Sox17 and β-catenin co-occupy Wnt-responsive enhancers to govern the endodermal gene regulatory network

Shreyasi Mukherjee[1,2,#], Praneet Chaturvedi[1,2,#], Scott A. Rankin[1,2,#], Margaret B. Fish[3] , Marcin Wizla[1, 6] , Kitt D. Paraiso[3,4], Melissa MacDonald[1,2], Xiaoting Chen[5], Matthew T. Weirauch[5,2], Ira L. Blitz[3], Ken W. Y. Cho[3] and Aaron M. Zorn[1,2*]

[1]Center for Stem Cell and Organoid Medicine (CuSTOM), Division of Developmental Biology, Perinatal Institute, Cincinnati Children's Hospital Medical Center,

[2]University of Cincinnati, College of Medicine, Department of Pediatrics, Cincinnati, OH, USA

[3]Department of Developmental and Cell Biology, University of California, Irvine, CA, USA

[4]Center for Complex Biological Systems, University of California, Irvine, CA, USA

[5] Center for Autoimmune Genomics and Etiology (CAGE), Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

[6]Present address: The National Xenopus Resource, Marine Biological Laboratory, Woods Hole, MA, USA

[#] Equal contribution

*Corespondence: aaron.zorn@cchmc.org

Key words: Sox, Wnt, Beta-catenin, endoderm, enhancers, GRNs, TCF, transcription

Key findings

- Sox17 regulates germ layer specification by promoting endoderm differentiation while simultaneously repressing mesectoderm fates.
- Functional interactions between Sox17 and Wnt-signaling is a major feature of the GRN controlling endoderm specification and patterning *in vivo*.
- Sox17 and β-catenin co-occupy a subset of enhancers to synergistically stimulate transcription in the absence of Tcfs.
- Sox17 regulates the spatial expression domains of Wnt-responsive transcription in the gastrula.

ABSTRACT

Lineage specification depends on the integration of lineage-determining transcription factors (TFs) and signaling effectors on enhancers to establish gene regulatory networks (GRNs). Sox17 is a key regulator of definitive endoderm development, and yet, its genomic targets remain largely uncharacterized. Here, using genomic approaches and epistasis experiments, we define the Sox17-governed endoderm GRN in *Xenopus* gastrulae. We show that Sox17 functionally interacts with the canonical Wnt pathway to specify and pattern the endoderm while repressing alternative mesectoderm fates. Together, Sox17 and β-catenin bind hundreds of key enhancers. In some cases, Sox17 and β-catenin synergistically activate transcription in the absence of Tcfs, whereas on other enhancers, Sox17 represses ß-catenin/Tcf-mediated transcription to spatially restrict gene expression domains. Our findings establish Sox17 as a tissue-specific modifier of Wnt responses and point to a novel paradigm where genomic specificity of Wnt/ß-catenin transcription is determined through functional interactions between lineage-specific Sox TFs and ß-catenin/Tcf transcriptional complexes. Given the ubiquitous nature of Sox TFs and Wnt signaling, this could be a general feature of Sox proteins across numerous contexts of development and disease.

INTRODUCTION

During embryogenesis, the pluripotent zygote progressively gives rise to specialized cell types expressing distinct sets of genes that, in turn, define the cell's identity and encode proteins necessary for its function. Lineage specific gene expression is controlled by the genomic integration of signaling pathways and transcription factors (TFs) on DNA cis-regulatory modules (CRMs), such as enhancers, that control transcription (Heinz et al., 2015; Stevens et al., 2017). An important goal of developmental biology is to elucidate how signaling effectors and TFs interact on distinct sets of CRMs within the chromatin landscape to form gene regulatory networks (GRNs) that activate lineage-specific transcriptional programs whilst repressing expression of alternative fates. This will provide a deeper understanding of how transcriptional cell states are established, dysregulated in disease, and how GRNs might be manipulated for therapeutic purposes.

We have addressed this in the context of endoderm germ layer specification: one of the earliest cell fate decisions in vertebrate development that provides a relatively simple model to elucidate how GRNs control lineage-specific transcriptional programs (Charney et al., 2017b). In embryos and pluripotent stem cells (PSCs), Nodal/Smad2 and Wnt/ß-catenin (Ctnnb1; hereafter Bcat) signaling cooperate to initiate the mesendoderm program and subsequent development of definitive endoderm (DE) progenitors, which gives rise to the epithelia of the digestive and respiratory systems (Sumi et al., 2008; Zorn and Wells, 2009). Downstream of Nodal and Wnt, a core set of endoderm TFs: Sox17, Gata4-6, Eomes, Foxa1/2 and Mix family of homeodomain TFs execute the endoderm GRN (Arnold et al., 2008; Charney et al., 2017b; Engert et al., 2013; Sinner et al., 2006). In *Xenopus* where endoderm specification is well studied, the Nodal and Wnt pathways interact at several levels. First, maternal (m) Wnt/Bcat, active on the dorsal side of the blastula, is required for high levels of *nodal* expression at the onset of zygotic transcription (Hyde and Old, 2000). Then Nodal and mWnt cooperate to promote the expression of endoderm TFs Sox17, Foxa2 and many dorsal mesendoderm organizer genes (Xanthos et al., 2002). A few hours later zygotic (z) *wnt8* (a Nodal target) is expressed on the ventral side of the gastrula where, together with Nodal signaling, it promotes ventral and posterior mesendoderm identities (Charney et al., 2017b; Stevens et al., 2017). In mammals early Wnt/Bcat similarly activates expression of Nodal and core endoderm TFs, with prolonged Wnt promoting posterior hindgut fate (Engert et al., 2013; Zorn and Wells, 2009).

Functional interactions between the core endoderm TFs and the Wnt pathway are critical to: 1) segregate the transient mesendoderm into endoderm and mesoderm, 2) pattern the nascent endoderm into spatially distinct subtypes and 3) execute the downstream differentiation program to give rise to endoderm-derived lineages. But mechanistically, how the Wnt signaling machinery interacts with core endoderm TFs to execute this differentiation cascade is unresolved. The transcriptional targets of the endoderm TFs are largely unknown, and it is unclear exactly how Wnt regulates distinct spatiotemporal transcription programs in the embryo. Indeed, how the canonical Wnt pathway elicits context-specific transcription in its multitude of different biological roles from development, homeostasis and cancer is still poorly understood. According to the dogma of canonical Wnt signaling, Wnt-activated Fzd-Lrp5/6 receptor complexes stimulate the stabilization and translocation of Bcat to the nucleus. There, it interacts with one of four Tcf HMG-box TFs (Tcf7, Tcf7L1, Tcf7L2 and Lef1) at Wnt-responsive CRMs (Cadigan and Waterman, 2012; Schuijers et al., 2014) to assemble a "Wnt-enhanceosome" complex with the scaffold proteins Bcl9, Pygopus and Ldb (Gammons and Bienz, 2018). In the absence of Bcat, Tcfs are coupled with co-repressor proteins Tle/Groucho and Hdac to inhibit transcription. Bcat displaces Tle and recruits a co-activator complex including histone acetyltransferases Ep300 or Cbp to stimulate transcription (Cadigan and Waterman, 2012). How distinct context-specific Wnt target genes are selected is unclear since Tcfs have nearly identical DNA-binding specificities (Badis et al., 2009; Ramakrishnan and Cadigan, 2017) and for the most part they are ubuiqitously expressed. An emerging idea is that the Wnt-enhanceosome also interacts with other lineage-specific TFs (Gammons and Bienz, 2018; Trompouki et al., 2011); yet, how these impact genomic specificity *in vivo* is largely untested and the idea of Tcf-independent Wnt-mediated transcription remains controversial.

Here we investigated the possibility that Sox17 functionally interacts with Wnt/Bcat to regulate transcription in the Xenoups endoderm GRN. In all vertebrate embryos, Sox17 is specifically expressed in the gastrula endoderm where it is required for early gut development (Clements et al., 2003; Hudson et al., 1997; Kanai-Azuma et al., 2002). *SOX17* is also a commonly used marker of definitive endoderm fate in differentiating PSC cultures (Wang et al., 2011). Despite the critical role for Sox17 in endoderm development, only few of its direct transcriptional target genes have been identified (e.g.: *hnf1b*, *foxa1* and *dhh*) (Ahmed et al., 2004; Sinner et al., 2004; Yagi et al., 2008). In Xenopus, ectopic Sox17 is sufficient to initiate endoderm development in pluripotent blastula animal cap cells (Clements et al., 2003; Hudson et al., 1997), and co-injection of stabilized Bcat can enhance this activity (Sinner et al., 2004). Sox17 can physically interact with Bcat *in vitro* and suppress the transcriptional activity of generic Tcf/Bcat-

multimer driven reporter constructs (pTOPflash) in tissue culture experiments (Sinner et al., 2007; Zorn et al., 1999). However, the biological relevance of these interactions and whether Sox17 and Bcat functionally interact on chromatin to regulate the endoderm GRN remains unknown.

Here, we defined the genomic targets of Sox17 in the Xenopus gastrula. In addition to promoting expression of endoderm genes, Sox17 also represses ectoderm and mesoderm gene transcription, and acts in a negative feedback loop to restrain Nodal signaling. We demonstrate that functional interaction with canonical Wnt signaling is a key feature of the Sox17-regulated GRN. Over a third of all Bcat and Sox17 genomic binding in the gastrula occur at the same loci. In some instances, Sox17 suppressed Bcat-Tcf mediated transcription, while in other cases, Sox17 and Bcat synergistically activated other enhancers apparently independently of Tcfs, indicating a novel mode of regulation These results provide new insight into the GRN controlling endoderm development and have implications for how Sox proteins and Bcat might interact across diverse biological contexts from development to cancer.

Results

Sox17 regulates a genomic program controlling germ layer segregation and endoderm development

To identify the transcriptional program regulated by Sox17, we performed RNA-sequencing (RNA-seq) on control and Sox17-depleted *Xenopus tropicalis* embryos at multiple time points during blastula and gastrula stages (NF9-12) when the endoderm germ layer is being specified. In *Xenopus tropicalis* there are three redundant genes; *sox17a*, *sox17b.1* and *sox17b.2* (collectively *sox17*) with indistinguishable activities and identical expression in presumptive vegetal endoderm cells of gastrula embryos (Fig. 1A and Fig. S1A) (D'Souza et al., 2003; Hellsten et al., 2010). Microinjection of a combination of antisense morpholino oligos (sox17aMO and sox17bMO) targeting all 3 paralogs resulted in a robust knockdown of Sox17 protein as confirmed by immunostaining (Fig. 1C,B and Fig. S1). The Sox17-MO phenotype was consistent with previous reports where about half of the embryos arrested in gastrulation with (Clements et al., 2003; Sinner et al., 2006) and remaining embryos phenocopied mouse mutants with defective gut development (Kanai-Azuma et al., 2002). Injection of mRNA encoding mouse Sox17 rescued both the anatomical and gene expression phenotypes confirming the MOs efficacy and specificity (Fig. S1C,D).

Differential expression analysis of control-MO and Sox17-MO embryos identified 1023 Sox17-regulated genes (> 2-fold change, FDR$\leq$ 5%) (Fig. 1D and Table S1). Gene Ontology (GO) enrichment was consistent with the Sox17-regulated transcriptome being involved in “endoderm

formation", "epithelial differentiation" and "digestive track morphogenesis" (Fig. S1E). In total 493 genes were downregulated in Sox17-depleted embryos and 530 genes were upregulated. The time course data revealed that >75% of the differentially expressed genes were Sox17-dependent during the mid-late gastrula (NF10.5-12), consistent with a role in maintaining the endoderm fate after initial induction by Nodal signaling in the blastula (NF9) (Fig.1D and Fig. S1A). Sox17-regulated genes comprised of 73 TFs, including known targets *foxa2* and *hnf1b* (Sinner et al 2004; Sinner et al 2006), and paracrine signaling components (Table S1) such as the Hedgehog pathway (*dhh*, *hhip*, *gli1*); a key epithelial signal in gut organogenesis. This confirms that Sox17 sits atop of the regulatory hierarchy regulating endoderm differentiation.

We next investigated how Sox17-regulated genes were spatially expressed; to do so, we leveraged previously published RNA-seq data set of different tissues dissected from gastrula embryos (Blitz et al., 2017). As predicted, a majority of Sox17-dependent genes were enriched in the vegetal endoderm and dorsal mesendoderm (organizer) (Fig. 1E) including *osr1, hnf1b, dhh* and *slc5a8,* as confirmed by in situ hybridization (Fig. 1F and Fig. S1D). Interestingly, over 30% of the genes upregulated early (NF9-10) in the Sox17-depleted embryos were normally enriched in the equatorial or animal cap tissue; examples include ectoderm-promoting TFs *tfap2a* and *lhx5* (Houston and Wylie, 2003) (Fig. 1D-F). These data suggest that Sox17 plays an important role in repressing ectoderm and mesoderm fate in vegetal endoderm cells. Unexpectedly Sox17 also negatively regulated genes that are normally enriched in the vegetal endoderm and dorsal mesendoderm. Some of these vegetally enriched genes encoded components of the endoderm-promoting Nodal pathway including *nodal1, nodal6, gdf3, foxh1, mix1* and *bix1.1*, all of which were upregulated in Sox17-depleted embryos (Fig. 1D-F and Fig. S1A,D). Indeed, even *sox17* transcripts were modestly increased in Sox17-MO embryos although previous work has demonstrated that Sox17 directly promotes its own transcription (Howard et al., 2007) after initial induction by Nodal/Smad2. These data indicate that Sox17 acts as a negative feedback inhibitor that restricts excessive endoderm induction by restraining Nodal activity.

The observation that over 10% of Sox17-regulated genes are enriched in the organizer mesendoderm while Sox17 is present throughout the endoderm, suggested that in addition to Nodal, Sox17 functionally interacts with the dorso-ventral patterning Wnt and/or Bmp pathways to control spatial expression. Intriguingly, we found that Sox17 regulates the expression of several Wnt/Bcat pathway components and targets including: *dkk1, dkk2, fzd5, nodal3.1, pygo1, sia1, ror1*, *rsop2* and *wnt11* (Fig. 1D,F and Table S1). This suggests that Sox17 regulates both endoderm specification and patterning while suppressing mesectoderm fate and that to do so, Sox17 participates in feedback loops with the Wnt and Nodal pathways.

Sox17 ChIP-Seq reveals direct endodermal targets that are Nodal and Wnt pathway components

To identify direct Sox17-targets we generated and validated anti-Sox17 antibodies (Fig. S2) and performed ChIP-seq analysis (biological replicates) of early gastrula (NF10.5) embryos, identifying 8436 statistically significant Sox17-bound CRMs (Irreproducible Discovery Rate ($p<0.05$). These were associated with 4801 genes (Fig. 2A). 88% of the Sox17-binding regions were in introns or intergenic regions more than 1kb away from the transcription start site, consistent with the presence of distal CRMs (Fig. S3A). A comparison to publically available p300 ChIP-seq data at the same stage (Hontelez et al., 2015) showed that most of the Sox17-bound genomic loci were also bound by Ep300, (Fig. 2B, E) indicative of active enhancers. *De novo* motif analysis of the ChIP-seq peaks confirmed that Sox17 DNA-binding sites were the most enriched motif as expected (Fig. 2C and Fig. S3B). Other significantly enriched motifs included Tbx and Gata, suggesting that Sox17 may coordinately engage enhancers with other core endoderm GRN TFs.

A comparison to published ChIP-Seq data from human PSC-induced definitive endoderm (Tsankov et al., 2015) revealed that 20% of the *Xenopus* Sox17-bound genes were also SOX17-bound in human DE (Fig S3). GO analysis of the conserved Sox17-bound genes show an enrichment for 'Tgfb receptor activity' and 'Bcat binding' (Fig S3) reinforcing the notion that Sox17 interactions with components of Nodal and Wnt pathways is a conserved feature of the endoderm GRN.

Intersecting the RNA-seq and ChIP-seq data identified 315 genes, corresponding to ~30% of the Sox17-regulated transcriptome, that showed Sox17 chromatin binding (609 peaks) and are therefore likely to be direct transcriptional targets (Fig. 2A). These putative direct targets were significantly enriched (44%, 139/315 of transcripts) for expression in the endoderm or dorsal mesendoderm (Fig. 2D). Of the Sox17-bound and regulated genes, 197 genes (383 peaks) were positively regulated by Sox17 (down in Sox17MOs) consistent with Sox17 directly activating their transcription including *slc5a8, hnf1b*, *foxa1* and *dhh* (Fig. 2E), which were previously suggested to be direct Sox17-targets. 118 putative direct targets were negatively regulated by Sox17 (up in Sox17MOs) including ectodermal genes (*lhx5, foxi2* and *tfap2a)*, endoderm-enriched Nodal pathway genes (*nodal1*, *gdf3*, *gdf6* and *mix1*) and Wnt-regulated organizer genes (*dkk* and *fst)*, suggesting direct transcriptional repression (Fig. 2E). Motif analysis (Fig S3B) revealed that ~45% of peaks associated with Sox17-activated genes were also enriched for LIM-homeodomain TFs, whereas Sox17-repressed genes were enriched for Tbx or Pou TFs, suggesting a likely regulatory

logic where Sox17 has distinct modes of transcriptional activation or repression depending on context-specific cofactor TFs.

