BMP4 triggers regulatory circuits specifying the cardiac mesoderm lineage

Pavel Tsaytler[1,*], Jinhua Liu[1], Gaby Blaess[1], Dennis Schifferl[1], Jesse V. Veenvliet[1], Lars Wittler[1], Bernd Timmermann[2], Bernhard G. Herrmann[1,*] and Frederic Koch[1,*]

[1]Max Planck Institute for Molecular Genetics, Department of Developmental Genetics, 14195 Berlin, Germany
[2]Max Planck Institute for Molecular Genetics, Sequencing Core Facility, 14195 Berlin, Germany

[*]Correspondence: tsaytler@molgen.mpg.de; herrmann@molgen.mpg.de; koch@molgen.mpg.de

Keywords: Cardiac mesoderm, Bmp4, differentiation, heart, WNT signaling, cardiac enhancers.

Summary statement

Bmp4 induces cardiac mesoderm formation in mESCs, whereby Bmp4 activates Nodal and Wnt3, and all three signaling pathways cooperate in activating mesodermal enhancers and cardiac progenitor genes.

Abstract

Cardiac lineage specification in the mouse is controlled by TGFβ and WNT signaling. From fly to fish, BMP has been identified as indispensable heart inducer. A detailed analysis of the role of Bmp4 and its effectors Smad1/5, however, were still missing. We show that Bmp4 induces cardiac mesoderm formation in murine ESCs *in vitro*. Bmp4 first activates *Wnt3* and up-regulates *Nodal*. pSmad1/5 and the WNT effector Tcf3 form a complex, and together with pSmad2/3 activate mesoderm enhancers and *Eomes*. They then cooperate with Eomes to consolidate the expression of many mesoderm factors, including *T*. Eomes and T form a positive feedback loop and open additional enhancers regulating early mesoderm genes, including the transcription factor *Mesp1* establishing the cardiac mesoderm lineage. In

parallel the neural fate is suppressed. Our data confirm the pivotal role of Bmp4 in cardiac mesoderm formation in the mouse. We describe in detail the consecutive and cooperative actions of three signaling pathways, BMP, WNT and Nodal, and their effector transcription factors, during cardiac mesoderm specification.

Introduction

In mammals, mesodermal cell lineages are derived from pluripotent epiblast cells that undergo epithelial to mesenchymal transition (EMT) in the primitive streak (PS) or tailbud, and ingress [1]. Cells originating from the proximal PS give rise to cardiac mesoderm progenitors, while the distal PS generates paraxial mesoderm [2]. Mesoderm formation is controlled by several inducers, in particular members of the TGFβ and WNT families, each activating its own signaling pathway and target transcription factors, primarily the essential regulators Eomes and T (Brachyury) [3-7].

TGFβ signaling in the PS is activated by Bmp4 and Nodal, and their downstream effectors - Smad transcription factors [8-10]. Both *Nodal* and *Smad2/3* (the Nodal effectors) deficient embryos do not form mesoderm and fail to gastrulate [9,11]. *Bmp4* deficient embryos display variable phenotypes; most mutant embryos do not proceed beyond the egg cylinder stage and form little or no mesoderm [8]. Furthermore, embryos lacking the Bmp4 receptor *Bmpr1a* do not form mesoderm and die before E9.5, suggesting that BMP signaling is essential for gastrulation [12]. Bmp4 induces phosphorylation of partially redundant Smad1/5 proteins [13]. *Smad5* mutant embryos showed delayed cardiac development and abnormal heart looping, but neither *Smad1* nor *Smad5* knockout prevented mesoderm formation *in toto* [14,15]. Similarly, a high proportion of *Smad1/5* double heterozygous mutant embryos displayed heart looping defects [13]. *Smad1/5* double homozygous mutant embryos have not been reported.

Smad1/5 and Smad2/3 interact with the common co-factor Smad4, allowing DNA binding and transcriptional regulation of target genes [10,16]. Embryos lacking *Smad4* do not form mesoderm and fail to gastrulate [17]. However, conditional depletion of *Smad4* in the early epiblast did not prevent cardiac

lineage specification and formation of a rudimentary heart, as well as somite formation, whereas primordial germ cells or an allantois did not form [3]. This data suggested that *Smad4* is dispensable for heart development in the mouse. In contrast, human *Smad4* mutant ES cells differentiated *in vitro* fail to form cardiac mesoderm or cardiomyocytes [18]. Moreover, in other vertebrates such as Xenopus, zebrafish and chick, BMP has been shown to play an essential role in cardiac mesoderm induction and heart development as well [19-21]. In Drosophila, the BMP homolog dpp is essential for heart induction too, pointing to a conserved role of BMP in heart development in evolution [22]. Therefore, it appears highly unlikely that BMP is dispensable for cardiac mesoderm formation in mouse.

TGFβ pathways crosstalk with other signaling cascades, and Smad TFs interact with various factors in a context-dependent manner [23,24]. WNT signaling, activated concurrently with TGFβ in the PS, is essential for proper PS and mesoderm formation [25]. WNT and TGFβ pathways have been shown to interact in various model organisms [26,27]. In differentiating ESCs, Smad2/3 and TCF factors co-occupy mesendodermal gene enhancers, suggesting cooperation between WNT and SMAD pathways in mesendoderm [28,29].

Early mesoderm formation is orchestrated by the T-box family members *Eomes* and *T* [30,31]. Although T is a pan-mesodermal transcription factor highly expressed in the PS and posterior mesoderm, it is not essential for cardiac mesoderm formation, as embryos lacking *T* develop beating heart structures [32]. Embryos with epiblast-specific *Eomes* deletion, on the other hand, fail to express cardiac marker genes, suggesting that *Eomes* is essential for specification of cardiac progenitors [30]. In addition, *Eomes* and *T* are required for suppressing the neural differentiation program in nascent mesoderm [33].

In this study, we have analyzed the process of Bmp4-induced early cardiac mesoderm specification in mESCs by functional characterization of the control mechanisms employed by the consecutively activated mesodermal key transcription factors Smad1/5, Smad2, Eomes and T.

Results

Cardiac progenitor genes are co-bound by pSmad1/5 and pSmad2, and require Smad4 for expression

In the mouse embryo, cardiac mesoderm is formed from cells located in the posterior epiblast, which are exposed to Bmp4 secreted from the adjacent extraembryonic ectoderm. However, in *in vitro* studies, cardiac mesoderm is commonly induced by Activin A (ActA) and/or WNT signaling [33,34]. We set out to generate cardiac mesoderm (ME) formation by exposing mESCs to Bmp4, thereby inducing processes occurring in the posterior PS (Fig. 1A; S1A,B). The dynamics of marker gene expression reflected mesoderm formation in the embryo. The pluripotency genes *Dppa3*, *Sox2* and *Esrrb* were downregulated after day 1 (Fig. S1C), followed by upregulation of nascent mesoderm control genes *Nodal, Eomes*, *Mixl1*, *Wnt3*, and *T* on day 2 and 3. The latter was paralleled by upregulation of the EMT genes *Zeb2* and *Snai1*, as well as of the lateral mesoderm and cardiac progenitor genes *Mesp1*, *Foxf1*, *Kdr*, *Gata4*, *Hand1*, and *Isl1* on day 4 (Fig. S1C,D). Neuroectoderm (e.g. Sox1, Pax6), endoderm (Sox17, Sox7), trophoblast (Cga, Gcm1), allantois (Cdx4, Tie1), and hematopoietic (Tal1, Runx1) lineage genes were not or hardly expressed, confirming the specificity of the approach (Fig. S1E). On day 6, around two third (65.3%) of the cells expressed the cardiomyocyte-specific gene Tnnt2 (Fig. S1F), and on day 8, all of the colonies were contracting and expressed markers of committed cardiomyocytes, Actc1 and Myl7, reflecting the high efficiency of the differentiation scheme (Fig. S1G, Video S1). Thus, our *in vitro* system faithfully copied mesoderm induction, EMT and cardiac mesoderm formation in the mouse embryo.

We assessed Smad1/5 and Smad2/3 activation by monitoring their phosphorylation levels, which peaked on day 2 or day 3, respectively (Fig. S1H). To gain insight into the mechanisms of SMAD-mediated transcriptional regulation, we first detected DNA binding sites (peaks) of Smad TFs in wild type (WT) cells on differentiation day 2 (ME, day 2) using ChIP-seq, yielding 7,575 phospho-Smad1/5 (pSmad1/5) and 12,423 phospho-Smad2 (pSmad2) peaks (Fig. S1I). Upon phosphorylation, pSmad1/5 and pSmad2/3 form complexes with Smad4, essential for regulation of target genes [16]. Therefore, depletion of *Smad4* is expected to reduce or eliminate DNA-bound

pSmad1/5 and pSmad2 complexes. Indeed, using Bmp4-treated Smad4 knock out (KO) cells, we detected only 493 pSmad1/5 and 554 pSmad2 peaks. These peaks showed low peak-calling confidence and were excluded from further analyses (Fig. S1I). 4418 (58%) of 7501 pSmad1/5 and 8425 (68%) of 12342 pSmad2 peaks were located outside of gene promoters indicating putative enhancer regions (Fig. 1B). To substantiate this assumption we plotted the corresponding profiles of enhancer-associated chromatin marks detected in the posterior PS of the E7 mouse embryo (ref. 35; Fig. S1J,K). A large fraction of the promoter-far binding sites of pSmad1/5 and pSmad2 were flanked by H3K27Ac and H3K4me1 marks, confirming that the majority of pSmad binding sites in ME cells are located in active enhancers. pSmad1/5 and pSmad2 peaks were enriched in previously reported GC-rich Smad1 or FoxH1 motifs, respectively, validating the specificity of the antibodies (ref. 36, Fig. 1C). Notably, both peak sets displayed strong enrichment of the Tcf/Lef motif associated with WNT signaling. The Oct4/Sox2 motif was significantly enriched as well, pointing to occupancy of these enhancers by pluripotency factors prior to differentiation.

Next, we compared the target genes of the two SMAD pathways and assessed their biological functions using Gene Ontology (GO) term analysis. Whereas genes bound by either pSmad1/5 or pSmad2 were only enriched in GO terms associated with nervous system development, genes co-bound by both pSmad1/5 and pSmad2 were also enriched in WNT signaling, mesoderm and heart development, and EMT terms (Fig. 1D, Table S1). This suggests that induction of the cardiac mesodermal program and EMT is mediated by the integration of both, BMP and Nodal signaling.

We identified putative direct target genes of SMAD by intersecting pSmad1/5 and pSmad2 bound genes with genes differentially expressed (DE) between Smad4 KO and WT ME cells. *Smad4* depletion disrupted the majority of pSmad1/5 and pSmad2 DNA binding (Fig. S1I), and neither Bmp4 nor ActA induced Eomes or T in KO cells (Fig. 1E, Fig. S1L). Therefore, in *in vitro* differentiated cells, loss of *Smad4* blocks both BMP and Nodal signaling, and thus allows the identification of direct target genes of both Smad1/5 and Smad2 (Fig. S1M).

Most putative direct targets were co-bound by pSmad1/5 and pSmad2 (Fig. 1D). Using GO term analysis, direct targets of pSmad1/5 and pSmad2 (Smad1/5/2) again were enriched in WNT signaling, mesoderm formation and heart development terms, whereas repressed targets were associated with nervous system development (Table S2). In contrast, direct target genes of either pSmad1/5 or pSmad2 alone did not show enrichment in any biological functions. Furthermore, the mesoderm control genes *Eomes*, *T*, *Tbx3*, *Fgf8*, *Mixl1*, *Evx1*, and *Hand1*, and WNT family members *Wnt3*, *Wnt8a*, and *Wnt5b*, were among the most strongly activated Smad1/5/2 targets, and their expression was almost entirely abolished in Smad4 KO cells. In contrast, the neural fate genes *Sox1*, *Sox2, Tal2,* and *Pax6* were among the most strongly repressed Smad1/5/2 targets (Fig. 1E,F; Fig. S1M).

Since *Wnt3* displayed transient upregulation during ME differentiation (Fig. S1C), and together with *Wnt8a* and *Wnt5b* was activated by Smads, we assessed whether all WNT signaling is abolished in the absence of SMAD signaling. We found that nuclear accumulation of active Ctnnb1 (b-catenin) peaked in day 2 WT ME cells and was strongly reduced in Smad4 KO cells (Fig. 1G). Thus, in Bmp4-induced mesoderm WNT activation requires SMAD signaling.

The combined data demonstrate that during mesoderm formation *in vitro* Smad4 is essential for transcriptional activity of pSmad1/5 and pSmad2, and suggest a crucial role of pSmad1/5/2 cooperation in the activation of early cardiac progenitor and repression of neural genes.

Cardiac progenitor and cardiomyocyte formation require pSmad1/5, pSmad2/3, and WNT signaling

Next, we addressed the effects of BMP4 and Nodal signaling on *Wnt3* and *Nodal* induction by selectively inhibiting the Smad1/5 pathway with LDN193189 (LN), or the Smad2/3 pathway with SB431542 (SB) (Fig. S2A). *Wnt3* expression was abolished in LN treated cells already one day after BMP4 induction, whereas SB treatment reduced *Wnt3* levels only on day 2 (Fig. 2A). This result is corroborated by the observation that nuclear accumulation of Ctnnb1 on day 2 is reduced in LN but not SB treated cells (Fig. S2B). Our data is in agreement with the finding that recombinant Bmp4

is sufficient to induce *Wnt3* in isolated embryonic epiblast explants [37]. We conclude that *Wnt3* is induced by Bmp4, possibly via the pSmad1/5 binding enhancer located in the second intron of the *Wnt3* gene (Fig. 1F).

Similarly, *Nodal* expression is reduced in LN treated cells on day 1 and further on day 2. In contrast, in SB treated cells, *Nodal* is strongly reduced only on day 2 (Fig. 2A). Wnt3 has been shown to enhance *Nodal* expression in the epiblast [37]. Accordingly, inhibition of WNT with XAV939 (XAV) resulted in reduced *Nodal* levels on day 3 (Fig. 2B). However, activation of WNT with CHIR99021 (CHIR) in the absence of SMAD signaling (LN/SB/CHIR) was not sufficient to restore *Nodal* expression (Fig. 2B). These data suggest that *Nodal* upregulation requires both BMP and WNT activity.

Both LN and SB treatment alone abolished *Eomes* expression on day 2, just as the Smad4 KO did (Fig. 2A). Different inhibitors of BMP or Nodal signaling confirmed the effect (Fig. S2C). Thus, pSmad1/5 is required for Eomes expression. SB treatment did not affect pSmad1/5 or WNT activity (Fig. S2A,B). Therefore, our data confirm the previous finding that pSmad2/3 activity is essential for *Eomes* expression [38]. Similarly, inhibition of WNT resulted in strongly reduced *Eomes* levels, whereas WNT activation upon SMAD inhibition with LN/SB was not sufficient to rescue *Eomes* expression (Fig. 2B,C). Thus, WNT signaling is also required but not sufficient for *Eomes* induction. Furthermore, activation of WNT signaling by CHIR in the presence of LN only slightly restored *Eomes* expression (Fig. 2D,E; Fig. S2D). Together, these data show that WT levels of *Eomes* expression require BMP signaling via pSmad1/5 and WNT signaling, in addition to pSmad2/3 activity.

Next, we assayed the effects of pSmad1/5, pSmad2/3 and WNT on *T* expression (Fig. 2B-E; Fig. S2D). Inhibition of either pathway alone resulted in strong down-regulation of T on day 3, and WNT induction by CHIR was not sufficient to rescue T expression to WT expression levels neither in LN nor in SB treated cells. Thus, proper *T* expression also requires pSmad1/5, pSmad2/3, and WNT signaling.

In line with these data, expression of the Eomes target and cardiac progenitor gene *Mesp1* [39], as well as of the cardiomyocyte-specific marker *Actc1,* were strongly down-regulated by LN or SB treatment (Fig. 2F). However, LN/CHIR or XAV treated differentiated cells showed low levels of *Mesp1* and *Actc1*

expression. Of note, none of these conditions yielded contracting aggregates on differentiation day 8, except for the control and for LN/CHIR treatment, the latter in about 10% of the colonies (Fig. 2G). This data indicates that WNT signaling can partially compensate for the loss of BMP signaling in producing cardiac mesoderm, albeit at low efficiency compared to control conditions.

Together, our data demonstrate that activities of pSmad1/5 and WNT signaling are required for proper cardiac mesoderm induction and efficient differentiation into contracting cardiomyocytes, in addition to pSmad2/3.

Smad transcription factors activate mesodermal enhancers

To gain insight into the mechanism of SMAD-mediated transcriptional regulation, pSmad1/5 and pSmad2 binding sites were examined for the chromatin accessibility state in day 2 ME cells. ATAC-seq profiling revealed that sites bound by pSmad1/5 alone had higher accessibility than those bound by pSmad2 (Fig. S3A). However, the strongest correlation between pSmad binding and chromatin opening was observed at sites co-occupied by pSmad1/5/2 (Fig. S3A). To examine the effect of pSmad binding on global chromatin accessibility, we compared ATAC-seq profiles in WT and Smad4 KO ME cells on day 2 and detected a total of 4731 differentially accessible (DA) regions (Fig. S3B). To quantify the occupancy of TFs at DA regions we defined six groups. We detected DA regions bound by pSmad1/5-solo, pSmad1/5/2, or pSmad2-solo. The remaining DA regions were grouped into those bound by Oct4, Sox2 and Nanog in mESCs (OSN), by one or two of these factors (O/S/N), or by none of the factors (None) (Fig. 3A). The majority of pSmad1/5-solo, pSmad1/5/2, or pSmad2-solo bound regions, but only a small fraction of O/S/N or OSN bound regions were activated by Smads (Fig. 3B). Notably, DA sites were mainly located at distal putative enhancer regions (more than 10kb from TSS) (Fig. S3C). We then plotted the TF occupancy and ATAC-seq density at pSmad1/5-solo, pSmad2-solo, and pSmad1/5/2 regions in WT and Smad4 KO ME cells (Fig. 3C). Activated regions were strongly bound by the corresponding Smad TFs at day 2 of differentiation, repressed regions displayed an enrichment of pluripotency TF binding in mESCs, suggesting that Smads repress OSN-occupied sites upon differentiation (Fig. 3C; Fig. S3D).

To gain insight into the mechanism of SMAD-mediated opening of enhancers, we scanned the largest groups (Smad1/5/2 and Smad2-solo) for TF motifs and found that activated regions were enriched in SMAD and Tcf motifs suggesting direct integration of TGFβ and WNT signaling (Fig. 3D). To substantiate this finding, we made use of Tcf3 ChIP-seq data derived from mesendodermal cells [28] and found that Tcf3 binding was strongly enriched at activated regions of the pSmad1/5/2 cluster (Fig. 3C; Fig. S3D). Notably, using immunoprecipitation we showed that pSmad1/5 and Tcf3 physically interact, suggesting that pSmad1/5 associates with Tcf3 at mesodermal and mesendodermal enhancers, and both cooperatively activate mesodermal gene expression (Fig. 3E).

