Genomic integration of Wnt/β-catenin and BMP/Smad1 signaling coordinates foregut and hindgut transcriptional program

AUTHORS

Mariana L. Stevens[1], Praneet Chaturvedi[1], Scott A. Rankin[1], Melissa Macdonald[1], Sajjeev Jagannathan[2], Masashi Yukawa[2], Artem Barski[2] and Aaron M. Zorn[1*]

AFFILIATIONS

[1]Division of Developmental Biology, Perinatal Institute [2] Division of Allergy & Immunology and Human Genetics, Cincinnati Children's Research Foundation and Department of Pediatrics College of Medicine, University of Cincinnati, Cincinnati, Ohio 45229, USA.

* Correspondence: aaron.zorn@cchmc.org

KEY WORDS

BMP, Smad1, Wnt, beta-catenin, foregut, hindgut, ChIP-seq, RNA-seq, Xenopus, endoderm, mesoderm

SUMMARY STATEMENT

Foregut and hindgut transcriptional programs are regulated by BMP and Wnt signaling through integration of Smad1 and β-catenin on cis-regulatory elements.

ABSTRACT

Digestive system development is orchestrated by combinatorial signaling interactions between endoderm and mesoderm, but how these signals are integrated in the genome is poorly understood. Here we identified the transcriptomes of *Xenopus* foregut and hindgut progenitors, which are conserved with mammals. Using RNA-seq and ChIP-seq we show that BMP/Smad1 regulates dorsal-ventral gene expression in both the endoderm and mesoderm, whereas Wnt/β-catenin acts as a genome-wide toggle between foregut and hindgut programs. Unexpectedly β-catenin and Smad1 binding were associated with both transcriptional activation and repression, with Wnt-repressed genes often lacking canonical Tcf DNA-binding motifs, suggesting a novel mode of direct repression. Combinatorial Wnt-BMP signaling was mediated by Smad1 and β-catenin co-occupying hundreds of DNA cis-regulatory elements, and by a crosstalk where Wnt negatively regulated BMP ligand expression in the foregut. These results extend our understanding of gastrointestinal organogenesis and how Wnt and BMP may coordinate genomic responses in other contexts.

INTRODUCTION

Embryonic development of the digestive and respiratory systems is controlled by a reiterative series of growth factor interactions between the epithelium and mesenchyme (reviewed in Zorn and Wells, 2009). Our understanding of how combinatorial signals orchestrate organogenesis in animal models has been the foundation for strategies to direct the differentiation of human pluripotent stem cells (hPSCs) into organoids for disease modeling and regenerative medicine (reviewed in Lancaster and Knoblich, 2014). Despite our growing knowledge of which growth factors act where and when during organogenesis, how combinatorial signals are integrated at the genomic level to coordinate gene expression through DNA cis-regulatory modules (CRMs) is still poorly understood. Here we investigated how spatially restricted BMP and Wnt signals coordinate the genomic transcriptional programs of foregut (FG) and hindgut (HG) progenitors in *Xenopus* embryos.

In post-gastrula vertebrate embryos and during hPSC differentiation, Wnt and BMP pattern the naïve endoderm (endo) and mesoderm (meso) germ layers along the anterior-posterior (A-P) axis into FG and HG progenitors (Loh et al., 2014; Zorn and Wells, 2009). Bmp4/7 and Wnt8 ligands expressed in the ventral-posterior mesendoderm promote HG fate and inhibit FG lineages (McLin et al., 2007; Rankin et al., 2011; Sherwood et al., 2011; Spence et al., 2011), whereas the anterior mesendoderm secretes Wnt- and BMP-antagonists (e.g., Dkk1, Sfrp5, Noggin and Chordin) that protect the FG from these posteriorizing signals (De Robertis, 2009; Green et al., 2011; Li et al., 2008). Within the meso, these same Wnt-antagonists also promote anterior lateral plate and cardiac fates (reviewed in Gibb et al., 2013; Klaus and Birchmeier, 2009), thus coordinating the development of the FG endo and meso lineages. The effects of these pathways on patterning are temporally restricted such that several hours later, spatially distinct Wnt and/or BMP signals no longer suppress FG identity, but promote lung, thyroid, liver, pancreas and heart organogenesis (Kenny et al., 2012; Klaus and Birchmeier, 2009; Zorn and Wells, 2009).

In both the BMP and Wnt pathways, ligand-receptor binding stimulates the translocation of transcriptional effectors to the nucleus. Activated BMP receptors phosphorylate cytosolic Smad1,5,8 (Smad1), which forms a complex with Smad4 and enter the nucleus to interact with DNA-binding transcription factors, such as Schnurri, Gata or Runx (reviewed in Gaarenstroom and Hill, 2014). In the canonical

Wnt pathway, receptor binding results in stabilization and nuclear translocation of β-catenin, which interacts with DNA-binding Tcf/Lef transcription factors, displacing a co-repressor complex containing Groucho/TLE and recruiting transcriptional co-activators (reviewed in Cadigan and Waterman, 2012). Tcf/β-catenin and Smad1/Smad4 both recruit co-activator complexes containing the p300 or CBP histone acetyltransferases (HATs), which acetylate H3 histones, to promote chromatin opening, RNA polymerase binding and transcription (Cadigan and Waterman, 2012; Gaarenstroom and Hill, 2014).

Combinatorial Wnt and BMP signaling governs cellular responses in a variety of development and disease contexts, and in a few well-characterized target genes β-catenin and Smad1 converge on the same DNA cis-regulatory modules (CRMs) to stimulate transcription (reviewed in Itasaki and Hoppler, 2010). ChIP-seq of *in vitro* differentiated myeloid cells indicates that β-catenin and Smad1 can co-occupy many genomic loci suggesting this may be a widespread mechanism of signal integration (Trompouki et al., 2011). Recent genome-scale studies in *Xenopus* embryos and hPSCs have begun to reveal how β-catenin and Smad2, which transduces Activin/Nodal signals, regulate transcription during germ layer formation and gastrulation (Estaras et al., 2015; Gupta et al., 2014; Kim et al., 2011; Kjolby and Harland, 2016; Nakamura et al., 2016; Tsankov et al., 2015). Despite these advances it is still unknown how Wnt/β-catenin and BMP/Smad1 signals are integrated in the genome to regulate cell fate choices during digestive system organogenesis.

Here we identify the transcriptional program of FG and HG progenitors in *Xenopus laevis* embryos, which are largely conserved with mammals. RNA-seq and ChIP-seq revealed how BMP and Wnt signals coordinate spatially restricted FG and HG gene expression with β-catenin and Smad1 co-binding CRMs of hundreds of key cell identity regulators. We identify a Wnt-BMP crosstalk in the FG and unexpectedly find that many genes inhibited by BMP or Wnt are associated with Smad1 or β-catenin binding, suggesting direct repression. These findings advance our understanding of how combinatorial signaling is integrated in the genome during gastrointestinal development and serve as a paradigm for other development and disease contexts.

RESULTS

Transcriptional program of foregut and hindgut progenitors

In order to examine how BMP and Wnt regulate early gut development at the genomic level, we first defined the FG and HG transcriptomes. Taking advantage of large and abundant *Xenopus laevis* embryos we microdissected the ventral FG-endo, FG-meso, HG-endo and HG-meso tissues from 50 sibling embryos at stage NF20 (22 hours post fertilization; ~ 6-7 somites), and performed RNA-seq (Fig. 1A,B). This time point is similar to E8.5 in mice, when the FG is being patterned and is still plastic. Differential expression analysis of FG (endo + meso) versus the HG (endo + meso), identified 906 FG-enriched genes and 987 HG-enriched genes, whereas endoderm (FG + HG) compared to the mesoderm (FG + HG) identified 3439 endo-enriched and 4829 meso-enriched transcripts (log2FC ≥1, FDR ≤5%) (Fig. 1A-C, Fig. S1A, Tables S1, S2). These gene lists contained over 98% of well-known FG and HG transcripts (n=74) manually curated from mouse and differentiated hPSCs, confirming the extensive conservation (Fig. S1B, Table S3). FG-endo included the transcription factors *hhex* and *pdx1*, while the FG-meso included key regulators of cardiac (*hand2*), myeloid (*spib* and *cebpa*) and endothelial (*fli1*) lineages (Fig. 1D). Gene Ontology (GO) analysis showed epithelial, vasculature and circulatory system development among the top FG-enriched terms (Fig. S1C). The HG transcriptome was enriched for GO terms related to A-P patterning and included the key intestinal regulators *cdx1, 2 and 4* in the HG-endo, and homeobox genes *hox5-11* in the HG-meso (Fig. 1D, Fig. S1B,C, Table S1).

BMP and Wnt regulate gut tube patterning

Consistent with a role in patterning, GO terms related to Wnt and BMP were enriched in the FG and HG datasets (Fig. S1C). Examination of BMP and Wnt pathway components in the total RNA-seq data indicated that BMP ligands (*bmp2, 4 and 7)*, receptors (*bmpr1a*, *bmpr2)* and BMP-target genes (*id3* and *szl)* were variably expressed in both FG and HG tissues, with no obvious A-P difference (Fig. S1D). In contrast, Wnt ligands were enriched in the HG while Wnt-antagonists (*dkk1*, *sfrp2* and *sfpr5*) were restricted to the FG-endo (Fig. S1D).

We next compared the levels of nuclear (n)β-catenin and pSmad1 immunostaining in FG and HG domain, which are marked by *hhex* and *ventx2* respectively (Fig. 1E) (McLin et al., 2007). Consistent with previous reports, the gastrula NF11 anterior mesendoderm (presumptive FG) had low levels of both pSmad1 and (n)β-catenin, while the posterior mesendoderm (presumptive HG) was the opposite (Schohl and Fagotto, 2002). After gastrulation, at stage NF20, (n)β-catenin was still low in the FG and high in the HG, however we observed an up-regulation of pSmad1 in the ventral (but not dorsal) FG (Fig. 1F), consistent with the known *de novo* expression of *bmp2/4/7* in the pre-cardiac FG-meso (Kenny et al., 2012). Thus during FG-HG patterning, BMP and Wnt are differentially active along orthologous axes; pSmad1 is high in the ventral and low in the dorsal FG and HG, whereas (n)β-catenin is low in the FG and high in the HG (Fig. 1G).

We next examined the impact of BMP or Wnt inhibition on progenitor patterning by treating embryos with the BMP-receptor inhibitor DMH1 (40 μM), or inhibiting Wnt with a heatshock inducible *Dkk1* transgenic line *Tg(hsp70:dkk1)* (Lin and Slack, 2008). We added DMH1 or heatshock between stages NF12-20, during FG-HG patterning but after gastrulation to avoid disruption of axial patterning. BMP-inhibition reduced both *hhex* and *ventx2.1*, whereas Wnt-inhibition expanded *hhex* and reduced *ventx2.1* expression (Fig. S1E). To test whether these changes in patterning impacted subsequent organogenesis, we analyzed organ lineages at NF35. BMP inhibition between NF12-20 resulted in an expansion of the dorsal esophageal marker *sox2*, and loss of liver (*nr1h5),* lung (*nkx2-1*) and heart (*nkx2-5)*. In contrast, Wnt-inhibition expanded the liver (*nr1h5)* and reduced the intestinal marker (*darmin*) (Fig. S1E). Thus spatially restricted Wnt and BMP pattern the FG and HG progenitors.

BMP-regulated foregut and hindgut transcriptome

To identify BMP-regulated transcripts, we isolated FG and HG explants (containing both endo and meso) from DMH1 or vehicle treated NF20 embryos and analyzed these by RNA-seq (Fig. 2A). Transcripts with reduced expression in DMH1 compared to controls were classified as normally activated by BMP (n=697), while increased expression in DMH1 indicated that they are repressed by BMP (n=1063) (log2FC ≥1, FDR ≤5%) (Fig. 2B, Fig. S2A,B, Tables S2, S4). Eight genes showed

variable regulation by DMH1 and were not included in further analysis (Fig. S2B), resulting in 1760 unique BMP-regulated genes. Approximately 17% (155/906) of FG-enriched genes were activated by BMP, whereas 21% (185/906) were repressed. Of the HG-enriched transcripts, 10% (97/987) were activated by BMP and 9% (89/987) were repressed (Fig. 2B). BMP-activated genes were enriched for GO terms related to cardiovascular, blood vessel and digestive system development, while BMP-repressed genes were enriched for skeletal and renal system, indicative of dorsal gene expression, normally low in FG and HG explants (Fig. S2C).

BMP signaling was required to maintain expression of key posterior homeobox genes *cdx2*, *cdx4*, *ventx2* and *msx1* in both the HG-endo and HG-meso (Fig. 2C-E). The role of BMP was more complex in the FG. Approximately 85% of the BMP-regulated FG-endo genes appear to be repressed by BMP (Fig. 2C,D), in contrast to the FG-meso where BMP was required for ~60% of the BMP-regulated genes including known regulators of the heart (*tbx20* and *hand2*) and myeloid (*spib* and *cebpa*) lineages (Fig. 2D).

BMP regulates dorsal-ventral patterning of the early foregut

Although most BMP-regulated FG-endo genes were repressed, a few FG-endo genes required BMP for their expression, including *hhex* and *sfrp5* (Fig. 2D,E), which are implicated in liver and ventral pancreas development (Li et al., 2008; McLin et al., 2007). Together with the observation that the ventral FG has higher nuclear pSmad1 levels than the dorsal FG (Fig. 1F) this prompted us to examine dorso-ventral (D-V) patterning in more detail.

Since some digestive organs, including the esophagus, stomach, intestine and pancreas originate from both ventral and dorsal endoderm cells (Chalmers and Slack, 2000), we hypothesized that ventral and dorsal transcripts might be differentially regulated by BMP. Unsupervised clustering of BMP-regulated transcripts in control and DMH1-treated FG explants along with isolated dorsal explants (which contain a thin layer of dorsal endoderm), revealed that the DMH1-treated FG had an expression profile similar to dorsal tissue, suggesting that BMP represses dorsal lineages (Fig. S2D). *In situ* hybridization confirmed that DMH1 treatment reduced the expression of ventral FG genes *sfrp5* (endo), *cebpa* (myeloid) and *nkx2.5* (cardiac), whilst causing the expansion of dorsal endo transcripts *mnx1* and *hrg* and the paraxial meso gene *foxc2*. Injection of BMP2 protein into the FG had

the opposite effect causing an expansion of ventral genes, and repression of dorsal markers (Fig. 2E, Fig. S2E). Thus in the FG between stages NF12-20, BMP promotes ventral (presumptive liver, lung, cardiac) and represses dorsal fates (presumptive esophagus, kidney and paraxial mesoderm).

Smad1 chromatin binding to BMP-regulated hindgut and foregut genes

In efforts to identify direct BMP target genes we performed Smad1 ChIP-seq on NF20 embryos and identified 7976 Smad1 peaks, located within +/- 20kb of 5252 genes (Fig. 3A, Table S5). This represents ~18% of the 45,099 predicted genes in the allotetraploid *Xenopus laevis* genome (Session et al., 2016). Smad consensus DNA-binding sites, as well as Gata and Tbx motifs were enriched in Smad1 peaks (Fig. 3B). Specificity of the Smad1 antibody was confirmed by reduction of Smad1 binding to known BMP-target genes *msx1*, *id3* and *ventx2.1* (Karaulanov et al., 2004) upon DMH1 treatment (Fig. S3A,B). Of the Smad1 peaks, 27% overlapped with p300 ChIP-seq peaks indicative of active transcription, and ~56% of the Smad1 associated genes had detectable expression in NF20 embryos (Fig. S3C,D).

Of the total 1760 BMP-regulated genes in the FG or HG (from Fig. 2B), ~35% (615/1760) were associated with Smad1-binding, a statistically significant enrichment based on a hypergeometric test (HGT) (2.7 fold enrichment (FE), $p<0.05$) (Fig. 3A). These included 48% of the BMP-activated HG genes (47/97, 3.7 FE, HGT *$p<0.05$), such as *cdx2*, *ventx2.1* and *msx1,* consistent with direct activation by Smad1 (Fig. 3C,D). In contrast, Smad1-binding was only associated with 17% (26/155) of BMP-activated FG genes such as *hand2* (meso) and *hhex* (endo) (Fig. 3C,D), suggesting that most BMP-activated FG genes are indirect targets. Unexpectedly, Smad1-binding was more associated with BMP-repressed genes (n=410) than with BMP-activated genes (n=205). Indeed, ~55% of BMP-repressed FG genes (103/185, 4.3 FE, HGT *$p<0.05$) and ~49% of BMP-repressed HG genes (44/89, 3.8 FE, HGT *$p<0.05$) were associated with Smad1 peaks, suggesting direct BMP repression, including dorsal genes such as *foxc2* and *mnx1* (Fig. 3D). Analysis of p300 co-occupancy on different classes of Smad1 peaks did not however show a statistically significant enrichment on activated versus repressed genes, presumably because the ChIP was done on whole embryos containing both expressing and non expressing cells. In sum, about a third of the BMP regulated FG and HG

transcriptome was associated with Smad1-binding suggesting direct transcriptional activation and repression.

