

# Developmental Cell

## Antagonistic Activities of *Sox2* and *Brachyury* Control the Fate Choice of Neuro-Mesodermal Progenitors

### Highlights

- *Brachyury* is essential for the maintenance of neuro-mesodermal progenitors
- *Sox2* and *Brachyury* antagonistically control the neural versus mesodermal fate choice
- *Brachyury* is essential for chromatin remodeling promoting the mesodermal lineage
- *Tbx6* sustains the mesodermal lineage choice and commits cells to the paraxial fate

### Authors

Frederic Koch, Manuela Scholze,  
Lars Wittler, ..., Bernd Timmermann,  
Karol Macura, Bernhard G. Herrmann

### Correspondence

herrmann@molgen.mpg.de

### In Brief

The mammalian trunk arises from neuro-mesodermal progenitors (NMPs) giving rise to two fundamentally different cell lineages, one forming the spinal cord and the other mesodermal tissues. Analyzing the lineages derived from NMPs, Koch et al. find antagonistic roles for *Sox2* and *Brachyury* in the specification of neural and mesodermal fates.

# Antagonistic Activities of Sox2 and *Brachyury* Control the Fate Choice of Neuro-Mesodermal Progenitors

Frederic Koch,<sup>1</sup> Manuela Scholze,<sup>1</sup> Lars Wittler,<sup>1</sup> Dennis Schifferl,<sup>1</sup> Smita Sudheer,<sup>1,4</sup> Phillip Grote,<sup>1,5</sup> Bernd Timmermann,<sup>2</sup> Karol Macura,<sup>1</sup> and Bernhard G. Herrmann<sup>1,3,6,\*</sup>

<sup>1</sup>Department of Developmental Genetics, Max Planck Institute for Molecular Genetics, Ihnestr. 63–73, 14195 Berlin, Germany

<sup>2</sup>Sequencing Core Facility, Max Planck Institute for Molecular Genetics, Ihnestr. 63–73, 14195 Berlin, Germany

<sup>3</sup>Institute for Medical Genetics, Charité – University Medicine Berlin, Campus Benjamin Franklin, Hindenburgdamm 30, 12203 Berlin, Germany

<sup>4</sup>Present address: Central University of Kerala, Tejaswini Hills, Periyar. P.O., Kasaragod District, Kerala 671316, India

<sup>5</sup>Present address: Institute of Cardiovascular Regeneration, Goethe University, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany

<sup>6</sup>Lead Contact

\*Correspondence: herrmann@molgen.mpg.de

<http://dx.doi.org/10.1016/j.devcel.2017.07.021>

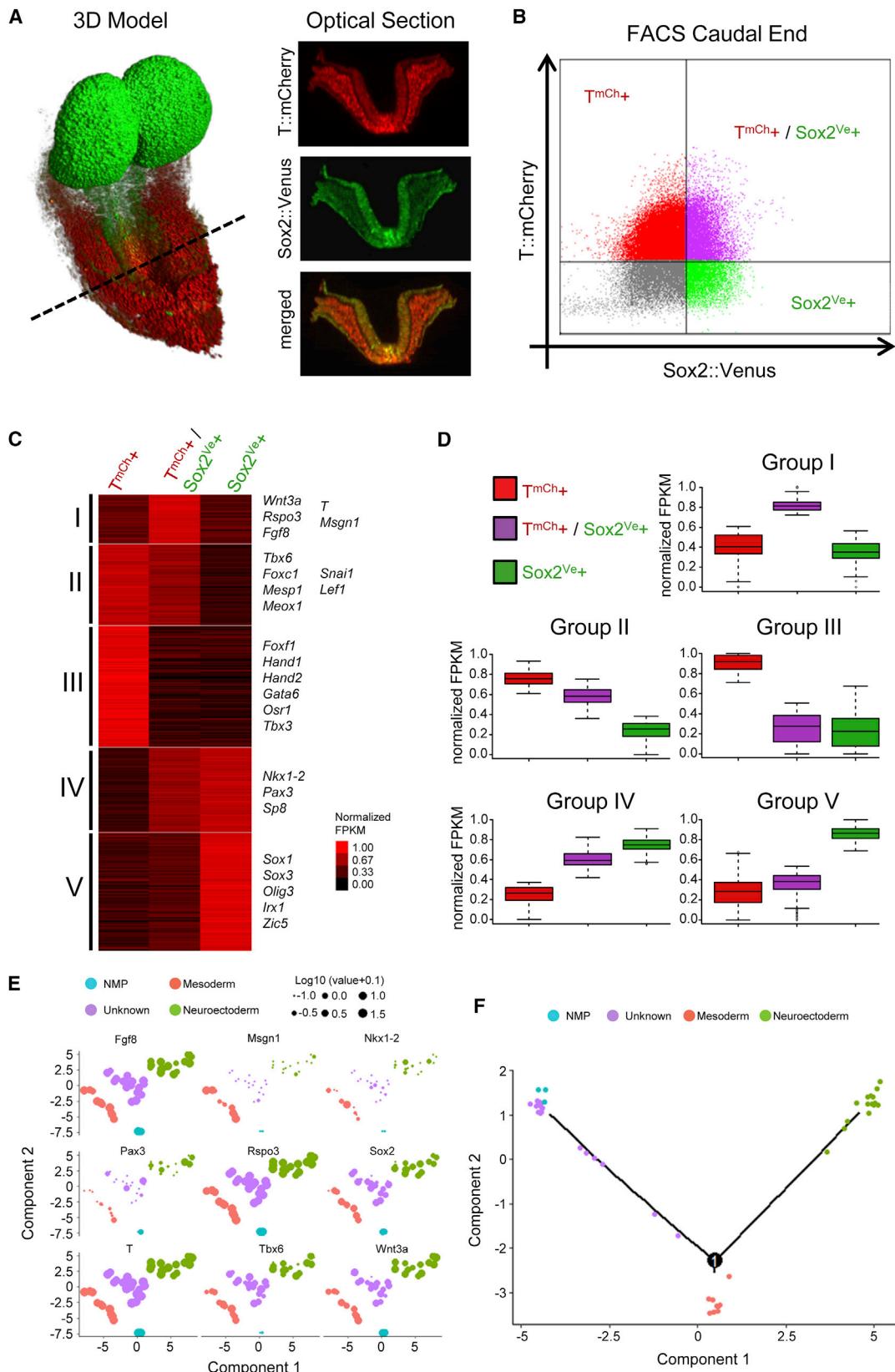
## SUMMARY

The spinal cord and mesodermal tissues of the trunk such as the vertebral column and skeletal musculature derive from neuro-mesodermal progenitors (NMPs). *Sox2*, *Brachyury* (*T*), and *Tbx6* have been correlated with NMP potency and lineage choice; however, their exact role and interaction in these processes have not yet been revealed. Here we present a global analysis of NMPs and their descending lineages performed on purified cells from embryonic day 8.5 wild-type and mutant embryos. We show that *T*, cooperatively with WNT signaling, controls the progenitor state and the switch toward the mesodermal fate. *Sox2* acts antagonistically and promotes neural development. *T* is also involved in remodeling the chromatin for mesodermal development. *Tbx6* reinforces the mesodermal fate choice, represses the progenitor state, and confers paraxial fate commitment. Our findings refine previous models and establish molecular principles underlying mammalian trunk development, comprising NMP maintenance, lineage choice, and mesoderm formation.

## INTRODUCTION

The trunk of murine embryos forms by continuous recruitment of cells generated in the primitive streak (PS), node-streak border (NSB), and caudal lateral ectoderm (CLE), located at the caudal end of the embryo, into the neural or mesodermal lineage thereby elongating the body anlage (Wilson et al., 2009). The source of cells giving rise to the spinal cord and mesodermal tissues, comprising the vertebral column, skeletal musculature, ventral body wall, kidneys, gonads, limbs, and others, is a resident progenitor cell type with self-renewing capability, the

neuro-mesodermal progenitor (NMP) (Garriock et al., 2015; Gouti et al., 2014; Henrique et al., 2015; Rodrigo Albors and Stoye, 2016; Tzouanacou et al., 2009; Wymeersch et al., 2016). NMPs are characterized by co-expression of the stem cell factor and key neural progenitor cell (NPC) regulator *Sox2*, and of the pan-mesodermal control factor *Brachyury* (*T*) (Bergslund et al., 2011; Boyer et al., 2005; Herrmann et al., 1990; Kispert et al., 1995; Wilkinson et al., 1990; Wymeersch et al., 2016). Recent grafting experiments carried out in embryonic day 8.5 (E8.5) embryos showed that the fate of NMP descendants correlates with the relative levels of *Sox2* or *T* protein expression and with their position in the PS/CLE (Wymeersch et al., 2016). This study also confirmed previous reports demonstrating the dependence of mesoderm formation in the trunk on WNT signaling (Jurberg et al., 2014; Martin and Kimelman, 2012; Takada et al., 1994; Tsakiridis et al., 2014). *T* was the first regulator shown to play an essential role in trunk development: *T* knockout embryos show a bulky PS, lack posterior mesoderm, arrest axis elongation, and therefore lack the trunk and tail (Chesley and Dunn, 1936; Herrmann, 1991; Yanagisawa et al., 1981). *T* acts together with *Wnt3a*, *Fgf1*, and *Fgf4/Fgf8* in trunk mesoderm development and is a target of WNT and fibroblast growth factor (FGF) signaling (Chapman and Papaioannou, 1998; Ciruna and Rossant, 2001; Naiche et al., 2011; Yamaguchi et al., 1999). Studies in zebrafish have shown that *T*, by controlling *Wnt3a* and the retinoic acid inhibitor *Cyp26a1*, establishes the mesodermal progenitor niche (Martin and Kimelman, 2010). Work in *Xenopus* demonstrated the importance of the *T* ortholog *Xbra* in neuro-mesodermal stem cell maintenance and in mesoderm formation (Gentsch et al., 2013). Thus, the combined genetic and fate mapping data would suggest that *T* is involved in the maintenance and mesodermal fate choice of NMPs. Recent reports, however, claimed that *Tbx6* in mouse and its functional homolog in zebrafish *Tbx16* suppress *Sox2* and the neural fate, and recruit progenitors into the mesodermal lineage (Bouldin et al., 2015; Takemoto et al., 2011). To resolve these partly conflictive conclusions, we have carried out a broad investigation of the role and interaction of *T*, *Sox2*, and *Tbx6* in NMPs and in fate decisions of NMP descendants during trunk development, based



(legend on next page)

on genome-wide datasets derived from purified embryonic cell populations.

## RESULTS

### Transcriptome Analysis Reveals Five Distinct Gene Groups Characterizing NMPs and Their Descending Lineages

To characterize NMPs *in vivo* and to investigate the fate choice of their descendants toward mesodermal or neural differentiation, we established a double reporter (*T*:mCherry, hereafter  $T^{mCh}$ ; *Sox2*:Venus, hereafter  $Sox2^{Ve}$ ) embryonic stem cell (ESC) line for the generation of embryos via tetraploid complementation (Eakin and Hadjantonakis, 2006). To this end, BAC clones containing the *T* gene or the *Sox2* gene were modified by replacing their respective start codons with fluorescent reporter cassettes via recombineering (Muyrers et al., 1999). Tetraploid complementation assays showed that the expression of the fluorescent reporters in E8.5 embryos reflects the endogenous pattern of *T* or *Sox2* expression and mark the mesodermal or the neural lineage, respectively (Figure 1A; Wymeersch et al., 2016). We detected the *T* reporter in the caudal epiblast, PS, nascent and early differentiating mesoderm, as well as the node and notochord, while the *Sox2* reporter is expressed in the epiblast, primitive streak, neuroectoderm, and brain, as well as in endoderm.  $T^{mCh+}/Sox2^{Ve+}$  cells with low reporter activity are prominent in the caudal epiblast, recently identified as the major site where NMPs reside, while high activity of both reporters was detected in the anterior streak near the node. We isolated caudal ends (up to pre-somite –2) of Theiler stage 13 (TS13; E8.5) embryos and subjected the single-cell suspension to fluorescence-activated cell sorting (FACS) analysis (Figures 1B and S1A). The majority (~50%) of labeled cells expressed the *T* reporter ( $T^{mCh+}$ ), a substantial fraction (~25%) was double positive ( $T^{mCh+}/Sox2^{Ve+}$ ), and a minor fraction (~10%) expressed the *Sox2* reporter ( $Sox2^{Ve+}$ ), representing early mesodermal, putative NMP, and neuroectodermal cells, respectively.

We subjected the three cell pools to transcriptome analysis by RNA sequencing (RNA-seq). We filtered the data for genes, which are differentially expressed across the three samples. Clustering of the RNA-seq data using per gene normalized FPKM values (fragments per kilobase of transcript per million mapped reads) revealed five distinct gene groups (Figure 1C and Table S1; see Figure S1B for examples). Group I comprises 154 genes, which are most highly expressed in  $T^{mCh+}/Sox2^{Ve+}$  cells and downregulated in  $T^{mCh+}$  and  $Sox2^{Ve+}$  cells. This group contains genes known to be expressed in nascent mesoderm

(*Fgf3*, *Fgf4*, *Fgf8*, *Fgf15*, *Rspo3*, *Dkk1*; see <http://mamep.molgen.mpg.de> for expression patterns), and genes essential for axial elongation and trunk development (*Wnt3a*, *T*, *Msn1*). Group II genes (250) are already significantly expressed in  $T^{mCh+}/Sox2^{Ve+}$  cells, mostly further upregulated in  $T^{mCh+}$  but downregulated in  $Sox2^{Ve+}$  cells. This group contains early mesodermal control genes (e.g., *Tbx6*, *Snai1*, *Lef1*, *Meox1*, *Mesp1*, *Foxc1*). A different set of mesodermal control genes is collected in group III (375 genes). Their expression is low in  $T^{mCh+}/Sox2^{Ve+}$  cells and high in  $T^{mCh+}$  cells, suggesting later activation in the mesodermal lineage. This group contains several regulators of the lateral and intermediate mesoderm (e.g., *Foxf1*, *Tbx3*, *Gata1*, *Gata3*, *Gata6*, *Hand1*, *Hand2*, and *Osr1*, respectively). A mirror image of groups II and III is provided by groups IV (261 genes) and V (362 genes) containing regulators of the neural lineage. Early neural lineage genes expressed in  $T^{mCh+}/Sox2^{Ve+}$  and mostly further upregulated in  $Sox2^{Ve+}$  cells (group IV) but downregulated in  $T^{mCh+}$  cells, comprise the control factors *Sp8*, *Nkx1-2*, and *Pax3*. Later neural regulators, which are strongly expressed in  $Sox2^{Ve+}$  cells (group V) but not in  $T^{mCh+}$  cells, comprise *Sox1*, *Sox3*, *Sox21*, *Olig3*, *Irx1*, and *Zic5*. Due to its strong upregulation in the neural lineage *Sox2* also clustered in this group. To verify our clustering approach, we generated box plots representing the distribution of expression values within each group (Figure 1D). The graphs confirm the visual interpretation of gene clustering shown in Figure 1C.