These analyses provide new insight into the endoderm GRN and reveal previously unappreciated roles for Sox17 in germ layer segregation and endoderm patterning involving functional interactions with the Nodal and Wnt pathways. These findings, together with our previous work demonstrating that Sox17-Bcat can physically interact *in vitro* (Sinner et al., 2007; Zorn et al., 1999) prompted us to investigate the biological relevance of any such Sox17/Bcat interaction further.

The β-catenin-regulated gastrula transcriptome is enriched for endodermal genes

To test the hypothesis that Sox17 and Bcat functionally interact *in vivo*, we set out to identify Bcat-regulated genes and compare these to the Sox17-regulated transcriptome. Injection of a well-characterized Bcat-MO (Heasman et al., 2000) resulted in depletion of nuclear Bcat, reduced expression of a transgenic Wnt-responsive reporter (Tran et al., 2010) and the expected ventralized phenotype, which was rescued by co-injection of stabilized human Bcat mRNA (Fig. S4), thus confirming the efficacy and specificity of the Bcat knockdown. RNA-seq of control-MO and Bcat-MO depleted embryos at eight time points from blastula and gastrula stages (NF7-12) identified a total of 2568 Bcat-regulated genes (> 2-fold change, FDR $\leq$ 5%), 1321 of which were downregulated and 1247 upregulated in the Bcat-MO embryos (Fig. 3A and Table S2). Remarkably the Bcat-dependent genes encoded 251 TFs, (~15% of all the TFs in the genome), reinforcing the notion that Wnt/Bcat is involved in multiple transcriptional cascades in the early embryo.

Intersecting the Bcat-MO RNA-seq data with previously published Bcat ChIP-seq data from *Xenopus tropicalis* gastrula (Nakamura et al., 2016) identified 898 putative direct targets genes associated with 2616 Bcat chromatin-binding peaks. 546 genes were downregulated in the Bcat-MO and 352 had increased expression in Bcat-MO embryos (Fig. 2B). These 898 genes included almost all of the previously known Bcat targets in early *Xenopus* embryos (e.g, *nodal1-6*) (Blythe et al., 2010) and had extensive overlap with other recent genomic analysis of *Xenopus* gastrula Wnt targets (Ding et al., 2017; Kjolby and Harland, 2017; Nakamura et al., 2016) (Fig. S5B).

As expected, many of the positively regulated Bcat targets had enriched expression in the dorsal mesoderm (Fig. 3D) including most known organizer genes such as; *cer1, chrd, dkk1, frzb, fst and gsc*. In contrast ~30% of Bcat-bound genes with increased expression in the early Bcat-MO embryos (NF7-9) were enriched for ectoderm specific transcripts, consistent with the

expansion of ectoderm and loss of neural tissue in ventralized embryos (Fig. 3D). We also identify many direct Bcat targets that are likely to be regulated by zygotic Wnt8, (Kjolby and Harland, 2017; Nakamura et al., 2016) including *sp5*, axin2, *cdx1/2, fzd10, msgrn1* and eight *hox* genes, all of which were downregulated in Bcat-MOs (Fig. 3A, E). A number of zygotic Wnt-targets with enriched expression in the ventral mesoderm were components of BMP pathway including; *bambi*, *bmp7*, *id2*, *msx1, smad7, szl*, *ventx1,* and *ventx3* (collectively known as the BMP-synexpression group) (von Bubnoff et al., 2005); most of these genes are known to be directly activated by both Bmp4/Smad1 and zWnt8/Bcat (Itasaki and Hoppler, 2010; Stevens et al., 2017). Consistent with this , these were upregulated in Bcat-depleted embryos, likely due to the increase in BMP/zWnt8 signaling in ventralized embryos.

Most importantly almost 200 Bcat-bound and regulated genes had endoderm enriched expression (Fig. 3D and Table S2). *De novo* motif analysis of showed that Bcat bound peaks were significantly more enriched for Sox than canonical Tcf motifs (Fig. 3C; Fig. S5?), consistent with the hypothesis that Sox17 and Bcat co-regulate endoderm gene transcription *in vivo*. Contrary to the prevailing view that Bcat promotes endoderm fate primarily by activating Nodal ligand expression, these results also suggest a direct input by Wnt/Bcat into the endoderm GRN.

Sox17 and β-catenin co-occupy CRMs to regulate the endoderm GRN

Comparison of the Sox17-MO and Bcat-MO RNA-seq datasets identified 415 transcripts that were regulated by both Sox17 and Bcat, representing ~40% of the Sox17-regulated and ~15% of the Bcat-regulated transcriptome (Fig. 3A). Intersection with their ChIP-seq datasets identified 3956 genomic loci bound by both Sox17 and Bcat, suggesting a direct co-regulation of at least a subset of these enhancers(Fig. 4A). Peak density analysis using published ChIP-seq data (Charney et al., 2017a; Hontelez et al., 2015; Paraiso et al., 2019) indicated that the cobound loci had epigenetic signatures of active enhancers (Fig. 4B). Homer motif analysis showed that these were much more enriched for Sox DNA-binding sites than Tcf sites (Fig S5A).

In total, 84 genes (191 Peaks) were bound and regulated by both Sox17 and Bcat, representing ~30% of direct Sox17 targets and ~10% of direct Bcat targets (Fig. 4A). This is likely to be an underestimation of the number of direct coregulated genes, as these genes had to pass a stringent thresholding comprising of four independent statistical tests to be included in this list. A detailed motif analysis of the 191 cobound and coregulated peaks using the comprehensive CIS-BP database (Weirauch et al., 2014) revealed that 78% (149/191) were predicted to contain both Sox17 and Tcf DNA-binding motifs, often > 5 sites in each peak (e.g., *dkk1*, *lhx5* and *osr1*) (Fig. 4D). In contrast ~21% (40/191) of the peaks had Sox motifs but no obvious Tcf consensus

sites (e.g., *six1),* suggesting that Bcat might be recruited to these loci independent of Tcfs. Only 1/191 peaks had Tcf but no Sox motifs. In most cases individual genes were associated with multiple cobound peaks suggesting that combinatorial binding and integration of several CRMs is crucial to control transcription.

Since the ChIP-seq experiments were performed on whole embryos, it was possible that Sox17 and Bcat chromatin binding might occur in different cell populations. To test whether Sox17 and Bcat bound the same enhancers in the same cells we performed ChIP-reChIP experiments. First anti-Bcat ChIP was performed on 500 gastrulae. DNA eluted from the first ChIP was then re-precipitated with either anti-Sox17 or a negative control IgG antibody. All 4 loci tested by ChIP-reChIP QPCR (*six1*, *dkk1*, *lhx5* and *gli1*) showed enrichment compared to IgG demonstrating that these genomic loci (~150 bp in length) were simultaneously bound by Sox17 and Bcat in endodermal cells. In contrast a negative control locus *gtse1* which is bound by Bcat but not Sox17 was not recovered in the ChIP-reChIP experiment (Fig. 4F).

The 84 cobound and coregulated genes fell into 4 regulatory categories (Fig. 4E); 32/84 were activated by both Sox17 and Bcat (e.g., *six1*, *slc5a8* and *osr1*), 27/84 were activated by Bcat but repressed by Sox17 (e.g., *dkk1*, *lhx5* and *nodal1*), 19/84 were positively regulated by Sox17 and negatively regulated by Bcat (up in the Bcat MO) and finally 6/84 appeared to be negatively regulated by both Sox17 and Bcat. The observation that Bcat positively regulated most of these genes (59/84, 70%) was consistent with its known role in transcriptional activation. Sox17 positively regulated 60% of targets and repressed 40% of targets, suggesting context-dependent activity. For further analysis we focused on the two major regulatory groups; genes/enhancers activated by both Sox17-Bcat and activated by Bcat but repressed by Sox17.

Sox17 and Bcat epistasis suggests a functional interaction between these two factors

It was possible that the overlap in Sox17- and Bcat-regulated genes was due entirely by Sox17 and Wnt interacting in a regulatory feedback loop. Bcat might be required for the robust expression of *nodal* and *sox17* and Sox17 protein, in turn, regulates the expression of Wnt-pathway components. While this is likely to partially account for Bcat/Sox interactions, we directy tested the functional necessity for both Sox17 and Bcat in epistasis experiments, asking whether injection of Sox17 RNA could rescue candidate coregulated genes in Bcat depleted embryos and vice versa. We found that co-injection of mSox17 mRNA could not rescue the normal expression patterns of *six1*, *osr1*, *lhx5* or *dkk1* in Bcat-MO embryos, but it did rescue *foxa1* expression (Fig. 5). This indicates that a reduction of Sox17 downstream of Bcat depletion cannot account for the disrupted expression of *six1*, *osr1*, *lhx5* or *dkk1*. Simillarly co-injection of stabilized Bcat mRNA

rescued the normal expression of *foxa1* but not *six1*, *osr1*, *lhx5* or *dkk1* in Sox17-morphants indicating that the impact on gene expression was not just due to disruption of Wnt signaling downstream of Sox17 depletion (Fig. 5). These data support the notion that Sox17 and Bcat functionally interact to regulate transcription of these key genes by directly engaging their enhancers.

Sox17 and Bcat synergistically regulate transcription of endodermal enhancers.

To examine how Sox17 and Bcat directly regulate endodermal transcription, we selected two exemplar cobound enhancers: a) a proximal *six1* enhancer located at -1kb from the transcription start site (TSS) and b) a *dkk1* enhancer located -1.6kb from its TSS (Fig. 6B). *six1* and and *dkk1* are both expressed in the dorsal mesendoderm of the organizer (Fig. 5), but have opposite regulation. Sox17-Bcat promotes *six1*, whilst Sox17 represses Bcat activation of *dkk1.*

The -1kb *six1* enhancer is evolutionarily conserved from Xenopus to human containing three Sox-binding sites but no predicted Tcf motifs (Fig. 6B). In mouse, this enhancer can drive transgenic expression in the embryonic gut tube (although the TFs that modulate it were previously unknown) (Sato et al., 2012). Analysis of published ChIP-seq data indicate that, like in Xenopus, the human *SIX1* CRM is also bound by SOX17 in human PSC-derived endoderm (Fig. S6B) (Tsankov et al., 2015).

Mammalian *Dkk1* is known to be a direct Bcat-Tcf target in human tissue culture (Niida et al., 2004) but the CRMs that regulate its expression in the gastrula are have not been characterized. Unlike *six1*, the -1.6kb *dkk1* enhancer harbors both Tcf and Sox DNA-binding sites (Fig. 6B, file S1)

We cloned these CRMs into luciferase (luc) reporter constructs and tested whether they require Sox17 and Bcat for their enhancer activity in *Xenopus* embryos. Targeted microinjection into different cells of 32-cell-stage *Xenopus* embryos (Fig 6A) showed that the *-1six1:luc* or *-1.6dkk1:luc* reporters recapitulated the endogenous spatial expression. They were both active in the dorsal-anterior endoderm, but not the ventral endoderm or ectoderm, indicating that they are bona fide enhancers (Fig.6C, E). Depletion of endogenous Sox17 or Bcat abolished expression of the *-1six1*:luc (Fig.6C), and this was rescued by adding back mSox17 to Sox17-MO or by adding back constitutively active (ca)Bcat mRNA to Bcat-MO embryos. However, co-injection of mSox17 could not rescue expression in Bcat-MO embryos and nor could ca-Bcat RNA rescue expression in Sox17-MOs (Fig. 5). These epistasis experiments indicate that both Sox17 and Bcat are required for activation of the -1kb *six1* enhancer. The human *SIX1*:luc construct exhibited identical activity to the Xenopus six1:luc enhancer suggesting that the coordinated SOX17-BCAT

regulation is evolutionarily conserved. In contrast, in the case of dkk1:luc construct, depletion of Bcat abrogated reporter activity in dorsal endoderm, whereas Sox17 depletion resulted in significantly elevated reporter activity, indicating that Sox17 restricts Bcat-mediated *dkk1* activity.

To test whether Sox17 and Bcat were sufficient to regulate the *-1six1* and *-1.6dkk1* enhancers, we assayed their activity in ectoderm (animal cap) tissue which does not express endogenous Sox17 and has low Wnt/Bcat activity. Injection of mSox17 but not ca-Bcat mRNA was sufficient to stimulate the *-1kbsix1:luc* reporter in the ectoderm, but co-injection of Bcat along with mSox17 mRNA was required to robustly and synergistically activate the reporter. On the other hand, Bcat but not mSox17 mRNA was able to activate the *dkk1:luc* in ectodermal cells, and co-injection of mSox17 suppressed the ability of Bcat to stimulate the reporter, consistent with the endogenous *dkk1* regulation. The ability of Sox17 to activate the *-1six1:luc* or suppress Bcat-induced *-1.6dkk1:luc* activation required the DNA-binding function of Sox17. A M76A point mutation in mSox17, known to disrupt the DNA-binding HMG domain (Sinner et al., 2007) abolished Sox17's activity in both cases(Fig. 6C). The *six1* enhancer is predicted to have only Sox sites with no evidence of Tcf binding sites. To formally test whether or not it is regulated by Tcfs, we injected a Tcf3-VP16 fusion construct which constitutively activates Tcf-target gene transcription (Darken and Wilson, 2001). As predicted, this had little to no impact on the *six1:luc* reporter. From these data, we conclude that *six1* expression is regulated by a mechanism wherein Sox17 and Bcat synergize to promote activation, likely in a Tcf-independent manner. Additionally, we mutated all three predicted Sox17 motifs in the *six1* enhancer, and this abolished all reporter activity, further confirming that Sox17 is necessary to regulate this enhancer (Fig S6C).

In contrast, the *dkk1* enhancer is predicted to have both Sox17 and Tcf DNA-binding sites. To formally test whether or not this enhancer is regulated by Tcfs, we mutated two of the predicted Tcf binding sites which abolished reporter activity, consistent with the idea of Tcf/Bcat mediated transcription (Fig 6F). Co-injection of mSox17 significantly downregulated Bcat-dependent activation at this enhancer. Consistent with this, mutating two predicted Sox17 binding sites, as well as coinjection with Sox17MO led to increased reporter activity (Fig 6F).

These experiments demonstrate multiple modes through which Sox17 and Bcat coordinately regulate the endoderm GRN. At a subset of enhancers lacking Tcf sites, Sox17 and Bcat are sufficient to synergistically activate transcription. At other enhancers containing both Sox and Tcf sites however, Sox17 repressed Bcat/Tcf mediated transcriptional activation. Taken together, our experiments provide insight into the role of Sox17 in specifying endodermal fate and suggest a novel mechanism through which Sox17 provides genomic specificity to Wnt/Bcat

mediated transcriptional responses by either a) synergistically regulating enhancers with Bcat or by b) modulating Bcat/Tcf mediated transcription.

DISCUSSION

Overview of findings

Sox17 is a key regulator of vertebrate endoderm development and morphogenesis, yet surprisingly the transcriptional program that it regulates has been poorly characterized until now. Here we defined the Sox17-regulated GRN in Xenopus gastrulae, which includes many conserved genes also defined as genomic targets of SOX17 in PSC-derived human DE. We show that in addition to acting atop a transcriptional hierarchy promoting endoderm differentiation, Sox17 has previously unappreciated roles in endoderm patterning amd germ layer segregation. We demonstrate that functional interaction with canonical Wnt signaling pathway is a major feature of the Sox17-regulated GRN. Over a third of all Bcat and Sox17 genomic binding in the gastrula occur at the same loci, which have epigenetic signatures of active enhancers. Sox17 and Bcat co-occupy these enhancers to regulate endoderm gene transcription and modulate spatial expression in the dorsal organizer mesendoderm. Our data suggests that Sox17 and Bcat synergistically activate a subset of endodermal enhancers in the absence of Tcfs, whereas at others enhancers, Sox17 suppresses Bcat-Tcf mediated transcriptional activation. Together these results provide novel insights into the establishment of the endoderm GRN and suggest a new paradigm where Sox17 acts as a tissue specific modifier of Wnt-signaling responses.

The Sox17 regulated endodermal GRN

The Sox17-regulated transcriptome includes over 40 other TFs and signaling factors. In addition to promoting an endodermal fate, Sox17 has a key role in repressing alternative lineages in the vegetal cells. This finding indicates a novel role for Sox17 as a dual-function TF, a feature shared by endodermal maternal TFs during zygotic genome activation like Foxh1 (Chiu et al, 2014), Vegt and Otx1 (Reid et al., 2016, Paraiso et al., 2019). For example, Sox17 appears to directly repress expression of the ectoderm specifying TFs *lhx5* and *tfap2a* (Houston and Wylie, 2003), and the mesoderm TFs *foxf1*, *tbx20* and *hlx*. Sox17 also regulates expression of other core endoderm GRN components (Arnold et al., 2008; Charney et al., 2017b; Zorn and Wells, 2009) promoting expression of *foxa1*, and *gata4*, whilst negatively regulating the expression of some *nodal* and *mix*/*bix* genes. Considering that Nodal, Gata4-6 and Mix can promote *sox17* expression (Afouda et al., 2005; Zorn and Wells, 2009), this suggests a series of positive and negative feedback loops where on one hand, Sox17 promotes endodermal fate, while on the other, it restrains excesive

Nodal signaling. Indeed, there is precedence for Sox factors influencing *nodal* ligand transcription. Maternal Sox3 localized in the blastula ectoderm directly binds the *nodal5/6* enhancers to repress transcription, whereas vegetally localized maternal Sox7 promotes *nodal5/6* expression in the vegetal cells (Zhang and Klymkowsky, 2007). This suggests a potential handoff model where different Sox TFs bind the same enhancers in different cells to activate or repress transcription in a context-dependent manner.