We monitored the accessibility changes of SMAD-dependent DA regions by ATAC-seq on days 1 to 3 of the differentiation course (Fig. 3C). Activated loci exhibited low accessibility in ESCs and transiently opened during differentiation, reaching the peak on day 2, with the highest chromatin accessibility levels detected at pSmad1/5/2 sites (Fig. 3F). Compared to pSmad1/5/2 sites, the 410 pSmad2-solo sites displayed low levels of Tcf3 binding (Fig. S3D) and only a moderate increase in accessibility levels on day 2 (Fig. 3C,F). These sites were associated with 739 genes enriched in GO terms related to cell migration (49 genes), endoderm (9), and nervous system development (51) (Fig. 3G, Table S3). Most of the latter 51 genes were not expressed (Table S3), and the associated 52 sites displayed reduced accessibility compared to all other (358) pSmad2-solo regions (Fig. S3E). In contrast, genes associated with pSmad1/5/2 bound sites were enriched in mesoderm and heart development, and in WNT pathway GO terms (Fig. 3G). Genes coupled to activated Smad1/5/2/Tcf3 enhancers comprised nascent mesoderm control factors, such as *Eomes*, *T*, *Wnt3*, *Fgf8*, *Mixl1*, and *Kdr*, strongly upregulated during early stages of differentiation (Fig. S1C), (ref. 40; Fig. 3C; Table S3). The comparison of active enhancer-close genes with pSmad1/5/2 direct targets revealed 63 SMAD-activated genes (Fig. S3F,G) related by GO analysis to mesoderm development, WNT signaling and heart morphogenesis terms, comprising many of the most strongly activated mesodermal targets of pSmad1/5/2, e.g. *Id1*, *Eomes*, *T*, *Pitx2*, *Mixl1*, *Fgf8*,

Evx1, *Lefty1/2*, *Wnt3*, *Wnt5b*, *Tbx3*, *Axin2*, *Nodal*, *Irf1*, and *Foxb1* (Fig. 3H; Fig. S3H,I; Table S3).

To summarize, we show that upon cardiac mesoderm induction with BMP4, pSmad1/5 forms a complex with Tcf3, which co-localizes with pSmad2 at mesodermal enhancers. These enhancers are associated with the strongest direct targets of pSmad1/5/2 and undergo a SMAD-dependent opening during the course of differentiation, correlated with activation of mesodermal gene expression.

Cooperation of BMP, Nodal and WNT signaling is pivotal to opening of mesodermal enhancers

We then set out to determine the specific roles for BMP, Nodal and WNT signaling in the activation of mesodermal enhancers. We chose to monitor accessibility levels of 409 DA Smad4-dependent enhancers, which are co-occupied by pSmad1/5, pSmad2, and Tcf3, and display the strongest opening during mesoderm formation (Fig. 3C,F). To assess the effect of WNT signaling, we induced WT and Smad4 KO cells with BMP4 followed by inhibition of the WNT pathway with XAV, 18 hours post induction. Quantification of the ATAC-seq density revealed that WNT inhibition significantly reduced enhancer accessibility, however not to the level of the Smad4 KO (Fig. 4A,C). Activation of WNT with CHIR in Smad4 KO cells increased enhancer accessibility, but was not sufficient to restore accessibility to WT levels. Thus, binding of WNT effectors to mesodermal enhancers contributes to their opening, but without SMAD signaling is not sufficient to establish average accessibility levels equal to WT. This result is in line with our data showing only partial rescue of cardiac gene expression and cardiomyocyte differentiation by WNT signaling in the absence of pSmad1/5 activity (Fig. 2B-G).

Furthermore, inhibition of either pSmad1/5 (LN) or pSmad2/3 (SB) resulted in closure of mesodermal enhancers, with average accessibility levels similar to Smad4 KO (Fig. 4B,C). WNT activation in combination with individual pSmad1/5 (LN/CHIR) or pSmad2/3 (SB/CHIR) inhibition increased enhancer opening to some degree, but not sufficiently to achieve accessibility levels observed in control WT cells (Fig. 4B,C). Together, these results demonstrate

that BMP and Nodal branches of SMAD signaling as well as WNT signaling are required for opening of mesodermal enhancers and subsequent induction of cardiac mesoderm.

Upregulation of cardiac progenitor genes involves a positive feedback loop between Eomes and T

Above we showed that WNT and Smads cooperatively induce *Eomes* and *T* expression. Similar to their expression in the posterior part of the PS, Eomes and T expression patterns overlap *in vitro*, with Eomes reaching its highest levels on day 2 and T on day 3 (Fig. S1H). To provide insight into Eomes function on day 2, we first examined its DNA binding pattern, using Eomes KO cells to detect and discard false-positive peaks (Table S3). Eomes binding sites were enriched in Oct4/Sox2 and Eomes (T-box) motifs, and approximately half the Eomes peaks (5168) were localized at promoter-far sites (Fig. 5A). In the posterior PS of the E7 embryo these sites are associated with H3K27Ac and H3K4me1 marks, suggesting that they represent active and/or poised enhancers (ref. 35; Fig. S4A). We identified 371 putative direct target genes of Eomes on day 2 (Fig. S4B). Among the activated targets were cardiac progenitor genes *Fgf8*, *Hand1*, *Nrcam*, *Mixl1*, *Apln* [40], and genes involved in cell migration and circulatory system development, whereas repressed targets included the neural regulators *Sox2*, *Pax6*, *Tal2*, and *Sox1*, and were most significantly enriched in the GO term central nervous system development (Fig. S4C,D).

The overlap of activity levels, similarities in biological functions of direct target genes, and enrichment of the Oct4/Sox2 motif in the binding sites suggested that in day 2 ME cells Smads and Eomes function cooperatively. In support, Eomes has been shown to physically interact with Smad1/5 and Smad2/3 in differentiated human (h) ESCs [41]. Accordingly, Eomes peaks showed significant overlap with pSmad peaks, and the majority of pSmad1/5/2-bound genes also displayed Eomes binding (Fig. S4E,F). Yet, the putative direct targets of pSmad1/5, pSmad2 and Eomes showed only a moderate overlap on day 2 (Fig. 5B), whereas on day 3 most of the Smad1/5/2 direct targets were associated with Eomes binding and were down-regulated in Eomes KO cells (Fig. S4G,H; Table S3). We suggest that this moderate overlap on day 2

might be due to compensation of the loss of Eomes by Smads on day 2, when SMAD activity is at its maximum.

Next, to gain insight into the role of T in early cardiac mesoderm formation, we examined its DNA binding pattern in day 3 ME, using T KO cells to detect and discard false-positive peaks (Table S3). We detected 26631 binding sites, of which 19869 (74%) are located outside the promoter at putative enhancers, as supported by the enrichment of active enhancer chromatin marks associated with a large fraction of sites occupied by T in the PS of E7.5 embryos (ref. 35; Fig. 5C; Fig. S4I). We detected 530 putative direct targets of T (Fig. S4J), most of which were positively regulated, suggesting that at day 3 of differentiation T mainly acts as an activator of transcription (Fig. S4K). Activated T targets are involved in mesoderm development, WNT signaling, cell migration, and circulatory system development, and include cardiac progenitor genes, e.g. *Eomes*, *Fgf8*, *Mixl1*, *Rspo3*, *Mesp1*, *Apln*, *Evx1*, *Lhx1*, *Isl1* and *Aldh1a2* (Fig. S4K,L).

Apart from the palindromic T motif, T peaks were also strongly enriched in the Eomes motif (Fig. 5C). The comparison of their DNA occupancy revealed that 4545 (45%) of Eomes day 2 peaks were also occupied by T on day 3 (Fig. 5D). Moreover, 8169 (88%) of Eomes-bound genes were also bound by T. 91 direct targets were common between Eomes on day 2 and T on day 3, and another 144 on day 3 (Fig. 5E). Unique T or Eomes targets were associated with distinct but related functions (Fig. 5F). Genes co-activated by T and Eomes play a role in gastrulation, mesoderm formation, cell migration, WNT signaling pathway, and circulatory system development. These common direct targets comprise most of the cardiac mesoderm regulators, among them Mesp1, whose expression depends on both Eomes and T (Fig. S4H). Our data indicate that in day 3 ME cells, Eomes and T function mostly cooperatively, but have additional separate roles.

To address the mechanism of cooperative target gene regulation, we identified DE genes between day 3 Eomes/T double KO (dKO) and WT cells. T- and Eomes-bound DE genes in dKO cells (1018 genes) were grouped in 7 clusters based on their expression patterns in WT, single and double KO cells (Fig 5G; Table S3). Genes in clusters 3 and 6 are regulated by Eomes alone. Clusters 2, 4, 5 and 7 contain genes, which are similarly either activated or

repressed by Eomes or T. Repressed genes in cluster 5 include the neural fate marker *Pax6* and the Bmp4 inhibitor *Nog*. Cluster 1 genes require the activity of both Eomes and T, since either E-KO or T-KO alone causes downregulation relative to wild-type to a similar degree as observed in dKO. These include most of the critical cardiac progenitor genes, such as *Mesp1*, *Eomes*, *T*, *Mixl1*, *Fgf8*, *Wnt3*, *Evx1*, *Isl1* and *Aldh1a2.* In contrast, *Foxa2*, expressed besides other cell types also in a small subpopulation of early cardiovascular progenitors, is regulated by Eomes alone, in line with roles for Eomes and Foxa2 in endodermal lineage specification (ref. 42,43; Fig. 5G). Of note, we show here that the cardiac lineage transcription factor Mesp1 requires both Eomes and T for its full activation (Fig. 5G,H). Importantly, Eomes and T directly up-regulate each other in day 2 and day 3 ME cells (Fig. 5G; Fig. S4M). Nevertheless, down-regulation of Eomes in T KO cells is not complete (Fig. S4M), and differentiated T KO cells express *Actc,* however at a low level, and only 30% of the colonies are contracting on day 8 of differentiation, suggesting that cardiac mesoderm formation is affected by loss of T (Fig. 5H,I).

In summary, we show that in Bmp4-induced cardiogenic mesoderm Eomes activates and forms a positive feedback loop with T. They control partially overlapping sets of target genes. Our data confirm the pivotal role of Eomes for proper cardiac mesoderm development, but in addition suggest an important novel role for T, as major cardiac progenitor genes including Mesp1 require both transcription factors for WT expression levels, and the T KO impairs cardiomyocyte development *in vitro*.

Eomes and T cooperatively activate mesoderm enhancers

Recent studies have shown that T is required for chromatin remodeling of neuro-mesodermal progenitors (NMPs), and that Eomes and T open chromatin in mesendodermal cells [33,44]. To examine whether Eomes and T remodel chromatin also in early cardiac lineage cells, we performed ATAC-seq in E-KO, T-KO and dKO cells. Individual ablation of Eomes or T resulted in 4361 or 5366 DA regions, respectively, whereas dKO changed the accessibility of more than 12000 sites, emphasizing Eomes/T cooperation

(Fig. 6A; Fig. S5A-D). A large fraction of the DA regions was bound by Eomes and/or T, suggesting that their accessibility is controlled by these TFs directly (Fig. 6B). Notably, the vast majority of activated DA sites were located outside of gene promoters at putative distal (>10 kb) enhancer regions (Fig. S5A-D). Thus, Eomes and T cooperatively regulate the accessibility of a set of putative enhancers, of which the majority become active.

To assess the functional relevance of Eomes/T-mediated enhancer accessibility regulation, we determined the fraction of direct target genes located next to DA sites. While 36% (136 out of 370) of Eomes and 40% (215 put of 530) of individual direct T targets were located next to DA regions, for common targets there was an increase to 58% (dKO day 2; 305 out of 520) on day 2 and to 68% (dKO day 3; 696 out of 1018) on day 3, further supporting the finding of Eomes/T cooperation, in particular in day 3 ME cells (Fig. 6C). The majority of DA regions occupied by Eomes and/or T displayed positive regulation (Fig. 6D,E). Among the genes associated with the activated enhancer regions were cardiac progenitor (*T*, *Fgf8*, *Mixl1*, *Nanog*, *Lhx1*, *Lef1*, *Apln* and *Isl1*) and EMT (*Twist1*, *Prrx1*, *Crb2*, *Cdh2*, and *Zeb2*) genes (Fig. 6D; Table S3). Of note, activated enhancers occupied by T were associated with markers of further differentiated cardiac mesoderm (*GATA6*, *Tbx20*, *Myl4*), suggesting that T might poise these enhancers for subsequent activation (Fig. 6D). While most of the T or Eomes activated regions are already open in ESCs, they undergo transient T and/or Eomes-dependent increase in accessibility during differentiation (Fig. 6F; Fig. S5E,F). Thus, in early cardiac progenitor cells, the majority of direct Eomes and T target genes, including cardiac progenitor genes, are regulated by individual or cooperative activation of enhancers by T and/or Eomes.

Discussion

Here we present the regulatory network controlling the first stages of cardiac mesoderm formation induced by Bmp4 in murine ESCs (Fig. 6G). We show that the Bmp4 effectors Smad1/5 are the prime TFs initiating the process, comprising a double feedforward mechanism involving WNT and Nodal signaling followed by *Eomes* activation, and a positive feedback loop between

Eomes and T, finally resulting in activation of *Mesp1* and the cardiac lineage program.

In the mouse embryo, cardiac mesoderm specification is initiated in epiblast cells located in the proximal PS, which are exposed to BMP4 signals secreted by the neighboring extra-embryonic ectoderm and express low Nodal activity. Bmp4 and Nodal activate Smad1/5 and Smad2/3, respectively, which form active TFs by hetero-trimerization with Smad4. Whereas Smad4 plays a central role in TGFβ signaling *in vitro*, its requirement for cardiac mesoderm formation in mouse embryos has been challenged [3,17,18]. Here, we show that cells lacking *Smad4*, irrespective of induction by Bmp4 or by ActA, fail to express mesodermal factors *in vitro* (Fig. 1E; S1L). Furthermore, *Smad4* ablation greatly reduced DNA binding of both pSmad1/5 and pSmad2 (Fig. S1I). These results underline the essential role of *Smad4* in BMP and Nodal signaling during cardiac mesoderm formation. In agreement with our results, *Smad4* has also been shown to be required for cardiac differentiation in hESCs [18]. In sharp contrast, cardiac lineage specification and formation of a rudimentary heart structure was not prevented by conditional Sox2-driven depletion of *Smad4* in the mouse epiblast [3]. However, Smad4 loss in mutant embryonic cardiac mesoderm has not been checked in this experiment, leaving the possibility that the conditional KO was not sufficiently effective. Alternatively, the discrepancy between our *in vitro* and the *in vivo* data might be explained by the fact that the PS of conditional Smad4 KO embryos expresses T, indicating that WNT signaling is active in the mutant [3], in line with the expression of *T* in Smad1/5 double KO cells upon induction of WNT signaling [45]. Eomes expression has not been examined in this context either. But as we have shown here, Eomes and T control each other in early cardiac mesoderm *in vitro*. Thus, it is conceivable that in the conditional Smad4 mutant embryo T might activate Eomes to sufficient levels allowing at least some cardiac mesoderm to form. Alternatively, T might be able to activate early cardiac mesoderm genes by itself and partially compensate for a possible lack of Eomes. The latter mechanism has been shown for ActA-induced mesendoderm [33]. In any case, the fact that Smad4 KO embryos develop only a rudimentary heart indicates that loss of SMAD activity does have a dramatic effect on heart development *in vivo* [3].

The requirement of Nodal and Smad2/3 in cardiac mesoderm specification is well established, whereas the role of Bmp4-activated Smad1/5 has not been well understood, mainly due to the lack of double Smad1/5 mutant embryos. Genetic studies have shown that Bmp4 activates WNT signaling in the PS *in vivo* [37]. Similarly, in a previous study, Bmp4-treated ESCs formed posterior PS-like cells, accompanied by enhanced *Wnt3* and *Nodal* expression *in vitro* [46]. Using selective inhibitors for the two SMAD branches, we show here that it is pSmad1/5 that specifically activates expression of *Wnt3* and *Nodal* at the early stages of cardiac mesoderm specification (Fig. 1F; 2A).

We observed that pSmad1/5 and pSmad2 co-localize with the WNT effector Tcf3 at enhancers of cardiac progenitor genes (Fig. 3C). It has been shown in hESCs that upon PS induction Smad2 interacts with Ctnnb1 [29], and that Smad2/3 and Tcf3 converge on regulatory regions of mesendodermal genes in mESCs [28]. Moreover, Smad2/3-bound enhancer regions displayed opening during ESC differentiation [38,47]. We report here that in Bmp4-induced mESCs pSmad1/5 physically interacts with Tcf3, and that the opening of mesodermal enhancers co-occupied by pSmad1/5/Tcf3 complexes coincides with pSmad1/5 and WNT activity (Fig. 3E; 4A-C). We found that inhibition of BMP signaling by LN only led to partial opening of mesodermal enhancers, low level expression of *Eomes*, *T* and *Mesp1*, and very low efficiency of functional cardiomyocyte formation (Fig. 2D-G; 4B-C). However, opening of mesodermal enhancers to WT levels, proper induction of the cardiac mesoderm fate and efficient formation of cardiomyocytes does not occur without concurrent pSmad1/5, pSmad2/3 and WNT activity (Fig. 2D-G; 4A-C).

Previous approaches of *in vitro* differentiation of ESCs to cardiac mesoderm have employed ActA, Activin/BMP or WNT signaling [33,34,48-50]. ActA has been shown to activate expression of *Wnt3* [49], and induction of WNT signaling resulted in expression of ligands and receptors of both Nodal and BMP pathways [50]. Thus, apparently all induction schemes involve all three pathways in one or another way.

Several enhancers controlling *Eomes*, in particular the proximal enhancers PSE and VPE, and an enhancer located 45 kb upstream of the *Eomes* promoter have been identified in endoderm cells [51]. PSE was dispensable,

and VPE deletion resulted in markedly reduced Eomes expression, suggesting that optimal *Eomes* expression requires other enhancers in addition [51]. In ME cells, we found that VPE was still accessible upon depletion of SMAD and/or WNT signaling (Fig. 4B). In contrast, the -45 kb enhancer, which displayed the strongest binding of pSmad1/5 and pSmad2 at the *Eomes* locus, was closed upon ablation of *Smad4* as well as by inhibition of pSmad1/5 or pSmad2/3 activity, and its accessibility was reduced upon inhibition of WNT signaling (Fig. 4B). Collectively, these findings suggest that in Bmp4-induced nascent mesoderm, pSmad1/5, pSmad2, and WNT signaling together activate *Eomes* transcription via the -45 kb enhancer.

Eomes plays essential roles in several embryonic and extraembryonic tissues, and its specificity is conferred upon by context-dependent cooperation with other TFs [5,6,30,52,53]. In the anterior PS, Eomes cooperates with Smad2/3 to promote mesendoderm and definitive endoderm formation [30,51,54]. In hESC-derived mesoderm, Eomes physically interacts with Smad1/5 and Smad2/3 [41]. We show here that in ME cells, pSmad1/5, pSmad2, and Eomes cooperate to promote the cardiac fate, whereby on day 2 SMAD activity is sufficient to compensate for the loss of *Eomes*, whereas on day 3 Eomes becomes essential (Fig. 5B; S4E-H). Therefore, pSmad1/5 and pSmad2/3, together with the effectors of WNT signaling, first activate *Eomes*, and then cooperate with Eomes to promote expression of mesoderm factors, such as *T*, *Fgf8*, *Evx1*, *Mixl1*, and *Kdr*. This feedforward mechanism has also been referred to as self-enabling mechanism of SMAD [55].