Canonical Wnt promotes hindgut and represses foregut transcriptomes

To determine the genomic basis of Wnt-mediated A-P patterning, we performed RNA-seq on FG and HG explants (endo + meso) from NF20 embryos where Wnt/β-catenin activity was either activated with BIO treatment or inhibited with secreted Wnt-antagonist Dkk1 expressed from *Tg(hsp70:dkk1)* from stages NF12-20 stages (Fig. 4A). Transcripts repressed by Dkk1 or induced by BIO, relative to controls, were considered Wnt-activated, while transcripts up regulated by Dkk1 or repressed by BIO were considered Wnt-repressed (Fig. S4A). Differential expression analysis identified 959 Wnt-activated transcripts and 2032 Wnt-repressed transcripts (log2FC ≥1, FDR ≤5%). 41 transcripts showed variable regulation by BIO and Dkk1 and were not included in further analysis (Fig. 4B, Fig. S4A, Tables S2, S6). Thus we classified a total of 2991 Wnt-regulated genes. Interestingly, over half of the FG-enriched genes (496/906) were repressed by Wnt, whereas only 3% (28/906) were activated. HG-enriched genes exhibited the opposite behavior, with 25% (247/987) being Wnt-activated and only 6% (57/987) repressed (Fig. 4B). GO term enrichment showed A-P patterning among the top terms for Wnt-activated genes, while Wnt-repressed genes were enriched for terms related to the circulatory system, consistent with the known role of Wnt in repressing early cardiac fate (Fig. S4B). Unsupervised clustering of FG and HG transcript levels in the various experimental conditions revealed that BIO-treated FG samples cluster with HG samples, indicating a genome-scale switch from FG to HG upon Wnt stimulation (Fig. 4C), which is also clearly shown in the scatter plots (Fig. 4D, Fig. S4C,D).

Wnt had the same impact on endo and meso transcripts - repressing in the FG and activating in the HG - regardless of germ layer. Wnt-repressed FG transcripts included *hhex* and *sfrp2* in the FG-endo, as well as *spib* and *hand2* in the FG-meso. HG transcripts activated by Wnt included *ventx2.1 and cdx2* in the HG-endo, as well as *hoxa9 and hoxd10* in the HG-meso (Fig. 4E). *In situ* hybridization confirmed that Dkk1 expanded expression of the FG genes *gata4, cebpa* and *tbx1,* whilst reducing the HG transcripts *cdx2* and *msx1*, with BIO treatment having the

opposite effect (Fig. 4F). Thus Wnt/β-catenin acts as a genome wide switch, promoting the HG transcriptional program and repressing the FG program.

Chromatin binding of β-catenin is associated with activation of HG and repression of FG genes

To determine how Wnt regulates A-P patterning at the genomic level, we performed β-catenin ChIP-seq on stage NF20 control and BIO-treated FG and HG explants (Fig. S5A). We merged sequence files from the four different ChIPs to call β-catenin peaks. This identified 16303 β-catenin peaks associated with 11007 genes (+/-20kb) (Fig. 5A, Fig. S5B, Table S7), which represents ~24% of the genes in the genome. As expected, β-catenin peaks were enriched for Tcf/Lef DNA-binding motifs and ChIP-PCR of known Wnt-target genes in *Tg(hsp70:dkk1)* embryos confirmed the expected reduction in β-catenin binding (Fig. S5C,D).

Of the 11,007 genes associated with β-catenin peaks only 1,243 (~11%) were Wnt regulated based on our RNA-seq data (Fig. 5A), suggesting that ~90% of all genomic β-catenin binding events in the FG or HG tissue were not associated with Wnt-regulated transcription, similar to recent findings in the *Xenopus* gastrula (Nakamura et al., 2016). Of all Wnt-regulated genes, ~41% (1243/2991, 1.5 FE, HGT $p<0.05$) were associated with β-catenin peaks, including 73% (180/247, 2.7 FE, HGT *$p<0.05$) of the HG-enriched Wnt-activated genes (Fig. 5Aa), suggesting direct regulation. Unexpectedly, 42% (208/496, 1.6 FE, HGT *$p<0.05$) of all FG-enriched Wnt-repressed genes were also associated with β-catenin peaks (Fig. 5Ab), suggesting direct repression. This was surprising since β-catenin is usually thought to stimulate transcription and thus we expected that Wnt would indirectly repress FG genes. We next assessed whether activation or repression correlated with changes in recruitment of β-catenin to chromatin through differential peak enrichment analysis of control versus BIO explants (Fig. 5B). As expected, 88% (159/180) of the Wnt-activated HG genes had increased β-catenin binding in BIO-treated FG tissue, as exemplified by *cdx2* (Fig. 5Ba',C). Surprisingly, among the 208 Wnt-repressed FG genes (Fig. 5Ab), we observed different behaviors; 59 genes were associated with increased β-catenin binding upon BIO treatment (Fig. 5Bb'), 43 genes experienced reduced β-catenin binding (Fig. 5Bb"), whereas 91 Wnt-repressed genes had no change in β-catenin (Fig. 5B). For example, β-catenin binding near the *sfrp2*

promoter was increased upon BIO treatment, whereas the *nkx2-3* peak was reduced (Fig. 5C). Plotting the average tag density of these different classes of peaks confirmed the significant changes in β-catenin recruitment upon BIO treatment (Fig. 5D, Fig. S5E,F). These data suggest that, in the context of Wnt-repressed genes, elevated nuclear β-catenin levels (from BIO) do not necessarily correlate with increased chromatin recruitment.

We next performed p300 ChIP-seq in control and BIO treated FG and HG tissue to determine how recruitment of the HAT co-activator complex correlated with changes in β-catenin binding (Fig. 5E, Table S7). Examination of individual genes such as *cdx2* (Fig. 5C), as well as average tag density analysis confirmed that p300 was recruited to β-catenin peaks of HG Wnt-activated genes upon BIO treatment (Fig. 5E, Fig. S5E). For both classes of Wnt-repressed FG genes, with either increased β-catenin (e.g. *sfrp2*) or reduced β-catenin (e.g. *nkx2-3*) recruitment upon BIO, we observed a trend of reduced p300 recruitment to β-catenin peaks, consistent with repression (Fig. 5C), although this was not statistically significant (Fig. 5E, Fig. S5F). To further investigate how β-catenin recruitment might activate some genes and repress others, we performed motif enrichment analysis on different classes of peaks. Tcf was the most enriched motif in HG Wnt-activated peaks. Surprisingly, Tcf DNA-binding sites were not enriched in β-catenin peaks from FG Wnt-repressed genes; rather these were enriched for Gata, Sox and TEAD motifs (Fig. 5F), suggesting that other DNA-binding proteins might form co-repressor complexes with β-catenin.

Smad1 and β-catenin converge on common CRMs to coordinate transcription

We next investigated the extent to which BMP and Wnt cooperate to regulate FG and HG transcription. A comparison of BMP-regulated and Smad1-bound genes with Wnt-regulated and β-catenin-bound genes revealed 229 coordinately regulated genes (Fig. 6A, Fig. S6A). Of these, 33 genes had a total 111 non-overlapping Smad1 and β-catenin peaks suggesting regulation through separate CRMs, while the other 196 genes had a total of 369 overlapping β-catenin and Smad1 peaks suggesting co-occupied CRMs (Fig. 6B). This included 21 genes activated by both pathways, primarily regulatory factors known to confer HG fate like *cdx2*, *ventx2* and *msx1* (Fig. 6C,D). 62 genes with overlapping β-catenin/Smad1 peaks were Wnt-

activated but BMP-repressed, many of which had a paraxial and intermediate mesoderm signature, whereas 55 genes were repressed by both Wnt and BMP including FG genes *irx1-3*, *sfrp2* and *tbx1*. Finally, a group of 56 genes were BMP-activated and Wnt-repressed, including the ventral FG-endo gene *hhex,* cardiac genes such as *nkx2-3* as well as *bmp4, bmp7* and the BMP-target gene *szl* (Fig. 6C,D, Table S8). Although comparable FG and HG datasets are not available in mammals, examination of public ChIP-seq data from human ES and transformed cell lines (Estaras et al., 2015; Tsankov et al., 2015; Watanabe et al., 2014) revealed that ~47% (92/194) of the human orthologs have β-CATENIN and/or SMAD1 peaks in regions of the genome syntenic to the co-occupied peaks in *Xenopus* (Fig. S7, Table S9) suggesting considerable conservation.

Motif analysis of these distinct classes of co-regulated genes revealed differential enrichment for Tcf, Smad, Gata, Lhx and/or Foxa sites (Fig. 6E) suggesting that different transcription factors probably influence whether β-catenin and Smad1 recruitment to chromatin results in transcriptional activation or repression. As a proof-of-principle that co-bound CRMs are coordinately regulated by Wnt and BMP, we tested the *hhex* distal region (lacking a Tcf-motif) and a *cdx2* intronic region (that contains a Tcf-motif) in embryo injection luciferase reporter assays (Fig. 6F). The *cdx2:luc* reporter was more active in the HG than the FG as expected, whereas the *hhex:luc* was active in both the FG and the HG, indicating that this CRM alone was not sufficient to confer FG restricted expression. As predicted, both *hhex:luc* and *cdx2:luc* were repressed upon DMH1 treatment (Fig. 6F), consistent with both genes being BMP-activated. On the other hand, BIO suppressed the *hhex:luc* construct but stimulated *cdx2:luc* activity consistent with Wnt repressing the FG gene *hhex* and activating the HG gene *cdx2* (Fig. 6F).These data indicate that binding of Smad1 and β-catenin to the same CRMs coordinate Wnt and BMP responsive gene expression to pattern the FG and HG progenitors.

The observation that the *bmp4* and *bmp7* loci are repressed by Wnt and activated by BMP, and that they both had CRMs with overlapping β-catenin and Smad1 peaks suggested an additional layer of Wnt-BMP signaling crosstalk (Fig. S6B). *In situ* hybridization confirmed that Wnt negatively regulates expression of *bmp4/7* as well as *bmp2* in the FG, and to a lesser extent in the HG. In contrast BMP signaling was required to maintain robust *bmp2/4/7* expression in both the FG and

HG (Fig. S6C,D). Interestingly, in the gastrula (when the BIO and DMH1 treatment is started) *bmp2* is expressed in the anterior mesendoderm (future FG), a low Wnt environment, whereas *bmp4/7* are expressed in the ventral/posterior mesendoderm (future HG) a high Wnt environment (Hoppler and Moon, 1998). Taken together this suggests that low Wnt is required to maintain *bmp2* in the presumptive FG, which then initiates a known positive BMP feedback loop to promote *bmp4/7* expression in the FG (Karaulanov et al., 2004; Kenny et al., 2012) (Fig. 7A).

Overall these data reveal a complex BMP-Wnt gene regulatory network that coordinates A-P and D-V patterning of the FG and HG progenitors (Fig. 7A), with β-catenin and Smad1 converging on CRMS to regulate both transcriptional activation and repression (Fig. 7B, Table S10).

DISCUSSION

In embryonic development and hPSC differentiation, combinatorial growth factor signaling controls cell fate decisions, but how these signals coordinate gene expression at the genomic level is poorly understood. We investigated how BMP and Wnt signaling, differentially active along orthologous embryonic axes, are integrated in the genome to coordinate the FG and HG transcriptomes (Fig. 7A). BMP/Smad1 activity, high in the ventral and low in the dorsal gut tube, regulates D-V identity in the FG, and promotes expression of key HG genes. On the other hand, low Wnt/β-catenin activity is required for FG fate, while high Wnt induces HG and represses FG transcription. We defined Smad1- and β-catenin-bound CRMs of many Wnt- and BMP-target genes, including lineage specifying transcription factors. Our data suggest a much more complicated regulatory landscape than previously appreciated with Smad1 or β-catenin binding correlated with either activation or repression, and combinatorial Wnt and BMP signaling converging on hundreds of common CRMs to regulate FG and HG specific responses (Fig. 7, Table S10). While there are many aspects of the model to be resolved with detailed cis-regulatory analysis, this provides a framework for understanding how β-catenin and Smad1 are integrated in the genome to control early gut patterning.

We found that combinatorial Wnt and BMP signaling coordinates A-P and D-V patterning in the Xenopus embryo in a manner, almost identically to recent hPSC differentiation protocols where BMP^{low}/Wnt^{low} specify general FG-endo, but

BMP^{high}/Wnt^{low} promote ventral FG lineages (Green et al., 2011; Loh et al., 2014). BMP^{high}/Wnt^{low} also promotes cardiac, myeloid and endothelial fates, whilst repressing intermediate and paraxial mesoderm fates, which require BMP^{low}/Wnt^{high} (Loh et al., 2016). Finally, BMP^{high}/Wnt^{high} promotes intestinal fate (Spence et al., 2011).

Our ChIP-seq results suggest a model for how this differential Wnt and BMP signaling regulates FG and HG transcription (Fig. 7B), and identified hundreds of putative β-catenin and Smad1-bound CRMs associated with known and novel target genes. In a comparison to recent β-catenin ChIP-seq of *Xenopus* gastrula, we found that 81% (695/849) of *X. laevis* (Kjolby and Harland, 2016) and 43% (1970/4529) of *X. tropicalis* (Nakamura et al., 2016) genes associated with β-catenin occupancy at the gastrula were also in our dataset. These included HG homeobox genes such as *cdx2*, *ventx1*, *hoxd1* and *msx1*, consistent with Wnt promoting the expression of these posterior genes starting at gastrulation. We also identified several hundred genes co-regulated by Wnt and BMP that exhibit overlapping β-catenin and Smad1 peaks, many of which appear to be conserved in humans based on public ChIP-seq data from hPSCs and transformed cell lines (Benahmed et al., 2008; Gaunt et al., 2003; Tsankov et al., 2015; Watanabe et al., 2014), suggesting a conserved paradigm.

Most BMP-activated genes in the FG were not associated with Smad1-binding, suggesting indirect regulation. Despite this, a small cohort of genes encoding lineage-specifying transcription factors such as, *nkx2-3*, *nkx2-5*, *hand2*, *spib, gata5*, and *hhex* were associated with Smad1-bound CRMs. In the case of *Nkx2-5* this Smad1-bound CRM is conserved in mammals (Lien et al., 2002), and public ChIP-seq data from hPCS indicates that SMAD1-binding can also occur at the human *HHEX* loci (Tsankov et al., 2015) (Fig. S7). This suggests that in the FG, BMP/Smad1 initiates a cascade of transcription factors that promote cardiac, myeloid and hepatic fates. Gata motifs were enriched in our Smad1 peaks, suggesting they might be key components of this regulatory cascade. Gata factors are known to be required for cardiac and foregut development and can physically interact with Smad1 to regulate target gene transcription (Benchabane and Wrana, 2003; Brown et al., 2004; Haworth et al., 2008; Rossi et al., 2001; Trompouki et al., 2011; Xuan et al., 2012).

For most BMP- and Wnt-activated genes, our data support the canonical model of Smad1 and β-catenin associating with DNA-binding proteins (Tcf in the case of β-catenin) to recruit HAT co-activator complex and stimulate transcription. Unexpectedly about 55% of BMP-repressed and 73% of Wnt-repressed FG genes associated with Smad1 and β-catenin peaks, respectively, suggesting direct repression. Motif analysis indicated that these peaks were not enriched in consensus Tcf or Smad DNA-binding sites; Wnt-repressed genes were enriched for Gata, Sox and TEAD motifs, whilst BMP-repressed genes were enriched for Gata, Lhx and Nkx. Transcriptional repression by Smad or β-catenin has only been documented in a handful of cases. For example, Smad1 can bind to Nkx2-3 on DNA and recruit Sin3/HDAC1 co-repressors to inhibit reporter construct expression in response to BMP signaling (reviewed in Blitz and Cho, 2009; Kim and Lassar, 2003; Marty et al., 2000), consistent with Nkx motifs in many of our Smad1 peaks (Fig. S3E). In mammalian and *Drosophila* cells, β-catenin can recruit co-repressor complexes to repress *e-cadherin* and *dpp (bmp)* transcription respectively (Jamora et al., 2003; Olson et al., 2006; Theisen et al., 2007), consistent with Wnt repression of *bmp4/7* that we observed in the FG. In recent years, a number of transcription factors have been shown to interact with β-catenin in different cellular contexts, including Sox, homeobox and TEAD (Estaras et al., 2015; reviewed in Kormish et al., 2010), suggesting that different DNA-binding proteins determine whether Smad1 or β-catenin recruit co-activator or co-repressor complexes.

Interestingly most Wnt-repressed FG genes had β-catenin peaks in both FG and HG tissues. One possibility is that the FG explants contained some HG cells, consistent with low levels of *cdx2* mRNA and expression of the *cdx2:luc* reporter in the FG. Alternatively, β-catenin activity levels might impact a switch between activation and repression, since we have previously shown that low levels of Wnt/Fzd7 are required for *hhex* expression, while high levels are inhibitory (Zhang et al., 2013). However, unlike Wnt-activated genes, BIO treatment (which inhibits GSK3 and stabilizes β-catenin) did not strictly correlate with increased β-catenin recruitment to Wnt-repressed FG genes. Since GSK3, has other substrates besides β-catenin (Ding et al., 2000; Wu and Pan, 2010), is it possible that some of the BIO regulated gene expression is Wnt-independent.

Another striking observation was that the majority of Smad1 and β-catenin peaks were not associated with BMP or Wnt regulated transcription. This is similar to recent findings in *Xenopus* gastrula and hPSCs (Nakamura et al., 2016; Tsankov et al., 2015), and is consistent with an emerging concept that transcription factor binding is pervasive throughout the genome even when they are not engaged in productive transcription (Nakamura et al., 2016; reviewed in Skalska et al., 2015). One possibility is that these β-catenin/Smad1 binding events are due to earlier Wnt/BMP signaling. Indeed ~40% of our stage NF20 β-catenin-bound genes were also reported in β-catenin ChIP-seq from *Xenopus tropicalis* gastrula (NF10.5) (Nakamura et al., 2016). This is consistent with the idea that β-catenin and/or Smad binding can prime genes for future activation, perhaps by modulating epigenetic poising and/or interacting with pioneering factors, which are important for gut tube lineages (Blythe et al., 2010; Loh et al., 2014; Tsankov et al., 2015; Wang et al., 2015)

In summary this study has advanced our understanding of how BMP/Smad1 and Wnt/β-catenin signaling are integrated in the genome to regulate FG and HG transcriptional programs, which should inform hPSC differentiation mechanisms and how Wnt and BMP interact in other development and disease contexts.

