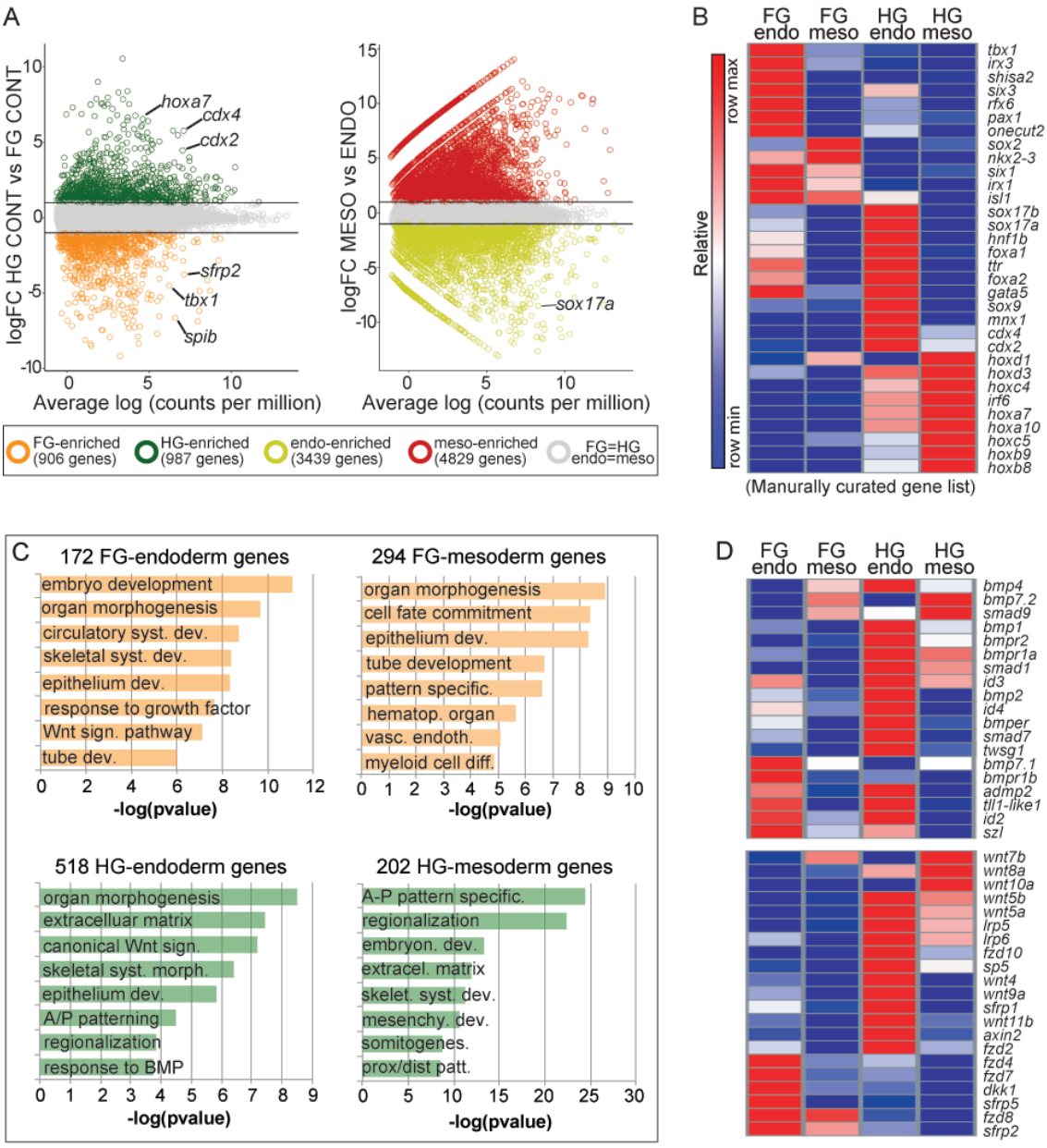


Fig. S1: Transcriptional program of FG and HG progenitors in vivo.

