alters the response to nodal signals in the vegetal endoderm domain. *Dev. Biol.* 237, 159–172.
- Fornes, O., Castro-Mondragon, J.A., Khan, A., Van Der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D., et al. (2020). JaspAr 2020: update of the open-access database of transcription factor binding profiles. *Nucleic. Acids. Res.* 48, D87–D92.
- Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic. Acids. Res.* 32, W273–W279.
- Geusz, R.J., Wang, A., Lam, D.K., Vinckier, N.K., Alysandratos, K.D., Roberts, D.A., Wang, J., Kefalopoulou, S., Ramirez, A., Qiu, Y., et al. (2021). Sequence logic at enhancers governs a dual mechanism of endodermal organ fate induction by FOXA pioneer factors. *Nat. Commun.* 12, 6636.
- González-Hernández, S., Gómez, M.J., Sánchez-Cabo, F., Méndez-Ferrer, S., Muñoz-Cánoves, P., and Isern, J. (2020). Sox17 controls emergence and remodeling of nestin-expressing coronary vessels. *Circ. Res.* 127, e252–e270.
- Goode, D.K., Obier, N., Vijayabaskar, M.S., Lie-A-Ling, M., Lilly, A.J., Hannah, R., Lichtinger, M., Batta, K., Florkowska, M., Patel, R., et al. (2016). Dynamic gene regulatory networks drive hematopoietic specification and differentiation. *Dev. Cell.* 36, 572–587.
- Heinke, J., Patterson, C., and Moser, M. (2012). Life is a pattern: vascular assembly within the embryo. *Front. Biosci.* 4, 2269–2288.
- Hosking, B., François, M., Wilhelm, D., Orsenigo, F., Caprini, A., Svingen, T., Tutt, D., Davidson, T., Browne, C., Dejana, E., and Koopman, P. (2009). Sox7 and Sox17 are strain-specific modifiers of the lymphangiogenic defects caused by Sox18 dysfunction in mice. *Development* 136, 2385–2391.
- Howard, L., Rex, M., Clements, D., and Woodland, H.R. (2007). Regulation of the *Xenopus* *xsox17alpha(1)* promoter by co-operating VegT and Sox17 sites. *Dev. Biol.* 310, 402–415.
- Huelsken, J., Vogel, R., Brinkmann, V., Erdmann, B., Birchmeier, C., and Birchmeier, W. (2000). Requirement for beta-catenin in anterior-posterior axis formation in mice. *J. Cell. Biol.* 148, 567–578.
- Hungerford, J.E., and Little, C.D. (1999). Developmental biology of the vascular smooth muscle cell: building a multilayered vessel wall. *J. Vasc. Res.* 36, 2–27.
- Jacob, A.G., and Smith, C.W.J. (2017). Intron retention as a component of regulated gene expression programs. *Hum. Genet.* 136, 1043–1057.
- Jung, H.S., Uenishi, G., Park, M.A., Liu, P., Suknuntha, K., Raymond, M., Choi, Y.J., Thomson, J.A., Ong, I.M., and Slukvin, I. (2021). SOX17 integrates HOXA and arterial programs in hemogenic endothelium to drive definitive lympho-myeloid hematopoiesis. *Cell Rep.* 34, 108758.
- Kamachi, Y., and Kondoh, H. (2013). Sox proteins: regulators of cell fate specification and differentiation. *Development* 140, 4129–4144.
- Kanai-Azuma, M., Kanai, Y., Gad, J.M., Tajima, Y., Taya, C., Kurohmaru, M., Sanai, Y., Yonekawa, H., Yazaki, K., Tam, P.P.L., and Hayashi, Y. (2002). Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development* 129, 2367–2379.
- Katoh, M. (2002). Molecular cloning and characterization of human SOX17. *Int. J. Mol. Med.* 9, 153–157.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome. Res.* 12, 996–1006.
- Kim, I., Saunders, T.L., and Morrison, S.J. (2007). Sox17 dependence distinguishes the transcriptional regulation of fetal from adult hematopoietic stem cells. *Cell* 130, 470–483.
- King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., et al. (2008). The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451, 783–788.
- Lange, A.W., Haitchi, H.M., Lecras, T.D., Sridharan, A., Xu, Y., Wert, S.E., James, J., Udell, N., Thurner, P.J., and Whitsett, J.A. (2014). Sox17 is required for normal pulmonary vascular morphogenesis. *Dev. Biol.* 387, 109–120.
- Lee, K., Cho, H., Rickert, R.W., Li, Q.V., Pulecio, J., Leslie, C.S., and Huangfu, D. (2019). FOXA2 is required for enhancer priming during pancreatic differentiation. *Cell Rep.* 28, 382–393.e7.
- Lee, S., Kim, I.K., Ahn, J.S., Woo, D.C., Kim, S.T., Song, S., Koh, G.Y., Kim, H.S., Jeon, B.H., and Kim, I. (2015). Deficiency of endothelium-specific transcription factor Sox17 induces intracranial aneurysm. *Circulation* 131, 995–1005.
- Lefebvre, V., Dumitriu, B., Penzo-Méndez, A., Han, Y., and Pallavi, B. (2007). Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. *Int. J. Biochem. Cell. Biol.* 39, 2195–2214.
- Liao, W.P., Uetzmann, L., Burtscher, I., and Lickert, H. (2009). Generation of a mouse line expressing Sox17-driven Cre recombinase with specific activity in arteries. *Genesis* 47, 476–483.
- Lie, A.L.M., Marinopoulou, E., Lilly, A.J., Challinor, M., Patel, R., Lancrin, C., Kouskoff, V., and Lacaud, G. (2018). Regulation of RUNX1 dosage is crucial for efficient blood formation from hemogenic endothelium. *Development* 145.