Motif analysis revealed that Sox17 bound loci were also enriched for Tbx and Gata motifs indicating that in addition to regulating each others expression, the core endoderm TFs act together in combinatorial manner, perhaps to establish endodermal super enhancers crucial for downstream lineage specification (Paraiso et al., 2019). Interestingly, enhancers that were negatively regulated by Sox17 were enriched for Tbx and Oct motifs whereas enhancers activated by Sox17 were enriched for homeodomain motifs. This suggests that Sox17 interactions with distinct lineage-specific TFs could elicit a context-specific transcriptional response. In the future, it will be important to identify and integrate the genomic targets of all the core endoderm-specifying TFs and examine how combinatorial enhancer binding is integrated into a unified endodermal GRN.

DNA Binding Specificity of SOX TFs

An important consideration is how Sox17 selects specific target enhancers. In DNA-binding experiments, all Sox TFs bind to similar variants of the DNA sequence 5'-(A/T)(A/T)CAA(A/T)3', which is similar but distinct from Tcf DNA-binding sites 5'-T(A/T)(A/T)CAAG 3' (Hou et al., 2017; She and Yang, 2015). Our preliminary analysis of large-scale TF protein-binding microarray data (Badis et al., 2009) suggest that Sox and Tcf might compete for a subset of divergent non-consensus sites, some of which are found in our ChIP-Seq peaks. Future experiments will test whether Sox17 and Tcfs compete for endoderm enhancers in vivo and determine how this would impact Wnt/Bcat responsive transcription.

Typically SOX TFs can act either as monomers or heterodimers with different cell-specific cofactors, which can impact their DNA-binding affinity and ability to activate or repress transcription (Bernard and Harley, 2010; Hou et al., 2017).The composite heterodimer DNA-binding site provide more sequence specificity. For example, in PSCs, Oct4-Sox2 heterodimes promote pluripotentcy and bind to different DNA sites than Oct4-Sox17 complexes which initiate mesendoderm development (Aksoy et al., 2013). It is therefore, conceivable, that Sox17 and Bcat cooperate with other TFs to provide cell-type specificity to Wnt-responsive transcription during germ layer segregation.

Sox-Bcat-Tcf Interactions

Many SOX proteins, including Sox17, can bind to recombinant Bcat and Tcfs *in vitro* and modulate Wnt/Bcat-simulated transcription of synthetic reporters assays with multimerized Tcf-sites (pTOPFlash) (Sinner et al., 2007; Zorn et al., 1999). While a number of models have been proposed to explain how Sox factors might impact Bcat activity, (Kormish et al., 2010; She and Yang, 2015; Tan et al., 2019), the biological relevance of these interactions was unclear. Although the resolution of ChIP-reChIP assays only allow us to demonstrate that Sox17 and Bcat bind chromatin within 150bp of each other, taken together with *in vitro* data, our results suggest that Sox-Bcat or Sox-Bcat-Tcf transcriptional complexes can be assembled on enhancers in vivo.

At enhancers like *six1* where Sox17 and Bcat synergistically stimulate transcription, we postulate that Sox17 recruits Bcat-Ep300/Cbp to chromatin, analogous to Tcf-Bcat in canonical Wnt-signaling (Gammons and Bienz, 2018). Consistent with this, structure-function studies have found that a conserved 10 amino-acid domain (VDXXEFEQYK) in the Sox17 C-terminus is required for both Bcat-binding and the transcriptional transactivation (Sinner et al., 2004). On the other hand, most Bcat and Sox17 cobound enhancers contain multiple Sox and Tcf sites clustered with 50 bp, suggesting that Sox and Tcf, together with Bcat and scaffolding proteins like Bcl9, (Gammons and Bienz, 2018; Masuda and Ishitani, 2017) aid in the assembly of higher-order complexes. Both Soxs and Tcfs are HMG-domain TFs, capable of binding to the minor groove of DNA and induce DNA bending. This could indicate that they are brought into close spatial proximity to establish the assembly of functional enhanceosomes, despite the distance between their binding sites on linear DNA (Bernard and Harley, 2010; Hou et al., 2017).

The possibility of Sox-Bcat-Tcf ternary complexes raise the question of competitive vs. or cooperative DNA binding, that in turn could influence transcriptional repression versus activation. For example, Sox17 might stabilize the formation of Bcat-Tcf-Ep300 complexes or destabilize the binding of co-repressors such as Tle and Hdac to promote transcription (e.g., *osr1*). To negatively regulate Bcat-Tcf activity (e.g.,*dkk1*) Sox17 might destabilize Bcat-Tcf-Ep300-DNA or stablize Tcf-Tle complexes. This idea is supported by work in mouse cortical neuron progenitors where Sox2 binds within 4-10bp of Tcf/Lef sites and recruits Tle/Gro corepressors to repress Tcf/Bcat-mediated *Ccnd1* expression. (Hagey and Muhr, 2014).

Another possibility is that Sox17 might also impact Wnt transcriptional responses by recruiting cofactors that post-translationally modify Bcat-complexes or alter the local chromatin landscape. Post-translational modifications of both nuclear Bcat or Tcfs on different residues have been shown to impact DNA-binding, stability and their ability to interact with components of the

basal transcriptional machinery (Gao et al., 2014; Masuda and Ishitani, 2017; Ramakrishnan and Cadigan, 2017; Valenta et al., 2012). For example, acetylation of Bcat at lysine 49 by Cbp (K49Ac) promotes lineage-specific transcriptional activation, whereas tri-methylation of the same residue by Ezh2 blocks activation of alternate lineage genes (Hoffmeyer et al., 2017). It is unclear how these co-factors are targeted to Bcat at specific genomic loci, and we postulate that Sox17 could play a role in recruiting transcriptional coactivators/repressors to a subset of lineage-specific endoderm enhancers.

Broader Implications

Although a growing number of other Bcat-binding TFs have been reported, (Gammons and Bienz, 2018; Notani et al., 2010; Trompouki et al., 2011), how these impact genomic specificity of Wnt-responsive transcription has previously not been well characterized and the idea of Tcf-independent transcription remains controversial (Doumpas et al., 2019; Schuijers et al., 2014). Our data suggests that Sox17, Bcat and Tcfs interact on enhancers to regulated Wnt-responsive transcription in the gastrula endoderm and in some cases Sox17 and Bcat appear to act independent of Tcfs. Considering that there are over 20 Sox TFs encoded in the vertebrate genome with most cell types expressing at least one, it is possible that SOX factors are widespread and previously unappreciated accessory effectors of Bcat that finetune Wnt-regulated transcription in many biological contexts.

Methods

***Xenopus* embryo manipulations:**

Xenopus laevis and *Xenopus tropicalis* were purchased from Nasco or the National *Xenopus* Resource Center and housed according to established CCHMC IACUC protocols.

Plasmids used to produce linear templates for mRNA synthesis using the mMessage mMachine SP6 transcription kit (Thermo Fisher AM1340) according to manufacturer's instructions were as follows: pCS2+ human B-catenin S37A (Zorn et al 1999; referred to in the text as ca-Bcat or stabilized Bcat; linearized with NotI); pcDNA6-V5-mouse Sox17 and pcDNA6-V5-Sox17 M76A (Sinner et al 2007; both linearized with PmeI); pCS2+ XTCF3-VP16 (Darken and Wilson, 2001; linearized with NotI).

Morpholino oligos were purchased from GeneTools,LLC and were as follows:

Standard control MO: 5'CCTCTTACCTCAGTTACAATTTATA-3';

X.trop ß-*catenin* MO: 5' TTTCAACAGTTTCCAAAGAACCAGG 3';

X.trop sox17a MO: 5'-AGCCACCATCAGGGCTGCTCATGGT -3'

X.trop sox17b1/2 MO: 5'-AGCCACCATCTGGGCTGCTCATGGT -3';

X.laevis ß *-catenin* MO: 5'TTTCAACCGTTTCCAAAGAACCAGG-3';

X.laevis sox17a MO: 5'-ATGATGAGGAGGGTTGGACAGGAGA-3';

X.laevis sox17b1/2 MO: 5'-TGATTCGGAGTGCTGTGGTGATTAG-3';

Total amounts of MO injected in to *X.tropicalis* embryos were as as follows: 2ng Bcat MO, 8ng sox17 MO (4ng each a+b1/2 MO), 8ng standard control MO.

***In-situ* hybridization:** In situ hybridization of *Xenopus tropicalis* embryos was performed as described (Sive et al 2000) with the following minor modifications. Embryos were fixed overnight (12-16 hours) at 4°C in MEMFA (0.1 M MOPS, 2 mM EGTA, 1 mM MgSO4, 3.7% formaldehyde), washed in MEMFA buffer without the formaldehyde 2x10 minutes, dehydrated directly into 100% ethanol, washed 4-6 times in 100% ethanol, and stored at -20°C for at least 24 hours. On day 1, after 7 minute re-hydration steps through a graded ethanol series (95%, 75%, 60%, 40% on ice, 25% in PBS on ice, PBS), embryos were bisected through the dorsal-ventral axis in PBS+0.01% Tween-20 on 2% agarose coated dishes using a fine double-edge razor blade (Electron Microscopy Services 72003-01). Proteinase K (ThermoFisher AM2548) treatment was performed at 1ug/mL for 10 minutes on bisected gastrula embryos. TEA/Acetic Annhydride steps were skipped on day1. The RNAse A step was 0.25ug/mL for 5 minutes on day 2; and finally the anti-DIG-alkaline phosphatase antibody (Sigma 11093274910) was used at a 1:6,000 dilution in MAB buffer (maleic acid buffer, 100 mM Maleic acid, 150 mM NaCl, pH7.5) + 20% heat-inactivated

lamb serum (Gibco 16070096) + 2% blocking reagent (Sigma 11096176001) on day2/3. Anti-sense DIG labeled in situ probes were generated using linearized plasmid full-length cDNA templates with the 10X DIG RNA labeling mix (Sigma 11277073910) according to manufacturer's instructions. Details on plasmids used for probe synthesis are listed in supplementary table S7.

Luciferase Assays: Putative enhancers of interest were cloned into the pGL4.23 *luc2*/miniP vector. (Promega E8411). 5pg of pRL-TK-renilla luciferase vector (Promega E2241) was co-injected with 50pg of the pGL4.23 *luc2*/miniP:enhancer of interest:*firefly* luciferase constructs. Injected *X.laevis* embryos were cultured overnight at 15^0C until stage NF10.5. Pools of 5 embryos were collected in a single microfuge tube and either snap frozen on dry ice, stored indefinitely at -80^0C until lysis, or lysed directly in 100uL of 100mM TRIS-Cl pH7.5. Lysates were centrifuged for 10 minutes at ~13,000 x *g* and then 25uL of the clear supernatant lysate was used separately in Firefly (Biotium #30085-1) and Renilla (Biotium 30082-1) luciferase assays according to the manufacturer's instructions. To determine relative luciferase activity, Firefly levels were divided by Renilla levels for each sample; each black dot in the graphs in Fig.6 and supplemental fig.S6 represents the relative luciferase activity of a pool of 5 injected embryos. Total amounts of mRNAs and MOs injected into *X.laevis* embryos in luciferase assays were as follows: 5pg ca-Bcat, 25pg mSox17, 25pg mSox17 M76A, 10pg Tcf3-VP16, 20ng *Xl sox17a+b1/2* MO(10ngeach), 10ng *Xl bcat* MO.

RNA isolation: Total RNA was extracted from 10 embryos in three independent biological replicates with the Nucleo-spin RNA kit (Machery-Nagel). **RNA-seq:** RNA-seq libraries were generated using Smart-seq2 cDNA synthesis followed by tagmentation, quality-tested using an Agilent Bioanalyzer 2100, quantified using KAPA qPCR and sequenced using Illumina sequencers at the UC Irvine Genomics High Throughput Facility.

Chromatin Immunoprecipitation

ChIP-qPCR experiments were carried out largely as described in (Hontelez et al., 2015) with some modifications. Briefly, NF 10.5 *Xenopus tropicalis* embryos were fixed in 1% formaldehyde in 0.1x MBS at room temperature for 30 minutes. Sonication was carried out in a Diagenode Bioruptor Pico instrument for 30 cycles of 30 seconds ON, 30 seconds OFF. The following antibodies were then added to the chromatin extract: 10 μl anti-Sox17-Cterminal antibody(affinitiy purified by Bethyl labs, FigS2) or 20 μl anti-β-catenin (Life Technologies, 712700) and the lysate with the antibody was incubated overnight with end over end rotation at 4°C. After washes and

elution, DNA was then purified using a MinElute PCR Purification kit and QIAQuick purification columns and eluted in 20μl TE. **ChIP-qPCR:** For qPCR, the ChIP DNA was diluted 25 fold in nuclease-free dH_2O. The input sample was diluted equally and dilution series of 0.1%, 0.01% and 0.001% was prepared to assess DNA recovery after ChIP. qPCR was carried out using a QuantStudio 3 Real-Time PCR machine. Relative fold change was plotted using GraphPad Prism v8. **ChIP-reChIP:** Before the first ChIP, the antibodies were crosslinked to beads. The antibody solution (20 ul anti-β-catenin (Life Technologies, 712700)) was prepared in 5%BSA/PBS and then added to the bead slurry at a 1:1 ratio. The solution was mixed and end-over-end rotated for 2 hours at 4°C. The antibody bound beads were washed with PBS and then with 0.2M sodium borate pH 9. Crosslinking solution was prepared by adding 1ml of 13mg/ml dimethyl pimelimidate to 1ml of 0.2M sodium borate pH9. This solution was mixed well and 10 volumes were added to the antibody bound beads. Crosslinking took place by end-over-end rotation for 30 min at room temperature. The reaction was then quenched by washing the beads with 0.2M ethanolamine pH 8 for an hour at room temperature. Excess unlinked antibody was removed by washed with 1M glycine pH 3 for 10 minutes at room temperature. ChIP assay was then carried out as described above. At the end of the first ChIP, DNA was eluted with elution buffer supplemented with 10mM DTT. The eluate was then diluted in 10 volumes of wash buffer (50mM Tris-HCl pH 8.0, 100mM NaCl, 2mM EDTA, 1% NP-40) supplemeted with 1x Protease Inhibitor Cocktail and 1mM DTT. The 2nd ChIP was then carried out as a normal ChIP as described above. **ChIP-Sequencing:** Large scale ChIP was carried out as previously described (Charney et al., 2017a) and ChIP-seq libraries were generated using Nextflex ChIP-seq kit (Bioo Scientific), analyzed using an Agilent Bioanalyzer 2100, quantified using KAPA qPCR and sequenced using Illumina instruments at the UC Irvine Genomics High Throughput Facility.

Bioinformatic Analysis:

RNA-Seq: Xenopus genome and annotations of version 9.0 were obtained from Xenbase.org (Karimi et al., 2018) **RNA-Seq Analysis:** Raw RNA-Seq reads were generated in duplicates for control and mutant conditions across timepoints ranging from NF stages 7 to 12. Raw reads were quality checked using FastQC and adapters/low quality reads were removed with Trimmomatic (Bolger et al., 2014). Reads were mapped to the *Xenopus tropicalis* genome version 9.0 (Hellsten et al., 2010) with bowtie2, quantified using RSEM (Li and Dewey, 2011) and reported in transcripts per million (TPM). Differentially expressed genes (LogFC >= |1| and FDR <= 0.05) between control-MO and Sox17-MO, control-MO and Bcat-MO embryos were identifed at each stage using RUVSeq (R package) (Risso et al., 2014).

To assess the spatial expression of transcripts we examined the TPM expression values in dissected stage 10.5 embryos from published data in GEO GSE81458 (Blitz et al., 2017). Normalized expression was then used to obtain genes with enriched expression in a specific tissue if it was 1.5x of mean expression across dissections. statistical significance of enriched expression pattern in groups of genes was determined by chi-square test $p<0.05$ performed in R package MASS.

ChIP-Seq Analysis: Raw reads were assessed for quality-control by FastQC and read trimming was performed wherever necessary with Trimmomatic. Reads were aligned to *Xenopus tropicalis* genome version 9.0 using bowtie2 v2.3.4.7 (Langmead and Salzberg, 2012). Duplicate and multi-mapped reads were removed by Picard [http://broadinstitute.github.io/picard/] and samtools (Li et al., 2009). Peaks were called with MACS2 (Zhang et al., 2008) against NF10.5 input DNA with default options. Irreproducible discovery rate (IDR) analysis was carried out to determine highly reproducible peaks between replicates (Landt et al., 2012). For IDR analysis aligned reads with MAPQ $\geq$ 30 were retained using samtools and converted to bam files. Bam files were then converted to tagAlign format and each replicate was shuffled to generate two pseudoreplicates using unix shuf and awk. Each pseudoreplicate was then converted to a tagAlign file. MACS2 was used to call peaks for each pseudoreplicate using a low confidence threshold of $p \leq 0.01$ against the merged input tagAlign file. These peaks were then used as input for IDR at a threshold of 0.01 to generate highly reproducible peaks. Peaks were annotated using HOMER v4.1.0 and default parameters using the function annotatePeaks.pl (Heinz et al., 2010). To visualize the data, aligned bam files generated as above were converted to bigwig format using Bedtools v2.27.1(Quinlan and Hall, 2010) and visualized on IGV v2.4.1.3 (Thorvaldsdottir et al., 2013).