In the PS, *Eomes* expression precedes and partially overlaps with the mesoderm TF T. Opposite modes of Eomes/T interaction have been reported: in the posterior PS Eomes activates *T* [51,56], whereas in the anterior PS-derived endoderm Eomes represses *T* [57]. A recent study demonstrated that in ActA-treated mESCs, Eomes induced the mesendodermal program, but upon *Eomes* ablation, *T* was upregulated and compensated for loss of *Eomes* [33]. In contrast, we show that in Bmp4-induced mesoderm cells, Eomes and T form a positive feedback loop (Fig. 5G; S4M). Further, we found that Eomes and T cooperate to open a large set of enhancers regulating cardiac progenitor genes (e.g. *Mesp1*, *Fgf8*, *Apln*, *Mixl1*, *Lhx1*, *Isl1*) and EMT (e.g. *Twist1*, *Zeb2*, *Cdh2*, *Crb2*) genes (Fig. 6A-F). Moreover, concurrent

expression of *Eomes* and *T* is required to activate major cardiac mesodermal fate genes, including *Mesp1*, previously only shown to be regulated by either Eomes or T (ref. 30,33,58; Fig. 5G,H). Conforming with the formation of abnormally looped heart structures observed in *T* mutant embryos [32], we show that T-KO cells expressed low levels of *Actc1*, and only 30% of the colonies contracted on differentiation day 8 (Fig. 5H,I). In agreement with this data, *T* ablation resulted in down-regulation of *Eomes* expression, confirming the effect of T on Eomes expression (Fig. S4M). The combined data therefore support our conclusion that T and Eomes cooperate during cardiac mesoderm formation in a positive feedback loop. We propose that the latter is true also for early cardiac mesoderm formation in the embryo. Collectively, these findings provide a new facet of *Eomes* and *T* interaction in controlling differential mesodermal/endodermal lineage specification.

We show that apart from promoting the cardiac fate in ME cells, SMAD signaling, Eomes and T also repressed neural fate genes, in particular *Sox1*, *Sox2, Otx2, Pax6,* and *Tal2* (Fig. 1E,F; 5F; S4D,L). It has been shown that during trunk development T promotes the mesodermal fate of NMP descendants and acts antagonistically to Sox2 driving the neural fate [44]. In addition, neural fate suppression by Eomes and T has also been demonstrated in ActA-induced early ME differentiation of mESCs [33]. Therefore, Sox2 is antagonized by mesoderm promoting TFs consecutively during development, starting with pSmad1/5/2 at the onset of gastrulation in the epiblast, followed by Eomes and T during head, trunk and tail development. Thus, activation of the mesodermal program and parallel suppression of the neural fate by mesoderm control genes take place throughout development.

Together, we show that Bmp4 induces cardiac mesoderm formation in murine ESCs, confirming a pivotal role for BMP in heart development in mouse. We present a model comprising a double feedforward mechanism involving Bmp4/Smad1/5, Wnt3/Tcf3, and Nodal/Smad2/3, which need to cooperate for opening mesoderm enhancers and effecting expression of mesodermal genes, including *Eomes* and *T*. The combined action of all three signaling pathways and their target mesoderm control factors Eomes and T is needed for cardiac lineage specification and proper cardiac mesoderm development. None of the factors can be omitted without affecting cardiac mesoderm formation in vitro, suggesting that heart development in the embryo employs the

same mechanisms. In sum, our data provide a detailed mechanistic model of the regulatory network controlling cardiac mesoderm formation in mouse (Fig. 6G).

Materials and methods
RESOURCE AVAILABILITY
Lead contact:

Information and requests for resources and reagents should be directed the lead contact, Bernhard G.

Herrmann (herrmann@molgen.mpg.de).

Materials availability:

Please contact Bernhard G. Herrmann for requests and inquiries.

EXPERIMENTAL MODEL AND SUBJECT DETAILS
Smad4 WT (conditional alleles) and knock-out mESCs were generated from mice kindly provided by Elizabeth Robertson. F1G4 mESCs were used to generate Eomes, T, and Eomes/T double KO cell lines. All cell lines were regularly tested for mycoplasma contamination.

Culture of mouse embryonic stem cells (mESCs)

All mESC lines were maintained on plates coated with 0.1% gelatin (Sigma-Aldrich; G1393) on a layer of mitotically inactive primary mouse embryo fibroblasts (feeder cells) in ESC growth medium: knockout Dulbecco's Modified Eagle's medium (DMEM) (ThermoFisher; 10829018) supplemented with 2mM L-glutamine (Lonza; BE17-605E), 15% (v/v) heat-inactivated FCS, 1% (v/v) non-essential amino acids (ThermoFisher; 11140050), 0.1 mM β-mercaptoethanol (Sigma-Aldrich; M3148), 1% (v/v) nucleosides (Sigma-Aldrich; ES-008-D), 1% (v/v) penicillin and streptomycin (Lonza; DE17-603E), and 1000 U/ml LIF (Chemicon; ESG1107) at 37°C and 7.5% CO_2 .

Generation of KO and reporter mESC lines
Wild type mESCs were from a F1G4 background [59]. Conditional Smad4 knock-out (KO) cells (Smad4CA) were provided by Prof. Dr. Elizabeth J. Robertson [3]. To generate Smad4 KO cells, the first exon of Smad4 gene flanked by loxP sites was excised by transient expression of Cre-recombinase in Smad4CA cells, followed by expanding single clones on 96-well plates and

screening for homozygous clones by PCR using primers outside of the deleted region.

Eomes and T KO cell lines were generated using double-nicking CRISPR/Cas9 approach with Cas9(D10A) nickase mutant and two pairs of guide RNAs surrounding the promoter and the first exon of the corresponding gene [60]. Single stranded oligonucleotides (Supplementary Information, Table S4) were annealed and cloned into the BbsI site of px335A_hCas9_D10A_G2P plasmid (gift from Dr. Boris Greber). Wild type F1G4 cells were transfected with a modified plasmid, and transiently selected with 1 µg/ml puromycin (Gibco; 10130127) for 2 days. Single clones were picked 7-9 days after transfection, plated onto 96-well plates and screened for genomic DNA deletions by PCR using primers outside of the deleted region, listed in Supplementary Information, Table S4. Double Eomes and T KO (dKO) cell line was generated from Eomes KO cells using the same approach as for generation of T KO cell line. The absence of Eomes and T at transcript and protein levels in the differentiated E-KO, T-KO, and dKO cells was confirmed using RT-PCR and western blotting, respectively.

Fluorescent reporter cell line was generated as previously described [44]. In short, we used the mouse T BAC (RP24-530D23) modified by replacing the starting codon of the T gene by a reporter gene containing H2B-mCherry, followed by the rabbit *b-globin* polyadenylation signal (provided by Manuela Scholze-Wittler). The modified T BAC was linearized, followed by random integration into F1G4 cells. Single clones with BAC integration were detected by PCR.

In Vitro Differentiation

mESCs were deprived of feeder cells by dissociating in trypsin and passaging for four consecutive 25 minute periods, followed by a 24 hour period. Cells were then dissociated, and 40,000 cell/ml cell suspension was plated onto tissue culture plates in 5 µl droplets using electronic multi-channel pipettes. Plates were then inverted and incubated for 24 hours. Drops containing mES cell aggregates (corresponding to differentiation day 0, or ES) were then pooled, washed twice with DMEM (ThermoFisher; 12491015), and resuspended in differentiation medium: 48% (v/v) DMEM/F12 (ThermoFisher;

11320033), 48% (v/v) Neurobasal medium (ThermoFisher; 21103049), 1% (v/v) 200 mM L-glutamine (Lonza; BE17-605E), 1% (v/v) penicillin and streptomycin, 1% (v/v) B27 (Gibco; 17504044), 0.5% (v/v) N2 (Gibco; 17502048), 0,5% (v/v) 7.5% BSA fraction V (Gibco; 15260037), 0.1 mM thioglycerol (Sigma-Aldrich; M1753) supplemented with 20 ng/ml mouse recombinant Bmp4 (R&D Systems; 5020-BP-010). To start differentiation, mESC aggregates resuspended in differentiation medium containing Bmp4 were immediately plated on fibronectin-coated plates (Calbiochem; 341631). Differentiation medium supplemented with Bmp4 was refreshed every 24 hours.

Alternatively, for ActA treatment, drops containing mES cell aggregates (corresponding to differentiation day 0, or ES) were pooled, washed twice with DMEM and resuspended in differentiation medium containing 20 ng/ml recombinant human/mouse/rat Activin A protein (R&D Systems; 338-AC-010). For selective inhibition of Smad signaling pathways, the mES cell aggregates were preincubated for 30 minutes with LN193189 (Sigma-Aldrich; SML0559), K0288 (Sigma-Aldrich; SML1307), SB431542 (Biogems; 3014193), or A-83-01 (Sigma-Aldrich; SML0788) at the indicated concentrations. Afterwards, the inhibitors we also added to the differentiation medium. For activation of inhibition of WNT signaling, CHIR99021 (TOCRIS; 4423) or XAV939 (Sigma-Aldrich; X3004) were added to the medium at 18 hours of differentiation at the indicated concentrations.

Antibodies

The antibodies used for western blotting: Phospho-Smad1/5 (Cell Signaling, 9516; RRID: AB_491015), Phospho-Smad2/3 (Cell Signaling, 3108; RRID: AB_490941), Smad1 (Cell Signaling, 9743; RRID: AB_2107780), Smad2/3 (Cell Signaling, 5339; RRID: AB_10626777), Eomes (Abcam, ab23345; RRID:AB_778267), T (rabbit polyclonal anti-Brachyury, custom), Tcf3 (Santa Cruz, sc-166411), GAPDH (Cell Signaling, 5174), H3 (Abcam, ab1791; RRID:AB_302613), active Ctnnb1 (Cell Signaling, 8814; RRID: AB_11127203). The antibodies used for immunoprecipitation: Tcf3 (Santa Cruz, sc-166411; RRID: AB_2302942), normal mouse IgG (Santa Cruz, sc-2025; RRID: AB_737182). The antibodies used for ChIP: Phospho-Smad1/5/9

(Cell Signaling, 13820; RRID: AB_2493181), Phospho-Smad2 (Cell Signaling, 18338; RRID: AB_2798798), Eomes (Abcam, ab23345), T (Santa Cruz, sc-17743; RRID:AB_634980). The antibodies used for immunofluorescence and Flow Cytometry: Myl7 (Proteintech, 17283-1-AP; RRID: AB_2250998; 1:250 dilution), Actc1 (Proteintech, 66125-1-Ig; RRID: AB_2881524; 1:200 dilution), Tnnt2 (Abcam, ab209813; 1:200 dilution), donkey anti-mouse IgG, Alexa Fluor 488 (ThermoFisher, A-21202; RRID: AB_141607; 1:500 dilution), donkey anti-rabbit IgG, Alexa Fluor 488 (ThermoFisher, A-21206; RRID: AB_2535792; 1:500 dilution).

Analysis of Protein Expression and Phosphorylation Levels by Western Blotting

Whole cell extracts were prepared from mESC or differentiating cells using 1X Novex NuPAGE LDS sample buffer (ThermoFisher; NP0007) supplemented with 100 mM DTT (Sigma-Aldrich; 43815). DNA was digested using Benzonase Nuclease (Millipore; E1014). Samples were heated at 95°C for 5 minutes, run on 4-12% NuPAGE Bis-Tris Protein Gel, transferred to nitrocellulose membrane, and subjected to western blotting using primary antibodies at 1:1000 dilution and the appropriate HRP-conjugated secondary antibodies at 1:5000 dilution. Detection was performed using the Amersham ESL reagents (GE Healthcare) and scanned using a Fusion SL chemiluminescent detection system (Vilber). At least two biological replicates were performed for each experiment. For nuclear extraction, Nuclear Complex Co-IP Kit (Active Motif; 54001) was used following the manufacturer's instructions.

Immunoprecipitation

Differentiated cells were collected and lyzed using Nuclear Complex Co-IP Kit. 2.5 µl primary Tcf3 antibody or mouse IgG antibody was added to the 300 µl of the nuclear fraction, followed by rotation overnight at 4°C. 50 µl protein G-coated Dynal beads were washed three times with PBS, resuspended in the nuclear fractions containing primary antibodies and rotated for 1 hour at 4°C. The beads were then washed 5 times with 1 ml of the cold IP washing buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100),

bound proteins were eluted using 100 µl 1X Novex NuPAGE LDS sample buffer and analyzed using Western Blotting.

FACS

To detect the percentage of cells expressing T during differentiation, we subjected F1G4 cells containing H2B-mCherry reporter to FACS analysis. Cells at various differentiation stages (ES/day 0, day 1, day 2, day 3, day 4, and day 5) were trypsinized, resuspended in 3% BSA in PBS and analyzed on a FACS Aria II (Becton Dickinson). For detection of cardiomyocytes expressing Tnnt2, differentiating cells on day 6 were trypsinized and counted. Aliquots containing 2e6 cells were washed with 1 ml PBS, incubated with 2% formaldehyde in PBS for 10 minutes at room temperature and washed twice with 1 ml PBS. The aliquots were then incubated in 0.5 ml PBS containing 0.2% Triton X-100, followed by incubation with the primary Tnnt2 antibody for 60 minutes at room temperature. After three washing steps with 1 ml PBS/0.1% Triton X-100, cells were incubated with the fluorescently labeled secondary antibody for 30 minutes at room temperature. For the control experiment, primary Tnnt2 antibody was omitted. After three washing steps with 1 ml PBS/0.1% Triton X-100, cells were analyzed on a FACS Aria II.

Immunofluorescence

For immunofluorescence, cells were washed two times with PBS and fixed in 4% PFA in PBS for 10 minutes at room temperature while shaking, followed by two washes with PBS for 5 minutes and one wash with 0.1% Triton X-100 in PBS. Cells were then incubated for one hour with 2% BSA in PBST (0.05% Tween 20 in PBS), rinsed with PBS, and incubated overnight with the primary antibodies diluted at 1:100 in PBST containing 0.2% BSA. After five washes with PBST, cells were incubated with the corresponding secondary antibodies (1:1000 dilution), washed five times with PBS. The images were collected with ZEISS Axio Observer Z1 microscope.

qRT-PCR analysis

For RNA extraction, cells were collected at the indicated times and processed with the RNeasy Micro kit (Qiagen). For each sample, 500 ng of total RNA was used for cDNA synthesis with QuantiTect Reverse Transcription Kit (Qiagen). Quantitative PCR was performed on a StepOnePlus Real-Time PCR System (Applied Biosystems) using Pmm2 as internal control gene for calculating relative expressions. Sequences of primers used for qRT-PCR are listed in Table S4. At least three biological replicates were performed for each experiment.

RNA extraction and RNA-seq Library Preparation

RNA extraction and library preparation was performed as previously described [44]. Briefly, total RNA was extracted from 40,000 mESCs or differentiated cells using Trizol reagent (ThermoFisher; 15596026) following the manufacturer's instructions and purified using the RNeasy Micro kit (Qiagen). Residual genomic DNA was digested on a column. The RNA was quantified using Qubit RNA HS assay (Life Technologies), and the integrity of the RNA was assessed using Bioanalyzer RNA pico chips (Agilent).

Strand-specific RNA-seq libraries were generated from 100 ng of total RNA using the ScriptSeq v2 (Epicentre) low input library preparation kit according to manufacturer's instructions. The library was amplified using 15 PCR cycles. RNA-seq libraries were quantified using the Qubit HS DNA assay (Life Technologies) and the size distribution was assessed using the DNA HS Bioanalyzer chips (Agilent). Libraries were pooled and paired-end sequenced on a HiSeq 2500 with 50 bp read lengths.

Chromatin Immunoprecipitation

ChIP experiments were carried out as previously described [44]. Briefly, crosslinking was performed directly on differentiating cells in differentiation medium with the addition of 1/10th volume of crosslinking solution (11% formaldehyde, 50 mM Hepes pH 7.8, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) for 10 minutes at room temperature, while shaking. The crosslinking reaction was quenched with the addition of 1/10th volume of 2.5 M glycine and 5 minutes incubation. Cells were washed twice with cold PBS, scraped in

cold PBS containing 0.05% Triton X-100, and pelleted in aliquots of 5×10^7. For sonication, complete protease inhibitors without EDTA (Roche) at 1x final concentration was added to all lysis buffers (LB) prior to use. Each pellet was resuspended in 2.5 ml LB1 (50 mM Hepes pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.75% NP-40, 0.25% Triton X-100) and rotated at 4°C for 20 minutes. The cell suspension was homogenized using a douncer. The chromatin was pelleted by centrifugation at 1400g and 4°C for 5 minutes and resuspended in 2.5 ml LB2 (10 mM Tris pH 8, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA). After 10 minutes rotation at 4°C, the centrifugation step was repeated and each pellet was resuspended in 1.5 ml LB3 (10 mM Tris pH 8, 1 mM EDTA, 0.5 mM EGTA, 100 mM NaCl, 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine), transferred to 15 ml Falcon tubes and sonicated using a W-450D Digital Sonifier (Branson) sonicator for 14 cycles of 10 seconds on/50 seconds off in a 4°C cold room with tubes chilled in ice water. After sonication, 150 µl of Triton X-100 was added per tube, transferred to two 1.5 ml Eppendorf tubes and debris was pelleted by centrifugation at 20,000g and 4°C for 10 minutes. The solubilized chromatin was then pooled and mixed thoroughly. After taking 50 µl as an input control, the chromatin was distributed into 1.5 ml aliquots, snap frozen and stored at -80°C until use.

The input was reverse-crosslinked with the addition of 50 µl 2x elution buffer (100 mM Tris pH 8, 20 mM EDTA pH 8, 2% SDS) and 13-15 hour incubation at 65°C. 100 µl of TE buffer were added, and RNA was digested using 0.2 µg/ml RNase A at 37°C for 2 hours. Proteins were digested with the addition of 0.2 µg/ml proteinase K and incubation at 55°C for 2 hours. The DNA was purified with two subsequent phenol:chloroform:isoamyl alcohol (25:24:1, pH 8) extractions and a subsequent MinElute (Qiagen) purification. The input DNA was quantified using a NanoPhotometer (Implen).

For ChIP, 100 µl protein G-coated Dynal beads (Life Technologies) were washed 3x with 1 ml of blocking buffer (PBS, 0.5% BSA), resuspended in 500 µl of blocking solution containing 5 µg ChIP antibody and rotated overnight at 4°C. Beads were then washed 3x with 1 ml of blocking buffer and resuspended in 100 µl blocking buffer. Chromatin equivalent to 5×10^7 cells was added and rotated overnight at 4°C. The following day, 6 washing steps (9 for T) with 1 ml RIPA buffer (50 mM Hepes pH 7.6, 500 mM LiCl, 1 mM

EDTA, 1% NP-40, 0.7% Na-Deoxycholate) and one washing step with 1 ml of TEN (10 mM Tris pH 8, 1 mM EDTA, 50 mM NaCl) were performed. The elution was performed in two subsequent steps using 100 µl of 1x elution buffer (50 mM Tris pH 8, 10 mM EDTA, 1% SDS) and incubation at 65°C while shaking for 10 minutes each. The eluates were combined and incubated for 13-15 hours at 65°C. 200 µl of TE were added and the ChIP DNA was purified as described for the input above. ChIP DNA was quantified using the Qubit (Life Technologies) DNA HS assay.