- Lilly, A.J., Lacaud, G., and Kouskoff, V. (2017). SOXF transcription factors in cardiovascular development. *Semin. Cell. Dev. Biol.* **63**, 50–57.
- Liu, F., Li, D., Yu, Y.Y.L., Kang, I., Cha, M.J., Kim, J.Y., Park, C., Watson, D.K., Wang, T., and Choi, K. (2015). Induction of hematopoietic and endothelial cell program orchestrated by ETS transcription factor ER71/ETV2. *EMBO. Rep.* **16**, 654–669.
- Liu, P., Wakamiya, M., Shea, M.J., Albrecht, U., Behringer, R.R., and Bradley, A. (1999). Requirement for Wnt3 in vertebrate Axis formation. *Nat. Genet.* **22**, 361–365.
- Lizama, C.O., Hawkins, J.S., Schmitt, C.E., Bos, F.L., Zape, J.P., Cautivo, K.M., Borges Pinto, H., Rhyner, A.M., Yu, H., Donohoe, M.E., et al. (2015). Repression of arterial genes in hemogenic endothelium is sufficient for haematopoietic fate acquisition. *Nat. Commun.* **6**, 7739.
- Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., Sahin, S., Abugessaisa, I., Fukuda, S., Hori, F., Ishikawa-Kato, S., et al.; FANTOM consortium (2015). Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome. Biol.* **16**, 22.
- Lowe, L.A., Yamada, S., and Kuehn, M.R. (2001). Genetic dissection of nodal function in patterning the mouse embryo. *Development* **128**, 1831–1843.
- Matsui, T., Kanai-Azuma, M., Hara, K., Matoba, S., Hiramatsu, R., Kawakami, H., Kurohmaru, M., Koopman, P., and Kanai, Y. (2006). Redundant roles of Sox17 and Sox18 in postnatal angiogenesis in mice. *J. Cell. Sci.* **119**, 3513–3526.
- Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S., and Dubchak, I. (2000). Vista : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **16**, 1046–1047.
- Murtomaki, A., Uh, M.K., Choi, Y.K., Kitajewski, C., Borisenko, V., Kitajewski, J., and Shawber, C.J. (2013). Notch1 functions as a negative regulator of lymphatic endothelial cell differentiation in the venous endothelium. *Development* **140**, 2365–2376.
- Naumova, N., Smith, E.M., Zhan, Y., and Dekker, J. (2012). Analysis of long-range chromatin interactions using Chromosome Conformation Capture. *Methods* **58**, 192–203.
- Niakan, K.K., Ji, H., Maehr, R., Vokes, S.A., Rodolfa, K.T., Sherwood, R.I., Yamaki, M., Dimos, J.T., Chen, A.E., Melton, D.A., et al. (2010). Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes. Dev.* **24**, 312–326.
- Nonaka, H., Tanaka, M., Suzuki, K., and Miyajima, A. (2007). Development of murine hepatic sinusoidal endothelial cells characterized by the expression of hyaluronan receptors. *Dev. Dyn.* **236**, 2258–2267.
- Pijuan-Sala, B., Wilson, N.K., Xia, J., Hou, X., Hannah, R.L., Kinston, S., Calero-Nieto, F.J., Poirion, O., Preissl, S., Liu, F., and Göttgens, B. (2020). Single-cell chromatin accessibility maps reveal regulatory programs driving early mouse organogenesis. *Nat. Cell. Biol.* **22**, 487–497.
- Pundhir, S., Bagger, F.O., Lauridsen, F.B., Rapin, N., and Porse, B.T. (2016). Peak-valley-peak pattern of histone modifications delineates active regulatory elements and their directionality. *Nucleic. Acids. Res.* **44**, 4037–4051.
- Reim, G., Mizoguchi, T., Stainier, D.Y., Kikuchi, Y., and Brand, M. (2004). The POU domain protein spg (pou2/Oct4) is essential for endoderm formation in cooperation with the HMG domain protein casanova. *Dev. Cell.* **6**, 91–101.
- Saba, R., Kitajima, K., Rainbow, L., Engert, S., Uemura, M., Ishida, H., Kokkinopoulos, I., Shintani, Y., Miyagawa, S., Kanai, Y., et al. (2019). Endocardium differentiation through Sox17 expression in endocardium precursor cells regulates heart development in mice. *Sci. Rep.* **9**, 11953.
- Sakamoto, Y., Hara, K., Kanai-Azuma, M., Matsui, T., Miura, Y., Tsunekawa, N., Kurohmaru, M., Saijoh, Y., Koopman, P., and Kanai, Y. (2007). Redundant roles of Sox17 and Sox18 in early cardiovascular development of mouse embryos. *Biochem. Biophys. Res. Commun.* **360**, 539–544.
- Sarkar, A., and Hochedlinger, K. (2013). The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell* **12**, 15–30.
- Schepers, G.E., Teasdale, R.D., and Koopman, P. (2002). Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. *Dev. Cell* **3**, 167–170.
- Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286.