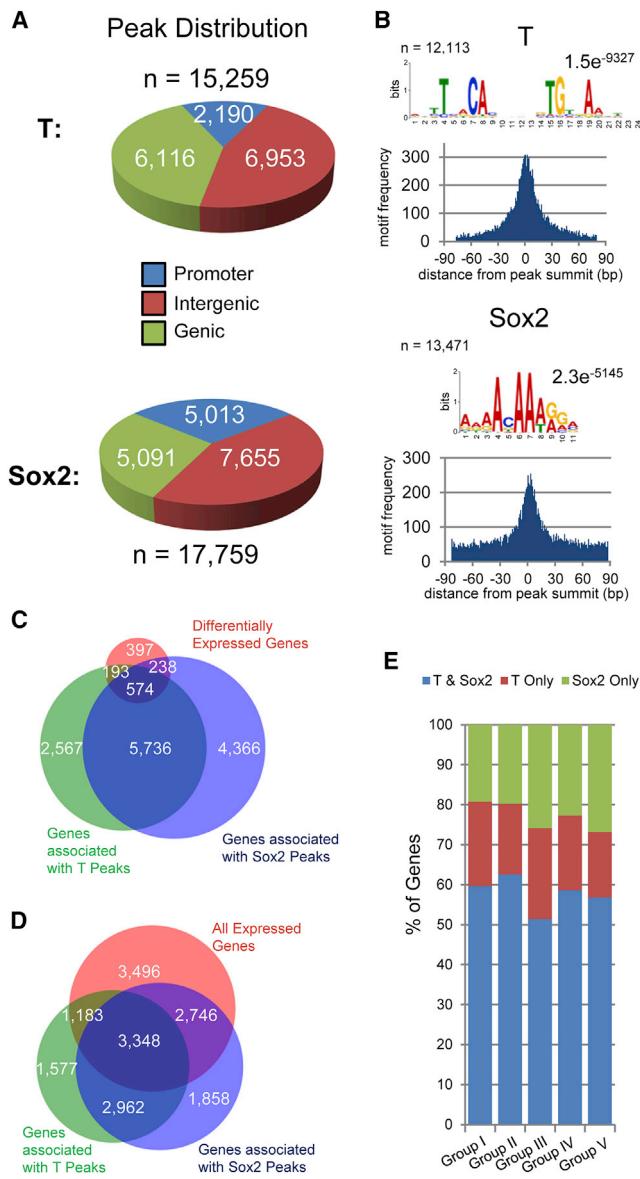
Thus, the expression profiles of control genes, which are essential for axial elongation, mesoderm formation, or neurogenesis during trunk development, confirmed that the neural and mesodermal lineage derive from *T/Sox2* double-positive cells.

### Single-Cell Transcriptome Analyses Distinguish NMPs from Cells Undergoing the Lineage Choice

The distinct expression profiles of group I genes (downregulated in both lineages) and group II/group IV genes (upregulated in one and downregulated in the other lineage) suggested that  $T^{mCh+}/Sox2^{Ve+}$  cells might represent a heterogeneous cell population comprising NMPs and cells undergoing lineage choice. In addition, *Msn1* is a mesoderm control gene and not supposed to be expressed in NMPs (Chalamalasetty et al., 2014; Yoon and Wold, 2000). To analyze the cellular composition of the  $T^{mCh+}/Sox2^{Ve+}$  cells, we performed single-cell transcriptome analysis.  $T^{mCh+}/Sox2^{Ve+}$  cells were isolated from TS13 embryos, FACS sorted, and subjected to single-cell capture on a Fluidigm C1 IFC. We obtained quality-controlled transcriptome data for 45 single cells by RNA-seq. Principal component analysis identified four distinguishable cell groups representing NMPs, prospective

**Figure 1. Transcriptome Analysis of NMPs, Nascent Mesoderm, and Neuroectoderm Isolated from Mouse Embryos Reveals Five Differentially Expressed Gene Groups**

- (A) TS12 (E8.25) embryo expressing *T*:mCherry and *Sox2*:Venus reporters imaged by lightsheet microscopy. Left: 3D model; an optical section through the node region (right) is indicated by a dashed bar.
  - (B) FACS profile of single cells derived from caudal ends of TS13 (E8.5) embryos, used for transcriptome analysis of  $T^{mCh+}/Sox2^{Ve+}$ ,  $T^{mCh+}/Sox2^{Ve-}$ , and  $T^{mCh-}/Sox2^{Ve+}$  subpopulations.
  - (C) Heatmap of k-means clustered differentially expressed genes, using per gene normalized FPKM values; groups (I–V) are indicated to the left, with important regulators within each group to the right.
  - (D) Box plots representing the distribution of FPKM values within each group, validating the clustering approach.
  - (E) PCA plot of cell types identified using a semi-supervised clustering and cell type imputation approach, displaying levels of marker gene expression.
  - (F) PCA plot showing the cell trajectories based on differential gene expression between cell types identified in (E).
- See also Figure S1 and Table S1.



**Figure 2. T and Sox2 Co-occupy a Large Fraction of NMP and Lineage Control Genes**

- (A) Genomic distribution of T (top) and Sox2 (bottom) peaks from NMPs derived by *in vitro* differentiation of ESCs.
  - (B) Most significant *de novo* motifs detected at T (top) or Sox2 (bottom) peaks, and corresponding frequency distribution plots  $\pm 90$  bp around peak summits.
  - (C) Venn diagram overlapping genes differentially expressed between NMPs and descending lineages (as in Figure 1C) (orange), and genes associated with T and/or Sox2 peaks.
  - (D) Same analysis as in (C), but for all genes expressed across the three cell types irrespective of differential expression.
  - (E) Bar plots showing the percentage of 1,005 genes associated with T and/or Sox2 sites, split into the five gene groups.
- See also Figure S2 and Table S2.

neural or mesodermal cells, and cells, which could not be assigned to any group (Figure 1E). All except three cells expressed the NMP markers *T*, *Sox2*, *Wnt3a*, *Rspo3*, and *Fgf8*. The NMPs were identified by high-level expression of the NMP marker

genes, but no expression of *Tbx6* or other lineage-specific control genes from groups II–V. The prospective mesodermal cells were characterized by the mesoderm markers *Tbx6* and *Msgn1*, and the prospective neural cell group by co-expression of *Sox2* (14/15), *Pax3* (6/15), and *Nkx1-2* (3/15). The set of 17 cells, which could not be assigned to any cell group, showed expression of the NMP marker genes plus *Tbx6*, as well as *Pax3* or *Nkx1-2* in some of them. The latter cells and the prospective mesodermal and neural cells showed overlapping expression of lineage-specific regulators at variable intensity.

The cell trajectory, based on differential gene expression between the four cell groups, illustrates the distinct nature but close relationship of NMPs with the undefined cells (Figures 1F and S1C). Some of the latter are distributed along the trajectory toward the branch point between the prospective mesodermal and neural cells. Thus, we conclude that this undefined cell group represents a transient cell state consisting of NMP descendants undergoing lineage choice.

Overall, the single-cell analysis demonstrated the heterogeneous nature of *T<sup>mCh+</sup>/Sox2<sup>ve+</sup>* cells consisting of NMPs, a distinct transient cell type expressing lineage control genes, and cells already determined for neural or mesodermal differentiation.

### T and Sox2 Co-occupy Genes of Both Lineages

Next we investigated the involvement of *T* and *Sox2* in the control of the NMP and lineage-specific gene expression profiles. To address this point we analyzed the genomic binding sites of *T* protein and *Sox2* protein in NMPs by chromatin immunoprecipitation sequencing (ChIP-seq). Due to limited embryonic material we used an *in vitro* differentiation model to obtain sufficient numbers of progenitor cells (Gouti et al., 2014). We differentiated the same reporter ESC line *in vitro* that we had utilized for the generation of embryos. FACS profiling of single-cell suspensions revealed that at day 3 of differentiation more than 80% of the cells are *T<sup>mCh+</sup>/Sox2<sup>ve+</sup>* double positive (Figures S2A and S2B). Western blotting confirmed that *T* protein and *Sox2* protein are co-expressed in reporter as well as in wild-type ESCs at this stage (Figure S2C). We therefore differentiated wild-type ESCs for 3 days and performed *T* and *Sox2* ChIP-seq experiments on bulk cultures. Initial ChIP-peak analyses revealed 15,259 binding sites for *T* and 17,759 sites for *Sox2*, mostly (around 85% for *T*, 72% for *Sox2*) in genic and intergenic regions, at putative regulatory elements (Figure 2A). The motif analysis of binding regions revealed a palindromic *T*-box motif and the canonical *Sox2* motif as enriched at their respective peaks (Figure 2B) (Kispert and Herrmann, 1993; Luna-Zurita et al., 2016; Marson et al., 2008; Shen et al., 2011). We defined potential target genes by associating the ChIP-peaks of either factor with genes located in their immediate neighborhood (Figure 2C). 13,674 genes were associated with either factor, about half of them with binding sites of both. Among the 1,402 differentially expressed genes identified by cluster analysis of RNA-seq data from FACS-purified embryonic cells (Figure 1C), 193 genes (13.8%) were found to be associated with binding sites for *T* and 238 (17%) with binding sites for *Sox2*. For 397 genes (28.3%) an association with *Sox2* or *T* peaks was not detected by our method. Strikingly, 574 (40.9%) of the differentially expressed genes are associated with binding sites of both factors, many of them located within

1 kb (Figures 2C [see Figure S2D for examples] and S2E; Table S2). Thus, in total 71.7% of the differentially expressed genes are associated with either T or Sox2 binding. Moreover, even among all 10,773 genes expressed in at least one of the three embryonic RNA samples we found 7,277 (67.5%) genes associated with peaks of either factor, 3,348 (46%) of the latter with peaks of both (Figure 2D). In addition, we found that all five differentially expressed gene groups (Figure 1C) show a similar distribution of binding sites of T (16%–23%), Sox2 (19%–27%), or both factors (51%–63%), and thus no enrichment of either factor in any of the five groups is evident (Figure 2E).

Together, these data show that the vast majority of genes that are differentially expressed in NMPs and their neural or mesodermal descendants are occupied by T or Sox2, and mostly by both transcription factors, irrespective of the cell type or lineage. This suggests that T and Sox2 together control the lineage choice of NMP descendants.

### T Activates NMP and Mesodermal Control Genes, in Cooperation with WNT Signaling, and Represses Neural Control Genes Including Sox2

We further investigated the role of T in the mesodermal fate choice of NMP descendants. Previous reports had shown that T is essential for mesoderm formation in the trunk and expressed in NMPs (Gouti et al., 2014; Herrmann, 1991; Wymeersch et al., 2016). However, whether T controls the lineage choice of NMPs toward mesoderm, and if so by what molecular mechanism this is achieved, has not been addressed. We generated embryos from wild-type and  $T^{2J/2J}$  mutant ESCs expressing the T reporter (Figure S3A) and isolated  $T^{mCh+}$  cells from stage-matched embryos (Figure S3B). We dissected the caudal ends of TS13 embryos at the onset of a visible mutant phenotype. We sorted  $T^{mCh+}$  cells by FACS and subjected the cell preparations to RNA-seq analysis. The data revealed 689 dysregulated genes, 333 downregulated and 356 upregulated, in the  $T^{2J}$  mutant (Figure 3A). We found NMP markers, such as *Wnt3a*, *Fgf8*, *Rspo3*, and control genes of all mesodermal sublineages, such as *Tbx6* and *Msx1* (paraxial mesoderm), *Foxf1* (lateral plate mesoderm), and *Osr1* (intermediate mesoderm) among the downregulated genes. We found important neural lineage regulators for early and later neural differentiation, including *Sox2*, *Sox3*, and *Sox1*, among the upregulated genes (Bergsland et al., 2011). We identified direct T targets by overlapping the sets of dysregulated genes and of genes associated with T peaks (Figure 3B). Among 335 direct T targets, we found the NMP markers *Sox2*, *Wnt3a*, *Rspo3*, and *Fgf8*, the mesodermal control genes *Tbx6*, *Osr1*, and *Msx1*, and the neural regulators *Nkx1-2*, *Sox1*, and *Sox3* (Figure 3B). We then divided the direct targets into activated genes and repressed genes and analyzed their relationship with the five groups in Figure 1C (Figure 3C). This analysis revealed that most direct target genes activated by T belong to group I comprising NMP genes, and to the mesodermal groups (II and III). In contrast, genes directly repressed by T are found in the neural groups (IV and particularly group V), including *Sox2*, *Sox1*, and *Sox3*.

In summary, these data show that key NMP, mesodermal, and neural regulators are directly targeted and controlled by T, and demonstrate a dual role of T as activator of NMP and mesodermal control genes, and as repressor of neural control genes.

