Dissecting the Role of *Brachyury* in Axial Elongation of the Mouse Embryo

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry,
Pharmacy
of Freie Universität Berlin

by

Dennis Schifferl

February 2021

The work here described was performed at the
Max-Planck-Institute for Molecular Genetics in Berlin
between November 2015 and February 2021
under supervision of Prof. Dr. Bernhard G. Herrmann

1st reviewer: Prof. Dr. Bernhard G. Herrmann

Department for Developmental Genetics
Max-Planck-Institute for Molecular Genetics
Ihnestraße 63-73, 14195 Berlin

2nd reviewer: Prof. Dr. Peter Robin Hiesinger

Division of Neurobiology, Institute for Biology
Free University Berlin
Königin-Luise-Str. 1-3, 14195 Berlin

Date of Defense: July 1st 2021

Statement of Authorship

I hereby declare that the work presented in this thesis has been conducted and authored independently. Experimental contributions of others are specified and acknowledged in the results section. All sources of information are referenced. I further declare that this thesis has not been submitted, either in the same or in a different form, to this or any other University for a degree.

Summary

The vertebrate body is laid down in a head to tail fashion by cells emerging from a growth zone at the posterior end of the embryo. This domain contains progenitor reservoirs providing the cellular material for axial elongation as well as the posterior notochord, which organizes the nascent tissues. The developmental master regulator Brachyury (T) controls progenitor maintenance, mesoderm formation and the specification of the notochord in a dosage dependent manner. How these different activities of T are regulated is not fully understood. In this study, I systematically dissected the gene regulatory landscape of the mouse *T* locus, where an enhancer cluster upstream of the *T* gene was mapped using ChIP-Seq. Employing the CRISPR/Cas9 system, I show that deletion of this regulatory region disrupts notochord development and tail outgrowth. Within this region, I identified a critical notochord enhancer. The enhancer mutants display axis truncation phenotypes that are consistent with the necessity of high T levels for notochord maintenance and the essential role of the notochord in axis extension.

Progenitors for axial tissues of different lineages are closely associated with the posterior notochord in a conserved topological arrangement. It has been proposed that the presumptive notochord provides the niche for axial progenitors and that these two cell groups together function as the organizer of trunk and tail development. To explore this concept, I analyzed the expression profile of nascent notochord, marked by a *Noto* reporter. Noto+ cells are the source of combinatorial Shh, Nodal, Fgf, Wnt and apelinergic signaling as well as BMP antagonists. The expression of some of these signals peaks at the stages of development when progenitor cells are amplified, indicating a role for Noto+ cells in this process. Further, I investigated the activity of the pluripotency factor Oct4 by light sheet microscopy throughout axial elongation, revealing a novel expression domain in a subset of progenitor cells in the early tail bud. RNA-Seq suggests that co-expression of *Oct4*, *Sox2* and *T* represents an axial stem cell signature.

Taken together, axial elongation is driven by interactions of Noto+ cells and progenitors. Differentially regulated activities of *Brachyury* are essential for every aspect of this process.

Zusammenfassung

Der Körper von Wirbeltieren wird entlang der rostrokaudalen Achse von Zellen aus einer Wachstumszone am hinteren Ende des Embryos angelegt. Diese Domäne enthält Reservoirs von Vorläuferzellen, die das Zellmaterial für die Achsenverlängerung generieren, sowie das posteriore Notochord, das die entstehenden Gewebe organisiert. Der für die Entwicklung maßgebliche Transkriptionsfaktor Brachyury (T) steuert dosisabhängig die Erhaltung der Vorläuferzellen, die Mesodermbildung und die Spezifizierung des Notochords. Wie diese verschiedenen Aktivitäten von T reguliert werden, ist nicht vollständig verstanden. In dieser Studie wurden genregulatorische Elemente des *T*-Locus in der Maus mittels des CRISPR/Cas9-Systems systematisch zerlegt und so ein Enhancer-Cluster upstream von *T* anhand von ChIP-Seq Daten lokalisiert. Die Deletion dieser regulatorischen Region verhindert die Entwicklung des Notochords und das Auswachsen der Schwanzknospe. Innerhalb der Region wurde ein essentieller Notochord-Enhancer identifiziert. Die Enhancer-Mutanten zeigen neben anderen Defekten einen vorzeitigen Abbruch der Achsenbildung auf. Folglich werden bestimmte Schwellenwerte an T Expression für die Aufrechterhaltung des Notochords benötigt. Das Notochord wiederum spielt eine essentielle Rolle bei der Aufrechterhaltung der Achsenbildung.

Vorläuferzellen der axialen Gewebe sind in einer konservierten Anordnung eng mit dem posterioren Notochord verbunden. Es ist möglich, dass das präsumptive Notochord die Nische für axiale Vorläuferzellen bildet und diese beiden Zellgruppen zusammen als Organisator der Rumpf- und Schwanzentwicklung fungieren. Um dieses Konzept zu untersuchen, wurde das Expressionsprofil des frühen Notochords mittels eines *Noto*-Reporters analysiert. Noto+-Zellen sind die Quelle von sekretierten Shh-, Nodal-, Fgf-, Wnt- und Apelin-Signalen, sowie BMP-Antagonisten. Die Expression einiger dieser Signale erreicht ihren Höhepunkt in den Entwicklungsstadien, in denen die Vorläuferzellen expandiert werden, was auf eine Rolle der Noto+-Zellen in diesem Prozess hinweist. Des weiteren wurde die Aktivität des Pluripotenzfaktors Oct4 während der Achsenbildung durch Mikroskopie analysiert, wobei eine bisher unbeschriebene Expressionsdomäne innerhalb der Vorläuferzellen in der frühen Schwanzknospe entdeckt wurde. Transkriptomdaten legen nahe, dass Oct4, Sox2 und T die Genexpressionssignatur einer neuen Klasse axialer Stammzellen bilden.

Zusammenfassend legen die Daten nahe, dass Interaktionen von Notochord- und Vorläuferzellen die Achsenbildung antreiben. Unterschiedlich regulierte Aktivitäten von *Brachyury* sind für jeden Aspekt dieses Prozesses wesentlich.

Contents

1 Introduction

1.1 The organizer concept in developmental biology

The metazoan body is a highly complex, stable system comprising specialized cell types, tissues and organs unfolding from a single totipotent cell during embryonic development. In the early 20th century, scientists pioneering the field of experimental biology established methods to manipulate, dissect and analyze embryos during this dynamic process. Grafting experiments in amphibian and sea urchin embryos, where fragments of embryos were tested for their potential to contribute to tissues in different contexts, coined basic principles of developmental biology, such as differentiation and self-regulation (Huxley and De Beer, 1934).

The Organizer is probably the most defining concept of that era and was introduced by Hans Spemann based on the findings of his doctorate student Hilde Mangold. They famously identified a structure at the amphibian upper blastopore lip, which, upon transplantation to an ectopic site in a host embryo, would initiate the development of a twinned embryo (Spemann and Mangold, 1924; Figure 1.1). Other than previous studies with similar results, they used differentially pigmented newt species and were able to distinguish between host and donor cells (Harland and Gerhart, 1997; Holtfreter, 1988). They found that the secondary body axis was formed by host cells, whereas contribution of the donor cells was mainly limited to axial mesoderm (Spemann and Mangold, 1924). These observations suggested that rather than integrating into the host environment, the transplanted cells from this specific region instructed their neighboring cells, which otherwise would have developed into epidermis, to become "dorsalized", adopt a new fate and give rise to the tissues of a duplicate axis.

