

RESEARCH ARTICLE

Genome-wide identification of notochord enhancers comprising the regulatory landscape of the brachyury locus in mouse

Dennis Schifferl, Manuela Scholze-Wittler, Alba Villaronga Luque*, Milena Pustet, Lars Wittler, Jesse V. Veenvliet*, Frederic Koch[‡] and Bernhard G. Herrmann[‡]

ABSTRACT

The node and notochord are important signaling centers organizing the dorso-ventral patterning of cells arising from neuro-mesodermal progenitors forming the embryonic body anlage. Owing to the scarcity of notochord progenitors and notochord cells, a comprehensive identification of regulatory elements driving notochord-specific gene expression has been lacking. Here, we have used ATAC-seq analysis of FACS-purified notochord cells from Theiler stage 12-13 mouse embryos to identify 8921 putative notochord enhancers. In addition, we established a new model for generating notochord-like cells in culture, and found 3728 of these enhancers occupied by the essential notochord control factors brachyury (*T*) and/or *Foxa2*. We describe the regulatory landscape of the *T* locus, comprising ten putative enhancers occupied by these factors, and confirmed the regulatory activity of three of these elements. Moreover, we characterized seven new elements by knockout analysis in embryos and identified one new notochord enhancer, termed *TNE2*. *TNE2* cooperates with *TNE* in the trunk notochord, and is essential for notochord differentiation in the tail. Our data reveal an essential role of *Foxa2* in directing T-expressing cells towards the notochord lineage.

KEY WORDS: Mouse, Embryo, Development, Notochord, Brachyury, Enhancer

INTRODUCTION

The notochord is a rod-like structure situated below the neural tube and extending all along the trunk and tail of the embryo. As the defining characteristic of chordates, it serves two major functions: to provide structural stability to the developing embryo and to secrete signals for patterning of neighboring tissues (Stemple, 2005). It comprises an epithelium wrapped around a core of vacuolated cells, which makes the structure both solid and laterally flexible. In aquatic animals, these characteristics are essential and allow the embryo to swim. In fact, in some primitive fish the notochord serves as the main axial skeleton. In most chordates and all mammals, however, the notochord is a transient embryonic signaling center,

which regresses during fetal development and forms the nucleus pulposus integrated in the intervertebral discs (Choi et al., 2008; McCann et al., 2012).

Adjoining organ anlagen of all three germ layers, the notochord lies ventral to the neuroectoderm, dorsal to the gut endoderm, and is flanked by somitic mesoderm. Patterning of the neural tube via secretion of Hedgehog ligands and Bmp antagonists from the notochord is a well-described signaling function (Briscoe and Ericson, 1999; McMahon et al., 1998). Further, the node induces the left-right axis and the notochord provides signals for patterning the endoderm during organogenesis (Cleaver and Krieg, 2001; Shiratori and Hamada, 2006). In the mesodermal lineage, the notochord is required for somite patterning (Brand-Saberi et al., 1993; Fan and Tessier-Lavigne, 1994), sclerotome induction, vertebral column differentiation and segmentation (Ward et al., 2018).


Axial mesoderm emerges during gastrulation and forms the prechordal plate, the anterior head process, and the node, which gives rise to trunk and tail notochord (Sulik et al., 1994; Tam and Beddington, 1987; Wymeersch et al., 2019; Yamanaka et al., 2007). Morphogenesis of these structures is controlled by a set of developmental transcription factors, in particular *Foxa2* and brachyury (*T*; also known as *Tbxt*). The endodermal master regulator *Foxa2* is essential for notochord specification at all axial levels (Ang and Rossant, 1994; Weinstein et al., 1994). Node formation and specification of trunk and tail notochord is controlled by the pan-mesodermal regulator brachyury in a dosage-dependent manner (Herrmann, 1995; Herrmann et al., 1990; Stott et al., 1993). In addition, the homeobox transcription factor *Noto* is required for tail notochord and disrupted in the truncate mutant, which displays a shortened tail phenotype (Abdelkhalik et al., 2004; Plouhinec et al., 2004; Zizic Mitrecic et al., 2010). During trunk notochord morphogenesis, *Noto* functions synergistically with *Foxa2*, before it becomes essential for tail notochord maintenance (Yamanaka et al., 2007). Of the three transcription factors, *Noto* is the only factor that is exclusively expressed in axial mesoderm precursors from embryonic day (E) 7.5.

The combined activities of *T*, *Foxa2*, *Noto* and other transcription factors establish the gene regulatory network that governs notochord morphogenesis (Di Gregorio, 2020; Matsumoto et al., 2007; Passamaneck et al., 2009; Song et al., 2023; Yamanaka et al., 2007). So far, few notochord enhancers have been identified in mice, including *TNE*, *Foxa2 NE*, *NOCE* and *Sspe2* at the *T*, *Foxa2*, *Noto* and *Shh* loci respectively (Alten et al., 2012; Jeong and Epstein, 2003; Nishizaki et al., 2001; Schifferl et al., 2021). We hypothesized that, in addition to *TNE*, which is active during trunk notochord specification and essential for tail development, a second enhancer compensating for its loss must be located upstream of the *T* gene (Schifferl et al., 2021). Enhancers can be predicted by assessing chromatin accessibility, indicative histone modifications and transcription factor binding (Heintzman et al., 2007;

Max Planck Institute for Molecular Genetics, Department Developmental Genetics, Ihnestr. 63-73, 14195 Berlin, Germany.

*Present address: Max Planck Institute of Molecular Cell Biology and Genetics, StemBryogenesis Group, Pflotenhauerstraße 108, 01307 Dresden, Germany.

[‡]Authors for correspondence (koch@molgen.mpg.de; herrmann@molgen.mpg.de)

 D.S., 0000-0003-3507-5734; M.P., 0000-0003-3825-6530; F.K., 0000-0002-0142-2411; B.G.H., 0000-0002-2192-8188

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Handling Editor: James Briscoe

Received 22 June 2023; Accepted 17 October 2023

Rada-Iglesias et al., 2011; Visel et al., 2009). Previous studies on notochord enhancers were limited owing to the small number of cells available from embryonic material and the restricted accessibility of axial mesoderm (Tamplin et al., 2011).

In this study, we present a comprehensive and integrated approach utilizing assay for transposase-accessible chromatin with sequencing (ATAC-seq), chromatin immunoprecipitation with sequencing (ChIP-seq) and transcriptome profiling to identify notochord enhancers throughout the genome and their corresponding target genes. We elucidate the cis-regulatory landscape of the *T* locus comprising multiple enhancers, and identify crucial enhancers required for *T* expression in the notochord and essential for notochord formation and differentiation.

RESULTS AND DISCUSSION

Genome-wide identification of notochord-specific enhancers bound by T and/or Foxa2

The notochord is characterized by co-expression of brachyury and Foxa2, which are also expressed in neuro-mesodermal progenitors (NMPs) and mesoderm or endoderm, respectively. Noto, by contrast, is notochord specific. To characterize the enhancer landscape of the notochord, we engineered a Noto::H2B-mCherry/T::Venus/Foxa2::mTurquoise triple reporter mouse embryonic stem cell (mESC) line allowing the isolation of putative notochord (Noto⁺/T⁺/Foxa2⁺) progenitor cells (NotoPs; Fig. 1A). The reporter line was used for the generation of embryos by diploid morula aggregation (Fig. 1B). The caudal ends of E8.5 and E9.5 embryos were dissected at the somite border, and Noto^{mC+}/T^{ve+}/Foxa2^{mT+} cells were purified by fluorescence-activated cell sorting (FACS) (Fig. 1C). For comparison, we also purified trunk notochord (Fig. S1A) and paraxial mesoderm progenitors (MPs; T^{ve+}/Noto^{mC-}/Foxa2^{mT-}; Fig. 1C). E8.5- and E9.5-derived cells were used for transcriptome profiling (Fig. 1C, Fig. S1A), cell pools from Theiler stage 12 and 13 embryos for ATAC-seq analysis (Fig. S1B-D).

Embryonic notochords do not provide sufficient cell numbers for generating ChIP-seq data from FACS-sorted material. To overcome this limitation, we developed an *in vitro* model for differentiation of mESCs into NotoPs. In parallel, we produced MPs under similar experimental conditions. Notochord cells require high levels of T in combination with Foxa2. Therefore, we generated an mESC line allowing doxycycline (Dox)-inducible expression of Foxa2 and eGFP, driven by reverse tetracycline-controlled transactivator (rtTA; Gossen et al., 1995) expressed under control of the T promoter and enhancers. In addition, this line carries a Noto^{mT} reporter (Fig. 1D). We followed a well-established protocol for generating NMPs on day 3 of culture (Fig. 1E; Gouti et al., 2014). CHIR 99021 (CHIR) treatment for one more day generated MPs (herein termed MP^{diff} to distinguish them from embryonic MPs), as previously shown. However, parallel treatment with CHIR and Dox for 2 days (d2-d4) caused Foxa2 and eGFP expression, and efficiently generated Noto^{mT+} cells, as shown by FACS and fluorescent microscopy (Fig. 1F; Fig. S2A). To investigate whether these cells resemble embryonic NotoPs, we isolated Noto^{mT+}/T^{mC+} (NotoP^{diff}) cells (Fig. 1F) and performed RNA sequencing (RNA-seq). In parallel, we analyzed MP^{diff} cells (T^{mC+}/Noto^{mT-}). We compared the transcriptomes of *in vitro*-generated cells with embryonic cells isolated from E8.5 and E9.5 embryos. For embryonic NotoP and MP cells, we identified 319 and 109 specifically expressed genes, respectively [3-fold change of fragments per kilobase of transcript per million mapped reads (FPKM) value in NotoP versus MP at E8.5 and E9.5], including many known markers for both cell types (Fig. 1G,

Fig. S2B, Table S1). The comparison revealed striking similarity between the expression profiles of embryonic and *in vitro*-generated cells. Considerable overlap was observed between notochord gene sets identified in NotoP cells and those reported in earlier studies (Fig. S2C, Table S2) (Tamplin et al., 2011; Wymeersch et al., 2019; Peck et al., 2017). Immunofluorescent staining (Fig. S2D) and qPCR confirmed that T and Foxa2 are upregulated in NotoP^{diff} cells, whereas expression of the paraxial mesoderm marker *Tbx6* is low (for detailed qPCR data analysis, see Fig. S2E). We conclude that our *in vitro* model is suitable for efficiently generating notochord-like cells strongly resembling embryonic NotoPs.

Next, we identified putative notochord enhancers. We performed ATAC-seq on embryonic NotoPs and MPs (Fig. S1B-D). Differential peak detection identified 8921 open regions outside of promoters (± 5 kb from the transcription start site) with higher accessibility in NotoPs compared with 4876 regions with higher accessibility in MP cells (Fig. 2A; Fig. S3A). To characterize these accessible regions with respect to T and/or Foxa2 binding, we generated cultures enriched for NotoP^{diff} cells (CHIR⁺; Dox⁺; d4). We performed ChIP-seq for T and Foxa2 on bulk cultures. In addition, for determining specific chromatin signatures we FACS-purified NotoP^{diff} [CHIR⁺; Dox⁺; d4; cell population (P)7] and MP^{diff} (CHIR⁺; Dox⁻; d4; P5) cells as above (Fig. S2E) and analyzed several histone marks (H3K27ac, H3K4me3, H3K4me1 and H3K9me2) by ChIP-seq. For comparison, we used T ChIP-seq data previously generated from *in vitro*-differentiated NMPs (Koch et al., 2017).

The co-expression of T and Foxa2 in our *in vitro* model strikingly changed the chromatin landscape from a mesoderm signature (MP^{diff}; CHIR⁺; Dox⁻; d4) to a notochord signature (NotoP^{diff}; CHIR⁺; Dox⁺; d4) (Fig. 2B). Among the 8921 open regions differentially accessible in NotoPs versus MPs, we identified 3728 sites bound by T and/or Foxa2 in NotoP^{diff} cells (Fig. 2B). Using H3K27ac data as a strong indicator of enhancer activity (Creighton et al., 2010), we found the highest signal in enhancers binding both T and Foxa2 [category (Cat) 3, followed by 5 and 1; Fig. 2C; Fig. S3B]. Cat 3 enhancers also showed the highest ATAC signal in embryonic NotoPs, and were also bound by T in NMPs. Cat 5 enhancers specifically occupied by T and Foxa2 in NotoP^{diff} cells showed somewhat lower ATAC and H3K27ac signal. These data suggest that Foxa2 and T might have additive effects on enhancer activation. Based on the H3K27ac signal, Foxa2 binding alone (Cat 7) seems to be more effective than T alone (Cat 2, 4, 6), whereas the accessibility of Cat 4, 5 and 6 enhancers appears to be equal. However, some of these differences may be caused by co-binding of additional factors not analyzed in this study.

A considerable fraction (2393/3728; 64.2%) of notochord enhancers showed T binding alone (Cat 2, 4 and 6), and a smaller fraction (757/3728; 20.3%) only Foxa2 binding (Cat 7). Most of the T-binding enhancers (2131/3728; 57.1%) were also T bound in NMPs, but 840 (22.5%) T-bound enhancers were only occupied in notochord cells. Thus, Foxa2 co-expression with T opens chromatin at enhancers not occupied by T when Foxa2 is not expressed, and this cooperative action changes the genomic landscape to a notochord signature.

Next, we determined the enhancer presence in the vicinity of 319 genes specifically upregulated in NotoP cells (Fig. 1G). We found at least one notochord enhancer each in the gene bodies or genomic intervals comprising the intergenic regions and gene bodies of the immediate neighbors of 146 (46.4%) of these genes (Fig. 2D). Fig. 2E shows the number of genes in relation to the category(s) and number of enhancers found in their neighborhood, with examples of

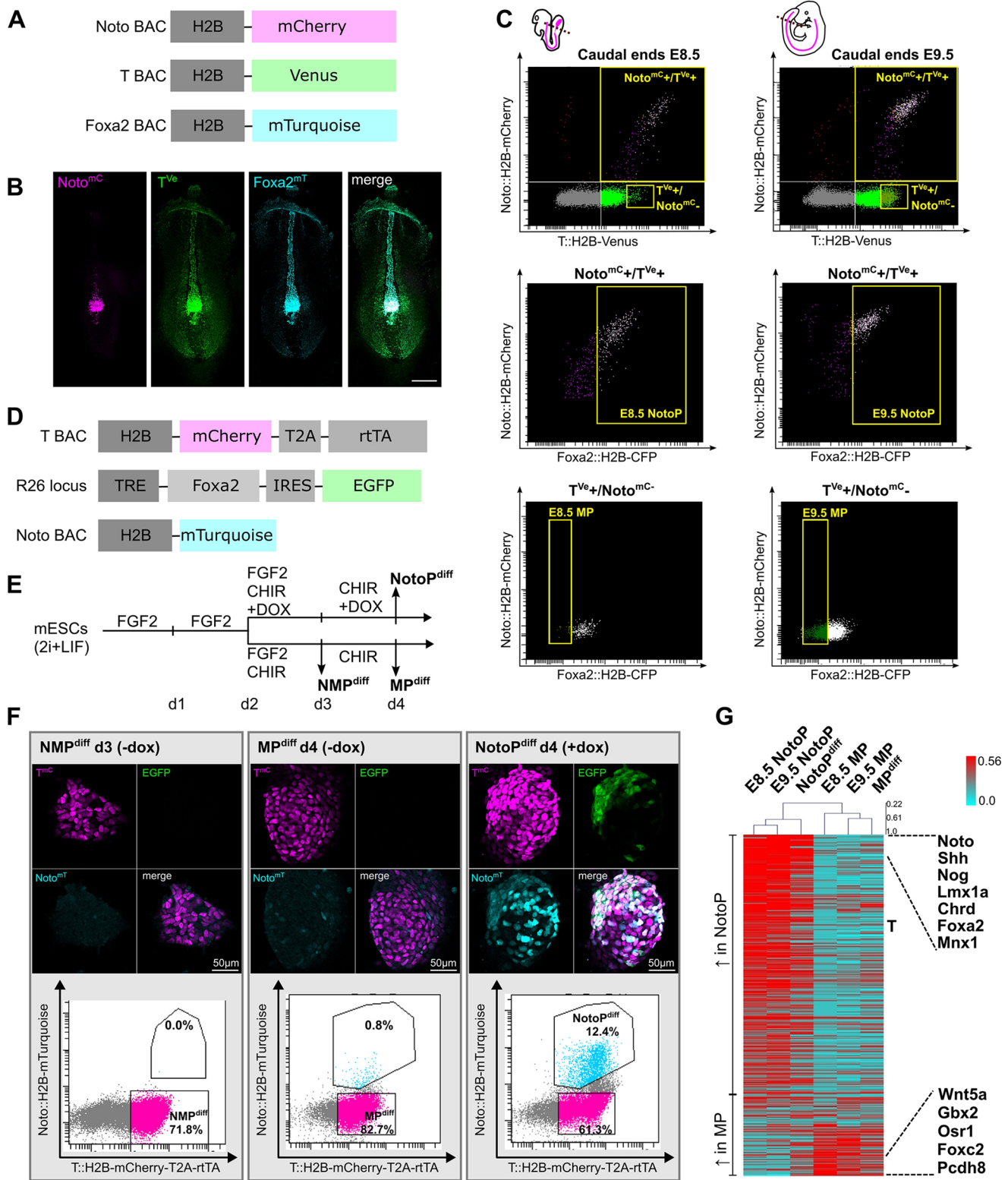


Fig. 1. Derivation of NotoP and MP progenitor cells *in vivo* and *in vitro*. (A) BAC constructs integrated into an mESC line for the establishment of a Noto^{mC}/T^{Ve}/Foxa2^{mT} triple reporter line. (B) Maximum intensity projection of images acquired by confocal microscopy of an E8.5 Noto^{mC}/T^{Ve}/Foxa2^{mT} mouse embryo. Scale bar: 200 μm. (C) FACS gating for purification of NotoPs and MPs from caudal ends of E8.5 and E9.5 embryos dissected at the somite border (as indicated by dotted lines in schematics; pink lines indicate the notochord). (D) Genetic modifications of an mESC line used for T-driven, Dox-inducible overexpression of Foxa2. (E) Differentiation scheme for the *in vitro* generation of NMPs, MPs and NotoPs from mESCs carrying the modifications shown in D; modified from Gouti et al., 2014. (F) Images of differentiated colonies acquired by confocal microscopy, and corresponding FACS profiles. (G) Ranked heat map showing per gene normalized FPKM values of 418 differentially expressed genes (3-fold change in NotoPs versus MPs at both E8.5 and E9.5) and corresponding *in vitro*-generated cells (NotoP^{diff}, MP^{diff}). Genes are sorted by average fold change in E8.5 and E9.5 samples. Examples of notochord (top) and mesoderm (bottom) progenitor marker genes are indicated on the right.