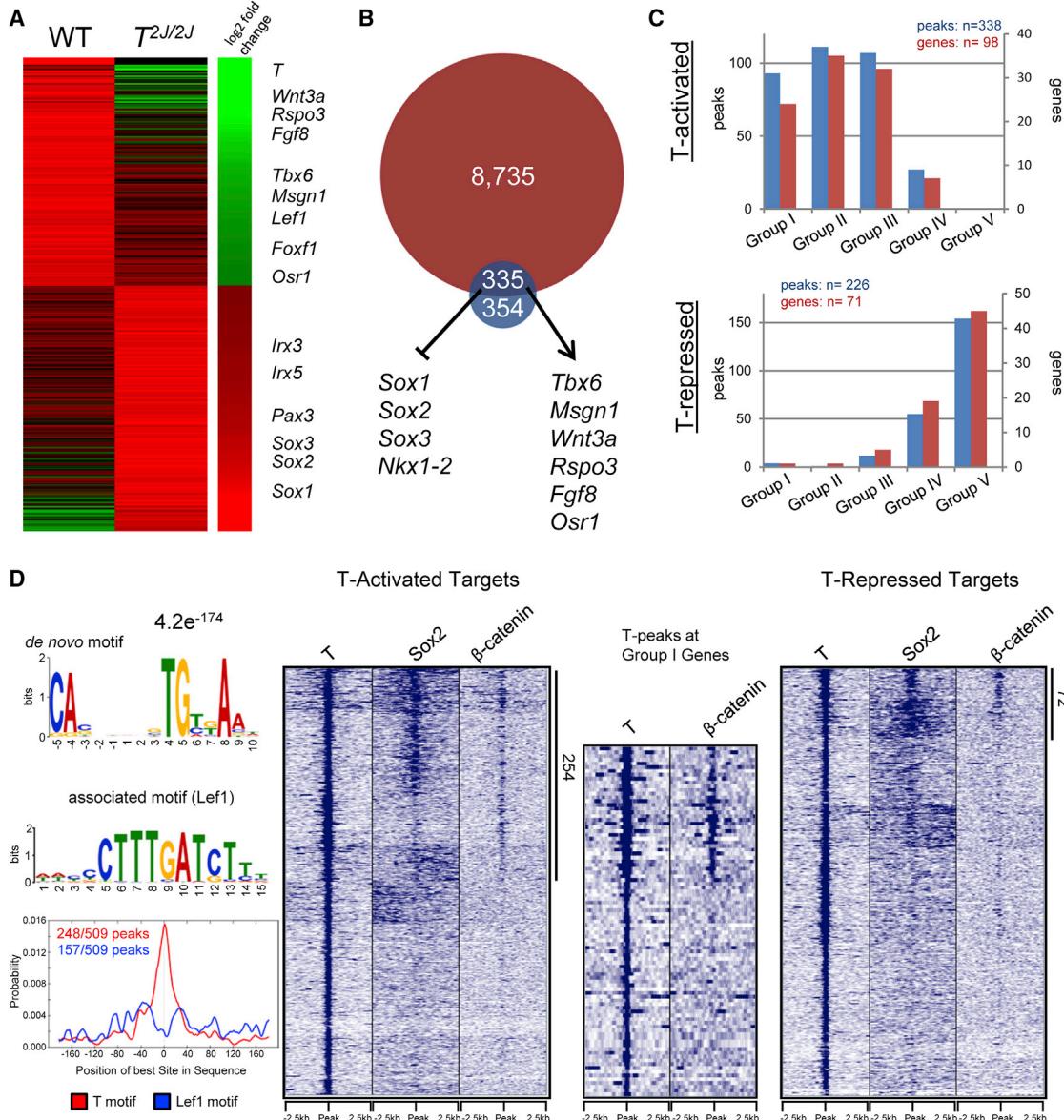
Next we asked how T might distinguish activated and repressed targets. A *de novo* motif analysis within a 400-bp window around T peaks revealed the palindromic T motif as most enriched at activated (248/509 peaks) as well as at repressed target genes (311/444 peaks; data not shown) (Figure 3D). Since co-factor binding might be involved in differential target control, we scrutinized T peak regions for co-occurring motifs. Among the most prominent motif co-occurring with the T site at T-activated genes is the LEF/TCF site (at 157/509 peaks) mediating the WNT signaling response in the nucleus. To further underpin this finding, we performed a ChIP-seq experiment to identify the genomic binding sites of *Ctnnb1* ( $\beta$ -catenin) protein in NMP cells differentiated *in vitro*. We identified 4,411 binding sites, the vast majority (89%) located in intergenic or genic regions (Figure S3C). The motif analysis of binding regions revealed the consensus Lef1 site at 65% of the peaks. To check for co-localization of T bound and  $\beta$ -catenin bound sites at T-activated genes, we clustered the latter sites identified in the vicinity of the former (Figure 3D). We found that  $\beta$ -catenin bound near 254/509 T sites, including 36/93 sites associated with NMP group genes. Thus, WNT signaling indeed cooperates with T in controlling a large fraction of T-activated genes.

However, when we clustered Sox2 bound sites detected by ChIP-seq in the vicinity of T bound sites in NMPs, we found a substantial co-occurrence of both sites at both the T-repressed and the T-activated targets (Figure 3D; example shown in Figure S3D).  $\beta$ -Catenin binding also co-occurred together with Sox2 binding near 72/444 T sites associated with T-repressed genes, suggesting cooperation of T and  $\beta$ -catenin also in repressing at least some of the neural lineage genes.

The combined data suggest a mutual interaction of both T and Sox2 in NMPs on target control regions of genes promoting the neural or mesodermal lineage. Since our data show that T is required for activating mesodermal and for suppressing neural control genes, and Sox2 is a known neural control factor binding to both neural and mesodermal control genes in NMPs, we conclude that T and Sox2 antagonize each other in cells undergoing the lineage choice toward either the neural or mesodermal fate. However, the molecular mechanism underlying the antagonism between T and Sox2 remains unclear and needs to be addressed in future work.

### T Controls Chromatin Remodeling Promoting Mesodermal Lineage Development

Since we found several thousand T binding sites in intergenic regions, while only a few hundred genes were affected by T loss of function, we wondered whether T might be involved in shaping the chromatin for mesodermal development. Therefore we asked whether T is able to control the chromatin state. We introduced the Sox2:Venus reporter into  $T^{2J/2J}$  mutant ESCs expressing the T:mCherry reporter and generated embryos (Figure S4A). We isolated the  $T^{mCh+}/Sox2^{Vet+}$ ,  $T^{mCh+}$ , and  $Sox2^{Vet+}$  subpopulations from TS12 and TS13 embryos by FACS (Figure S4B) and performed ATAC-seq (assay for transposase-accessible chromatin with high-throughput sequencing) analysis (Buenrostro et al., 2013). After verifying comparable fragment length distributions (Figure S4C), the analysis was performed using 50-kb bins of normalized ATAC-seq data to investigate global changes in chromatin accessibility (Figure S4D). We computed  $\log_2$  ratios



**Figure 3. T Activates Mesodermal and Represses Neural Lineage Genes**

(A) Ranked heatmap showing  $\log_2$  fold changes of genes dysregulated in  $T^{2J/2J}$  embryos. Examples of important regulators are indicated to the right. WT, wild-type.

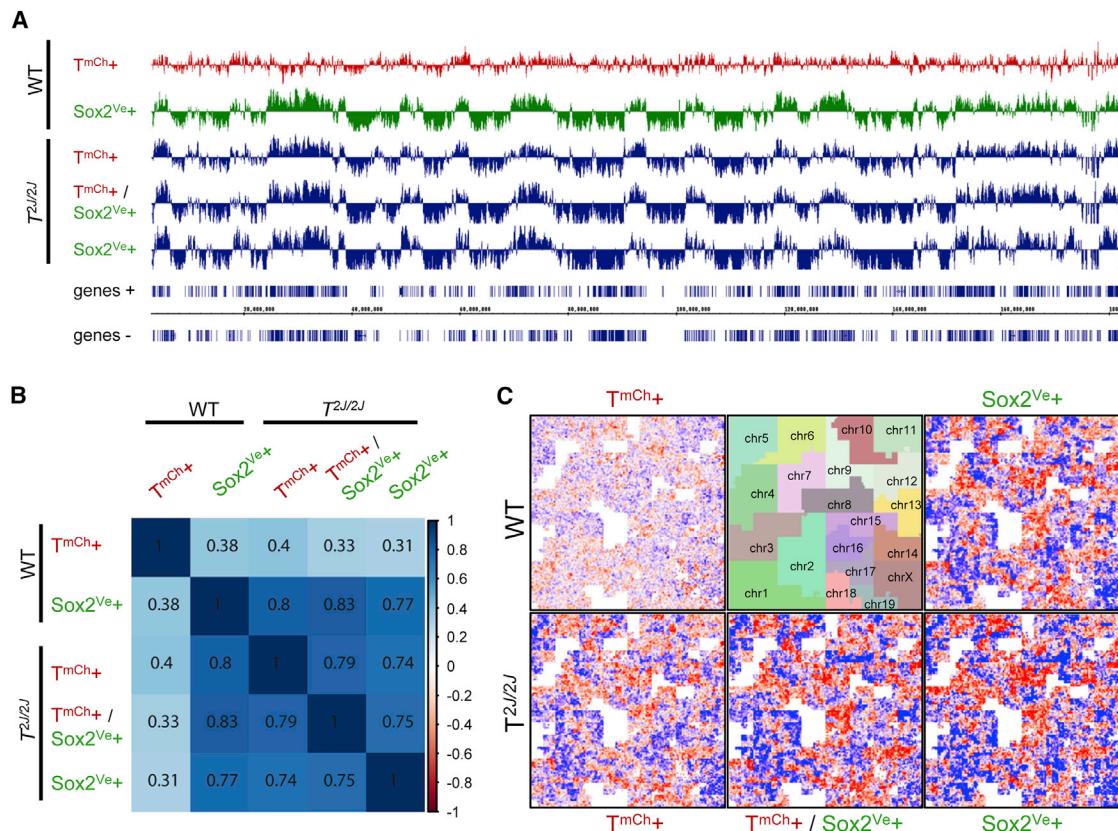
(B) Venn diagram showing the overlap between genes associated with T peaks and genes dysregulated in  $T^{2J/2J}$  embryos; examples of direct T targets are shown.

(C) Bar plots showing the number of T peaks (blue, left axis) and associated genes (red, right axis) in groups I–V separately for activated (top) and repressed (bottom) T targets.

(D) T binding sites and heatmaps of clustered ChIP-seq signals for T, Sox2, and  $\beta$ -catenin at T peaks associated with genes activated (left) or repressed (right) by T. For T-activated genes, the motif analysis and a heatmap illustrating the overlap between T and  $\beta$ -catenin binding sites at group I genes are shown. See also Figure S3.

using  $T^{mCh+}/Sox2^{ve+}$  cell data from TS12 and TS13 wild-type embryos as a reference. At both developmental stages, mesodermal ( $T^{mCh+}$ ) and neuroectodermal ( $Sox2^{ve+}$ ) cells display distinct global chromatin conformation patterns indicating extensive differential chromatin remodeling between the NMP state and either cell lineage (Figures 4A and S5A). While we observe wider regions of chromatin accessibility changes in  $Sox2^{ve+}$  cells, the changes revealed in the  $T^{mCh+}$  population

are mostly more local. Strikingly, the  $T^{mCh+}$  cell population of  $T^{2J/2J}$  mutants displayed the neuroectodermal ( $Sox2^{ve+}$  cell) chromatin signature, suggesting that T is required for generating the distinct chromatin signature of the mesodermal lineage (Figure 4A). Moreover,  $T^{mCh+}/Sox2^{ve+}$  cells derived from  $T^{2J/2J}$  mutants also displayed the chromatin signature of the neural lineage, suggesting that in T mutants NMPs and their descendants undergo chromatin remodeling for neural differentiation.



**Figure 4. T Controls the Chromatin State of NMPs**

(A) Genome browser snapshot of chromosome 2 showing ATAC-seq data profiles derived from FACS-purified cells of wild-type (WT) or  $T^{2J/2J}$  embryos relative to stage-matched wild-type NMP cells.

(B) Genome-wide correlation matrix of all ATAC-seq datasets derived from indicated cell types and genotypes at TS13 relative to their respective stage-matched wild-type NMP dataset.

(C) Heatmap of the genome-wide Hilbert curve of global ATAC-seq data (relative to the NMP dataset) derived from FACS-purified  $T^{mCh+}$ ,  $T^{mCh+}/Sox2^{Ve+}$ , or  $Sox2^{Ve+}$  cells from TS13 wild-type (WT) or mutant ( $T^{2J/2J}$ ) embryos. The map of the underlying Hilbert curves shows the location of each chromosome. See also Figures S4 and S5.

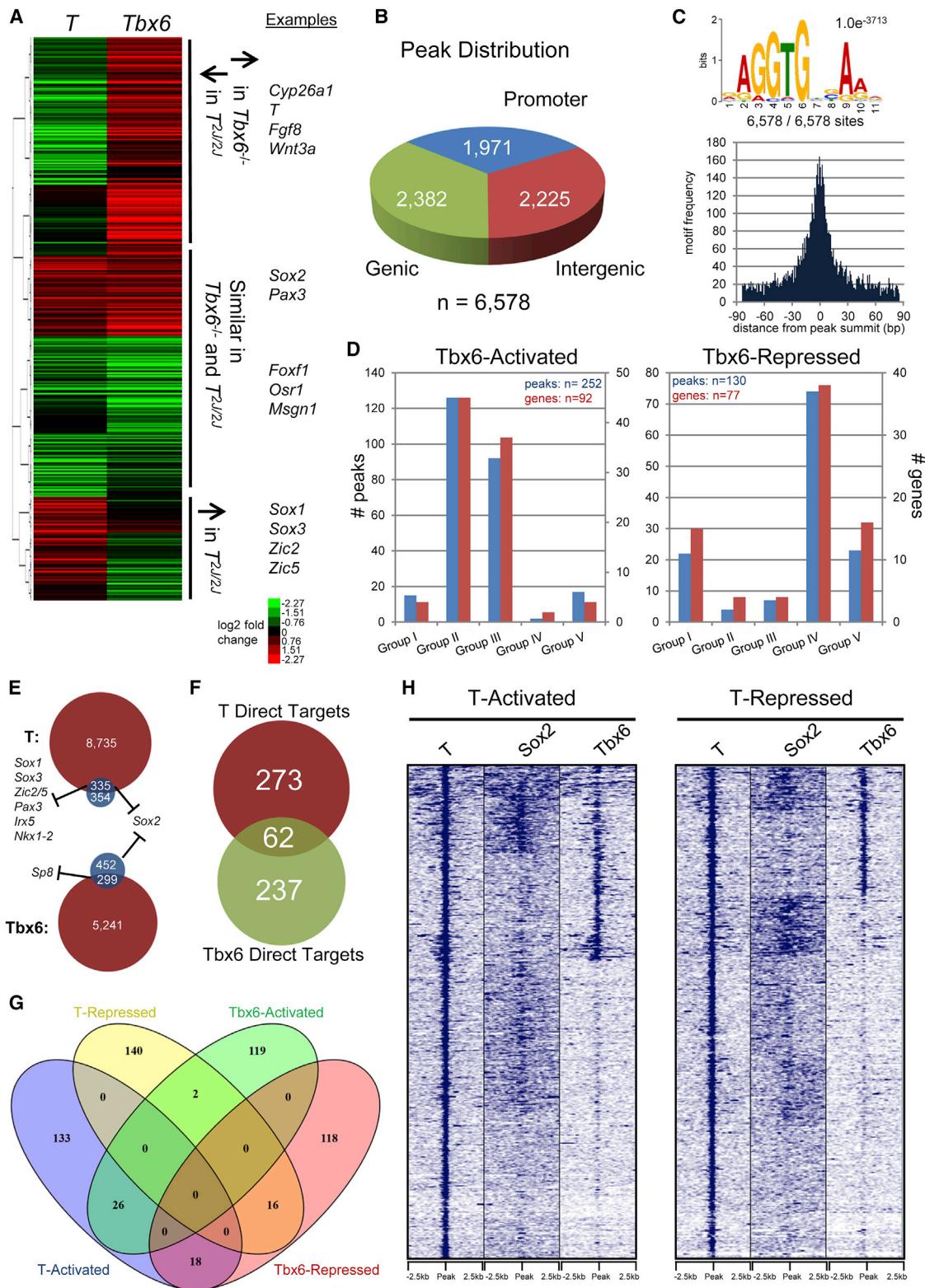
Calculating global correlations across the whole genome showed a strong correlation of  $T^{mCh+}$  or  $T^{mCh+}/Sox2^{Ve+}$  cells from mutant with  $Sox2^{Ve+}$  cells from wild-type embryos, both at TS13 ( $r = 0.8$  or  $r = 0.83$ , respectively) (Figure 4B) and TS12 ( $r = 0.71$ ) (Figure S5B). These correlations are also apparent upon plotting of the ATAC-seq data for TS13 (Figure 4C) and TS12 (Figure S5C) across the genomes using Hilbert curves.