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.

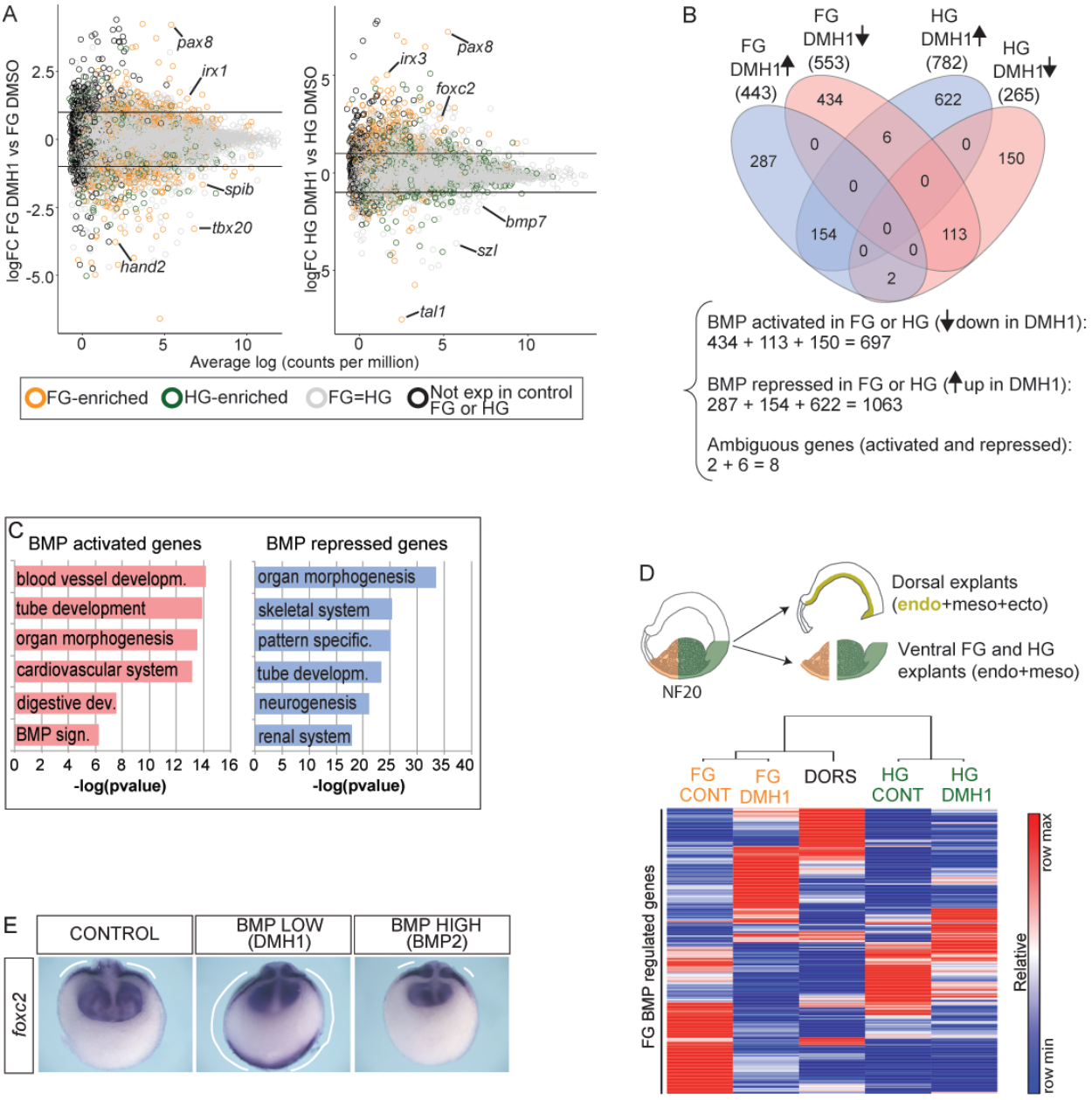
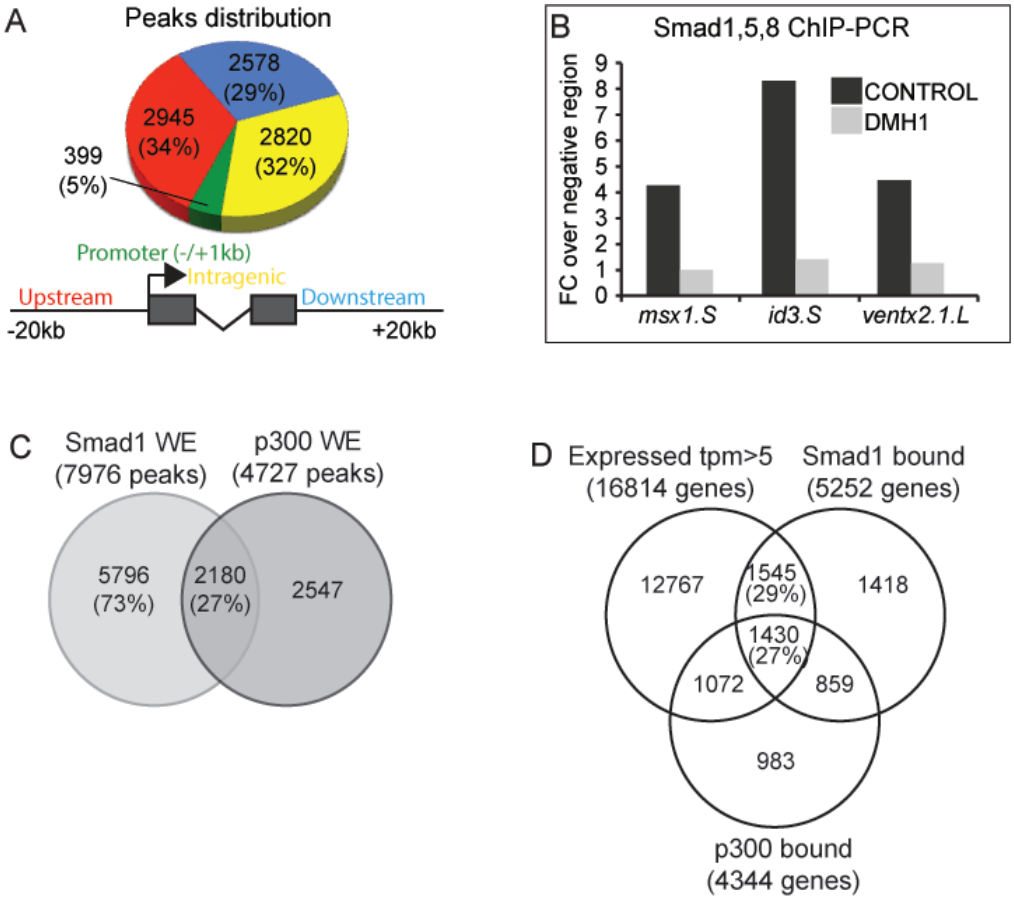