ChIP-seq Library Preparation

The ChIP-seq libraries were generated using the TruSeq ChIP library kit (Illumina) with the previously described modifications [44]. After adapter ligation, 0.95x of AMPure XP beads (Beckman Coulter) were used for a single purification and the DNA was eluted using 14 µl of resuspension buffer (RSB, Illumina). After the addition of 1 µl primer mix (25 mM each, Primer 1: 5'-AATGATACGGCGACCACCGA*G-3'; Primer2: 5'-CAAGCAGAAGACGGCATACGA*G-3') and 15 µl 2x Kapa HiFi HotStart Ready Mix (Kapa Biosystems), amplification was performed for 45 seconds at 98°C, 5 cycles of [15 seconds at 98°C, 30 seconds at 63°C and 30 seconds at 72°C] and a final 1 minute incubation at 72°C. The PCR products were purified using 0.95x of beads and eluted using 21 µl of RSB. Ligation products were then separated using a 1.5% agarose gel. Post-run staining was performed using SYBR Gold (Life Technologies) under agitation for 30 minutes. Gel slices corresponding to ~250-400 bp fragment size were cut out using a Dark Reader (Clare Chemical Research) transilluminator. The gel extraction was performed using 5 gel volumes of QG buffer (Qiagen) with the addition of one gel volume of isopropanol. The MinElute (Qiagen) columns were washed once with QG buffer and twice with PE buffer, air-dried for at least 10 minutes and eluted using 21 µl of EB buffer. 19 µl of the eluate was used in the final amplification, with the addition of 1 µl primer mix and 20 µl 2x Kapa HiFi Hotstart premix. The same protocol as for the pre-amplification was used, with the exception of using 13 amplification cycles. The libraries were quantified using the Qubit DNA HS assay and the library size was validated using DNA HS Bioanalyzer chips. Sequencing was performed on either the

HiSeq 2500 or NextSeq 500 (Illumina) using 2x 50 bp or 2x 75 bp read lengths, respectively.

ATAC-seq

ATAC-seq was performed as previously described [61]. The cells were trypsinized, and the trypsin was inactivated by washing with cold PBS containing 2% BSA. For each sample, 50,000 cells were collected, washed with 1 ml cold PBS and pelleted by centrifugation. The cells were lysed in 50 µl of cold lysis buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl$_2$, 0.1% IGEPAL CA-630) and immediately centrifuged at 500g, 4°C for 10 min. The pellet was resuspended in the transposition reaction mix (25 µl 2x TD buffer, 2.5 µl Tn5 transposase, 22.5 µl H$_2$O) and incubated at 37°C for 30 minutes. After incubation, the reaction was stopped with the addition of PB buffer (Qiagen) and the tagmented DNA was purified using the MinElute kit (Qiagen). The DNA was combined with the ATAC index PCR primers and 2x Kapa HiFi Hotstart Ready Mix and pre-amplified (98°C 30 seconds, 5x [98°C 10 seconds, 63°C 30 seconds, 72°C 1 minute] in a 50 µl volume. To determine the remaining cycles to avoid potential over-amplification, 5 µl of the pre-amplification mix was combined with the primers, 1x Evagreen SYBR green (Jena Biosciences) and 2x Kapa HiFi Hotstart Ready Mix in a 15 µl total volume and run for 30 cycles on a StepOne Plus. The remaining 45 µl of pre-amplified samples were amplified for a further 6-7 cycles and the libraries were purified using MinElute column (Qiagen). The libraries were quantified using the Qubit DNA HS assay and the library sizes were validated using DNA HS Bioanalyzer chips. Samples were pooled and paired-end sequencing was performed on either the HiSeq 2500 or HiSeq 4000 (Illumina) using 2x 50 bp or 2x75 bp read lengths, respectively.

Genome Assembly

All datasets were mapped to the mouse mm10 reference genome containing chromosomes 1-19, X, Y, and M.

Analysis of RNA-seq data

RNA-seq reads were subjected to quality control using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed using cutadapt (version 2.9) (https://doi.org/10.14806/ej.17.1.200) to remove adapter sequences. Alignment was performed using Hisat2 [62] (version 2.1.0) with the following arguments: *--no-mixed --no-discordant*. The index files for Hisat2 were built using mm10 reference genome containing chromosomes 1-19, X, Y and M, and splice sites and exon data from refSeq annotations. The resulting sam format files were then converted to bam format, sorted and indexed using samtools [63] (version 1.9). FPKM values were calculated using Cuffdiff [64] (version 2.2.1) with the following arguments: *-u -no-effective-length-correction -b*. To identify genes that showed differential expression (DE) upon depletion of Smad4, Eomes or T, normalized FPKM values were filtered, and only genes with FPKM greater than 1 in at least one of the samples were examined. Log_2 fold change values were calculated using normalized FPKM comparing WT to KO, and genes with log_2 fold change greater than 1 or less than -1 were selected.

For visualization, the reads from biological replicates were combined, and the data were normalized and converted to bigWig (bw) format using bamCoverage tool from deepTools [65] (version 3.1.3) with the following arguments: *--normalizeUsing BPM -bs 5 --ignoreDuplicates*. Genome browser snapshots containing RNA-seq data were generated using pyGenomeTracks [66] (version 3.1). The expression data were visualized in heatmaps using heatmap.2 function of gplots package (version 3.0.1) (https://CRAN.R-project.org/package=gplots) in R.

For clustering of genes regulated by T and/or Eomes, genes bound by T and DE in day 3 T-KO as compared to day 3 WT cells were combined with genes bound by T and DE in Eomes/T dKO cells as compared to day 3 WT cells, resulting in a gene list of 1018 genes. The k-means clustering into 7 groups was performed using Cluster3.0 [67], and visualized using Java TreeView [68].

Analysis of ChIP-seq data

For 75 bp read length ChIP-seq, the reads were first trimmed to 50 bp length. The reads were then aligned to the reference genome using bowtie2 [69] (version 2.2.4) with the following arguments: *-3 5 -I 100 -X 500 --no-discordant --no-mixed*. Sam files containing mapped paired-end reads were converted to bam files, sorted and indexed using samtools. For visualization, the data were normalized and converted to bigWig (bw) format using bamCoverage tool from deepTools with the following arguments: *--normalizeUsing CPM --extendReads 200 --smoothLength 9 -bs 3 -ignore chrM*. Peak detection was performed using MACS2 [70] (version 2.1.1.20160309) with q-value cutoffs of 1e-4 (for T), 5e-4 (pSmad1/5/9), 1e-7 (pSmad2), 5e-4 (Eomes day 2 and day 3), and 1e-2 for all TFs in the corresponding KO cells. MACS2-generated q-values were used to assess the peak-calling confidence. Genome browser snapshots containing ChIP-seq data were generated using pyGenomeTracks. ChIP-seq data from public repositories were aligned, and bw files were generated as described above. For Sox2, Oct4, and Nanog ChIP-seq data, peaks were called using MACS2 with q value cutoff of 1e-4.

To identify genes potentially regulated by TF binding sites, we first overlapped the ChIP-seq peak summits of every TF with promoters (transcription start sites +/-2 kb) of refSeq annotated genes using bedtools [71] (version 2.27.1). This identified genes containing promoter-associated peaks. The remaining peak summits were associated to the genes with the closest up- and down-stream promoters. These genes were combined with the genes containing promoter-associated peaks to form a total list of genes bound by a TF. A subset of TF-bound genes that displayed differential expression between WT and the corresponding TF KO cells, were called putative direct target genes of the corresponding TF.

For comparison of binding sites of various TFs, ChIP-seq peaks were called overlapped if peak summit distances were less than 300 bp. Distribution of peaks based on distances from the nearest promoters based on peak summit and TSS coordinates was calculated using bedtools.

For *de novo* motif discovery within ChIP-seq peaks, we extracted genomic sequences in 100 bp (+/-50 bp) regions around peak summits using bedtools and used the resulting fasta files as input for Meme (version 5.1.1) in the MEME suite [72]. To find known motifs enriched within defined subsets of ChIP-seq peaks, such as peaks overlapping with the differentially accessible ATAC-seq regions, genomic sequences in 120 bp (+/- 60 bp) regions around corresponding peak summits were used as input for MEME-chip in the MEME suite.

Clustered ChIP-seq heatmaps were generated using computeMatrix and plotHeatmap tools from deepTools. Grouping into clusters was performed based either on peak overlaps of various TFs or on appropriate clusters of ATAC-seq regions. Coordinates of peak summits within each cluster and normalized bw ChIP-seq files were used as input. Boxplots indicating distributions of ChIP-seq densities within heatmap clusters were generated using ggplot2 (version 3.1.1, https://ggplot2.tidyverse.org/) package in R. The values were obtained from the output file of *--outFileNameMatrix* argument of the computeMatrix tool run with additional argument *-bs 1500.*

Analysis of ATAC-seq data

All ATAC-seq experiments were performed with two biological replicates. ATAC-seq reads were trimmed to 40 bp length and adapter sequences were removed using cutadapt. The trimmed reads were aligned to the reference genome using bowtie [73] (version 1.1.2) with the following arguments: *-y -v 2 --best --strata -m 3 -k 1 -S -X 2000 --allow-contain*. Sam files containing mapped paired-end reads were converted to bam files, sorted and indexed using samtools. Reads mapped to chr M and Y or known ATAC artifact regions (ENCODE) were removed using samtools. Possible PCR duplicates were removed using Picard (version 1.103, https://broadinstitute.github.io/picard).

For visualization, the reads from biological replicates were combined, and the data were normalized and converted to bw format using bamCoverage tool from deepTools with the following arguments: *--normalizeUsing RPCG --effectiveGenomeSize 2652783500 --extendReads 200 --smoothLength 9 -bs*

3 -ignore chrY chrM. All accessible regions were detected as narrowPeaks using MACS2 with q value cutoffs of 5e-2.

To identify regions that are differentially accessible between WT and KO cells, we used DiffBind package (version 2.12.0, https://bioconductor.org/packages/release/bioc/html/DiffBind.html) in R. As input, for each replicate of WT and KO ATAC-seq data, bam file containing filtered mapped reads and list of genomic coordinates of accessible regions were used. DiffBind was run with default parameters, except consensus peak list was set to include peaks detected in at least two of the four samples.

Detected ATAC-seq regions were considered to overlap with ChIP-seq peaks if the summit of a ChIP-seq peak was located within an ATAC-seq region. Clustered ATAC-seq heatmaps, with clustering based on the differential accessibility and/or overlap with TFs, were generated using computeMatrix and plotHeatmap tools from deepTools. Normalized ATAC-seq bw files and coordinates of ATAC-seq region centers were used as input. Boxplots indicating distributions of ATAC-seq densities within heatmap clusters were generated using ggplot2. The values were obtained from the output file of *--outFileNameMatrix* argument of the computeMatrix tool run with additional argument *-bs 1500.*

GO term analysis

GO term biological process enrichment analyses were performed using the PANTHER classification system [74] (version 14.0).

T-test

Statistical significance of qRT-PCR as well as ATAC-seq and ChIP-seq density differences was assessed using two-tailed t-test as indicated in figure legends.

Competing interests

The authors declare no competing interests.

Data availability

Raw and processed sequencing data were deposited in Gene Expression Omnibus under accession
number GSE168580 (https://www.ncbi.nlm.nih.gov/geo/). Table S5 contains RNA-seq data (FPKM values) of the RefSeq genes for the entire differentiation course and indication of genes associated with pSmad1/5, pSmad2, Eomes, and T ChIP-seq peaks; all direct target genes of pSmad1/5, pSmad2, Eomes, and T; and all genes associated with sites opened or closed in Smad4, Eomes and/or T KO cells compared to WT cells.

Author contributions

B. G. H. initiated and designed the study, P.T. designed experiments, P.T., G.B. and D.S. performed experiments, P.T. and J.L. performed data analysis, L.W. generated Smad4 KO ESCs, B.T. supervised the sequencing, P.T., F.K., J.V.V. and B.G.H. interpreted data, P.T., B.G.H. and F.K. supervised the

project, P.T. wrote the first draft of the manuscript, P.T., B.G.H. and F.K. reviewed and edited the manuscript.

References

1. Tam, P.P.L. & Loebel, D.A.F. Gene function in mouse embryogenesis: get set for gastrulation. Nat. Rev. Genet. 2007; 8:368-381.

2. Lawson, K.A. Fate mapping the mouse embryo. Int. J. Dev. Biol. 1999;43:773-775.

3. Chu, G.C., Dunn, N.R., Anderson, D.C., Oxburgh, L. & Robertson, E.J. Differential requirements for Smad4 in TGFbeta-dependent patterning of the early mouse embryo. Development 2004;131:3501-3512.

4. Brennan, J. et al. Nodal signalling in the epiblast patterns the early mouse embryo. Nature 2001;411:965-969.

5. Russ, A.P. et al. Eomesodermin is required for mouse trophoblast development and mesoderm formation. Nature 2000;404:95-99.

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

7. Herrmann, B.G., Labeit, S., Poustka, A., King, T.R. & Lehrach, H. Cloning of the T gene required in mesoderm formation in the mouse. Nature 1990;343:617-622.

8. Winnier, G., Blessing, M., Labosky, P.A. & Hogan, B.L. Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. Genes Dev. 1995;9(17):2105-2116.

9. Conlon, F.L. et al. A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. Development 1994; 120:1919-1928.

10. Gaarenstroom, T. & Hill, C.S. TGF-B signaling to chromatin: How Smads regulate transcription during self-renewal and differentiation. Sem. In Cell and Dev. Biol. 32, 107-118 (2014).

11. Dunn, N.R., Vincent, S.D., Oxburgh, L., Robertson, E.J. & Bikoff, E.K. Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo. Development 2004;131:1717-1728.

12. Mishina, Y., Suzuki, A., Ueno, N. & Behringer R.R. Bmpr encodes a type I bone morphogenetic protein receptor that is essential for gastrulation during mouse embryogenesis. Genes Dev. 1995;9(24):3027-3037.

13. Arnold, S.J., Maretto, S., Islam, A., Bikoff, E.K. & Robertson, E.J. Dose-dependent Smad1, Smad5 and Smad8 signaling in the early mouse embryo. Dev. Biol. 2006; 296:104-118.

14. Tremblay, K.D., Dunn, N.R. & Robertson, E.J. Mouse embryos lacking Smad1 signals display defects in extra-embryonic tissues and germ cell formation. Development 2011;128(18):3609-21.

15. Chang, H., Zwijsen, A., Vogel, H., Huylebroeck, D. & Matzuk, M.M. Smad5 is essential for left-right asymmetry in mice. Dev. Biol. 2000;219:71-78.

16. Massague, J., Seoane, J. & Wotton, D. Smad transcription factors. Genes & Dev. 2005;19:2783-2810.

17. Sirard, C. et al. The tumor suppressor gene Smad4/Dpc4 is required for gastrulation and later for anterior development of the mouse embryo. Genes Dev. 1998;12:107-119.

18. Xu, J., Gruber, P.J. & Chien, K.R. SMAD4 is essential for human cardiac mesodermal precursor cell formation. Stem Cells 2019;37(2):216-225.

19. Shi, Y., Katsev, S., Cai, C. & Evans, S. BMP signaling is required for heart formation in vertebrates. Dev. Biol. 2000;224:226-237.

20. de Pater, E., et al. Bmp signaling exerts opposite effects on cardiac differentiation. Circulation Res. 2012;110:578-587.

21. Schultheiss, T.M., Burch, J.B.E. & Lassar A.B. A role for bone morphogenetic proteins in the induction of cardiac myogenesis. Genes and Dev. 1997;11:451-462.

22. Frasch, M. Induction of visceral and cardiac mesoderm by ectodermal Dpp in the early Drosophila embryo. Nature 1995;374:464-467.

23. Chen, X. et al. Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. Cell 2008;133:1106-1117.

24. Morikawa, M. et al. BMP Sustains Embryonic Stem Cell Self-Renewal through Distinct Functions of Different Krüppel-like Factors. Stem Cell Rep. 2046;6(1):64-73.

25. Liu, P., Wakamiya, M., Shea, M.J., Albrecht, U., Behringer, R.R. & Bradley, A. Requirement for Wnt3 in vertebrate axis formation. Nature Genetics 1999;22:361-365.

26. Reid, C.D., Zhang, Y., Sheets, M.D. & Kessler, D.S. Transcriptional integration of Wnt and Nodal pathways in establishment of the Spemann organizer. Dev. Biol. 2012;368:231-241.

27. Trompouki, E. et al. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. Cell 2011;14:577-589.

28. Wang, Q. et al. The p53 Family Coordinates Wnt and Nodal Inputs in Mesendodermal Differentiation of Embryonic Stem Cells. Cell Stem Cell 2017;20:70-86.

29. Funa, N.S. et al. b-Catenin Regulates Primitive Streak Induction through Collaborative Interactions with SMAD2/SMAD3 and OCT4. Cell Stem Cell 2015;16:639-652.

30. Costello, I. et al. The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. Nat. Cell Biol. 2011;13:1084-1091.

31. Papaioannou, V.E. The T-box gene family: emerging roles in development, stem cells and cancer. Development 2014;141:3819-3833.

32. Concepcion, D. & Papaioannou V.E. Nature and extent of left/right axis defects in TWis/TWis mutant mouse embryos. Dev. Dynamics 2014l;243:1046-1053.

33. Tosic, J. et al. Eomes and Brachyury control pluripotency exit and germ-layer segregation by changing the chromatin state. Nat. Cell Biol. 2019;21:1518-1531.

34. Zhao, M. et al. Deciphering role of Wnt signalling in cardiac mesoderm and cardiomyocyte differentiation from human iPSCs: Four-dimentional control of Wnt pathway for hiPSC-CMs differentiation. Sci. Rep. 2019;9:19389.

35. Yang, X. et al. Distinct enhancer signatures in the mouse gastrula delineate progressive cell fate continuum during embryo development. Cell Research 2019;29:911-926.

36. Morikawa, M. et al. ChIP-seq reveals cell type-specific binding patterns of BMP-specific Smads and a novel binding motif. Nucl. Acid Res. 2011;39(20):8712-8727.

37. Ben-Haim, N. et al. The nodal precursor acting via activin receptors induces mesoderm by maintaining a source of its convertases and BMP4. Dev. Cell 2006;11:313-323.

38. Senft, A.D. et al. Combinatorial Smad2/3 activities downstream of nodal signaling maintain embryonic/extra-embryonic cell identities during lineage priming. Cell Reports 2018;24:1977-1985.

39. Saga, Y., Kitajima, S. & Miyagawa-Tomita, S. Mesp1 expression is the earliest sign of cardiovascular development. Trends Cardiovasc. Med. 2000;10(8):345-352.

40. Meilhac, S.M. & Buckingham, M.E. The deployment of cell lineages that form the mammalian heart. Nat. Rev. Cardiology. 2018;15(11):705-724.

41. Faial, T. et al. Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells. Development 2015;142:2121-2135.

42. Burtscher, I. & Lickert, H. Foxa2 regulates polarity and epithelialization in the endoderm germ layer if the mouse embryo. Development 2009;136:1029-1038.

43. Bardot, E. et al. Foxa2 identifies a cardiac progenitor population with ventricular differentiation potential. Nat. Comm. 2017;8:14428.

44. Koch, F. et al. Antagonistic activities of Sox2 and Brachyury control the fate choice of neuro-mesodermal progenitors. Dev. Cell. 2017;42:1-13.

45. Fernandes, M.G., et al. BMP-SMAD signaling regulates lineage priming, but is dispensable for self-renewal in mouse embryonic stem cells. Stem Cell Reports 2016;6:85-94.