In summary, the chromatin accessibility data reveal genome-wide chromatin remodeling from NMPs to the neural or mesodermal lineage, which display distinct chromatin patterns. T is essential for remodeling the chromatin for mesodermal development.

#### Tbx6 Represses Progenitor Genes and Promotes Mesoderm Formation Downstream of T

It has been proposed that the lineage decision for mesodermal development is determined by Tbx6-dependent regulation of *Sox2* (Takemoto et al., 2011). However, in *Tbx6* mutants mesoderm formation is initiated, whereas the paraxial tissue formed in the trunk eventually switches to the neural fate (Chapman and Papaioannou, 1998). The latter suggests that *Tbx6* is not

essential for the fate decision as such, but is sustaining the paraxial fate. To investigate the different roles of T and Tbx6 in the lineage choice and in mesoderm development in more detail, we compared the sets of direct target genes of either factor. For this purpose we introduced a *Tbx6:mCherry* reporter BAC into wild-type and *Tbx6*<sup>-/-</sup> ESCs and generated embryos (Figure S6A). We isolated *Tbx6:mCherry* cells from caudal ends of dissected TS14 embryos (up to pre-somite -2 in wild-type embryos) by FACS (Figure S6B) and performed transcriptome analyses. Genes dysregulated in either *Tbx6* or *T* mutant embryos were compared (Figure 5A). We found a large overlap of genes commonly regulated by T and Tbx6, in particular mesodermal control genes of the paraxial (*Msgn1*), intermediate (*Osr1*), and lateral (*Foxf1*) sublineages, which are downregulated in either mutant (see Figure S6C for examples). However, two gene groups showed opposite behavior. Many genes showing highest expression in NMPs (group I, Figure 1C) are downregulated in the *T* mutant, but upregulated in *Tbx6*<sup>-/-</sup> embryos, indicating positive regulation by T and repression by Tbx6 (see Figure S6D for examples). In contrast, many neural transcription factor genes, among them *Sox1* and *Sox3*, are upregulated in



**Figure 5. Direct Target Comparisons Reveal Different Roles of T and Tbx6 in NMPs, Lineage Choice, and Mesoderm Formation**

(A) Heatmap showing hierarchical clustering of  $\log_2$  fold-change expression values of genes dysregulated in  $T^{2J/2J}$  and  $Tbx6^{-/-}$  embryos compared with wild-type. Major clusters and examples of important regulators are shown to the right; arrows indicate up- or downregulation.

(B) Peak analysis of Tbx6 ChIP-seq data from *in vitro* differentiated paraxial (pre-somitic) mesoderm shows a similar prevalence of Tbx6 binding sites in promoters, genic, and intergenic regions.

*T* mutants and thus are repressed by *T*, but unchanged in *Tbx6*<sup>-/-</sup> embryos (see Figure S6E for examples). An important exception is *Sox2*, which is upregulated in either mutant, indicating repression by both *T* and *Tbx6*.

To identify *Tbx6* binding sites and direct targets (dysregulated genes in *Tbx6*<sup>-/-</sup> embryos associated with *Tbx6* sites), we performed ChIP-seq on ESCs differentiated to pre-somitic mesoderm *in vitro* (Sudheer et al., 2016). In comparison with *T*, a larger proportion of *Tbx6* binding sites was found at promoters rather than in genic or intergenic regions (Figure 5B), and the exclusive *Tbx6* binding motif is not a palindrome but a single T-box motif (Figure 5C). Thus *Tbx6* and *T* differ in their genomic binding site distribution patterns and DNA binding characteristics. To compare *T* and *Tbx6* targets in more detail, we first divided *Tbx6* targets into activated and repressed genes and also assigned them to the five groups derived in Figure 1C (Figure 5D). Like *T* (Figure 3C), *Tbx6* activates many mesodermal genes (groups II and III). Unlike *T*, however, only four NMP genes (group I) are activated by *Tbx6*. Instead NMP genes including *T* are mostly repressed by *Tbx6*. Also, several neural lineage genes (group IV and a few from group V) are repressed by *Tbx6*. However, most key neural transcription factors such as *Sox1*, *Sox3*, *Zic2*, *Zic5*, *Irx5*, *Pax3*, or *Nkx1-2* are not repressed by *Tbx6* (Figure 5E and data not shown). Exceptions are *Sp8*, and *Sox2*, which acts as both key NMP and neural control gene.

The comparison of direct targets of *T* and *Tbx6* revealed 62 genes regulated by both factors (Figure 5F; see Figures S6F–S6H for examples); 42 are either activated or repressed by both regulators (Figure 5G). Among 26 genes activated by *T* and *Tbx6* are the paraxial control genes *Msgn1*, *Foxc1*, *Meox1*, and *Lef1*, and components of the Notch pathway. Eighteen genes were found to be activated by *T* but to be repressed by *Tbx6* (Figure 5G), and those comprised the key genes acting in NMPs and early mesoderm development (*T*, *Wnt3a*, *Cyp26a1*, and *Fgf8*, *Fgf4*) (Gouti et al., 2014; Henrique et al., 2015; Martin and Kimelman, 2010; Naiche et al., 2011; Wymeersch et al., 2016).

Clustering of *T*, *Sox2*, and *Tbx6* ChIP-seq signals around *T* peaks associated with direct targets shows a strong overlap of *T* and *Tbx6* bound sites at genes activated by *T*, while there is less overlap at *T*-repressed genes (Figure 5H).

The combined data demonstrate that *Tbx6* acts downstream of *T* in the mesodermal lineage.

## DISCUSSION

Our study addresses the molecular mechanism of NMP maintenance and lineage choice in mouse based on global analyses of purified embryonic cell populations. This approach allowed a

more refined analysis of NMPs and of the molecular changes occurring from the progenitor state to the neural or mesodermal lineage than previously possible, and thus establishes principal mechanisms controlling early mammalian trunk development.

We provide convincing evidence that the fate choice of NMP descendants is determined by antagonistic interaction between *T* and *Sox2*. Both neural and mesodermal lineage-specific control genes are co-occupied by both transcription factors. We show that *T* activates mesodermal and represses neural lineage genes. Conversely, *Sox2* is a well-known regulator of neural progenitors (Bergsland et al., 2011) and might (directly or at least indirectly) act as repressor of mesodermal genes. The latter is based on the finding that mesodermal genes occupied by *Sox2* are downregulated in the neural lineage and that in *T* mutants, cells induced to form mesoderm (*T*<sup>mCh+</sup> cells in *T*<sup>2J/2J</sup> embryos) switch to the neural fate.

Using ATAC-seq data we show that the chromatin accessibility signatures change between the NMP state and the neural or mesodermal lineage, indicating that remodeling of the chromatin is involved in the formation of these distinct lineages. Strikingly, in the absence of *T* the chromatin landscape of *T*<sup>mCh+</sup>/ *Sox2*<sup>Ve+</sup> cells changes to the neural pattern. Mutant cells induced to form the mesodermal lineage (*T*<sup>mCh+</sup> cells in *T*<sup>2J/2J</sup> embryos) fail to undergo proper chromatin remodeling; they also show the neural chromatin signature. These data suggest that *T* is essential for global chromatin remodeling accompanying or required for mesodermal lineage development.

Several reports have provided evidence for an essential requirement of *Wnt3a* in the maintenance of NMPs and in mesoderm formation (Garriock et al., 2015; Jurberg et al., 2014; Martin and Kimelman, 2012; Takada et al., 1994; Wymeersch et al., 2016). Studies in zebrafish have revealed an autoregulatory loop between *Wnt3a* and *T* establishing the progenitor niche (Martin and Kimelman, 2010). *Wnt3a*, the extracellular WNT activator *Rspo3*, *Fgf8*, and other factors of the WNT and FGF pathways are direct targets of *T* in NMPs (Figures 1E and 3B), and *T* is a target of WNT and FGF signaling (Arnold et al., 2000; Ciruna and Rossant, 2001; de Lau et al., 2011; Martin and Kimelman, 2010; Yamaguchi et al., 1999). However, the lack of *T*, *Wnt3a*, or FGF signaling disrupts trunk formation (Chesley and Dunn, 1936; Ciruna and Rossant, 2001; Takada et al., 1994). Thus *T*, WNT, and FGF signaling together form a positive feedback (autoregulatory loop) mechanism required for NMP maintenance (Figure 6).

However, each of these factors/pathways is also essential for mesoderm formation, which requires exit from the progenitor state. How can these opposite roles co-exist, and how can NMP maintenance work despite the antagonism between *Sox2* and *T*?

(C) *De novo* peak analysis identifies the (single) T-box motif as most significant and enriched below *Tbx6* peaks.

(D) Bar plots showing the distribution of peaks (blue, left axis) and associated genes (red, right axis) among groups I–V separately for target genes activated (left) or repressed (right) by *Tbx6*.

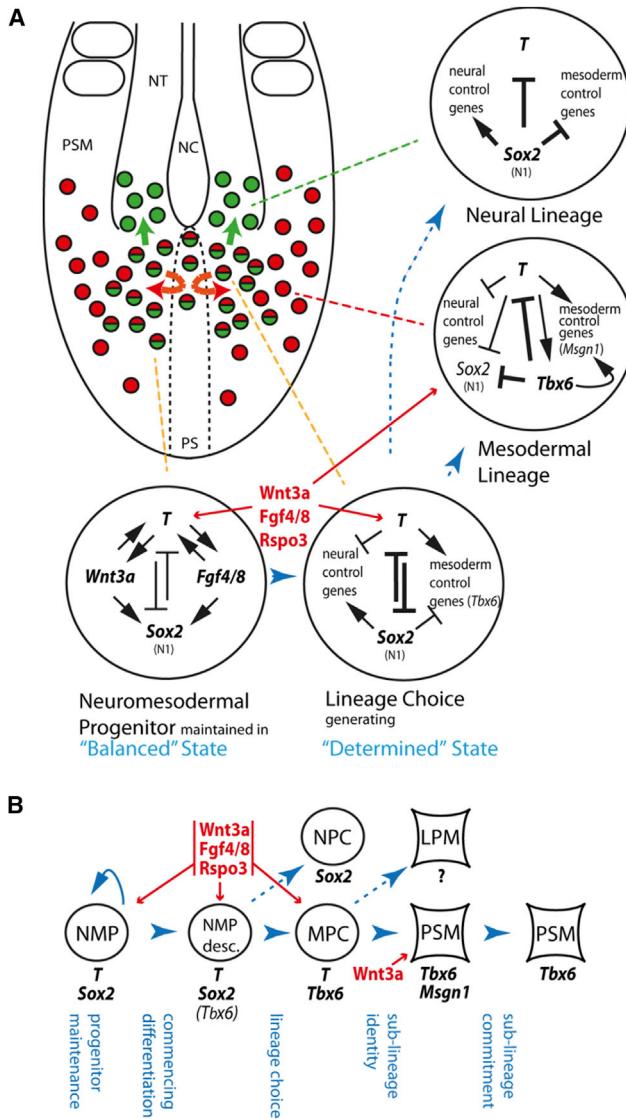
(E) Venn diagram showing the overlap between genes associated with *T* or *Tbx6* peaks and genes dysregulated in the respective mutant embryos. Important neural lineage regulators are indicated.

(F) Venn diagram showing overlap between direct *T* and *Tbx6* target genes.

(G) A 4-way Venn diagram displaying the overlaps between activated and/or repressed direct *T* and *Tbx6* target genes.

(H) Heatmap representing k-means clustered ChIP-seq signals of *T*, *Sox2*, and *Tbx6* binding around *T* peaks associated with genes activated (left) or repressed (right) by *T*.

See also Figure S6.



**Figure 6. Model: Basic Mechanisms of Trunk Development Comprising Progenitor Maintenance, Lineage Choice, and Commitment**

(A) NMPs (green/red circles) are maintained by an autoregulatory loop involving T and WNT/FGF signaling, which also control Sox2, together generating a “balanced” state. NMP descendants undergo the lineage choice involving antagonistic activities of T/WNT/FGF and Sox2, generating a “determined” state; lineage control genes, e.g., *Tbx6*, are upregulated and reinforce the lineage choice. Pre-neural cells (green) remain in the epiblast (straight green arrow), while nascent mesodermal cells (red) detach and migrate through the primitive streak (PS, indicated by red curved arrows).

(B) Paraxial (pre-somitic) mesoderm (PSM) formation is a multi-step process controlled by the WNT targets *T*, *Tbx6*, and *Msn1*. *Tbx6* supports *T* in the mesodermal fate choice; their target *Msn1* confers PSM identity and *Tbx6* PSM commitment.

For details see text. Black arrows indicate activation; black bars indicate inhibition; Wnt3a, Fgf4/8, Rspo3 (red color) indicate extracellular growth factors; gene symbols are shown in black. NMP, neuro-mesodermal progenitor; MPC, mesoderm progenitor cell; LPM, lateral plate mesoderm; NC, notochord; NPC, neural progenitor cell; NT, neural tube.