- Spence, J.R., Lange, A.W., Lin, S.C.J., Kaestner, K.H., Lowy, A.M., Kim, I., Whitsett, J.A., and Wells, J.M. (2009). Sox17 regulates organ lineage segregation of ventral foregut progenitor cells. *Dev. Cell* **17**, 62–74.
- Theiler, K. (1989). *The House Mouse Atlas of Embryonic Development* (Springer Berlin).
- Tosic, J., Kim, G.J., Pavlovic, M., Schröder, C.M., Mersiowsky, S.L., Barg, M., Hofherr, A., Probst, S., Köttgen, M., Hein, L., and Arnold, S.J. (2019). Eomes and Brachyury control pluripotency exit and germ-layer segregation by changing the chromatin state. *Nat. Cell. Biol.* **21**, 1518–1531.
- Tremblay, K.D., Hoodless, P.A., Bikoff, E.K., and Robertson, E.J. (2000). Formation of the definitive endoderm in mouse is a Smad2-dependent process. *Development* **127**, 3079–3090.
- Uemura, M., Ozawa, A., Nagata, T., Kurasawa, K., Tsunekawa, N., Nobuhisa, I., Taga, T., Hara, K., Kudo, A., Kawakami, H., et al. (2013). Sox17 haploinsufficiency results in perinatal biliary atresia and hepatitis in C57BL/6 background mice. *Development* **140**, 639–648.
- Vincent, S.D., Dunn, N.R., Hayashi, S., Norris, D.P., and Robertson, E.J. (2003). Cell fate decisions within the mouse organizer are governed by graded Nodal signals. *Genes. Dev.* **17**, 1646–1662.
- Zhu, Y., Li, Y., Jun Wei, J.W., and Liu, X. (2012). The role of Sox genes in lung morphogenesis and cancer. *Int. J. Mol. Sci.* **13**, 15767–15783.
- Zorn, A.M., and Wells, J.M. (2007). *Molecular Basis of Vertebrate Endoderm Development* (Elsevier).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Goat anti-Sox17 (1:100 dilution for immunofluorescence staining – IF, 1:200 dilution for western blot - WB)	R&D Systems	Cat# AF1924, RRID:AB_355060
Rabbit anti-FoxA2/HNF3β (D56D6) (1:400 dilution for IF)	Cell Signaling Technology	Cat# 8186, RRID:AB_10891055
Goat anti-human FoxA2/HNF3β (1:100 dilution for IF)	R&D Systems	Cat# AF2400, RRID:AB_2294104
Rat anti-mouse CD31 (1:100 dilution for IF)	BD Biosciences	Cat# 550,274, RRID:AB_393571
Guinea pig anti-Pdx1 (1:1000 dilution for IF)	Christopher V.E. Wright's lab	N/A
Rabbit anti-HNF4α (C11F12) (1:500 dilution for IF)	Cell Signaling Technology	Cat# 3113, RRID:AB_2295208
Rabbit anti-GAPDH (D16H11) (1:1000 dilution for WB)	Cell Signaling Technology	Cat# 5174, RRID:AB_10622025
Alexa Fluor 488 conjugated Donkey anti-Rat IgG (1:1000 dilution for IF)	Thermo Fisher Scientific	Cat# A21208, RRID:AB_2535794
Alexa Fluor 647 conjugated Donkey anti-Goat IgG (1:1000 dilution for IF)	Thermo Fisher Scientific	Cat# A21447, RRID:AB_2535864
Alexa Fluor 555 conjugated Donkey anti-Rabbit IgG (1:1000 dilution for IF)	Thermo Fisher Scientific	Cat# A31572, RRID:AB_162543
Alexa Fluor 488 conjugated Donkey anti-Goat IgG (1:1000 dilution for IF)	Thermo Fisher Scientific	Cat# A11055, RRID:AB_2534102
Alexa Fluor 647 conjugated Donkey anti-Guinea pig IgG (1:1000 dilution for IF)	Jackson ImmunoResearch	Cat# 706-605-148, RRID: AB_2340476
HRP conjugated Mouse anti-Rabbit light chain (1:1000 dilution for WB)	Jackson ImmunoResearch	Cat# 211-032-171, RRID: AB_2339149
HRP conjugated Donkey anti-Goat IgG (1:1000 dilution for WB)	Jackson ImmunoResearch	Cat# 705-035-003, RRID: AB_2340390
PE conjugated Rat anti-Mouse Lyve-1 (156 ng/10 ⁶ cells for flow cytometry analysis)	R&D Systems	Cat# FAB2125P, RRID:AB_10889020
APC conjugated Rat anti-Mouse Cd31 (7.8 ng/10 ⁶ cells for flow cytometry analysis)	BD Biosciences	Cat# 561814, RRID:AB_10893351
PE conjugated Mouse anti-Human EphrinB2 (60 ng/10 ⁶ cells for low cytometry analysis)	Santa Cruz Biotechnology	Cat# sc-398735, RRID:AB_2895232
PE conjugated Rat anti-Human EphB4 (50ng/10 ⁶ cells for flow cytometry analysis)	R&D Systems	Cat# FAB3038, RRID:AB_2293626
Bacterial and virus strains		
DH5 alpha competent cells	Thermo Fisher Scientific	Cat# EC0112
Chemicals, peptides, and recombinant proteins		
Mitomycin C	Tocris	Cat# 3258/10
DMEM	Thermo Fisher Scientific	Cat# 11960-044
Fetal Bovine Serum – Premium Select	R&D Systems	Cat# 11550
Non-essential amino acids	Thermo Fisher Scientific	Cat# 11140-050
L-glutamine	Thermo Fisher Scientific	Cat# 25030-081
Gentamicin	Thermo Fisher Scientific	Cat# 15750-060
MLIF	MilliporeSigma	Cat# ESG1107
β-mercaptoethanol	Thermo Fisher Scientific	Cat# 21985-023
0.25% Trypsin	Thermo Fisher Scientific	Cat# 15050057