46. Nostro, M.C., Cheng, X., Keller, G.M. & Gadue, P. Wnt, Activin, and BMP Signaling Regulate Distinct Stages in the Developmental Pathway from Embryonic Stem Cells to Blood. Cell Stem Cell 2008;2(1):60-71.

47. Xu, X. Wang, L., Liu, B., Xie, W. & Chen, Y.-G. Activin/Smad2 and Wnt/-catenin up-regulate HAS2 and ALDH3A2 to facilitate mesendoderm differentiation of human embryonic stem cells. J. Biol. Chem. 2018;293(48):18444-18453.

48. Kattman, S.J. et al. Stage-specific optimization of Activin/Nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. Cell Stem Cell 2011; 8:228-240.

49. Yoney, A. et al. WNT signaling memory is required for ACTIVIN to function as a morphogen in human gastruloids. eLife 2018;7:e38279.

50. Kim, M-S. et al. Activin-A and Bmp4 levels modulate cell type specification during CHIR-induced cardiomyogenesis. PLoS ONE 2015; 10(2):e0118670.

51. Simon, C.S. et al. Functional characterisation of cis-regulatory elements governing dynamic Eomes expression in the early mouse embryo. Development 2017;144:1249-1260.

52. Nowotschin, S. et al. The T-box transcription factor Eomesodermin is essential for AVE induction in the mouse embryo. Genes and Dev. 2013;27:997–1002.

53. Alexanian, M. et al. A transcribed enhancer dictates mesendoderm specification in pluripotency. Nat. Comm. 2017;8:1806.

54. Chia, C.Y. et al. GATA6 cooperates with EOMES/SMAD2/3 to deploy the gene regulatory network governing human definitive endoderm and pancreas formation. Stem Cell Reports 2019;12:57-70.

55. Hill, C.S. Transcriptional control by the SMADs. Cold Spring Harb. Perspect. Biol. 2016;8(10):1-17.

56. Pfeiffer, M.J. et al. Cardiogenic programming of human pluripotent stem cells by dose-controlled activation of EOMES. Nat. Comm. 2018;9:440.

57. Teo, A.K.K. et al. Pluripotency factors regulate definitive endoderm specification through eomesodermin. Genes Dev. 2011;25:238-250.

58. David, R. et al. Induction of Mesp1 by Brachyury(T) generates the common multipotent cardiovascular stem cell. Cardiovascular Res. 2011;32:115-122.

59. George, S.H.L. et al. Development and adult phenotyping directly from mutant embryonic stem cells. PNAS 2007;104:4455-4460.

60. Ran, F.A. et al. Genome engineering using the CRISPR-Cas9 system. Nat. Protocols 2013;8(11):2281-2308.

61. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 2013;10:1213-1218.

62. Kim, D., Paggi, J.M., Park, C., Bennett, C. & Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT genotype. Nat. Biotechnology 2019;37:907-915.

63. Li, H. et al. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 2009;25:2078-2079.

64. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 2010;28:511-515.

65. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nuc. Acid Res. 2016;44:160-165.

66. Ramirez, F. et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat. Comm. 2018;9(189):1-15.

67. De Hoon, M.J., Imoto, S. & Miyano, N.S. Open source clustering software. Bioinformatics 2004;20:1453-1454.

68. Saldanha, A.J. Java Treeview - extensible visualization of microarray data. Bioinformatics 2004;20:3246-3248.

69. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 2012;9:357-359.

70. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9:R137.

71. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841-842.

72. Bailey, T.L. et al. MEME SUITE: tools for motif discovery and searching. Nuc. Acids Res. 2009;37:202-208.

73. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.

74. Mi, H. et al. Protocol update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat. Protocols 2019;14:703-721.
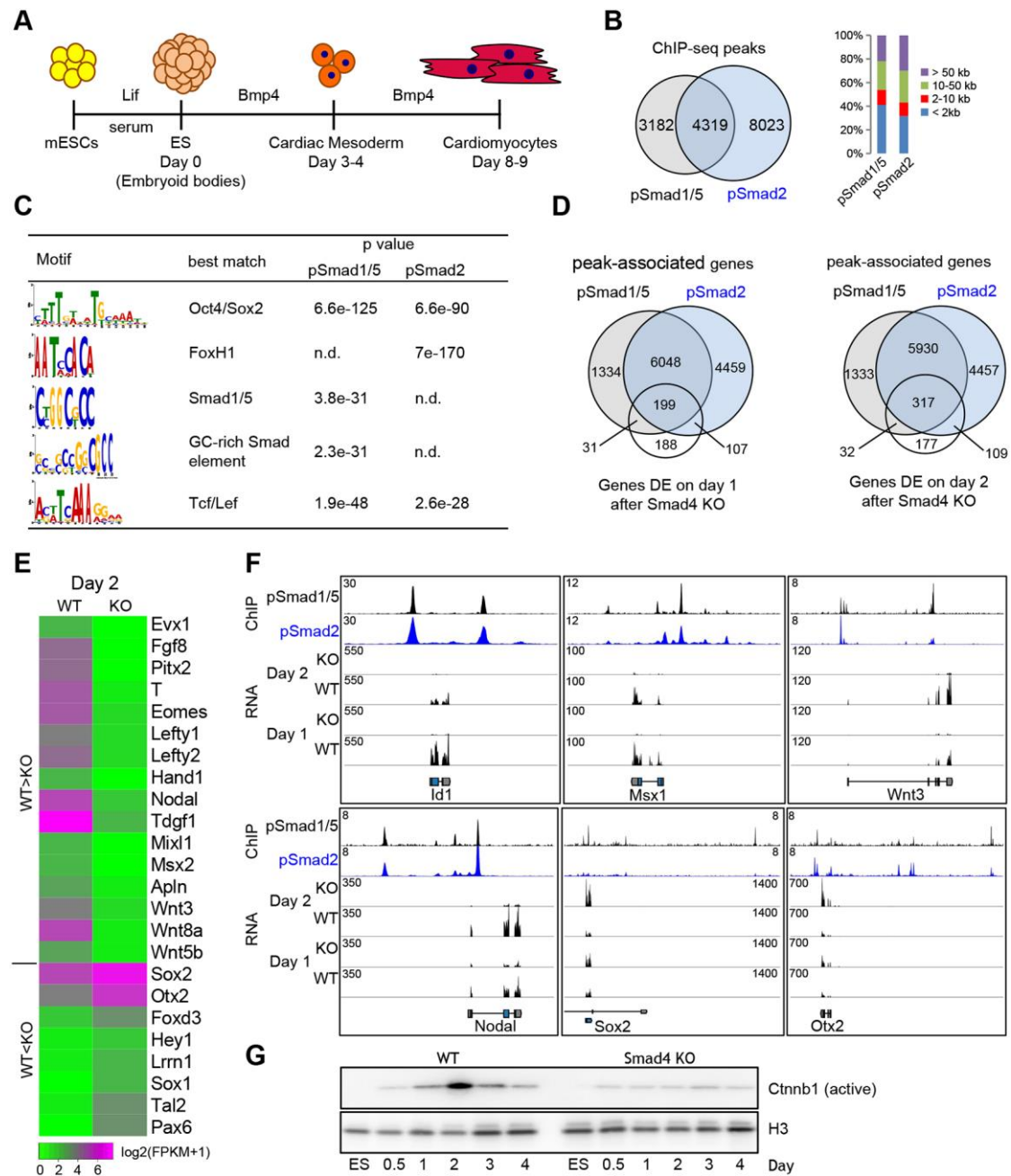
Figures

Fig. 1. Cardiac progenitor genes are co-bound by pSmad1/5 and pSmad2, and require Smad4 for expression

(**A**) Schematic representation of the *in vitro* differentiation course of mESCs to cardiac mesoderm.

(**B**) Venn diagram showing overlap between pSmad1/5 and pSmad2 DNA binding sites in day 2 ME cells (left). Barplot shows the quantification of distances between binding sites and nearest gene promoters (right).

(**C**) Top *de novo* motifs discovered at pSmad binding sites (n.d. = not detected).

(**D**) Venn diagram showing overlap between genes bound by pSmad1/5 and/or pSmad2, and genes DE between WT and Smad4 KO cells at day 1 or day 2 of differentiation.

(**E**) Heatmap representation of the expression of selected pSmad1/5/2 direct target genes in WT or Smad4 KO cells at day 2 of differentiation. (RNA-seq, average of two replicates).

(**F**) Snapshots of ChIP-seq and RNA-seq tracks in WT or Smad4 KO ME cells at selected genomic loci.

See also Supplementary Information, Fig. S1 and Tables S1-S3.

(**G**) Western blot analysis of time-dependent nuclear accumulation of active Ctnnb1 in WT and Smad4 KO cells. Histone H3 (H3) served as loading control.

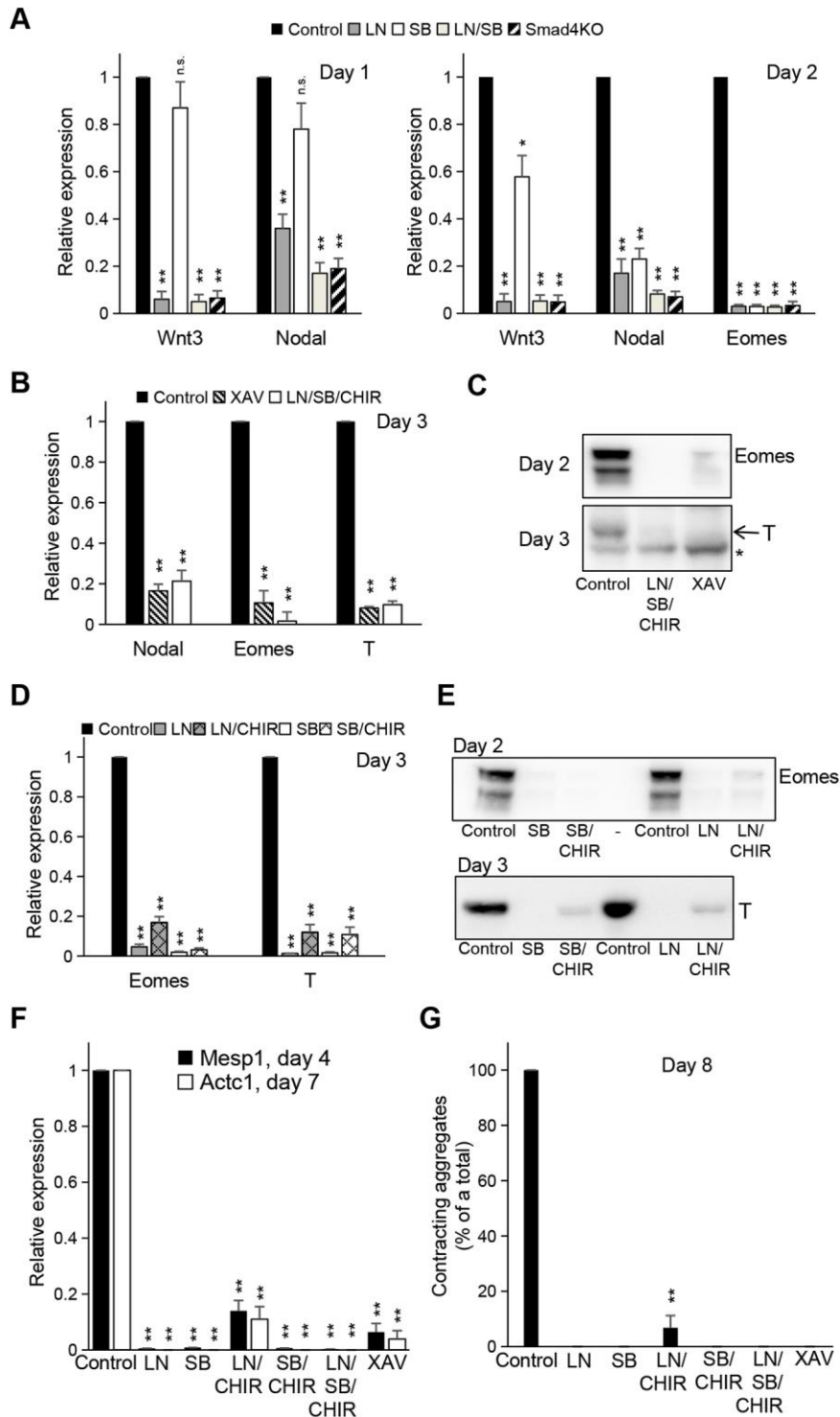See also Supplementary Information, Fig. S1 and Tables S1-3.

Fig. 2. Cardiac progenitor and cardiomyocyte formation require pSmad1/5, pSmad2/3, and WNT signaling

(**A**) Barplots showing the expression of *Wnt3, Nodal,* and *Eomes* in day 1 (left panel) and day 2 ME WT cells (control) or in cells treated with 1 μM LDN193189 (LN), 10 μM SB431542 (SB), a combination of LN and SB, or in

Smad4 KO cells. For all qRT-PCR experiments total RNA was used, expression levels were calculated relative to the corresponding WT control, and each bar represents the mean plus the Standard Error of the Mean (SEM) (Student's t-test, *p < 5e-2, **p < 1e-2, n.s., not significant, n=3).

(**B**) Barplot displaying the expression of *Nodal, Eomes*, or *T* in day 3 ME WT cells (control) or in cells treated with 10 µM XAV939 (XAV) or a combination of 1 µM LDN, 10 µM SB, and 5 µM CHIR99021 (LN/SB/CHIR).

(**C**) Western blot analysis of Eomes or T protein expressed in cells treated as in (B). Asterisk indicates a background band.

(**D**) Barplots of *Eomes* or *T* expression in day 3 ME WT cells (control) or in cells treated with 1 µM LDN alone or with 5 µM CHIR (LN or LN/CHIR), with 10 µM SB alone or with 5 µM CHIR (SB or SB/CHIR).

(**E**) Western blot analysis of Eomes or T protein expression in cells treated as in (D).

(**F**) Barplots showing the expression of *Mesp1* (day 4) or *Actc1* (day 7) in ME WT cells (control) or in cells treated with 1 µM LDN alone or in combination with 5 µM CHIR (LN or LN/CHIR), with 10 µM SB alone or with 5 µM CHIR (SB or SB/CHIR), with a combination of 1 µM LDN, 10 µM SB, and 5 µM CHIR (LN/SB/CHIR), or with 10 µM XAV.

(**G**) Barplot showing the fractions of contracting aggregates (% of total aggregates) at day 8 of differentiation. Cells were treated as in (F) and grown in 24-well plates. All aggregates in 6 wells were counted per treatment condition. Each bar represents the mean plus the SEM (Student's t-test compared to Control, **p < 1e-2, n=6).
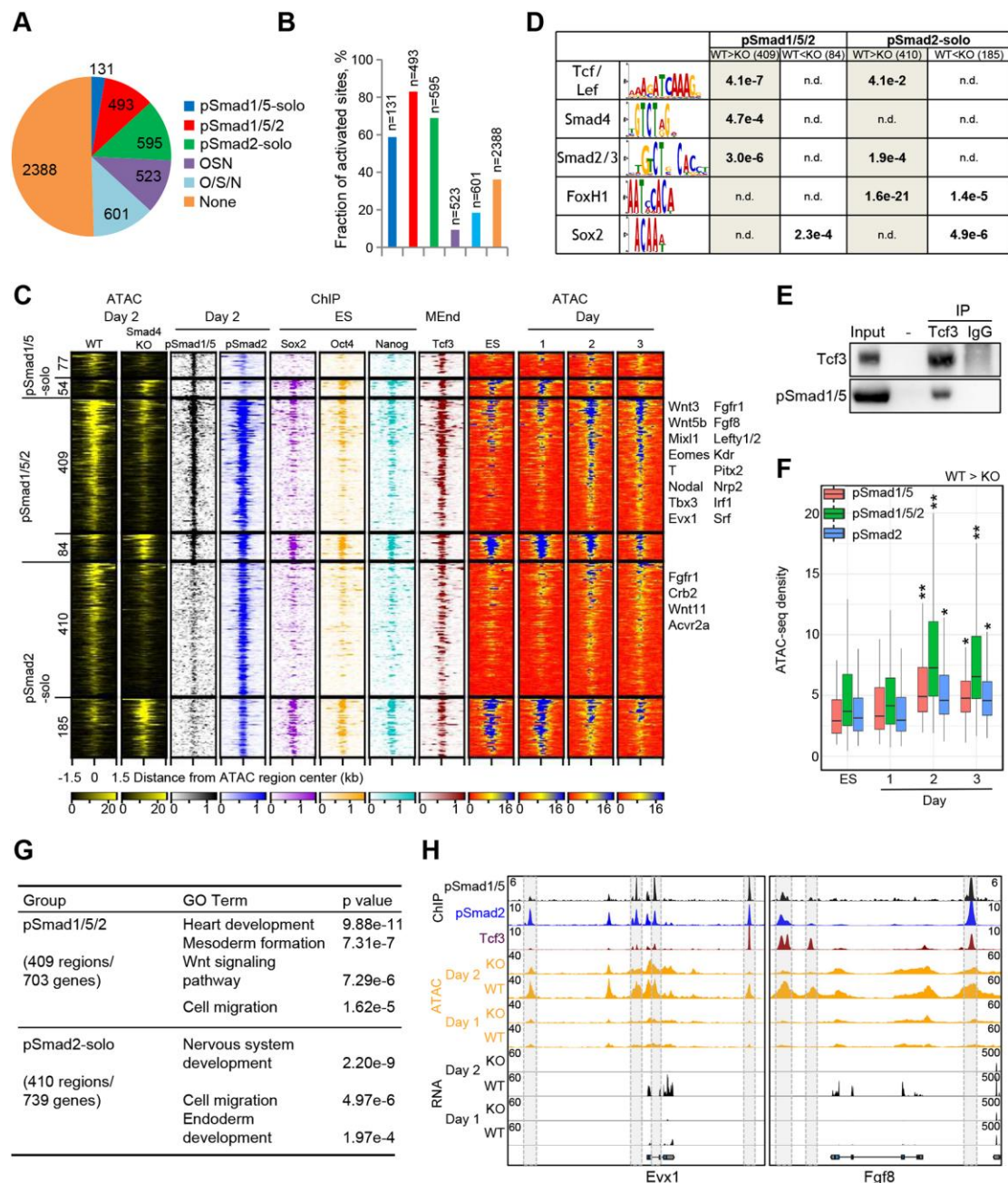
See also Supplementary Information, Fig. S2.

Fig. 3. Smad transcription factors activate mesodermal enhancers

(**A**) Pie chart showing regions differentially accessible (DA) between WT and Smad4 KO ME cells occupied by pSmad1/5 (Smad1/5-solo), pSmad1/5 and pSmad2 (pSmad1/5/2), or pSmad2 (pSmad2-solo) in day 2 ME cells, Oct4/Sox2 and Nanog (OSN), or any one or two of Oct4, Sox2 and/or Nanog (O/S/N) in ESCs, or none.

(**B**) Barplot of the fractions of activated (WT>KO) DA regions in groups from (A).

(**C**) Clustered heatmaps of ATAC-seq reads in WT and Smad4 KO day 2 ME cells, centered on DA regions, and corresponding heatmaps of ChIP-seq and ATAC-seq reads during differentiation. Clusters contain activated and repressed DA sites occupied by pSmad1/5-solo, pSmad2, or pSmad1/5/2. Selected genes bound by regions from corresponding clusters are listed on the right.