MATERIALS AND METHODS

Embryo experiments and manipulations

Animal experiments were performed according to CCHMC IACUC approved protocols. *Xenopus laevis* embryos were staged according to Nieuwkoop and Faber (Nieuwkoop and Faber, 1967). Injections, small molecule treatments, luciferase assays (see supplementary Materials and Methods), *in situ* hybridizations and immunostaining were performed as previously described (McLin et al., 2007). Protein injections into the closing blastocoel of the foregut were performed at stage NF12, with either 40 nl of recombinant human BMP2 (5.8 µM; R&D Systems) in PBS + 0.1%BSA or PBS + 0.1%BSA as control. Embryos were cultured from stages NF12-20 with either DMSO vehicle in 0.1XMBS or DMH1 (40 µM; TOCRIS) or BIO (60 µM; TOCRIS) dissolved in DMSO. Stage NF12 transgenic *Xenopus laevis* *Xla.Tg(hgem:Xtr.dkk1)*^{Jmws}*,* referred to as *Tg(hsp70:dkk1)* (Lin and Slack, 2008), were heatshocked at 37^{0}C for 30 minutes followed by incubation at 13^{0}C to NF20. For immunofluorescence we used anti-phospho-Smad1/5/8 (1:300, Cell Signaling Technology, 13820S) and anti-β-catenin (1:300, Santa Cruz Biotechnologies, sc-7199).

Genomic analysis

RNA/DNA-seq analyses were performed using the *Xenopus laevis* genome v9.1, (Session et al., 2016). Since *Xenopus laevis* is allotetraploid most genes contain two copies, designated as .L or .S (eg. *cdx2.L* and *cdx2.S*). For simplicity we dropped the .L or .S in most figures, however this is reported in all the supplementary gene lists.

For each RNA-seq sample, 50 explants were microdissected and when necessary cultured in 10 µg/ml dispase for 15-20 minutes to separate endo and meso. Total RNA was extracted from two or three independent biological replicates and libraries were sequenced with ~7-10 million reads/library with 75 bp length. Quality trimmed reads were mapped to the *X. laevis* genome 9.1, quantified using RSEM and mapped with bowtie2 using default thresholds (Li and Dewey, 2011). Differential gene expression analysis was carried out with RUVSeq (Risso et al., 2014) with log2FC ≥1 or ≤-1, p <0.05 and FDR ≤5%.

ChIP was carried out as previously described (Akkers et al., 2012; Blythe et al., 2009) with 25-50 whole embryos or 100 FG or HG explants using the following antibodies: anti-Smad1 (Invitrogen, 38-5400), anti-β-catenin (Life technologies, 712700) and anti-p300 (Santa Cruz sc-585 X). Libraries were sequenced with ~30 million reads/library. Reads were mapped to the *X. laevis* genome assembly v9.1 using Bowtie2 at default thresholds (Langmead and Salzberg, 2012). ChIP-seq peaks were called using MACS2 at default thresholds (Zhang et al., 2008). Irreproducibility Discovery Rate (IDR) was performed with standard thresholds (Li et al., 2011). Publicly available data for SMAD1 (GSM1505734), and β-CATENIN (GSM1579346 and GSM1303695) (Estaras et al., 2015; Tsankov et al., 2015; Watanabe et al., 2014) were processed with bowtie and MACS2 using BioWardrobe Toolkit at default thresholds (Kartashov and Barski, 2015). For further details of RNA-seq analysis, ChIP and ChIP-seq analysis, see the supplementary Materials and Methods.

ACKNOWLEDGEMENTS

We specially thank Dr. Ken Cho and Jin Cho for the ChIP-seq training, Dr. Gert Veenstra's lab for ChIP-seq advice and Dr. Matt Weirauch for valuable discussions. We are grateful to the Zorn and Wells labs and Endoderm Club for helpful discussions. We thank Brian Gebelein and Malcolm Fisher for comments on the manuscript.

AUTHORS CONTRIBUTION

MLS and AMZ designed the study, interpreted the data, and wrote the manuscript. MLS, SAR, MM and AMZ performed *Xenopus* experiments. SJ, MY and AB prepared the ChIP-seq libraries. PC and MLS performed the bioinformatics analysis.

FUNDING

This work was funded by NIH DK070858 to AMZ. MLS was supported by NIH T32HL007752-2, and a University of Cincinnati Research Council Summer fellowship. NGS library construction and sequencing was supported by NIGMS DP2 GM119134 and NCATS 5UL1TR001425-02 to AB. This project was supported in part by NIH P30 DK078392.

DATA AVAILABILITY

All RNA-seq and ChIP-seq datasets from this study are available in NCBI Gene Expression Omnibus (GEO) under accession number GSE87654.

COMPETING INTERESTS

A.B. is co-founders of Datirium, LLC. Which provides software development and bioinformatics support services, including installation of BioWardrobe.

REFERENCES

Akkers, R. C., Jacobi, U. G. and Veenstra, G. J. (2012). Chromatin immunoprecipitation analysis of Xenopus embryos. *Methods Mol Biol* **917**, 279-292.

Benahmed, F., Gross, I., Gaunt, S. J., Beck, F., Jehan, F., Domon-Dell, C., Martin, E., Kedinger, M., Freund, J. N. and Duluc, I. (2008). Multiple regulatory regions control the complex expression pattern of the mouse Cdx2 homeobox gene. *Gastroenterology* **135**, 1238-1247.

Benchabane, H. and Wrana, J. L. (2003). GATA- and Smad1-dependent enhancers in the Smad7 gene differentially interpret bone morphogenetic protein concentrations. *Mol Cell Biol* **23**, 6646-6661.

Blitz, I. L. and Cho, K. W. (2009). Finding partners: how BMPs select their targets. *Dev Dyn* **238**, 1321-1331.

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.

Blythe, S. A., Reid, C. D., Kessler, D. S. and Klein, P. S. (2009). Chromatin immunoprecipitation in early Xenopus laevis embryos. *Dev Dyn* **238**, 1422-1432.

Brown, C. O., 3rd, Chi, X., Garcia-Gras, E., Shirai, M., Feng, X. H. and Schwartz, R. J. (2004). The cardiac determination factor, Nkx2-5, is activated by mutual cofactors GATA-4 and Smad1/4 via a novel upstream enhancer. *J Biol Chem* **279**, 10659-10669.

Cadigan, K. M. and Waterman, M. L. (2012). TCF/LEFs and Wnt signaling in the nucleus. *Cold Spring Harb Perspect Biol* **4,** 1-22.

Chalmers, A. D. and Slack, J. M. (2000). The Xenopus tadpole gut: fate maps and morphogenetic movements. *Development* **127**, 381-392.

De Robertis, E. M. (2009). Spemann's organizer and the self-regulation of embryonic fields. *Mech Dev* **126**, 925-941.

Ding, V. W., Chen, R. H. and McCormick, F. (2000). Differential regulation of glycogen synthase kinase 3beta by insulin and Wnt signaling. *J Biol Chem* **275**, 32475-32481.

Estaras, C., Benner, C. and Jones, K. A. (2015). SMADs and YAP compete to control elongation of beta-catenin:LEF-1-recruited RNAPII during hESC differentiation. *Mol Cell* **58**, 780-793.

Gaarenstroom, T. and Hill, C. S. (2014). TGF-beta signaling to chromatin: how Smads regulate transcription during self-renewal and differentiation. *Semin Cell Dev Biol* **32**, 107-118.

Gaunt, S. J., Drage, D. and Cockley, A. (2003). Vertebrate caudal gene expression gradients investigated by use of chick cdx-A/lacZ and mouse cdx-1/lacZ reporters in transgenic mouse embryos: evidence for an intron enhancer. *Mech Dev* **120**, 573-586.

Gibb, N., Lavery, D. L. and Hoppler, S. (2013). sfrp1 promotes cardiomyocyte differentiation in Xenopus via negative-feedback regulation of Wnt signalling. *Development* **140**, 1537-1549.

Green, M. D., Chen, A., Nostro, M. C., d'Souza, S. L., Schaniel, C., Lemischka, I. R., Gouon-Evans, V., Keller, G. and Snoeck, H. W. (2011). Generation of

anterior foregut endoderm from human embryonic and induced pluripotent stem cells. *Nat Biotechnol* **29**, 267-272.

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.

Haworth, K. E., Kotecha, S., Mohun, T. J. and Latinkic, B. V. (2008). GATA4 and GATA5 are essential for heart and liver development in Xenopus embryos. *BMC Dev Biol* **8**, 74.

Hoppler, S. and Moon, R. T. (1998). BMP-2/-4 and Wnt-8 cooperatively pattern the Xenopus mesoderm. *Mech Dev* **71**, 119-129.

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

Jamora, C., DasGupta, R., Kocieniewski, P. and Fuchs, E. (2003). Links between signal transduction, transcription and adhesion in epithelial bud development. *Nature* **422**, 317-322.

Karaulanov, E., Knochel, W. and Niehrs, C. (2004). Transcriptional regulation of BMP4 synexpression in transgenic Xenopus. *EMBO J* **23**, 844-856.

Kartashov, A. V. and Barski, A. (2015). BioWardrobe: an integrated platform for analysis of epigenomics and transcriptomics data. *Genome Biol* **16**, 158.

Kenny, A. P., Rankin, S. A., Allbee, A. W., Prewitt, A. R., Zhang, Z., Tabangin, M. E., Shifley, E. T., Louza, M. P. and Zorn, A. M. (2012). Sizzled-tolloid interactions maintain foregut progenitors by regulating fibronectin-dependent BMP signaling. *Dev Cell* **23**, 292-304.

Kim, D. W. and Lassar, A. B. (2003). Smad-dependent recruitment of a histone deacetylase/Sin3A complex modulates the bone morphogenetic protein-dependent transcriptional repressor activity of Nkx3.2. *Mol Cell Biol* **23**, 8704-8717.

Kim, S. W., Yoon, S. J., Chuong, E., Oyolu, C., Wills, A. E., Gupta, R. and Baker, J. (2011). Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev Biol* **357**, 492-504.

Kjolby, R. A. and Harland, R. M. (2016). Genome-wide identification of Wnt/beta-catenin transcriptional targets during Xenopus gastrulation. *Dev Biol*.

Klaus, A. and Birchmeier, W. (2009). Developmental signaling in myocardial progenitor cells: a comprehensive view of Bmp- and Wnt/beta-catenin signaling. *Pediatr Cardiol* **30**, 609-616.

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.

Lancaster, M. A. and Knoblich, J. A. (2014). Organogenesis in a dish: modeling development and disease using organoid technologies. *Science* **345**, 1247125.

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, Q., Brown, J. B., Huang, H. and Bickel, P. J. (2011). Measuring reproducibility of high-throughput experiments. 1752-1779.

Li, Y., Rankin, S. A., Sinner, D., Kenny, A. P., Krieg, P. A. and Zorn, A. M. (2008). Sfrp5 coordinates foregut specification and morphogenesis by antagonizing both canonical and noncanonical Wnt11 signaling. *Genes Dev* **22**, 3050-3063.

Lien, C. L., McAnally, J., Richardson, J. A. and Olson, E. N. (2002). Cardiac-specific activity of an Nkx2-5 enhancer requires an evolutionarily conserved Smad binding site. *Dev Biol* **244**, 257-266.
Lin, G. and Slack, J. M. (2008). Requirement for Wnt and FGF signaling in Xenopus tadpole tail regeneration. *Dev Biol* **316**, 323-335.
Loh, K. M., Ang, L. T., Zhang, J., Kumar, V., Ang, J., Auyeong, J. Q., Lee, K. L., Choo, S. H., Lim, C. Y., Nichane, M., et al. (2014). Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. *Cell Stem Cell* **14**, 237-252.
Loh, K. M., Chen, A., Koh, P. W., Deng, T. Z., Sinha, R., Tsai, J. M., Barkal, A. A., Shen, K. Y., Jain, R., Morganti, R. M., et al. (2016). Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. *Cell* **166**, 451-467.
Marty, T., Muller, B., Basler, K. and Affolter, M. (2000). Schnurri mediates Dpp-dependent repression of brinker transcription. *Nat Cell Biol* **2**, 745-749.
McLin, V. A., Rankin, S. A. and Zorn, A. M. (2007). Repression of Wnt/beta-catenin signaling in the anterior endoderm is essential for liver and pancreas development. *Development* **134**, 2207-2217.
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.
Nieuwkoop, P. D. and Faber, J. (1967). Normal Table of Xenopus laevis (Daudin): A Systematical and Chronological Survey of the Development from the Fertilized Egg to the End of Metamorphosis.: New York: Garland Publishing.
Olson, L. E., Tollkuhn, J., Scafoglio, C., Krones, A., Zhang, J., Ohgi, K. A., Wu, W., Taketo, M. M., Kemler, R., Grosschedl, R., et al. (2006). Homeodomain-mediated beta-catenin-dependent switching events dictate cell-lineage determination. *Cell* **125**, 593-605.
Rankin, S. A., Kormish, J., Kofron, M., Jegga, A. and Zorn, A. M. (2011). A gene regulatory network controlling hhex transcription in the anterior endoderm of the organizer. *Dev Biol* **351**, 297-310.
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.
Rossi, J. M., Dunn, N. R., Hogan, B. L. and Zaret, K. S. (2001). Distinct mesodermal signals, including BMPs from the septum transversum mesenchyme, are required in combination for hepatogenesis from the endoderm. *Genes Dev* **15**, 1998-2009.
Schohl, A. and Fagotto, F. (2002). Beta-catenin, MAPK and Smad signaling during early Xenopus development. *Development* **129**, 37-52.
Session, A. M., Uno, Y., Kwon, T., Chapman, J. A., Toyoda, A., Takahashi, S., Fukui, A., Hikosaka, A., Suzuki, A., Kondo, M., et al. (2016). Genome evolution in the allotetraploid frog Xenopus laevis. *Nature* **538**, 336-343.
Sherwood, R. I., Maehr, R., Mazzoni, E. O. and Melton, D. A. (2011). Wnt signaling specifies and patterns intestinal endoderm. *Mech Dev* **128**, 387-400.
Skalska, L., Stojnic, R., Li, J., Fischer, B., Cerda-Moya, G., Sakai, H., Tajbakhsh, S., Russell, S., Adryan, B. and Bray, S. J. (2015). Chromatin signatures at Notch-regulated enhancers reveal large-scale changes in H3K56ac upon activation. *EMBO J* **34**, 1889-1904.

Spence, J. R., Mayhew, C. N., Rankin, S. A., Kuhar, M. F., Vallance, J. E., Tolle, K., Hoskins, E. E., Kalinichenko, V. V., Wells, S. I., Zorn, A. M., et al. (2011). Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. *Nature* **470**, 105-109.

Theisen, H., Syed, A., Nguyen, B. T., Lukacsovich, T., Purcell, J., Srivastava, G. P., Iron, D., Gaudenz, K., Nie, Q., Wan, F. Y., et al. (2007). Wingless directly represses DPP morphogen expression via an armadillo/TCF/Brinker complex. *PLoS One* **2**, e142.

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.

Wang, A., Yue, F., Li, Y., Xie, R., Harper, T., Patel, N. A., Muth, K., Palmer, J., Qiu, Y., Wang, J., et al. (2015). Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell* **16**, 386-399.

Watanabe, K., Biesinger, J., Salmans, M. L., Roberts, B. S., Arthur, W. T., Cleary, M., Andersen, B., Xie, X. and Dai, X. (2014). Integrative ChIP-seq/microarray analysis identifies a CTNNB1 target signature enriched in intestinal stem cells and colon cancer. *PLoS One* **9**, e92317.

Wu, D. and Pan, W. (2010). GSK3: a multifaceted kinase in Wnt signaling. *Trends Biochem Sci* **35**, 161-168.

Xuan, S., Borok, M. J., Decker, K. J., Battle, M. A., Duncan, S. A., Hale, M. A., Macdonald, R. J. and Sussel, L. (2012). Pancreas-specific deletion of mouse Gata4 and Gata6 causes pancreatic agenesis. *J Clin Invest* **122**, 3516-3528.

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.

Zhang, Z., Rankin, S. A. and Zorn, A. M. (2013). Different thresholds of Wnt-Frizzled 7 signaling coordinate proliferation, morphogenesis and fate of endoderm progenitor cells. *Dev Biol* **378**, 1-12.

Zorn, A. M. and Wells, J. M. (2009). Vertebrate endoderm development and organ formation. *Annu Rev Cell Dev Biol* **25**, 221-251.