The "Organizing center" was postulated as a structure with three major features: First, the potential to instruct morphogenesis, induce the establishment of the body plan and initiate neural activation. Second, self-differentiation into notochord and partly somites and floor plate. Finally, the activity of a signaling center patterning surrounding tissues and organ anlagen (Spemann and Mangold, 1924).
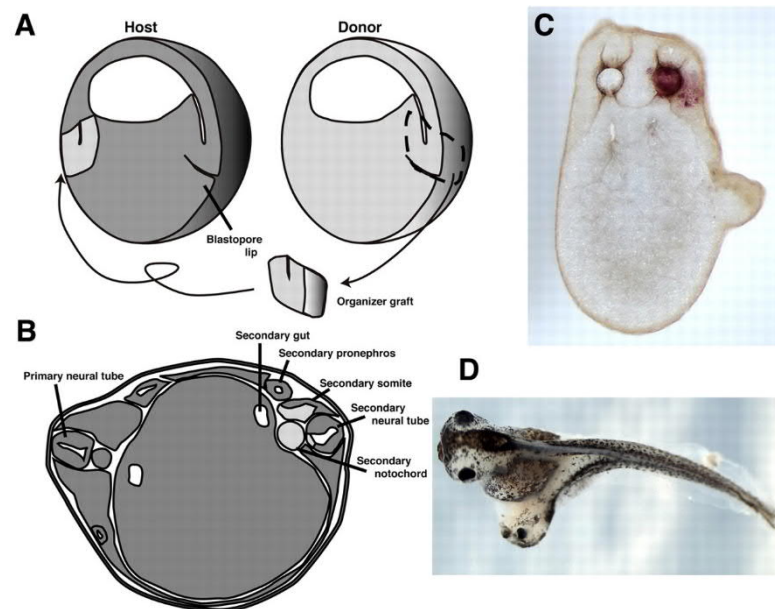
Figure 1. 1 The Organizer experiment

(A) Schematic of the grafting experiment using differently colored *Triturus* newt species (B) Section through the trunk of a secondary embryo. The graft (light-gray) contributes to notochord and partly medial somites and floor plate. The graft has an induced neural tube, somites, a pronephros and a secondary archenteron cavity in the host (dark-gray). (C, D) Contemporary organizer grafts from Andrea E. Wills (UC Berkeley, CA, USA). (C) The section shows a rafted organizer labeled with *lacZ* mRNA and stained with Red-Gal. (D) Twinned *Xenopus* embryo, resulting from an organizer transplantation. Figure adapted from (Harland, 2008).

In the era of pre-molecular biology, the compounds mediating cell-cell communication and signaling remained elusive. It took until the end of the twentieth century to find the first parts of the molecular cocktail secreted by the Spemann-Mangold Organizer. Isolation of genes from organizer derived cDNA libraries identified a range specifically expressed factors, which can induce the neural axis in a gastrulating embryo mainly by antagonizing BMP, Wnt and Nodal signaling (De Robertis and Kuroda, 2004; Niehrs, 1999).

The group of cells introduced as the Organizer in the initial graft experiment has been further dissected in sub-regions with specific activities that become sequentially important during development (Smith and Slack, 1983). An organizer from an early gastrula embryo can induce a complete axis, whereas grafts from a late gastrula embryo only result in formation of posterior tissues, which led to the proposition that there are separate organizers for head, trunk and tail (Zoltewicz and Gerhart, 1997). Alternatively, in contrast to a model where different portions of the organizer code for different positional information, Peter Nieuwkoop proposed a two-step model in which first, formation of anterior neural structures is activated

2

and in a second transformation step, signals from the organizer gradually "posteriorize" neural tissue (Nieuwkoop, 1952). Experimental evidence of sub-regions with distinct organizing properties and the activation-transformation model were integrated in a three-step-model, in which a pre-neural state is transiently activated, then stabilized by signals from the organizer, which finally posteriorizes the neural territory (Stern, 2001).

Considering the rapid progression of embryogenesis in newts, the Spemann-Mangold Organizer is an accumulation of distinct dynamic activities that spatio-temporally overlap in amphibian embryos, rather than a uniform and universal structure. Some decades after publication of the Organizer experiment, the importance of the competence of the host tissue to respond to external signaling and self-organize after induction was emphasized (Holtfreter, 1988). Nevertheless, the basic phenomenon is widely conserved and structures with analogous functions are found in other vertebrates, like the embryonic shield in zebrafish, Hensen's node in birds and arguably the node in mammalian embryos (Joubin and Stern, 2001).

1.2   The Organizer in the mouse

1.2.1   Initial axis development and the role of early and late gastrula organizer

In the mouse, the homologous organizer structures appear after implantation into the uterus at a stage when the embryo has already completed some axial organization. The earliest rounds of zygotic cleavages establish a polarity that is epigenetically engraved into the cells and determines which blastomeres of the morula become trophoectoderm and which will contribute to the inner cell mass (Torres-Padilla et al., 2007; Zernicka-Goetz, 2002; Zernicka-Goetz et al., 2009). The inner cell mass further segregates into epiblast, a pluripotent epithelium that will form the embryo proper, and primitive endoderm, the second extraembryonic lineage which covers the epiblast. By the time the resulting blastocyst embryo implants into the uterine wall at embryonic day (E) 4.5, two asymmetries are specified: The proximal-distal (PD) axis with the epiblast on one pole and the trophoectoderm on the other as well as a bilaterality apparent by a slight tilt in the blastocyst. While the role of the latter

remains unclear, the PD polarity is converted to the anterior-posterior (AP) axis in a series of cell interactions during implantation. On the distal pole, high Nodal signaling from the epiblast induces expression of Lefty1, a Nodal antagonist in a subset of primitive endoderm cells (Takaoka et al., 2017). These specified cells will give rise to the distal visceral endoderm (DVE; Figure 1.2), a signaling center expressing Lefty1, Cerberus (Cer1) and Dkk1, which act as extracellular antagonists on Nodal and Wnt signaling in the overlying epiblast (Arnold and Robertson, 2009). At about E5.5, the DVE cells collectively migrate into one direction and form the anterior visceral endoderm (AVE), which establishes a Nodal and Wnt gradient throughout the epiblast and is the foundation of the AP axis (Antonica et al., 2019; Morris et al., 2012; Takaoka et al., 2011). The posterior side is determined by high levels of Nodal, Wnt3 and Wnt3a, as well as BMP4 and Nodal potentiating convertases from the extraembryonic ectoderm (ExE) (Ben-Haim et al., 2006; Perea-Gomez et al., 2001; Perea-Gomez et al., 2002). The AVE secretes Wnt and Nodal inhibitors which further regionalize the embryo (Glinka et al., 1997; Perea-Gomez et al., 2001).
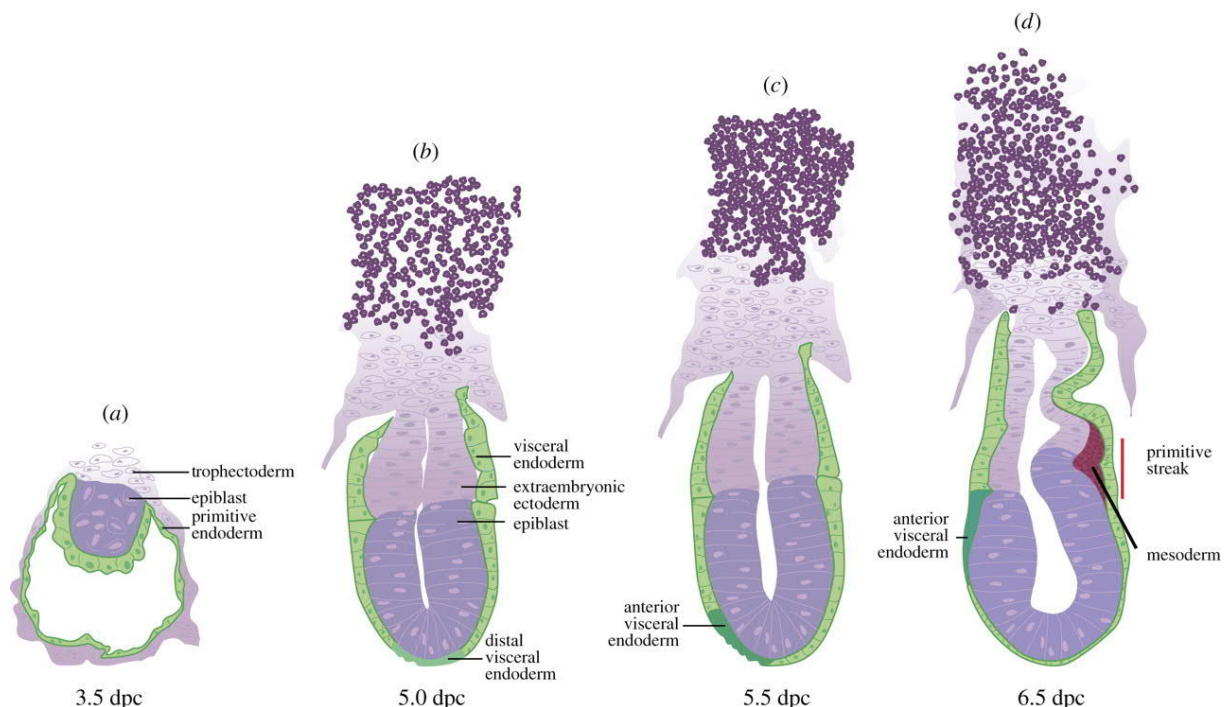
Figure 1. 2 Early embryogenesis from implantation to primitive streak formation

(a-d) Post-implantation development from 3.5-6.5 dpc (days post coitum). Adapted from (Arkell and Tam, 2012).