Fig. S2: RNA-seq of DMSO and DMH1 treated embryos identified BMP-regulated genes.
(A) Scatter plot showing log2FC in expression between DMSO and DMH1 treated FG (left) and between DMSO and DMH1 treated HG (right) samples. Transcripts are colored based on the normal control expression; HG-enriched in 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.

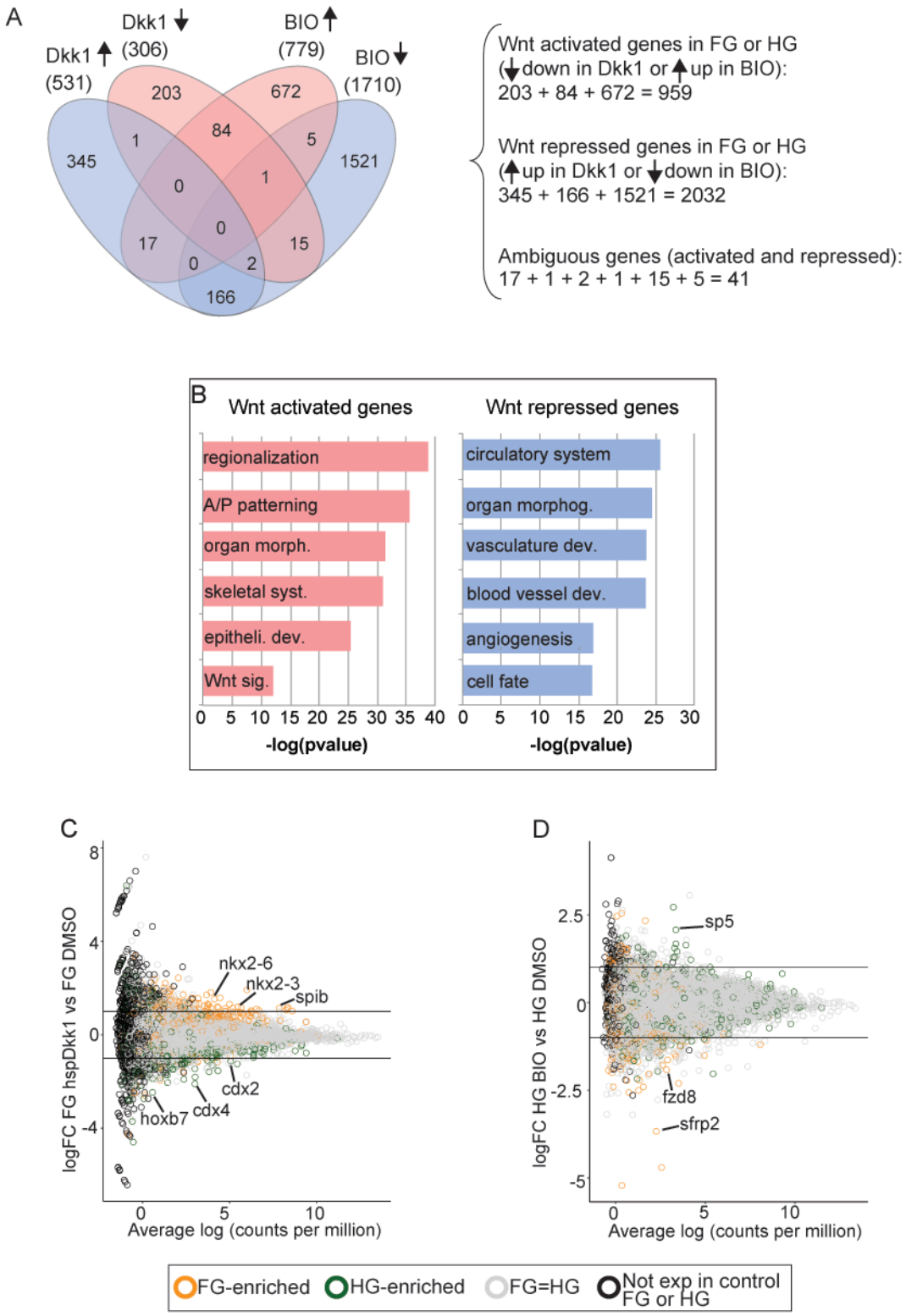


Fig. S4: RNA-seq of control, *Tg(hsp70:dkk1)* and BIO treated embryos.
(A) Venn diagram illustrates transcripts with expression increased (↑) or decreased (↓) upon *Tg(hsp70:dkk1)* or BIO treatment in FG or HG tissues. Wnt-activated genes were log2FC ≤-1 upon heat-shock or log2FC ≥1 upon BIO treatment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