Figures

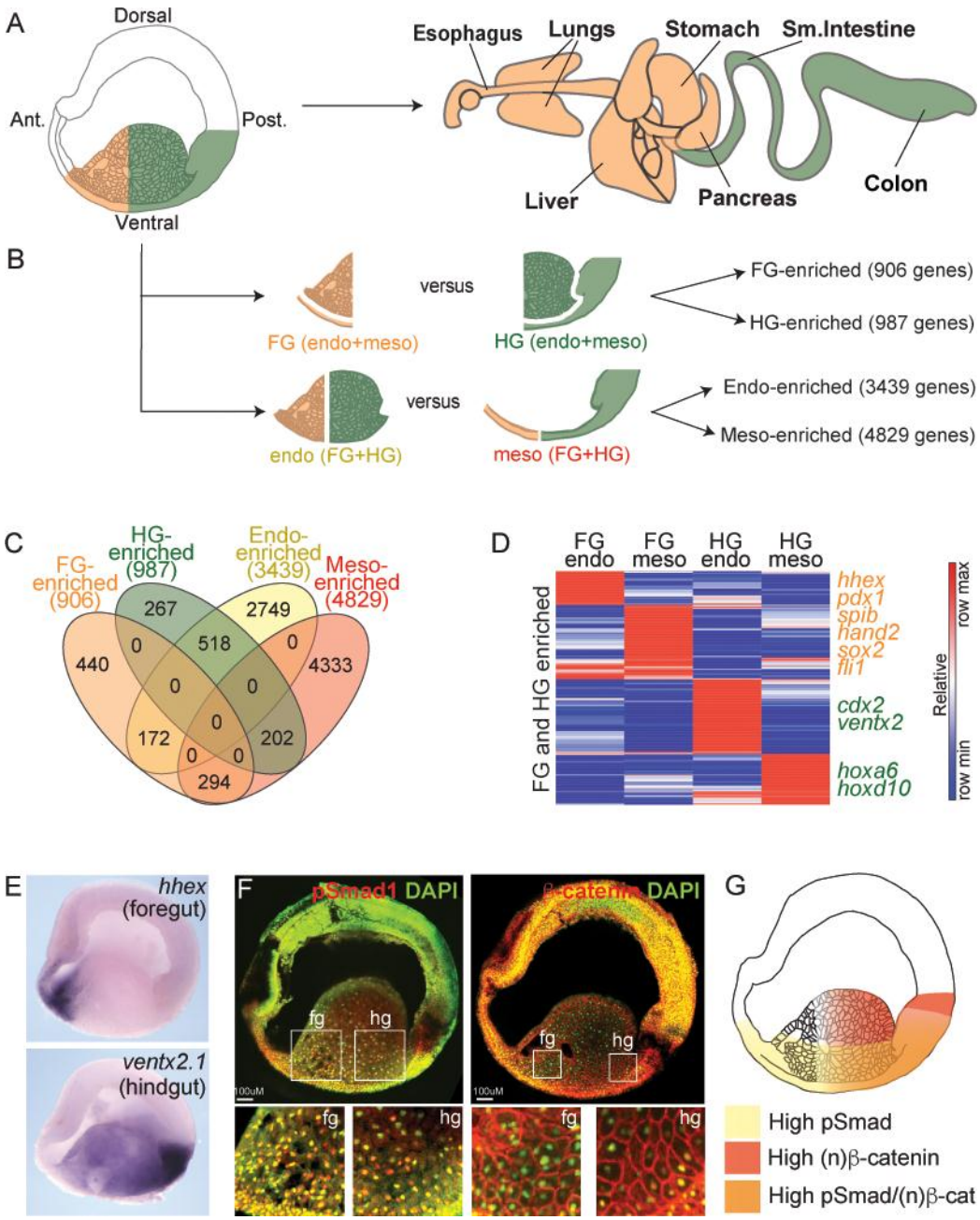


Fig. 1: Transcriptional program of FG and HG progenitors correlates with differential BMP and Wnt signaling.
(A) Fate map showing that FG progenitors (yellow) give rise to lungs, liver, pancreas and stomach, whereas HG progenitors (green) give rise to intestine (Chalmers and Slack, 2000). (**B**) Experimental design. RNA-seq was performed on FG-endo, FG-meso, HG-endo and HG-meso explants dissected from stage NF20 embryos. Differentially expressed transcripts were identified by pairwise comparisons of FG (endo and meso) versus HG (endo and meso), as well as endo (FG and HG) versus meso (FG and HG) tissue (log2 fold change (FC) ≤-1 or ≥1, false discovery rate (FDR) ≤5%). **(C)** Venn diagram showing the intersection of two separate differential expression analysis: FG versus HG and endo versus meso, showing mutually exclusive lists of transcripts with enriched expression. **(D)** Heatmap

clustering of the 906 FG-enriched and 987 HG-enriched genes showing expression in the indicated tissues with representative FG (orange) and HG (green) genes listed on the right. **(E)** *In situ* hybridization of sagittal bisected stage NF20 embryos with *hhex* and *ventx2.1* marking FG and HG domains, respectively. **(F)** BMP and Wnt activity shown by pSmad1 (red) and nuclear (n)β-catenin (red) immunostaining in NF20 embryos. Nuclei staining with DAPI (green) **(G)** pSmad1 is high in the ventral and low in the dorsal FG and HG, whereas (n)β-catenin is low in the FG and high in the HG.

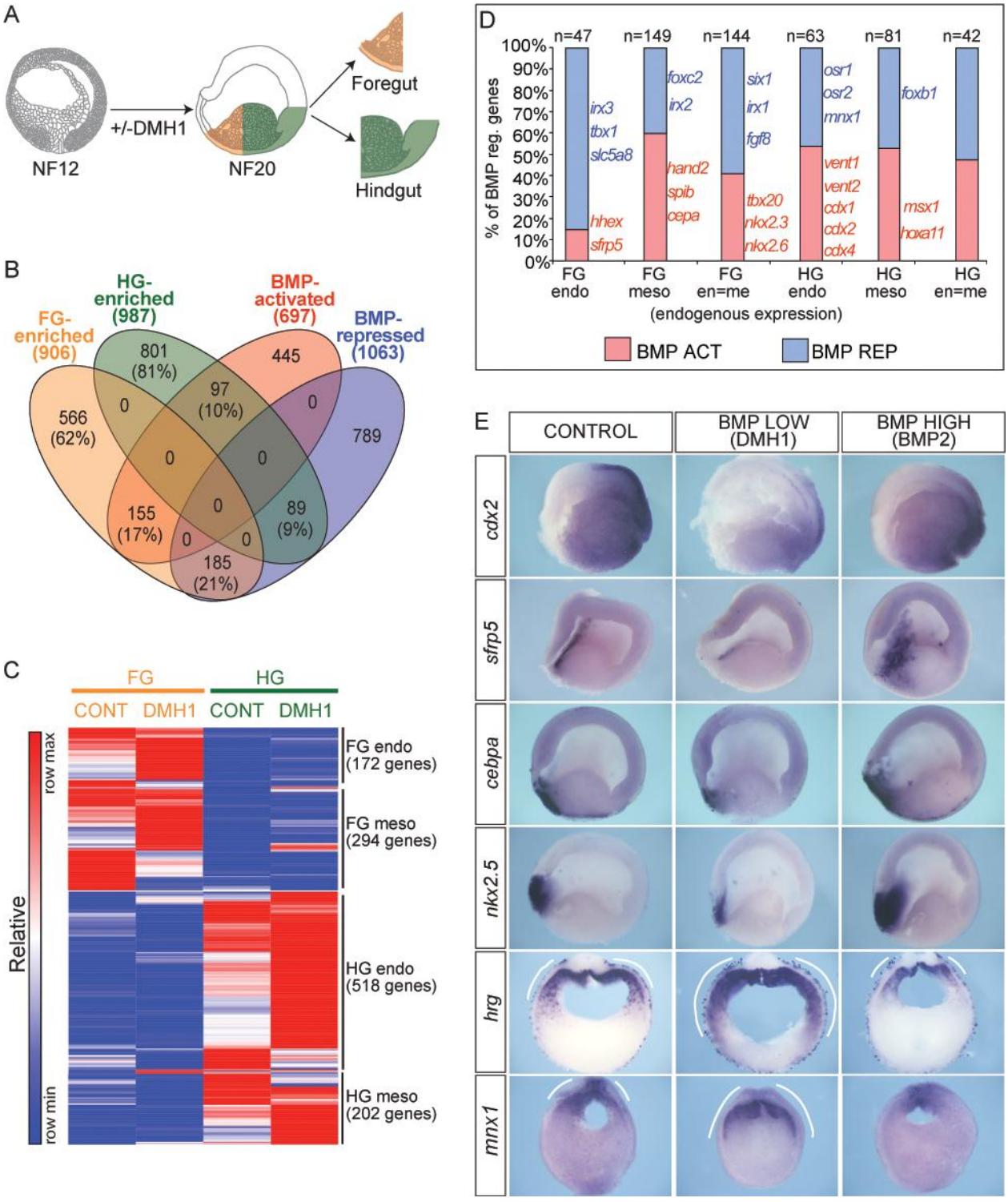


Fig. 2: BMP signaling coordinates D-V patterning.
(A) Experimental design. FG and HG (endo + meso) explants were dissected from DMSO and DMH1 treated NF20 embryos and submitted for RNA-seq (in triplicate). **(B)** Venn diagram showing overlap of FG-enriched genes, HG-enriched genes and BMP-activated and repressed transcripts (log2FC ≥1, FDR ≤5%); see Fig. S2B for details. **(C)** Expression heatmap clustering of FG-endo, FG-meso, HG-endo and HG-meso enriched transcripts in vehicle and DMH1-treated explants. **(D)** Different categories of BMP activated (ACT) or repressed (REP) transcripts based on whether the gene is normally enriched in the endo, meso or expressed in both endo and meso (en=me). **(E)** *In situ* hybridization of vehicle

control, DMH1-treated or BMP2 injected stage NF20 embryos in mid-sagittal section (*cdx2, sfrp5, cebpa* and *nkx2-5;* anterior left and dorsal up) or cross section (*mnx1* and *hrg;* dorsal up, white lines indicate expression domain), n>20 for each probe.

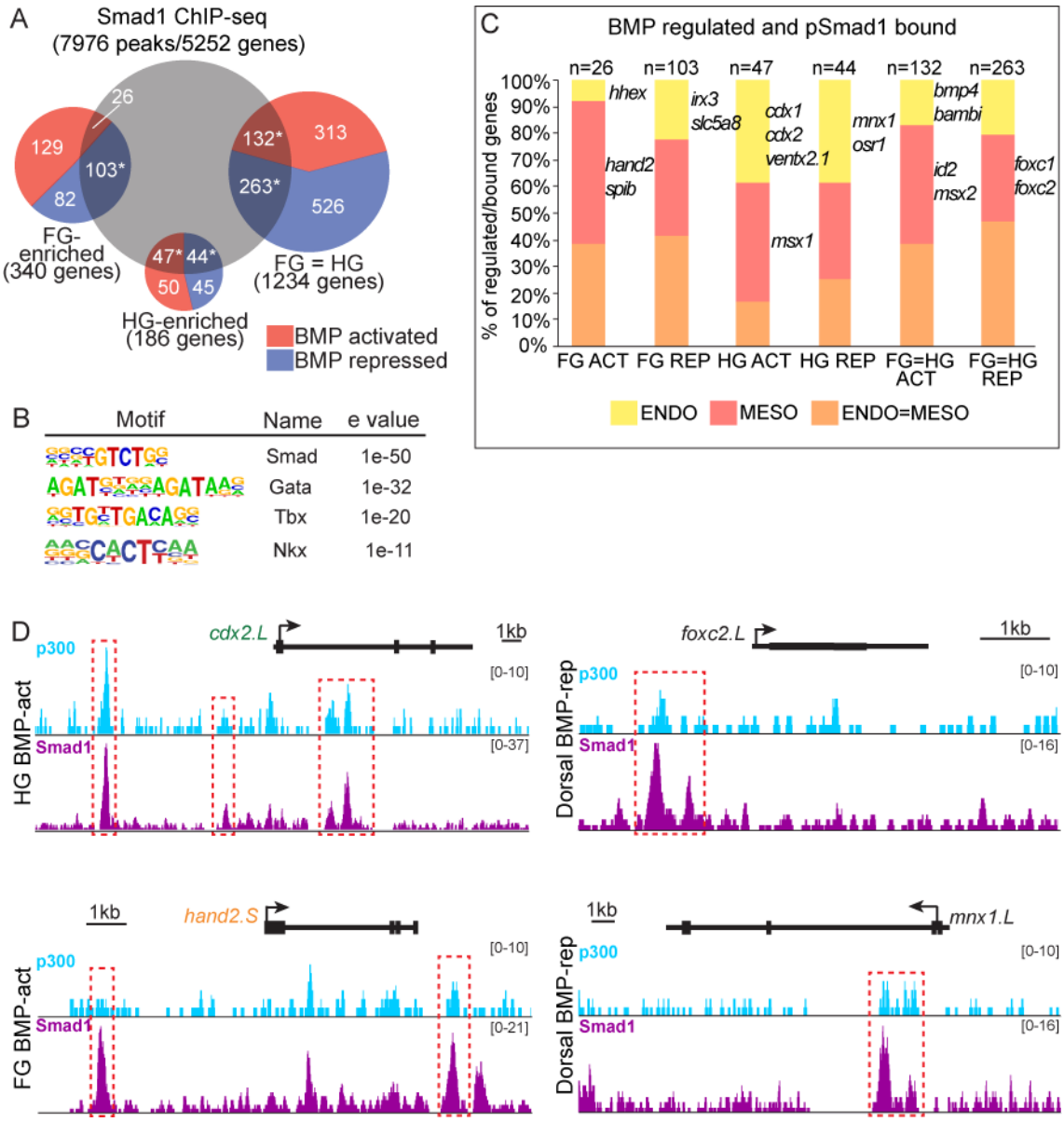


Fig. 3: Smad1 chromatin binding to BMP-regulated genes.
(A) Smad1 ChIP-seq analysis of NF20 whole embryos identified 7976 peaks within +/- 20 Kb of 5252 genes. The Venn diagram intersecting 5252 Smad1-bound genes with the 1760 BMP-regulated genes (from Fig. 2B) grouped as FG-enriched (n=340), HG-enriched (n=186) or expressed at similar levels in the FG and HG (FG=HG; n=1234), identifies 615 BMP-regulated genes associated with Smad1-binding, BMP-activated genes in red and BMP-repressed in blue. *Statistically enriched based on hypergeometric test ($p<0.05$). **(B)** DNA-binding protein motif enrichment analysis of 7976 Smad1 peaks. **(C)** Chart illustrates Smad1-bound genes within the different categories of BMP-activated (ACT) and repressed (REP) genes. **(D)** Genome browser view of Smad1 peaks on BMP-activated HG gene *cdx2,* and ventral mesoderm *hand2*, as well as BMP-repressed dorsal genes *foxc2* and *mnx1*. Red boxes indicate statistically significant Smad1 peaks.

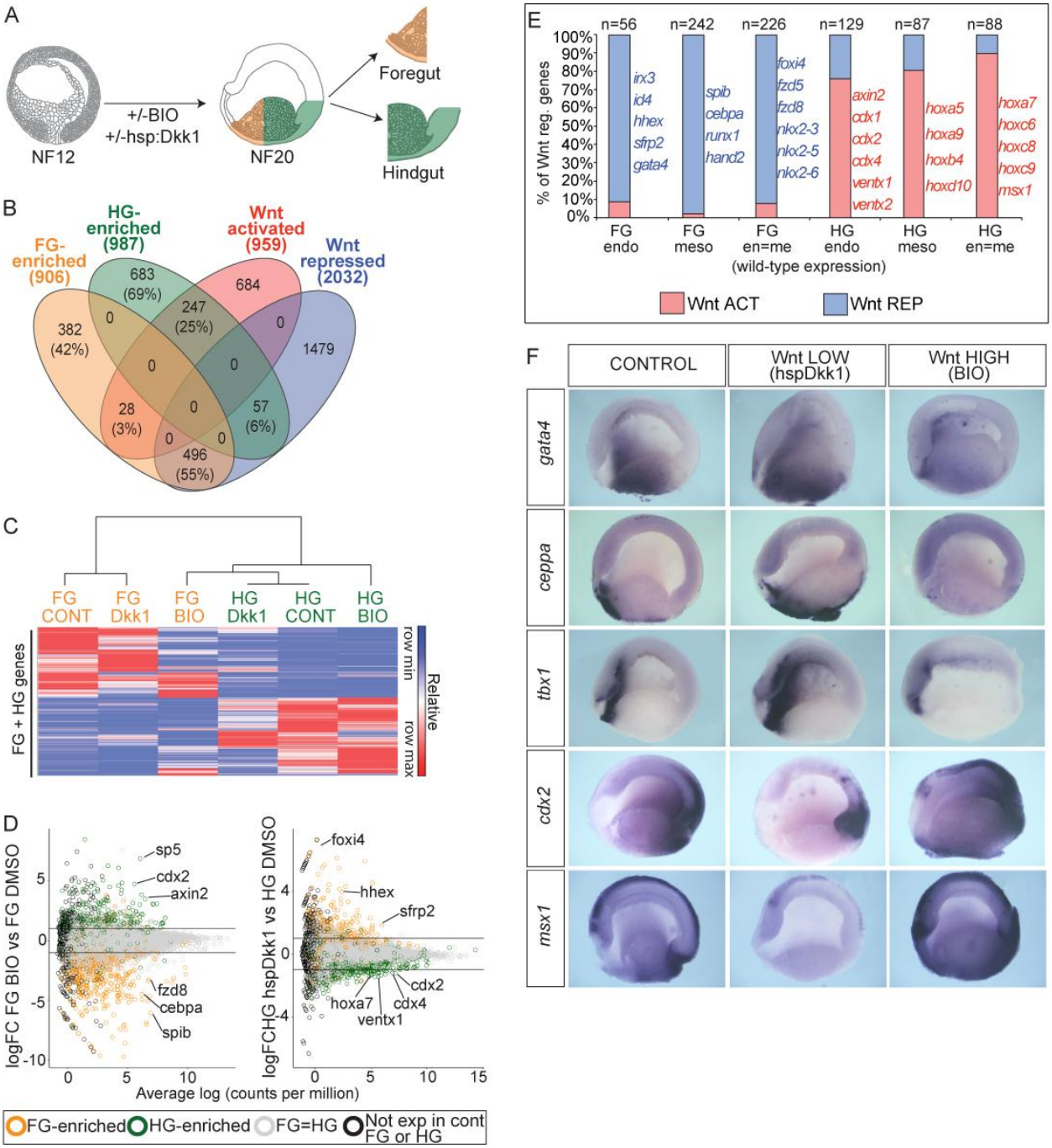


Fig. 4: Wnt signaling promotes HG transcriptional program and represses FG transcriptional program.