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
IDE2	Stemcell Technologies	Cat# 72522
Advanced RPMI 1640	Thermo Fisher Scientific	Cat# 12633-012
DMEM/F12	Thermo Fisher Scientific	Cat# 11330-032
Neurobasal medium	Thermo Fisher Scientific	Cat# 21103-049
B27 minus vitamin A	Thermo Fisher Scientific	Cat# 12587010
N2	Thermo Fisher Scientific	Cat# 17502-048
GSK3 inhibitor (CHIR99021)	LC Laboratories	Cat# C-6556
Recombinant Activin A	Thermo Fisher Scientific	Cat# PHG9014
Recombinant human FGF-basic (FGF-2)	PeproTech	Cat# 100-18B
Recombinant Bmp4	R&D systems	Cat# 5020-BP/CF
Recombinant Vegf	Novus Biologicals	Cat# NBP2-35189
Bc-cAMP	MilliporeSigma	Cat# B5386
ALK inhibitor	Stemcell Technologies	Cat# 72234
Proteinase K	Thermo Fisher Scientific	Cat# EO0491
EcoRV Restriction enzyme	New England BioLabs	Cat# R0195
NheI-HF Restriction enzyme	New England BioLabs	Cat# R3131
XmaI Restriction enzyme	New England BioLabs	Cat# R0180
Phosphate Buffer Saline 10X	Corning	Cat# 46-013-CM
Tissue-Tek OCT compound	Sakura	Cat# 4583
Bovine Serum Albumin	MilliporeSigma	Cat# A3059
RIPA buffer	MilliporeSigma	Cat# R0278
Protease Inhibitor Cocktail/Mammalian	MilliporeSigma	Cat# P8340
Phosphatase inhibitor cocktail II	MilliporeSigma	Cat# P5726
Phosphatase inhibitor cocktail III	MilliporeSigma	Cat# P0044
4X Laemmli Sample Buffer	Bio-Rad Laboratories	Cat# 1610747
Mini-Protean TGX 10% precast gel	Bio-Rad Laboratories	Cat# 4561035
Ponceau S solution	MilliporeSigma	Cat# P7170
SuperSignal West Dura Extended Duration Substrate	Thermo Fisher Scientific	Cat# 34075
EconoTaq PLUS 2X Master Mix	Biosearch Technologies	Cat# 30035
SYBR™ Green PCR Master Mix	Thermo Fisher Scientific	Cat# 4309155
VECTASHIELD Vibrance Antifade Mounting Medium	Vector Laboratories	Cat# H-1700
HF Phusion polymerase	New England BioLabs	Cat# M0530
DNase I	Thermo Fisher Scientific	Cat# AM2222
Accumax	MilliporeSigma	Cat# A7089
Histodenz	MilliporeSigma	Cat# D2158
Sodium azide	MilliporeSigma	Cat# S2002
DABCO (1,4-diazabicyclo[2.2.2]octane)	MilliporeSigma	Cat# D27802
BglII Restriction enzyme	New England BioLabs	Cat# R0144
T4 DNA ligase	New England BioLabs	Cat# M0202
Phenol-chloroform-isoamyl alcohol	MilliporeSigma	Cat# P2069