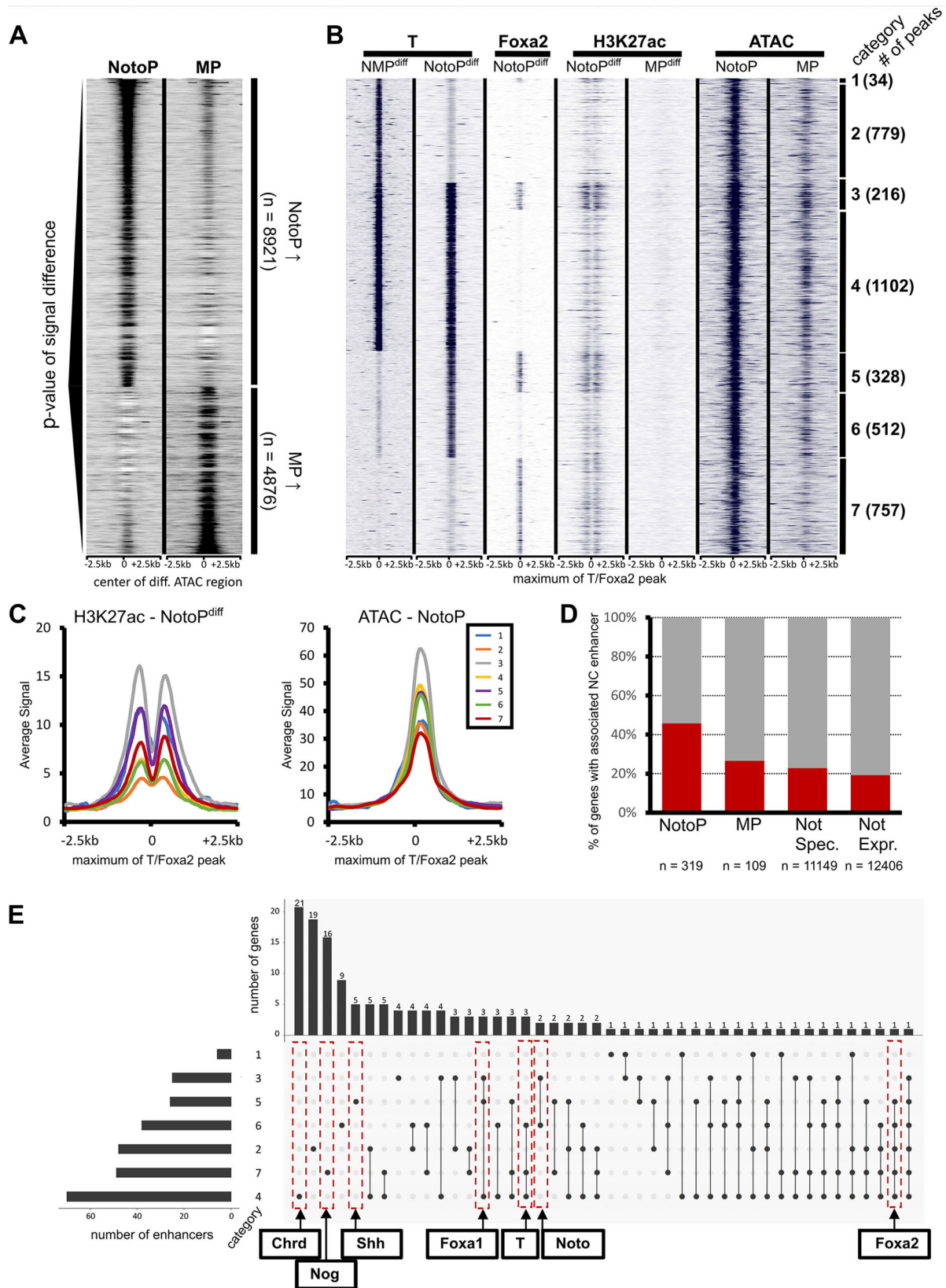


Fig. 2. Genome-wide prediction and analysis of notochord enhancers. (A) Heatmap showing ATAC-seq signals of NotoP and MP cells around 8921 notochord enhancer candidates defined by differential peak detection, sorted by *P*-values. (B) Heatmap of 3728 differential NotoP accessible sites bound by T and/or Foxa2 in NotoP^{diff} cells. Putative enhancers are categorized into all possible combinations of T and Foxa2 binding combinations (1-7, right) and the number of putative enhancers per category is shown in parentheses. (C) Average profiles of H3K27ac and ATAC-seq signals for all enhancer categories. (D) Bar plot quantifying the association of putative notochord enhancers to NotoP, MP, not specifically expressed (Not Spec.) and not expressed (Not Expr.) genes. (E) UpSet plot for the 146 NotoP-specific genes showing their associated enhancers according to categories 1-7.

prominent notochord marker genes. Strikingly, most of the latter are associated with several putative enhancers falling into different enhancer categories and none of them is associated with a T-NMP^{diff} peak (Cat 2) alone. Genome browser screenshots of important markers are given in Fig. S4, and of housekeeping genes for control in Fig. S5. In addition to the notochord enhancers described previously, we found new putative enhancers, for example for *Noto*, *Foxa2*, *Chrd*, *Nog* and *Foxa1* (Schifferl et al., 2021; Alten et al., 2012; Nishizaki et al., 2001; Jeong and Epstein, 2003). A list of all notochord enhancer candidates identified here and their associated genes, as well as the list of the 319 NotoP genes and the categories of their associated enhancers, is given in Tables S3 and S4, respectively. Our data provide an important resource of putative notochord enhancers.

The regulatory landscape of the *T* locus in notochord and mesoderm

Next, we used our integrated ATAC and ChIP data to characterize further the regulatory landscape of the *T* locus in notochord and mesoderm. We previously identified the *TNE* enhancer, which is essential for tail notochord development; however, this enhancer alone does not fully explain the severe loss of notochord and tailbud phenotype of a 37 kb deletion (*T^{UD}*), suggesting that additional notochord and mesoderm elements are present in this region (Schifferl et al., 2021). Our ATAC-seq data of the *T* locus showed ten peaks, including *TNE* (Fig. 3).

Three ATAC-seq peaks in close vicinity to each other were bound by T and *Foxa2* and showed enrichment for the active enhancer mark H3K27ac in NotoP^{diff} cells (Fig. 3). Therefore, we used a 4.4 kb fragment encompassing these T- and *Foxa2*-binding sites to perform an enhancer assay. The embryonic activity assay at E9.0 showed Venus reporter expression in the posterior notochord, decreasing towards the head, identifying the 4.4 kb fragment as a notochord-specific enhancer (Fig. 3B-B''). We designated this enhancer *T locus notochord enhancer 2* (*TNE2*).

In addition, based on ATAC peaks and, in a subset, also H3K27 acetylation, we defined several more candidate *T* locus enhancers, *TE1*, *TE3*, *TE9* and *TE8*, within the *T^{UD}* region (Schifferl et al., 2021), as well as *TE6* –5 kb upstream and *TE7* 13 kb downstream of the *T* transcription start site (Fig. 3). We integrated ChIP-seq data for T, Sox2, β -catenin and Tbx6 published previously, to relate these enhancers to regulatory processes occurring in NMPs and during mesoderm formation (Koch et al., 2017). For example, Sox2 acting antagonistically to T in the neural versus mesodermal lineage choice, binds to *TE6* and *TE9* in NMPs, whereas the WNT signal mediator β -catenin, which cooperates with T in NMPs and NotoPs, binds to *TE3* and *TNE*. In contrast, Tbx6, which has a repressive effect on NMP maintenance and *T* expression, is detected at *TE1*, *TNE*, *TNE2* and possibly *TE3*. All four TFs also bind at the *T* promoter.

We assayed the activity of *TE3*, characterized by T peaks flanked by *Foxa2* binding in NotoP^{diff} cells, using the HSP68-Venus reporter. This element showed expression in the tailbud mesoderm, posterior neural tube and gut of the tail at E10.5, but not in the notochord (Fig. S6A-A''). These data suggest that *TE3* acts as enhancer in tailbud NMPs. The significance of the flanking *Foxa2* peaks is unclear.

Moreover, we assayed the putative enhancer *TE7*, which is located downstream of the *T* gene and was not identified by chromatin marks or differential ATAC analysis, but showed a small ATAC peak and T binding in NotoP^{diff} cells. β -Galactosidase reporter activity was detected in the entire notochord of E9.0

embryos, albeit with increasing staining towards the head (Fig. S6B). Thus, *TE7* identifies another notochord enhancer. The data suggest that the activity of *TE7* is increasing during notochord differentiation, which might explain why it was not detected in our differential ATAC-seq data derived from caudal end NotoPs.

A new *T* locus notochord enhancer, *TNE2*, cooperates with *TNE* in notochord development

Next, we investigated the function of several putative *T* locus enhancers by employing the CRISPR/Cas9 system to generate a series of knockout (KO) mESC lines and embryos carrying the *Noto::H2B-mCherry* (Noto^{mC}) reporter (Fig. S7, Table S5). We generated mutant embryos by tetraploid complementation assays and evaluated the phenotype between E9.5 and E12.5 by visualizing the Noto^{mC} reporter as well as T and Sox2 protein expression. Deletions of *TE1*, *TE3*, *TE9* and *TE7* did not result in embryonic defects enhancing the phenotypes of the corresponding parental line, which contained either the wild-type *T* locus or the heterozygous *T^{LD}* deletion covering the entire *T* locus (Schifferl et al., 2021). The data suggest that these cis control elements either have no critical function or are redundant (Table S6).

However, homozygous deletion of *TNE2* resulted in a notochord deficiency phenotype. In E9.75 *T^{ΔTNE2/ΔTNE2}* mutant embryos, less T protein was detected in the trunk notochord marked by Noto^{mC} than in wild-type embryos, and the notochord was partially disrupted at the hindlimb level (Fig. 4A,A',E,E'). At E11.5, the Noto^{mC+} notochord progenitor domain was present in the outgrowing tailbud, but the number of Noto^{mC+} cells in the midline was reducing quickly towards the anterior and the notochord was not formed (Fig. 4B-D,F-H). The reduced T protein level in the trunk notochord was apparently still sufficient for trunk development, but tail notochord formation was not supported resulting in a tail-less phenotype at E12.5 (Fig. 4M-T).

In E9.75 *T^{ΔTNE; ΔTNE2/ΔTNE; ΔTNE2}* double KO embryos, T expression was detected in the tailbud, but not in notochord progenitors, and the notochord was not formed (Fig. 4I,I'). A few cells with weak Noto^{mC+} signal were visible in gut endoderm, indicating improper specification (Fig. 4I'). Neither Noto^{mC+} cells nor a notochord were detected in the tailbud at E11.5 (Fig. 4J-L). Immunofluorescence staining for Olig2 and Nkx2-2 revealed neural tube patterning defects (Fig. S8). Olig2 expression was shifted ventrally, whereas Nkx2-2 was not detected. Consistent with the latter, neural tube differentiation was severely affected at E12.5, resulting in embryonic lethality (Fig. 4U-X).

Our data show that *TNE* can compensate for the loss of *TNE2* during trunk notochord formation and vice versa (Fig. 5; Schifferl et al., 2021). The double KO phenotype shows that the combined activity of these two enhancers is required for notochord formation and differentiation in the trunk and tail. *TE7* is not able to compensate for the loss of *TNE* and *TNE2*. The reduced T protein level in either KO suggests that the two enhancers cooperate during trunk notochord development. The partial disruption of the notochord in the posterior trunk of either single KO suggests an increasing requirement for T expression in notochord progenitors along the axis, as has been reported previously (Stott et al., 1993; Herrmann, 1995). Alternatively, higher T activity may be required during progenitor formation in the node in order to generate sufficient cell numbers supporting notochord development throughout the trunk. Tail notochord formation, however, needs the combined activity of both enhancers.

The single loss-of-function phenotypes reveal different roles of *TNE* and *TNE2* in the tail. There, *TNE* is mainly required for notochord progenitor maintenance, notochord cell specification and

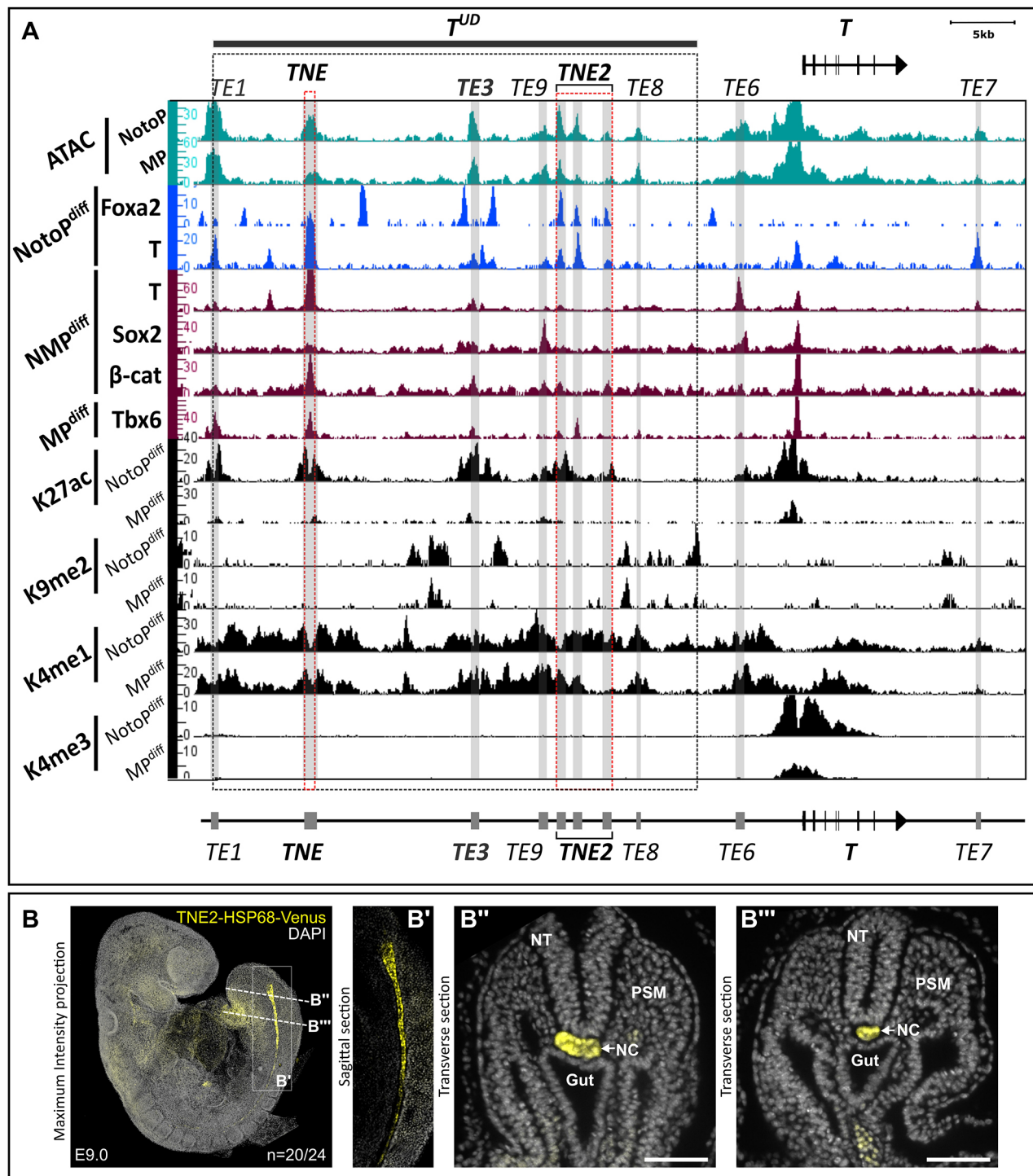


Fig. 3. Identification of putative enhancers at the mouse *T* locus. (A) Genome browser snapshot showing the mouse *T* locus with signal density tracks of ATAC-seq, transcription factor ChIP-seq and histone 3 ChIP-seq data. Track maxima are normalized to the number of mapped reads for each antibody. Enhancer elements are indicated as gray boxes and shown schematically below the ChIP-seq tracks; TNE and TNE2 are highlighted by red dashed lines. The T^{UD} and TNE regions were described previously (Schifferl et al., 2021). (B) *TNE2*-driven Venus reporter expression in the notochord of an E9.0 embryo. Nuclei are stained with DAPI (gray). Left: maximum intensity projections of confocal microscopy. (B') Sagittal optical midline section. (B'') Transverse optical sections acquired by light-sheet microscopy at the axial levels indicated in B. NC, notochord; NT, neural tube; PSM, presomitic mesoderm. Scale bars: 100 μ m.