(A) Experimental design. FG and HG (endo + meso) explants were dissected from DMSO, BIO and *Tg(hsp70:dkk1)* NF20 embryos and submitted for RNA-seq (in triplicates). **(B)** Venn diagram illustrates overlap of FG-enriched, HG-enriched, Wnt-activated and Wnt-repressed genes (log2FC ≤-1 or ≥1, FDR ≤5%); see Fig. S4A for details. **(C)** Unsupervised clustering of FG- and HG-enriched genes, showing that BIO-treated FG has an expression profile similar to HG control. **(D)** Scatter plot showing log2FC in expression between control, BIO and *Tg(hsp70:dkk1)* FG and HG explants. Transcripts are colored based on the normal control expression; HG-enriched in green, FG-enriched in orange, transcripts expressed similarly FG and HG (FG=HG) in grey and normally not expressed in FG or HG in black. **(E)** Different categories of BMP activated (ACT) or repressed (REP) transcripts group based on whether

the gene is normally enriched in the endo, meso or expressed in both endo and meso (en=me). **(F)** *In situ* hybridization of Control, *Tg(hsp70:dkk1)* or BIO treated embryos in mid-sagittal section, anterior left and dorsal up. Anterior genes are *gata4*, *cebpa* and *tbx1* and posterior genes are *cdx2,* and *msx1*, n>20 for each probe.

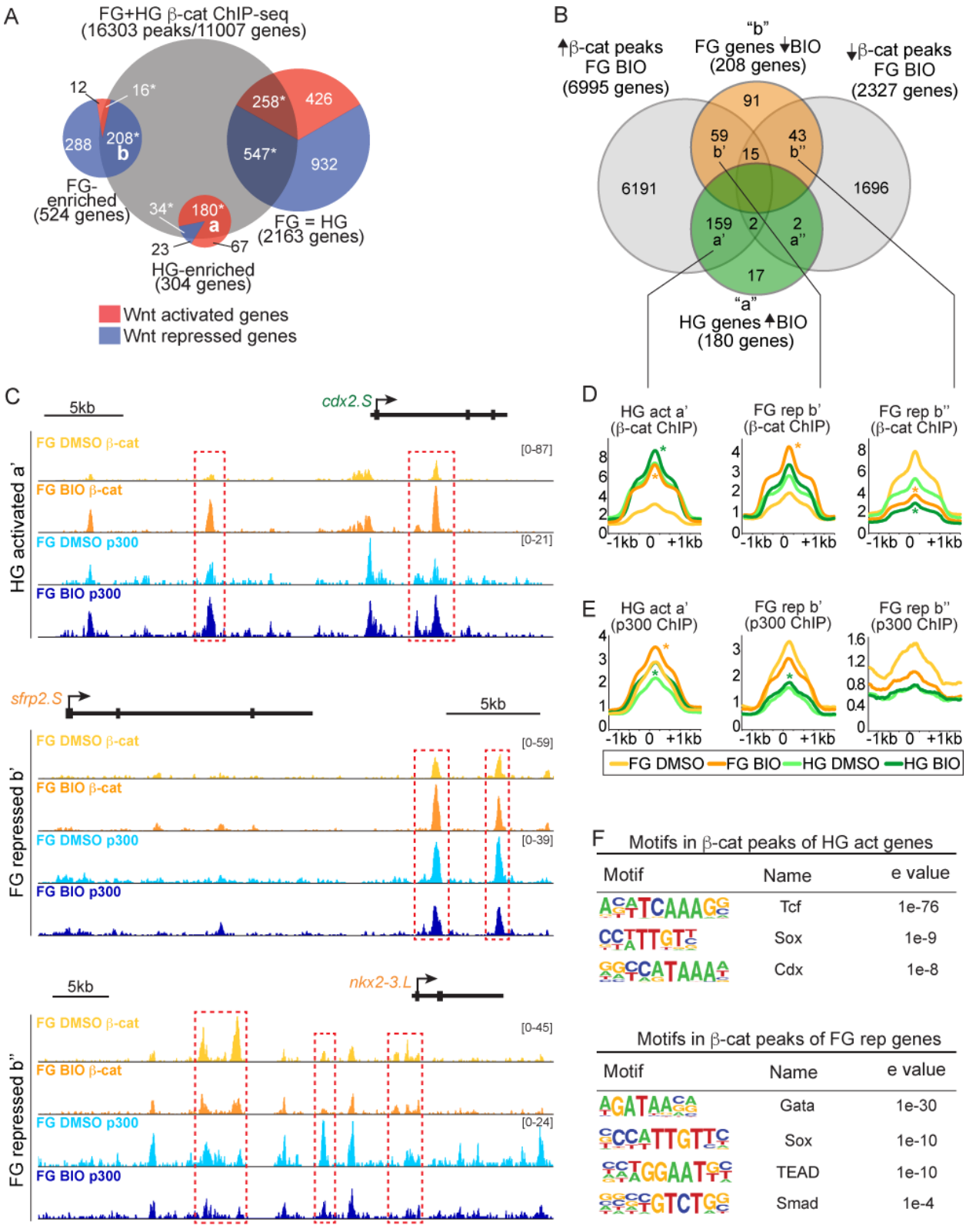


Fig. 5: β-catenin chromatin binding to Wnt-regulated genes.

(A) β-catenin ChIP-seq of FG and HG explants identified 16303 peaks within +/- 20kb of 11007 genes. The Venn diagram shows the overlap between 11007 β-catenin-bound genes and 2991 Wnt-regulated genes (from Fig. 4B) grouped as FG-enriched (n=524), HG-enriched (n=304) or expressed at similar levels in the FG and HG (FG=HG; n=2163) based on RNA-seq data. BMP-activated genes in red and BMP-repressed in blue. *Statistically

enriched based on hypergeometric test ($p<0.05$). **(B)** Overlap between 180 HG Wnt-activated genes (a), 208 FG Wnt-repressed genes (b), and genes with gain or loss of β-catenin ChIP enrichment upon BIO treatment. **(C)** Genome browser view of β-catenin peaks on HG-activated gene *cdx2* and FG-repressed genes *sfrp2* and *nkx2-3*. Red boxes indicate β-catenin significant peaks. **(D-E)** Average tag density of β-catenin **(D)** and p300 **(E)** peaks on HG Wnt-activated genes (a'), and two classes of FG Wnt-repressed genes (b' and b") comparing control to BIO treatment. *$p<0.05$, Wilcoxon test. **(F)** Motif enrichment analysis of β-catenin peaks.

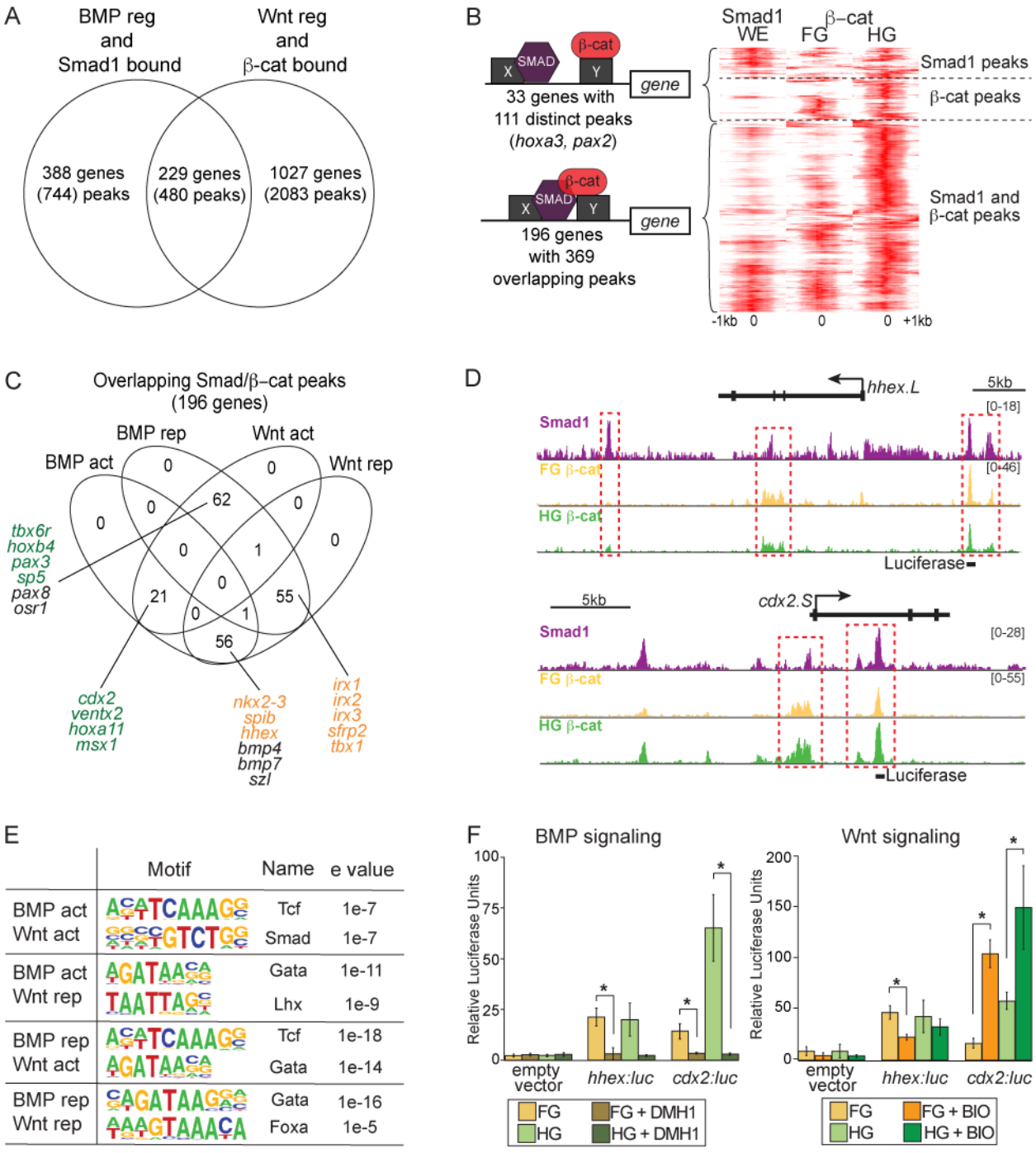


Fig. 6: Smad1 and β-catenin converge on common CRMs.
(A) Overlap of BMP-regulated/Smad1-bound genes with Wnt-regulated/β-catenin bound genes. **(B)** Schematic of 33 genes with distinct β-catenin and Smad1 peaks and 196 genes with overlapping β-catenin and Smad1 peaks. Right panel shows Smad1 and β-catenin read density in the corresponding peaks. FG genes (yellow) and HG genes (green). (WE = Whole Embryo). **(C)** Venn shows 196 genes with overlapping Smad1 and β-catenin peaks categorized based on activation (act) or repression (rep) by BMP and Wnt signaling. **(D)** Browser view of β-catenin and Smad1 peaks on *hhex* and *cdx2* with illustration of luciferase constructs. Red boxes indicate overlapping Smad1 and β-catenin peaks. **(E)** Motif enrichment analysis of gene sets based on BMP/Wnt regulation. **(F)** Luciferase assays of reporter constructs with CRMs depicted in panel D. FG cells were injected in the C1

blastomere and HG cells in the C4 blastomere at 32-cell stage embryos. Error bars represent standard deviation of three biological replicates, * $p<0.05$ in student T-test. .

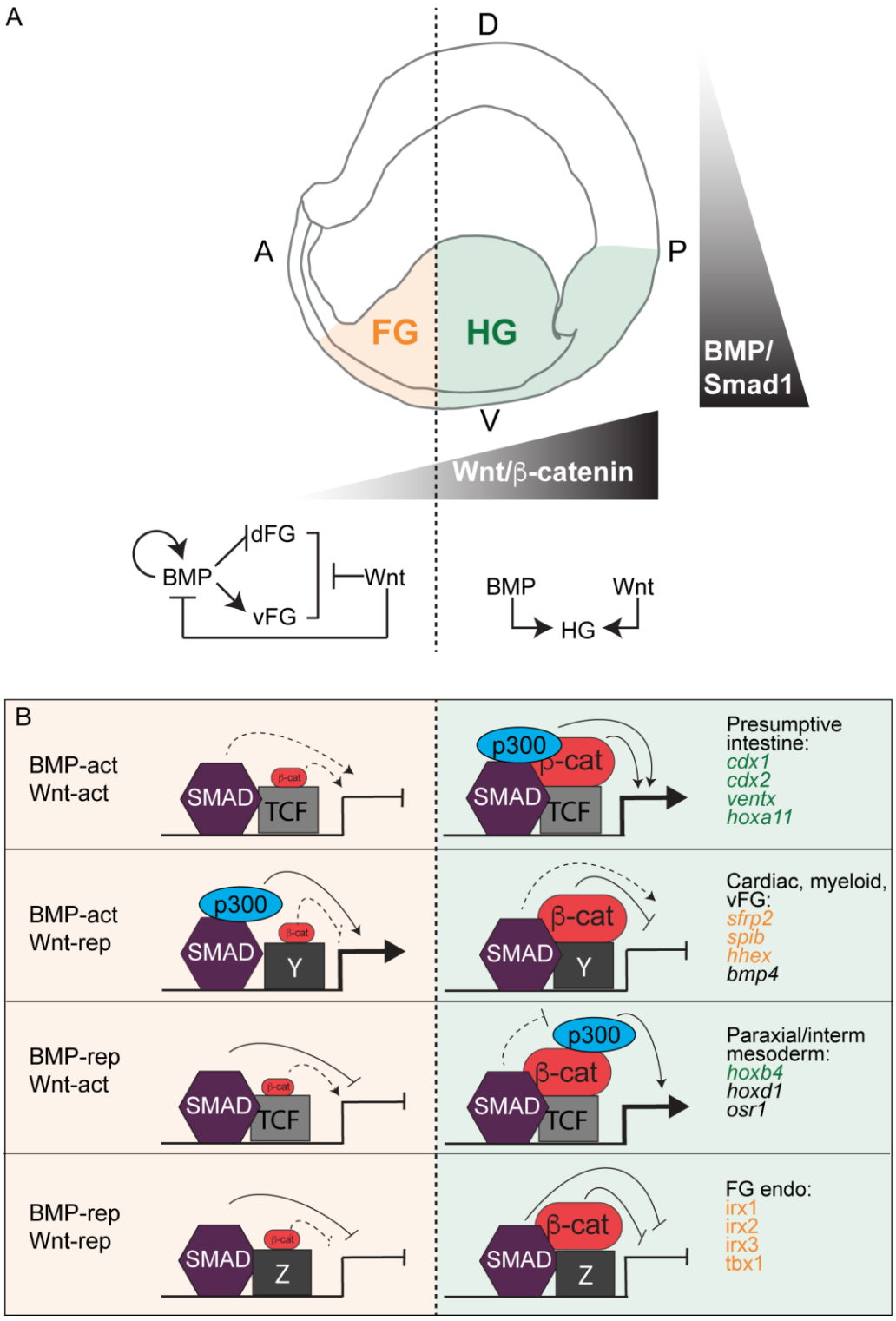


Fig. 7: BMP/Smad1 and Wnt/β-catenin converge on the same FG and HG CRMs.
(A) A model of how spatially restricted Wnt and BMP activity coordinate A-P and D-V patterning of FG and HG progenitors. A signaling crosstalk in the FG with low Wnt promoting BMP ligand expression in the pre-cardiac mesoderm. **(B)** Schematic of how overlapping Smad1 and β-catenin peaks regulate FG and HG transcription.