Sox2 is essential for the “stemness” of ESCs (Boyer et al., 2005). It is conceivable that Sox2 is required for maintaining the progenitor state of NMPs and for preventing differentiation. In NMPs (in CLE) Sox2 is activated by the N1 enhancer, which is controlled by WNT and FGF signaling (Takemoto et al., 2011). Thus, NMP maintenance might directly depend on the activation of Sox2 by the T/WNT/FGF autoregulatory loop. Both activities might be necessary to generate a balanced state (Figure 6). When T, WNT, or FGF signaling are impaired by loss of function the progenitor state is exited, and cells undergo differentiation and take the neural fate. The latter implies that Sox2 is not sufficient to maintain the progenitor state without the T/WNT/FGF autoregulatory loop.

During lineage choice, however, Sox2 and T oppose each other. In this process, high Wnt3a/Fgf8/Fgf4/Rspo3 levels in the cellular environment activating the WNT/FGF signaling cascades and T promote the mesodermal fate choice of NMP descendants. WNT activity is highest close to the node, node/streak border, and anterior PS (Ukita et al., 2009; Wymeersch et al., 2016).

The lineage choice apparently is not taking place in the NMPs themselves but in their immediate descendants (Figures 1E, 1F, and 6). We infer this because it is hardly conceivable that self-renewal and differentiation can happen concurrently in the same cell. A switch between the two states is required. Experimental evidence for a distinct nature of NMPs and of cells commencing differentiation comes from our single-cell transcriptome data on *T<sup>mCh+</sup>/Sox2<sup>ve+</sup>* cells. We identified a separate group of cells expressing NMP control genes without *Tbx6* or other lineage-specific regulators, which therefore most likely represent NMPs. The lineage trajectory analysis identified cells of an intermediate state between NMPs and prospective mesodermal or neural cells, which express NMP genes together with *Tbx6* and neural lineage genes. The latter cells most likely represent NMP descendants undergoing lineage choice. The single-cell data therefore confirm that *T<sup>mCh+</sup>/Sox2<sup>ve+</sup>* cells comprise self-renewing NMPs and distinct cells undergoing lineage choice followed by differentiation.

Our data show that Rspo3 is expressed together with T, Sox2, Wnt3a, and Fgf8 in NMPs and in cells undergoing lineage choice, and thus suggest that Rspo3, an activator of WNT signaling, might be an important NMP regulator. Since Rspo3, just as Wnt3a and Fgf4/Fgf8, is secreted and thus can act in an autocrine as well as a paracrine fashion, it might be able to promote self-renewal of NMPs as well as promote the mesodermal fate in NMP descendants undergoing lineage choice (Chal et al., 2015).

In previous reports, *Tbx6* (in mouse) and its functional equivalent in zebrafish, *Tbx16*, have been suggested to play a major role in mesodermal fate choice and neural suppression (Bouldin et al., 2015; Takemoto et al., 2011). *Tbx6* has been shown to be involved in the repression of Sox2 in the mesodermal lineage via (indirect) inactivation of the Sox2 N1 enhancer (Takemoto et al., 2011). However, an earlier report had shown that *Tbx6* mutants initiate mesoderm formation but fail to maintain the paraxial mesoderm fate (Chapman and Papaioannou, 1998). Thus the combined data suggested the action of another factor involved in mesoderm formation. *Tbx16* has been shown to lock cells into the mesodermal state and to repress progenitor genes including Sox2 (Bouldin et al., 2015). This conclusion was based

on data derived by ectopic overexpression of *Tbx16* in the progenitors. This ectopic activity might have forced NMPs to prematurely exit the progenitor state and take the mesodermal fate. However, *Tbx16* was not sufficient for initiating the epithelial-to-mesenchymal transition required for mesoderm formation, suggesting an additional factor(s) involved in the exit from the progenitor state.

Thus both reports demand an additional (upstream?) factor besides *Tbx6* to be involved in mesoderm formation. We argue that this factor is T, but also come to different conclusions with respect to the role of *Tbx6*. We show that *Tbx6* is a direct target of T, and both T and *Tbx6* activate mesoderm control genes. We also provide strong arguments that T is the major antagonist of *Sox2* during lineage choice and is a repressor of the neural fate. This is based on the fact that T represses most key neural transcription factors and that T and *Sox2* co-occupy sites at a large fraction of T-repressed genes. In contrast, *Tbx6* represses only *Sox2* and *Sp8* among many neural transcription factors, and several neural lineage genes of various functions. In addition, *Tbx6* represses progenitor genes including *T* and *Sox2*. Moreover, whereas all trunk mesoderm formation is arrested in *T* mutants, *Tbx6* mutants do produce paraxial tissue and lateral mesoderm (Chapman and Papaioannou, 1998).

The combined data strongly suggest that *Tbx6* is upregulated by T during lineage choice and supports T in sustaining the mesodermal fate choice of NMP descendants, but is not essential for the latter. Our single-cell transcriptome data showing *Tbx6* expression in cells undergoing lineage choice, but not in NMPs, support this conclusion.

Interestingly, the single-cell analysis showed that *Msgn1* expression is first detected in prospective mesodermal cells, whereas *Tbx6* is already expressed in NMP descendants undergoing lineage choice. These data confirm previous reports demonstrating that *Msgn1* is a T target and is synergistically controlled by WNT signaling and *Tbx6* (Chalamalasetty et al., 2011; Wittler et al., 2007). Thus, *Msgn1* clearly acts downstream of *Wnt3a*, T, and *Tbx6*. *Msgn1* knockout embryos show loss of paraxial mesoderm in the trunk, while lateral mesoderm appears unaffected (Yoon and Wold, 2000). The combined genetic and single-cell transcriptome data therefore suggest that *Msgn1* is required to assign paraxial mesoderm identity to prospective mesoderm cells, which are exposed to sufficiently high WNT activity.

T is only transiently active in nascent mesoderm and, at least in the paraxial lineage, not sufficient to maintain mesoderm differentiation. The latter requires *Tbx6*, since in the absence of *Tbx6* paraxial tissue eventually switches to the neural fate (Chapman and Papaioannou, 1998). This suggests that T and WNT/FGF signaling act as lineage selectors rather than as differentiation factors for the mesodermal lineage. The lineage decision, however, needs to be locked in, which is termed commitment, and at least in the paraxial mesoderm this requires *Tbx6* (Figure 6). However, lateral mesoderm apparently is not affected in the *Tbx6* mutant (Chapman and Papaioannou, 1998), and therefore may require other commitment factors. To our knowledge these have not been reported.

In summary, the combined data suggest that T together with WNT and FGF signaling maintain the NMPs via a positive feedback mechanism, repress the neural fate choice by antagonizing

*Sox2*, and promote the mesodermal fate (Figure 6B). The upregulation of *Tbx6* by these factors reinforces the latter, and together they control *Msgn1* expression, which is required to confer paraxial mesoderm identity. Paraxial mesoderm again needs *Tbx6* for sublineage commitment.

While this article was under review, a similar work addressing the gene regulatory network controlling neural and mesodermal lineage specification from NMPs was published (Gouti et al., 2017). In contrast to our study, this work is mainly based on data derived from NMPs differentiated *in vitro*, in combination with single-cell transcriptome data. In this work the authors establish a core NMP signature comprising T and *Sox2*, as well as *Cdx2*, *Nkx1-2*, and *Sp8*, and suggest an important role for retinoic acid (RA) signaling in the induction and differentiation of NMPs. In their model, opposing activities of WNT and RA signaling, which control mutually inhibiting T and *Sox2*, respectively, generate a switch resolving either into mesoderm or neural differentiation, depending on the relative level of each inducer. They conclude that mesoderm precursor cells produce RA, which promotes *Sox2* expression and differentiation of NMPs located adjacent to the NSB to pre-neural tube and NPC identity. To our knowledge, currently there is no evidence that RA signaling plays a role in NMP induction, maintenance, or lineage choice *in vivo*. However, RA is a well-known inducer of neural differentiation, and it will be interesting to dissect the exact role of RA in spinal cord formation during trunk development in the embryo.

## STAR METHODS

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

- **KEY RESOURCES TABLE**
- **CONTACT FOR REAGENT AND RESOURCE SHARING**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
- **METHOD DETAILS**
  - ESC Derivation
  - Embryonic Stem Cell Cultivation
  - Reporter ESC Generation
  - Tetraploid Complementation Assay and Embryo Dissection
  - Microscopy and Clearing
  - *In Vitro* Differentiation
  - Western Blot
  - Cell Sorting
  - RNA-Seq
  - Single-Cell RNA-Seq
  - Antibody Generation Against TBX6
  - ChIP-Seq Library Preparation and Sequencing
  - ATAC-Seq
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Genome Assembly
  - RNA-Seq
  - Single Cell RNA-Seq
  - ChIP-Seq
  - ATAC-Seq
  - Venn Diagrams
- **DATA AVAILABILITY**

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes six figures and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.devcel.2017.07.021>.

**AUTHOR CONTRIBUTIONS**

B.G.H. conceived and supervised the project; F.K. and B.G.H. designed experiments and evaluated data; F.K., M.S., L.W., D.S., P.G., B.T., and K.M. performed experiments; F.K. performed bioinformatics analyses; S.S. carried out pre-somitic mesoderm differentiation *in vitro*; B.G.H. and F.K. wrote the manuscript.

**ACKNOWLEDGMENTS**

We thank Virginia E. Papaioannou for providing the *Tbx6* mutant. We thank members of the Department of Developmental Genetics, in particular Anna D. Senft for generating the *Tbx6::mCherry* reporter BAC, ESC lines, and microscopy of derived embryos; Anna Anurin for support in NMP differentiation and ChIP-seq experiments; Lisette Lange, Sandra Piehl, and Gaby Bläss for support in cell culture and embryo preparation; Judith Fiedler from the Transgenic Unit for generating chimeric embryos; Ludger Hartmann, Mirjam Peetz, Christin Franke, Dijana Micic, and Heike Schlenger from the animal facility for support in generating chimeric embryos and for animal husbandry; Uta Marchfelder for FACS support; Norbert Mages for technical support with single-cell sequencing; and all members of the Sequencing Core Facility. This work was supported by an EMBO Long-Term Fellowship to F.K. (co-funded by the European Commission EMBOCOFUND2010, GA-2010-267146) and funded by the Max Planck Society.

Received: January 26, 2017

Revised: June 7, 2017

Accepted: July 24, 2017

Published: August 17, 2017

**REFERENCES**

- Arnold, S.J., Stappert, J., Bauer, A., Kispert, A., Herrmann, B.G., and Kemler, R. (2000). Brachury is a target gene of the Wnt/beta-catenin signaling pathway. *Mech. Dev.* 91, 249–258.
- Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.
- Benoukraf, T., Cauchy, P., Fenouil, R., Jeanniard, A., Koch, F., Jaeger, S., Thieffry, D., Imbert, J., Andrau, J.C., Spicuglia, S., et al. (2009). CoCAS: a ChIP-on-chip analysis suite. *Bioinformatics* 25, 954–955.
- Bergsland, M., Ramskold, D., Zaouter, C., Klum, S., Sandberg, R., and Muhr, J. (2011). Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev.* 25, 2453–2464.
- Bouldin, C.M., Manning, A.J., Peng, Y.H., Farr, G.H., 3rd, Hung, K.L., Dong, A., and Kimelman, D. (2015). Wnt signaling and *tbx16* form a bistable switch to commit bipotential progenitors to mesoderm. *Development* 142, 2499–2507.
- Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122, 947–956.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218.
- Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523, 486–490.
- Chal, J., Oginuma, M., Al Tanoury, Z., Gobert, B., Sumara, O., Hick, A., Bousson, F., Zidouni, Y., Mursch, C., Moncuquet, P., et al. (2015). Differentiation of pluripotent stem cells to muscle fiber to model Duchenne muscular dystrophy. *Nat. Biotechnol.* 33, 962–969.
- Chalamalasetty, R.B., Dunty, W.C., Jr., Biris, K.K., Ajima, R., Iacovino, M., Beisaw, A., Feigenbaum, L., Chapman, D.L., Yoon, J.K., Kyba, M., et al. (2011). The Wnt3a/beta-catenin target gene Mesogenin1 controls the segmentation clock by activating a Notch signalling program. *Nat. Commun.* 2, 390.
- Chalamalasetty, R.B., Garriock, R.J., Dunty, W.C., Jr., Kennedy, M.W., Jailwala, P., Si, H., and Yamaguchi, T.P. (2014). Mesogenin 1 is a master regulator of paraxial presomitic mesoderm differentiation. *Development* 141, 4285–4297.
- Chapman, D.L., and Papaioannou, V.E. (1998). Three neural tubes in mouse embryos with mutations in the T-box gene *Tbx6*. *Nature* 391, 695–697.
- Chesley, P., and Dunn, L.C. (1936). The inheritance of taillessness (anury) in the house mouse. *Genetics* 21, 525–536.
- Ciruna, B., and Rossant, J. (2001). FGF signalling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. *Dev. Cell* 1, 37–49.
- de Hoon, M.J., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. *Bioinformatics* 20, 1453–1454.
- de Lau, W., Barker, N., Low, T.Y., Koo, B.K., Li, V.S., Teunissen, H., Kujala, P., Haegebarth, A., Peters, P.J., van de Wetering, M., et al. (2011). Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. *Nature* 476, 293–297.
- Eakin, G.S., and Hadjantonakis, A.K. (2006). Production of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. *Nat. Protoc.* 1, 1145–1153.
- Freese, N.H., Norris, D.C., and Loraine, A.E. (2016). Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* 32, 2089–2095.
- Garriock, R.J., Chalamalasetty, R.B., Kennedy, M.W., Canizales, L.C., Lewandoski, M., and Yamaguchi, T.P. (2015). Lineage tracing of neuromesodermal progenitors reveals novel Wnt-dependent roles in trunk progenitor cell maintenance and differentiation. *Development* 142, 1628–1638.
- Gentsch, G.E., Owens, N.D., Martin, S.R., Piccinelli, P., Faial, T., Trotter, M.W., Gilchrist, M.J., and Smith, J.C. (2013). In vivo T-box transcription factor profiling reveals joint regulation of embryonic neuromesodermal bipotency. *Cell Rep.* 4, 1185–1196.
- Gouti, M., Tsakiridis, A., Wymeersch, F.J., Huang, Y., Kleinjung, J., Wilson, V., and Briscoe, J. (2014). In vitro generation of neuromesodermal progenitors reveals distinct roles for wnt signalling in the specification of spinal cord and paraxial mesoderm identity. *PLoS Biol.* 12, e1001937.
- Gouti, M., Delile, J., Stamatakis, D., Wymeersch, F.J., Huang, Y., Kleinjung, J., Wilson, V., and Briscoe, J. (2017). A gene regulatory network balances neural and mesoderm specification during vertebrate trunk development. *Dev. Cell* 41, 243–261.e7.
- Gu, Z., Eils, R., and Schlesner, M. (2016). HilbertCurve: an R/Bioconductor package for high-resolution visualization of genomic data. *Bioinformatics* 32, 2372–2374.
- Henrique, D., Abrantes, E., Verrier, L., and Storey, K.G. (2015). Neuromesodermal progenitors and the making of the spinal cord. *Development* 142, 2864–2875.
- Herrmann, B.G. (1991). Expression pattern of the Brachury gene in whole-mount TWis/TWis mutant embryos. *Development* 113, 913–917.
- Herrmann, B.G., Labeit, S., Poustka, A., King, T.R., and Lehrach, H. (1990). Cloning of the T gene required in mesoderm formation in the mouse. *Nature* 343, 617–622.
- Hulsen, T., de Vlieg, J., and Alkema, W. (2008). BioVenn—a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* 9, 488.
- Jurberg, A.D., Aires, R., Novoa, A., Rowland, J.E., and Mallo, M. (2014). Compartment-dependent activities of Wnt3a/beta-catenin signalling during vertebrate axial extension. *Dev. Biol.* 394, 253–263.

- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36.
- Kispert, A., and Herrmann, B.G. (1993). The Brachyury gene encodes a novel DNA binding protein. *EMBO J.* 12, 3211–3220.
- Kispert, A., Koschorz, B., and Herrmann, B.G. (1995). The T protein encoded by Brachyury is a tissue-specific transcription factor. *EMBO J.* 14, 4763–4772.
- Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C., et al. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.* 18, 956–963.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Luna-Zurita, L., Stirnimann, C.U., Glatt, S., Kaynak, B.L., Thomas, S., Baudin, F., Samee, M.A., He, D., Small, E.M., Mileikovsky, M., et al. (2016). Complex interdependence regulates heterotypic transcription factor distribution and coordinates cardiogenesis. *Cell* 164, 999–1014.
- Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* 27, 1696–1697.
- Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134, 521–533.
- Martin, B.L., and Kimelman, D. (2010). Brachyury establishes the embryonic mesodermal progenitor niche. *Genes Dev.* 24, 2778–2783.
- Martin, B.L., and Kimelman, D. (2012). Canonical Wnt signaling dynamically controls multiple stem cell fate decisions during vertebrate body formation. *Dev. Cell* 22, 223–232.
- Muyrers, J.P., Zhang, Y., Testa, G., and Stewart, A.F. (1999). Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.* 27, 1555–1557.
- Naiche, L.A., Holder, N., and Lewandoski, M. (2011). FGF4 and FGF8 comprise the wavefront activity that controls somitogenesis. *Proc. Natl. Acad. Sci. USA* 108, 4018–4023.
- Pease, S., Saunders, T.L., and International society for transgenic technologies. (2011). Advanced Protocols for Animal Transgenesis: An ISTT Manual (Springer).
- Pettitt, S.J., Liang, Q., Raidan, X.Y., Moran, J.L., Prosser, H.M., Beier, D.R., Lloyd, K.C., Bradley, A., and Skarnes, W.C. (2009). Agouti C57BL/6N embryonic stem cells for mouse genetic resources. *Nat. Methods* 6, 493–495.
- Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.A., and Trapnell, C. (2017). Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* 14, 309–315.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Rodrigo Albors, A., and Storey, K.G. (2016). Mapping body-building potential. *Elife* 5, e14830.
- Saldanha, A.J. (2004). Java Treeview—extensible visualization of microarray data. *Bioinformatics* 20, 3246–3248.
- Shen, T., Aneas, I., Sakabe, N., Dirschinger, R.J., Wang, G., Smemo, S., Westlund, J.M., Cheng, H., Dalton, N., Gu, Y., et al. (2011). Tbx20 regulates a genetic program essential to adult mouse cardiomyocyte function. *J. Clin. Invest.* 121, 4640–4654.
- Sudheer, S., Liu, J., Marks, M., Koch, F., Anurin, A., Scholze, M., Senft, A.D., Wittler, L., Macura, K., Grote, P., et al. (2016). Different concentrations of FGF ligands, FGF2 or FGF8 determine distinct states of WNT-induced presomitic mesoderm. *Stem Cells* 34, 1790–1800.
- Takada, S., Stark, K.L., Shea, M.J., Vassileva, G., McMahon, J.A., and McMahon, A.P. (1994). Wnt-3a regulates somite and tailbud formation in the mouse embryo. *Genes Dev.* 8, 174–189.
- Takemoto, T., Uchikawa, M., Yoshida, M., Bell, D.M., Lovell-Badge, R., Papaioannou, V.E., and Kondoh, H. (2011). Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. *Nature* 470, 394–398.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515.
- Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2012). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31, 46–53.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381–386.
- Tsakiridis, A., Huang, Y., Blin, G., Skylaki, S., Wyneersch, F., Osorno, R., Economou, C., Karagianni, E., Zhao, S., Lowell, S., et al. (2014). Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. *Development* 141, 1209–1221.
- Tzouanacou, E., Wegener, A., Wyneersch, F.J., Wilson, V., and Nicolas, J.F. (2009). Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. *Dev. Cell* 17, 365–376.
- Ukita, K., Hirahara, S., Oshima, N., Imuta, Y., Yoshimoto, A., Jang, C.W., Oginuma, M., Saga, Y., Behringer, R.R., Kondoh, H., et al. (2009). Wnt signaling maintains the notochord fate for progenitor cells and supports the posterior extension of the notochord. *Mech. Dev.* 126, 791–803.
- Wilkinson, D.G., Bhatt, S., and Herrmann, B.G. (1990). Expression pattern of the mouse T gene and its role in mesoderm formation. *Nature* 343, 657–659.
- Wilson, V., Olivera-Martinez, I., and Storey, K.G. (2009). Stem cells, signals and vertebrate body axis extension. *Development* 136, 1591–1604.
- Wittler, L., Shin, E.H., Grote, P., Kispert, A., Beckers, A., Gossler, A., Werber, M., and Herrmann, B.G. (2007). Expression of Msgn1 in the presomitic mesoderm is controlled by synergism of WNT signalling and Tbx6. *EMBO Rep.* 8, 784–789.
- Wyneersch, F.J., Huang, Y., Blin, G., Cambray, N., Wilkie, R., Wong, F.C., and Wilson, V. (2016). Position-dependent plasticity of distinct progenitor types in the primitive streak. *Elife* 5, e10042.
- Yamaguchi, T.P., Takada, S., Yoshikawa, Y., Wu, N., and McMahon, A.P. (1999). T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. *Genes Dev.* 13, 3185–3190.
- Yanagisawa, K.O., Fujimoto, H., and Urushihara, H. (1981). Effects of the brachyury (T) mutation on morphogenetic movement in the mouse embryo. *Dev. Biol.* 87, 242–248.
- Ye, T., Krebs, A.R., Choukallah, M.A., Keime, C., Plewniak, F., Davidson, I., and Tora, L. (2011). seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res.* 39, e35.
- Yoon, J.K., and Wold, B. (2000). The bHLH regulator pMesogenin1 is required for maturation and segmentation of paraxial mesoderm. *Genes Dev.* 14, 3204–3214.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Goat polyclonal anti-Brachyury (N-19)	Santa Cruz Biotechnology	sc-17743x; RRID: AB_634980
Goat polyclonal anti-Sox2	R&D Systems	AF2018; RRID: AB_355110
Rabbit polyclonal anti-Tbx6	This paper	N/A
Rabbit polyclonal anti- $\beta$ -catenin	Invitrogen	71-2700; RRID: AB_2533982
Rabbit polyclonal anti-Brachyury	Kispert et al., 1995	N/A
Rabbit polyclonal anti-H3	Abcam	ab1791; RRID: AB_302613
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
LIF	Chemicon	ESG1107
bFGF	Peprotech	100-18B
CHIR99021	Merck	361571
PD0325901	Stemgent	04-0006-02
CHIR99021	Stemgent	04-0004-02
<b>Critical Commercial Assays</b>		
TotalScript RNA-Seq Kit	Epicentre	TSRNA12924
ScriptSeq Complete (Human/Mouse/Rat) low input RNA-Seq Kit	Epicentre	SCL24H
TruSeq ChIP Library Preparation Kit	Illumina	IP-202-1012
Nextera DNA Library Preparation Kit	Illumina	FC-121-1030
C1 Reagent Kit for mRNA Seq	Fluidigm	100-6201
SMARTer Ultra Low RNA Kit for the Fluidigm C1 System	Clontech	634833
<b>Deposited Data</b>		
Raw and analyzed data	This paper	GEO: GSE93524
Mouse reference genome GRCm38/mm10	Genome Reference Consortium	<a href="http://hgdownload.soe.ucsc.edu/goldenPath/mm10/">http://hgdownload.soe.ucsc.edu/goldenPath/mm10/</a>
Mouse mm10 RefSeq annotation	NCBI RefSeq project	<a href="http://hgdownload.soe.ucsc.edu/goldenPath/mm10/">http://hgdownload.soe.ucsc.edu/goldenPath/mm10/</a>
<b>Experimental Models: Cell Lines</b>		
Mouse JM8A1.N3 (C57BL/6N) ESCs	Pettitt et al., 2009	N/A
Mouse C57BL/6J ESCs	This paper	N/A
Mouse C57BL/6J T <sup>2j</sup> /T <sup>2j</sup> ESCs	This paper	N/A
Mouse C57BL/6J Tbx6 <sup>-/-</sup> ESCs	This paper	N/A
<b>Software and Algorithms</b>		
Bowtie	Langmead et al., 2009	<a href="http://bowtie-bio.sourceforge.net/index.shtml">http://bowtie-bio.sourceforge.net/index.shtml</a>
Tophat2	Kim et al., 2013	<a href="https://ccb.jhu.edu/software/tophat/index.shtml">https://ccb.jhu.edu/software/tophat/index.shtml</a>
Cufflinks	Trapnell et al., 2010; Trapnell et al., 2012	<a href="http://cole-trapnell-lab.github.io/cufflinks/">http://cole-trapnell-lab.github.io/cufflinks/</a>
Samtools	Li et al., 2009	<a href="http://www.htslib.org/">http://www.htslib.org/</a>
BEDTools	Quinlan and Hall, 2010	<a href="http://bedtools.readthedocs.io/">http://bedtools.readthedocs.io/</a>
Picard	N/A	<a href="https://broadinstitute.github.io/picard/">https://broadinstitute.github.io/picard/</a>
Ea-utils	N/A	<a href="https://expressionanalysis.github.io/ea-utils/">https://expressionanalysis.github.io/ea-utils/</a>
Meme	Bailey and Elkan, 1994	<a href="http://meme-suite.org/">http://meme-suite.org/</a>
Meme-ChIP	Machanick and Bailey, 2011	<a href="http://meme-suite.org/">http://meme-suite.org/</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
R: A Language and Environment for Statistical Computing	R Core Team	<a href="https://www.R-project.org">https://www.R-project.org</a>
SeqMiner	Ye et al., 2011	<a href="https://sourceforge.net/projects/seqminer/">https://sourceforge.net/projects/seqminer/</a>
Cluster 3.0	de Hoon et al., 2004	<a href="http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm">http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm</a>
Java TreeView	Saldanha, 2004	<a href="https://sourceforge.net/projects/jtreeview">https://sourceforge.net/projects/jtreeview</a>
Monocle	(Qiu et al., 2017; Trapnell et al., 2014)	<a href="https://bioconductor.org/packages/release/bioc/html/monocle.html">https://bioconductor.org/packages/release/bioc/html/monocle.html</a>

**CONTACT FOR REAGENT AND RESOURCE SHARING**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Bernhard G. Herrmann ([herrmann@molgen.mpg.de](mailto:herrmann@molgen.mpg.de)).

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

All animal procedures were conducted as approved by the local authorities (LAGeSo Berlin) under the license numbers G0368/08 and G0247/13. All mouse ESC lines used in this study were male and either from a JM8A1.N3 (C57BL/6N) or C57BL/6J genetic background and were cultured according to standard conditions.

**METHOD DETAILS****ESC Derivation**

The *T<sup>2J/2J</sup>*, *Tbx6<sup>-/-</sup>* and C57BL/6J WT ESC lines were established from C57BL/6J homozygous blastocysts using N2B27 Medium supplemented with FGF/Erk and Gsk3 pathway inhibitors (2i) and LIF according to Nagy and Nichols ([Pease et al., 2011](#)).

**Embryonic Stem Cell Cultivation**

All mouse ESC lines used in this study were male and either from a JM8A1.N3 ([Pettitt et al., 2009](#)) (C57BL/6N) or C57BL/6J genetic background. Cells were maintained on gelatinized plates and mitotically inactive primary mouse embryo fibroblasts with standard ESC medium containing 15% FCS and 1000 U/ml leukemia inhibitory factor (LIF, Chemicon ESG1107).

**Reporter ESC Generation**

To generate the fluorescent reporter constructs, we genome-engineered the mouse BACs *RP24-530D23* (*T*), *RP23-249O15* (*Sox2*) and *RP23-421P23* (*Tbx6*), containing ~180 - 230kb of the C57BL/6 mouse genome surrounding the respective loci, via Red/ET recombination ([Muyrers et al., 1999](#)). Briefly, we replaced the starting codon (ATG) of the genes with a reporter gene containing mCherry (H2B-mCherry in the case of *T<sup>2J/2J</sup>* *T*::mCherry / *Sox2*::Venus ESCs) or H2B-Venus, followed by the rabbit *b-globin* polyadenylation signal and an *FRT*-flanked selection cassette (hygromycin or neomycin), driven by the *Pgk* promotor. For random integration, 5µg of the modified BACs were linearized with PI-SceI (NEB) and electroporated into 3x10<sup>6</sup> ESCs. Selection was performed using ESC medium containing 250µg/ml neomycin or 150µg/ml hygromycin. After selection, single clones were picked, expanded and checked for BAC integration by PCR.

**Tetraploid Complementation Assay and Embryo Dissection**

All animal procedures were conducted as approved by the local authorities (LAGeSo Berlin) under the license numbers G0368/08 and G0247/13. The modified ESCs were used to generate embryos via the tetraploid complementation assay ([Eakin and Hadjantonakis, 2006](#)). For embryo isolation, mice were sacrificed by cervical dislocation and uteri were dissected in PBS. Isolated embryos were transferred to M2 medium (Sigma) for staging and microdissection.