4

The molecular patterning in the implanted embryo sets up the coordinate system for gastrulation, in which the apparent radial symmetry is broken. During gastrulation, cells from the posterior epiblast undergo epithelial to mesenchymal transition (EMT) and ingress through the primitive streak (PS) to form the three germ layers, ectoderm, mesoderm and definitive endoderm (Keller et al., 2003; Solnica-Krezel and Sepich, 2012; Tam and Loebel, 2007). This represents a major morphological transformation, which lays out both the AP axis and the dorso-ventral (DV) axis with the ectoderm layer on the dorsal and endoderm layer on the ventral side.

In the mouse, the first Organizer structure is specified after the PS is specified. Epiblast cells close to the anterior tip of the PS form the Early Gastrula Organizer (EGO), marked by expression of *Gsc* and *Foxa2* (Filosa et al., 1997). Lineage tracing of EGO cells suggest that it is a transient cell population with main contributions to head process notochord and foregut endoderm (Kinder et al., 2001). In contrast to its equivalents in amphibians, fish and birds, the mouse EGO does not evoke the complete organizer response in transplantations experiment (Tam et al., 1997). However, when co-transplanted with the AVE and its adjacent epiblast, it can induce a secondary axis expressing anterior forebrain markers (Tam and Steiner, 1999). Proper positioning of the AVE is required for head formation and mutants lacking the secreted factors or transcription factors for AVE specification have a headless phenotype (Albazerchi and Stern, 2007; Perea-Gomez et al., 2002). Still, the AVE is not sufficient to induce the neuraxis but rather synergistically interacts with signals from the EGO and PS (Perea-Gomez et al., 2001; Tam and Steiner, 1999).

At mid streak stage, the Gastrula Organizer was shown to acquire a different molecular signature and expresses BMP antagonists Chordin and Noggin (Kinder et al., 2001). Grafting experiments of the Mid Gastrula Organizer (MGO) revealed tissue contributions similar to the EGO, but a slightly increased capacity to induce expression of anterior neural markers like *Otx2* (Kinder et al., 2001).

In general, the Organizer activities that are condensed in amphibians are spatially and temporally dispersed in mammalian embryos. After initial self-organization of the blastocyst, different signaling centers, namely the AVE, the EGO and MGO synergistically regulate head formation and positioning of the primitive streak. There is no single structure in the mouse gastrula that would meet all of Spemann's stringent Organizer characteristics.

1.2.2 The node

The primitive node is the third region with organizing activity in the mouse and arises after primary gastrulation (E7.5) at the anterior end of the PS as an indentation at the surface of the cup-shaped embryo. In contrast to the EGO and MGO, the node is a morphologically distinct structure, which is distinguishable from squamous endoderm by its columnar cell shape. Transplants from the node region were shown to differentiate into trunk notochord and to evoke a limited organizer response inducing posterior tissues (Beddington, 1994; Kinder et al., 2001; Sulik et al., 1994). Therefore, the node was postulated to function as trunk and tail organizer and an equivalent of the LGO (Kinder et al., 2001).
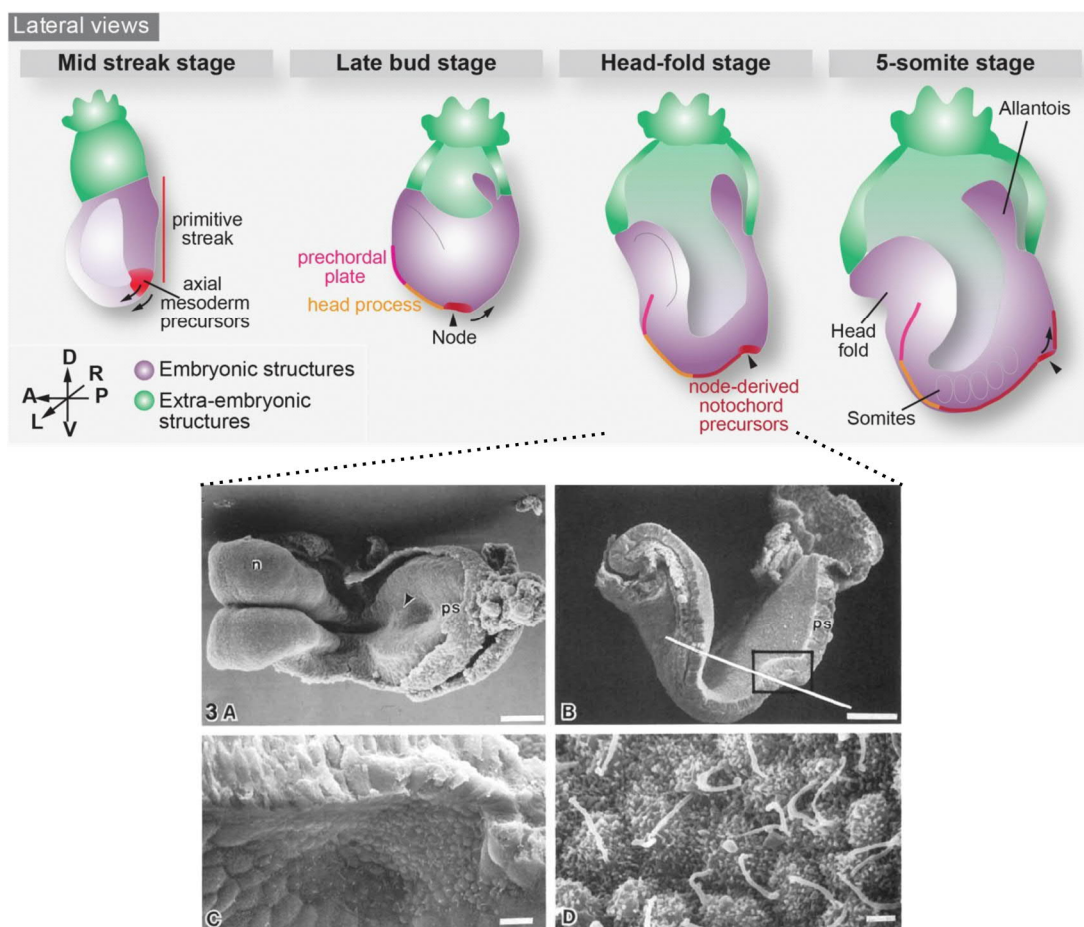
Figure 1. 3 Node and notochord morphogenesis in the mouse

Top: Schematic representation of axial mesoderm formation. Figure adapted from (Balmer et al., 2016). Bottom: Scanning electron microscopy of an E8.25 node. (A) Top view of the whole embryo. (B) Sagittal section. (C) Zoom of the node area indicated by the square in (B). (D) Close-up of the ciliated ventral node. Figure adapted from (Sulik et al., 1994).

The phenotype of mutants that do not form a node (Ang and Rossant, 1994; Weinstein et al., 1994) as well as ablation experiments (Davidson et al., 1999) demonstrate that it does not have an essential function in AP axis establishment. The axis truncation phenotype in these mutants however suggests that the node structure could have a function in maintenance of axial elongation.

Additionally, the ventral node establishes of left-right (LR) asymmetry, the final major body axis. This function has been attributed monociliated cells in the node pit (Figure 1.3), which propel a unilateral flow of Nodal protein (Hirokawa et al., 2006; Lee and Anderson, 2008; Nonaka et al., 2002).

There is some controversy concerning the terminology of "the node". A careful morphological and genetical analysis showed that ciliation and expression of organizer signature genes are displayed by different entities: Columnar cells in the indentation giving the node its typical gestalt are ciliated, but expression of *Nodal* is limited to the crown cells, a more posterior, squamous cell population (Blum et al., 2007). It was suggested to revise the nomenclature and refer to the ciliated cells in the indentation as notochordal plate or posterior notochord (PNC) and reserve the term "node", which implies organizer activity, for the crown cells (Blum et al., 2007). In their function specifying LR asymmetry, the PNC is more analogous to the gastrocoel roof in anamniotes or Kupffer's vesicle in fish than to the late gastrula organizer. Again, organizer activities are distributed between separate signaling centers, which in some, but not all cases temporally and spatially overlap between different animal clades (Martinez Arias and Steventon, 2018).

1.2.3   The notochord

After the nascent embryonic linages of the gastrula are specified by the primary organizer structures discussed above, dorso-ventral and medio-lateral (ML) patterning is established and maintained by axial mesoderm. Axial mesoderm is positioned in the middle of the DV and ML axes and can be partitioned in three parts: prechordal plate, head process notochord and node-derived trunk and tail notochord (Figure 1.3).