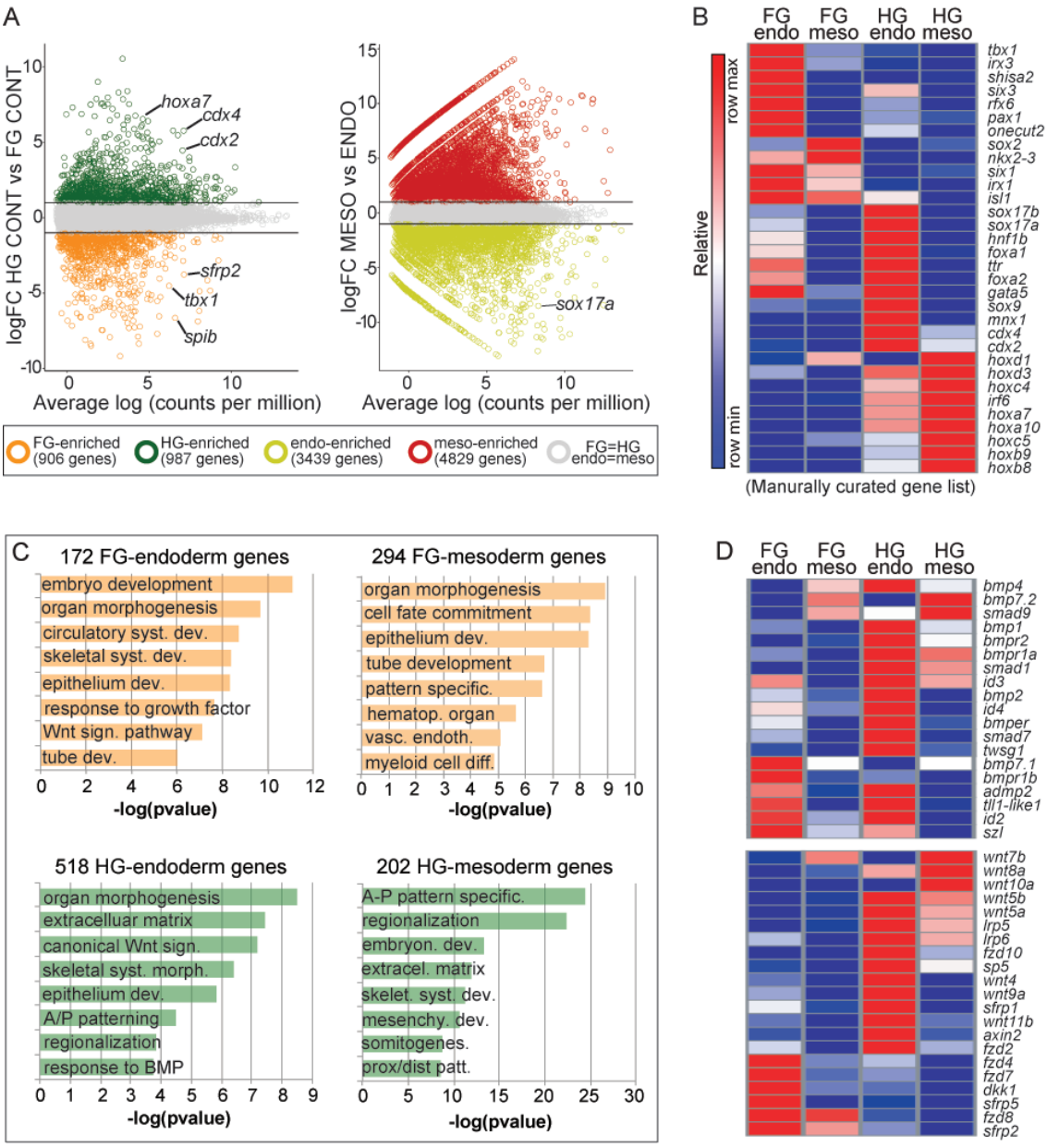


Fig. S1: Transcriptional program of FG and HG progenitors in vivo.

Fig. S1: Transcriptional program of FG and HG progenitors *in vivo*.
(A) Scatter plot of log fold change (log2FC) in expression between HG versus FG samples and endo versus meso samples. HG-enriched transcripts (green), FG-enriched transcripts (orange) meso-enriched transcripts (red) and endo-enriched transcripts (yellow) based on log2 fold change (FC) ≤-1 or ≥1 and false discovery rate (FDR) ≤5%. **(B)** Xenopus orthologs of genes known to be involved in human and mouse GI development (manually curated list from the literature) are present in our FG-enriched and HG-enriched gene lists. The heatmap shows that the Xenopus transcripts have restricted expression in manner predicted from the mouse and human literature, illustrating high conservation across species. **(C)** GO term enrichment analysis of 172 FG-endo, 294 FG-meso, 518 HG-endo and 202 HG-meso genes from Fig. 1C. **(D)** BMP and Wnt pathway components that are expressed in any sample (FG-endo, FG-meso, HG-endo or HG-meso) above one transcripts per-million reads (TPM >1; lower than this is considered not expressed). The heatmap shows that BMP pathway genes are expressed in both the FG and HG, whereas Wnt ligands are generally restricted to the HG and Wnt-antagonists enriched in the FG. **(E)** *In situ* hybridization of mid-sagittal section stage NF20 (*hhex* and *ventx2.1)* or NF35 embryos (*nr1h5*, *nkx2-1*, *nkx2-5*, *sox2* and *darmin)* in DMH1 or *Tg(hsp70:dkk1)* embryos; anterior left and dorsal up.

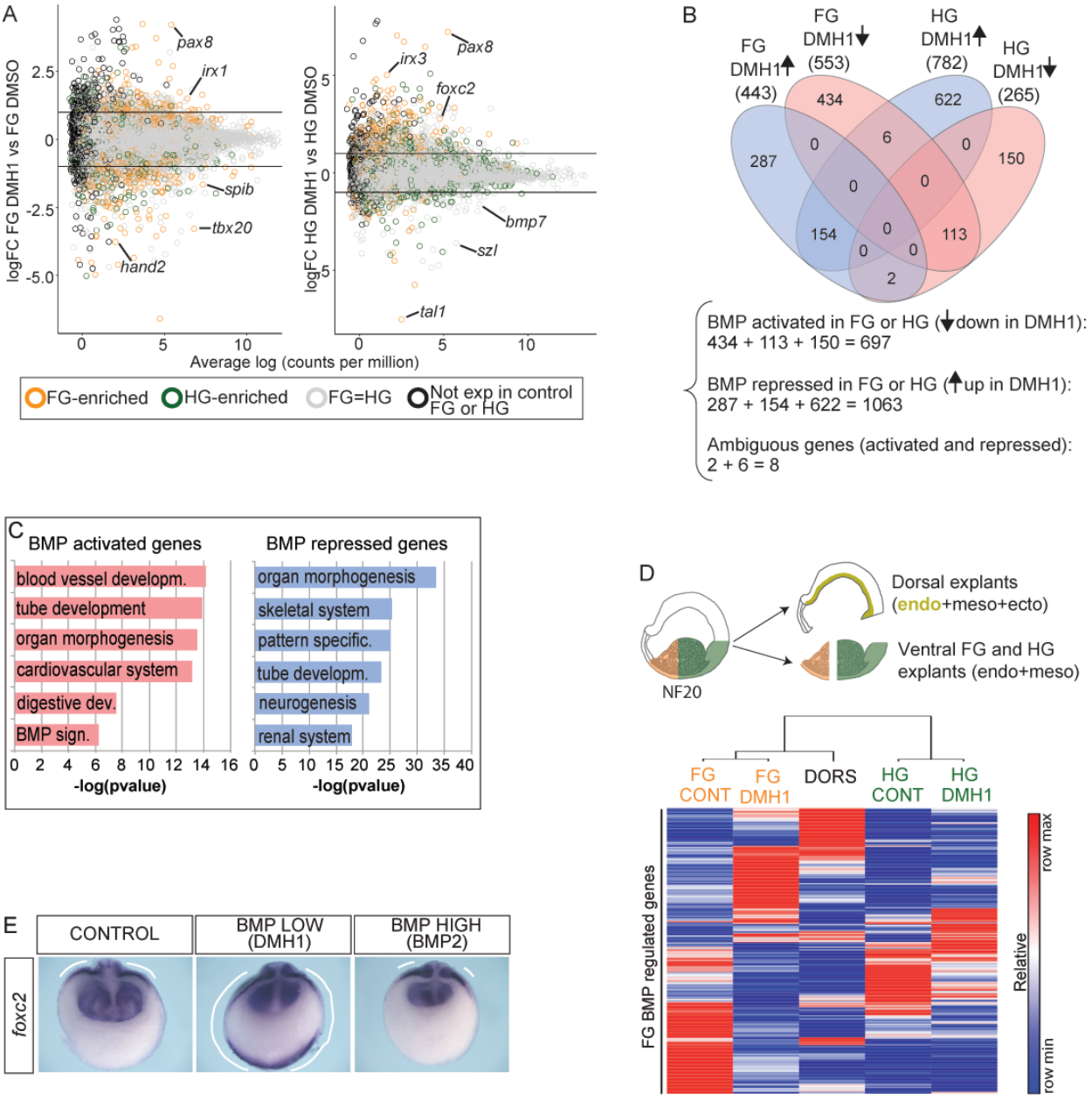
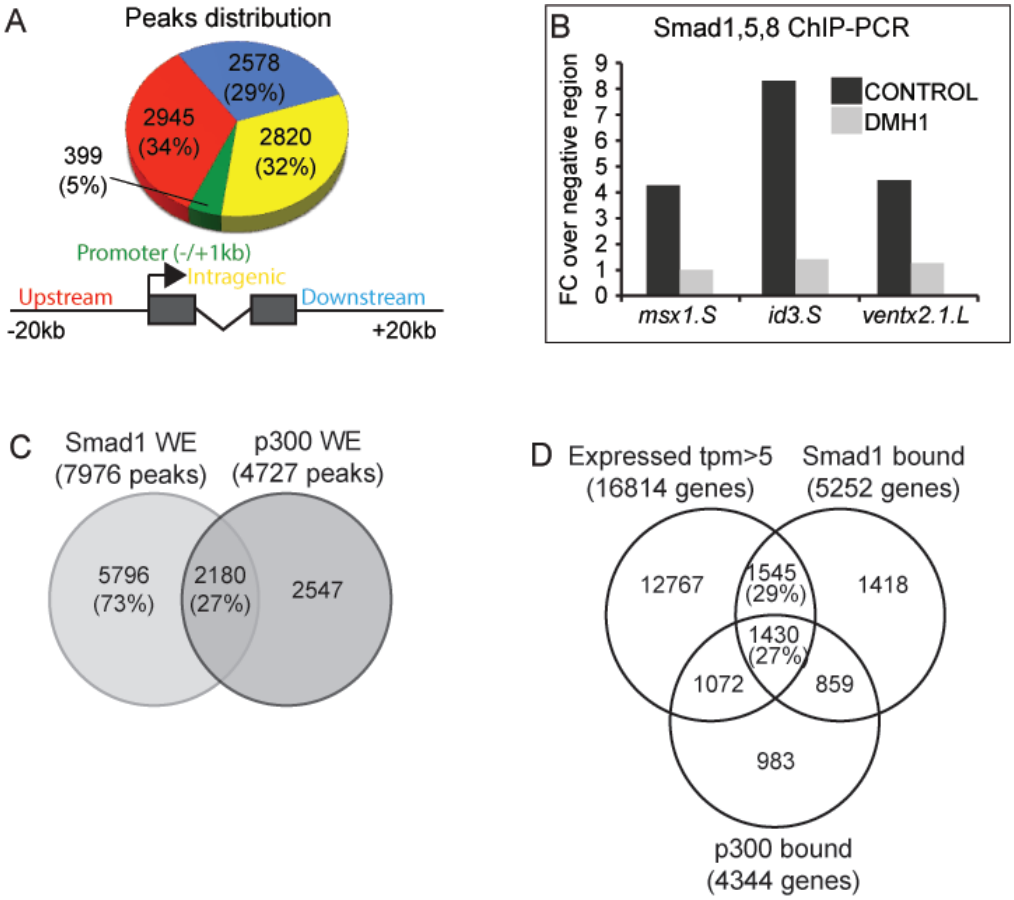


Fig. S2: RNA-seq of DMSO and DMH1 treated embryos identified BMP-regulated genes.
(A) Scatter plot showing log2FC in expression between DMSO and DMH1 treated FG (left) and between DMSO and DMH1 treated HG (right) samples. Transcripts are colored based on the normal control expression; HG-enriched in green, FG-enriched in orange, genes expressed similarly in FG and HG (FG=HG) in grey and normally not expressed in control FG or HG in black. **(B)** Venn diagram illustrates overlap between transcripts up regulated (↑) or down regulated (↓) upon DMH1 treatment in FG and HG tissues. FG or HG transcripts repressed upon DMH1 treatment (log2FC ≤-1, FDR ≤5% relative to controls) are considered to be BMP-activated genes (n=697), whereas FG or HG transcripts that are increased upon DMH1 (log2FC ≥1, FDR ≤5%) are classified as BMP-repressed genes (n=1063). Eight transcripts had ambiguous regulation being both activated and repressed by DMH1 in FG or HG tissues, and were excluded from further analysis. Overall we categorized a total of 1760 (697+1063) BMP-regulated genes in the FG and HG tissue. **(C)** GO term analysis of BMP-activated and -repressed genes. **(D)** Unsupervised clustering of BMP-regulated genes in control (CONT) and DMH1 treated FG and HG

samples compared to dorsal explants (DORS), which contain a thin layer of dorsal endoderm (yellow) along with neural and somite tissue. The DMH1-treated FG showed similarities to the dorsal tissue suggesting that BMP induces ventral mesendoderm fate and represses dorsal fate. **(E)** *In situ* hybridization of control, DMH1 treated or BMP2 injected embryos, in a cross section confirms that expression of the dorsal mesoderm gene *foxc2* is expanded ventrally with DMH1 and restricted dorsally upon BMP injection. White line indicates expression domain.



E

Motifs in Smad1 peaks of BMP act genes

Motif	Name	e value
	Gata	1e-35
	Lhx	1e-30
	Tcf	1e-17
	Smad	1e-11

Motifs in Smad1 peaks of BMP rep genes

Motif	Name	e value
	Gata	1e-60
	Lhx	1e-54
	Nkx	1e-40
	Tcf	1e-25

Fig. S3: Smad1 ChIP-seq of embryos stage NF20.
(A) Genomic distribution of Smad1 ChIP-seq peaks in stage NF20 embryos categorized as upstream (-20kb), downstream (+20kb), intragenic and promoter (-1kb to +1kb) regions. **(B)** Smad1 ChIP-PCR of known BMP-target genes showing reduced Smad1-binding to CRMs of *msx1*, *id3* and *ventx2.1* promoters in DMH1 treated embryos compare to DMSO controls. **(C)** Peak overlap between Smad1 and p300 ChIP-seq of stage NF20 whole embryos. **(D)** Venn showing the overlap between Smad1-bound genes, p300-bound genes and genes expressed in NF20 embryo at levels higher than 5 transcripts per million (TPM>5) based on RNA-seq. **(E)** Motif enrichment analysis of Smad1 ChIP-seq peaks associated with activated (act) or repressed (rep) genes.

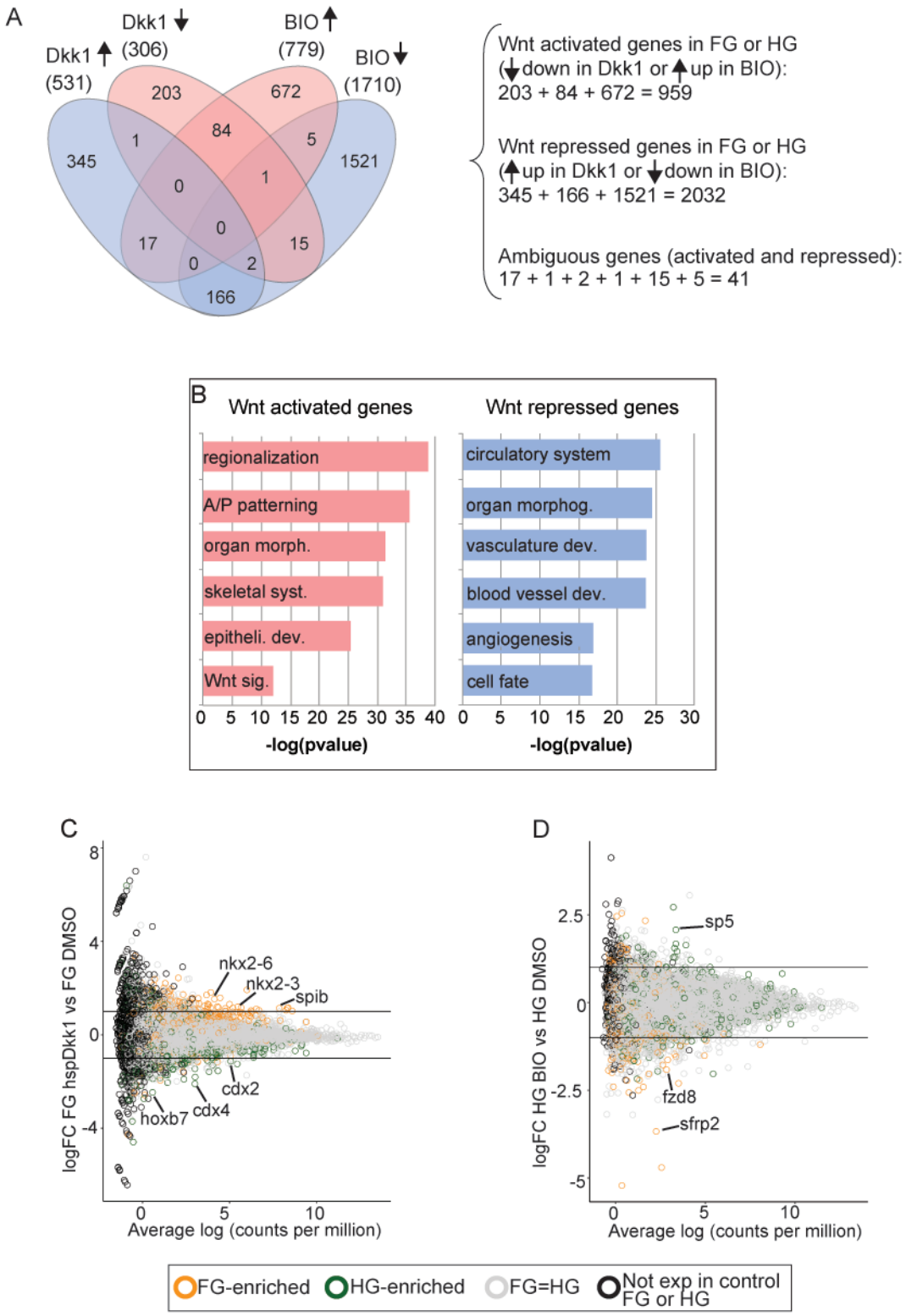


Fig. S4: RNA-seq of control, *Tg(hsp70:dkk1)* and BIO treated embryos.
(A) Venn diagram illustrates transcripts with expression increased (↑) or decreased (↓) upon *Tg(hsp70:dkk1)* or BIO treatment in FG or HG tissues. Wnt-activated genes were log2FC ≤-1 upon heat-shock or log2FC ≥1 upon BIO treatment FDR ≤5% (n=959). Wnt-repressed genes were log2FC ≥1 upon

heat-shock or log2FC ≤-1 upon BIO treatment FDR ≤5% (n=2032). Forty-one transcripts had ambiguous regulation with evidence of being both Wnt-activated and Wnt-repressed, and were excluded from further analysis. Overall we categorized a total of 2991 (959+2032) Wnt-regulated genes in the FG and HG tissue. **(B)** GO term enrichment analysis of Wnt-activated and Wnt–repressed genes. **(C-D)** Scatter plot showing log2FC in expression between FG non-heatshock and FG *Tg(hsp70:dkk1)* (**C**) and HG DMSO and HG BIO (**D**) explants. Transcripts are colored based on the normal control expression; HG-enriched in green, FG-enriched in orange, expressed similarly FG and HG (FG=HG) in grey and normally not expressed in FG or HG in black.