Critical commercial assays

Maxwell® 16 LEV simplyRNA Purification Kit	Promega	Cat# AS1280
High Capacity cDNA Reverse Transcription Kit	Thermo Fisher Scientific	Cat# 4368814
Vector TrueVIEW autofluorescence quenching Kit	Vector Laboratories	Cat# SP-8400-15
Xfect Transfection Reagent Kit	Takara Bio	Cat# 631318

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Dual-Luciferase Reporter Assay kit	Promega	Cat# E1910
Pierce BCA protein assay	Thermo Fisher Scientific	Cat# 23225
Amicon ultra-0.5 30k purification device	MilliporeSigma	Cat# UFC503024

Experimental models: Cell lines

TL1 mESCs	Gifted to Vanderbilt Genome Editing Resource	
DR4 MEFs	Derived at Vanderbilt Genome Editing Resource	

Experimental models: Organisms/strains

CD1-Sox17 ^{em1Mgn} /Vu (Sox17 CR1 line)	Vanderbilt Cryopreserved Mouse Repository	ID: 16563
CD1-Sox17 ^{em2Mgn} /Vu (Sox17 CR2 line)	Vanderbilt Cryopreserved Mouse Repository	ID: 16564
Sox17 ^{tm1.3(Cre.GFP)Mgn}	Mutant Mouse Resource & Research Centers	RRID:MMRRC_036463-UNC
B6.Cg-Gt(ROSA)26Sor ^{tm14(CAG-tdTomato)Hze/J}	The Jackson Laboratory	RRID:IMSR_JAX:007914

Oligonucleotides

sgRNA and ssDNA sequences to generate mouse models	See Table S2
PCR primers for genotyping and to generate cloning inserts	See Table S3
qPCR primers	See Table S4
Mutant CR2 oligonucleotides	See Table S6

Recombinant DNA

pGL4.14[luc2/Hygro]	Promega	Cat# E6691
pRL-SV40	Promega	Cat# E2231
pGL2-Control vector	Promega	Cat# E1611
Sox17 BAC (RPCI-22)	BACPAC (CHORI)	Index# 4916

Software and algorithms

FIJI	http://fiji.sc	RRID:SCR_002285
Prism 9	GraphPad	RRID:SCR_002798
BioRender	http://biorender.com	RRID:SCR_018361
Integrative Genomics Viewer	http://www.broadinstitute.org/igv/	RRID:SCR_011793
UCSC Genome Browser	http://genome.ucsc.edu/	RRID:SCR_005780
VISTA Browser (VISTA-point)	http://pipeline.lbl.gov/cgi-bin/gateway2	RRID:SCR_011808
Imaris	http://www.bitplane.com/imaris/imaris	RRID:SCR_007370

RESOURCE AVAILABILITY

Lead contact

Additional information and requests for resources should be directed to the lead contact, Mark A. Magnuson (mark.magnuson@vanderbilt.edu).

Materials availability

Mouse lines generated in this study are available from the Vanderbilt Cryopreserved Mouse Repository (https://labnodes.vanderbilt.edu/resource/view/id/14860/community_id/2613).

Plasmids generated in this study are available upon request. Please contact the lead author, Mark A. Magnuson (mark.magnuson@vanderbilt.edu)

Data and code availability

This study did not generate any new datasets or software code. Original datasets and processed visualization files of chromatin immunoprecipitation analyzed by sequencing (ChIP-seq) used in this paper are available at Gene Expression Omnibus (GEO). Detailed GEO accession numbers for the used datasets were provided in corresponding figure legends.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Experimental mice

Two mutant mouse lines were designed and produced using CRISPR/Cas9 in collaboration with the Vanderbilt Genome Editing Resource in Nashville, TN, USA. The Vanderbilt Institutional Animal Care and Use Committee approved all experimental procedures in accordance with the ethics guidelines at Vanderbilt. Mice are socially housed within Vanderbilt's animal facility with a 12-hour light-dark cycle. Both mouse lines are maintained in an outbred Crl:CD1(ICR), or CD-1® IGS background (Charles River, model #022) by breeding heterozygous males with wild-type CD1 females obtained from Charles River.

For stage-specific embryo collections, timed matings were performed by placing 8-week-old heterozygous male mice that had been single-housed for 1-2 weeks with 1 to 2 group-housed heterozygous females that were also 6-8 weeks old. Female mice with vaginal plugs were defined as being 0.5 days *post coitum* (or embryonic day 0.5 – E0.5). At desired developmental time points, embryonic tissues were micro-dissected in cold phosphate buffered saline (PBS) and staged according to Theiler staging criteria.

Mouse genotypes were determined using DNA extracted from adult tail biopsy or embryonic yolk sac lysates. Wild-type embryos from the same litters were used as controls for CR-null embryos.