The notochord is a rod-like structure, which runs through the midline of the embryo from head to tail. As the defining feature of chordates it serves two major functions: Patterning and providing structural stability to the developing embryo (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 gives rise to the *nucleus pulposus* of the vertebral discs (Choi et al., 2008; McCann et al., 2012).

Adjoining organ anlagen of all 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 possibly its best described signaling function (Placzek and Briscoe, 2018). Further, critical signals from the notochord maintain the LR axis and provide signals for endoderm derivatives during organogenesis (Cleaver and Krieg, 2001). In the mesodermal lineage, the notochord is required for somite patterning (Brand-Saberi et al., 1993; Fan and Tessier-Lavigne, 1994) as well as vertebral column differentiation and segmentation (Ward et al., 2018; Wopat et al., 2018).

While the exact mechanism of notochord formation remains unclear (Balmer et al., 2016), there are three phases that can be distinguished. First, precursors of head process notochord and node re-emanate from the PS and epithelialize at its anterior end. Head process cells move towards the anterior pole, then, in the second phase, cells from the node form the trunk notochord. At E7.5, node and notochordal plate cells have a basement

membrane continuous with the definitive endoderm until subsequently they submerge from the surface of the embryo and elongate in a convergence and extension process. The convergent extension required for trunk notochord formation is controlled by the planar cell polarity (PCP) and non-canonical Wnt signaling (Andre et al., 2015; Minegishi et al., 2017; Song et al., 2010; Ybot-Gonzalez et al., 2007). Cells from the relatively broad node region straighten and intercalate to form notochord. Lengthening of the node structure may be sufficient to produce the more anterior trunk notochord, but since in parallel AP axis of the embryo drastically extends between E7.5 and E13.5, an additional source of cells is required to generate posterior trunk and tail notochord. Thus, after the quiescent node cells are exhausted, progenitor reservoirs are thought to give rise to the third section of the notochord (Ukita et al., 2009; Wymeersch et al., 2019; Yamanaka et al., 2007). The origin of these cells remains to be understood, however live-imaging data suggests that progenitor cells residing at the caudal tip of the notochord actively migrate towards the posterior to extend the midline axis (Yamanaka et al., 2007). The initially quiescent notochord progenitors proliferate at E9.5 after a stimulus, which probably involves β-catenin (Ukita et al., 2009) and EMT (Andre et al., 2015).

Genetically, the different levels of notochord specification are controlled by a hierarchy of transcription factors. The endodermal master regulator *Foxa2* is essential for the formation of prechordal plate, the head process as well as trunk and tail notochord and throughout the AP axis (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 T* (discussed in more detail in 1.5) in a dosage dependent manner (Herrmann et al., 1990; Kispert and Herrmann, 1994). *Noto* is required for tail notochord and disrupted in the *truncate* mutant, which displays a shortened tail phenotype (Abdelkhalek et al., 2004; Plouhinec et al., 2004; Zizic Mitrecic et al., 2010). Of the three, *Noto* is the only factor that is exclusively expressed in axial mesoderm precursors from E7.5. During trunk notochord morphogenesis, *Noto* functions in a synergistically with *Foxa2*, before it becomes essential for tail notochord maintenance (Yamanaka et al., 2007).

During development of trunk and tail, the notochord patterns the axial and paraxial tissues. Both notochord and neighboring tissues are derivatives of progenitor cell pools residing in close proximity to the caudal end of the notochord (Wymeersch et al., 2021).

1.3 Neuromesodermal progenitors and the axial stem cell niche

Vertebrates are highly variable in their length, which depends on the duration of axial elongation. This process generates a species-specific number of segments comprising the major proportion of the body in most animals. Axial elongation occurs during mid gestation in a head to tail fashion and is fueled by a supply of cells from a posterior growth zone. Constant recruitment of cells from the caudal growth zone leads to a drastic extension of the body between E7.5 and E13.5. Tissues of all germ layers are continuously generated: Musculature, bones, kidney and heart from mesoderm, spinal cord and epidermis from ectoderm and lungs and organs of the digestive tract from endoderm. In the last decades, the paradigm that the three germ layers segregate during gastrulation was challenged with the discovery of multi-potent progenitors that retain germ layer plasticity, self-renew and therefore behave like stem cells (Wilson et al., 2009). In the mouse, such populations have been identified through serial transplantation in specific regions, namely the node-streak border (NSB) and caudal lateral epiblast (CLE) in the trunk (Cambray and Wilson, 2007) and the chordo-neural hinge (CNH) in the tail bud (Figure 1.4; Cambray and Wilson, 2002). Cells in these compartments retain their stemness after passage through multiple embryos. The CNH is a continuum of the NSB and both regions have a shared arrangement of topological domains: An ectodermal layer expressing mesodermal characteristics lying dorsally adjacent to notochord precursors. Clonal analysis revealed that the stem-like cells can contribute to both somitic mesoderm and neuroectoderm throughout embryogenesis and have hence been termed neuro-mesodermal progenitors (NMPs) (Henrique et al., 2015; Tzouanacou et al., 2009). NMPs reside in an environment that protects them from differentiation promoting growth factors and signals. Co-expression of mesodermal master regulator *Brachyury T* and pluripotency and neuro-ectodermal factor *Sox2* creates an undifferentiated state, which is maintained by high Wnt and Fgf (Delfino-Machín et al., 2005; Garriock et al., 2015; Henrique et al., 2015; Koch et al., 2017; Tsakiridis and Wilson, 2015). Antagonistic activities of *Sox2* and *T* create a bi-stable switch controlling the progenitor exit and neural versus mesodermal fate choice (Gouti et al., 2017; Koch et al., 2017). In the posterior trunk and tail NMPs, lineage decisions and differentiation are inhibited by a network involving the Lin28/let-7 pathway, which genetically maintains the progenitor state (Aires et al., 2019; Robinton et al., 2019).

NMPs can be generated *in vitro* from embryonic stem cells (ESCs) by Wnt and FGF application in neural-lineage promoting medium (Gouti et al., 2014; Turner et al., 2014). In addition to 2D cell culture approaches, 3D systems for studying embryonic development have been gaining momentum with the development of protocols for the generation of "gastruloids", ESC derived elongating cellular ensembles with a T+/Sox2+ double positive progenitor pool producing neural and mesodermal derivative tissue (van den Brink et al., 2014). Upon addition of the ECM surrogate Matrigel to the culture medium, these neural and mesodermal derivatives were shown to give rise to neural tube and segment into somites, respectively (van den Brink et al., 2020; Veenvliet et al., 2020).

Further, it was recently postulated that NMPs are in a partial EMT like, transitory state that allows them to switch between neuroepithelium and mesodermal mesenchyme (Dias et al., 2020; Goto et al., 2017; Kinney et al., 2020).

NMPs are bicompetent precursor cells, which are regulated by a specific signaling environment, an intrinsic transcriptional balance and an intermediate morphological state. It is still a matter of debate whether NMPs reflect a self-renewing axial stem cell population that persists throughout axial elongation or if an NMP is a cell in a transient bipotent cell state (Sambasivan and Steventon, 2020).