Motif analysis: *De-novo* motif analysis was performed with HOMER using the function findMotifsGenome.pl. For a comprehensive Sox and Tcf motif analysis, we divided our datasets into 3 groups: a) Peaks bound and regulated by Sox17 only, b) Peaks bound and regulated by Bcat Only and c) Peaks that were cobound and coregulated by Sox17 and Bcat. DNA sequences of 150 bp centered on the peaks summits were extracted and motif analysis was carried out as follows: 1. To find which peaks have instances of Sox17 and Tcf motifs, HOMER function annotatePeaks.pl was used with the parameters: -size 150 -m motif sox17.motif tcf7.motif lef1.motif tcf3.motif tcf7l2.motif. 2., To perform an exhaustive motif scan PWMs identified from direct binding experiments for Sox17, Tcf7, Tcf7l1 and Lef1 were downloaded from CIS-BP (Weirauch et al., 2014) . The 150 bp DNA sequences were scanned for the presence of these PWMs and a threshold of $\geq$ 70% was used to identify putative DNA binding sites. Results were

merged to generate BED files using bedtools merge function denoting the chromosome coordinates of each putative binding site with each 150bp sequence.. Bedtools intersect function was then used to extract coordinates into 'Sox only' 'Tcf only' and 'Sox and Tcf overlapping' categories.

ChIP-Seq Heatmaps: From the Sox17 and Bcat peaks called by MACS2, Bedtools intersect function was used to stratify the peaks into 'Sox bound only', 'Bcat bound only' and 'Sox and Bcat overlapping peaks'. For each category of peaks, the peak summit was extracted from the TF summit file generated by MACS2 and 2000 bp were added to both sides of the summit using awk. Bedtools computematrix was used to generate per genome regions. The generated matrix was then used with bedtools plotHeatmap to generate heatmaps and signal density plots.

Gene Ontology: Gene ontology enrichment analysis was performed using the Gene Ontology Consortium Online Resource (http://geneontology.org).

Analysis of public datasets: The following datasets were downloaded from GEO: GSE72657 (Nakamura et al., 2016), GSE85273 (Charney et al., 2017a), GSE56000 (Gupta et al., 2014) and GSE67974 (Hontelez et al., 2015). Wherever necessary, raw data was downloaded from GEO and aligned to *Xenopus* genome version 9.0, and post-alignment processing to generate bigwig files and heatmaps was carried out as describd above.

Data and software availability:

Raw and Processed RNA-seq and ChIP-seq data generated by this study are available at NCBI Gene Expression Omnibus GSE#pending

Acknowledgements

This work was supported by HD073179 to AMZ and KWYC and by NIH P30 DK078392 gene expression core of the Digestive Diseases Research Core Center in Cincinnati. We are grateful to members of the Zorn and Wells labs and the Endoderm Club for helpful discussions. We also wish to thank Xenbase.org (RRID: SCR_003280) and the National Xenopus Resource (RRID:SCR_013731) for critical community resources.

Author Contributions

AMZ, KWYC, ILB, MBF and SAR designed the experiments. SM, SAR, MBF, MW, MM and ILB performed the experiments. SM, PC, KDP, XC and MTW performed the bioinformatic analysis., SM, SAR, KWYC and AMZ wrote the manuscript.

REFERENCES

Afouda, B. A., Ciau-Uitz, A. and Patient, R. (2005). GATA4, 5 and 6 mediate TGFbeta maintenance of endodermal gene expression in Xenopus embryos. *Development* **132**, 763-774.

Ahmed, N., Howard, L. and Woodland, H. R. (2004). Early endodermal expression of the Xenopus Endodermin gene is driven by regulatory sequences containing essential Sox protein-binding elements. *Differentiation* **72**, 171-184.

Aksoy, I., Jauch, R., Chen, J., Dyla, M., Divakar, U., Bogu, G. K., Teo, R., Leng Ng, C. K., Herath, W., Lili, S., et al. (2013). Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *EMBO J* **32**, 938-953.

Arnold, S. J., Hofmann, U. K., Bikoff, E. K. and Robertson, E. J. (2008). Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. *Development* **135**, 501-511.

Badis, G., Berger, M. F., Philippakis, A. A., Talukder, S., Gehrke, A. R., Jaeger, S. A., Chan, E. T., Metzler, G., Vedenko, A., Chen, X., et al. (2009). Diversity and complexity in DNA recognition by transcription factors. *Science* **324**, 1720-1723.

Bernard, P. and Harley, V. R. (2010). Acquisition of SOX transcription factor specificity through protein-protein interaction, modulation of Wnt signalling and post-translational modification. *Int J Biochem Cell Biol* **42**, 400-410.

Blitz, I. L., Paraiso, K. D., Patrushev, I., Chiu, W. T. Y., Cho, K. W. Y. and Gilchrist, M. J. (2017). A catalog of Xenopus tropicalis transcription factors and their regional expression in the early gastrula stage embryo. *Dev Biol* **426**, 409-417.

Blythe, S. A., Cha, S. W., Tadjuidje, E., Heasman, J. and Klein, P. S. (2010). beta-Catenin primes organizer gene expression by recruiting a histone H3 arginine 8 methyltransferase, Prmt2. *Dev Cell* **19**, 220-231.

Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120.

Cadigan, K. M. and Waterman, M. L. (2012). TCF/LEFs and Wnt signaling in the nucleus. *Cold Spring Harb Perspect Biol* **4**.

Charney, R. M., Forouzmand, E., Cho, J. S., Cheung, J., Paraiso, K. D., Yasuoka, Y., Takahashi, S., Taira, M., Blitz, I. L., Xie, X., et al. (2017a). Foxh1 Occupies cis-

Regulatory Modules Prior to Dynamic Transcription Factor Interactions Controlling the Mesendoderm Gene Program. *Dev Cell* **40**, 595-607 e594.

Charney, R. M., Paraiso, K. D., Blitz, I. L. and Cho, K. W. Y. (2017b). A gene regulatory program controlling early Xenopus mesendoderm formation: Network conservation and motifs. *Semin Cell Dev Biol* **66**, 12-24.

Chiu, W.T., R. Charney, Le, I.L. Blitz, M.B. Fish, Y. Li, J. Biesinger, X. Xie, K.W.Cho. (2014) Genome-wide view of TGFβ/Foxh1 regulation of the early mesendoderm program Development, 141 (2014), pp. 4537-4547

Clements, D., Cameleyre, I. and Woodland, H. R. (2003). Redundant early and overlapping larval roles of Xsox17 subgroup genes in Xenopus endoderm development. *Mech Dev* **120**, 337-348.

Darken, R.S. and Wilson, P.A. (2001). Axis induction by Wnt Signaling: Target promoter responsiveness regulates competence. Dev. Biol., 234, 42-54

D'Souza, A., Lee, M., Taverner, N., Mason, J., Carruthers, S., Smith, J. C., Amaya, E., Papalopulu, N. and Zorn, A. M. (2003). Molecular components of the endoderm specification pathway in Xenopus tropicalis. *Dev Dyn* **226**, 118-127.

Ding, Y., Ploper, D., Sosa, E. A., Colozza, G., Moriyama, Y., Benitez, M. D., Zhang, K., Merkurjev, D. and De Robertis, E. M. (2017). Spemann organizer transcriptome induction by early beta-catenin, Wnt, Nodal, and Siamois signals in Xenopus laevis. *Proc Natl Acad Sci U S A* **114**, E3081-E3090.

Doumpas, N., Lampart, F., Robinson, M. D., Lentini, A., Nestor, C. E., Cantu, C. and Basler, K. (2019). TCF/LEF dependent and independent transcriptional regulation of Wnt/beta-catenin target genes. *EMBO J* **38**.

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.

Gammons, M. and Bienz, M. (2018). Multiprotein complexes governing Wnt signal transduction. *Curr Opin Cell Biol* **51**, 42-49.

Gao, C., Xiao, G. and Hu, J. (2014). Regulation of Wnt/beta-catenin signaling by posttranslational modifications. *Cell & bioscience* **4**, 13.

Gupta, R., Wills, A., Ucar, D. and Baker, J. (2014). Developmental enhancers are marked independently of zygotic Nodal signals in Xenopus. *Dev Biol* **395**, 38-49.

Hagey, D. W. and Muhr, J. (2014). Sox2 acts in a dose-dependent fashion to regulate proliferation of cortical progenitors. *Cell Rep* **9**, 1908-1920.

Heasman, J., Kofron, M. and Wylie, C. (2000). Beta-catenin signaling activity dissected in the early Xenopus embryo: a novel antisense approach. *Dev Biol* **222**, 124-134.

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

Heinz, S., Romanoski, C. E., Benner, C. and Glass, C. K. (2015). The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* **16**, 144-154.

Hellsten, U., Harland, R. M., Gilchrist, M. J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N. H., Shu, S., Taher, L., et al. (2010). The genome of the Western clawed frog Xenopus tropicalis. *Science* **328**, 633-636.

Hoffmeyer, K., Junghans, D., Kanzler, B. and Kemler, R. (2017). Trimethylation and Acetylation of beta-Catenin at Lysine 49 Represent Key Elements in ESC Pluripotency. *Cell Rep* **18**, 2815-2824.

Hontelez, S., van Kruijsbergen, I., Georgiou, G., van Heeringen, S. J., Bogdanovic, O., Lister, R. and Veenstra, G. J. (2015). Embryonic transcription is controlled by maternally defined chromatin state. *Nature communications* **6**, 10148.

Hou, L., Srivastava, Y. and Jauch, R. (2017). Molecular basis for the genome engagement by Sox proteins. *Semin Cell Dev Biol* **63**, 2-12.

Houston, D. W. and Wylie, C. (2003). The Xenopus LIM-homeodomain protein Xlim5 regulates the differential adhesion properties of early ectoderm cells. *Development* **130**, 2695-2704.

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.

Hudson, C., Clements, D., Friday, R. V., Stott, D. and Woodland, H. R. (1997). Xsox17alpha and -beta mediate endoderm formation in Xenopus. *Cell* **91**, 397-405.

Hyde, C. E. and Old, R. W. (2000). Regulation of the early expression of the Xenopus nodal-related 1 gene, Xnr1. *Development* **127**, 1221-1229.

Itasaki, N. and Hoppler, S. (2010). Crosstalk between Wnt and bone morphogenic protein signaling: a turbulent relationship. *Dev Dyn* **239**, 16-33.

Kanai-Azuma, M., Kanai, Y., Gad, J. M., Tajima, Y., Taya, C., Kurohmaru, M., Sanai, Y., Yonekawa, H., Yazaki, K., Tam, P. P., et al. (2002). Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development* **129**, 2367-2379.

Karimi, K., Fortriede, J. D., Lotay, V. S., Burns, K. A., Wang, D. Z., Fisher, M. E., Pells, T. J., James-Zorn, C., Wang, Y., Ponferrada, V. G., et al. (2018). Xenbase: a genomic, epigenomic and transcriptomic model organism database. *Nucleic Acids Res* **46**, D861-D868.

Kjolby, R. A. S. and Harland, R. M. (2017). Genome-wide identification of Wnt/beta-catenin transcriptional targets during Xenopus gastrulation. *Dev Biol* **426**, 165-175.

Kormish, J. D., Sinner, D. and Zorn, A. M. (2010). Interactions between SOX factors and Wnt/beta-catenin signaling in development and disease. *Dev Dyn* **239**, 56-68.

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B. E., Bickel, P., Brown, J. B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**, 1813-1831.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359.

Li, B. and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.

Masuda, T. and Ishitani, T. (2017). Context-dependent regulation of the beta-catenin transcriptional complex supports diverse functions of Wnt/beta-catenin signaling. *J Biochem* **161**, 9-17.

Nakamura, Y., de Paiva Alves, E., Veenstra, G. J. and Hoppler, S. (2016). Tissue- and stage-specific Wnt target gene expression is controlled subsequent to beta-catenin recruitment to cis-regulatory modules. *Development* **143**, 1914-1925.

Niida, A., Hiroko, T., Kasai, M., Furukawa, Y., Nakamura, Y., Suzuki, Y., Sugano, S. and Akiyama, T. (2004). DKK1, a negative regulator of Wnt signaling, is a target of the beta-catenin/TCF pathway. *Oncogene* **23**, 8520-8526.

Notani, D., Gottimukkala, K. P., Jayani, R. S., Limaye, A. S., Damle, M. V., Mehta, S., Purbey, P. K., Joseph, J. and Galande, S. (2010). Global regulator SATB1 recruits beta-catenin and regulates T(H)2 differentiation in Wnt-dependent manner. *PLoS Biol* **8**, e1000296.

Owens, N. D. L., Blitz, I. L., Lane, M. A., Patrushev, I., Overton, J. D., Gilchrist, M. J., Cho, K. W. Y. and Khokha, M. K. (2016). Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. *Cell Rep* **14**, 632-647.

Paraiso, K. D., Blitz, I. L., Coley, M., Cheung, J., Sudou, N., Taira, M. and Cho, K. W. Y. (2019). Endodermal Maternal Transcription Factors Establish Super-Enhancers during Zygotic Genome Activation. *Cell Rep* **27**, 2962-2977 e2965.

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842.

Ramakrishnan, A. B. and Cadigan, K. M. (2017). Wnt target genes and where to find them. *F1000Res* **6**, 746.

Risso, D., Ngai, J., Speed, T. P. and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* **32**, 896-902.

Sato, S., Ikeda, K., Shioi, G., Nakao, K., Yajima, H. and Kawakami, K. (2012). Regulation of Six1 expression by evolutionarily conserved enhancers in tetrapods. *Dev Biol* **368**, 95-108.

Schuijers, J., Mokry, M., Hatzis, P., Cuppen, E. and Clevers, H. (2014). Wnt-induced transcriptional activation is exclusively mediated by TCF/LEF. *EMBO J* **33**, 146-156.

She, Z. Y. and Yang, W. X. (2015). SOX family transcription factors involved in diverse cellular events during development. *Eur J Cell Biol* **94**, 547-563.

Sinner, D., Kirilenko, P., Rankin, S., Wei, E., Howard, L., Kofron, M., Heasman, J., Woodland, H. R. and Zorn, A. M. (2006). Global analysis of the transcriptional network controlling Xenopus endoderm formation. *Development* **133**, 1955-1966.

Sinner, D., Kordich, J. J., Spence, J. R., Opoka, R., Rankin, S., Lin, S. C., Jonatan, D., Zorn, A. M. and Wells, J. M. (2007). Sox17 and Sox4 differentially regulate beta-catenin/T-cell factor activity and proliferation of colon carcinoma cells. *Mol Cell Biol* **27**, 7802-7815.

Sinner, D., Rankin, S., Lee, M. and Zorn, A. M. (2004). Sox17 and beta-catenin cooperate to regulate the transcription of endodermal genes. *Development* **131**, 3069-3080.

Sive HL, Grainger RM, & Harland RM (2000) *Early development of Xenopus laevis : a laboratory manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.).

Stevens, M. L., Chaturvedi, P., Rankin, S. A., Macdonald, M., Jagannathan, S., Yukawa, M., Barski, A. and Zorn, A. M. (2017). Genomic integration of Wnt/beta-catenin and BMP/Smad1 signaling coordinates foregut and hindgut transcriptional programs. *Development* **144**, 1283-1295.

Sumi, T., Tsuneyoshi, N., Nakatsuji, N. and Suemori, H. (2008). Defining early lineage specification of human embryonic stem cells by the orchestrated balance of canonical Wnt/beta-catenin, Activin/Nodal and BMP signaling. *Development* **135**, 2969-2979.

Tan, D. S., Holzner, M., Weng, M., Srivastava, Y. and Jauch, R. (2019). SOX17 in cellular reprogramming and cancer. *Semin Cancer Biol*.

Thorvaldsdottir, H., Robinson, J. T. and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192.

Tran, H. T., Sekkali, B., Van Imschoot, G., Janssens, S. and Vleminckx, K. (2010). Wnt/beta-catenin signaling is involved in the induction and maintenance of primitive hematopoiesis in the vertebrate embryo. *Proc Natl Acad Sci U S A* **107**, 16160-16165.

Trompouki, E., Bowman, T. V., Lawton, L. N., Fan, Z. P., Wu, D. C., DiBiase, A., Martin, C. S., Cech, J. N., Sessa, A. K., Leblanc, J. L., et al. (2011). Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell* **147**, 577-589.

Tsankov, A. M., Gu, H., Akopian, V., Ziller, M. J., Donaghey, J., Amit, I., Gnirke, A. and Meissner, A. (2015). Transcription factor binding dynamics during human ES cell differentiation. *Nature* **518**, 344-349.

Valenta, T., Hausmann, G. and Basler, K. (2012). The many faces and functions of beta-catenin. *EMBO J* **31**, 2714-2736.

von Bubnoff, A., Peiffer, D. A., Blitz, I. L., Hayata, T., Ogata, S., Zeng, Q., Trunnell, M. and Cho, K. W. (2005). Phylogenetic footprinting and genome scanning identify vertebrate BMP response elements and new target genes. *Dev Biol* **281**, 210-226.