Cell lines

The TL1 mouse embryonic stem cell (mESC) line was derived from 129S6/SvEvTac mice by E4.5 blastocyst outgrowth. The cell line was not authenticated

TL1 mESCs were routinely cultured at 37°C in 5% CO₂ on mitomycin C-treated DR4 mouse embryonic fibroblast (MEF) feeder cells using mESC complete medium prepared from Dulbecco's Modified Eagle Medium (DMEM) supplemented with 15% heat-inactivated Fetal Bovine Serum (FBS), 1mM non-essential amino acids (Thermo Fisher Scientific), 2 mM L-glutamine, 1X gentamicin, 10³ U/mL mouse Leukemia Inhibitory Factor (mLIF) and 0.11 mM β-mercaptoethanol. Cells were split 1:6-1:8 at approximately 80% confluence using 0.25% trypsin (5 minutes at 37°C). For differentiation, feeder MEFs were depleted from mESCs by trypsinization and plating of mixed cells on gelatinized plates for 30 min to 1 h at 37°C, 5% CO₂. After incubation, floating MEF-depleted mESCs were pelleted by centrifugation at 1000 rpm for 5 minutes, resuspended at desired concentrations and plated on gelatin-coated plates for further differentiation.

Generation of definitive endoderm. mESCs were differentiated to definitive endoderm using IDE2, a small molecule, as previously described ([Borowiak et al., 2009](#)) with minor modifications. MEF-depleted mESCs were seeded at 5,000 cells/cm² overnight in mESC complete medium. After 24 hours (start day 1), mESC complete medium was replaced with advanced RPMI 1640 medium supplemented with 2 mM L-glutamine, 0.2% heat-inactivated FBS and 5 μM IDE2.

Generation of vascular endothelial cells. mESCs were differentiated to vascular endothelial cells using another previously described protocol ([Chiang and Wong, 2011](#)) with minor modifications. In short, MEF-depleted mESCs were plated at 250 cells/cm² in basal medium prepared from 1:1 ratio of DMEM/F12 and Neurobasal medium, supplemented with 1X B27 without vitamin A, 1X N2 and 55 μM β-mercaptoethanol. After 48 hours, differentiating cells were switched to mesodermal inducing media containing basal media supplemented with 3 μM GSK inhibitor, 4 ng/mL Activin A, 12.5 ng/mL Fgf2 and 5 ng/mL Bmp4. After 48 hours the cells were switched to vasculogenic medium containing basal media supplemented with 12.5 ng/mL Fgf2, 20 ng/mL Bmp4, 20 ng/mL Vegf, 0.25 mM Br-cAMP and 4 μM ALK inhibitor for the remainder of the differentiation time.

METHOD DETAILS

Bioinformatics analysis

Boundaries of deleted regions (CR1: chr1.4,496,413 - chr1.4,496,529 and CR2: chr1.4,493,622 - chr1.4,493,747) were determined based on sequence conservation analysis with UCSC genome browser (<http://genome.ucsc.edu>) on mouse assembly mm10 (Conservation track using Multiz alignments and PHAST package) and VISTA-Point multiple genomes alignment tool (<http://pipeline.lbl.gov/cgi-bin/gateway2>) (Mayor et al., 2000).

Open reading frames (ORFs) analysis was performed using Show Translations feature of SnapGene software. To predict all possible ORFs from +1/+2/+3 frames, Translation Options was set to show Top 3 frames and to define ORFs at a minimum length of 10 amino acids with AUG as the required start codon. SnapGene software was also used to search for consensus RNA splice sites listed in Table S7.

CRISPR/Cas9 mutagenesis

Mouse lines lacking CR1 and CR2 were generated by pronuclear microinjection of a Cas9 ribonucleoprotein complex into fertilized zygotes from the mating of CD-1 mice. The injection solution contained 100 ng/ μ l Cas9 protein, two crRNAs (crRNA + tracrRNA) at 25 ng/ μ l each, 25 ng/ μ l tracrRNA, and 50 ng/ μ l of a 180 base oligonucleotide in 10 mM Tris, 0.1 mM EDTA buffer at pH 7.6 (TEKnova #T0230). See Table S2 for crRNA and ssDNA sequences. Injected embryos were implanted into pseudo-pregnant CD-1 mice. Pups were weaned and tail biopsies performed at three weeks of age. Founder mice carrying desired deletions were identified by PCR (see Table S3 for primer sequences) and confirmed by Sanger sequencing. Founder animals were backcrossed to wild type CD-1 mice for 3 generations prior to interbreeding to produce homozygous mutant mice. All experimental procedures were approved by the Vanderbilt Institutional Animal Care and Use Committee.

Mouse genotyping

Genomic DNA was extracted from adult tail tips or embryonic yolk sacs digested with 0.5 mg/mL Proteinase K in lysis buffer (10 mM Tris pH 8.0, 100 mM NaCl, 10 mM EDTA and 0.5% SDS) at 55°C overnight. PCR reactions used the EconoTaq PLUS 2X Master Mix standard protocol (see Table S3 for primer sequences) and the resulting products were analyzed using 2% agarose gel electrophoresis.

Reverse transcription

RNA was extracted from embryos, sorted cells or cultured cells using Promega Maxwell® 16 LEV simplyRNA Purification Kit then converted to cDNA using a High Capacity cDNA reverse Transcription kit using manufacturer recommendations.