presumably differentiation, given that embryos lacking *TNE2* still form *Noto*^{mC+} notochord progenitor cells, whereas in *TNE* mutants notochord cell specification is lost (Schifferl et al., 2021). *TNE2*, by contrast, is required for proper notochord cell differentiation, as in

embryos lacking *TNE2* *Noto*^{mC+} cells are still present in the progenitor domain, but are quickly reducing in number in the midline toward the anterior and the notochord is lacking. Of note, our data suggest cooperative binding of T and Foxa2 at *TNE2*,

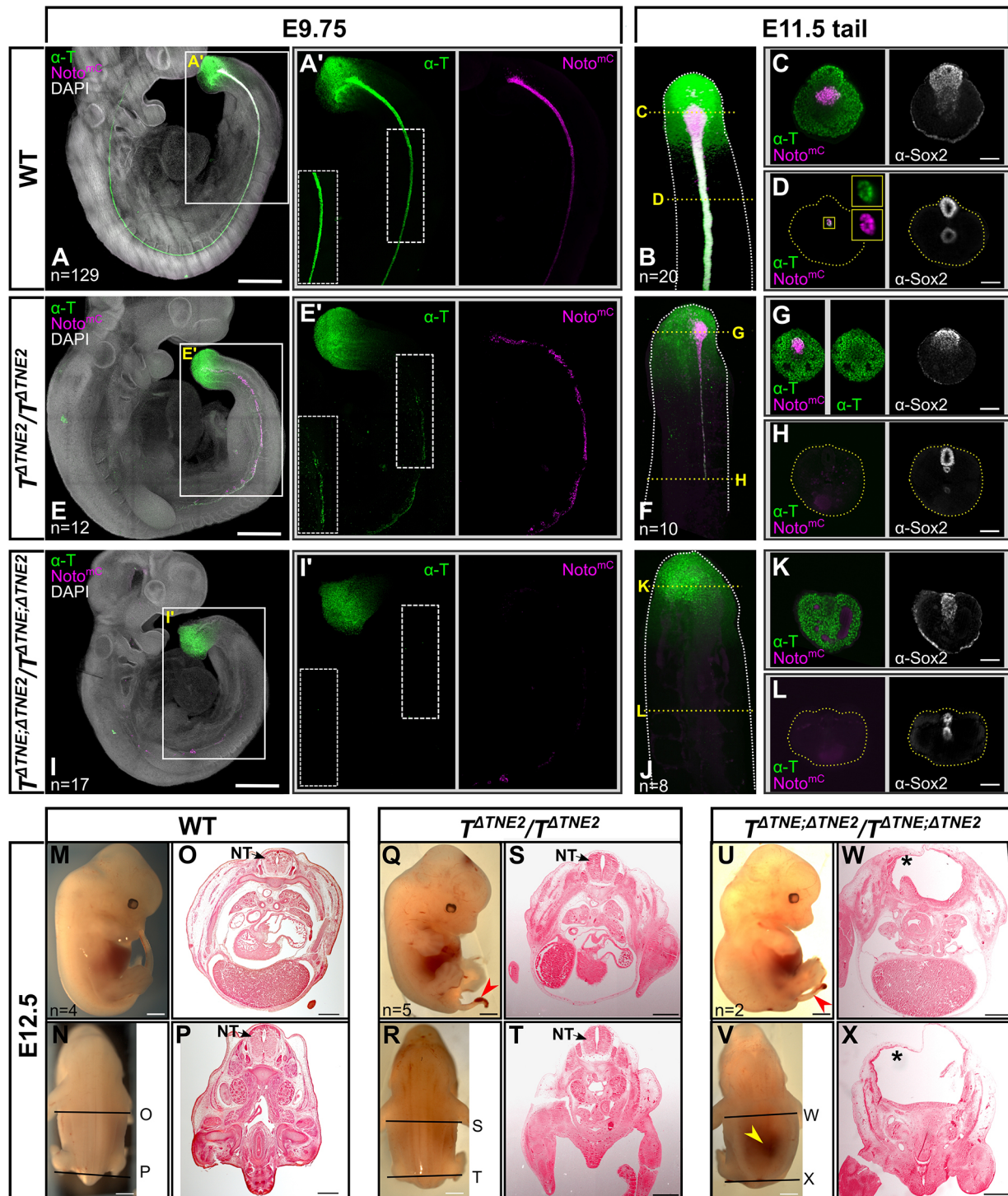


Fig. 4. *TNE2* is an essential notochord enhancer of brachyury and interacts with *TNE*. (A,A',E,E',I,I') Maximum intensity projections of E9.75 embryos with *Noto^{mC}* reporter signal, immunostaining for T (green) and DAPI nuclear staining (gray). Scale bars: 500 μ m. Boxed areas are magnified in single channels. To visualize the lower T signal of the mutant notochord in (E'), maximum brightness was equally adjusted in the insets at the bottom left of A',E',I'. (B-D,F-H,J-L) Maximum intensity projections of E11.5 tails with immunostaining for T (green) and the *Noto^{mC}* reporter signal (magenta) are shown in B,F,J. Yellow dashed lines indicate the position of optical sections shown in C,D,G,H,K,L. Immunostaining for Sox2 is shown in white, insets in D show a magnification of the notochord (boxed). Dotted lines indicate the circumference of the section. Scale bars: 100 μ m. (M,N,Q,R,U,V) Lateral and dorsal views of E12.5 embryos. Red or yellow arrowheads indicate the tail phenotype or neural tube defect, respectively. Scale bar: 1 mm. (O,P,S,T,W,X) Histological sections at the axial levels indicated in N,R,V. Asterisks indicate lack of a neural tube (NT) (W,X). Scale bars: 500 μ m. WT, wild type.

whereas *TNE* appears to bind T and β -catenin, consistent with the role of canonical Wnt signaling in notochord specification (Ukita et al., 2009). Thus, T appears to act earlier in the lineage through

TNE, followed by cooperative action with *Foxa2* through *TNE2*. Our enhancer reporter assays suggest that *TE7* is active even later, upon further differentiation and possibly in the maintenance of the

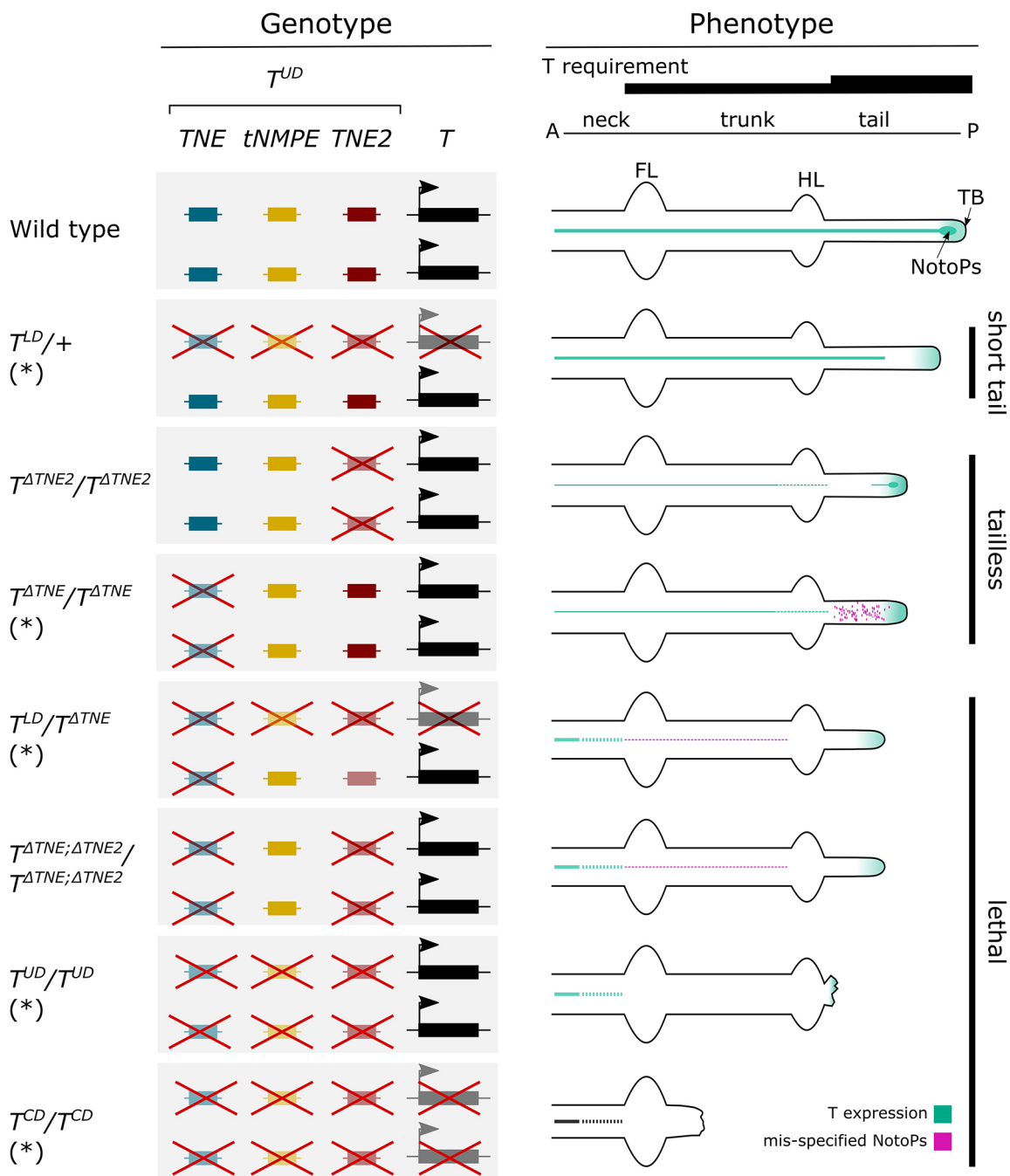


Fig. 5. Schematic view of *T* enhancer function in axial development. Axial elongation depends on *T* expression in the tailbud and in the notochord (green line). Notochord formation in the trunk requires at least one of the enhancers *TNE* or *TNE2*, tail notochord formation both. *T* enhancers located in the T^{UD} deletion and the *T* gene are shown as boxes on the left; loss-of-function mutations are indicated by red crosses. A, anterior; FL, forelimb bud; HL, hindlimb bud; P, posterior; TB, tailbud; tNMPE, tail NMP enhancer, defined by failure of tailbud outgrowth in T^{UD}/T^{UD} embryos. Dotted lines/areas indicate interrupted notochord (green) or mis-specified NotoPs (red), respectively. For details, see the text; data from Schifferl et al. (2021) are marked by an asterisk.

notochord, as during axis elongation the reporter activity is low in the posterior end and increasing towards the head. The strong T^{UD} homozygous deletion phenotype reported previously is, at least with respect to the lack of the notochord, explained by the loss of *TNE* and *TNE2*, but *TE7* is not involved (Schifferl et al., 2021). However, the missing tailbud in the deletion mutant is not explained, as *TNE/TNE2* double KO embryos still form the tailbud and show some tail outgrowth. *TE3* driving expression in tailbud NMPs might be responsible for tail outgrowth in the absence of *TNE/TNE2*.

However, our *TE3* deletion analysis did not show the expected loss of the tailbud in the absence of the enhancer, even in combination with the deletion of *TE9*, an enhancer candidate showing Sox2 and β -catenin binding in NMPs (Fig. S7; Table S6). Therefore, the control element(s) required for tailbud formation and tail outgrowth must remain undefined at this point.

The *in vitro* model for notochord progenitor formation introduced here is based on the essential role of both *T* and *Foxa2* in node and notochord development. Accordingly, WNT (Chiron) induction is

employed for parallel activation of *T* and *Foxa2*, the latter under control of the *T* promoter and enhancers comprised by a BAC construct. Because notochord is not derived from NMPs, our data suggest that *T* is already activated in epiblast stem cell-like cells. Strong upregulation of *Foxa2* supposedly directs a fraction of $T^+/Foxa2^+$ cells into the notochord lineage, whereas cells expressing no or low levels of *Foxa2* become NMPs and take the mesodermal fate. However, whether NMPs and notochord-like cells have a common progenitor or derive from different subsets of epiblast stem cell-like cells remains unresolved at this point.

MATERIALS AND METHODS

General mESC procedures

All cell lines used in this study were derived from male mESCs of the G4 hybrid line 129S6/SvEvTac×C57BL/6Ncr (George et al., 2007). mESC clones were regularly tested for possible *Mycoplasma* contamination, using PCR Mycoplasma Test Kit II (Appligene, A8994) according to the manufacturer's recommendations.

Generation and integration of BAC transgenes

For the generation of reporter and driver transgenes, BACs containing ~200 kb of the C57/BL6 genome surrounding the mouse brachyury (RP24-530D23), *Foxa2* (RP23-254G2) and *Noto* (RP23-289M19) genes were obtained from BACPAC resources. The *T::H2B-Venus*, *T::H2B-mCherry-T2A-rtTA*, *Foxa2::H2B-mTurquoise* and *Noto::H2B-mTurquoise* BACs carrying neomycin, hygromycin, puromycin and blasticidin resistance cassettes, respectively, were engineered via Red/ET recombineering (Muyers et al., 1999) as described previously (Schifferl et al., 2021). For random integration of BACs, 5 µg of each construct were linearized using *Pi*-*SceI* (New England Biolabs, R0696S) and electroporated into 3×10^6 mESCs. Approximately 30 h after electroporation, selection was initiated, as shown in Table S5. Selection medium was refreshed daily until single colonies were clearly visible after approximately 1 week. Single clones were picked and genotyped by PCR. Oligonucleotides used in this study are listed in Table S7.

Generation of enhancer mutants

Homozygous $T^{\Delta TNE2}/T^{\Delta TNE2}$ single and $T^{\Delta TNE; \Delta TNE2}/T^{\Delta TNE; \Delta TNE2}$ double mutants were generated from cells carrying the *Noto::H2B-mCherry* construct using CRISPR/Cas9 as reported previously (Schifferl et al., 2021).

Recombinase-mediated cassette exchange

For the generation of enhancer reporter and *Foxa2* overexpression cell lines, 3×10^5 mESCs with a modified *Rosa26* harboring locus (Vidigal et al., 2010) were co-transfected with 5 µg of linearized NE2-HSP68-Venus, TE3-HSP68-Venus, TE7-HSP68-bGal or TRE-*Foxa2*-IRES-EGFP constructs and 1 µg PGK-iCre vector using Lipofectamine 2000 (Invitrogen, 11668027). Cells were cultured in ES+LIF containing 350 µg/ml geneticin (Thermo Fisher Scientific, 10131027) for selection for successful recombination resulting in a switch from hygromycin to neomycin resistance.

Generation of transgenic embryos

Transgenic mouse embryos were generated by diploid or tetraploid morula aggregation by the transgenic unit of the Max Planck Institute for Molecular Genetics in Berlin as described previously (Eakin and Hadjantonakis, 2006). All animal experiments were performed according to local animal welfare laws and approved by local authorities (covered by LaGeSo licenses G0243/18 and G0247/13).

Embryo isolation

Timed pregnant foster mice were euthanized by carbon dioxide administration and cervical dislocation. For whole-mount immunofluorescence and tissue clearing, embryos were isolated from uteri in 4°C pre-cooled PBS, fixed using 4% paraformaldehyde (PFA) in PBS (Sigma-Aldrich, P6148) in 4 ml glass vials (Wheaton, 224882) and processed as described previously (Schifferl et al., 2021). For RNA-seq and

ATAC-seq, embryos were isolated in M2 medium (Sigma-Aldrich, MR-015P). Samples were further dissected into the sub-regions of interest using forceps (Dumont, 11251-10). Tissue samples were kept on ice in M2 medium and processed subsequently.

Whole-mount immunofluorescence and tissue clearing

Immunofluorescence staining and clearing procedures were performed on E9.75-E11.5 embryos as described previously (Schifferl et al., 2021). Antibodies are listed in Table S8.

Whole-mount β-galactosidase staining

Embryos carrying the *TE7-HSP68-bGal* reporter were fixed for 30 min at 4°C and subsequently washed three times for 15 min in Rinse Buffer (50 mM EGTA, 0.1% deoxycholate, 0.2% NP-40, 20 mM MgCl₂ in DPBS) at room temperature. After rinsing, embryos were incubated in staining solution [50 mM K₃Fe(CN)₆, 50 mM K₄Fe(CN)₆, 50 mM EGTA, 0.1% deoxycholate (100×), 0.2% NP-40 (100×), 0.2 M MgCl₂, 1 mg/ml X-gal in DPBS] at 37°C overnight. Stained embryos were washed three times with PBS and stored in 4%PFA in PBS at 4°C for secondary fixation.

FACS

For FACS of embryonic material, single-cell suspensions were prepared adding 100 µl Trypsin/EDTA to the sample. After incubation at room temperature for 5 min, trypsin was quenched by adding 200 µl PBS with 5% bovine serum albumin (BSA; Sigma-Aldrich, A8412).

For FACS of cell cultures, cells were washed twice in PBS and dissociated by trypsinization at 37°C for 10 min. Trypsin/EDTA was quenched using a double volume of 5% BSA in PBS, then cultures were resuspended and kept on ice until further procedure.