Their close spatial association to the node and posterior end of the notochord suggests that there is a functional connection between the two domains (Wymeersch et al., 2019; Wymeersch et al., 2021). One possible nature of such an interaction could be that the posterior notochord provides factors establishing the niche where the NMPs can self-renew. It is also conceivable that in addition to the DV and ML patterning of the axial tissues by the notochord, the very first derivatives of NMPs receive lineage instructing signals from the posterior-most notochord cells.
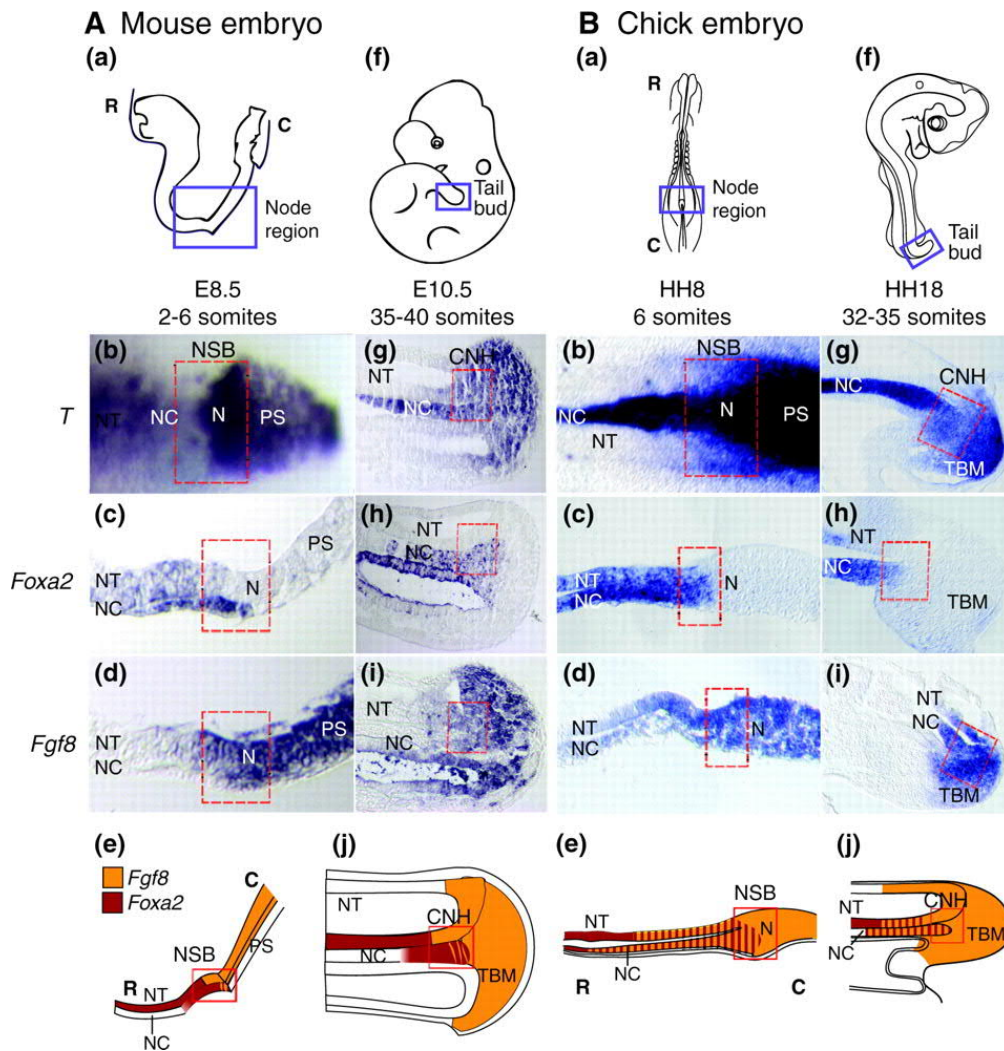
Figure 1. 4 Node-Streak-Border and Chordo-Neural-Hinge are conserved domains harboring progenitor cells

A conserved topological arrangement of morphologies and gene expression define the Node-Streak-Border (NSB) and Chordo-Neural-Hinge (CNH) in mouse (A) and chick (B). R = Rostral. C = Caudal. NC = notochord. N = node. PS = primitive streak. NT = neural tube. TBM = tail bud mesoderm. In the mouse, NSB and CNH can be defined as the axial position where *Foxa2* and *Fgf8* expression overlap. In chick, *Fgf8* extends more rostral to the NSB and CNH and cannot be used as a marker. Figure adapted from (Wilson et al., 2009).

1.4 Signals in the caudal growth zone

Embryonic development is driven by coordinated differentiation of competent progenitors towards increasingly specialized tissues, which is orchestrated by signaling centers, reciprocal signaling between tissues and self-organization. During vertebrate axis elongation, progenitor populations in the caudal end have to be protected from differentiation promoting signals. This is ensured by an interplay of signaling pathways (Wilson et al., 2009). In brief, Wnt and Fgf signals in the progenitor zone inhibit factors like retinoic acid (RA) that promote differentiation. At the core of this mechanism is a positive feedback loop between Wnt3a, T, and Fgf8/4 binding to Fgfr1. All of these compounds are essential and mutants show severe axis truncation phenotypes (Ciruna and Rossant, 2001; Herrmann, 1991; Naiche et al., 2011; Takada et al., 1994). RA is synthesized by Aldehyde dehydrogenase *Aldh1a2* in NMP derivatives, namely neural plate and somites, and opposes Fgf8 activity in the NMP domain (Diez del Corral et al., 2003). Wnt3a and T regulate Cyp26a1, a hydroxylase that degrades RA, thereby protecting the NMP domain from RA exposure (Gouti et al., 2017; Martin and Kimelman, 2010). The activation of the progenitor niche is further assisted by Cdx transcription factors, most prominently Cdx2, which links Wnt and Fgf signaling and positional information encoded in Hox collinearity (Amin et al., 2016; Savory et al., 2009; Young et al., 2009).

Within NMPs, the neural versus mesodermal lineage choice is controlled by a genetic antagonism between T and Sox2, with cells downregulating T and upregulating Sox2 to undergo neural differentiation and vice versa (Gouti et al., 2017; Koch et al., 2017; Romanos et al., 2020). Mesodermal lineage commitment is initiated by *T* and *Wnt3a* and genetically anchored by *Tbx6* expression, which downregulates both *T* and *Sox2* (Figure 1.5; Chapman and Papaioannou, 1998; Koch et al., 2017; Takemoto et al., 2011). Cells that are now committed to pre-somitic mesoderm (PSM) remain mesenchymal for about two cell cycles (Tam, 1981). During that time, Fgf8 and Fgf4 are thought to chemotactically guide the NMP derivatives from the posterior progenitor zone, with Fgf8 repelling the cells and Fgf4 attracting them towards somite border (Yang et al., 2002). Paraxial mesoderm is then segmented into somites, which are blocks of mesoderm giving rise to vertebrae, ribs, cartilage and muscle. The individual somites are defined by an outer epithelium formed in a mesenchymal to epithelial transition.

At a certain threshold level of the AP Wnt/Fgf gradient, referred to as the determination front, PSM becomes competent for instructive signals that govern the timing of somite formation. The determination front is periodically hit by signaling from the segmentation clock, a molecular oscillator. A tightly wired network of Wnt, Fgf and Notch signaling pathways generates waves of transcription, which start out of phase and become synchronized to induce somitogenesis (Aulehla et al., 2003; Aulehla et al., 2008; Bénazéraf and Pourquié, 2013; Sonnen et al., 2018).
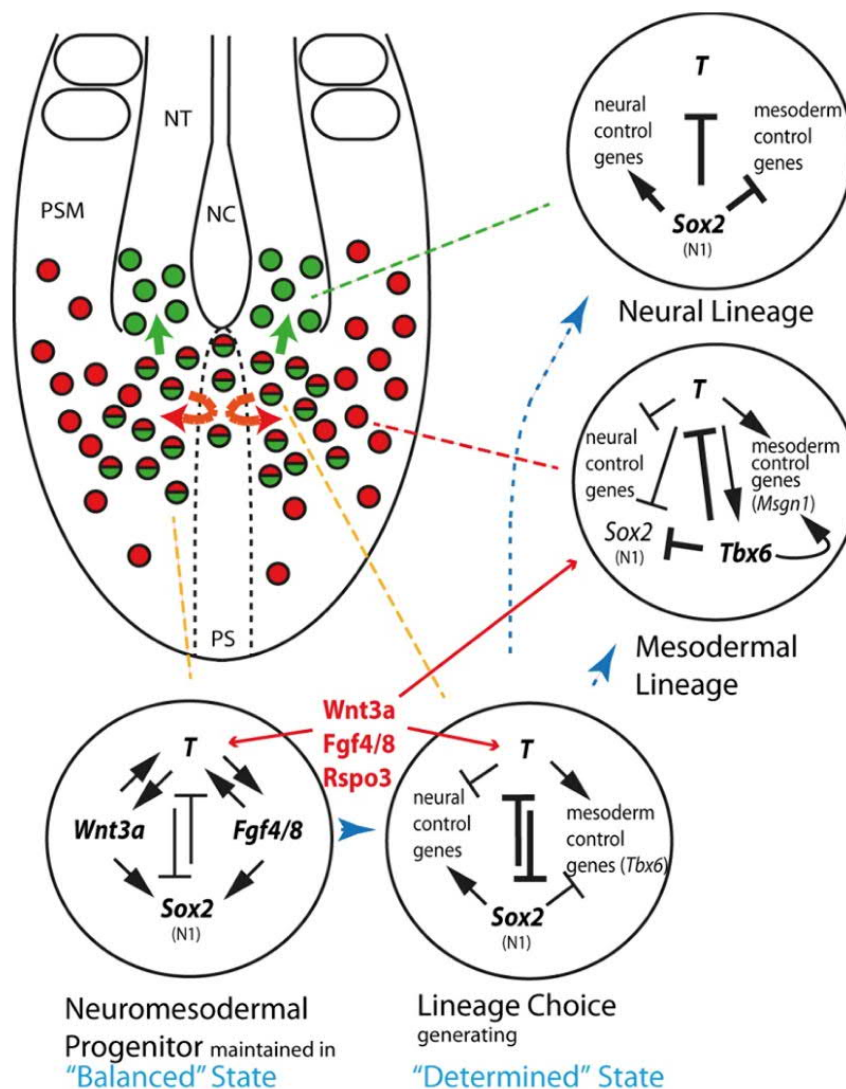
Figure 1. 5 Regulation of NMP maintenance and lineage choice

(A) NMPs (red + green) are maintained in a balanced state through *Wnt3a*, *Fgf4/8* as well as *T* and *Sox2*. Descendants acquire a determined state. Mutual inhibition of *T* and *Sox2* regulates the lineage choice. The determined state is reinforced by lineage control genes like *Tbx6*. Neural progenitors (green) remain in the neuroepithelium, while mesoderm progenitors (red) become mesenchymal. PS = primitive streak. NC = notochord. PSM = presomitic mesoderm. NT = neural tube. Figure adapted from (Koch et al., 2017).