Wang, P., Rodriguez, R. T., Wang, J., Ghodasara, A. and Kim, S. K. (2011). Targeting SOX17 in human embryonic stem cells creates unique strategies for isolating and analyzing developing endoderm. *Cell Stem Cell* **8**, 335-346.

Weirauch, M. T., Yang, A., Albu, M., Cote, A. G., Montenegro-Montero, A., Drewe, P., Najafabadi, H. S., Lambert, S. A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431-1443.

Xanthos, J. B., Kofron, M., Tao, Q., Schaible, K., Wylie, C. and Heasman, J. (2002). The roles of three signaling pathways in the formation and function of the Spemann Organizer. *Development* **129**, 4027-4043.

Yagi, Y., Ito, Y., Kuhara, S. and Tashiro, K. (2008). Cephalic hedgehog expression is regulated directly by Sox17 in endoderm development of Xenopus laevis. *Cytotechnology* **57**, 151-159.

Zhang, C. and Klymkowsky, M. W. (2007). The Sox axis, Nodal signaling, and germ layer specification. *Differentiation* **75**, 536-545.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137.

Zorn, A. M., Barish, G. D., Williams, B. O., Lavender, P., Klymkowsky, M. W. and Varmus, H. E. (1999). Regulation of Wnt signaling by Sox proteins: XSox17 alpha/beta and XSox3 physically interact with beta-catenin. *Mol Cell* **4**, 487-498.

Zorn, A. M. and Wells, J. M. (2009). Vertebrate endoderm development and organ formation. *Annual review of cell and developmental biology* **25**, 221-251.

FIGURES

Figure 1

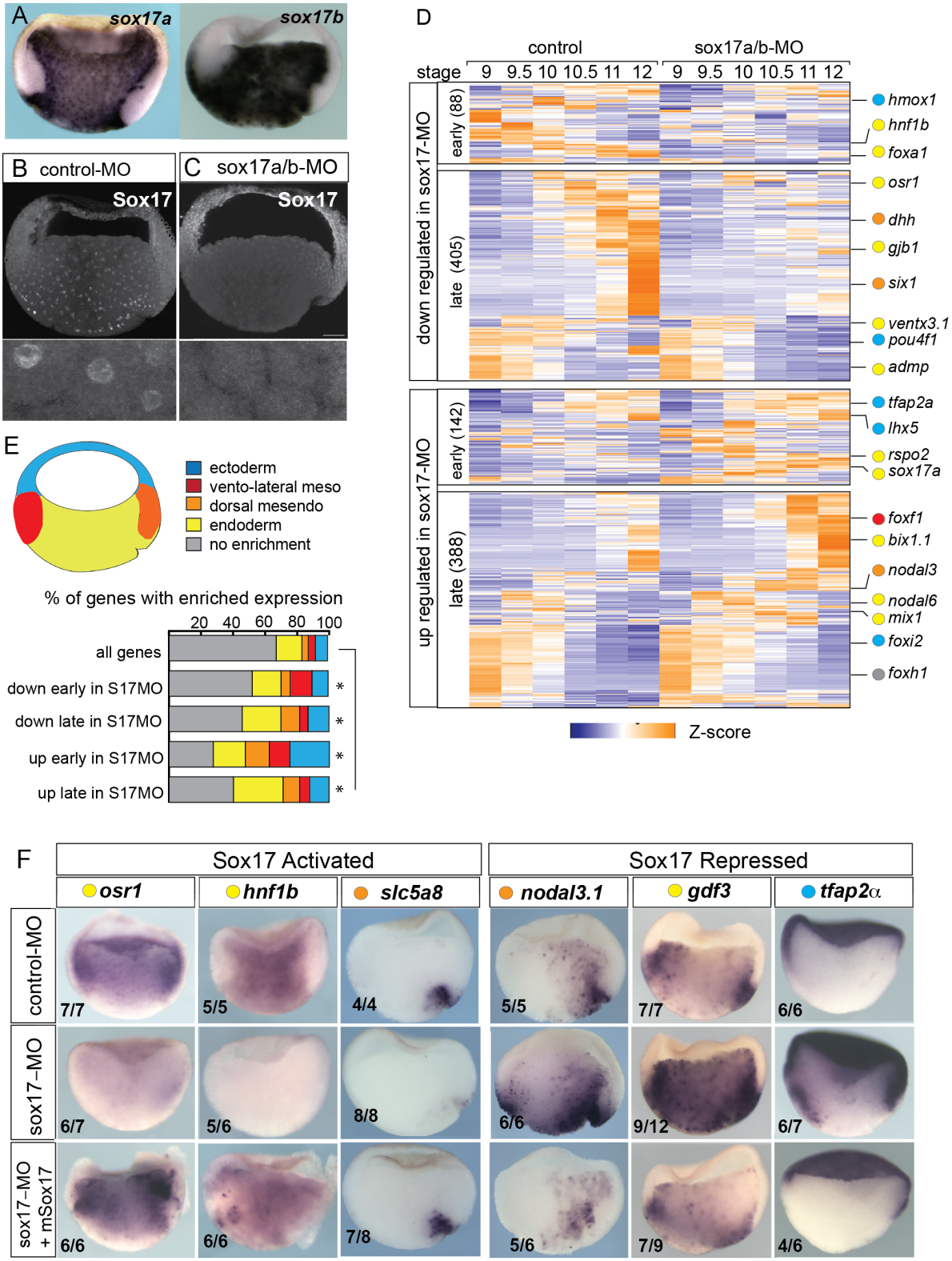
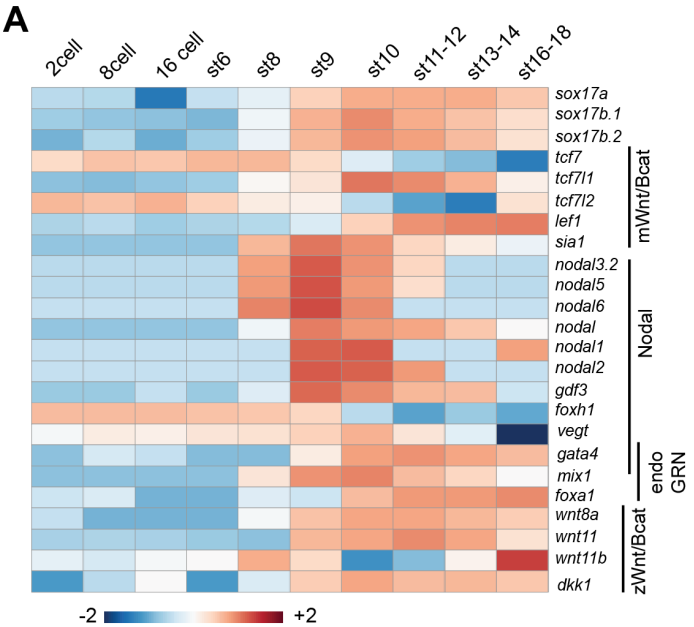
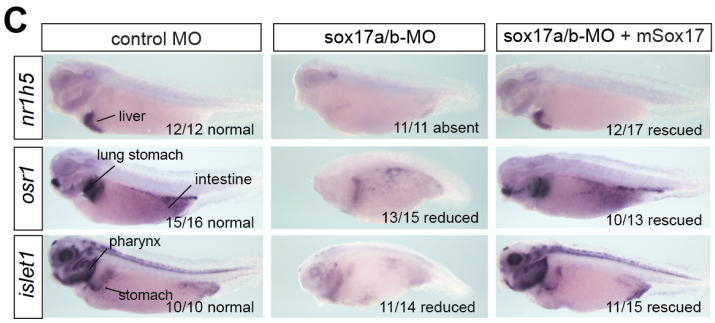


Fig. 1. Sox17-regulated endoderm transcriptome.

(A) Endoderm specific in situ expression of *sox17a* and *sox17b* in NF10.5 *Xenopus tropicalis* gastrula. **(B-C)** Sox17 immunostaining in control-MO (B) and sox17a/b-MO (C) injected gastrula with an antibody that recognizes both Sox17a and Sox17b shows MO effective knockdown. **(D)** Time course heatmap of Sox17-regulated transcripts. Differential expressed genes from a pairwise comparison of control-MO and sox17-MO (Log2 FC< |1|, FDR<5%) at each stage showing transcripts that are down regulated or up regulated early (NF9-10) or late (NF10.5-12). Key genes are color coded based on regional expression from fate map in (E). **(E)** Enriched expression of Sox17-regulated transcripts. Fate map of a bisected gastrula embryos (dorsal right) colored by different tissues. Histogram showing enriched expression in differnet regions of the embryos based on RNA-seq of dissected gastrula tissues (GSE81458; (Blitz et al., 2017). Enriched expression = 1.5 fold > than mean of expression across all tissues. *p<0.05 Chi squared test relative to all genes in the genome. **(F)** In situ validation of Sox17-regulated transcripts. Disrupted expression in sox17-MOs is rescued by co-injeciton of mouse (m) *Sox17* RNA.

Fig. S1

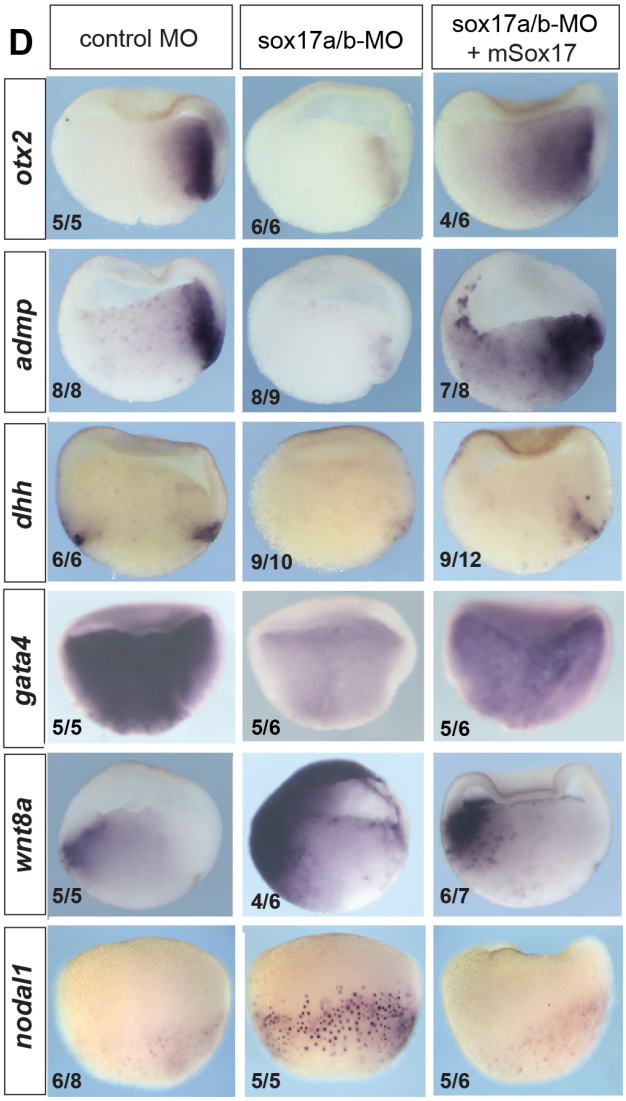




B

```
sox17a:     5-GTCCAACTCTACTCACCATGAGCAGCCCTGATGGTGGCTACG
sox17a-MO:                      3-TGGTACTCGTCGGGACTACCACCGA-5

sox17b.1:   5-GCGCTACTCTCCTCACCATGAGCAGCCCAGATGGTGGCTACG
sox17b.2:   5-GCGCTACTCTCCTCACCATGAGCAGCCCAGATGGTGGCTACG
sox17b-MO:                      3-TGGTACTCGTCGGGTCTACCACCGA-5
```



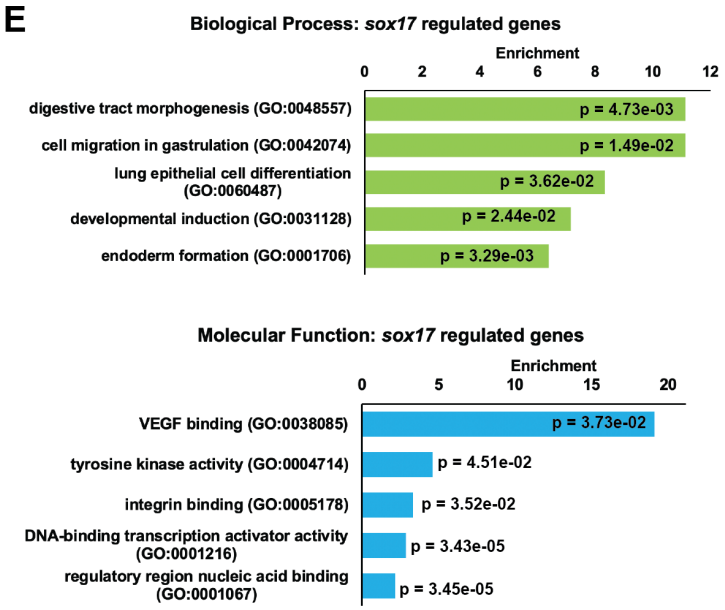


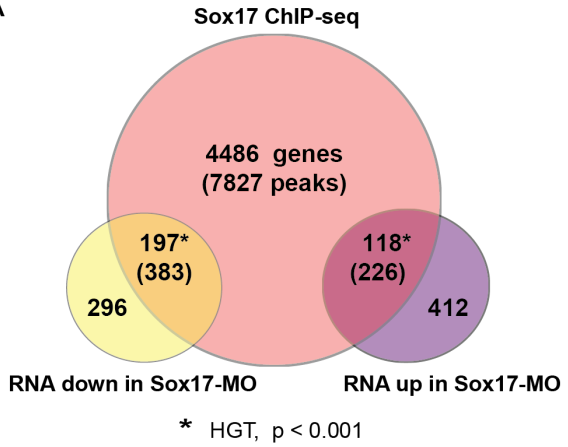
Fig. S1 Characterization of Sox17-MO embryos and Sox17-regulated transctiptome

(A) Temporal expression of key endoderm GRN transcripts (GSE65785; (Owens et al., 2016).

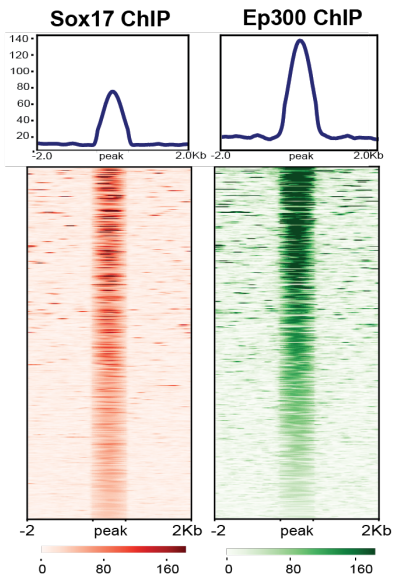
(B) Sequence alignment of antisense sox17a-MO and sox17b-MO (red) with 5' end of *Xenopus tropicalis sox17a, sox17b.1* and *sox17b.2* mRNAs (black) spanning the ATG translation start (green). **(C)** In situ hybridization of markers for liver (*nr1h5*), stomach, lung, intestine and pharynx (*osr1* and *islet1*) snowing that co-injection of mouse *Sox17* RNA rescues gut development in sox17-MO embryos at the tailbud stage. **(D)** Additional validation of gastrula stage trasncripts downregulated (*otx2, admp, dhh* and *gata4*) or upregulated (*wnt8* and *nodal1*) in sox17-MO embryos and resuce by co-injection of mSox17. **(E)** GO term enrichment analysis of Sox17-regulated genes.