(**D**) Motifs enriched in DA regions from the groups defined in (C) (n.d. = not detected).

(**E**) Western blot analysis of Tcf3 or pSmad1/5 in the input or eluates following immunoprecipitation with Tcf3 or IgG antibodies.

(**F**) Boxplot of the quantification of normalized ATAC-seq density in groups containing activated DA sites defined in (C) during the differentiation course. (Student's t-test compared to ES, *p < 5e-6, **p < 5e-10).

(**G**) GO terms enriched among genes associated with activated pSmad1/5/2 or pSmad2-occupied DA regions.

(**H**) Snapshots of pSmad1/5 or pSmad2 (day 2 ME), or Tcf3 (MEnd) ChIP-seq, ATAC-seq and RNA-seq tracks in WT or Smad4 KO ME cells at selected genomic loci. Grey boxes highlight selected DA regions.

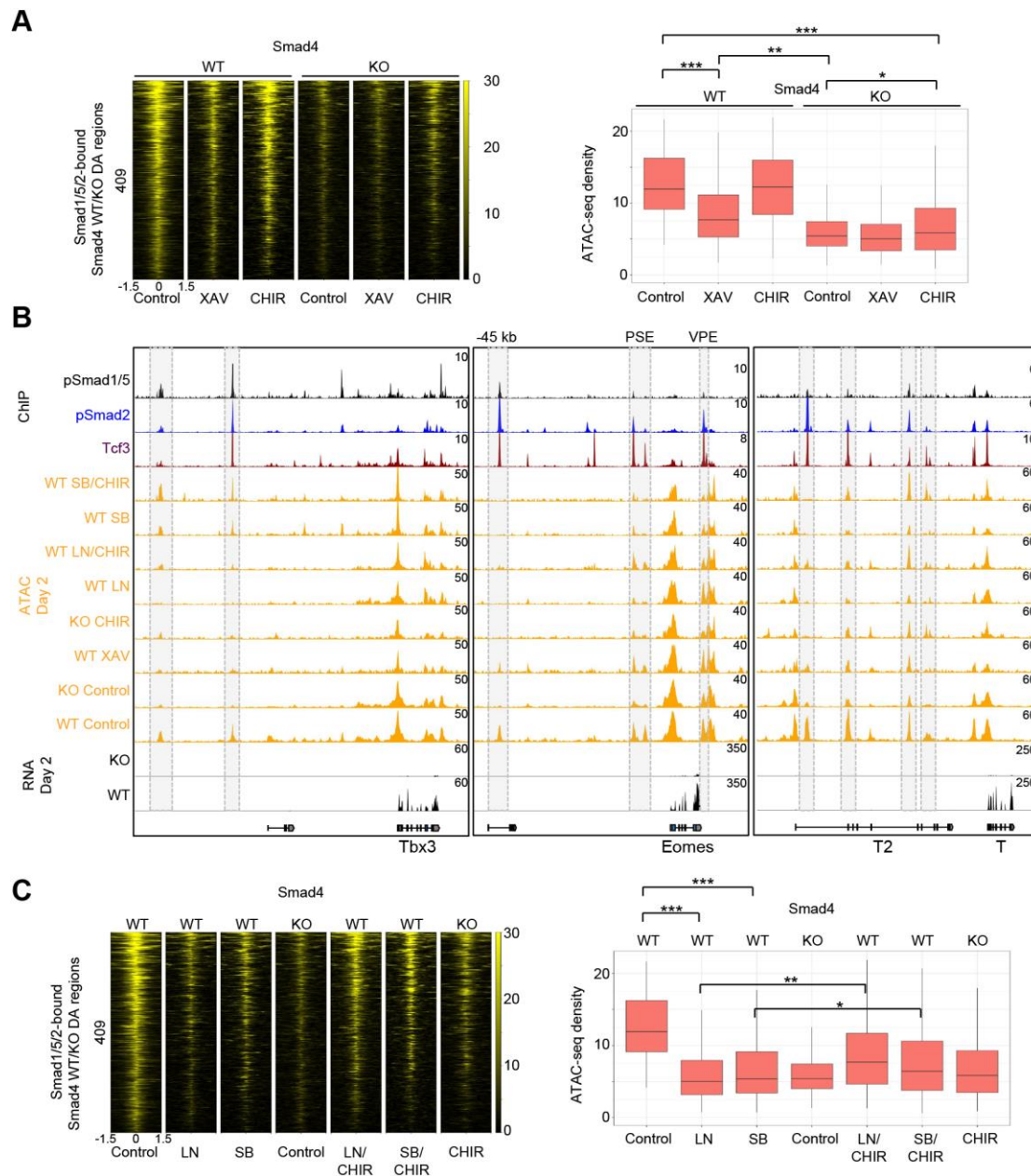See also Supplementary Information, Fig. S3 and Table S3.

Fig. 4. Cooperation of BMP, Nodal, and WNT signaling is pivotal to opening of mesodermal enhancers

(**A**) Heatmap of ATAC-seq reads and boxplot showing quantification of normalized ATAC-seq density in WT and Smad4 KO day 2 ME cells, centered on 409 DA regions occupied by pSmad1/5/2 defined in Fig. 3C. Cells treated with 10 μM XAV (XAV), 5 μM CHIR (CHIR), and control cells are depicted. (Student's t-test, *p < 5e-2, **p < 5e-6, ***p < 5e-10).

(**B**) Snapshots of pSmad1/5 and pSmad2 (day 2 ME) and Tcf3 (MEnd) ChIP-seq, ATAC-seq and RNA-seq tracks in WT or Smad4 KO ME cells at selected genomic loci. Grey boxes highlight selected DA regions.

(**C**) Heatmap of ATAC-seq reads and boxplot showing quantification of normalized ATAC-seq density in WT and Smad4 KO day 2 ME cells, centered on the same regions as in (A). Cells treated with 1 µM LDN alone or in combination with 5 µM CHIR (LN or LN/CHIR), with 10 µM SB alone or together with 5 µM CHIR (SB or SB/CHIR), with 5 µM CHIR (CHIR), and control cells are depicted. (Student's t-test, *p < 5e-2, **p < 5e-6, ***p < 5e-10).
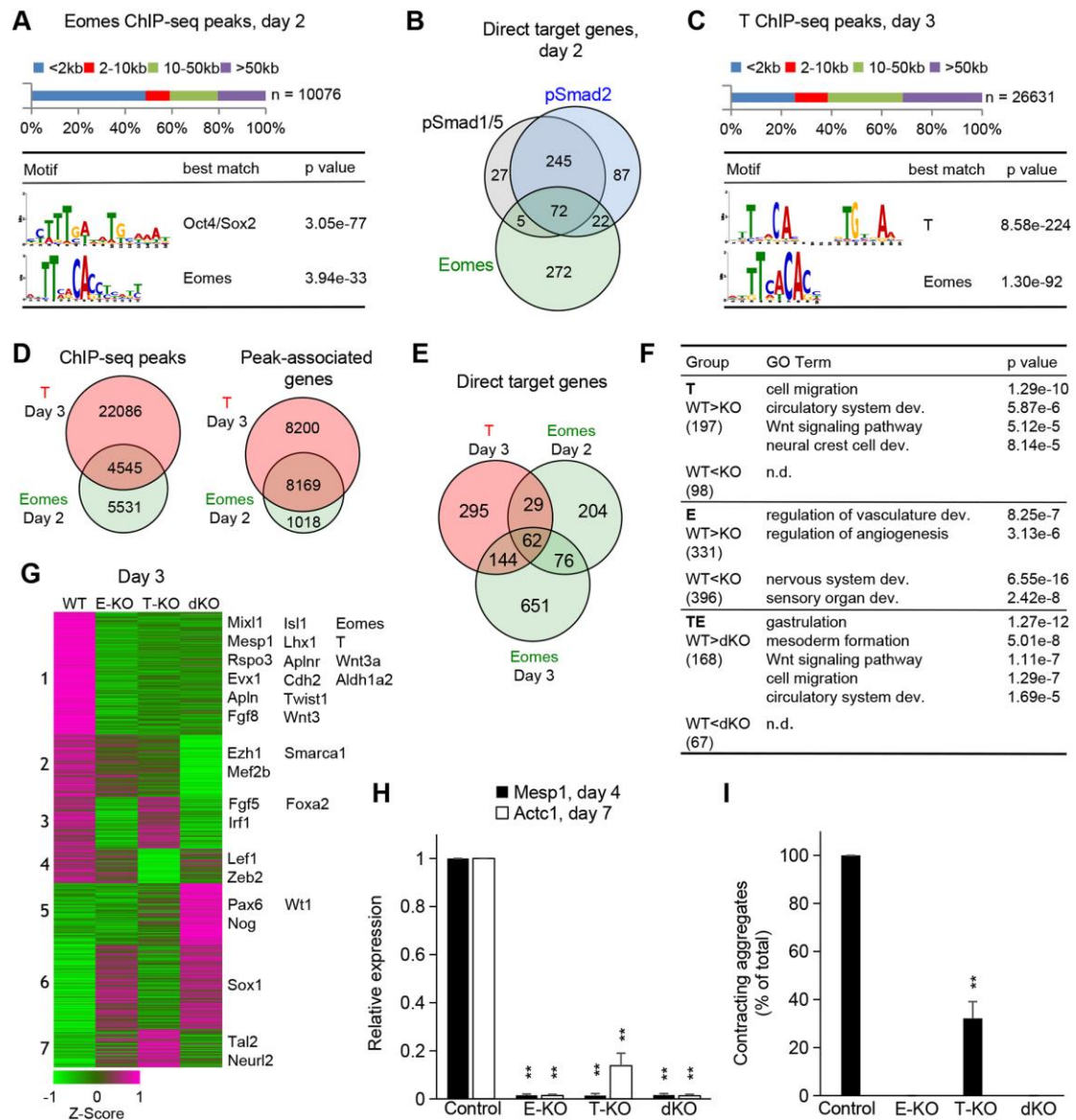
Fig. 5. Upregulation of cardiac progenitor genes involves a positive feedback loop between Eomes and T

(**A**) Barplot showing quantification of distances between Eomes binding sites in day 2 ME cells and nearest gene promoters (top). Top two *de novo* motifs discovered at Eomes binding sites (bottom).

(**B**) Venn diagram displaying the overlap between the direct target genes of pSmad1/5, pSmad2 and Eomes at day 2.

(**C**) Barplot showing quantification of distances between T binding sites in day 3 ME cells and nearest gene promoters (top). Top two *de novo* motifs discovered at T binding sites (bottom).

(**D**) Venn diagrams indicating the overlap between T and Eomes DNA binding sites or the overlap between genes bound by T and Eomes in ME cells.

(**E**) Venn diagram showing overlap between T and Eomes direct target genes.

(**F**) GO terms overrepresented in T, Eomes (E) and common (TE) day 3 direct target genes subdivided into activated (WT>KO) and repressed (WT<KO) groups (n.d. = not detected).

(**G**) K-means clustered heatmap representation of the expression of T and/or Eomes direct target genes in WT, E-KO, T KO, and dKO day 3 ME cells. Selected genes from each cluster are listed on the right. (RNA-seq, average of two replicates).

(**H**) Barplot of the expression of *Mesp1* (day 4) and *Actc1* (day 7) in ME WT, Eomes KO (E-KO), T KO (T-KO) and double KO (dKO) cells. Each bar represents the mean plus the SEM (Student's t-test, **p < 1e-2, n=3).

(**I**) Barplot of the fractions (% of total) of contracting aggregates at day 8. Cells were treated as in (H) and grown in 24-well plates. All aggregates in 6 wells per treatment condition were counted. Each bar represents the mean plus the SEM (Student's t-test, *p < 1e-2, n=6).

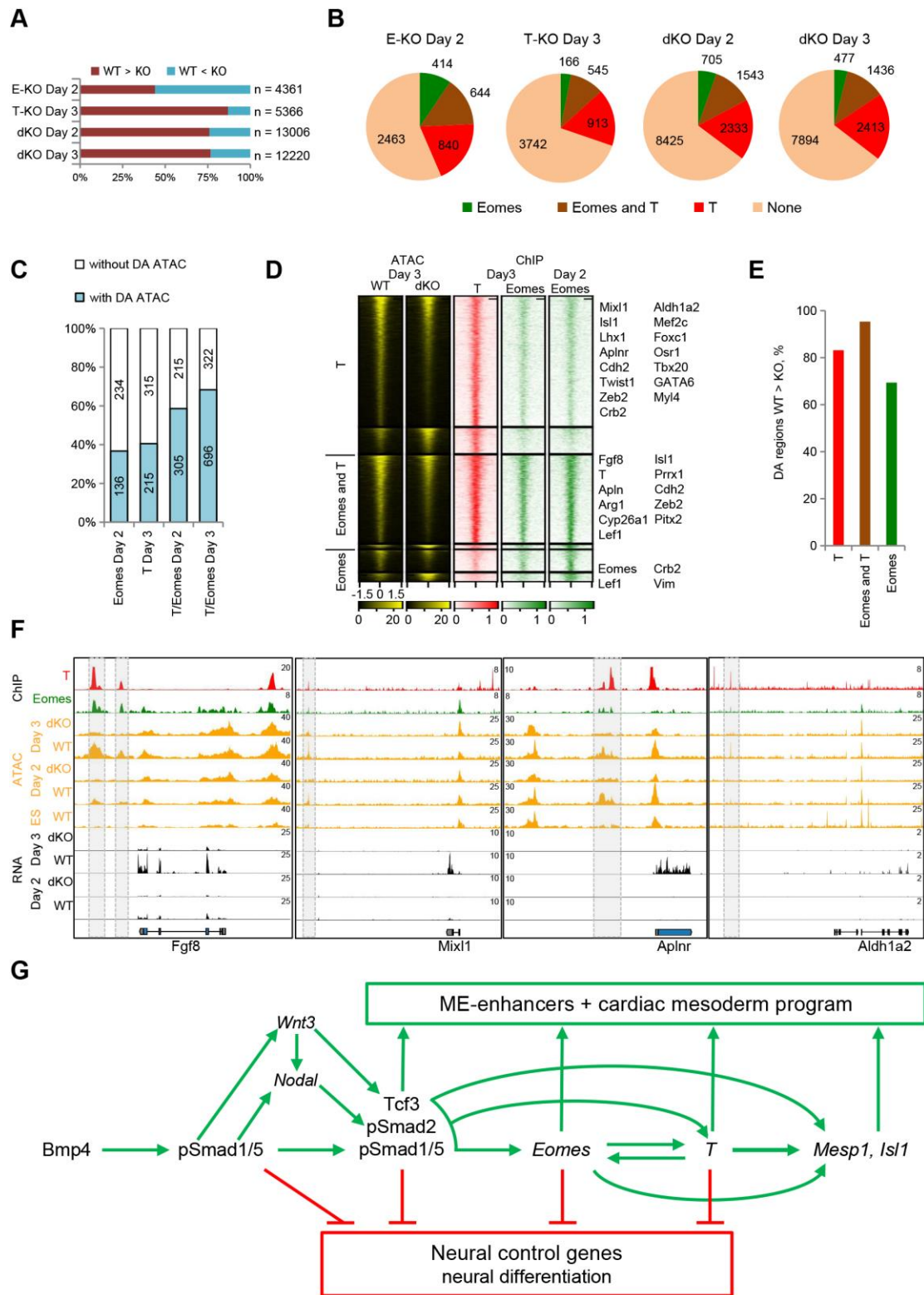See also Supplementary Information, Fig. S4 and Table S3.

Fig. 6. Eomes and T cooperatively activate mesoderm enhancers

(**A**) Barplot of the fractions of activated (WT > KO) or repressed (WT < KO) DA regions in E-KO, T-KO and dKO ME cells.

(**B**) Pie charts showing occupation of DA regions between WT and E-KO, T-KO or dKO ME cells by Eomes, T or both.

(**C**) Barplot of the fractions and number of Eomes, T, or T/Eomes direct targets bound by regions DA between WT and E-KO, T-KO or dKO ME cells.

(**D**) Clustered heatmaps of ATAC-seq reads in WT or dKO day 3 ME cells, centered on DA regions, and corresponding heatmaps of ChIP-seq reads. Clusters contain activated or repressed DA sites occupied by T, Eomes or both. Selected genes bound by regions from corresponding clusters are listed on the right.
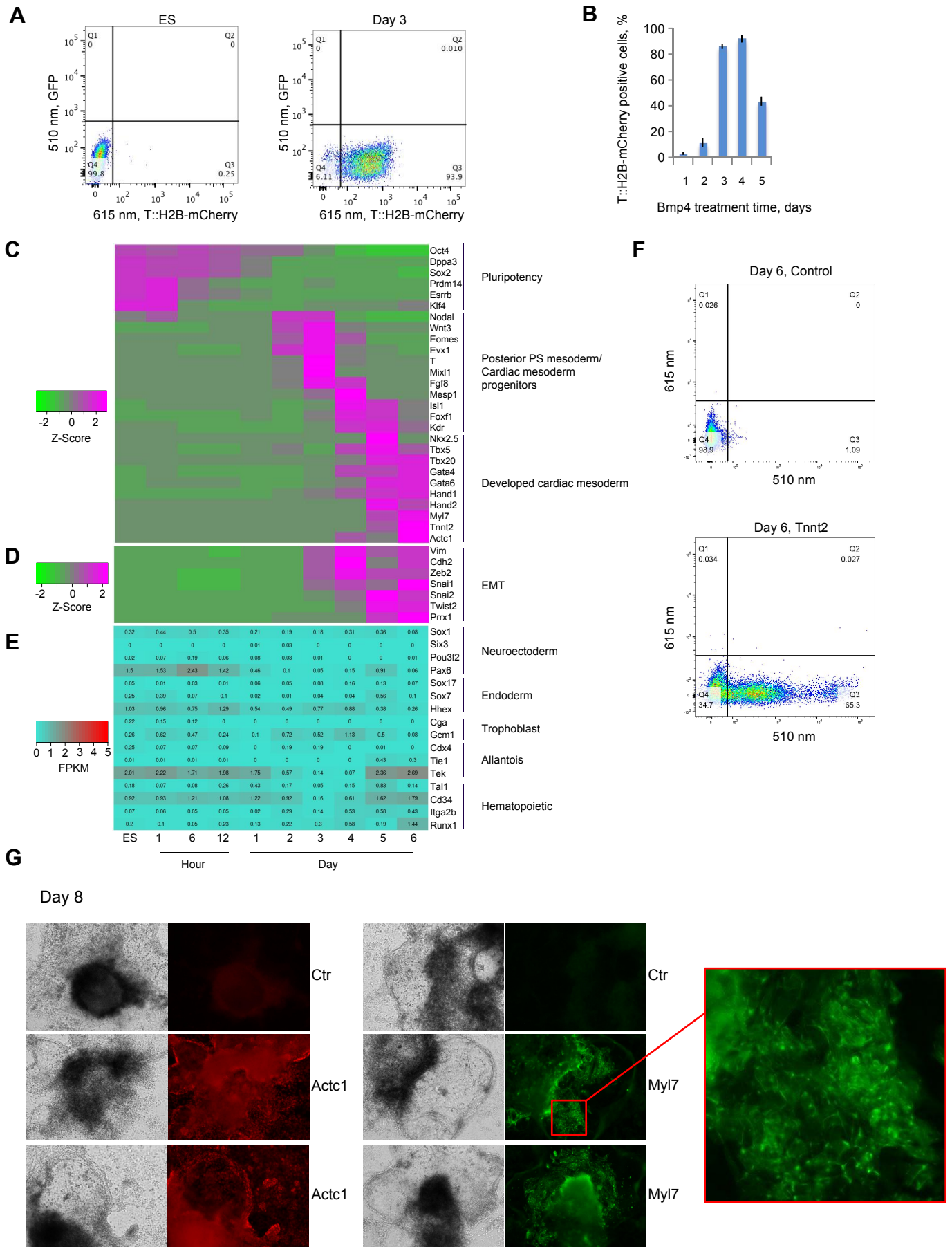
(**E**) Barplot showing fractions of activated (WT>KO) DA regions between WT and dKO day 3 ME cells occupied by T, Eomes, or both.