1.5 Brachyury functions in axial elongation

Brachyury (T) is a key regulator and essential for multiple aspects of axial elongation (Herrmann et al., 1990; Wilkinson et al., 1990). T was shown to act as a transcription factor binding to T-box motifs (Kispert et al., 1995; Muller and Herrmann, 1997). In the mouse, homozygous *T -/-* mutant embryos are truncated and die around E10.5 due to a dysfunctional chorio-allantoic connenction (Chesley, 1935; Gluecksohn-Schoenheimer, 1944; Inman and Downs, 2006). After formation of the first eight somites, *T -/-* embryos fail to generate paraxial mesoderm. Further, the mutants display morphological aberrations in the node and lack its derivative, the trunk notochord (Davidson et al., 1999; Fujimoto and Yanagisawa, 1983). Heterozygous *T+/-* mutants are viable, but have short tails with a variable phenotype between tailless and a tail of about half the normal length (Gluecksohn-Schoenheimer, 1944). The axis defects in *T* mutants result from the loss of different T functions at successive stages of trunk and tail morphogenesis (Figure 1.6).
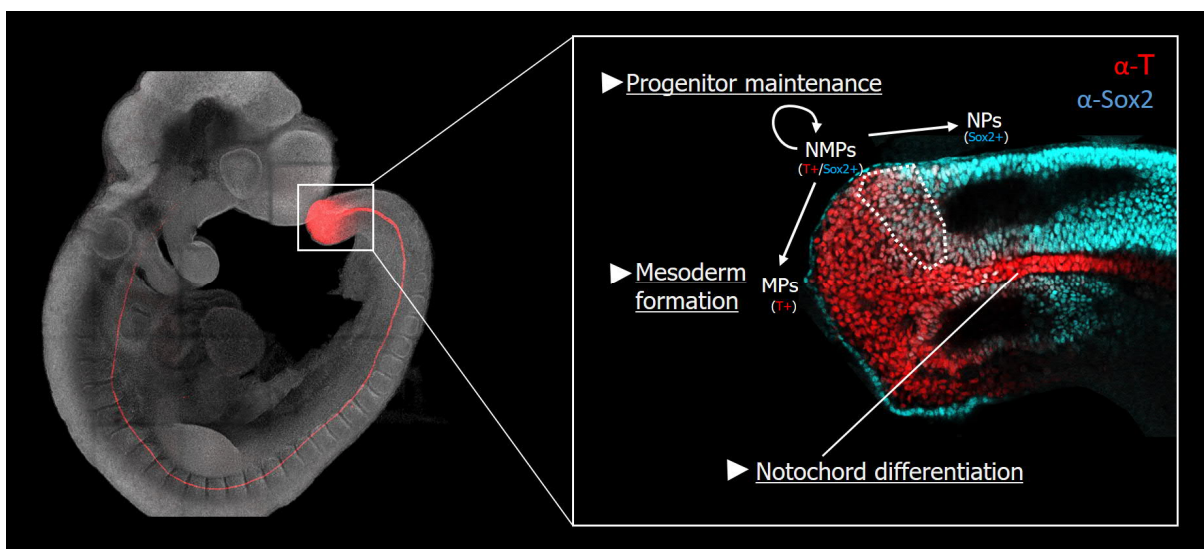
Figure 1. 6 Brachyury activities in axial elongation

Left: E9.75 mouse embryo. Maximum intensity projection of confocal microscopy stacks. Whole-mount immunofluorescence for T (red), nuclei stained using DAPI (grey). Right: Sagittal optical section through the tail bud. Immunofluorescence for T (red) and Sox2 (blue). Different activities and their corresponding expression domains are indicated.

T activity is required to maintain the pool of NMPs (Henrique et al., 2015; Li et al., 2007; Tsakiridis and Wilson, 2015; Wymeersch et al., 2016). The conserved Wnt/T loop promotes mesodermal lineage choice and acts antagonistic to the proneural activity of Sox2 (Garriock et al., 2015; Gouti et al., 2017; Koch et al., 2017; Martin and Kimelman, 2008; Martin and Kimelman, 2012; Turner et al., 2014). As a pan-mesodermal pioneer factor, T induces mesoderm formation in the primitive streak and tail bud (Beisaw et al., 2018; Tosic et al., 2019). Finally, somites, gut and neural tube are patterned by signals from the node and notochord, which require a high dosage of T for their maintenance (Chiang et al., 1996; Pennimpede et al., 2012; Stemple, 2005; Zhu et al., 2016).

Brachyury exerts these functions in adjacent domains and its dynamic expression requires robust spatio-temporal control. Expression in the nascent mesoderm is transient and downregulated as cells differentiate (Wilkinson et al., 1990). Integration of a 23 kb genomic fragment starting 8 kb upstream of the T gene was shown to restore the capacity of *T -/-* mESCs to colonize mesodermal tissues *in vivo* and completely rescue the short tailed phenotype of T +/- mice (Stott et al., 1993; Wilson and Beddington, 1997). This fragment contains the *T* promoter (-500 to 150 bp from the TSS) sufficient for primitive streak and tail bud expression as well as a recently discovered divergently transcribed non-coding RNA termed *yyT* (Clements et al., 1996; Frank et al., 2019). In addition to the dynamically regulated expression levels in mesoderm precursors, T expression in the resident progenitors, as well as in the notochord is persistent until the completion of the tail. At E13.5, NMPs are thought to be exhausted and the notochord regresses to give rise to the nucleus pulposus of the vertebral disk (Choi et al., 2008; McCann et al., 2012). The promoter and proximal elements within the 23kb genomic region are not sufficient for the high T dosage required for notochord specification and tail bud outgrowth (Stott et al., 1993). Interestingly, the *T^{BoB}* mutation, which has a 200 kb insertion mapped about 35 kb upstream of T, induces a partial, less severe *T -/-* phenotype with absent notochord and impaired tail bud outgrowth (Rennebeck et al., 1995). The *T^{BoB}* insertion was show to disrupt the rodent specific *T2* transcript (Rennebeck et al., 1998). However, the similarity of the axis phenotype as well as the genetic interactions of the *T* and *T^{BoB}* mutants suggest that rather than *T2* coding for an equally critical factor, the notochord enhancers of T might be embedded in the *T2* region and be deleted or insulated from the promoter by the large insertion in *T^{BoB}* (Wu et al., 2007).

The exact location of these enhancers however is not known and the complex transcriptional control of this developmental master regulator remains incompletely understood. Identifying these elements might reveal how during development and evolution, *T* is modulated to shape the Chordate body plan. With *T* being a tumor marker, oncogene and possibly a drug target, the investigation of its transcriptional regulation also has some clinical significance (Shah et al., 2017; Sharifnia et al., 2019; Tarpey et al., 2017; Tirabosco et al., 2008; Yang et al., 2009).