Fig. 2

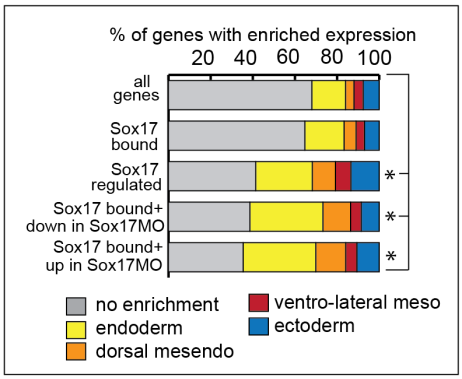
A



B



D

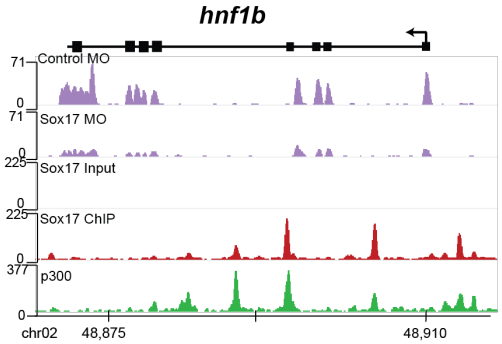


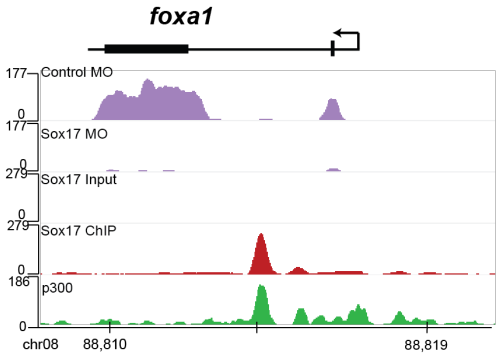
C

Top enriched motifs in 609 Sox17 bound and regulated peaks

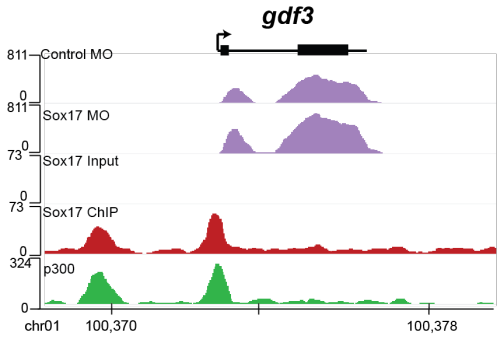
		motif	p-value	%peaks
1.	AACAATGG	Sox	1e-122	74.06
3.	ATTAGCACACAA	Oct4:Sox17	1e-37	9.85
4.	CTTCACACCT	Tbx	1e-37	30.21
5.	CCAGATAAGC	Gata	1e-25	16.09

E **Sox17 bound and activated:**





Sox17 bound and repressed:



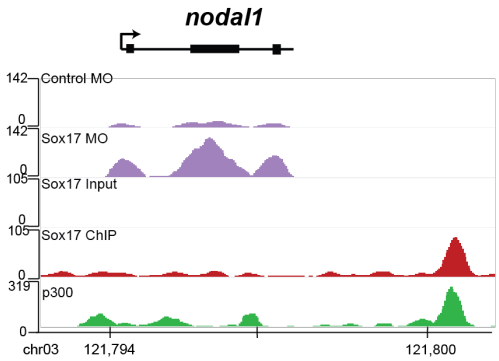


Fig. 2 ChIP-seq identifies Sox17-bound enhancers and direct transcriptional targets.

Sox17 ChIP-seq of gasturla embryos. Venn diagram of Sox17-bound genes (and associated peaks) from ChIP-seq intersected with Sox17-regulated genes. *p<0.001 in hypergeometric test. **(B)** Peak density plots showing that the Sox17-bound loci are cobound by the histone acetyletransferase Ep300, a marker of active enhancers (GSE67974; {Hontelez et al, 2015}. Average density of all peaks in top panel. **(C)** Motif analysis of TF DNA-binding sites enriched in Sox17 peaks. **(D)** Sox17-bound genes are enriched in the endoderm. The Histogram shows the regional expression pattern of Sox17-bound and regulated genes based on RNA-seq of dissected gastrula tissues (GSE81458; (Blitz et al., 2017). Enriched expression = 1.5 fold > than mean of expression in all tissues. *p<0.05 Chi squared test relative to all genes in the genome. **(D)** Genome browser views of representative genes showing RNA-seq expression in control-MO and sox17-MO embryos and the ChIP-seq tracks of Sox17 and Ep300 bound genomic loci.

Fig. S2. Validation of anti-Sox17 antibodies for ChIP

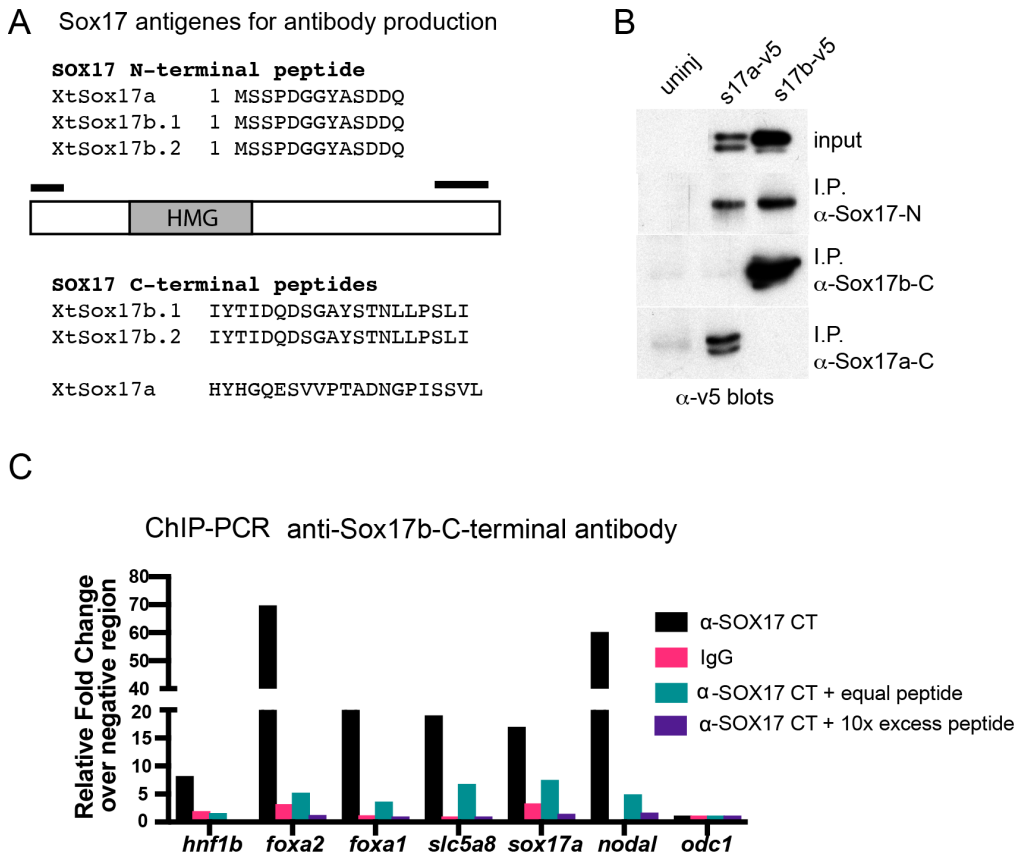
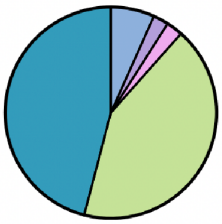


Fig. S2 Validation of anti-Sox17 antibodies.

(A) Schematic and amino acid sequences of the *Xenopus tropicalis* Sox17 peptides uses to immunize rabbits and generate polyclonal antibodies to; pan-Sox17 N-terminal, Sox17b C-terminal and Sox17a C-terminal. **(B)** Immunoprecipitation of *Xenopus laevis* embryos expressing V5 epitope tagged versions of Sox17a-V5 and Sox17b1-V5. Anti-V5 western blot showed that all antibodies specifically immunoprecipitated the target antigen but that the Sox17b C-terminal antibody was the most efficient. **(C)** ChIP-PCR validation of anti-Sox17 antibodies. In preliminary ChIP-PCR experiments the pan-Sox17 N-terminal and Sox17b C-terminal antibodies bound to all the same positive control loci, but the Sox17b C-terminal antibody was more efficient (data not shown) and was used for further optimization and the ChIP-seq analysis. The histogram shows the relative fold change of Sox17b binding versus IgG control in ChIP-PCR of positive control loci and the negative control loci *odc* all normalized to a second negative control loci *eef1a1*. Addition of the Sox17b C-terminal peptide reduced the ChIP signal in a dose-dependent manner indicating specific binding.

Fig. S3 Sox17 ChIP relevant to Fig. 2

A **Distribution of Sox17 peaks (n=8436)**

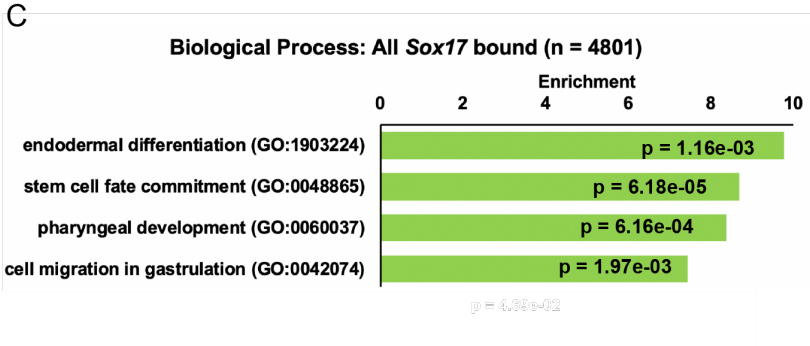


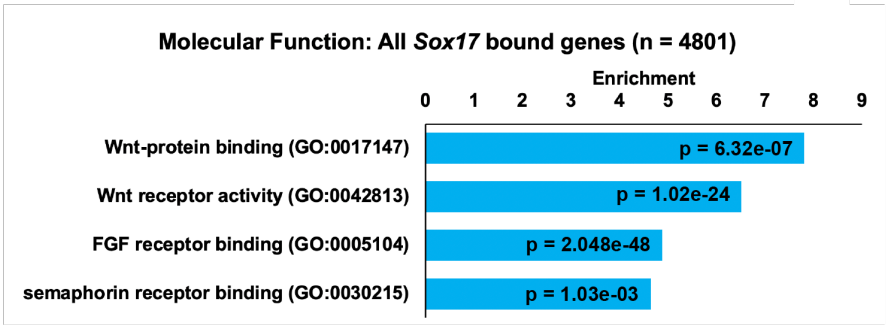
- 6.68% Promoter/TSS (-1kb to 100bp)
- 2.37% TTS (-100bp to +1kb)
- 2.51% Exon
- 42.51% Intron
- 45.94% Intergenic

B ***de novo*** **motif analysis of Sox17 peaks**

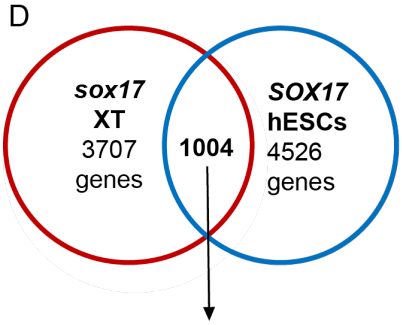
8436 Sox17 peaks	motif	p-value	%peaks
1.	Sox	1e-2460	64.49
2.	Lhx	1e-466	31.28
3.	Oct4:Sox17	1e-370	9.20
4.	Tbx	1e-366	20.06
609 bound and regulated peaks			
1.	Sox	1e-122	74.06
3.	Oct4:Sox17	1e-37	9.85
4.	Tbx	1e-37	30.21
5.	Gata	1e-25	16.09
383 bound and activated peaks			
1.	Sox	1e-82	60.07
2.	Lhx	1e-19	44.96
6.	Tbx	1e-15	7.47
7.	Gsc	1e-12	8.63
226 bound and repressed peaks			
1.	Sox	1e-57	62.39
2.	Tbx	1e-28	33.19
5.	Oct4	1e-13	27.43
6.	Foxh1	1e-13	9.29

C





D



20% of sox17 bound genes in xenopus are also bound by sox17 in hESCs.

E

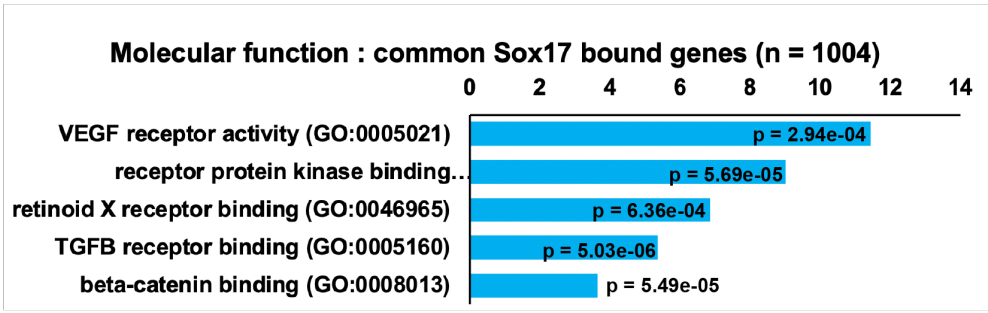
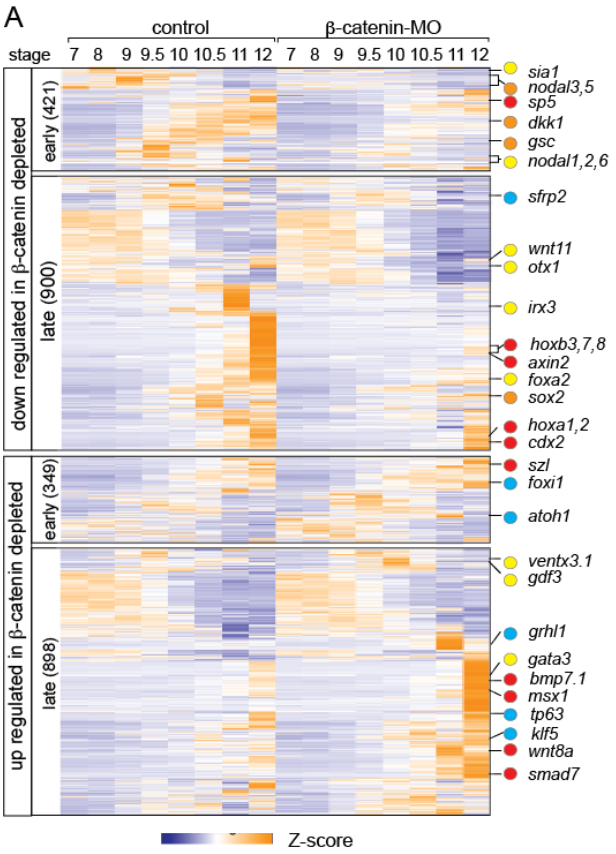


Fig. S3 Analysis of Sox17-bound chromatin loci

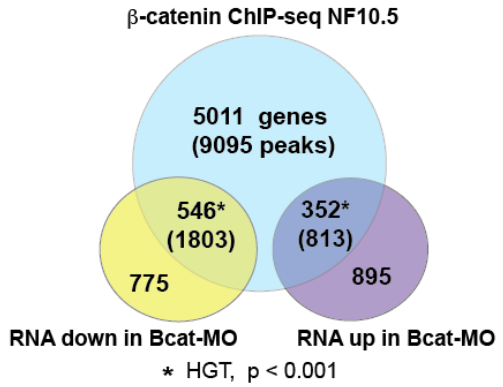
(A) Genomic distriputions of 8436 Sox17 bound chromatin loci. ChIP-seq peaks were associated to cognate promoters using Homer nearest gene function and the location of the peak was annotated relative to the transcription start site (TSS). 88% of the Sox17 peaks were located in either introns or elements more than 1 kb from the TSS, indicative of distal CRMs. **(B)** *De novo* motif analysis (Homer) of Sox17 peaks associated with genes regulated by Sox17 based on the RNA-seq analysis. Sox17-bound peaks that were activated by Sox17 (expression down regualted in Sox17-MOs) were more enriched for Lhx type homeodomain DNA-binding sites, compared to peaks associated with Sox17 repressed genes (expression up regulated in Sox17-MOs), which were enriched for Tbx and Oct motifs. **(C)** GO analysis of 4810 genes associated with Sox17 ChIP-seq peaks showed enrichment for endoderm development and Wnt signaling. **(D)** Comparison of the Xenopus genes associated with Sox17 ChIP-seq peaks to published human SOX17 ChIP-seq from human embryonic stem cell (hESC)-derived definitive endoderm (GSE61475; (Tsankov et al., 2015). The extensive overlap indicates that Sox17 regulates an evolutionarily conserved GRN. The venn only includes Xenopus genes that have human orthologs, unannotated Xenopus genes were excluded from the analysis. **(E)** GO term enrichment analysis of the 1004 Sox17-bound genes conserved between Xenopus and human.