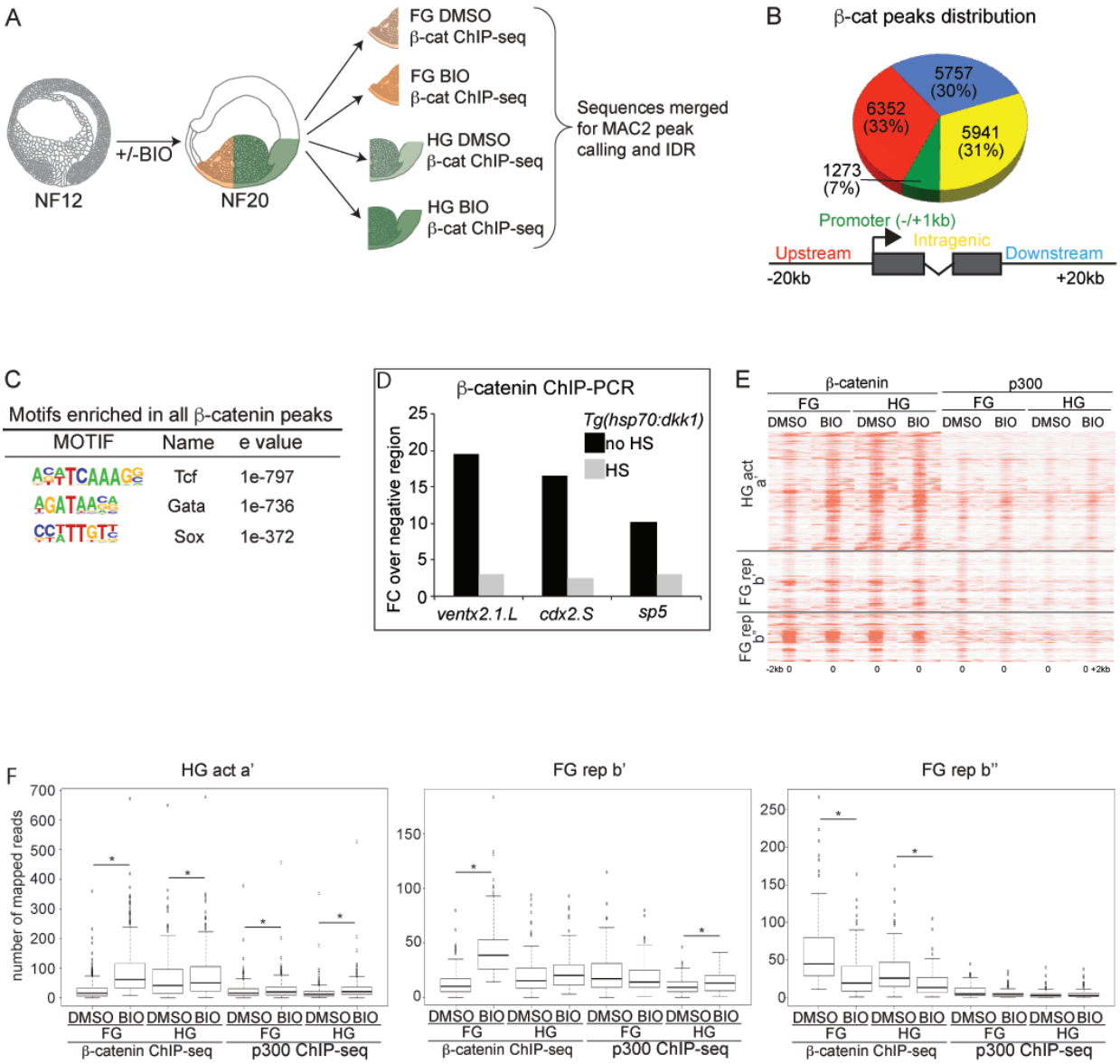


Fig. S5: β-catenin ChIP-seq of embryos stage NF20.

(A) Experimental design showing β-catenin ChIP-seq of 100 FG or 100 HG explants dissected from DMSO vehicle control or BIO treated NF20 embryos. Reads were merged and MACS2 peak calling was performed followed by irreproducibility discovery rate (IDR) filtering with standard thresholds (Li et al., 2011) identified 16303 statistically significant peaks associated with 11007 genes (+/- 20 Kb from transcription start site) in the FG and HG samples. **(B)** Genomic distribution of β-catenin ChIP-seq peaks classified as upstream (-20kb), downstream (+20kb), intragenic and promoter (-1kb to +1kb) regions. **(C)** DNA-binding protein motif enrichment analysis of all β-catenin ChIP-seq peaks in the genome. **(D)** β-catenin ChIP-PCR of known CRMs in Wnt-target genes *ventx2.1*, *cdx2* and *sp5* from *Tg(hsp70:dkk1)* embryos with and without heat shock (HS). **(E)** Read density of different classes of β-catenin and p300 peaks in DMSO or BIO treated FG and HG explants from Fig. 5B. +/-2kb centered on the β-catenin peak summit. **(F)** Box plots of average tag density of β-catenin and p300 peaks on HG Wnt-activated genes (a'), and FG Wnt-repressed genes (b' and b") upon BIO treatment. *p<0.05, Wilcoxon test.

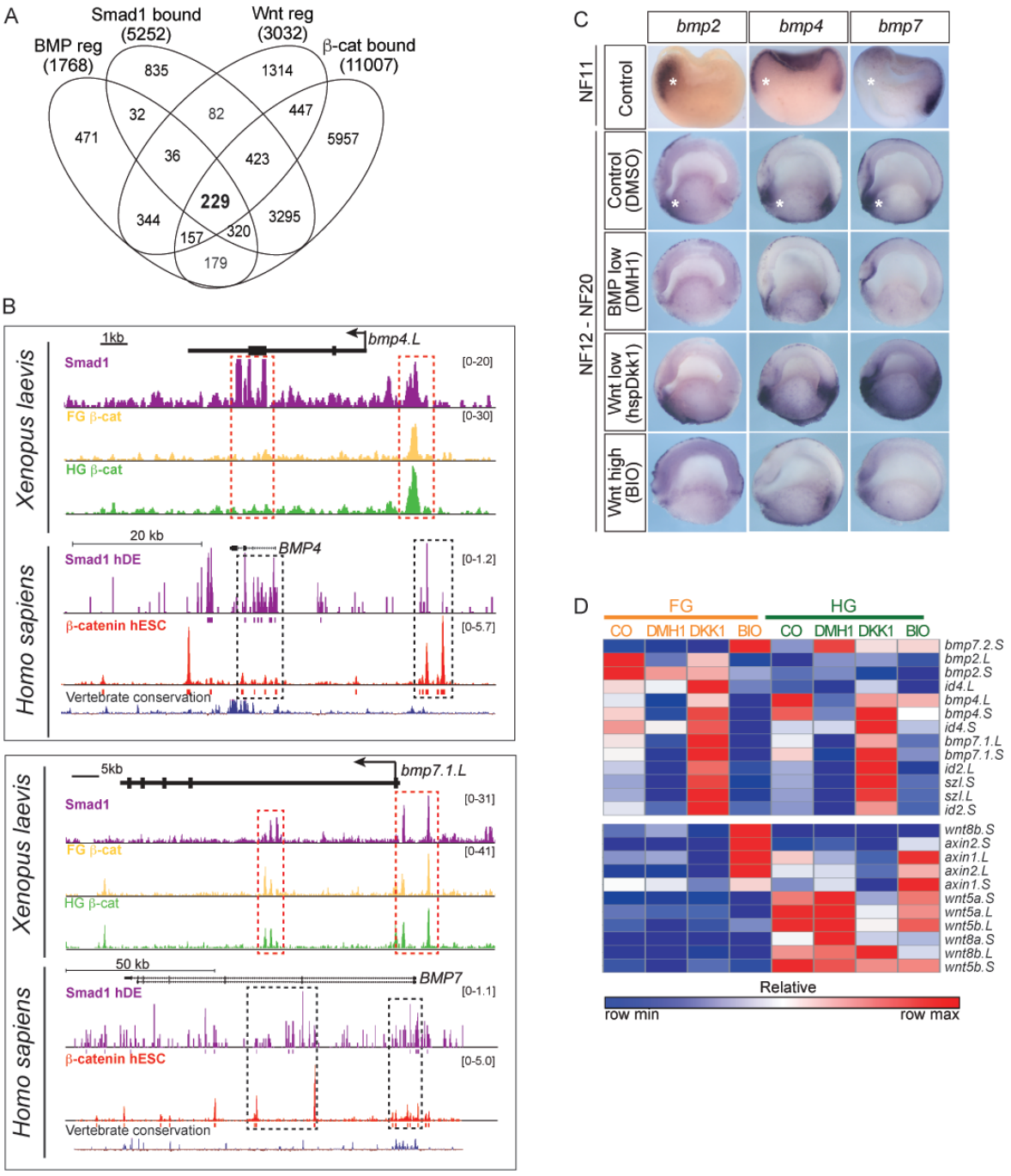


Fig. S6: BMP/Smad1 and Wnt/β-catenin crosstalk.
(A) Overlap between BMP-regulated/Smad1-bound and Wnt-regulated/β-catenin-bound genes. **(B**) Browser view of β-catenin and Smad1 peaks in *Xenopus* and human genes (from GSM1505734 and GSM1579346). Red boxes indicate overlapping Smad1 and β-catenin peaks. Black boxes indicate syntenic peaks. **(C)** *In situ* hybridization of DMH1, *Tg(hsp70:dkk1)* and BIO treated embryos in mid-sagittal section, anterior left and dorsal up. Embryos are either stage NF12 wild-type or treated at stage NF11 and fixed at stage NF20. * indicates *hhex*-expressing FG cells. **(D)** Expression heatmap of BMP and Wnt ligands and targets present in FG and HG samples of controls (CO), DMH1, *Tg(hsp70:dkk1)* and BIO-treated embryos.

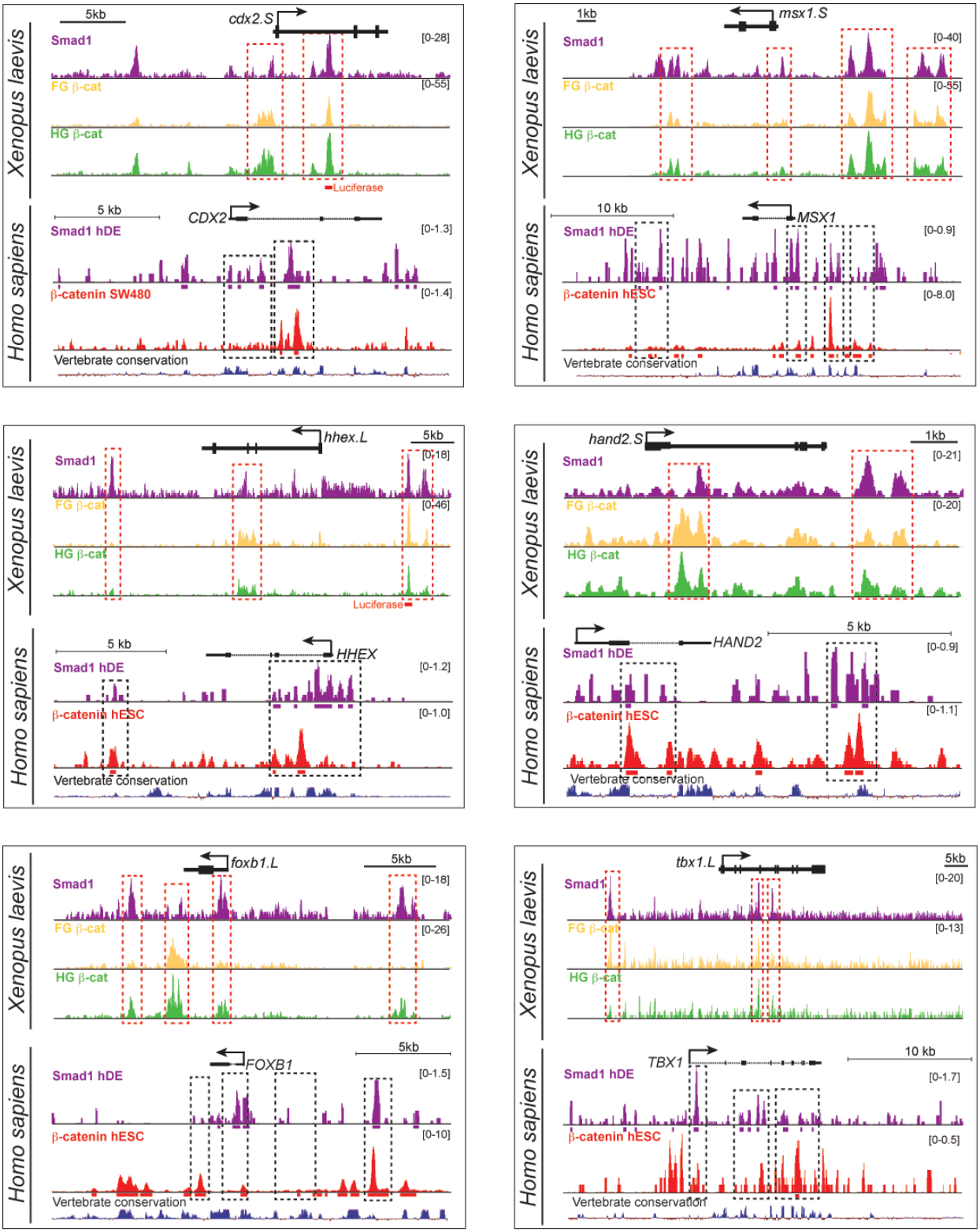


Fig.S7: Smad1 and β-catenin syntenic peaks in *Xenopus laevis* and *Homo sapiens*.

Fig.S7: Smad1 and β-catenin syntenic peaks in *Xenopus laevis* and *Homo sapiens*.
Browser view of β-catenin and Smad1 peaks in *Xenopus* and human genes (from the following public data: GSM1505734 and GSM1579346). Red boxes indicate overlapping Smad1 and β-catenin peaks. Black boxes indicate syntenic peaks.

Supplemental Tables

Table S1: Transcriptional program of FG and HG progenitors.
Differential expression analysis between FG and HG samples identified 906 FG-enriched and 987 HG-enriched genes. Comparison between FG+HG endo and FG+HG meso identified 3439 endo-enriched and 4829 meso-enriched genes. log2FC ≤-1 or ≥1, FDR ≤5%.

Click here to Download Table S1

Table S2: Tables show the number of transcripts overlapping in the following pairwise differential expression analyses. Enriched transcripts have log2 fold change ≤-1 or ≥1 difference in expression and false discovery rate ≤5%.
(A) Intersection of FG versus HG and endo versus meso transcripts. This table is supplementary to Fig. 1B-C. **(B)** Intersection of FG versus HG and BMP-activated versus BMP-repressed transcripts. This table is supplementary to Fig. 2B. **(C)** Intersection of FG versus HG and Wnt-activated versus Wnt-repressed. This table is supplementary to Fig. 4B.

A. Intersection of FG versus HG and endo versus meso transcripts (see Fig. 1B-C)

	Endo-enriched	Meso-enriched	Endo≅Meso
FG-enriched	172	294	440
HG-enriched	518	202	267
FG≅HG	2749	4333	

B. Intersection of FG versus HG and BMP-activated versus BMP-repressed transcripts (see Fig. 2B)

	BMP-activated	BMP-repressed	
FG-enriched	155	185	566
HG-enriched	97	89	801
FG≅HG	445	789	

C. Intersection of FG versus HG and Wnt-activated versus Wnt-repressed (see Fig. 4B)

	Wnt-activated	Wnt-repressed	
FG-enriched	28	496	382
HG-enriched	247	57	683
FG≅HG	684	1479	

Table S3: FG and HG transcriptome conservation among vertebrates.
Manually curated list of genes expressed in FG and HG tissue from mouse embryos and direct differentiation of human stem cells. Genes expressed with tpm≥1 in at least one of the frog samples are considered present.

Click here to Download Table S3

Table S4: BMP regulated genes from RNA-seq analysis.
BMP differentially expressed genes in DMH1 treated FG (Sheet 1) or HG (Sheet 2) samples compared to DMSO control. Differentially expressed genes are identified with their gene name, log2FC, p value and FDR. Log fold change and FDR indicate those values of DMH1 experiments compared to DMSO control. Experiments were done in triplicate with log2FC ≤-1 or ≥1, FDR ≤5%.