1.6   Developmental enhancers

Embryonic development produces about 200 specialized cell types from common progenitors in a relatively short time frame. Cell specification is driven by a dynamic activation of gene regulatory networks by transcription factors. Especially powerful developmental regulators, such as T for the mesodermal lineage, require a tight spatio-temporal control of gene expression, which is mediated by multiple enhancers. Enhancers are *cis*-regulatory DNA elements that contain transcription factor binding sites and can contact promoters and activate transcription over long distances, from typically about 30-70kb to 1Mb (Shlyueva et al., 2014; Spitz and Furlong, 2012). With hundreds of thousands of enhancers distributed across the genome, the estimated number of enhancer elements vastly exceeds the number of genes, implying that one gene is typically regulated by many enhancers acting in additive, synergistic or redundant ways (Kvon et al., 2021). Enhancer activity usually requires binding of multiple transcription factors, which facilitates lineage specificity and the integration of signaling pathways. Active enhancers recruit RNA Polymerase (Pol) II as well as transcriptional co-activators such as general transcription factors, p300 and the Mediator complex to promoters via DNA looping (Heintzman et al., 2007). Looping and DNA topology allows more remote enhancers to bridge closer genes or regulatory sites and selectively interact with specific promoters (Levine et al., 2014). The enhancer-promoter interaction is dependent on three dimensional chromatin organization and preferentially happens within loop domains formed by a ring-like cohesion complex and boundary elements. These topologically associating domains (TADs) may function as a scaffold for interactions of regulators elements (Furlong and Levine, 2018).
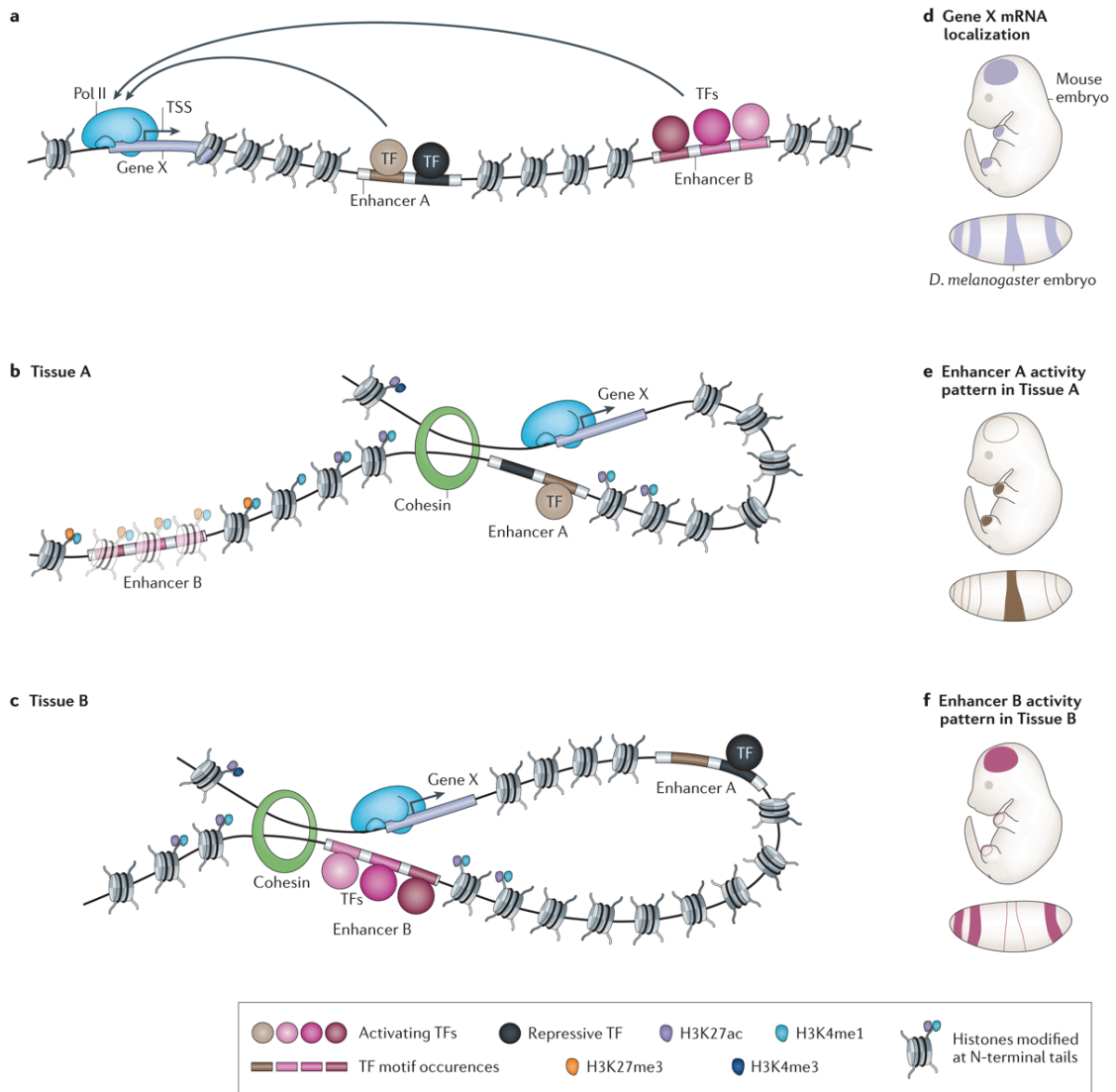
Figure 1. 7 Schematic illustration of tissue specific enhancer activation

(a) Schematic showing a Gene X regulated by a proximal Enhancer A and more distal Enhancer B, each containing a different combination of binding sites for tissue-specific activating and repressive TF (Transcription factor) binding sites. (d) Gene expression pattern resulting from regulated by combinatorial activity of Enhancers A/B. (b) and (c) Enhancer activation by tissue specific TF binding and recruitment of activating histone modifications resulting in enhancer activity patterns (e) and (f). Figure adapted from (Shlyueva et al., 2014)

The recruitment of the transcription machinery leads to transcription of active enhancers (Kim et al., 2010). A genome-wide analysis of PolII and co-activator binding and enhancer transcription revealed CpG high enhancer clusters of 0.4-10 kb especially in proximity of tissue-specific genes which were termed transcription initiation platforms (Koch et al., 2011). Recently, genomic features sharing similar characteristics are more often referred to as "super-enhancers" (Whyte et al., 2013). These regulatory regions are associated with oncogenes and key regulators of cell identity, including *Brachyury (TBX-T)* in human chordoma cell lines (Sharifnia et al., 2019; Sheppard et al., 2021).

In the last years, the phase separation hypothesis for super-enhancer activity has gained a lot of attention (Hnisz et al., 2017). In this model, chromatin architecture creates hubs in which transient clusters of PolII and complexes of the transcription machinery come together. The aggregations are thought to depend on liquid-liquid phase transitions created by interactions of intrinsically disordered regions in the TFs and parts of the transcription machinery (Boija et al., 2018). This model integrates observations like enhancer tracking, transcription from enhancers, linking, looping and transcriptional factories. It further would explain dosage effects of TFs and indirect interactions of TFs.

In general, enhancers can be predicted by assessing combinatorial high throughput datasets for transcription factor binding sites, chromatin accessibility or epigenetic markers such as histone modifications (Heintzman et al., 2007; Rada-Iglesias et al., 2011; Visel et al., 2009). Since enhancer function is independent of orientation and distance to promoters, the gold standard for enhancer activity has been the capacity to induce reporter expression when cloned upstream of a minimal promoter (Kvon, 2015). In addition, the CRISPR/Cas9 System has become a feasible tool for functional assays specifically disrupting gene regulation (Lopes et al., 2016).

1.7 Aim of this study

In essence, the generation of the vertebrate trunk and tail is initiated by two components: Progenitor reservoirs that produce the cellular material for the different lineages, as well as structures providing the signals that perpetuate these progenitors, specify their descendants and pattern the nascent tissues. The work presented here was driven by the hypothesis that the node and the posterior end of the notochord establish an instructive niche for axial progenitor cells. *Brachyury* activity is essential in different aspects of this system, namely for progenitor cell maintenance, mesoderm formation and the specification of the notochord. How the expression domains corresponding to these activities are controlled remains incompletely understood.

Therefore, in this study I set out to identify the enhancers that orchestrate T expression in trunk and tail development. To this end, I systematically dissected the cis-regulatory landscape of T using CRIPSR/Cas9. I generated and characterized loss-of-function mutants lacking different regulatory elements to provide insight into the linkage of notochord development and axial elongation.

Further, I investigated both components of the putative niche throughout trunk development. To explore the capacity of nascent notochord cells to form the niche and orchestrate lineage allocation via secreted signals, I profiled the transcriptome of *Noto* expressing cells. To localize the putative stem cells that might populate this niche, I modified the pluripotency factor Oct4 with a fluorescent protein tag and imaged the activity throughout mid-gestational development.

Finally, I frequently observed that some of the *Noto* expressing cells were not fully committed to notochord and contributed to paraxial mesoderm. This route of notochord progenitors towards vertebrae might represent an alternative mechanism for chordoma formation. Therefore, I investigated the possibility of a common precursor for the sub-mesodermal lineages using fate-mapping and RNA-Seq.