PCR

cDNA was amplified using Phusion High-Fidelity DNA polymerase following the manufacturer recommended protocol for GC-rich targets. Annealing temperature and primers used were listed in Table S3. Sox17 amplification products were confirmed with Sanger sequencing (see Table S5).

Quantitative PCR

cDNAs were subjected to qPCR using SYBR™ Green PCR Master Mix and a Bio-Rad CFX Real time PCR instrument. The amplification program consisted of 95°C for 10 minutes followed by 45 cycles of 95°C for 15 sec and 60°C for 1 minute (see Table S4 for primer sequences). Amplicons spanned at least one exon junction when possible and appeared as a single band at expected size when analyzed with agarose gel electrophoresis. Sox17 amplification products were confirmed with Sanger sequencing (see Table S5). Relative expression level was calculated using double delta Ct method with Gapdh as a housekeeping gene and wildtype (WT) embryonic samples or cells at day 0 (D0) of differentiation as controls for normalization.

Immuno-fluorescence staining

E9.5 embryos were dissected in PBS then fixed in 2% paraformaldehyde for at least 1 h at room temperature (RT) before incubation in 30% sucrose in 1X PBS solution at 4°C overnight on a shaker. Intact embryos were then embedded in Tissue-Tek® OCT compound on dry ice. Frozen embryos were cut into 8 μ m thick

sections using a Leica CM3050S cryostat, post-fixed with a cold acetone and air-dried for 30 minutes followed by permeabilization with 0.3% Triton X-100 in PBS for 10 minutes at RT. A 1 hour blocking step at RT was performed with 3% Bovine Serum Albumin (BSA) in PBS before primary antibodies in the same solution was applied to samples for overnight incubation at 4°C (see [Key resources table](#) for an antibody list and a dilution factor). Samples were washed 3 times (10 minutes each) with PBST (0.2% Tween in PBS) before being incubated with secondary antibodies in 1% BSA/PBS at RT for 1 hour. After 4 washes with PBST and 1 wash with PBS (10 minutes each), sections were incubated with Vector TrueVIEW autofluorescence quenching solution for 10 minutes at RT. Nuclei staining with DAPI and mounting with VECTASHIELD Vibrance Antifade Mounting Medium was performed as recommended by Vector TrueVIEW autofluorescence quenching kit. Images were acquired using a Zeiss LSM710 confocal microscope. Display settings were adjusted based on negative staining controls (no primary or secondary antibodies) and applied similarly to every image using Zen 2 lite or Fiji software.

Light-sheet microscopy

Dissected E9.5 embryos that were fixed with 2% PFA for 1 hour at RT were washed with 1X PBS and cleared overnight at RT in refractive index matching solution (RIMS). RIMS was prepared by dissolving 40g of Histodenz in 30 mL of 0.02 M phosphate buffer, followed by adding of 0.01% sodium azide, 0.1% Tween20 and 1 g DABCO (1,4-diazabicyclo[2.2.2]octane) and adjusting to pH7.5 with NaOH. Cleared embryos were mounted with 1.2–1.5% low-melt agarose in RIMS using either capillary tubes or syringe depending on embryo size. Samples were imaged in RIMS solution using Z1 light-sheet microscopy with 5× objective. 3D images and surface renderings were generated using Imaris software.

Luciferase reporter assay

CR1 and CR2 were amplified from wild-type mouse genomic DNA using primers and conditions listed in [Table S3](#) with Phusion High-Fidelity DNA polymerase for a GC-rich target. Purified PCR products were cloned in random direction into the pGL4.14 *luciferase* vector using blunt end ligation into the EcoRV restriction site. Mutant CR2 oligonucleotides were annealed then cloned into pGL4.14 vector by replacing wild-type CR2 sequence using NheI and XmaI restriction sites (oligonucleotide sequences are in [Table S6](#)). All constructs were confirmed with Sanger sequencing. Test vectors, pGL2 *Firefly* positive control vector and pRL-SV40 *Renilla* control vector were transfected to cell culture at day 6 of differentiation using standard protocol of Xfect Transfection Reagent kit. Transfection medium was replaced after 24 hours with fresh differentiation media. Samples were collected a day later with passive lysis protocol of Dual-Luciferase Reporter Assay kit and luciferase activity was measured and GloMax Discover Microplate Reader.

Fluorescence-activated cell sorting

Dissected E9.5 mouse embryos were dissociated at 37°C for 10 minutes with Accumax solution containing 2U/mL DNase I. On ice, cells were filtered through 35 μm strainer cap of Falcon tube, counted, pelleted by centrifugation at 1,000 rpm for 3 minutes at 4°C and resuspended in flow cytometry staining buffer containing 2 U/mL DNase I (Flow/DNaseI buffer). Samples were Fc-blocked with mouse IgG for 15 minutes, followed by incubation with fluorophore conjugated-antibodies in blocking solution for 30 minutes at RT. Stained cells were washed once with flow cytometry staining buffer and stained with DAPI in Flow/DNaseI buffer 15 minutes at room temperature, before being sorted into homogenization buffer (from Maxwell® 16 LEV simplyRNA Purification Kits) if RNA extraction is needed at the VUMC Flow Cytometry Shared Resource Core using a 4-laser FACARIA III.