All samples were immediately filtered (35 µm mesh) and sorted on a FACS Aria II (Becton Dickinson) flow cytometer. For transcriptome analysis, cells were sorted into 350 µl RLT Plus buffer (QIAGEN, 1053393) containing 1% β-mercaptoethanol (Sigma-Aldrich, M6250) in 1.5 ml low-binding tubes (Thermo Fisher Scientific, 90410) and stored at -80°C until further procedure. For ChIP, cells were sorted into 5% BSA in PBS in BSA-coated glass tubes.

Histology

PFA-fixed E12.5 embryos were dehydrated through an ethanol series (30%, 50%, 70%, 70%, 15 min each), processed in a MICROM STP 120 processor (Microm, 813150) and embedded in paraffin (Leica, 3801320) utilizing an EC 350-1 embedding station (Microm). Sections of 10 µm thickness were prepared using a rotary microtome (Microm, HM355S), transferred onto adhesion microscope slides (Menzel, K5800AMNZ72) and dried overnight at 37°C. Eosin (Merck, 109844) counterstaining was performed according to standard procedures and specimens were mounted in Entellan (Sigma-Aldrich, 107960). Sections were imaged using an AxioZoom V16 stereomicroscope (Zeiss).

Microscopy

Embryos were imaged using a Zeiss LSM880 laser-scanning microscope with Airyscan detector or Zeiss Light sheet LS Z1 with appropriate filters for DAPI, mTurquoise, EGFP, Venus, mCherry, Alexa 488 or Alexa 647. For light-sheet microscopy, specimens were cleared and embedded in 1.5% low melting agarose (Sigma-Aldrich, A9414) in PBS. Agarose columns containing the samples were cleared in RIMS overnight before acquisition. z-stacks of approximately 500 µm range were acquired with 8-10 µm intervals.

Processing was performed using ZEN Blue Version 3.1 and ZEN Black Version 2.3 (Zeiss) software. For the visualization of volumetric data, maximum intensity projections (Wallis et al., 1989) were generated using the maximum intensity projection volume rendering tool in ZEN Black. This tool generates 2D images from stacks of optical sections, displaying the brightest pixel (x,y), along the z-axis in the according position (x,y) on a 2D plane.

To visualize the weaker signal in mutant notochord cells (Fig. 4E',I'), the maximum brightness parameter was adjusted, reducing the difference (maximum brightness–minimum brightness) by 50%.

In vitro differentiation

NMP^{diff}, MP^{diff} and Noto^{diff} cells were derived from *T^{mC}-2A-irTA/TRE::Foxa2/Noto^{mT}* mESCs following an established protocol (Gouti et al., 2014) with previously described modifications (Koch et al., 2017). For Noto^{diff} generation, 1 ng/ml Dox was applied from d2 to d4.

RNA-seq

For transcriptome analysis of FACS-purified subpopulations, total RNA was isolated from 250 cells (or fewer; see Table S9) using the RNeasy MinElute kit (QIAGEN, 74204). RNA extraction was performed according to the manufacturer's protocol with an additional DNase digest step between two washes with 350 µl RW1. Therefore, reaction mixes of 10 µl DNase I (QIAGEN, 79254) and 1 µl (=10 U) DNase I (Roche, 4716728001) in 70 µl buffer RDD (QIAGEN, 1011132) were applied to the spin columns for a 15 min incubation at room temperature. Membranes were air dried for 10 min to remove the remains of ethanol and eluted in 15 µl H₂O.

Sequencing libraries were prepared using the Ovation SoLo RNA-seq system (NuGEN) according to the manufacturer's recommendations, starting at step A.9 with 12 µl purified DNA. After each amplification step, libraries were quantified with the Qubit High Sensibility DNA assay (Thermo Fisher Scientific, 12102). Library size was validated using DNA High Sensitivity Bioanalyzer chips (Agilent, 5067-4626).

cDNA library pools (150 nmol in 15 µl) with 16 bp barcode length (8 bp barcode+8 bp UMI) were sequenced on using the Illumina HiSeq4000 (for E8.5/E9.5 NotoP, MP and Noto trunk) or NextSeq 2000 (for Noto^{diff}, MP^{diff}, NMP^{diff}) platforms.

Prior to mapping, the first five nucleotides of the forward and reverse reads were trimmed using *fastx_trimmer* (http://hannonlab.cshl.edu/fastx_toolkit/index.html) according to the manufacturer's instructions (NuGEN). The resulting reads were mapped to chromosomes 1-19, X, Y and M of the mouse mm10 genome using TopHat2 (v2.1.1) and bowtie (v1.2.2) (Kim et al., 2013; Langmead et al., 2009) and the RefSeq annotation in gtf format (UCSC), providing the options '-no-coverage-search -no-mixed -no-discordant -g1 -mate-inner-dist 250 -mate-std-dev 100 -library-type fr-secondstrand'. Read duplications resulting from the PCR amplification of the library were removed using the NuDup deduplication script provided by NuGEN (<https://github.com/tecangenomics/nudup/>). Wiggle files were generated with BEDTools version 2.23.0 (Quinlan and Hall, 2010), converted to bigwig format and visualized in the Integrated Genome Browser (Freese et al., 2016). FPKM values were calculated using Cufflinks version 2.2.1 (Trapnell et al., 2013) with options '-u -no-effective-length-correction -b'. Using FPKM values, per gene normalization, generation of heatmaps and non-hierarchical clustering of samples was performed in MeV (Howe et al., 2011).

Gene set comparison

For comparison with our NotoP gene set, we obtained published microarray (Tamplin et al., 2011; Wymeersch et al., 2019) and RNA-seq datasets (Peck et al., 2017). For the former, we used the provided lists of notochord-specific genes (Tamplin et al.) or genes enriched in E8.5 rostral node and node-streak border relative to remaining E8.5 samples ('extended spatial analysis' in Wymeersch et al.). Gene annotations not found in our dataset were converted. For the Tamplin et al. dataset, seven genes (*Defc-rs2*, *EG639426*, *ENSMUSG00000074335*, *H3073F06*, *LOC672711*, *LOC674134* and *Pkd11l1*), and for the Wymeersch et al. dataset 12 genes (*4930533K18Rik*, *B930045J24Rik*, *Defa1*, *Defc-rs10*, *Defc-rs2*, *LOC100040592*, *LOC100044289*, *LOC100046120*, *LOC100048721*, *LOC381284*, *OTTMUSG00000017677* and *sc10003799.1_2*) were not found in our mm10 RefSeq annotation. These genes were removed from the comparative analysis.

The RNA-seq dataset was mapped to chromosomes 1-19, X, Y and M of the mouse mm10 genome using TopHat2 (v2.1.1) and bowtie (v1.2.2) (Kim et al., 2013; Langmead et al., 2009) and the RefSeq annotation in gtf format (UCSC), using the options '-no-coverage-search -g1 -library-type fr-firststrand'. FPKM values and differential gene expression were calculated using Cufflinks version 2.2.1 (Trapnell et al., 2013) with the options '-u -no-effective-length-correction -b'. We then selected genes with an adjusted *P*-value of <0.01 and a >4-fold increased expression in the notochord of

E12.5 embryos compared with nucleus pulposus from postnatal day 0, resulting in 429 differentially expressed genes. The four gene sets were then plotted using nVennR (Perez-Silva et al., 2018).

qPCR

For the qPCR analysis of different T^{mC}/Noto^{mT}/Foxa2^{EGFP} populations by FACS, 20,000 cells were sorted for each sample, with the exception of P7 in Dox⁺ d3 (4994 cells), P10 in Dox⁺ d3 (9977 cells) and d3.25 (12,297 cells), and P11 in Dox⁺ d3.5 (15,299 cells). For P11 in Dox⁺ d3 and d3.25, not enough cells were present to be sorted. The RNA was extracted as described for RNA-seq. For cDNA synthesis, 14 µl of each sample was reverse-transcribed using the QuantiTect Reverse Transcription Kit (QIAGEN, 205311) according to the manufacturer's instructions, including the optional increased incubation time of 30 min at 42°C. The resulting cDNA was then diluted with H₂O, adjusted to the amount of cells that were sorted in each sample. qPCR was performed on a StepOnePlus (Thermo Fisher Scientific) system using the GoTaq qPCR premix (Promega, A6001) according to the manufacturers' recommendations. Each sample and primer combination was loaded in triplicate. Primer sequences are listed in Table S7. Potential residual genomic DNA contamination was first tested using a primer pair detecting genomic DNA in a gene desert region and a melt curve analysis was performed for all reactions to control for potential off-target amplifications. The quantification was performed using StepOne Software v2.3 using the relative quantification (ΔΔCt) method. The -Dox P4 sample at d3 served as the reference sample for all calculations. Expression values were normalized using *Pmm2* as a reference gene.

ATAC-seq

ATAC-seq experiments were conducted following an established protocol (Buenrostro et al., 2013). A total of 2000 cells were utilized per ATAC-seq experiment. After FACS sorting, the cells were centrifuged (100 g for 5 min at 4°C) and the resulting pellets were resuspended in 50 µl of lysis buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% NP-40) before being subjected to centrifugation at 500 g at 4°C for 10 min. The supernatant was discarded, and each pellet was then resuspended in the transposition reaction mix (25 µl 2× TD buffer, 2.5 µl Tn5 transposase, 22 µl H₂O), followed by incubation at 37°C for 30 min. Subsequently, the reaction was halted by adding PB buffer (QIAGEN), and the tagmented DNA was purified using the MinElute kit (QIAGEN). The purified DNA was combined with ATAC index PCR primers and 2× Kapa HiFi Hotstart Readymix and subjected to pre-amplification (98°C for 30 s, followed by eight cycles of 98°C for 10 s, 63°C for 30 s, and 72°C for 1 min) in a 50 µl reaction volume. To determine the optimal number of additional cycles and prevent overamplification, 5 µl of the pre-amplification mix was combined with primers, 1× Evagreen Sybr green (Jena Biosciences), and 2× Kapa HiFi Hotstart Readymix in a 15 µl total volume and subjected to 30 cycles on a StepOne Plus instrument. The remaining pre-amplified samples (45 µl) were subjected to an additional seven (NotoP sample) or six (MP sample) cycles, resulting in a total of 15 (NotoP sample) or 14 (MP sample) cycles. The libraries were purified using MinElute columns (QIAGEN), and the concentration was determined using the DNA HS Qubit assay (Life Technologies). Approximately 4 ng of each library was subjected to DNA HS Bioanalyzer chip (Agilent) to assess library size and calculate molarities. The samples were pooled and subjected to paired-end sequencing on an Illumina NovaSeq 6000 platform with 2×100 bp read length.

ChIP-seq

For histone ChIP-seq on sorted Noto^{diff} and MP^{diff} cells, the iDeal ChIP-Seq kit (Diagenode) was used according to the manufacturer's instructions. Approximately 200,000 cells were used per ChIP. ChIP on bulk d4 *in vitro*-differentiated notochord-like cells for the identification of T- and Foxa2-binding sites was performed as described previously (Koch et al., 2011). Antibodies are listed in Table S8. ChIP-Seq sequencing libraries were generated using the TrueSeq ChIP-Seq kit (Illumina) following the manufacturer's instructions with minor modifications. After adapter ligation, 0.95× of AMPure XP beads (Beckman Coulter, A63880) were used for a single purification and the DNA was eluted using 15 µl of resuspension buffer (RSB, Illumina). After the addition of 1 µl primer mix

(25 mM each; Primer 1: 5'-AATGATACGGCGACCACCGA*G-3'; Primer 2: 5'-CAAGCAGAAGACGGCATAACGA*G-3') and 15 µl 2× Kapa HiFi HotStart Ready Mix (Kapa Biosystems), amplification was performed for 45 s at 98°C; five cycles of 15 s at 98°C, 30 s at 63°C and 30 s at 72°C; and a final 1 min incubation at 72°C. The PCR products were purified using 0.95× of beads and eluted using 21 µl of RSB. Libraries were directly amplified for an additional 13 cycles and purified using AMPure XP beads. The libraries were quantified using the Qubit DNA HS assay and the library size was validated using DNA HS bioanalyzer chips (Agilent, 5067-4626). The samples were pooled and subjected to paired-end sequencing on an Illumina NextSeq 500 platform with 2×75 bp read length.

Reads were mapped to chromosomes 1-19, X, Y and M of the mouse mm10 genome using bowtie version 1.3.1 (Langmead et al., 2009), providing the options '-y -m 1 -S -I 100 -X 500'. The mapping information of the paired-end reads was used to elongate each fragment to its original size using a custom perl script, with the result stored as a BED file. Reads were then sorted and deduplicated such that only one fragment with the same starting and end position was retained. For visualization, wiggle files were generated with BEDTools version 2.23.0 (Quinlan and Hall, 2010), converted to bigwig format and analyzed in the Integrated Genome Browser (Freese et al., 2016).

Peak detection was performed using MACS version 3.0.0b1 (<https://github.com/macs3-project/MACS>) using the elongated and deduplicated bed files as inputs and setting a q-value cutoff of 0.1.

ATAC-seq data processing

For ATAC-seq mapping, adapters were detected and removed using fastq-mcf of ea-utils (<https://github.com/ExpressionAnalysis/ea-utils>, version 1.04.738 was used) and mapped to chromosomes 1-19, X, Y and M of the mouse mm10 using bowtie (Langmead, et al., 2009) version 1.3.1, with the options '-y -m 1 -S -X 2000 -allow-contain'. Mapped paired-end reads were converted to a bed file by generating the original fragment using a custom perl script and duplicates were removed such that only one fragment with the same starting and end position was retained. Owing to the repetitive nature of the Y chromosome and non-informative mitochondrial genome, reads mapped to either of them were removed and .wig files were generated using BEDTools and converted to bigwig format.

Differential ATAC-seq analysis

In order to identify regions with differential accessibilities between the NotoP and MP samples, we employed diffReps version 1.55.4 (Shen et al., 2013), using the elongated and deduplicated ATAC-NotoP and ATAC-MP samples as treatment and control, respectively. We used the default parameters, with the exception of performing the statistical analysis using a G-test ('-meth gt') and disabling the DNA fragment shifting ('-frag 0') owing to the use of already elongated reads. This resulted in 16,680 regions that displayed either a significantly higher (more open in NotoPs) or lower (more open in MPs) accessibility. We then removed all those regions overlapping known promoters (mm10 UCSC refseq genes), defined as ±5 kb from known transcription start sites, resulting in 13,890 differential regions.

Transcription factor peaks overlaps

In order to avoid inclusion of transcription factor peaks that fall within regions displaying mapping artifacts in ATAC-seq data, we removed all peaks falling within those 35 previously identified regions (Koch et al., 2017; Buenrostro et al., 2015). We then overlapped the peak regions of the T-NMPdiff, T-NotoPdiff and Foxa2-NotoPdiff datasets to obtain all seven possible combinatorial binding profiles. We then used the single bp maxima of each peak and intersected those with ATAC regions displaying a significant increase in accessibility in NotoP cells. Finally, each peak was assigned potential target genes depending on the binding location. Intergenic peaks (at least 5 kb away from any gene annotation) were assigned to both the closest up- and downstream gene, whereas genic peaks (those located between -5 kb of the promoter and +5 kb after the gene end) were assigned to that gene (or multiple in case of overlapping genes) as well as the closest up- and downstream gene.

Heatmaps and average profiles

Heatmaps and average profiles were generated using SeqMINER (Ye et al., 2011). For plotting the ATAC-seq NotoP versus MP comparison, the reference bed files were generated by using the center of differential ATAC regions and sorting them by the generated diffReps *P*-values. For plotting the T, Foxa2, H3K27ac and ATAC profiles, we used the peak maxima of T-NMPdiff, T-NotoPdiff or Foxa2-NotoPdiff (in that respective order if present) and sorted the peaks first by the seven different combinations and then randomized the peaks within each category. The corresponding elongated and deduplicated bed files were used as inputs and hence set the extension size to 0.

UpSet plot

The UpSet plot was generated using UpSetR (Conway et al., 2017) to visualize associations between peak categories and associated genes.

Note added in proof

While this work was prepared for publication, complementary work dissecting the regulatory landscape of the *T* locus with respect to notochord control elements was reported on the bioRxiv preprint website (Kemmler et al., 2023 preprint). In this work, the notochord enhancers *TNE*, *TNE2* and *TE7* were termed *T3*, *C* and *I*, respectively. The genomic fragments defining these enhancers are different from ours, but with considerable overlaps.

Acknowledgements

We are grateful to the MPIMG transgenic unit and animal facility, especially Karol Macura, Judith Fiedler, Adrian Landsberger, Mirjam Peetz, Dijana Micić and Christin Franke, for performing the morula aggregation experiments and animal husbandry. We thank Sandra Währisch and Gaby Bläß for tissue culture support, Claudia Giesecke-Thiel, Erwin Weiß and Uta Marchfelder from the MPIMG flow cytometry unit for FACS support, and all members of the MPIMG microscopy and sequencing core facilities for technical advice and library sequencing, respectively. Some of the data in this paper formed part of Dennis Schifferl's PhD thesis in the Institute of Biology, Department of Biology, Chemistry and Pharmacy of the Freie Universität Berlin (2021).

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: B.G.H.; Methodology: D.S., B.G.H.; Software: F.K.; Validation: D.S., F.K.; Formal analysis: D.S., F.K.; Investigation: D.S., M.S.-W., A.V.L., M.P., J.V.V., F.K.; Resources: L.W.; Data curation: D.S., F.K.; Writing - original draft: D.S., B.G.H.; Writing - review & editing: D.S., F.K., B.G.H.; Visualization: D.S.; Supervision: F.K., B.G.H.; Project administration: B.G.H.; Funding acquisition: B.G.H.

Funding

All funding was provided by the Max-Planck-Gesellschaft. Open Access funding provided by the Max Planck Society. Deposited in PMC for immediate release.