Fig. 3

A



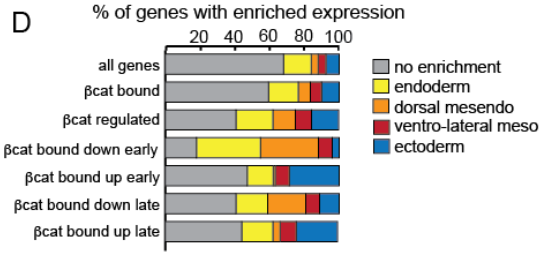
B



C

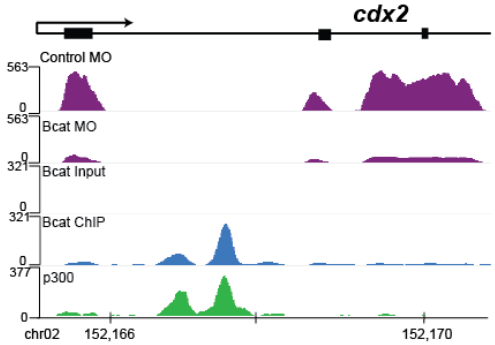
11711 Bcat peaks	motif	p-value	%peaks
1.	AP-2	1e-903	18.35
2.	Sox	1e-862	47.76
3.	Tbx	1e-371	21.22
5.	Lef1	1e-306	9.63

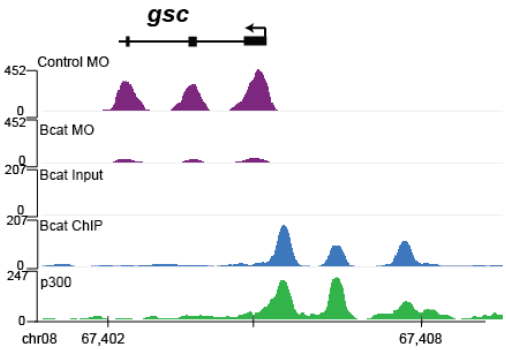
D



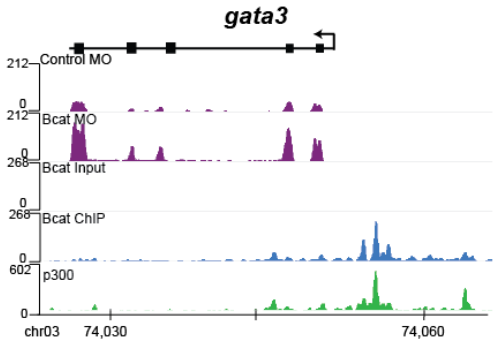
E

Bcat bound and down in Bcat-MO:





Bcat bound and up in Bcat-MO:



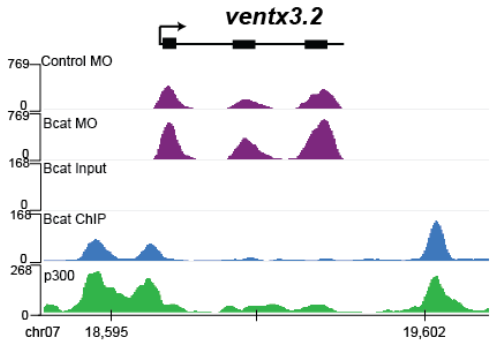


Fig. 3 The Bcat regulated genomic program

(A) Time course heatmap of Bcat regulated transcripts. Differential expressed genes from a pairwise comparison of control-MO and Bcat-MO embryos (Log2 FC< |1|, FDR<5%) at each stage identified 2568 Bcat-regulated transcripts; 1321 downregulated and 1247 upregulated in early (NF7-9.5) or late (NF10-12) Bcat-MO embryos. Key genes are indicated color coded based on regional expression from (D). **(B)** Analysis of published Bcat ChIP-seq data (GSE72657; (Nakamura et al., 2016). identified 11711 peaks associated with 5,909 genes. Venn diagram of Bcat-bound genes (and peaks) intersected with Bcat-regulated genes identified 898 putative direct targets associated with 2616 peaks. *p<0.001 in hypergeometric test. **(C)** Motif aenrichment nalysis of Bcat peaks. **(D)** Histogram showing the expression pattern of Bcat-bound and regulated genes in differnet regions of the embryos based on RNA-seq of dissected gastrula tissues (GSE81458; (Blitz et al., 2017)). Enriched expression = 1.5 fold > than mean of expression in all tissues. **(E)** Genome browser views of representative genes showing the RNA-seq expression in control-MO and Bcat-MO embryos and the ChIP-seq tracks of Bcat and Ep300 bound genomic loci.

Fig. S4

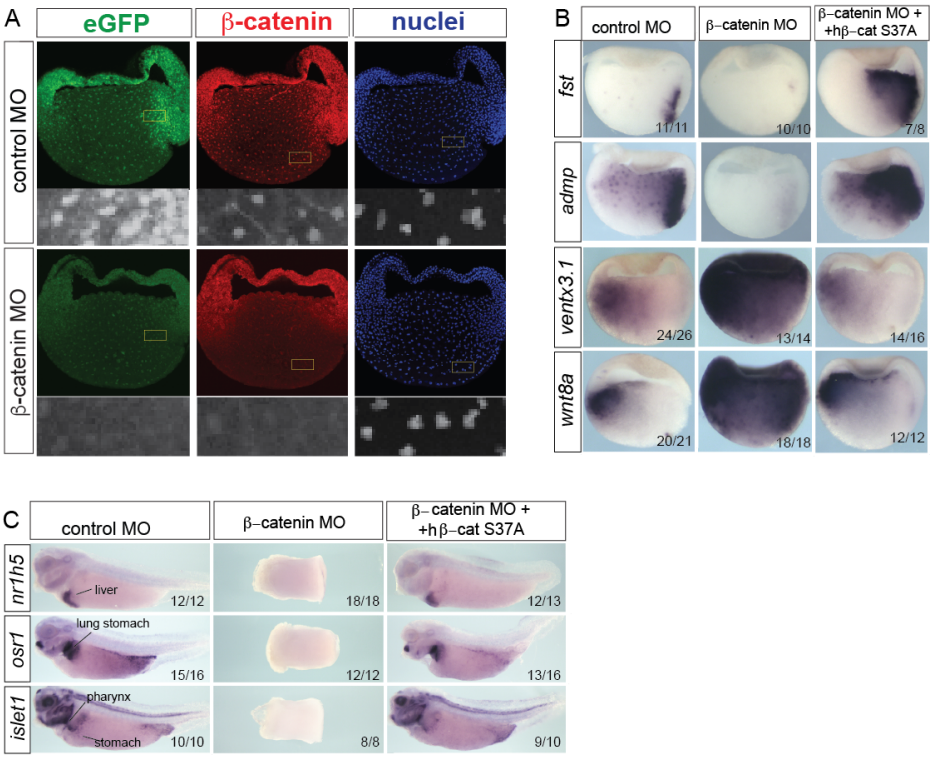


Fig. S4 Validation of Bcat-MO embryos

(A) Immunostaining of *Tg(WntREs:dEGFP)*^{Vlem} *Xenopus tropicalis* gastrula shows the loss of nuclear Bcat and GFP expression from the Wnt-reproter transgene in Bcat-MO embryos demonstrating effective knockdown. **(B-C)** Co-injection of RNA encoding a human S37A stabilized Bcat recues the Bcat-MO ventralized phenotype. In situ hybridization of gastrula (B) and tailbud embryos (C). In Bcat-MO gastrula *fst* and *admp* are downregulated in the dorsal organizer,whereas *wnt8a* and *ventx3.1* are upregulated consistent with a ventralized phenotype. (C) In situ hybridization of markers for liver (*nr1h5*), stomach, lung, intestine and pharynx (*osr1* and *islet1*) snowing that co-injection of human S37A stabilized Bcat RNA rescues the ventralized phenotype and disrupted gut development in bcat-MO tailbud embryos.

Fig. S5 Additional Bcat ChIP-seq analysis

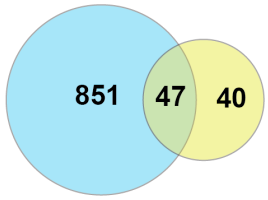
A

de novo motif analysis of Bcat peaks

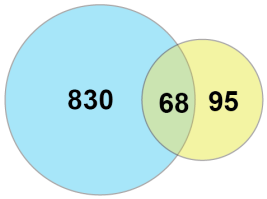
	11711 Bcat peaks	motif	p-value	%peaks
1.		AP-2	1e-903	18.35
2.		Sox	1e-862	47.76
4.		Tbx	1e-371	21.22
5.		Lef1	1e-306	9.63
	2092 bound and regulated peaks			
1.		Hox	1e-168	42.16
2.		Sox	1e-131	40.77
3.		AP-2	1e-112	16.25
4.		Tbx	1e-93	21.46
5.		Lef1	1e-64	29.73
	1426 bound and activated peaks			
1.		Eomes	1e-127	42.87
2.		Sox	1e-82	55.82
3.		AP-2	1e-67	17.46
5.		Lef1	1e-42	23.07
	615 bound and repressed peaks			
1.		AP-2	1e-70	29.11
2.		Oct	1e-50	36.59
3.		Sox	1e-43	35.93

B

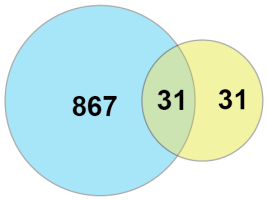
898 Bcat bound and regulated genes NF7-12 (this study)



laevis Bcat-MO regulated
Ding et. al., 2017

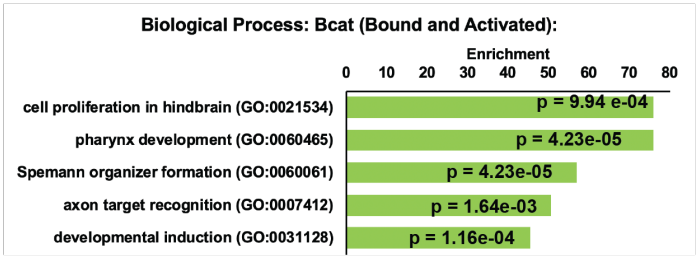


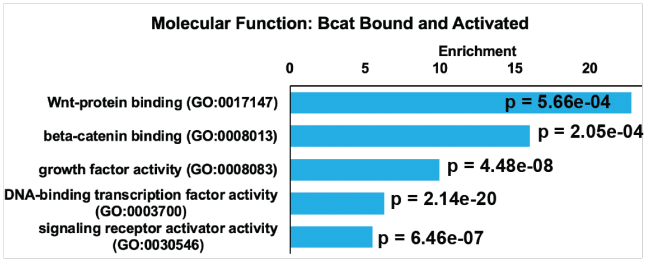
trop zWnt8a regulated NF10.25
Nakamura et. al., 2016

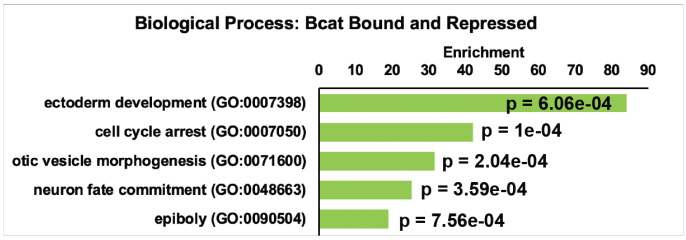


laevis zWnt/Dkk1-regulated
Kjolby and Harland 2017

C







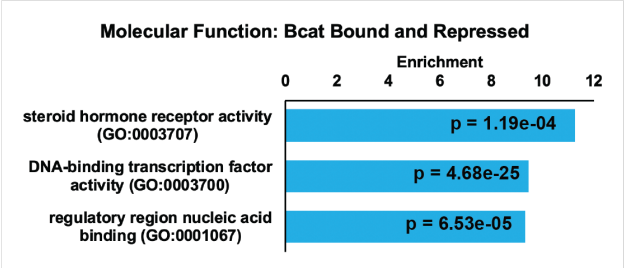


Fig. S5 Additional analysis of Bcat ChIP-Seq

(A) *de-novo* motif analysis of all Bcat bound, Bcat bound and regulated, Bcat bound and activated or repressed peaks. **(B)** Comparison of differentially expressed genes with other recent genomic analysis of gastrula-stage Wnt targets (Ding et al, 2017; Nakamura et al, 2016; Kjolby and Harland, 2017) **(C)** GO enrichment analysis of peaks corresponding to Bcat Bound and either activated or repressed genes.

Fig 4

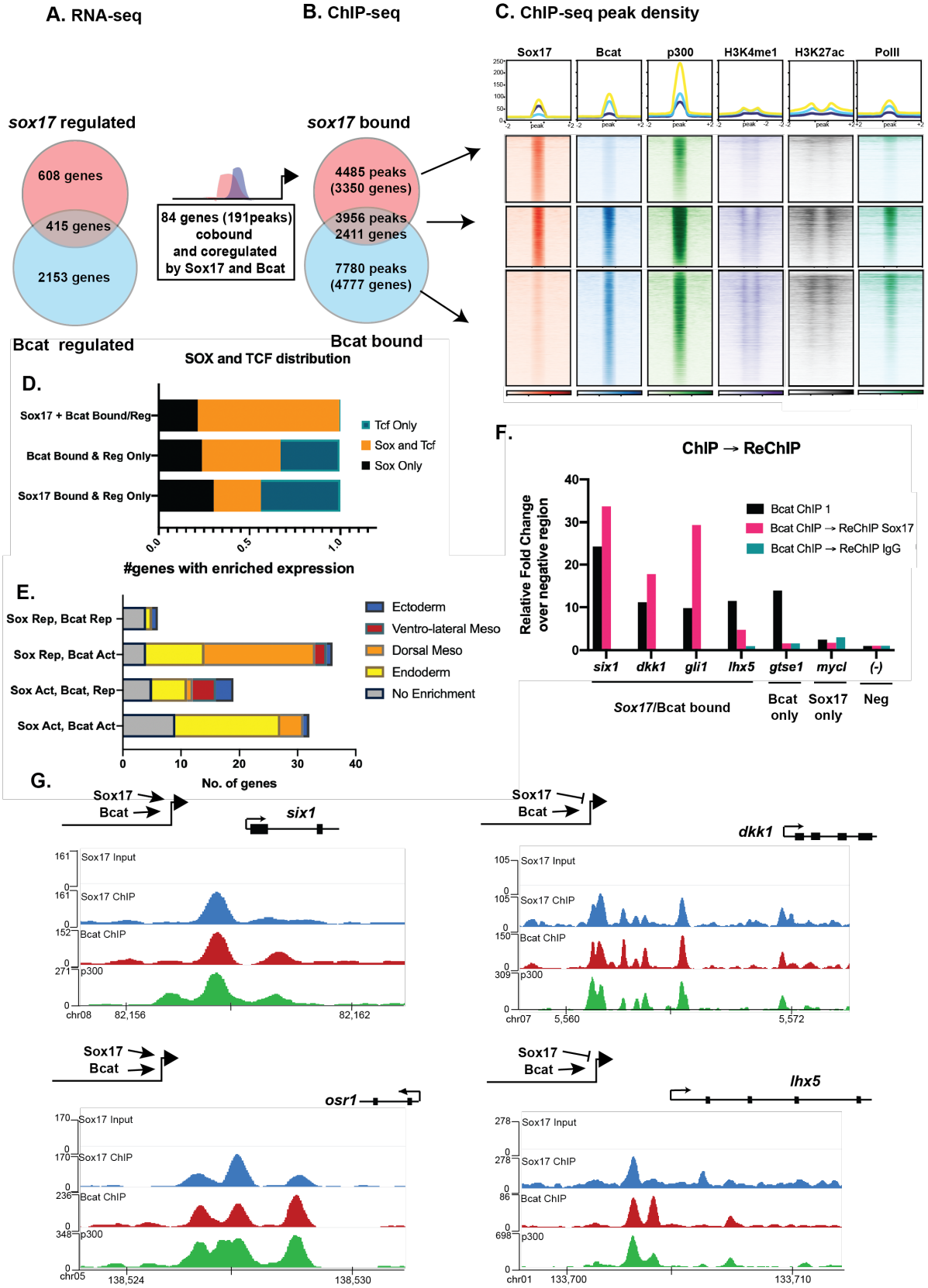

A. RNA-seq
B. ChIP-seq
C. ChIP-seq peak density
sox17 regulated
608 genes
415 genes
2153 genes
84 genes (191peaks) cobound and coregulated by Sox17 and Bcat
sox17 bound
4485 peaks (3350 genes)
3956 peaks 2411 genes
7780 peaks (4777 genes)
Sox17
Bcat
p300
H3K4me1
H3K27ac
PolII
Bcat regulated
Bcat bound
D.
SOX and TCF distribution
Sox17 + Bcat Bound/Reg
Bcat Bound & Reg Only
Sox17 Bound & Reg Only
Tcf Only
Sox and Tcf
Sox Only
#genes with enriched expression
E.
Sox Rep, Bcat Rep
Sox Rep, Bcat Act
Sox Act, Bcat, Rep
Sox Act, Bcat Act
Ectoderm
Ventro-lateral Meso
Dorsal Meso
Endoderm
No Enrichment
No. of genes
F.
ChIP → ReChIP
Relative Fold Change over negative region
Bcat ChIP 1
Bcat ChIP → ReChIP Sox17
Bcat ChIP → ReChIP IgG
six1
dkk1
gli1
lhx5
gtse1
mycl
(-)
Sox17/Bcat bound
Bcat only
Sox17 only
Neg
G.
Sox17
Bcat
six1
Sox17 Input
Sox17 ChIP
Bcat ChIP
p300
chr08
82,156
82,162
dkk1
chr07
5,560
5,572
osr1
chr05
138,524
138,530
lhx5
chr01
133,700
133,710

Fig. 4 Sox17 and Bcat co-bind enhancers to regulated the endoderm GRN

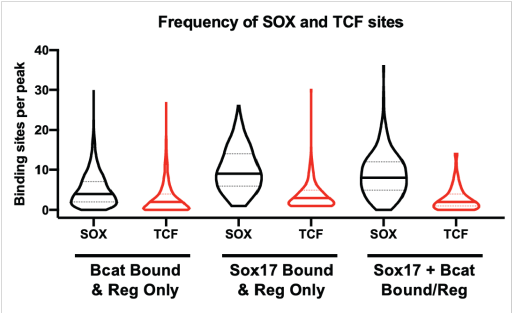
(A) 84 genes (191 Peaks) that are cobound and coregulated by Sox17 and Bcat. Venn diagram shows that 416 genes are regulated by both Sox17 and Bcat (left), whereas 2411 genes (3956 peaks) are bound by both Sox17 and Bcat based on Bedtools peak overlap. **(C)** Peak density plots of Sox17 only peaks, Bcat only peaks and peaks bound by both Sox17 and Bcat shows that the cobound peaks are enriched for acetyletransferase Ep300, H3K4me1 and H3K27Ac epigenetic histome marks and RNA polII, all markers of active enhancers (GSE67974, GSE56000, GSE85273) Average density of all peaks in top panel. **(D)** Distribution of Sox and Tcf binding sites in three groups of enhancers: those bound exclusively by Sox17 or Bcat, and those cobound/coregulated by both. **(E)** Regional enrichment of all subsets of enhancers cobound and coregulated by Sox17 and Bcat. **(F)** ChIP-reChIP experiments show that Sox17 and Bcat coocupy all the test CRMs (six1, dkk1, gli1 and lhx5) in the same cells, in contrast to the *gtse1* loci, which is bound only by Bcat. **(G)** Genome browser views of representative genes showing the RNA-seq expression in control-MO and Bcat-MO embryos and the ChIP-seq tracks showing the genomics loci bound by Bcat and Ep300.