Western blot analysis

Dissected E9.5 embryos were lysed in RIPA buffer supplemented with 1X protease inhibitor cocktail/mammalian and phosphatase inhibitor cocktails for 1 hour at 4°C with periodic mixing. Samples were centrifuged at 12,000 rpm for 20 minutes at 4°C to collect only supernatant. Protein concentration of samples were determined using Pierce BCA protein assay according to manufacture protocol. 10μg of protein of each sample were diluted in Laemmli buffer with β-mercaptoethanol, denatured at 95°C for 10 minutes and loaded on 10% Mini-PROTEAN TGX precast gel. Electrophoresis was run in MOPS running buffer at 100 V for about 40 minutes, then protein was transferred to methanol activated PVDF membrane overnight at 4°C at 0.3 mA using Mini-PROTEAN tetra vertical electrophoresis system. Membranes were inspected with Ponceau S staining, blocked for 1 hour at RT with blocking buffer (5% non-fat milk in 1X Tris buffer saline (20 mM, pH 7.4) containing 1mL/L Tween 20 (TBS-T). Primary antibodies against SOX17 or GAPDH were

diluted in the blocking buffer and applied to membrane overnight at 4°C. Next day, membrane was washed 3 times with TBS-T and secondary antibodies diluted in blocking buffer were applied to membranes at RT for 1 hour. Three washes were carried out using TBS-T to remove excess secondary antibody. Protein bands were visualized using SuperSignal West Dura Extended Duration Substrate and imaged with the digital ChemiDoc MP imaging system.

Chromatin conformation capture assay

A previously described protocol was used (Naumova et al., 2012). In brief, collected cells were fixed with 1% formaldehyde for 15 minutes at RT, quenched with 1X glycine for 5 minutes, and washes thoroughly with ice-cold 1X PBS. Fixed cells were then incubated in lysis buffer containing 5 mM MgCl₂, 10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA-630 (NP40) and 1X protease inhibitor for 2 hours on ice. Lysed cells were next incubated with digestion enzyme buffer (NEB 3.1) containing 0.3% SDS for 1 hour at 37°C, followed by an addition of 1.8% Triton X-100 to the same buffer for another one-hour incubation at 37°C. 500 U of BglII restriction enzyme were added to every 10⁷ cells and samples were incubated overnight at 37°C on rocking platform. 250 U of BglII restriction enzyme was added to each sample the next morning with an incubation of two more hours at 37°C to increase the digestion efficiency. 1.6% SDS was added to each sample, followed by an incubation of 25 minutes at 65°C to inactivate the enzyme. Every 10⁷-cell sample were then mixed with 8 mL of pre-chilled 1.1X ligation cocktail master mix including 1.1% Triton X-100, 1.1X NEB ligation buffer, 0.11 mg/ml BSA and 1.1 mM ATP. After one hour of incubation at 37°C, 800 U of T4 DNA ligase was added and samples were incubated at 16°C for four hours on a rocking platform. To terminate ligation step, proteinase K was added to the final concentration of 100 µg/mL for an overnight incubation at 55°C. DNA purification involves two rounds of twice phenol-chloroform-isoamyl alcohol extraction followed by once ethanol precipitation of DNA. After two times of precipitation, DNA pellet was resuspended in 500 µl of 1X TE buffer pH 8.0 and subjected to Amicon ultra-0.5 30k purification device as manufacturer recommendations. Final purified product was incubated with 1 µg of RNase A at 37°C for 15 minutes. Small aliquots of samples were saved before and after digestion step to extract DNA and used for qPCR assessment of digestion efficiency across digested sites. Only samples that achieved at least 60–70% of digestion efficiency were advanced for further steps. A control library to correct for primer pairs with different amplification efficiencies was constructed following similar digestion, ligation and DNA purification steps described earlier using bacterial artificial chromosomes (BAC) that contains interested genomic region. To determine the interaction frequencies between fragments in invested samples, qPCR was performed using templates from both a dilution series of control library and 3C ligation product library from collected samples. C_t results from studied samples were normalized against the titration curved built from C_t results of the control library for each primer pair, and then against a control genomic region, thus the interaction frequency between each pair of fragments could be compared against each other.

Graphical illustration, quantification and statistical analysis

Wherever applied, performed statistical test, excluded samples, exact n number and what n represents were indicated in the figure legend. Quantification of western blot signals was performed using ImageJ (FIJI). Processed ChIP-seq datasets were visualized using Integrative Genomics Viewer (IGV) software. Graphical illustrations in Figures 5, 6, 7 and S3 were created with BioRender.com. All other quantification visualization and statistics were carried out with GraphPad Prism 9 software. Results were shown as individual dots representing replications (details on replication types can be found in figure legends) and bars representing mean ± SEM. p values are represented as asterisk symbols: ns = non-significant, * p ≤ 0.05, ** p ≤ 0.01, *** p ≤ 0.001, **** p ≤ 0.0001.