Data availability

Sequencing data have been deposited in Gene Expression Omnibus under accession number GSE235030.

Peer review history

The peer review history is available online at <https://journals.biologists.com/dev/lookup/doi/10.1242/dev.202111.reviewer-comments.pdf>

References

- Abdelkhalek, H. B., Beckers, A., Schuster-Gossler, K., Pavlova, M. N., Burkhardt, H., Lickert, H., Rossant, J., Reinhardt, R., Schalkwyk, L. C., Muller, I. et al. (2004). The mouse homeobox gene *Not* is required for caudal notochord development and affected by the truncate mutation. *Genes Dev.* **18**, 1725-1736. doi:10.1101/gad.303504
- Alten, L., Schuster-Gossler, K., Eichenlaub, M. P., Wittbrodt, B., Wittbrodt, J. and Gossler, A. (2012). A novel mammal-specific three partite enhancer element regulates node and notochord-specific *Noto* expression. *PLoS One* **7**, e47785. doi:10.1371/journal.pone.0047785
- Ang, S. L. and Rossant, J. (1994). HNF-3β is essential for node and notochord formation in mouse development. *Cell* **78**, 561-574. doi:10.1016/0092-8674(94)90522-3

- Brand-Saberi, B., Ebensperger, C., Wilting, J., Balling, R. and Christ, B. (1993). The ventralizing effect of the notochord on somite differentiation in chick embryos. *Anat. Embryol.* **188**, 239-245. doi:10.1007/BF00188215
- Briscoe, J. and Ericson, J. (1999). The specification of neuronal identity by graded Sonic Hedgehog signalling. *Semin. Cell Dev. Biol.* **10**, 353-362. doi:10.1006/scdb.1999.0295
- 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. doi:10.1038/nmeth.2688
- 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. doi:10.1038/nature14590
- Choi, K. S., Cohn, M. J. and Harfe, B. D. (2008). Identification of nucleus pulposus precursor cells and notochordal remnants in the mouse: implications for disk degeneration and chordoma formation. *Dev. Dyn.* **237**, 3953-3958. doi:10.1002/dvdy.21805
- Cleaver, O. and Krieg, P. A. (2001). Notochord patterning of the endoderm. *Dev. Biol.* **234**, 1-12. doi:10.1006/dbio.2001.0214
- Conway, J. R., Lex, A. and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938-2940. doi:10.1093/bioinformatics/btx364
- Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A. et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* **107**, 21931-21936. doi:10.1073/pnas.1016071107
- Di Gregorio, A. (2020). The notochord gene regulatory network in chordate evolution: conservation and divergence from Ciona to vertebrates. *Curr. Top. Dev. Biol.* **139**, 325-374. doi:10.1016/bs.ctdb.2020.01.002
- 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. doi:10.1038/nprot.2006.173
- Fan, C. M. and Tessier-Lavigne, M. (1994). Patterning of mammalian somites by surface ectoderm and notochord: evidence for sclerotome induction by a hedgehog homolog. *Cell* **79**, 1175-1186. doi:10.1016/0092-8674(94)90009-4
- Freese, N. H., Norris, D. C. and Loraine, A. E. (2016). Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* **32**, 2089-2095. doi:10.1093/bioinformatics/btw069
- George, S. H., Gertszenstein, M., Vintersten, K., Korets-Smith, E., Murphy, J., Stevens, M. E., Haigh, J. J. and Nagy, A. (2007). Developmental and adult phenotyping directly from mutant embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **104**, 4455-4460. doi:10.1073/pnas.0609277104
- Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W. and Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766-1769. doi:10.1126/science.7792603
- 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. doi:10.1371/journal.pbio.1001937
- Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A. et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311-318. doi:10.1038/ng1966
- Herrmann, B. G. (1995). The mouse Brachyury (T) gene. *Semin. Dev. Biol.* **6**, 385-394. doi:10.1016/S1044-5781(06)80002-2
- 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. doi:10.1038/343617a0
- Howe, E. A., Sinha, R., Schlauch, D. and Quackenbush, J. (2011). RNA-Seq analysis in MeV. *Bioinformatics* **27**, 3209-3210. doi:10.1093/bioinformatics/btr490
- Jeong, Y. and Epstein, D. J. (2003). Distinct regulators of Shh transcription in the floor plate and notochord indicate separate origins for these tissues in the mouse node. *Development* **130**, 3891-3902. doi:10.1242/dev.00590
- Kemmler, C. L., Smolikova, J., Moran, H. R., Mannion, B. J., Knapp, D., Lim, F., Czarkwiani, A., Aguayo, V. H., Rapp, V., Fitch, O. E. et al. (2023). Conserved enhancers control notochord expression of vertebrate Brachyury. *Nat. Commun.* **14**, 6594. doi:10.1038/s41467-023-42151-3
- 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. doi:10.1186/gb-2013-14-4-r36
- Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T. K., Zacarias-Cabeza, J., Spicuglia, S., De La Chapelle, A. L., Heidemann, M., Hintermaier, C. et al. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.* **18**, 956-963. doi:10.1038/nsmb.2085
- Koch, F., Scholze, M., Wittler, L., Schifferl, D., Sudheer, S., Grote, P., Timmermann, B., Macura, K. and Herrmann, B. G. (2017). Antagonistic activities of Sox2 and brachyury control the fate choice of neuro-mesodermal progenitors. *Dev. Cell* **42**, 514-526.e7. doi:10.1016/j.devcel.2017.07.021
- 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. doi:10.1186/gb-2009-10-3-r25
- Matsumoto, J., Kumano, G. and Nishida, H. (2007). Direct activation by Ets and Zic is required for initial expression of the Brachyury gene in the ascidian notochord. *Dev. Biol.* **306**, 870-882. doi:10.1016/j.ydbio.2007.03.034
- McCann, M. R., Tamplin, O. J., Rossant, J. and Seguin, C. A. (2012). Tracing notochord-derived cells using a Noto-cre mouse: implications for intervertebral disc development. *Dis. Model. Mech.* **5**, 73-82. doi:10.1242/dmm.008128
- McMahon, J. A., Takada, S., Zimmerman, L. B., Fan, C. M., Harland, R. M. and McMahon, A. P. (1998). Noggin-mediated antagonism of BMP signaling is required for growth and patterning of the neural tube and somite. *Genes Dev.* **12**, 1438-1452. doi:10.1101/gad.12.10.1438
- Muyers, J. P. 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. doi:10.1093/nar/27.6.1555
- Nishizaki, Y., Shimazu, K., Kondoh, H. and Sasaki, H. (2001). Identification of essential sequence motifs in the node/notochord enhancer of Foxa2 (Hnf3 β) gene that are conserved across vertebrate species. *Mech. Dev.* **102**, 57-66. doi:10.1016/S0925-4773(01)00281-7
- Passamanek, Y. J., Katikala, L., Perrone, L., Dunn, M. P., Oda-Ishii, I. and Di Gregorio, A. (2009). Direct activation of a notochord cis-regulatory module by Brachyury and FoxA in the ascidian *Ciona intestinalis*. *Development* **136**, 3679-3689. doi:10.1242/dev.038141
- Peck, S. H., Mckee, K. K., Tobias, J. W., Malhotra, N. R., Harfe, B. D. and Smith, L. J. (2017). Whole Transcriptome Analysis of Notochord-Derived Cells during Embryonic Formation of the Nucleus Pulposus. *Sci. Rep.* **7**, 10504. doi:10.1038/s41598-017-0692-5
- Perez-Silva, J. G., Araujo-Voces, M. and Quesada, V. (2018). nVenn: generalized, quasi-proportional Venn and Euler diagrams. *Bioinformatics* **34**, 2322-2324. doi:10.1093/bioinformatics/bty109
- Plouhinec, J. L., Granier, C., Le Mentec, C., Lawson, K. A., Saberan-Djoneidi, D., Aghion, J., Shi, D. L., Collignon, J. and Mazan, S. (2004). Identification of the mammalian Not gene via a phylogenomic approach. *Gene Expr. Patterns* **5**, 11-22. doi:10.1016/j.modgep.2004.06.010
- Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842. doi:10.1093/bioinformatics/btq033
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A. and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-283. doi:10.1038/nature09692
- Schifferl, D., Scholze-Wittler, M., Wittler, L., Veenvliet, J. V., Koch, F. and Herrmann, B. G. (2021). A 37 kb region upstream of brachyury comprising a notochord enhancer is essential for notochord and tail development. *Development* **148**, dev200059. doi:10.1242/dev.200059
- Shen, L., Shao, N. Y., Liu, X., Maze, I., Feng, J. and Nestler, E. J. (2013). diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. *PLoS One* **8**, e65598. doi:10.1371/journal.pone.0065598
- Shiratori, H. and Hamada, H. (2006). The left-right axis in the mouse: from origin to morphology. *Development* **133**, 2095-2104. doi:10.1242/dev.02384
- Song, B. P., Ragsac, M. F., Tellez, K., Jindal, G. A., Grudzien, J. L., Le, S. H. and Farley, E. K. (2023). Diverse logics and grammar encode notochord enhancers. *Cell Rep.* **42**, 112052. doi:10.1016/j.celrep.2023.112052
- Stemple, D. L. (2005). Structure and function of the notochord: an essential organ for chordate development. *Development* **132**, 2503-2512. doi:10.1242/dev.01812
- Stott, D., Kispert, A. and Herrmann, B. G. (1993). Rescue of the tail defect of Brachyury mice. *Genes Dev.* **7**, 197-203. doi:10.1101/gad.7.2.197
- Sulik, K., Dehart, D. B., Langaki, T., Carson, J. L., Vrablic, T., Gesteland, K. and Schoenwolf, G. C. (1994). Morphogenesis of the murine node and notochordal plate. *Dev. Dyn.* **201**, 260-278. doi:10.1002/aja.1002010309
- Tam, P. P. and Beddington, R. S. (1987). The formation of mesodermal tissues in the mouse embryo during gastrulation and early organogenesis. *Development* **99**, 109-126. doi:10.1242/dev.99.1.109
- Tamplin, O. J., Cox, B. J. and Rossant, J. (2011). Integrated microarray and ChIP analysis identifies multiple Foxa2 dependent target genes in the notochord. *Dev. Biol.* **360**, 415-425. doi:10.1016/j.ydbio.2011.10.002
- Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L. and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46-53. doi:10.1038/nbt.2450
- 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. doi:10.1016/j.mod.2009.08.003
- Vidigal, J. A., Morkel, M., Wittler, L., Brouwer-Lehmitz, A., Grote, P., Macura, K. and Herrmann, B. G. (2010). An inducible RNA interference system for the

- functional dissection of mouse embryogenesis. *Nucleic Acids Res.* **38**, e122. doi:10.1093/nar/gkq199
- Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F. et al.** (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854-858. doi:10.1038/nature07730
- Wallis, J. W., Miller, T. R., Lerner, C. A. and Kleerup, E. C.** (1989). Three-dimensional display in nuclear medicine. *IEEE Trans. Med. Imaging* **8**, 297-230. doi:10.1109/42.41482
- Ward, L., Pang, A. S. W., Evans, S. E. and Stern, C. D.** (2018). The role of the notochord in amniote vertebral column segmentation. *Dev. Biol.* **439**, 3-18. doi:10.1016/j.ydbio.2018.04.005
- Weinstein, D. C., Ruiz i Altaba, A., Chen, W. S., Hoodless, P., Prezioso, V. R., Jessell, T. M. and Darnell, J. E. Jr.** (1994). The winged-helix transcription factor HNF-3 β is required for notochord development in the mouse embryo. *Cell* **78**, 575-588. doi:10.1016/0092-8674(94)90523-1
- Wymeersch, F. J., Skylaki, S., Huang, Y., Watson, J. A., Economou, C., Marek-Johnston, C., Tomlinson, S. R. and Wilson, V.** (2019). Transcriptionally dynamic progenitor populations organised around a stable niche drive axial patterning. *Development* **146**, dev168161. doi:10.1242/dev.168161
- Yamanaka, Y., Tamplin, O. J., Beckers, A., Gossler, A. and Rossant, J.** (2007). Live imaging and genetic analysis of mouse notochord formation reveals regional morphogenetic mechanisms. *Dev. Cell* **13**, 884-896. doi:10.1016/j.devcel.2007.10.016
- 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. doi:10.1093/nar/gkq1287
- Zizic Mitrecic, M., Mitrecic, D., Pochet, R., Kostovic-Knezevic, L. and Gajovic, S.** (2010). The mouse gene *Noto* is expressed in the tail bud and essential for its morphogenesis. *Cells Tissues Organs* **192**, 85-92. doi:10.1159/000291015

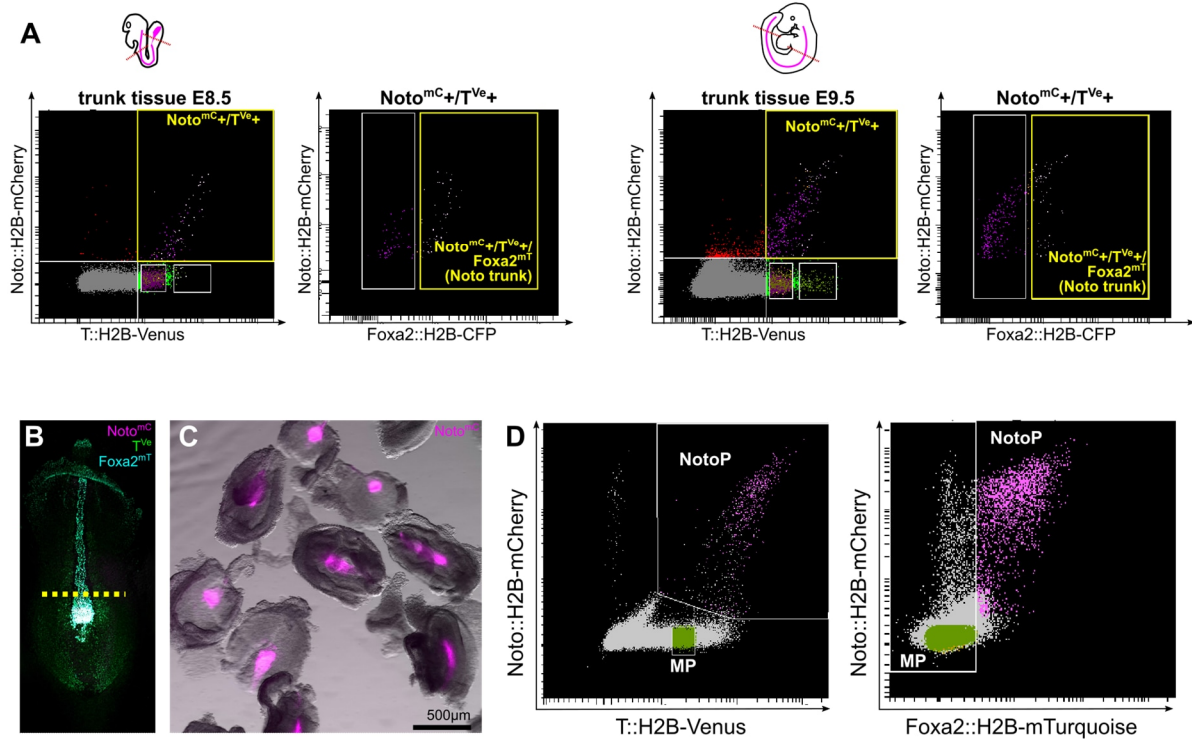


Fig. S1. Generation of ATAC-seq data from subpopulations from E8.5 caudal ends.(A) FACS profile with subpopulations based on $\text{Noto}^{\text{mC}/\text{T}^{\text{Ve}}}/\text{Foxa2}^{\text{mT}}$ reporter expression used for sorting of subpopulations for transcriptome analysis of trunk notochord (Noto trunk). **(B-D)** Workflow for sample preparation and FACS for ATAC-seq. **(B)** Maximum intensity projection of a E8.5 embryo generated from the $\text{Noto}^{\text{mC}/\text{T}^{\text{Ve}}}/\text{Foxa2}^{\text{mT}}$ mESC line. Caudal ends were dissected from E8.5 embryos at the somite border indicated by the yellow dotted line. **(C)** Isolated caudal ends used for the experiment. **(D)** FACS profile with subpopulations based on $\text{Noto}^{\text{mC}/\text{T}^{\text{Ve}}}/\text{Foxa2}^{\text{mT}}$ reporter expression.