Fig. S6 Additional analysis of cobound and co-reg peaks

A *de novo* motif analysis of Sox17 and Bcat cobound peaks

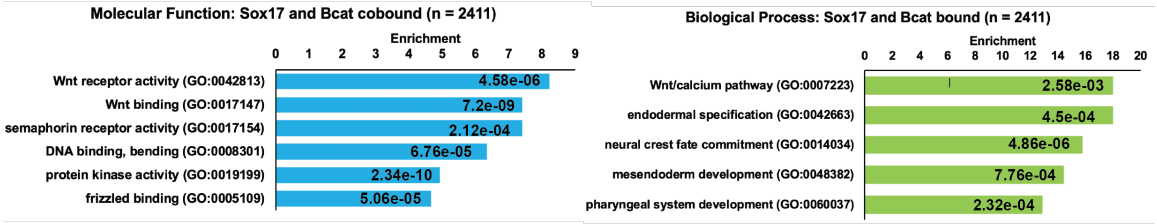
4485 Sox17 only peaks	motif	p-value	%peaks
1.	Sox	1e-1634	79.53
2.	Lhx	1e-212	46.80
3.	Tbx	1e-140	12.73
4.	Prop1	1e-139	17.28
7780 Bcat only peaks			
1.	AP-2	1e-828	24.2
2.	Oct	1e-391	36.63
3.	Lef1	1e-271	13.35
5.	Tbx	1e-225	28.15
14.	Sox	1e-68	13.43
3956 Sox17 and Bcat cobound peaks			
1.	Sox	1e-867	69.64
2.	Lhx	1e-332	34.20
3.	Eomes	1e-250	25.86
4.	Prop1	1e-192	34.20
15.	Tcf	1e-24	0.46
191 Sox17 and Bcat cobound/coregulated peaks			
1.	Sox	1e-23	48.76
8.	Isl1	1e-10	61.98
10.	Lhx	1e-8	44.63
11.	Eomes	1e-8	50.41
12.	Oct4-Sox17	1e-7	14.05

B



C *GO analysis:*

Sox17 and Bcat cobound



Sox17 and Bcat cobound and coregulated

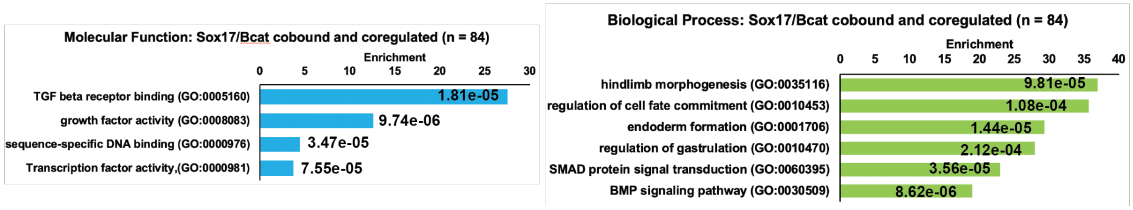


Fig. S6 Additional analysis of cobound and coreg peaks

(A) *de-novo* motif analysis of exclusively bound, cobound and cobound/coregulated Sox17 and Bcat peaks **(D)** Frequency of Sox or Tcf sites per peak in three data subsets: peaks exclusively bound and regulated by Bcat, Sox17 or those that are cobound/coregulated **(C)** Distribution of Sox and Tcf binding sites in three groups of enhancers: those bound exclusively by Sox17 or Bcat, and those cobound/coregulated by both. **(E)** GO enrichment analysis of cobound and cobound/coregulated Sox17 and Bcat peaks.

Fig. 5 Sox17-Bcat epistasis

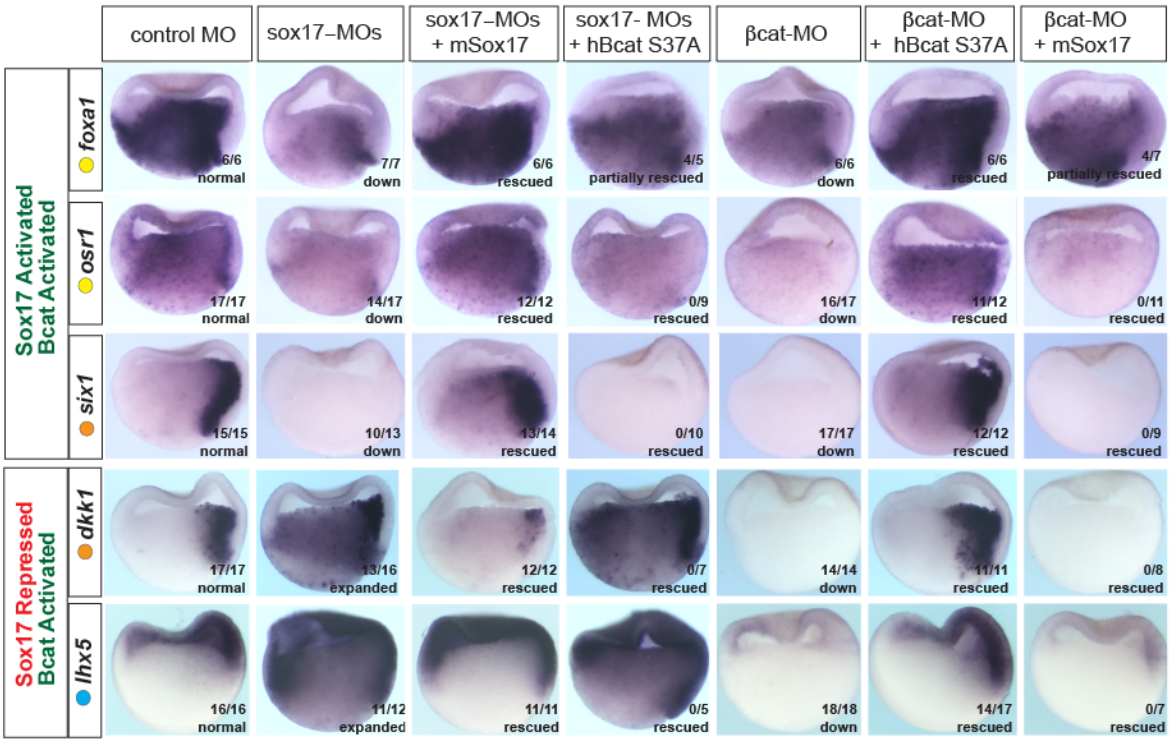


Fig. 5 Sox17 and Bcat epistatsis

In situ hybridization of gastrula embryos injected with sox17-MO or bcat-MO some of which were co-injected with RNA encoding either mSox17 or a human S37A stabilized Bcat. *osr1*, *six1* and *foxa1* are activated by both Sox17 and Bcat, whereas *dkk1* and *lhx5* are activated by Bcat but repressed by Sox17. Co-injection of mSox17 cannot resuce normal expression in Bcat-MOs, nor can hBcat S37A RNA rescue Sox17-MOs.

Fig. 6 Luciferase Assays testing coregulation of Sox17 and Bcat

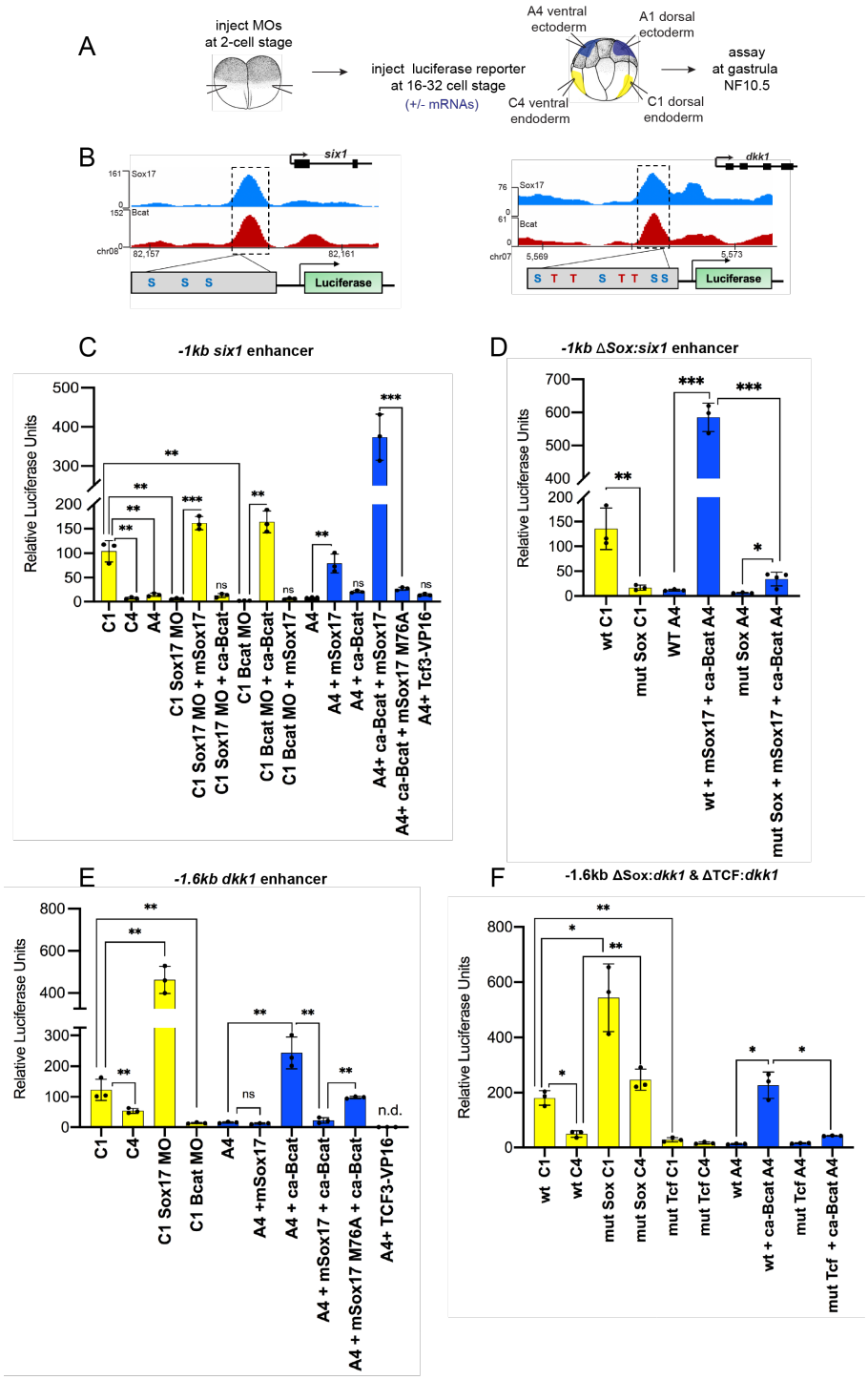


Fig. 6 Luciferase assays testing coregulation of Sox17 and Bcat

(A) Experimental design **(B)** IGV track showing predicted Sox17 binding sites at the -1kb*six1* and -1.6kb*dkk1* enhancers **(C)** Luciferase assay using reporter construct shown in B. **(D)** IGV track showing predicted the -1.6kb*dkk1* enhancer cobound by both Sox17 and Bcat. **(E)** Luciferase assay using reporter construct shown in D.

Fig. S7 – additional data for lucifierase assays

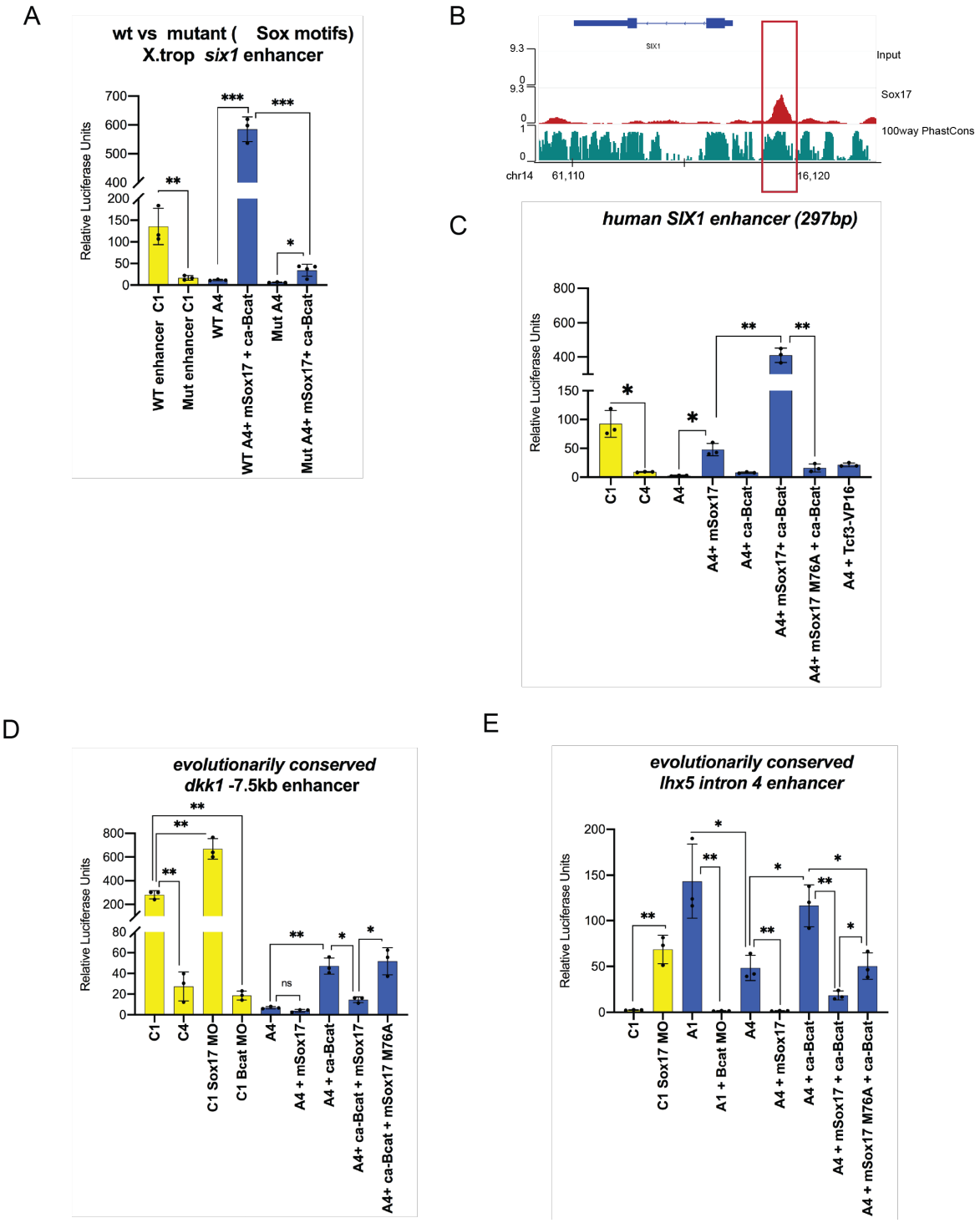


Fig. 7 Additional data for luciferase assays

(A) Luciferase assay of reporter constructed shown in Fig6B, with predicted Sox sites mutated **(B)** IGV track showing evolutionary conserved *SIX1* peak in human DE (Tsankov et al, 2015) **(C)** Additional reporter assay at an evolutionary conserved -7.5kb *dkk1* enhancer and at a **(D)***lhx5* enhancer.

List of Supplementary Data:

1. **Supplementary Table S1:** Differentially expressed genes in RNA-Sequencing of Control and and Sox17MO embryos across NF 9-12.
2. **Supplementary Table S2:** Coordinates of Sox17 bound peaks at NF10.5 with their gene annotations.
3. **Supplementary Table S3:** Differentially expressed genes in RNA-Sequencing of Control and BcatMO embryos across NF 9 -12.
4. **Supplementary Table S4:** Coordinates of Bcat bound peaks at NF10.5 with their gene annotations.
5. **Supplementary Table S5:** Coordinates of 191 Sox17 and Bcat cobound/coregulated enhancers and the frequency of Sox and Tcf sites per peak.
6. **Supplementary Table S6: a.** Plasmid and probe synthesis information **b.** ChIP-qPCR primer table
7. **Supplementary File S1:** List of wild-type and Sox/Tcf mutated sequences used for luciferase assays