(**F**) Snapshots of ChIP-seq, ATAC-seq and RNA-seq tracks in WT or dKO ME cells at selected loci. Gray boxes highlight selected DA T- and/or Eomes-bound enhancers.
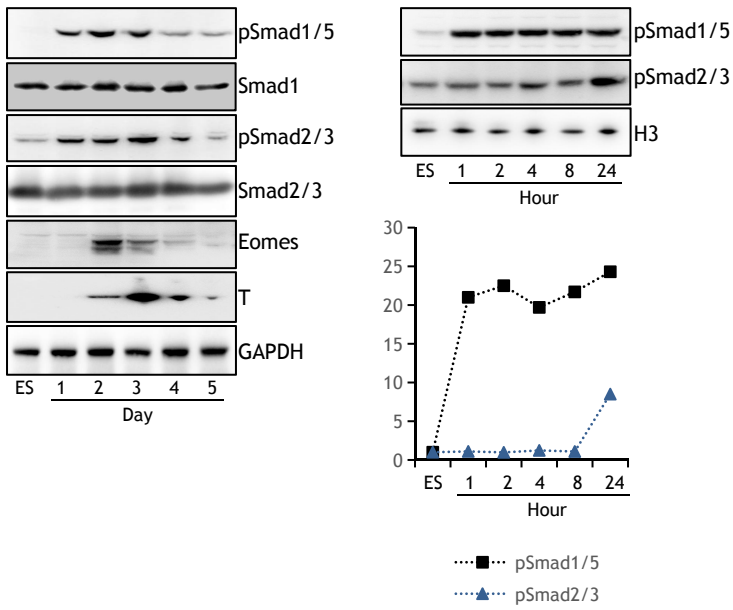
(**G**) Schematic model of the regulatory network orchestrating cardiac mesoderm formation.
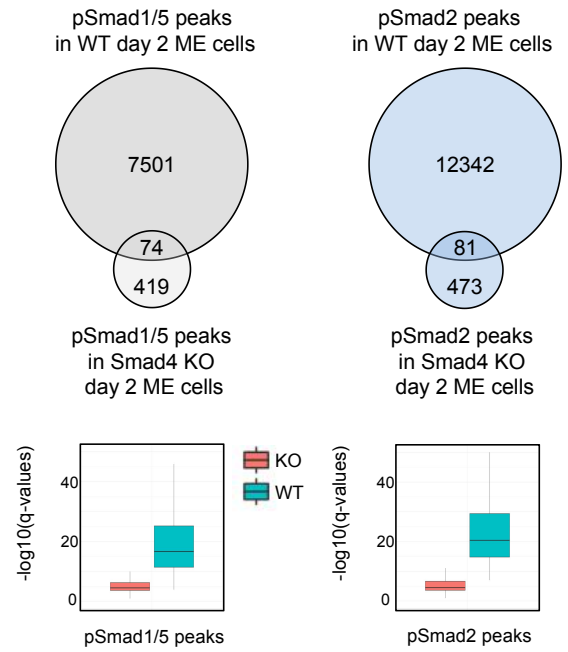
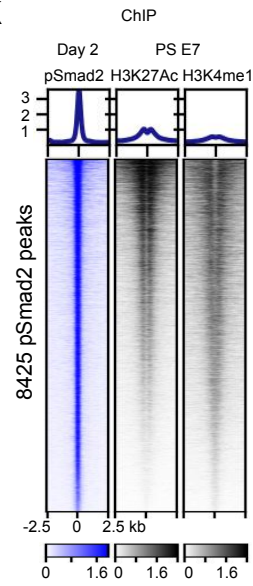See also Supplementary Information, Fig. S5 and Table S3.

Figure S1

A

B

C

D

E

F

G

Day 8

H

I

pSmad1/5 peaks in WT day 2 ME cells

7501

74

419

pSmad1/5 peaks in Smad4 KO day 2 ME cells

pSmad2 peaks in WT day 2 ME cells

12342

81

473

pSmad2 peaks in Smad4 KO day 2 ME cells

J

ChIP

Day 2 — pSmad1/5 — PS E7 — H3K27Ac — H3K4me1

4418 pSmad1/5 peaks

K

ChIP

Day 2 — pSmad2 — PS E7 — H3K27Ac — H3K4me1

8425 pSmad2 peaks

L

ActA, day 2 — Eomes

WT — Smad4 KO

ActA, day 3 — T

WT — Smad4 KO

M

Smad1/5/2 direct targets, day 2

log2(KO/WT)

Mesodermal fate

WT>KO (205)

WT<KO (112)

Sox3

(sorted by log2FC)

Evx1
Fgf8
Wnt8a
Tdgf1
Pitx2
Trim67
T
Hand1
Sp5
Foxb1
Mixl1
Msx2
Tbx3
Lefty2
Eomes
Nodal
Wnt3
Axin2
Fzd10
Lefty1
Msx1
Gsc
Apln
Wnt5b

Sox2
Hey1
Otx2
Foxd3
Lrrn1
Sox1
Tal2
Pax6
Npcd

Neural fate
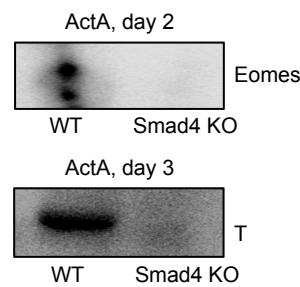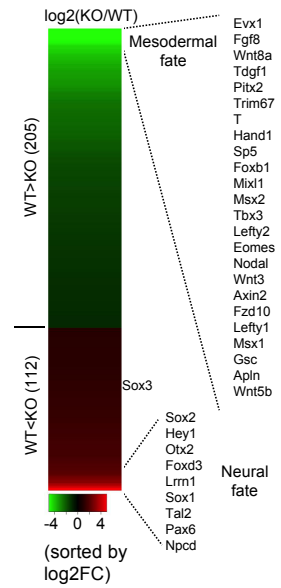
Fig. S1. Characterization of the *in vitro* differentiation system and DNA binding of Smads (related to Fig. 1)

(**A**) Differentiation efficiency assessed by FACS using ES cell line with T::H2B-mCherry reporter inserted into the translation start site of T located on a BAC. FACS dot plots indicating that over 90% of cells express T at differentiation day 3.

(**B**) Barplot showing fractions of cells expressing T::H2B-mCherry reporter during the differentiation time course assessed by FACS.

(**C-D**) Heatmap representation of the relative expression (z-score) of pluripotency factors, mesoderm progenitor and cardiac mesoderm marker genes (D), and EMT marker genes (E) throughout the course of differentiation. (RNA-seq, average of two replicates).

(**E**) Heatmap displaying the expression values (FPKM) of definitive endoderm and neuroectoderm marker genes throughout the course of differentiation. (RNA-seq, average of two replicates).

(**F**) Cardiomyocyte formation efficiency assessed by FACS using day 6 cells expressing cardiomyocyte marker Tnnt2 (510nm).

(**G**) Expression of cardiac markers Actc1 or Myl7 in differentiated WT aggregates at day 8 detected by immunofluorescence. Abbr.: Ctr, secondary antibody control.

(**H**) Time-course of Smad1/5 or Smad2/3 phosphorylation and Eomes or T expression detected by Western blot analysis. Quantification of the Smad1/5 and Smad2/3 phosphorylation levels relative to ES.

(**I**) Venn diagrams showing overlap of pSmad1/5 or pSmad2 peaks detected in WT and Smad4 KO day 2 ME cells. Peaks detected in Smad4 KO cells were excluded from further analyses. Boxplots showing distribution of the q-values for the corresponding peak sets.

(**J**) Heatmaps centered on pSmad1/5 peaks outside promoters, displaying ChIP-seq reads of pSmad1/5 in day 2 ME cells and H3K27Ac and H3K4me1 in E7 posterior primitive streak [35].

(**K**) Heatmaps centered on pSmad2 peaks outside promoters, displaying ChIP-seq reads of pSmad2 in day 2 ME cells and H3K27Ac and H3K4me1 in E7 posterior primitive streak [35].

(**L**) Expression of Eomes or T in WT or Smad4 KO cells treated with Activin A (ActA) for 2 or 3 days, respectively.

(**M**) Heatmap representation of the expression of pSmad1/5/2 direct target genes at day 2, sorted by log2 fold change (FC) of FPKM values in Smad4 KO or WT ME cells. The strongest activated and repressed target genes are listed on the right.

Figure S2

A

Day 1, Bmp4

pSmad1/5

H3

0	1	0	10	µM
LN		SB		

Day 1, ActA

pSmad2/3

H3

0	1	0	10	µM
LN		SB		

C

Day 2

Eomes

0	10	0	10	µM

K02288 (Smad1/5 inhibition) A-83-01 (Smad2/3 inhibition)

B

Day 2

← Ctnnb1 (active)

Ctr   SB   SB/CHIR   Ctr   LN   LN/CHIR

D

Day 2

Eomes

Ctr   SB   SB/CHIR   Ctr   LN   LN/CHIR

Eomes

Eomes

Ctr   SB   SB/CHIR   Ctr   LN   LN/CHIR

Day 3

← T

*

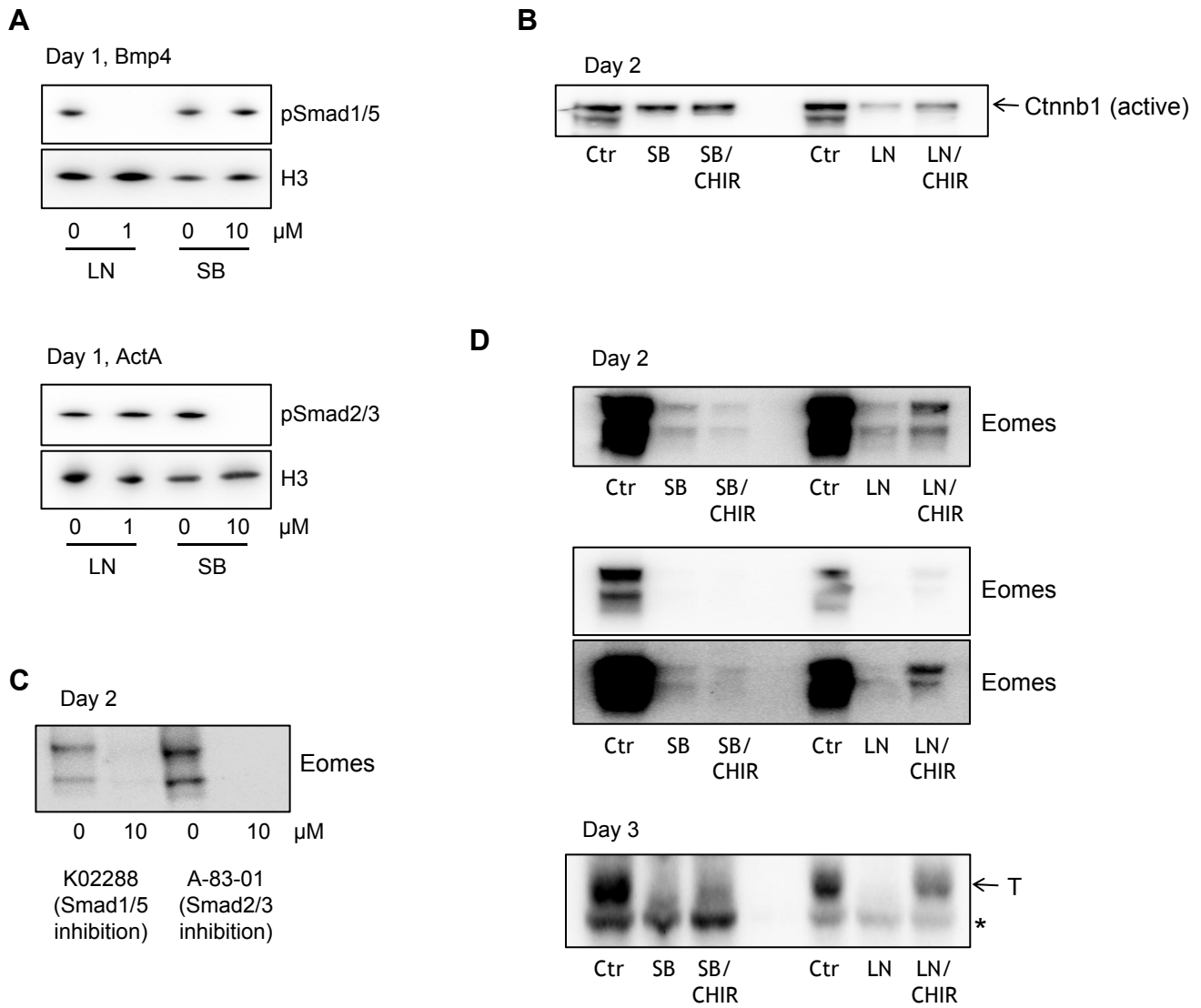Ctr   SB   SB/CHIR   Ctr   LN   LN/CHIR

Fig. S2. Formation of proper nascent and committed cardiac mesoderm requires simultaneous activity of pSmad1/5, pSmad2/3, and WNT signaling (related to Fig. 2)

(**A**) 1 µM LDN193189 (LN) or 10 µM SB431542 (SB) selectively inhibit pSmad1/5 or pSmad2/3 phosphorylation, respectively. pSmad1/5 and pSmad2/3 are detected by Western blot analysis. pSmad1/5 and pSmad2/3 phosphorylation was induced by Bmp4 or ActA, respectively.

(**B**) Nuclear accumulation of active Ctnnb1 is selectively blocked by 1 µM LDN (LN), as detected by Western blot analysis. Abbr.: Ctr, control.

(**C**) Eomes expression is repressed in day 2 ME cells by K02288 (inhibits Smad1/5 phosphorylation) as well as by A-83-01 (inhibits Smad2/3 phosphorylation).

(**D**) Western blot showing higher exposure of Fig. 2E, detecting Eomes expression (upper panel); Western blot analysis at low and high exposure showing Eomes protein expression in an independent experiment (middle panel). Independent experiment showing T expression by Western blot analysis (related to Fig. 2E; lower panel). Asterisk indicates a background band.
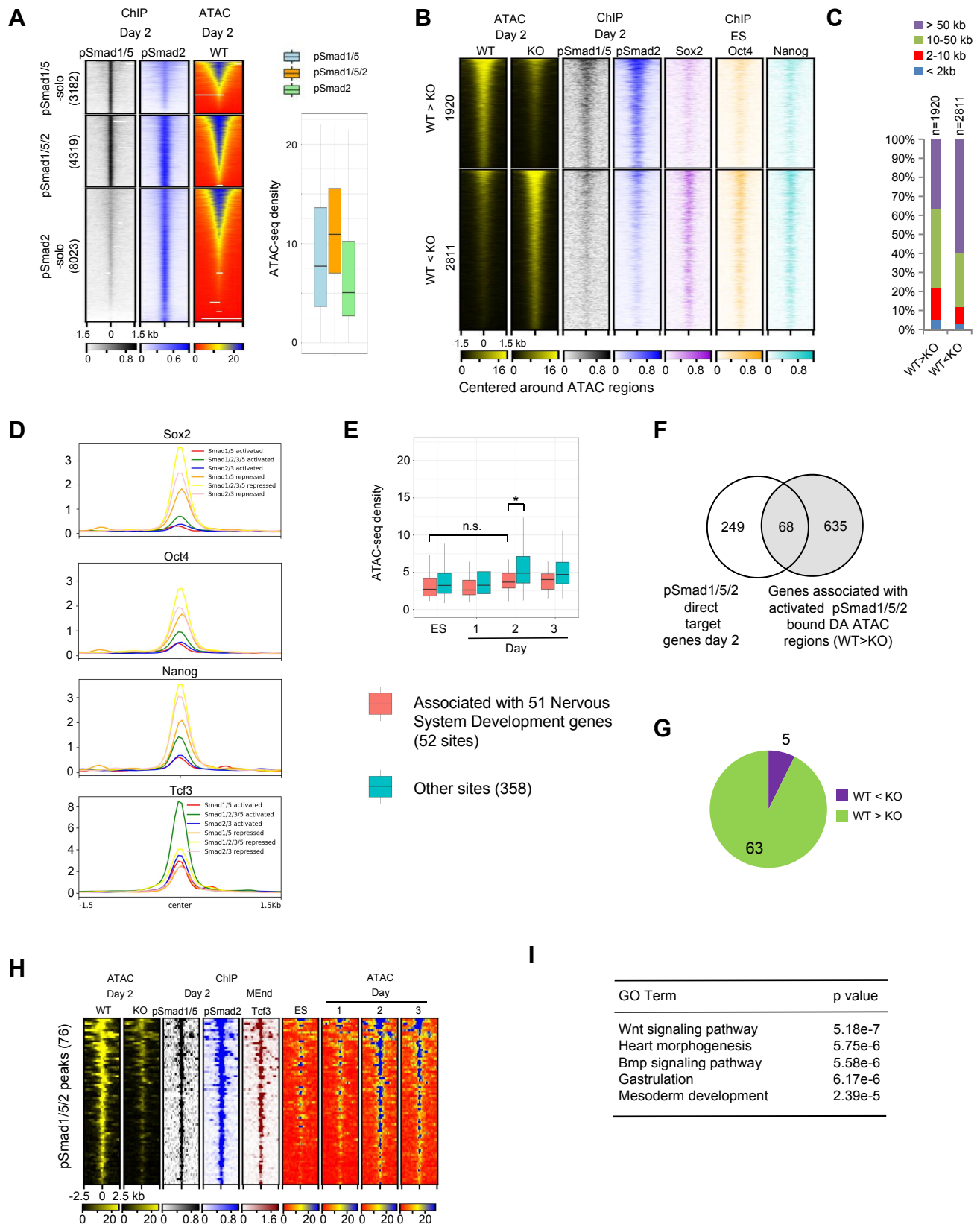
Figure S3

Fig. S3. Characterization of Smad-mediated chromatin regulation (related to Fig. 3)

(**A**) Heatmaps, clustered by pSmad1/5 and pSmad2 occupancy and centered on the peak summits, displaying ChIP-seq reads of pSmad1/5 or pSmad2 and ATAC-seq reads from the same sites in day 2 ME cells. Boxplot of the quantification of normalized ATAC-seq density is shown on the right.

(**B**) Clustered heatmaps displaying ATAC-seq reads in day 2 WT or Smad4 KO ME cells centered on sites activated or repressed by Smad activity, and the corresponding ChIP-seq reads of pSmad1/5, pSmad2, Sox2, Oct4 and Nanog.

(**C**) Barplot showing quantification of distances between DA regions in Smad4 KO relative to WT cells and nearest gene promoters for groups defined in (B).

(**D**) Average ChIP-seq density profiles of indicated TFs in clusters defined in Fig. 3C.