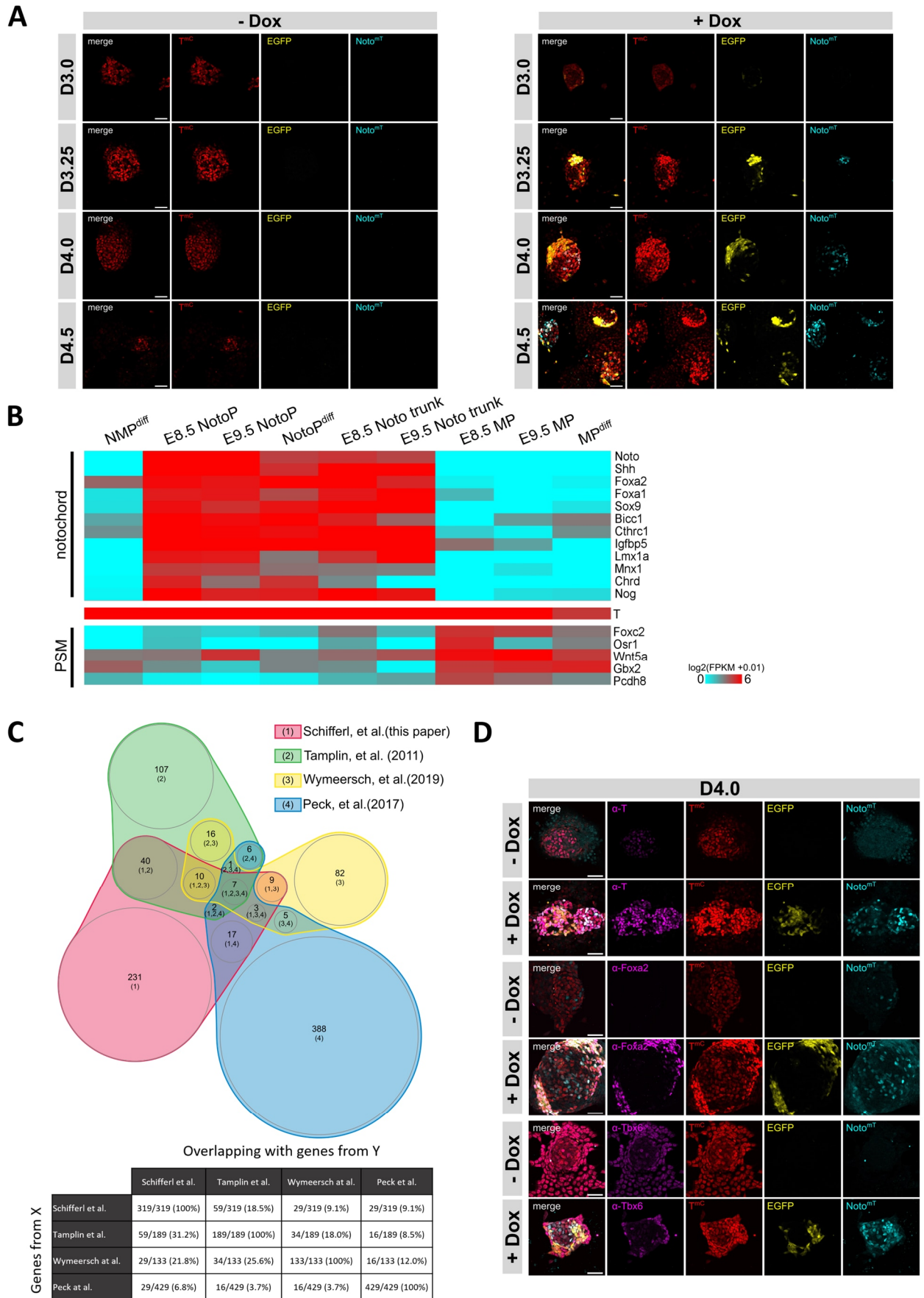


Fig. S2. Validation of *in vitro* derived NotoPs.

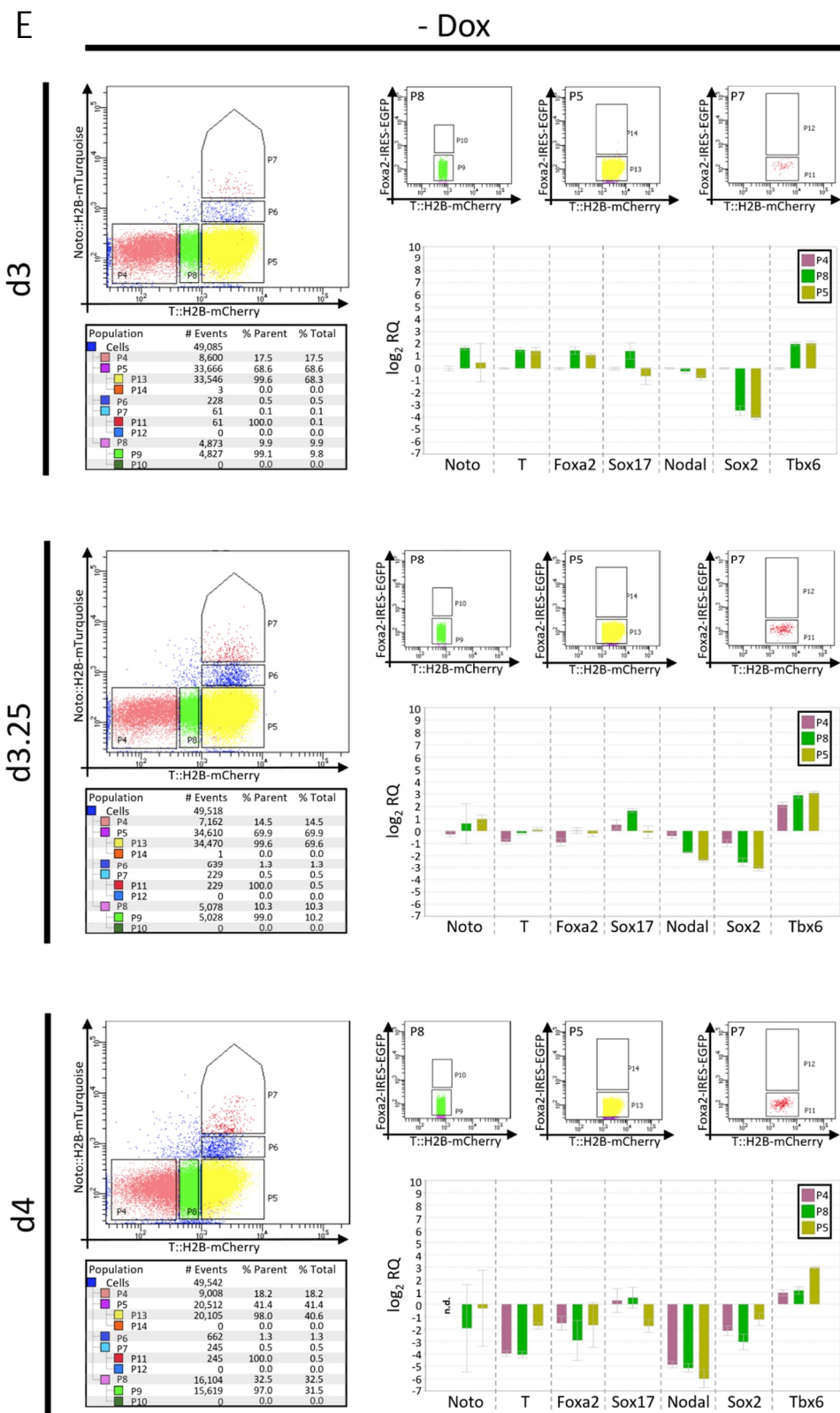


Fig. S2. Validation of *in vitro* derived NotoPs.

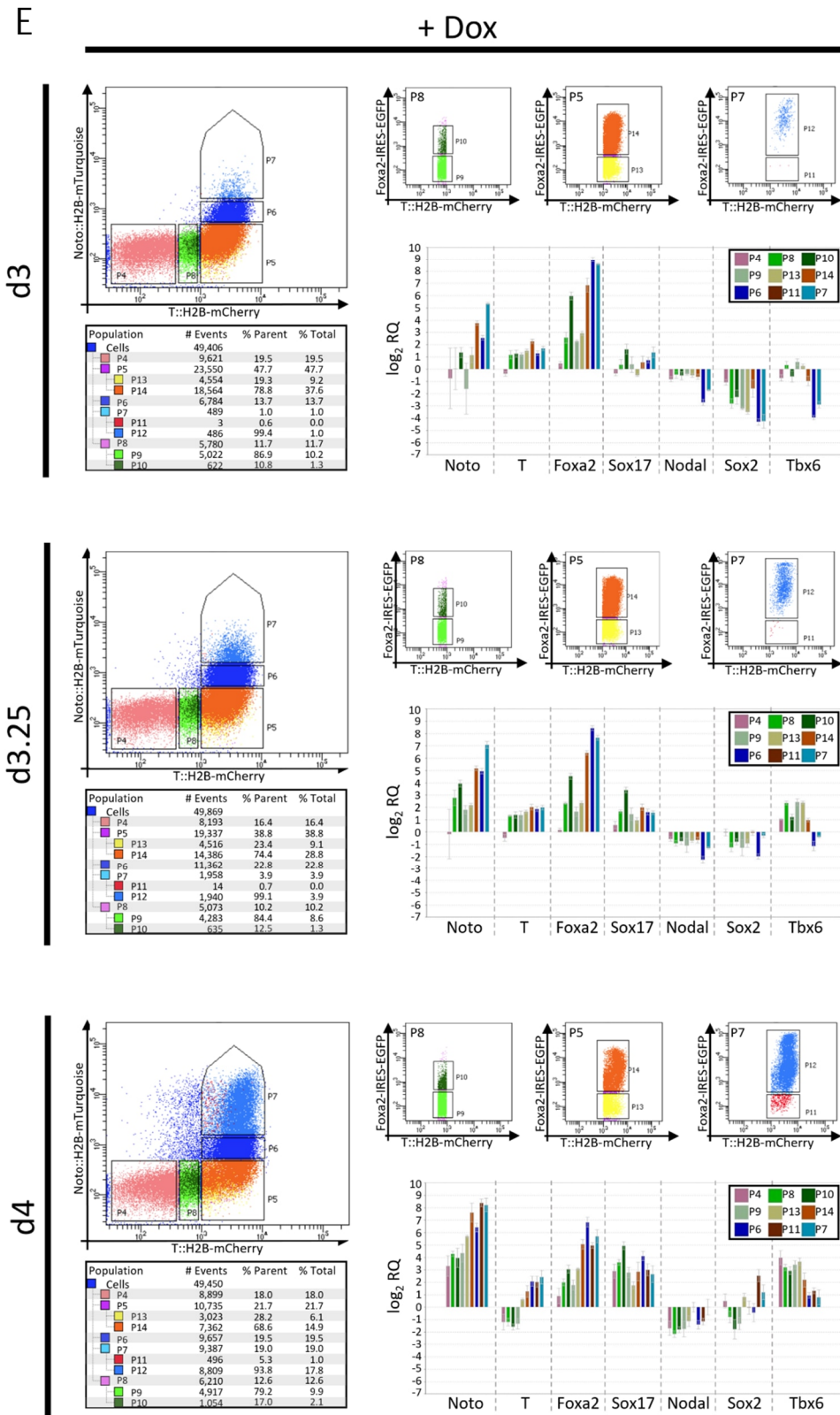


Fig. S2. Validation of *in vitro* derived NotoPs.

Fig. S2. Validation of *in vitro* derived NotoPs.

(A) Images of differentiated colonies acquired by confocal microscopy without and with doxycycline (Dox) induced *Foxa2* expression. Scale bar = 50 μm . **(B)** Log transformed FPKM values of selected marker genes across analyzed samples. **(C)** Top: Venn diagram showing the intersections between the NotoP gene set determined in this study and gene sets from previously published studies. Tamplin et al (2011): 189 genes enriched in E8.5 Noto-GFP+ vs. Noto-GFP- cells (microarray). Wymeersch et al (2019): 133 genes enriched in E8.5 node streak border and rostral node vs. remaining samples (microarray). Peck et al (2017): 429 genes enriched in E12.5 Shh-Cre+ notochord vs. P0 Shh-Cre+ nucleus pulposus (RNA-seq). Bottom: Table showing the pairwise overlap of datasets. The genes in the individual groups are listed in Supplementary Table S2. **(D)** Images of differentiated colonies with immunofluorescence for T, *Foxa2* or *Tbx6* acquired by confocal microscopy. Scale bar = 50 μm . **(E)** Time course FACS profiles and qPCR for marker genes on cDNA isolated from sorted subpopulations during differentiation with and without Dox-induction. NMP^{diff}, MP^{diff} or NotoP^{diff} cells used for ChIP-seq/RNA-seq experiments are derived from cell populations P5 d3.0 -Dox, P5 d4.0 -Dox, or P7 d4.0 +Dox, respectively. The NotoP^{diff} cell number of P7 (+Dox) is increasing from d3 to d4 and characterized by increasing Noto expression as well as high T and *Foxa2* expression. The NMP^{diff} cell population P5 (d3, -Dox) is decreasing in number and differentiating to MP^{diff} (P5, -Dox) on d4, characterized by increasing *Tbx6* and parallel decreasing T expression. *Nodal* expression decreases in all cell populations, however reaching initial levels again in P7 at d4. *Sox17* expression is highest in P10 (d4, +Dox), which might represent a small subpopulation of endodermal cells (2.1% of total). All qPCR data were calculated relative to the expression level of the respective gene in T^{Ve} negative cells (d3, -Dox). Error bars indicate the 95% confidence interval of 2-3 technical replicates.

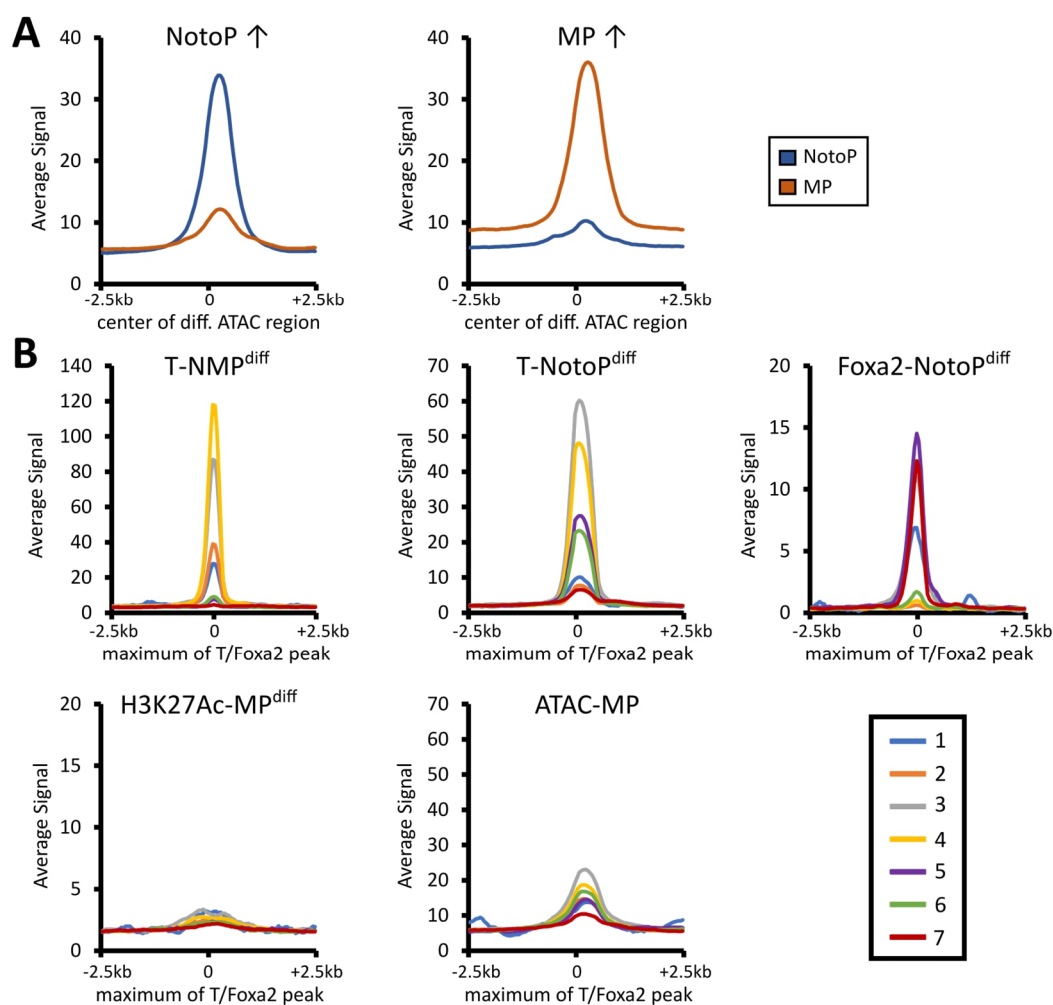


Fig. S3. Average Profiles genome-wide ATAC- and ChIP-seq datasets.

(A) ATAC-seq signal in NotoP (blue) and MP (orange) cells around differential ATAC regions displaying higher accessibility in NotoP (left) or MP (right) cells. (B) Remaining average profiles of ATAC- and ChIP-seq datasets for the different enhancer categories.