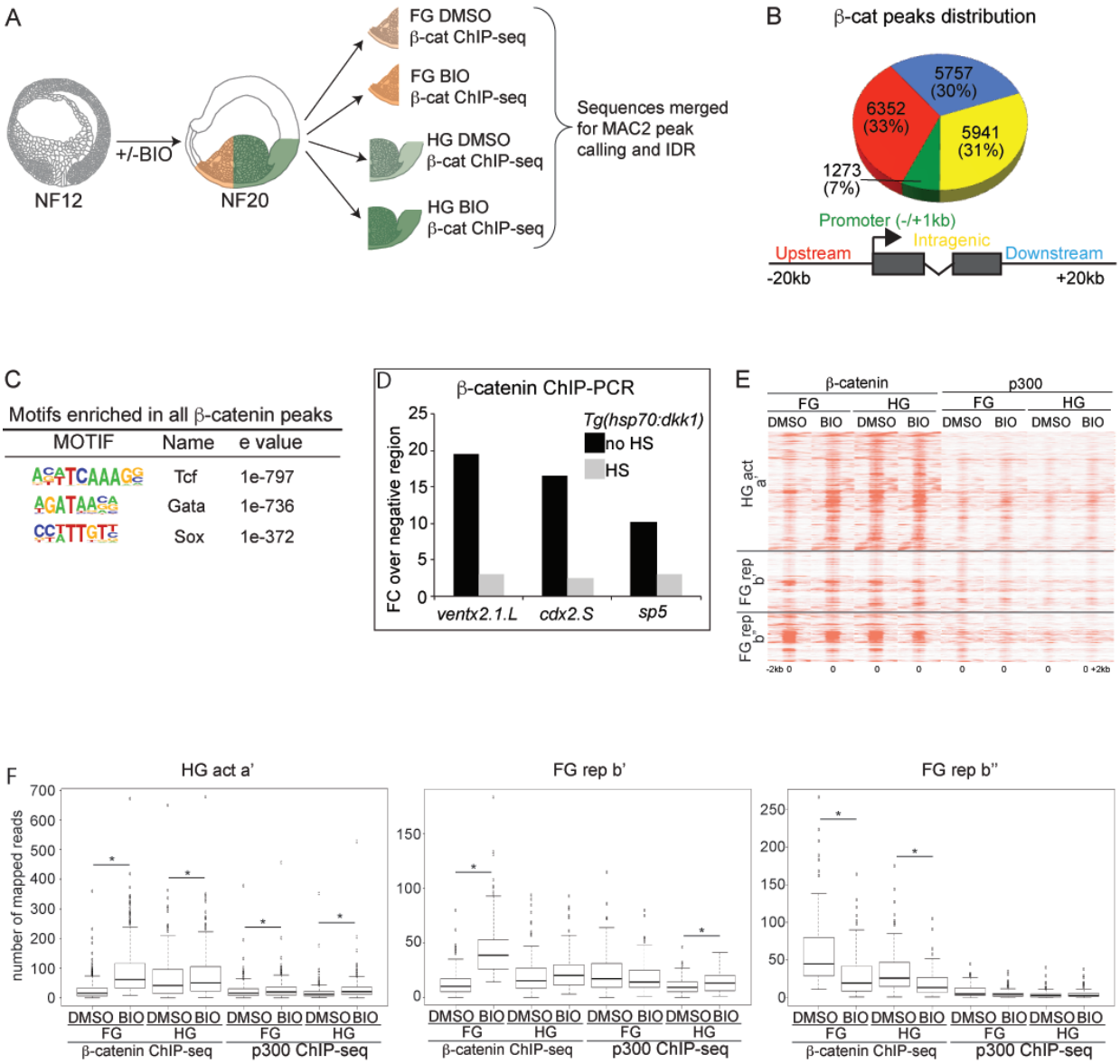


Fig. S5: β-catenin ChIP-seq of embryos stage NF20.

(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.