**Microscopy and Clearing**

Embryos containing the *T*, *Sox2* or *Tbx6* fluorescent reporters were either imaged using a Leica MZ 16FA or Zeiss Axiozoom microscope with appropriate filters for mCherry and Venus. For imaging of cleared *T*::mCherry/*Sox2*::Venus reporter positive embryos on a Zeiss Z.1 lightsheet microscope, embryos were fixed for 30 minutes in 4% PFA in PBS at 4°C under rotation. All samples were kept in the dark as much as possible. After stopping the fixation by washing with PBS, embryos were rinsed 3x 5 minutes at room temperature in 0.1M phosphate buffer (PB, 0.025M NaH<sub>2</sub>PO<sub>4</sub> and 0.075M Na<sub>2</sub>HPO<sub>4</sub>, pH 7.4) containing 0.01% Tween-20. Clearing was performed by incubation in RIMS (133% w/v Histodenz (Sigma-Aldrich) in 0.02M PB containing 0.01% Tween-20) under rotation at 4°C

for at least one to several days. Embryos were embedded in low-melting point agarose (1.5% in PBS) in a glass capillary of 1.5 mm diameter. The sample chamber of the Z.1 (Zeiss) was filled with RIMS and embryos were incubated in the chamber overnight. Imaging was performed using a 20x clearing objective. The image was processed using the Vision 4D software (Arivis).

### In Vitro Differentiation

For NMP differentiation, we used a previously described protocol ([Gouti et al., 2014](#)) with slight modifications. ESCs were removed from feeders by dissociation using 0.05% trypsin and then plated onto tissue culture plates for four short successive periods (25', 20', 15', 10') to remove feeders. We then cultivated the cells for at least two passages in 2i + Lif medium (1:1 DMEM/F12:Neurobasal medium, 1x Pen/Strep, 1x Glutamax, 1x non-essential amino-acids, 1x Na-Pyruvate, 0.11mM  $\beta$ -mercaptoethanol, 1x N2, 1x B-27 without vitamin A, 1 $\mu$ M PD0325901 (Stemgent), 3 $\mu$ M CHIR99021 (Stemgent), 1x Lif). Cells were then counted and plated on CellBINDSurface dishes (Corning) precoated with Synthemax II-SC (Corning) at a density of 1x10<sup>4</sup> cells per cm<sup>2</sup> in a medium consisting of 'N2B27' medium ([Gouti et al., 2014](#)), supplemented with 10ng/ml bFGF (Peprotech) for 3 days (d1-d3). From d2-d3 5 $\mu$ M CHIR99021 (CHI, Calbiochem) was added to the medium. Presomitic mesoderm differentiation was carried out as described ([Sudh-eer et al., 2016](#)) using CHIR99021 and high levels of FGF2 (250ng/ml for 2 days), following the conversion to EpiSCs.

### Western Blot

For Western blot analysis, cells were differentiated as described, harvested at the indicated time points and counted. After centrifugations, 1x NuPAGE LDS Sample Buffer (Life Technologies) was added to a final concentration of 2000 cells/ $\mu$ l. Samples were sonicated for 20 cycles (30s on/30s off) in a Bioruptor Pico (Diagenode) to fragment the DNA. For each sample, 5 $\mu$ l were mixed with 5 $\mu$ l of 1x NuPage LDS Sample Buffer containing 25 $\mu$ l/ml  $\beta$ -mercaptoethanol and boiled for 5 minutes at 95°C. The resulting 10 $\mu$ l were loaded onto a 12-well NuPAGE 4-12% Bis-Tris gradient gel (Life Technologies) according to the manufacturer's protocol. The wet transfer was performed using transfer buffer (25mM Tris, 192mM glycine, 400ml MeOH) for 1 hour at 4°C and constant 100V onto a PVDF membrane (Bio-Rad). The membrane was subsequently blocked for 1h in 5% milk (Sigma Aldrich) in TBS-T (10mM Tris pH8, 150mM NaCl, 0.05% Tween-20) at room temperature. The blot was cut into appropriate pieces, washed quickly twice with TBS-T and incubated with the appropriate antibodies in 2% milk overnight at 4°C under agitation. The antibodies used were rabbit anti-T (1:2000; [Kispert et al., 1995](#)), goat anti-Sox2 (1:2000; AF2018, R&D), rabbit anti-Tbx6 (1:500; see below) and rabbit anti-H3 (1:2000; ab1791, Abcam). The blots were washed 3x10 minutes with TBS-T and incubated with the appropriate antibodies in 2% milk for 1 hour at room temperature. The secondary antibodies used were a goat anti-rabbit-HRP (1:5000; #7074, Cell Signaling Technology) and a donkey anti-goat-HRP (1:5000, 705-035-003, Jackson ImmunoResearch Laboratories). The blots were washed 3x10 minutes with TBS-T. Detection was performed using the Amersham ECL reagents (GE Healthcare) and scanned using a Fusion SL chemiluminescence system (Vilber). The detection of T and Sox2 was performed on the same membrane, by first probing for T and subsequent stripping using Restore PLUS Western Blot Stripping buffer (Thermo Scientific) for 45 minutes at room temperature. The Tbx6 blot was performed using a separate gel and membrane.

### Cell Sorting

For Tbx6::mCherry embryos, the PSM up to the -2 somite was dissected from TS14 embryos. For the T::mCherry and T::mCherry/Sox2::Venus embryos, we used the caudal end containing the node-streak border and surrounding caudal lateral epiblast. In order to obtain single cell suspensions from embryo tissue for FACS sorting, 10 $\mu$ l of 5% BSA in PBS was added to the pooled samples and 100 $\mu$ l of ice-cold 0.05% trypsin was added. A homogenous suspension was achieved by slow pipetting and the reaction was stopped with the addition of 200 $\mu$ l of 5% BSA in PBS. Before cell sorting on a FACS Aria II (Becton Dickinson), the cell suspension was filtered through a 35 $\mu$ m mesh. For downstream RNA-seq experiments cells were sorted directly into Eppendorf tubes containing 350 $\mu$ l RLT buffer (Qiagen) and 3.5 $\mu$ l b-mercaptoethanol. For ATAC-seq, cells were sorted into FACS-tubes containing 5% BSA in PBS, which had previously been rotating for several hours to pre-coat the tube walls with BSA. Sorted cells were centrifuged at 1000rpm and 4°C for 10 minutes, transferred to Eppendorf tubes containing PBS with 0.05% Triton X-100 and centrifuged at 1000rpm and 4°C for 10 minutes. The supernatant was removed and cells were directly processed for ATAC-seq. For single-cell RNA-seq, T<sup>mCh+</sup>/Sox2<sup>Ve+</sup> cells were sorted from caudal ends of TS13 embryos and collected in Eppendorf tubes containing filtered 5% BSA in PBS at 4°C.

### RNA-Seq

For the T/Sox2 subpopulations, total RNA from approximately 2,500 cells was isolated using the RNeasy Micro kit (Qiagen) and genomic DNA was digested on column with the addition of an extra 1 $\mu$ l of RNase-free DNase I (Roche) to ensure complete digestion. The RNA was eluted from the columns using RNase-free water. RNA-seq libraries with the RNA equivalent of approximately 1,000 cells were generated using the TotalScript RNA-seq Kit (Epicentre) with random hexamers according to manufacturer's instructions with slight modifications. Briefly, all reagents during the cDNA synthesis steps were scaled down in order to achieve a total volume of 39 $\mu$ l, allowing all of the material to be used in the subsequent tagmentation procedure. To further improve strand-specificity of the libraries, Actinomycin D was added to the first strand cDNA synthesis reaction at a 5ng/ $\mu$ l final concentration and the libraries were finally amplified using 14 cycles.

For T and Tbx6 wild-type vs. knock-out comparisons, total RNA was isolated from approximately 20,000 cells using the RNeasy Micro kit (Qiagen) as described above. The RNA was quantified using the Qubit RNA HS Assay and the integrity was verified using

Bioanalyzer RNA Pico chips. Approximately 80-100ng of total RNA was used for the generation of strand-specific RNA-seq libraries using the ScriptSeq v2 (Epicentre) low input library preparation kit according to manufacturer's instructions. During library preparation, the purification after rRNA-depletion was performed using RNeasy Micro columns and the cDNA was purified using the MinElute PCR Purification Kit (Qiagen). The library was finally amplified with 15 PCR cycles.

Quantification of all RNA-seq libraries was performed using the Qubit high sensitivity DNA assay (Life Technologies) and the size distribution was verified using the DNA HS Bioanalyzer chips (Agilent). Libraries were pooled and paired-end sequenced on a HiSeq 2500 with 2x50bp read lengths.

### Single-Cell RNA-Seq

Sorted cells were pelleted by centrifugation and subsequently resuspended in 10ul PBS. The density and viability (>90%) were checked on a Luna cell counter (Logos Biosystems). The suspension was adjusted to a cell density of 250 cells/ $\mu$ l, resuspension buffer added (4.5:1 ratio of cells:buffer) and approximately 5 $\mu$ l were loaded onto a 5-10 $\mu$ m C1 IFC (Fluidigm). The IFC was processed according to the Fluidigm mRNA-seq protocol. A 1:4000 dilution of ERCC spike-ins (Ambion) was added to the lysis buffer. Upon visual inspection, 55 capture sites were called to contain cells. The amplified cDNA was measured using the Qubit DNA HS assay and a selection of samples was verified on a Bioanalyzer DNA HS chip (Agilent). The final libraries for those 55 capture sites were prepared with the Nextera XT (Illumina) kit according to the protocol, pooled and sequenced on a NextSeq (Illumina) sequencer using 2x75bp read lengths.

### Antibody Generation Against TBX6

A cDNA fragment of *Tbx6*, lacking the coding region for amino acid #103 to 347, was cloned in frame downstream of a maltose binding protein (MBP) sequence, controlled by an Arabinose inducible promoter in an *E. coli* expression vector. Recombinant, soluble MBP-TBX6( $\Delta$ 103-347) was produced in *E. coli* DH5 $\alpha$ , purified over an amylose column and two rabbits immunized (Pineda Antikörper-Service, Berlin, Germany). Specificity of the anti-TBX6 antibody was verified on western blot and immunohistochemistry.

### Chromatin Immunoprecipitation (ChIP)

ChIP experiments were carried out essentially as previously described (Koch et al., 2011). Crosslinking was performed directly on differentiating cells in growth medium with the addition of 1/10th volume of crosslinking solution (11% formaldehyde, 100mM NaCl, 1mM EDTA pH8, 0.5mM EGTA pH8, 50mM Hepes pH7.8) for 10 minutes at room temperature, with the dishes placed on a shaker. The crosslinking reaction was quenched with the addition of 1/10th volume of 2.5M glycine and 5 minutes incubation. Dishes were placed on ice or in a 4°C cold room and washed twice with cold PBS. Cells were scraped and washed off using cold PBS, containing 0.05% Triton X-100 to break the surface tension. Cells were pooled in aliquots of ~3-5x10<sup>7</sup> in Eppendorf tubes, snap frozen and stored at -80°C until sonication. 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 LB1 (50mM Hepes pH7.5, 140mM NaCl, 1mM EDTA pH8, 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 with a tight pestle with 3x 15 strokes, with short resting times on ice in between. The chromatin was pelleted by centrifugation at 1400g and 4°C for 5 minutes and resuspended in 2.5ml LB2 (200mM NaCl, 1mM EDTA pH8, 0.5mM EGTA pH8, 10mM Tris pH 8) per cell pellet. After 10 minutes rotation at 4°C, the centrifugation step was repeated and each pellet was resuspended in 1.5ml LB3 (1mM EDTA pH8, 0.5mM EGTA, 10mM Tris pH8, 100mM NaCl, 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine), transferred to 15ml Falcon tubes and sonicated using a W-450D Digital Sonifier (Branson) sonicator for 14 cycles of 10s on/50s off in a 4°C cold-room with tubes chilled in ice water. After sonication, 150 $\mu$ l of triton X-100 was added per tube, transferred to two 1.5ml 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 $\mu$ l as an input control, the chromatin was distributed into 1.5ml aliquots, snap frozen and stored at -80°C until use.

The input was reverse-crosslinked with the addition of 50 $\mu$ l 2x elution buffer (100mM Tris pH8, 20mM EDTA pH8, 2% SDS) and overnight (13h-15h) incubation at 65°C. The next day, 100 $\mu$ l of TE buffer were added to dilute the SDS of the elution buffer and RNA was digested using RNase A at a 0.2 $\mu$ g/ml final concentration at 37°C for 2 hours. Proteins were digested with the addition of 0.2 $\mu$ g/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, pH8) isolations and a subsequent MinElute (Qiagen) purification. Purification was performed according to manufacturer's recommendations, except that the columns were washed twice with PE and air dried for at least 10 minutes before elution with EB. The input DNA was quantified using a NanoPhotometer (Implen) and the shearing was verified by running 600ng of DNA on a 1.5% agarose gel. For sonication to be considered successful and sonicated chromatin to be usable for ChIP, the bulk DNA of the input had to be located between 100bp-500bp.

For ChIP, 50 $\mu$ l or 100 $\mu$ l (for  $\beta$ -catenin) of protein-coated Dynal beads (Life Technologies) were washed 3x using 1ml of blocking solution (PBS containing 0.5% BSA), resuspended in 500 $\mu$ l of blocking solution together with approximately 5 $\mu$ g or 10 $\mu$ g (for  $\beta$ -catenin) antibody and rotated overnight at 4°C. The antibodies used were a goat polyclonal anti-T (Santa Cruz, sc-17743), a goat polyclonal anti-Sox2 (R&D Systems, AF2018), a rabbit polyclonal anti- $\beta$ -catenin (Invitrogen, 71-2700) and a rabbit polyclonal anti-Tbx6 (this paper). The next day, three further washes using 1ml blocking buffer were performed and the beads were resuspended in 100 $\mu$ l blocking buffer. Chromatin equivalent to approximately 25x10<sup>6</sup> or 50x10<sup>6</sup> (for  $\beta$ -catenin) cells was added and rotated overnight at 4°C. The following day, washing steps (6, 8 and 9 washes for Tbx6, Sox2 /  $\beta$ -catenin and T respectively) were performed using

1ml RIPA buffer (50mM Hepes pH7.6, 500mM LiCl, 1mM EDTA pH8, 1% NP-40, 0.7% Na-Deoxycholate) containing 1x complete protease inhibitors without EDTA and finally with 1ml of TEN (10mM Tris pH8, 1mM EDTA pH8, 50mM NaCl). The immunoprecipitates were eluted in two subsequent steps using 100 $\mu$ l of 1x elution buffer (50mM Tris pH8, 10mM EDTA pH8, 1% SDS) and incubation at 65°C while shaking for 10 minutes each. The eluates were combined and incubated overnight (13h-15h) at 65°C under agitation. The next day, 200 $\mu$ l of TE were added to dilute the SDS of the elution buffer and the ChIPs were purified as described for the input above. ChIP DNA was quantified using the Qubit (Life Technologies) DNA HS assay.