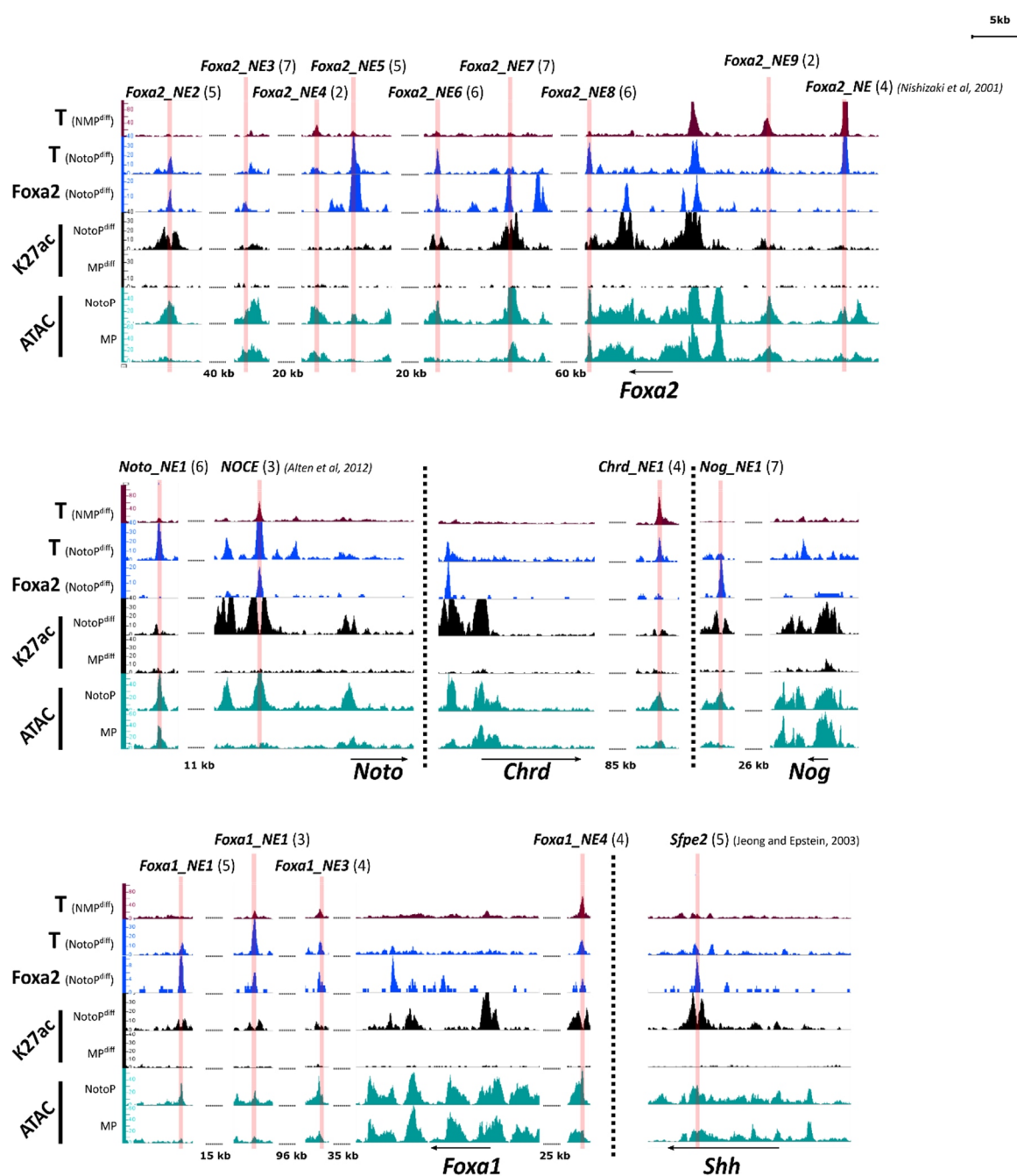


Fig. S4. Genome browser screenshots of identified enhancers at important notochord genes.

Genome browser snapshots showing loci of notochord genes with tracks for TF ChIP-seq, H3K27ac ChIP-seq and ATAC-Seq. Track maxima are normalized to the number of mapped reads for each antibody and ATAC experiment. Elements identified in this study are highlighted in red. The number in brackets indicates the enhancer category as shown in Fig.2B.

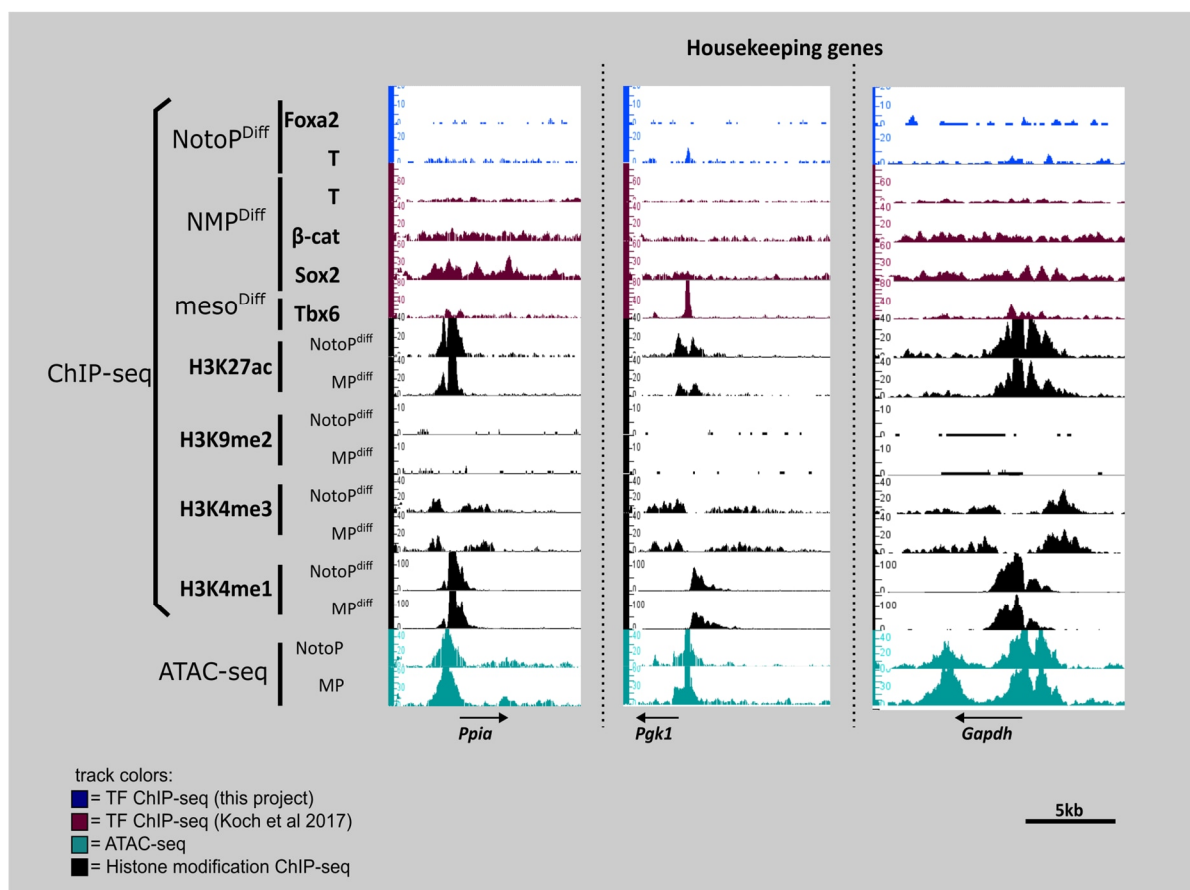


Fig. S5. Genome browser screenshots of exemplary housekeeping loci.

Genome browser snapshots showing loci of housekeeping genes with tracks for TF ChIP-seq, Histone 3 ChIP-seq and ATAC-seq. Track maxima are normalized to the number of mapped reads for each antibody and ATAC experiment. The same normalization was used in Figures 3,4, S4 and S5.

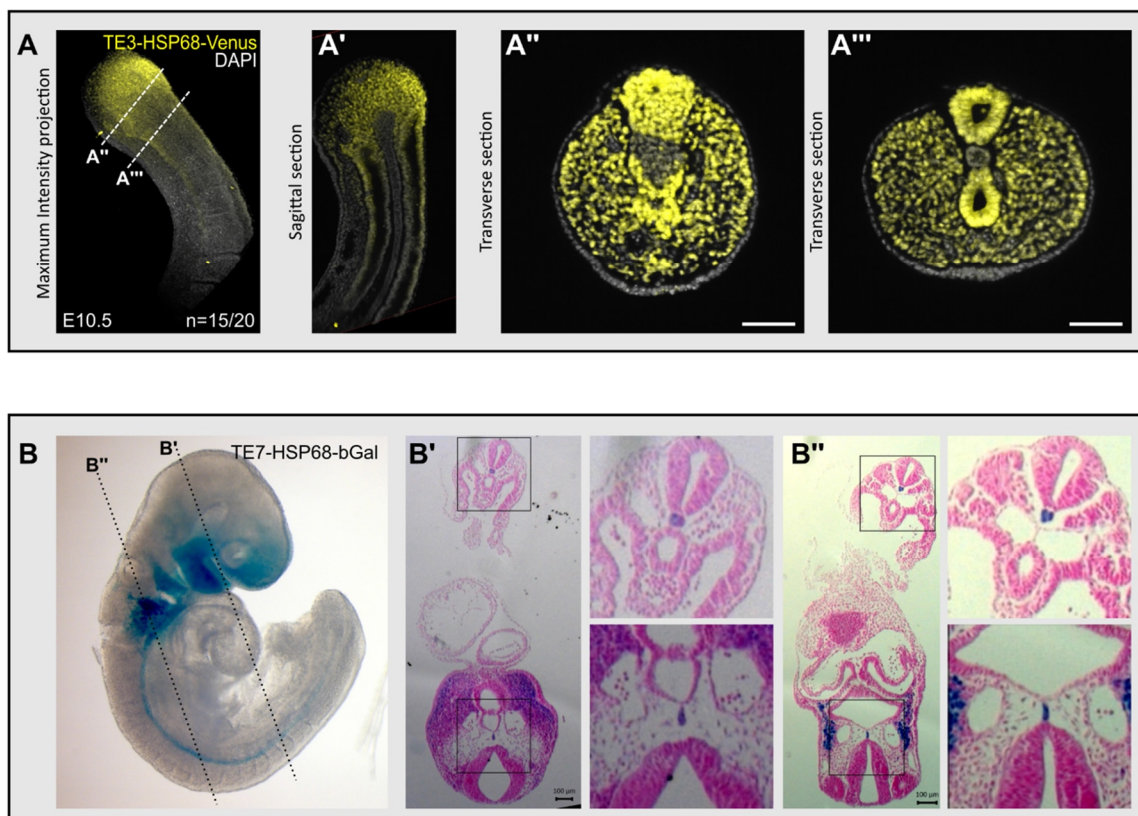


Fig. S6. Activity assays of enhancer elements in the murine *T* locus.

(A) Light sheet micrograph of *TE3*-driven Venus reporter expression in the tail bud of a E10.5 embryo. Nuclei stained with DAPI (grey). Left: maximum intensity projection. **(A')** Sagittal optical midline section showing activity in paraxial mesoderm, posterior neural tube and hindgut pocket. **(A''-A''')** Transverse optical sections acquired by light sheet microscopy at the axial levels indicated in (A). Scale bar = 100 μ m. **(B)** *TE7*-driven galactosidase activity visualized in a E9.25 embryo. Dotted lines indicate the axial position of transverse histological sections with eosin staining shown in **(B'-B'')**.

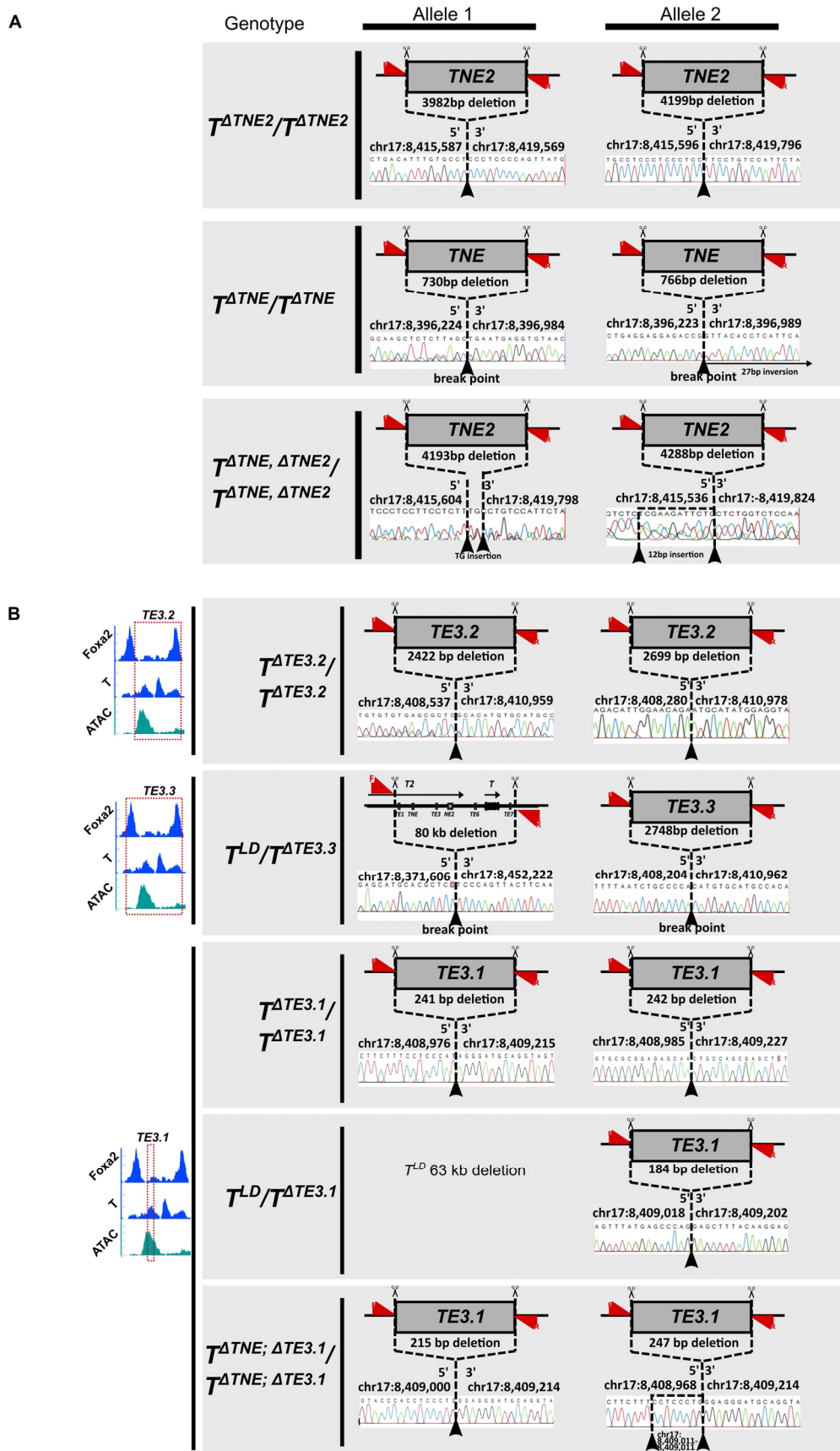


Fig. S7. Genotyping of Deletions

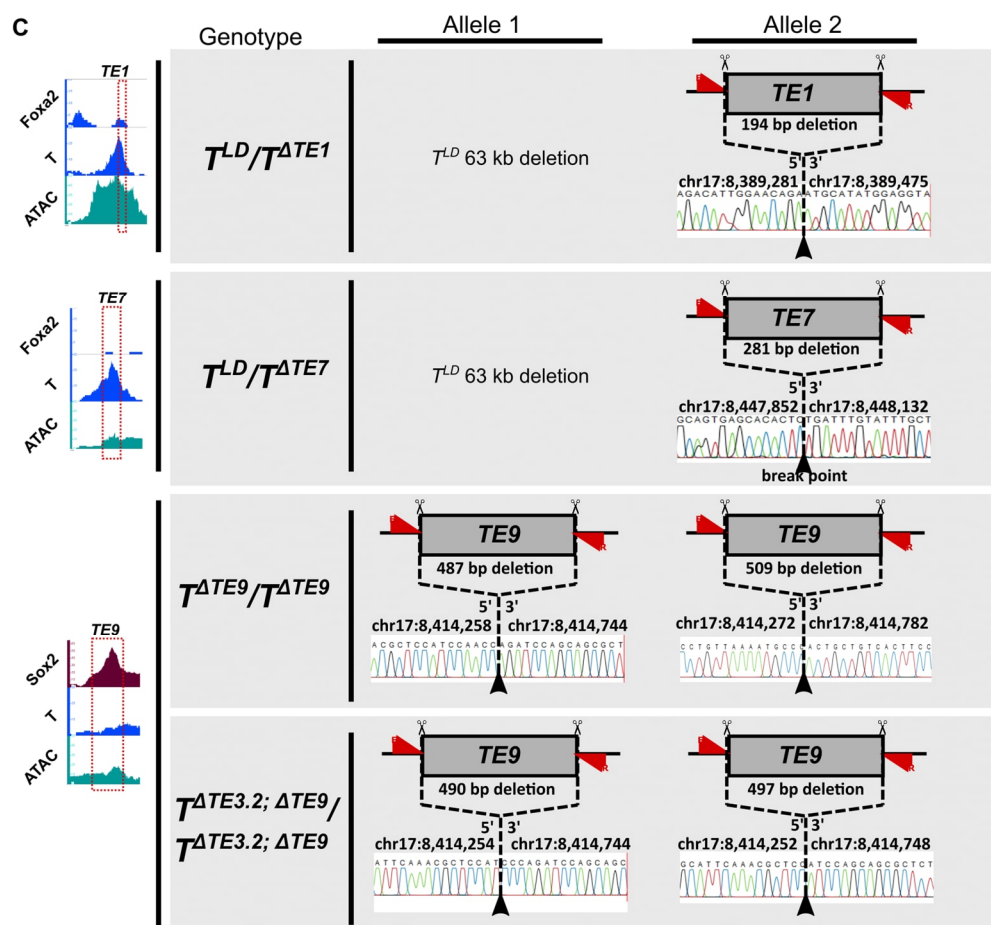


Fig. S7. Genotyping of Deletions

Red arrows indicate the position of the forward (F) and reverse (R) primers used for genotyping by PCR, not to scale. Scissors icons indicate the approximate positions of gRNA target sites. The dotted lines converging at the break points mark the CRISPR/Cas9 mediated deletions. The coordinates (mm10) of the bordering 5' and 3' edges of the respective deletions are specified next to the break point(s). Sanger sequencing tracks of enhancer mutant PCR fragment or subcloned PCR fragments in case of double deletions show the region flanking the deletions. For each genotype at least two clones were generated and checked for phenotypic identity. **(A)** TNE and TNE2 mutants shown in Main Figure 4. **(B)** TE3 mutants. **(C)** TE1, TE7 and TE9 mutants. The T^{LD} mutant was reported previously (Schifferl et al, 2021).