(**E**) Boxplot of the quantification of normalized ATAC-seq density during the differentiation course in sites associated with Nervous system Development GO term and other p-Smad2-solo bound sites defined in

(Fig. 3C, Fig. 3G). (Student's t-test, *p < 5e-2, n.s., not significant).

(**F**) Venn diagram showing overlap of pSmad1/5/2 direct target genes with genes bound by activated pSmad1/5/2-containing DA regions.

(**G**) Pie chart showing that the majority of pSmad1/5/2 direct targets associated with activated pSmad1/5/2-bound DA regions (see Fig. S3E) are Smad-activated (WT > Smad4 KO).

(**H**) Heatmaps centered on 76 pSmad1/5/2 peaks associated with 63 activated pSmad1/5/2 targets from (**F**), displaying ATAC-seq reads in WT or Smad4 KO day 2 ME cells, ChIP-seq reads of indicated TFs, and ATAC-seq reads during differentiation.

(**I**) GO terms enriched among the 63 activated pSmad1/5/2 target genes defined in (F).
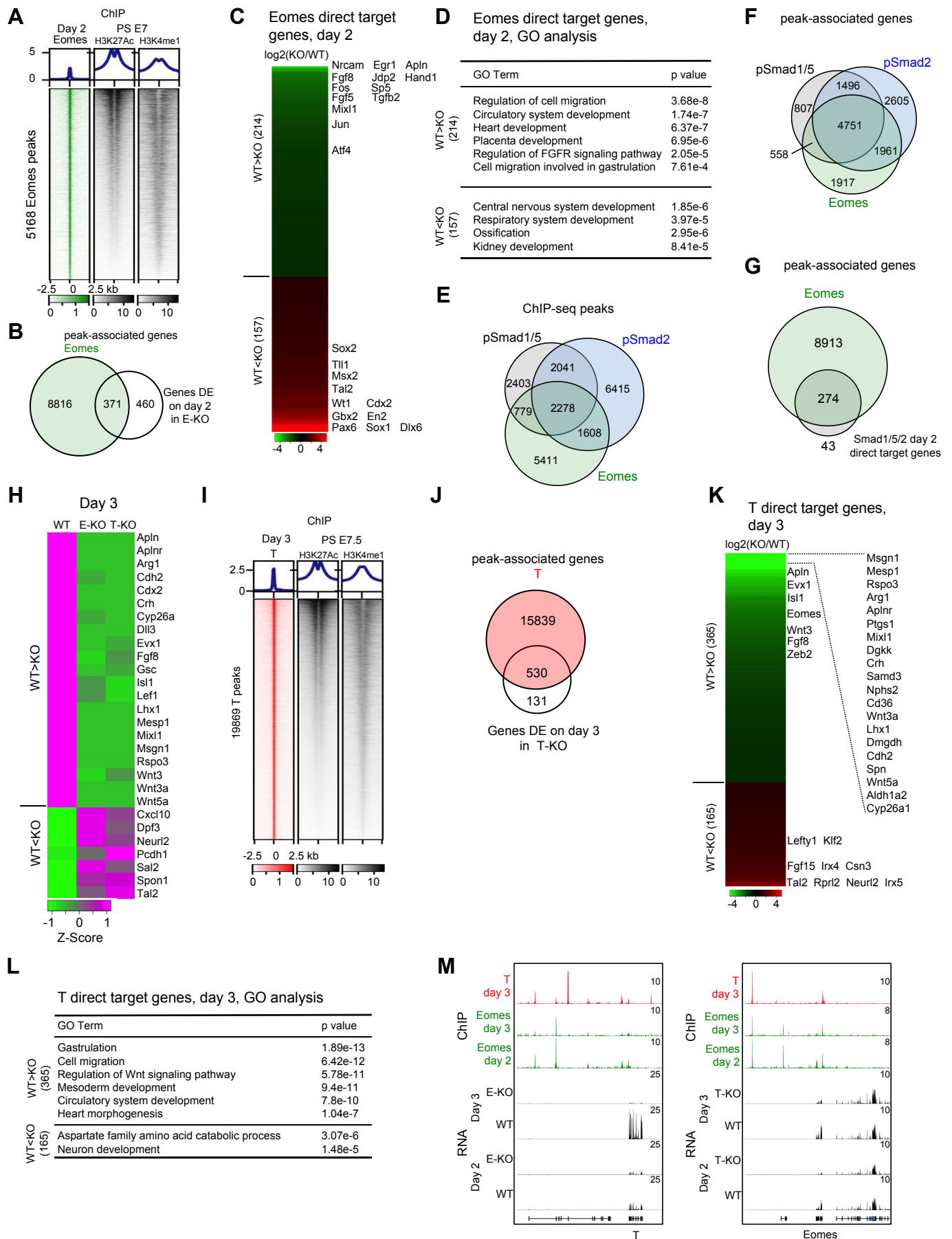
Figure S4

Fig. S4. DNA binding and target genes of Eomes and T (related to Fig. 5)

(**A**) Heatmaps centered on promoter-far Eomes peaks, displaying ChIP-seq reads of Eomes in day 2 ME cells, and H3K27Ac and H3K4me1 in E7 posterior primitive streak [35].

(**B**) Venn diagram showing overlap between genes bound by Eomes and genes DE between WT and Eomes KO cells. Genes in the intersection are putative direct targets of Eomes.

(**C**) Heatmap representation of Eomes day 2 direct target gene expression, sorted by log2 fold change (FC) of FPKM values in Eomes KO or WT ME cells. Selected activated and repressed target genes are listed on the right.

(**D**) GO terms enriched among activated (WT>KO) or repressed (WT<KO) Eomes direct target genes.

(**E**) Venn diagram showing overlap between pSmad1/5, pSmad2 and Eomes DNA binding sites in day 2 ME cells.

(**F**) Venn diagram showing overlap between genes bound by pSmad1/5, pSmad2, and Eomes in day 2 ME cells.

(**G**) Venn diagram showing overlap between genes bound by Eomes and direct target genes of
pSmad1/5/2 in day 2 ME cells.

(**H**) Heatmap representation of the expression of selected Eomes and T direct target genes in WT, Eomes KO (E-KO) or T KO (T-KO) day 3 ME cells. (RNA-seq, average of two replicates)

(**I**) Heatmaps centered on promoter-far T peaks, displaying ChIP-seq reads of T in day 3 ME cells, and H3K27Ac and H3K4me1 in E7.5 posterior primitive streak [35].

(**J**) Venn diagram showing overlap between genes bound by T and genes DE between WT and T-KO cells. Genes in the intersection are putative direct targets of T.

(**K**) Heatmap representation of the expression of T direct target genes at day 3, sorted by log2 fold change (FC) of FPKM values in T-KO relative to WT ME cells. Selected activated or repressed target genes are listed on the right.

(**L**) GO terms enriched among activated (WT>KO) or repressed (WT<KO) day 3 direct target genes of T.(**M**) T is a direct target of Eomes, and vice versa. Snapshots of ChIP-seq and RNA-seq tracks in WT, E-KO or T-KO ME cells at T and Eomes loci.
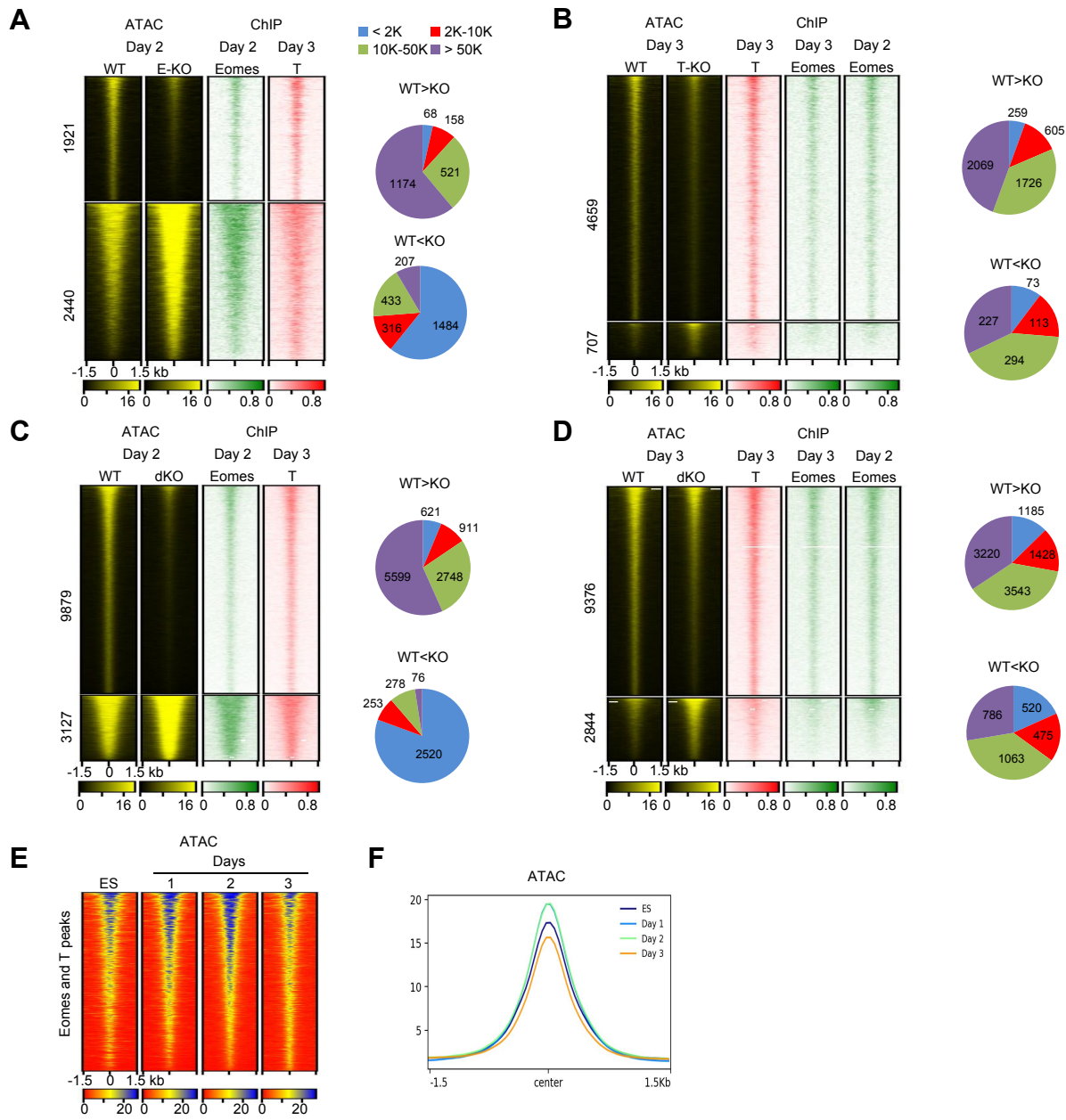
Figure S5

Fig. S5. Characterization of Eomes- and/or T-mediated chromatin regulation (related to Fig. 6)

(**A-D**) Clustered heatmaps displaying ATAC-seq reads in WT or (A) E-KO, (B) T-KO, (C,D) double KO

(dKO) ME cells, centered on activated or repressed DA sites, and the corresponding ChIP-seq reads of indicated TFs (left). Pie charts showing quantification of distances between the corresponding activated or repressed DA regions and nearest gene promoters (right).

(**E**) Heatmap displaying ATAC-seq reads centered on activated DA regions co-occupied by Eomes and T (see Fig. 6D).

(**F**) Average ATAC-seq density profiles in clusters from (E).

Table S1. GO term analysis for genes bound by pSmad1/5 and pSmad2 (related to Fig. 1).

GO Term	pSmad1/5 and pSmad2 common targets (6247)	pSmad1/5 only targets (1365)	pSmad2 only targets (4566)
Nervous system development	1.61e-27	2.84e-5	1.2e-5
Heart development	1.13e-18	ND	ND
Wnt signalling	4.4e-11	ND	ND
Mesoderm development	1.95e-8	ND	ND
Gastrulation	3.8e-8	ND	ND
TGFbeta signalling pathway	3.7e-7	ND	ND
Regulation of EMT	3.31e-5	ND	ND

Table S2. GO term analysis for direct target genes of pSmad1/5/2 (related to Fig. 1).

A   pSmad1/5/2 direct targets, day 1

	GO Term	p value
WT>KO (110)	Mesoderm formation	1.91e-9
	BMP signaling pathway	1.63e-5
	Wnt signaling pathway	2.29e-5
WT<KO (89)	Nervous system development	4.21e-14
	Neuron differentiation	3.98e-9

B   pSmad1/5/2 direct targets, day 2

	GO Term	p value
WT>KO (205)	Wnt signaling pathway	2.68e-11
	Heart development	3.17e-10
	Vasculature development	4.03e-8
	BMP signaling pathway	1.93e-7
	EMT	8.99e-7
	Mesoderm formation	4.87e-6
WT<KO (112)	Nervous system development	1.2e-11
	Neuron fate commitment	4.56e-5

Table S3. Lists of pSmad1/5, pSmad2, Eomes, and T direct target genes and genes associated with indicated sets of DNA loci (related to Figs. 1 - 6).

Click here to download Table S3

Table S4. Oligonucleotides used for generation and genotyping of knock-out cells. Primers used for RT-PCR.

Single stranded oligonucleotides for Eomes KO

| Pair 1: | CACCGACATGGTCTACCCACGGCAT | F |
	AAACATGCCGTGGGTAGACCATGTC	R
Pair 2:	CACCGTCCGGCGCCATAGAGCCACC	F
	AAACGGTGGCTCTATGGCGCCGGAC	R
Pair 3:	CACCGTCACGTAGATCTGCCCTCAA	F
	AAACTTGAGGGCAGATCTACGTGAC	R
Pair 4:	CACCGGTTTGCAGCAGGCGATTTG	F
	AAACCAAATCGCCTGCTGCAAACC	R

Eomes KO genotyping primers

| genoF: | TGAAGGCCAGGAAGTCAGAAG | F |
| genoR: | GTCACCAGCTCAAACCGACTTC | R |

Single stranded oligonucleotides for T KO

| Pair 1: | CACCGCGCCACCCTCTCCACCTTCC | F |
	AAACGGAAGGTGGAGAGGGTGGCGC	R
Pair 2:	CACCGTCGCCGGGCACAGAGAGCGC	F
	AAACGCGCTCTCTGTGCCCGGCGAC	R
Pair 3:	CACCGACGCACCTACCTGCCGTTCT	F
	AAACAGAACGGCAGGTAGGTGCGTC	R
Pair 4:	CACCGGCGAGTGCCCCTCGCCTCG	F
	AAACCGAGGCGAGGGGCACTCGCC	R
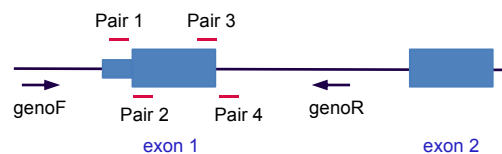
T KO genotyping primers

| genoF: | GCAGCTGCTCGGTACTTCAAAG | F |
| genoR: | CAGCAGAGGGACAAATCTAGCCAAG | R |

Primers used for qRT-PCR assays

CTTCTAATGGAGCCCCACCTG	F	Wnt3
GGGCCAGGGACCACCAAAT	R	
CTGGCGTACATGTTGAGCCT	F	Nodal
GGTCACGTCCACATCTTGCG	R	
GGGACAACTACGATTCATCCCA	F	Eomes
GGGCTTGAGGCAAAGTGTTG	R	
GGTATTCCCAATGGGGGTGG	F	T
TGTGTGGAGGGGAGAGAGAG	R	
GGATCCCCGCCTGCCTA	F	Mesp1
GCTGCTGAAGAGCGGAGAT	R	
CCCCGTCCATCAGAGAGCTA	F	Actc1
TGGGTTCTGTAGGCGTGCTA	R	
ACTGGGAAATGATGAGAATGGC	F	Pmm2
CCCCCAGATGCCCTTGAATA	R	

Eomes gene locus

T gene locus

Table S5. RNA-seq data (FPKM values) of the RefSeq genes for the entire differentiation course and indication of genes associated with pSmad1/5, pSmad2, Eomes, and T ChIP-seq peaks; all direct target genes of pSmad1/5, pSmad2, Eomes, and T; and all genes associated with sites opened or closed in Smad4, Eomes and/or T KO cells compared to WT cells.

Click here to download Table S5
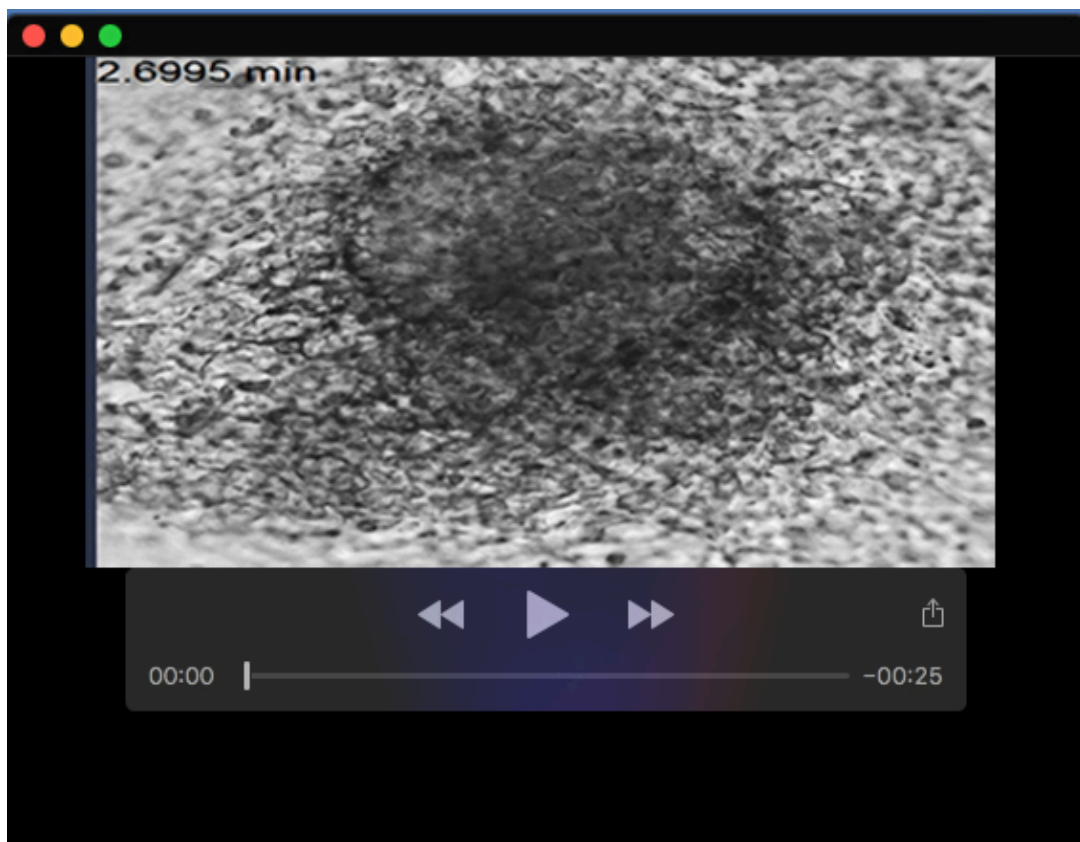
Movie 1. Capture of contracting cells on differentiation day 8.