### ChIP-Seq Library Preparation and Sequencing

The ChIP-seq libraries were generated using the TruSeq ChIP library kit (Illumina) with the following modifications. After adapter ligation, 0.95x of AMPure XP beads (Beckman Coulter) were used for a single purification and the DNA was eluted using 15 $\mu$ l of resuspension buffer (RSB, Illumina). In the case of Tbx6, 14 $\mu$ l of the eluate were used for PCR pre-amplification in order to convert adapters into double-stranded DNA. After the addition of 1 $\mu$ l primer mix (25mM each, Primer 1: 5'- AATGATAACGGCGACCACCGA\*G-3'; Primer2: 5'- CAAGCAGAACGGCATACGA\*G-3') and 15 $\mu$ l 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 $\mu$ l of RSB. After transfer of 20 $\mu$ l of eluate, 6 $\mu$ l of 5x loading buffer was added and the ligation products were separated using a 1.5% agarose gel. Post-run staining was performed using Sybr Gold (Life Technologies) under agitation for 30 minutes. Gel slices according to ~250-400bp 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, with each flow through added to the column a second time. 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 $\mu$ l of EB buffer. Of the eluate, 19 $\mu$ l was used in the final amplification, with the addition of 1 $\mu$ l primer mix and 20 $\mu$ l 2x Kapa HiFi Hotstart premix. The same protocol as for the pre-amplification was used, with the exception of using 13 amplification cycles. In the case of NMP ChIPs and input, the gel purification step was omitted and libraries were directly amplified for 18 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 2000 or NextSeq 500 (Illumina) using 2x50bp or 2x75bp read lengths respectively. In the case of 75bp read lengths, sequences and quality strings were trimmed to 50bp prior to alignment.

### ATAC-Seq

ATAC-seq was performed essentially as previously described (Buenrostro et al., 2013). In total, 1,000-5,000 cells were used per ATAC-seq experiment. After spinning down the FACS sorted cells, the pellets were resuspended in 50 $\mu$ l of lysis buffer (10mM Tris pH7.4, 10mM NaCl, 3mM MgCl<sub>2</sub>, 0.1% NP-40) and immediately spun down at 500g and 4°C for 10 min. The supernatant was discarded and each pellet was resuspended in the transposition reaction mix (25 $\mu$ l 2x TD buffer, 2.5 $\mu$ l Tn5 transposase, 22 $\mu$ l H<sub>2</sub>O) and incubated at 37°C for 30 minutes. After incubation, the reaction was stopped with the addition of PB buffer (Qiagen) and the fragmented DNA was purified using the MinElute kit (Qiagen). The DNA was combined with the ATAC index PCR primers and 2x Kapa HiFi Hotstart Readymix and pre-amplified (98°C 30 seconds, 8x [98°C 10 seconds, 63°C 30 seconds, 72°C 1 minute]) in a 50 $\mu$ l volume. To determine the remaining cycles to avoid potential overamplification, 5 $\mu$ l of the preamplification mix was combined with the primers, 1x Evagreen Sybr green (Jena Biosciences) and 2x Kapa HiFi Hotstart Readymix in a 15 $\mu$ l total volume and run for 30 cycles on a StepOne Plus. The remaining 45 $\mu$ l pre-amplified samples were amplified for a further 9 cycles, resulting in a total of 17 cycles. The libraries were purified using MiElute column (Qiagen) and the concentration was measured using the DNA HS Qubit assay (Life Technologies). Approximately 4ng of each library were run on a DNA HS Bioanalyzer ChIP (Agilent) to verify library size and calculate molarities. Samples were pooled and paired-end sequenced on an Illumina NextSeq 500 at 2x75bp read length.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Genome Assembly

All datasets were mapped to the mouse mm10 genome containing chromosomes 1-19, X, Y and M using bowtie version 1.1.2 (Langmead et al., 2009).

### RNA-Seq

For RNA-seq, reads were mapped with TopHat2 (version 2.1.0) (Kim et al., 2013) using bowtie (Langmead et al., 2009), providing refSeq annotations in gtf format (UCSC) and the options *-no-coverage-search -no-mixed -no-discordant -g1 -library-type* ("fr-first-strand" or "fr-secondstrand" for libraries generated using the TotalScript and Scriptseq kits respectively). Reads that were called to be spliced and containing ambiguous strand information (e.g. containing both XS:A:+ and XS:A:- tags) were removed together with their respective mates.

For visualization, wiggle tracks were generated with BEDTools (version 2.23.0) (Quinlan and Hall, 2010), normalized towards the sample with the lowest read count, converted into bigwig format and loaded into the Integrated Genome Browser (Freese et al., 2016). For calculation of FPKM's, we used Cuffdiff, part of Cufflinks (version 2.2.1) (Trapnell et al., 2012; Trapnell et al., 2010), with the options *-u -no-effective-length-correction -b*.

For initial clustering of the T/Sox2 subpopulations into 5 expression groups, we first performed two filtering steps. Firstly, we selected only those genes displaying FPKM greater 2 in at least one of the samples, and secondly, those genes with a fold-change greater than 2 between any two samples – resulting in a gene list of 1402 genes. We then employed a per gene normalization, such that the sum of the squared FPKM values in each row equal 1.0, and k-means clustering into 5 groups using Cluster 3.0 (de Hoon et al., 2004) which was visualized using Java TreeView (Saldanha, 2004). The boxplots of selected groups was performed using the same per gene normalized values and BoxPlotR (<http://boxplot.tyerslab.com>).

In order to compare the WT vs.  $T^{2J/2J}$  transcriptomes, we selected only those genes displaying FPKM greater 2 in at least one of the samples and with a fold-change greater than 2. We then sorted the genes from most highly repressed to most highly activated and plotted the FPKM values using Java TreeView (Saldanha, 2004).

For clustering of the *T* KO against the *Tbx6* KO transcriptomes, we first selected the most highly expressed genes with an FPKM > 10 in at least one of the samples. We then selected highly dysregulated genes with a fold-change of at least 3 in any pair-wise comparison. We then exported the log2 fold-changes between WT and KO and performed hierarchical clustering using Cluster 3.0 (de Hoon et al., 2004) and visualized the heatmap using Java TreeView (Saldanha, 2004).

### Single Cell RNA-Seq

Those libraries with very few reads (2 samples with less than  $1 \times 10^6$  reads) were excluded from further processing. Reads were mapped to the mm10 genome as previously described for the bulk RNA-seq analysis with TopHat2 (version 2.1.0) (Kim et al., 2013) using bowtie (Langmead et al., 2009), providing refSeq annotations in gtf format (UCSC) and the options *-no-mixed -no-discordant -g1 -mate-inner-dist 200 -mate-std-dev 100*, without initial trimming of reads. We then calculated FPKM values using Cuffquant and Cuffnorm from the Cufflinks (version 2.2.1) (Trapnell et al., 2012; Trapnell et al., 2010) package. Downstream analysis was carried out using the R Monocle package (version 2.4.0) (Qiu et al., 2017; Trapnell et al., 2014), by reading in the FPKM, sample and gene attributes tables from the Cufflinks output.

The FPKM values were converted into RPC values (Qiu et al., 2017) and initial quality control was performed. We filtered out genes displaying a low RPC value (< 0.1) in all the cells analyzed. In order to identify outliers, we calculated the mean and SD of RPC values across all the cells. Those with an RPC count greater than the mean + 1.5xSD (6 cells) or less than the mean - 1.5xSD (2 cells) were excluded from the analysis since they are likely to represent doublets or dead cells respectively. This resulted in a final amount of 45 cells used for downstream analysis. We then re-ran the quantifications using Cuffquant and Cuffnorm without the excluded cells.

We loaded the resulting files into Monocle, converted the FPKM values to RPC and repeated the quality control steps for verification. As expected, no cells were detected as outliers anymore. We then removed genes from the analysis with an RPC value less than 0.1 less than two cells. We then created a cell type hierarchy based on marker expression (RPC > 0.1) for neuroectoderm (co-expression of Sox1 and Sox3 or co-expression of Nkx1-2 and Pax3), mesoderm (co-expression of *Tbx6* and *Msgn1*) and putative NMP's (no expression of neuroectoderm and mesoderm markers, but co-expression of *T* and *Sox2*). Using these settings, we performed the initial cell type imputation, resulting in 4 neuroectodermal cells, 9 mesodermal cells, 3 NMP's and 29 cells of an unknown cell type. The dimensions were reduced using the tSNE method and cells were clustered into four groups. We then performed differential gene expression analysis between the 4 groups (p-value < 0.05) and selected the 500 most cell type-specific genes from this analysis. We checked for the variance explained by each component in the analysis and reduced the data to 2 dimensions accordingly using tSNE. We then repeated the cell type imputation using the selected marker genes and clustered the cells into 4 groups. This resulted in a final distribution of 15 neuroectodermal, 10 mesodermal, 3 NMP and 17 unclassified cells. These final cell groups are shown according to the PCA1 and PCA2 distance, displaying the expression levels of genes. In order to calculate a cell trajectory, we performed differential gene expression analysis between these final 4 groups and selected genes with a p-value < 0.04. The cell trajectory is shown based on PCA1 and PCA2 with expression values of marker genes or pseudotime.

### ChIP-Seq

ChIP-seq data was using bowtie. For *Tbx6*, we estimated the fragment size distribution of the library and determined the maximum length cutoff by performing an initial mapping of a limited number of reads using the options *-y -m 1 -S -I 100 -X 500*. The final mapping was performed using the same options and the determined maximum length parameter (-X). For the NMP ChIP's and input, we used *-X 1000* for the maximum length parameter.

Paired-end reads from the .sam file were combined in a bed file by generating the original fragment using a custom perl script, duplicates were removed and .wig files were generated using BEDtools. For peak detection, 10bp step .gff files were generated and loaded into CoCAS (Benoukraf et al., 2009). For NMP's, we made use of the input file and removed all reads overlapping with regions displaying enrichment in the input from downstream analysis.

In order to assign ChIP-seq peaks to genomic locations and putative target genes, we first overlapped the peaks with -5kb/+2kb of the transcription start sites of refseq annotated genes and assigned them as promoter associated. The remaining peaks were overlapped with +2kb from the transcription start site to +5kb after the end of refseq annotated genes and assigned them to be genic. The remaining peaks were called to be intergenic binding events. In order to associate intergenic peaks to putative targets, we isolated both the closest up- and down-stream gene. These peak/gene associations were also used for performing overlaps with WT vs. mutant transcriptomes and to assign peaks to the five groups obtained from the T/Sox2 subpopulations.

For *de novo* motif analysis, were isolated the 10bp bin with the highest binding-value from the peak detection and extracted genomic sequences +/- 85bp from their location using bedtools. The resulting fasta files were used for motif discovery in Meme

(Bailey and Elkan, 1994) and to plot the distribution of discovered motifs. For the detection of associated motifs, we enlarged the flanking regions to +/-195 bp and submitted the resulting fasta files to MemeChIP (Machanick and Bailey, 2011) using default parameters.

Clustering of ChIP-seq data was performed using SeqMiner (Ye et al., 2011), using bed files containing the 10bp summits of T peak coordinates associated to either up or downregulated genes. The original bed files obtained after mapping, containing the elongated fragments, were loaded and clustered into either 2 or 4 groups when using either only the T/Sox2 or T/Sox2/Tbx6 datasets using the KMeans linear method. Clustering was performed at +/-2500bp around the peaks, disabling the read extension setting. Due to the lower signal/noise ration of the  $\beta$ -catenin ChIP, we used a percentile threshold of 90% for the T/Sox2/ $\beta$ -catenin cluster. For the T/Sox2/Tbx6 cluster, we used a percentile threshold of 75%. The remaining settings were the same in both clusters: wiggle step of 10, percentile threshold of 90%, 400 runs and a T threshold of 5.

### ATAC-Seq

ATAC-seq data reads were trimmed to 50bp length and adapters were detected and removed using fastq-mcf of ea-utils (<https://expressionanalysis.github.io/ea-utils>, version 1.04.738). Those reads in which the adapter sequences were not found, were further trimmed by 5bp (to account for potentially undetected but present short adapter sequences) and paired-end mapped using bowtie (Langmead et al., 2009), allowing a maximum fragment size of 2000bp. Possible PCR duplicates were removed using Picard (<https://broadinstitute.github.io/picard>, version 1.103). Reads for which the adapter sequences were found and removed were independently mapped from their mates to the mm10 genome (containing chromosomes 1-19, X, Y and M) using bowtie and supplying the options -y -m 1 -S.. After merging the resulting output files, the samtools ((Li et al., 2009), version 1.2) fixmate function was used to combine corresponding reads into mates. Only those properly mapped mates with a maximum fragment length of 2000bp were filtered and potential PCR duplicates were removed using Picard. The resulting pairs were combined with the previously mapped reads, resulting in the final alignment file. Inspection of the data revealed 35 regions in the genome, mostly repetitive and on average 18kb in size, displaying mapping artefacts (Buenrostro et al., 2015). We used samtools to remove reads, together with their corresponding mates, falling into these regions. Due to the repetitive nature of the Y chromosome and non-informative mitochondrial genome, reads mapped to either of them were filtered out.

For global analysis of the ATAC-seq data, we first generated wig files representing the genomic read density binned into 50kb windows. We then calculated the total coverage across each chromosome in every sample and normalized the datasets toward the sample with the lower genome coverage. To verify the normalization, the coverage scores across each chromosome was calculated again and the pre- and post-normalized data was visualized using boxplots generated with BoxPlotR (<http://boxplot.tyerslab.com>). We then calculated a log2 ratio of each dataset compared to the stage-matched NMP sample, smoothed the data using a sliding window of 3 bins and calculated the Pearson correlation matrices using the “corrplot” function in R. Hilbert curves were generated using the “HilbertCurve” (Gu et al., 2016) R package for chromosomes 1-19 and X.

### Venn Diagrams

Proportional 2-way and 3-way Venn diagrams were generated using BioVenn (Hulsen et al., 2008). The 4-way Venn diagram for comparing the T and Tbx6 direct targets were generated using Venny version 2.1 (<http://bioinfogp.cnb.csic.es/tools/venny/index.html>).

### DATA AVAILABILITY

The accession number for the sequencing data reported in this paper is GEO: GSE93524.