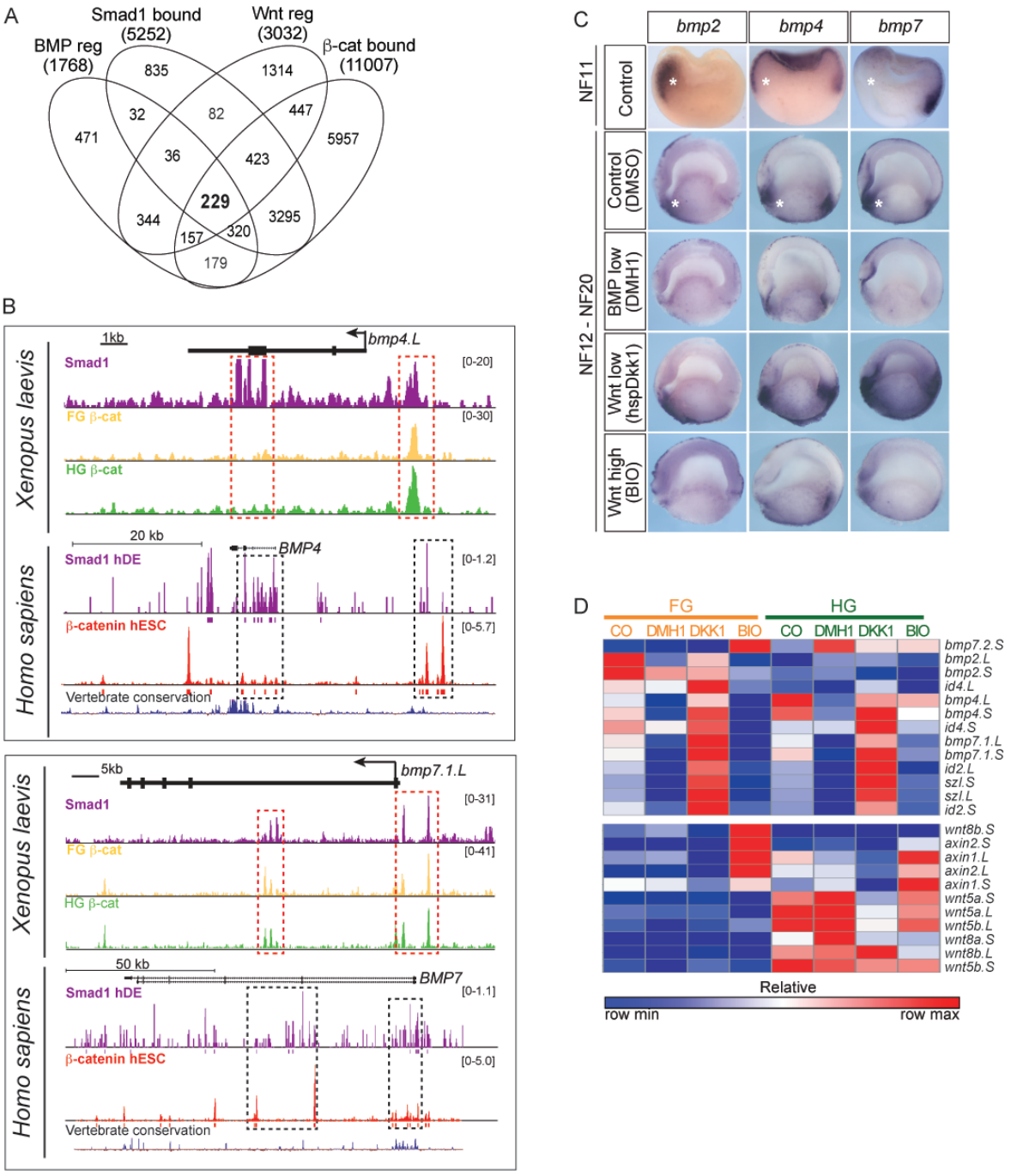


Fig. S6: BMP/Smad1 and Wnt/β-catenin crosstalk.
(A) Overlap between BMP-regulated/Smad1-bound and Wnt-regulated/β-catenin-bound genes. **(B**) Browser view of β-catenin and Smad1 peaks in *Xenopus* and human genes (from GSM1505734 and GSM1579346). Red boxes indicate overlapping Smad1 and β-catenin peaks. 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.