Click here to Download Table S4

Table S5: Smad1 and p300 peaks of whole embryos stage NF20
Smad1 (Sheet1) and p300 (Sheet2) ChIP-seq identified 7976 and 4727 peaks, respectively. The position of each Smad1 peak is indicated by “Chromosome/Scaffold”, “Peak_Start” and “Peak_Stop”, with “Summit” of the peak. The nearest genes are indicated by “gene” with the “Gene_Start” and “Gene_Stop” positions and “Distance_to_Gene_TSS_in_bp_from_Summit”. Peaks were categorized depending on where they fall related to each gene. Peaks inside the gene were categorized as ”intragenic”, peaks +1kb/-1kb of the TSS are “promoter”, peaks downstream of the gene are “proximal downstream” (+10kb) or “distal downstream” (+20kb), and peaks upstream of the gene are “proximal upstream” (-10kb) or “distal upstream” (-20kb).

Click here to Download Table S5

Table S6: Wnt regulated genes from RNA-seq analysis.
Wnt differentially expressed genes in *Tg(hsp70:dkk1)* or BIO treated FG (Sheet 1 and 3) or HG (Sheet 2 and 4) samples compared to non-heatshock or DMSO control. Differentially expressed genes are identified with their gene name, log2FC, p value and FDR. Log fold change and FDR indicate those values of experimental manipulation compared to control. Experiments were done in triplicate with log2FC ≤-1 or ≥1, FDR ≤5%.

Click here to Download Table S6

Table S7: β-catenin and p300 peaks of whole embryos and FG+HG tissues stage NF20
β-catenin (sheet1) and p300 (sheet2) ChIP-seq identified 16303 and 15146 peaks MACS2 IDR, respectively. The position of each β-catenin peak is indicated by "Chromosome/Scaffold", "Peak_Start" and "Peak_Stop", with "Summit" of the peak. The nearest genes are indicated by "gene" with the "Gene_Start" and "Gene_Stop" positions and "Distance_to_Gene_TSS_in_bp_from_Summit". Peaks were categorized depending on where they fall related to each gene. Peaks inside the gene were categorized as "intragenic", peaks +1kb/-1kb of the TSS are "promoter", peaks downstream of the gene are "proximal downstream" (+10kb) or "distal downstream" (+20kb), and peaks upstream of the gene are "proximal upstream" (-10kb) or "distal upstream" (-20kb).

Click here to Download Table S7

Table S8: Genes associated with Smad1 and β-catenin peaks. Overlapping peaks are considered when the overlap is of at least 1 nucleotide. Genes were separated in different lists according to BMP and Wnt regulation, based on the RNA-seq data. Activated (act), repressed (rep), FG-enriched (orange) and HG-enriched (green).

Click here to Download Table S8

Table S9: Syntenic Smad1 and β-catenin peaks between *Xenopus* and human ChIP-seq data. Peaks considered syntenic have similar positions in relation to both *Xenopus* and human genes. Human publicly available data for SMAD1 GSM1505734 (Tsankov et al., 2015) and β-CATENIN GSM1579346 (Estaras et al., 2015) and GSM1303695 (Watanabe et al., 2014).

Click here to Download Table S9

Table S10: Summary of FG- and HG-enriched genes indicating BMP and Wnt and association of Smad1 or β-catenin peaks

Table with 906 FG-enriched and 987 HG-enriched genes and how they were affected by the different BMP and Wnt manipulations, as well as whether they were associated with Smad1 or β-catenin peak within +/-20kb. FG-enriched genes in orange with log2FC ≤1 and HG-enriched genes in green with log2FC ≥1. Endoderm-enriched genes in yellow with log2FC ≤1 and mesoderm-enriched genes in red with logFC ≥1. BMP inhibition with DMH1 in FG and HG tissues with activated genes in pink with logFC ≥1 and repressed genes in blue with logFC ≥1. Wnt activation with BIO in FG and HG tissues with activated genes in pink with logFC ≥1 and repressed genes in blue with log2FC ≤1. Wnt inhibition with *Tg(hsp70:dkk1)* in FG and HG tissues with activated genes in pink with log2FC ≤1 and repressed genes in blue with log2FC ≥1. For simplicity tpm values are represented by average of the replicates. Smad1 and β-catenin peaks association within +/-20kb of each transcript is shown with genome coordinates. NS = non-significant log2FC (-1< log2FC >1).

Click here to Download Table S10

SUPPLEMENTARY MATERIALS AND METHODS

Luciferase reporter assay

The following CRMs were cloned into the pGL4.23 luc2 miniP vector (Promega) to generate *hhex:luc* and *cdx2:luc* luciferase constructs:

hhex CRM:

TTGTCTCTGCTCCCCTTGCTCATTACCTGCCCAGTCCCTATACACACCTTGCTGCT
CACACTGAGAGGGTAGAGACAAGGAATCTTCTCCCATCTGAGCGGCGCCGA

cdx2 CRM with Tcf motif in bold (based on Cis-BP TF binding tool (PWMs – LogOdds >10) (Weirauch et al., 2014):

CGGCGGCGTTTGTTCAGTAGTGGTAATTCCAAATATCTATAGGCCTGATAACATTTT
GCCTTGTAGCTCATTGTTAGCCCCTGTGTTCTCCATTCATTGACACTGCCCAAT**TCT**
CTCTGATCTGCCTTGTCCCCTCTCCA

One hundred picograms of *hhex:luc* and *cdx2:luc* luciferase constructs were co-injected with PRL-SV40 control Renilla vector (Promega) (25pg) into C1 (presumptive FG) or C4 (presumptive HG) cells of 32-cell stage embryos. At stage NF12 embryos were treated with DMSO, DMH1 or BIO (as described) and 3 embryos were frozen in triplicate at stage NF20. Embryos were lysed by pipetting in 75uL of 100mM TRIS-HCl pH7.4 + 0.2% NP-40 and then 25 uL of embryo Lysate was assayed using a dual luciferase assay kit (Biotium,Inc). Luciferase activity was normalized to co-injected TK:renilla and the mean relative activity of the triplicate samples was shown ±S.D. with pairwise student T-tests to determine significant differences in expression. Each experiment was repeated a minimum of two times and a representative result is shown.

RNA-seq analysis

For each RNA-seq sample, 50 explants were microdissected and when necessary cultured in 10 µg/ml dispase for 15-20 minutes to separate endo and meso. Total RNA was extracted from two or three independent biological replicates with the Nucleo-spin RNA kit (Machery-Nagel). Libraries were constructed with TruSeq Stranded mRNA Library Prep Kit and sequenced ~7-10 million reads/library with 75 bp length using Illumina HiSeq2500. FastQC reports identified adapters, over-represented

sequences, low quality bases and overall low quality reads. Trimmomatic was used to clip off adapters, over-represented sequences and low quality bases. Reads were trimmed keeping minimum length as 50, thereby after trimming, read lengths ranged from 50 to 75 base pairs. Quality trimmed reads were mapped to the *X. laevis* genome 9.1, quantified using RSEM and mapped with bowtie2 using default thresholds (Li and Dewey, 2011). Differential gene expression analysis was carried out using CSBB's [https://github.com/csbbcompbio/CSBB-v1.0] Differential Expression Module, which uses RUVSeq (Risso et al., 2014). With RUVSeq we performed two-way normalization on the count's matrix 1) Upper Quantile and 2) Empirical gene normalization with default settings, and differential expression analysis. Pairwise comparisons create mutually exclusive lists of enriched genes with log2FC ≤-1 or ≥1, $p < 0.05$ and FDR ≤5% differences in expression for each of the following pairwise comparison analysis:

- To define FG- and HG-enriched transcripts we merged fastq files from FG-endo and FG-meso from the same biological replicates, as well as HG-endo and HG-meso samples. This resulted in ~14-20 million reads for each biological replicate of FG (endo+meso) and HG (endo+meso). Correlation analysis indicated that merging fastq was very similar to sequencing intact FG and HG (with meso and endo not separated) r^2=0.93, compared to different biological replicates of intact FG or intact HG r^2=0.94-0.90, validating this approach. We then performed a differential expression analysis comparing FG and HG. Transcripts with log2FC ≤-1 are classified as FG-enriched (n=906) and log2FC ≥1 as HG-enriched (n=987).

- To define endo- and meso-enriched genes we merged fastq files from FG endo to HG endo from the same biological replicates to generate an "endo'" transcriptome as well as FG meso to HG meso samples to generate a "meso" sample, similar to our approach described above. We then performed a differential expression analysis comparing endo and meso. Transcripts with log2FC ≤-1 are classified as endo-enriched (n=3439) and log2FC ≥1 as meso-enriched (n=4829).

- To identify BMP-regulated genes we compared FG DMH1 with FG DMSO as well as HG DMH1 with HG DMSO samples. Transcripts with log2FC ≤-1 are classified as BMP-activated genes (n=697) and log2FC ≥1 as BMP-repressed genes (n=1063). Eight transcripts had ambiguous regulation being both activated and repressed by DMH1 in FG or HG tissues, and were excluded from further analysis. Overall we categorized a total of 1760 (697+1063) BMP-regulated genes in the FG and HG tissue.

- To identify Wnt-regulated genes we compared *Tg(hsp70:dkk1)* heatshocked with non-heatshocked FG, as well as heatshocked with non-heatshocked HG. We also compared FG BIO with FG DMSO as well as HG BIO with HG DMSO samples. Wnt-activated genes were log2FC ≤-1 upon heat-shock or log2FC ≥1 upon BIO treatment FDR ≤5% (n=959). Wnt-repressed genes were log2FC ≥1 upon heat-shock or log2FC ≤-1 upon BIO treatment FDR ≤5% (n=2032). Forty-one transcripts had ambiguous regulation with evidence of being both Wnt-activated and Wnt-repressed, and were excluded from further analysis. Overall we categorized a total of 2991 (959+2032) Wnt-regulated genes in the FG and HG tissue.

GO term enrichment analyses were performed using ToppGene Suite (Chen et al., 2009). Heatmaps were generated using GeneE from Broad Institute (https://software.broadinstitute.org/GENE-E/index.html). Scatter Plots were generated using CSBB's InteractiveScatterPlot module.

Chromatin immunoprecipitation

Embryos (25-50 whole embryos or 100 FG or HG explants) at stage NF20 were harvested and fixed at room temperature with 1% formaldehyde in 0.1XMBS for 45 minutes. Immediately after fixation, the embryos were incubated with 125 mM glycine/MBS for 10 minutes and washed three times with ice-cold RIPA buffer (50 mM Tris pH 7.6,150 mM NaCl, 1 mM EDTA, 1% IGEPAL CA-630, 0.25% Sodium deoxycholate, 0.1% SDS, 0.5 mM DTT, and supplemented with Protease Inhibitor

Cocktail (Sigma,P8340)) for 5 minutes. Batches of 50 embryos were snap-frozen in liquid nitrogen and stored at -80°C for future use. Embryos were thawed on ice, 1 ml of RIPA buffer was added, homogenized, and then kept on ice for 10 minutes. The lysate was centrifuged at 14,000 rpm for 10 minutes at 4°C, and the pellet was ressuspended in 1ml of RIPA buffer and transferred to a Bioruptor tube (Diagenode) for sonication. The lysate was sonicated for 15 cycles of 20 seconds ON and 60 seconds OFF on the Bioruptor Pico Instrument (Diagenode). The sonicated samples were centrifuged at 14,000 rpm for 10 minutes at 4°C, and the supernatant were transferred to a 1.5ml tube. The supernatant was blocked for 2 hours at 4°C with Dynabeads Protein G (Life technologies). In a separate tube, 20 µl of Dynabeads Protein G was blocked with 1 ml 5% BSA/PBS for 1 hour at 4°C. Followed by another 1 hour incubation with the following antibodies on ice: 20 µl of anti-Smad1 per IP (Invitrogen, 38-5400), 20 µl of anti-β-catenin per IP (Life technologies, 712700) and 3 µl of anti-p300 per IP (Santa Cruz sc-585 X). A small chromatin aliquot was saved for input (50 µl) and the rest was transferred to the tube with beads and antibody, and incubated overnight at 4°C. The input material was stored at -20°C for later usage. The beads were successively washed with ChIP buffer 1 (50 mM HEPES-KOH pH 7.5, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate), ChIP buffer 2 (50 mM HEPES-KOH pH 7.5, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate), ChIP buffer 3 (10 mM Tris pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% IGEPAL CA-630, 0.5% Sodium deoxycholate), ChIP buffer 4 (10 mM Tris pH 8.0, 1 mM EDTA) for 20 minutes each. Chromatin was eluted from the beads with 105 µl of elution buffer (50 mM Tris pH 8.0, 10 mM EDTA, 1% SDS) for 2 washes of 30 minutes at 65°C. At this stage, the frozen input samples were supplemented with elution buffer and incubated overnight at 65°C for reverse crosslinking. ChIP and input samples were incubated with RNase A at 37°C for 1 hour and treated with proteinase K for 1 hours at 55°C. The de-crosslinked DNA fragments were purified with phenol:chloroform:isoamylalcohol and precipitated in ethanol for qPCR. qPCR was performed using iQ SYBR Green Supermix (BIORAD) on a QuantStudio 3 Real-time PCR System (ThermoFisher). qPCR primers used were:
sp5 (F, 5'- TGT CCC GCC TTT TGT CAC CTC-3' and R, 5'- GCC GCC CAA TCA TCA AAG AAG-3');

ventx2.1 (F, 5'- CAT AGC CAG CTG AGC ATA ATA AA-3' and R, 5'- TCA AAG GCA GAG ATC ACT ACC A-3');
msx1 (F, 5'- CAT ATG TTT GGG TTT GGA GAG-3' and R, 5'-GTG CAG AAC ATG GGA GAT TAG-3');
id3 (F, 5'- TTC GGC GCC GTT GGT TAC TTT ACT -3' and R, 5'- GTC TCC ACG GGC AAC CAC TCC TT -3');
cdx2 (F, 5'- AGG TTT CGG CGG CGT TTG TT-3' and R, 5'- TTG GGC AGT GTT AGT GAA TGG AGA -3');
sp5 -15kb (F, 5'- GTG ATA AAG TAG TCC CAG CAG TGA-3' and R, 5'- AAG GGG GAA ATT TAA ACC AGA TA-3');
ventx2.1 -15kb (F, 5'- GTA GGA ACC CAC AGC CAA TAA TC-3' and R, 5'- GTC AGT AAG AAA ATC GCC CAT AAG-3');
id3 -8.5kb (F, 5'- TTC CCT GTG CCT GTG TTG AT-3' and R, 5'- TTG GGG GCA TTT ATT TAG TTA TT-3').

ChIP-seq analysis

ThruPLEX® DNA-seq libraries were constructed from ChIP and input control DNA and sequenced (~30 million reads/library) using Illumina HiSeq2500. Raw reads quality check and quality trimming was performed using FastQC and Trimmomatic. Duplicate mapped and multi-mapped reads were removed using picard and samtools respectively. Peaks were called with MACS2 at default thresholds [--qvalue 0.01, --mfold 5:50, --call-summits] (Zhang et al., 2008). IDR (Irreproducibility Discovery Rate) was performed with standard thresholds (Li et al., 2011) to identify high-confidence: Smad1, β-catenin and p300 reproducible peaks as follows:

- Smad1 and p300 whole embryo ChIP-seq were individually mapped to the *X. laevis* genome assembly v9.1 (Session et al., 2016) using Bowtie2 at default thresholds (Langmead and Salzberg, 2012).

- β-catenin and p300 in either FG or HG explants with or without BIO. Fastq files from individual ChIP-seq experiments were merged for FG/HG and DMSO/BIO

explants from β-catenin or p300 ChIP-seq datasets. Pooled fastq's reads were mapped to the *X. laevis* genome assembly v9.1 (Session et al., 2016) using Bowtie2 at default thresholds (Langmead and Salzberg, 2012).

Described bam files (from merged and not merged fastq) were converted to tagAlign format with only mapped reads with mapping quality ≥30 using samtools and bedtools. With this tagAlign file we created three replicates of equal sizes by shuffling and randomly placing tags in each replicate. We performed the IDR analysis on the 3 replicates and input using Rscript as described for ENCODE with a threshold of 0.01 [https://sites.google.com/site/anshulkundaje/projects/idr]. The merged input tagAlign file was used as the input for MACS2 peak calling with [-p 1e-3, --to-large] thresholds. Our IDR pipeline resulted in higher confidence peaks than with MACS2 peak calling alone. HOMER findMotifsGenome.pl script was used for motif analysis (Heinz et al., 2010). For the Hypergeometric test we used dhyper function in R (https://stat.ethz.ch/R-manual/R-devel/library/stats/html/Hypergeometric.html). Genome browser views were visualized with Integrative Genomics Viewer (IGV) (Robinson et al., 2011).

The sum of all of our genomic analysis is provided in the Table S10.

REFERENCES

Chen, J., Bardes, E. E., Aronow, B. J. and Jegga, A. G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* **37**, W305-311.

Estaras, C., Benner, C. and Jones, K. A. (2015). SMADs and YAP compete to control elongation of beta-catenin:LEF-1-recruited RNAPII during hESC differentiation. *Mol Cell* **58**, 780-793.

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.

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, Q., Brown, J. B., Huang, H. and Bickel, P. J. (2011). Measuring reproducibility of high-throughput experiments. 1752-1779.

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.

Session, A. M., Uno, Y., Kwon, T., Chapman, J. A., Toyoda, A., Takahashi, S., Fukui, A., Hikosaka, A., Suzuki, A., Kondo, M., et al. (2016). Genome evolution in the allotetraploid frog Xenopus laevis. *Nature* **538**, 336-343.

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.

Watanabe, K., Biesinger, J., Salmans, M. L., Roberts, B. S., Arthur, W. T., Cleary, M., Andersen, B., Xie, X. and Dai, X. (2014). Integrative ChIP-seq/microarray analysis identifies a CTNNB1 target signature enriched in intestinal stem cells and colon cancer. *PLoS One* **9**, e92317.

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.

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.