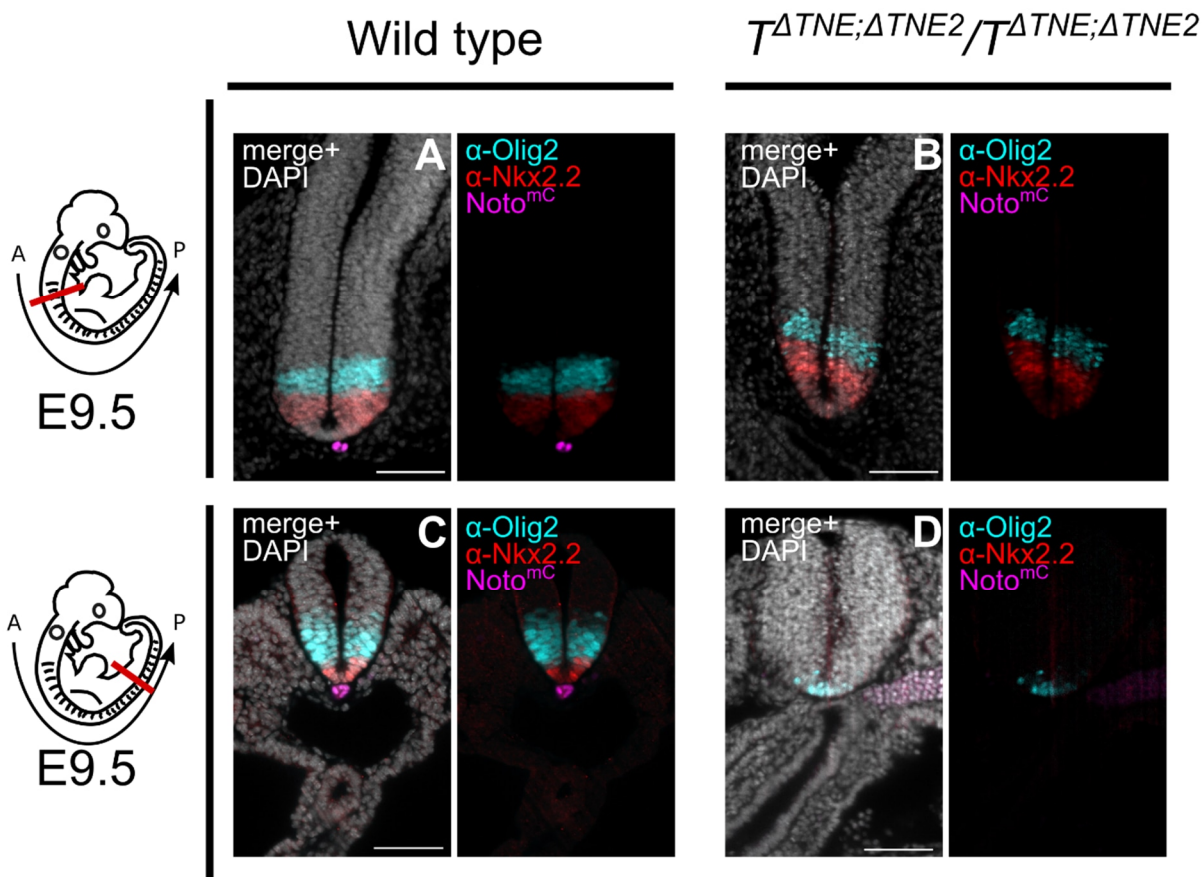


Fig. S8. Loss of trunk notochord in *TNE*; *TNE2* double enhancer mutants leads to neural tube patterning defects.

Light sheet micrographs of wild type and mutant embryos with immunofluorescence for Olig2 (blue) and Nkx2.2 (red). Transversal optical sections at different axial positions at E9.75 as indicated in the schematic on the left. **(A-B)**: Cervical. **(C-D)**: Lumbar. Scale bar = 100 μ m.

Table S1. Relative Expression values (FPKM) for RNA-seq samples.

Available for download at

<https://journals.biologists.com/dev/article-lookup/doi/10.1242/dev.202111#supplementary-data>

Table S2. Intersections of notochord gene sets shown in Figure S2C.

Available for download at

<https://journals.biologists.com/dev/article-lookup/doi/10.1242/dev.202111#supplementary-data>

Table S3. List of the 3728 candidate enhancers identified by differential ATAC-seq and TF binding analysis.

Available for download at

<https://journals.biologists.com/dev/article-lookup/doi/10.1242/dev.202111#supplementary-data>

Table S4. List of the 319 notochord genes with and number of associated enhancer categories.

Available for download at

<https://journals.biologists.com/dev/article-lookup/doi/10.1242/dev.202111#supplementary-data>

Table S5. Cell lines used in this work

modification	parental clone	construct	transfection method	selection	purpose	published
<i>Noto^{mC}</i>	F1G4 (George et al., 2007)	Noto::H2B-mCherry-pA-FRT-PGK-hygro-pA-FRT BAC	electroporation	hygromycin (150µg/ml)	wild type reference for mutant analysis	Schifferl et al 2021
<i>Noto^{mC}/T^{Ve}/Foxa2^{mT}</i>	<i>Noto^{mC}</i>	T::H2B-Venus-pA-FRT-PGK-neo-pA-FRT BAC Foxa2::H2B-mTurquoise-pA-FRT-PGK-puro-pA-FRT BAC	electroporation	G418 (250µg/ml), puromycin (1µg/ml)	Derivation of subpopulations for RNA-seq	this study
$T^{\Delta TNE2}/T^{\Delta TNE2}$	<i>Noto^{mC}</i>	px459-NE2-g1 (transient) px459-NE2-g2 (transient)	lipofectamine	puromycin (d1-2 2µg/ml; d3 1µg/ml)	generation of enhancer mutant alleles	this study
$T^{\Delta TNE}/T^{\Delta TNE}$	<i>Noto^{mC}</i>	px459-TNE-g1 (transient) px459-TNE-g2 (transient)	lipofectamine	puromycin (d1-2 2µg/ml; d3 1µg/ml)	generation of enhancer mutant alleles	Schifferl et al 2021
$T^{\Delta TNE; \Delta TNE2}/T^{\Delta TNE; \Delta TNE2}$	$T^{\Delta TNE}/T^{\Delta TNE}$	px459-NE2-g1 (transient) px459-NE2-g2 (transient)	lipofectamine	puromycin (d1-2 2µg/ml; d3 1µg/ml)	generation of enhancer mutant alleles	this study
<i>R26^{RMC}</i> (Vidigal et al 2010)	F1G4 (George et al., 2007)	loxP-PGK-Hygro-pA-loxP-neo-pA	electroporation	hygromycin (150µg/ml)	parental	Vidigal et al 2010
<i>TRE::Foxa2</i>	<i>R26^{RMC}</i>	pDonor-loxP-TRE-Foxa2-IRES-EGFP-pA-PGK-loxP PGK-Cre	lipofectamine	G418 (350µg/ml)	Foxa2 OE line	this study
<i>TRE::Foxa2/T^{mC-2A-rtTA}</i>	<i>TRE::Foxa2</i>	T::H2B-mCherry-T2A-irTA-pA-FRT-PGK-puro-pA-FRT BAC	electroporation	puromycin (1µg/ml)	driver for inducible Foxa2 OE	this study
<i>TRE::Foxa2/T^{mC-2A-rtTA}/Noto^{mT}</i>	<i>TRE::Foxa2/T^{mC-2A-rtTA}</i>	Noto::H2B-mTurquoise-pA-FRT-PGK-Bla-pA-FRT BAC	electroporation	blasticidin (5µg/ml)	purification of NotoPs	this study
<i>TNE2-HSP68-Venus</i>	<i>R26^{RMC}</i>	pDonor-loxP-NE2-HSP68-Venus-pA-PGK-loxP; PGK-Cre (transient)	lipofectamine	G418 (350µg/ml)	enhancer activity assay	this study
<i>TE3-HSP68-Venus</i>	<i>R26^{RMC}</i>	pDonor-loxP-TE3-HSP68-Venus-pA-PGK-loxP; PGK-Cre (transient)	lipofectamine	G418 (350µg/ml)	enhancer activity assay	this study
<i>TE7-HSP68-bGal</i>	<i>R26^{RMC}</i>	pDonor-loxP-TE7-HSP68-bGal-pA-PGK-loxP; PGK-Cre (transient)	lipofectamine	G418 (350µg/ml)	enhancer activity assay	this study

Table S6. Mutant phenotypes of enhancer knockout lines shown in Fig. S7-8

genotype	deletion (mm10)	size	E9.5 notochord phenotype	E11-12.5 tail outgrowth	embryos analyzed
$T^{\Delta TE3.2}/T^{\Delta TE3.2}$	chr17:8,408,537-8,410,959	2422	normal	yes	16 (E11.5)
	chr17:8,408,280-8,410,978	2699			
$T^{LD}/T^{\Delta TE3.3}$	chr17:8,371,606-8,452,222	80617	like parental clone (T^{LD})	yes	10 (E11.5)
	chr17:8,408,204-8,410,962	2748			
$T^{\Delta TE3.1}/T^{\Delta TE3.1}$	chr17:8,408,976-8,409,216	241	normal	yes	8 (E10.5) 4 (E12.5) 4 (E12.5)
	chr17:8,408,985-8,409,227	242			
$T^{LD}/T^{\Delta TE3.1}$	chr17:8,389,017-8,451,946	63kb	like parental clone (T^{LD})	yes	10 (E9.5) 4 (E11.5) 5 (E9.5) 4 (E12.5)
	chr17:8,409,018-8,409,204	184			
$T^{\Delta TNE; \Delta TE3.1}/T^{\Delta TNE; \Delta TE3.1}$	chr17:8,409,000-8,409,214	215	like parental clone (T^{LD})	yes	11 (E9.5) 12 (E11.5)
	chr17:8,408,968-8,409,214	247			
$T^{LD}/T^{\Delta TE1}$	chr17:8,389,017-8,451,946	63kb	like parental clone (T^{LD})	yes	7 (E9.5) 3 (E10.5)
	chr17:8,389,281-8,389,475	194			
$T^{LD}/T^{\Delta TE7}$	chr17:8,389,017-8,451,946	63kb	like parental clone (T^{LD})	yes	11 (E8.5) 10 (E9.5) 5 (E10.5) 1 (E12.5)
	chr17:8,389,017-8,451,946	281			
$T^{\Delta TE9}/T^{\Delta TE9}$	chr17:8,414,258-8,414,744	487	normal	yes	25 (E11.5)
	chr17:8,414,273-8,414,782	509			
$T^{\Delta TE3; \Delta TE9}/T^{\Delta TE3; \Delta TE9}$	chr17:8,414,254-8,414,744	490	normal	yes	9 (E9.5)
	chr17:8,414,252-8,414,749	497			

Table S7. List of oligonucleotides used for cloning and PCR

Name	5'-3' Sequence	Purpose
TNE2_fw	GTGGCCTGTCAGCTTTGTC	<i>T</i> ^{ΔTNE2} CRISPR genotyping
TNE2_R	CATGTGTGTATCTGTGCG	<i>T</i> ^{ΔTNE2} CRISPR genotyping
TNE2_g1_top	CACCGCCATCTCCATTTCGGGAAG	<i>T</i> ^{ΔTNE2} CRISPR genotyping
TNE2_g1_bot	AAACCTTCCGGAATGGGAGATGGC	<i>T</i> ^{ΔTNE2} CRISPR genotyping
TNE2_g2_top	caccgTCTAGAATGGACAGGAACGC	<i>T</i> ^{ΔTNE2} CRISPR genotyping
TNE2_g2_bot	aaacGCGTTCCTGTCCATTCTAGAc	<i>T</i> ^{ΔTNE2} CRISPR genotyping
TNE2_fw_BamHI	TTTTTGGATCCGTGGCCTGTCAGCTT TGTC	Enhancer reporter cloning
TE3_fw_BamHI	TTTTTGGATCCCGCGTCCTCAGCCTT TAC	Enhancer reporter cloning
TE3_rev_BamHI	TTTTTCCTAGGGGTTATCAGCCCTCC TCCTG	Enhancer reporter cloning
TE7_BamHI_fw_new	TTTTTGGATCCCACTCTGTCATGTGG CCC	Enhancer reporter cloning
TE7_BamHI_rv_new	TTTTTCCTAGGCAAATCACCGTGCAA GCC	Enhancer reporter cloning
T_5_homology_H2B_F	TGTTGGGTAGGGAGTCAAGACTCCT GGAAGGTGGAGAGGGTGGCGGGAG GATGCCAGAGCCAGCGAAGTC	BAC reporter recombineering
T_3_homolgy_FRT_R	TCCACTCGGTAAGTGCAGGCTCTTCC CTGCGCTCTCTGTGCCCGGCGAGCT ATTATGTACTGACTGATGAAGTTCC	BAC reporter recombineering
T_geno_F	GAAGGTGGCTGTTGGGTAGG	BAC reporter genotyping
T_geno_R	TCGCAGTTCGCGTTCGGTGG	BAC reporter genotyping
H2B-Venus_R	GGAATAGCTCTCCTTGCGGC	BAC reporter genotyping
Noto- 5_homology_H2B_F	CTCCATTGAGCTCCTTGCACAGCCT GGGAGGTCCCCTCAGGGTTCGCGCA ATGCCAGAACCAGCCAAATC	BAC reporter genotyping
Noto_3_homolgy_FRT_ R	GGGCGCAGGCTCCCGGGCTGGACC TGAGTGCCTGAGGGAGCAGGGCTG GATTCCCAGTCACGACGTTGTA	BAC reporter recombineering
Noto_geno_F	GGCCTCAATCAGCGATGATTAAG	BAC reporter genotyping
Noto_geno_R	CTGGACCTGAGTGCCTGAG	BAC reporter genotyping
Pmm2_F	CCCCTTTCTGAAGCACTCTG	qPCR
Pmm2_R	TAAGGCGTCATTTCCCAAAG	qPCR
Noto_F	ATGCACTCACCACCAGCAG	qPCR
Noto_R	CAGCTGCTGCAAGTTAAACG	qPCR
T_F	AACTGGTCTAGCCTCGGAGT	qPCR
T_R	CTCACAGACCAGAGACTGGG	qPCR
Foxa2_F	CGAGCACCATTACGCCCTTCAAC	qPCR
Foxa2_R	AGTGCATGACCTGTTTCGTAGGC	qPCR
Sox17_F	AGCCATTTCTCCGTGGTGT	qPCR
Sox17_R	AACACTGCTTCTGGCCCTCAG	qPCR
Nodal_F	GGTGGACTTCAACCTGATTGGC	qPCR
Nodal_R	GGTTGGTATCGTTTCAGCAGGC	qPCR
Sox2_F	TACAGCATGATGCAGGAGCAG	qPCR
Sox2_R	TCATGTAGGTCTGCGAGCTG	qPCR
Tbx6_F	CCTGAGCTTGGAGAACCAGG	qPCR
Tbx6_R	GGCCAGTGAAGTACTCGG	qPCR
gDNA_control_F	CCCCTTTCTGAAGCACTCTG	qPCR
gDNA_control_R	TAAGGCGTCATTTCCCAAAG	qPCR

Table S8. Antibodies used for Immunofluorescence and ChIP References and details on antibody validation are listed on the manufacturer's websites.

Description	Catalog number	Company	Host Organism	Concentration	Application
α -T	#81694	Cell Signaling Technology	Rabbit	1:250 in PBSTB	IF
α -Sox2	AF2018	R&D	Goat	1:250 in PBSTB	IF
α -Foxa2	sc-6554	Santa Cruz	Goat	1:250 in PBSTB	IF
α -Olig2	AF2418	R&D	Goat	Rabbit	IF
α -Nkx2.2	ab191077	Abcam	Rabbit	Rabbit	IF
Alexa Fluor 488 α -Rabbit	ab150073	Abcam	Donkey	1:250 in PBSTB	IF
Alexa Fluor 647 α -Goat	ab150135	Abcam	Donkey	1:250 in PBSTB	IF
α -T	AF2085	R&D	Goat	2.5 μ g reaction	ChIP
α -Foxa2	C15410343	Diagenode	Rabbit	2.5 μ g/reaction	ChIP
H3K4me1	pAb-194-050	Diagenode	Rabbit	1.5 μ g/reaction	ChIP
H3K4me3	ab8580	abcam	Rabbit	1.5 μ g/reaction	ChIP
H3K27ac	ab4729	abcam	Rabbit	1.5 μ g/reaction	ChIP
H3K27me3	C15410069	Diagenode	Rabbit	1.5 μ g/reaction	ChIP
H3k9me2	A90-0042	Diagenode	Rabbit	1.5 μ g/reaction	ChIP

Table S9. RNA-Seq experiments

sample	embryos	cells	de-duplicated mapped reads
NotoP Noto^{mC}+/T^{Ve}+/<i>Foxa2</i>^{mT}+ E8.5 9-11S	5	250	3343606
MP T^{Ve}_high/Noto^{mC}-/<i>Foxa2</i>^{mT}- 9-11S	5	24	1198064
NotoP Noto^{mC}+/T^{Ve}+/<i>Foxa2</i>^{mT}_high E9.5 24-28S	5	250	4788324
MP T^{Ve}_high/Noto^{mC}-/<i>Foxa2</i>^{mT} - 24-28S	5	250	6246472
Noto^{mC}+/T^{Ve}+/<i>Foxa2</i>^{mT}_high 4-11S	8	73	1664032
Noto^{mC}+/T^{Ve}+/<i>Foxa2</i>^{mT}_high 24-32 S	5	90	1581548