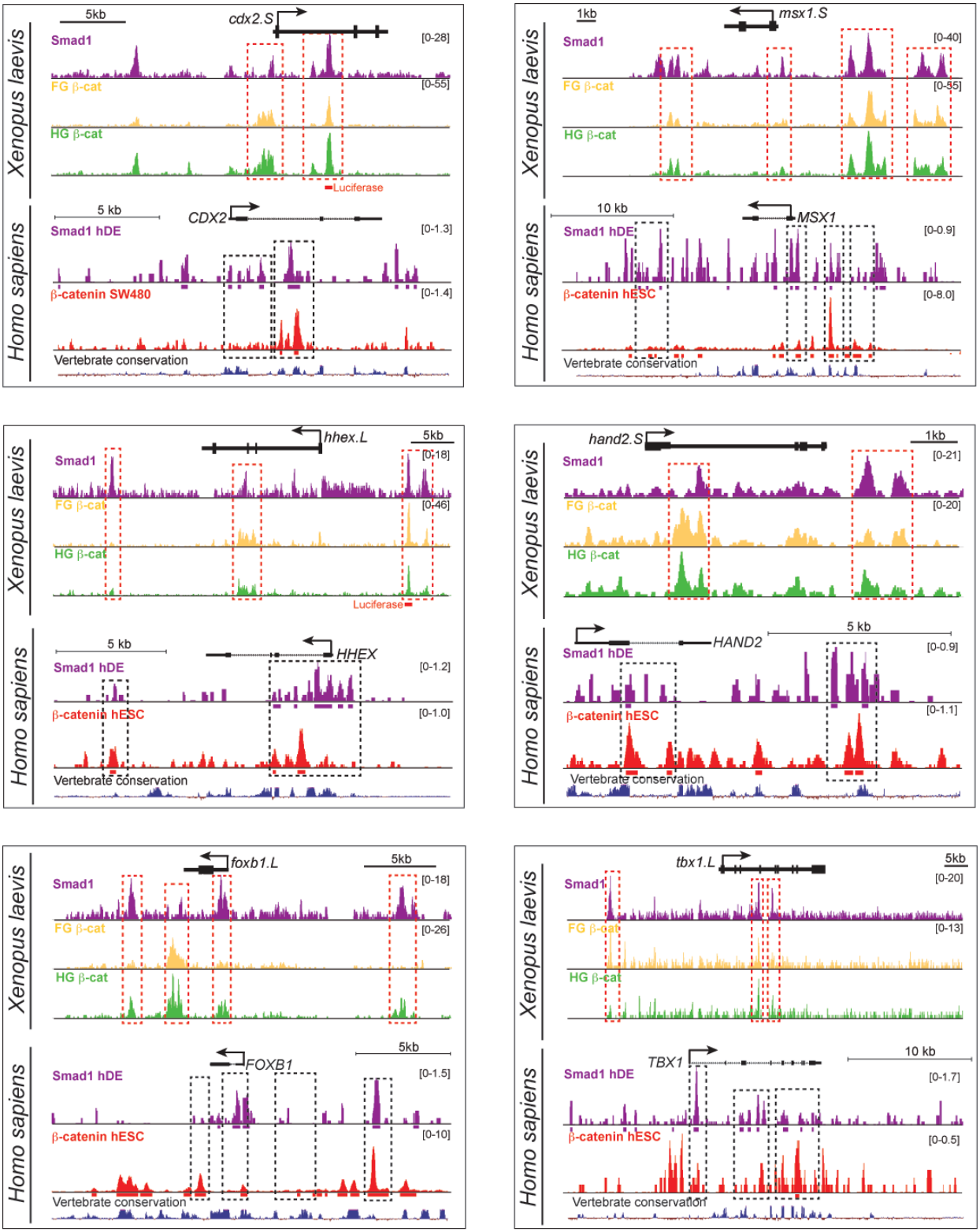


Fig.S7: Smad1 and β-catenin syntenic peaks in *Xenopus laevis* and *Homo sapiens*.

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.