2 Results

Part I– Investigating the functions of *Brachyury* in axis elongation

2.1 *Brachyury* activity in notochord development

2.1.1 Establishment of a *Noto* reporter line

The node and the posterior portion of trunk and tail notochord are distinguished by the co-expression of the transcription factors *Foxa2*, *T* and *Noto.* With *Foxa2* expression in endodermal derivatives and floor plate and *T* expression in pre-somitic mesoderm, hindgut roof and posterior neural tube, only *Noto* qualifies as a specific marker gene (Abdelkhalek et al., 2004; Plouhinec et al., 2004). In order to make the nascent axial mesoderm cells accessible for *in vivo* fluorescence microscopy and flow cytometry purification, a fluorescent reporter mouse embryonic stem cell (mESC) line was established.

To this end, a bacterial artificial chromosome (BAC) clone comprising the murine *Noto* locus was modified such that the ATG start codon was replaced by a reporter cassette via RET-recombineering (Muyrers et al., 1999). The cassette contains the coding sequence of a Histone 2b (H2B) fused to mCherry fluorescent protein for nuclear signal as well as a hygromycin resistance cassette for selection in mESCs (Figure 2.1.1A). The linearized *Noto::H2B-mCherry* BAC was randomly integrated into wildtype F1G4 hybrid mESCs (Nagy et al., 1993), single clones were picked after selection and genotyped by PCR.

In order to validate the expression pattern of *Noto::H2B-mCherry* (hereafter Noto[mC]) reporter and exclude secondary effects from the modification, mouse embryos of mESC clones were generated by tetraploid complementation assays (Figure 2.1.1B). Embryos were isolated at E7.5 and E9.5. The embryos did not display any signs of developmental defects. At E7.5, clear mCherry signal was detectable in the node and, in a more scattered pattern, in the head process notochord and pre-chordal plate. At E9.5, reporter expression is strongest at the posterior end of the notochord but remains detectable throughout the trunk notochord. In

general, the Noto^{mC} reporter adequately recapitulates the transcription domains known from *Noto* whole mount in situ hybridization (WISH), with Noto^{mC} signal extending somewhat further into the mature notochord, which can be attributed to the higher stability of the fluorescent reporter compared to the endogenous protein due to the H2B fusion (Figure 2.1.1B).

To further confirm that Noto^{mC} marks the right cell type, embryos were stained for T and Foxa2 by immunofluorescence. At all stages tested, Noto^{mC} overlaps with Foxa2 and high T expression and distinguishes axial mesoderm from neighboring cells (Figure 2.1.1 C). Therefore, the Noto^{mC} reporter is an appropriate tool to analyze the cells of the early notochordal lineage.

general, the Noto^{mC} reporter adequately recapitulates the transcription domains known from *Noto* whole mount in situ hybridization (WISH), with Noto^{mC} signal extending somewhat further into the mature notochord, which can be attributed to the higher stability of the fluorescent reporter compared to the endogenous protein due to the H2B fusion (Figure 2.1.1B).

To further confirm that Noto^{mC} marks the right cell type, embryos were stained for T and Foxa2 by immunofluorescence. At all stages tested, Noto^{mC} overlaps with Foxa2 and high T expression and distinguishes axial mesoderm from neighboring cells (Figure 2.1.1 C). Therefore, the Noto^{mC} reporter is an appropriate tool to analyze the cells of the early notochordal lineage.

A

B

Noto WISH   **_Noto::H2B-mCherry_**

C   **E7.25 embryo**

E9.0 caudal end

Figure 2.1.1 Establishment of a Noto reporter mESC line

(A) Schematic representation of the _Noto::H2B-mCherry_ reporter BAC. (B) Left: _Noto_ whole mount _in situ_ hybridization adapted from Pennimpede et al, 2012. Right: Noto[mC] reporter fluorescence in a E9.5 embryo acquired by stereomicroscopy (C) Confocal fluorescence microscopy of a E7.25 early headfold stage embryo and a E9.0 caudal end with Noto[mC] reporter expression and immunofluorescence for T and Foxa2. At E7.25, Noto[mC] can be detected in the ventral node, head process notochord, crown cells and cells in the midline of the epiblast. At E9.0. Noto[mC] marks the posterior tip of the notochord. Yellow arrows point out exemplary cells co-expressing all markers. Scale bar=500µm

2.1.2 ChIP-seq identifies conserved regulatory elements on the T locus

The multiple roles of *Brachyury (T)* activity during axial elongation require a robust and tight transcriptional control. The promoter and proximal elements which drive expression in the primitive streak and tail bud and induce paraxial mesoderm formation have been identified decades ago (Clements et al., 1996; Schmidt et al., 1997). There has been some evidence for one or more distal enhancers initiating the high expression levels required for notochord formation located at the T locus upstream of the T gene, the exact location however remains unknown (Stott et al., 1993; Wu et al., 2007).

Tissue specific transcription factor binding is a hallmark of enhancer activity. In order to identify novel notochord enhancer candidates, binding sites of T and Foxa2 were analyzed by Chromatin Immunoprecipitation DNA-Sequencing (ChIP-Seq). In order to achieve the high cell numbers required for this experiment, an *in vitro* notochord differentiation protocol was employed. The procedure was based on the previously published protocol by Winzi et al. (2011), which was optimized by Jesse Veenvliet (personal communication). The Noto^{mC} mESC line was seeded without feeders in N2B27 medium with Activin A to stimulate the Nodal-pathway and to differentiate the cells towards mesendoderm (Figure 2.1.2). After three days, cells were cultured in step II medium, which contains additional factors that activate FGF and Sonic Hedgehog and inhibit the BMP and retinoid acid pathways for an additional four days.
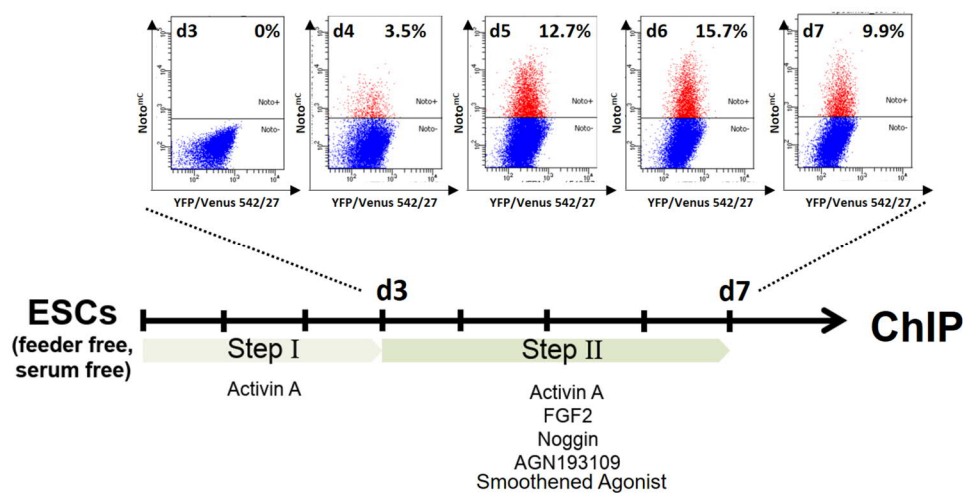

Figure 2.1.2 Schematic representation of the notochord differentiation procedure

FACS plots show the proportion Noto^{mC}+ expressing cells after application of step II medium at d3 in an exemplary experiment.

Between day 6 and day 7 of the protocol, these conditions yielded typically between 10-20% Noto^mC positive cells in the culture, as determined by fluorescence activated cells sorting (FACS; Figure 2.1.2).

ChIP-seq analysis was performed on bulk *in vitro* differentiated notochord-like cells for T and Foxa2. Binding sites of these notochord key transcription factors should be predictive for enhancers that are active early in the notochord lineage. To examine lineage specificity, these data were overlapped with published ChIP-seq data (Koch et al., 2017), where T, Sox2 and β-Catenin binding was assayed in *in vitro* generated NMPs at d3 (Figure 2.1.3 A). Indeed, the T locus is highly auto-regulated with multiple tissue specific T binding sites, of which many are co-occupied by other key transcription factors. Co-binding of T and Foxa2 suggests a role for the element in axial mesoderm or hindgut formation, co-binding of T and Sox2 would indicate a role in the neural vs. mesodermal fate choice in NMPs. The most prominent peaks were annotated as putative T enhancers (*TE1-TE7*) for further analysis. Interestingly, the sequences of these binding sites are largely conserved in mammals, a characteristic that is often seen in developmental enhancers (Figure 2.1.3 B).

Taken together, with an optimized protocol, the cell numbers for transcription factor binding profiling in notochord cells can be derived by *in vitro* differentiation. ChIP-Seq identified a number of evolutionary conserved putative enhancer elements which were characterized in further experiments.
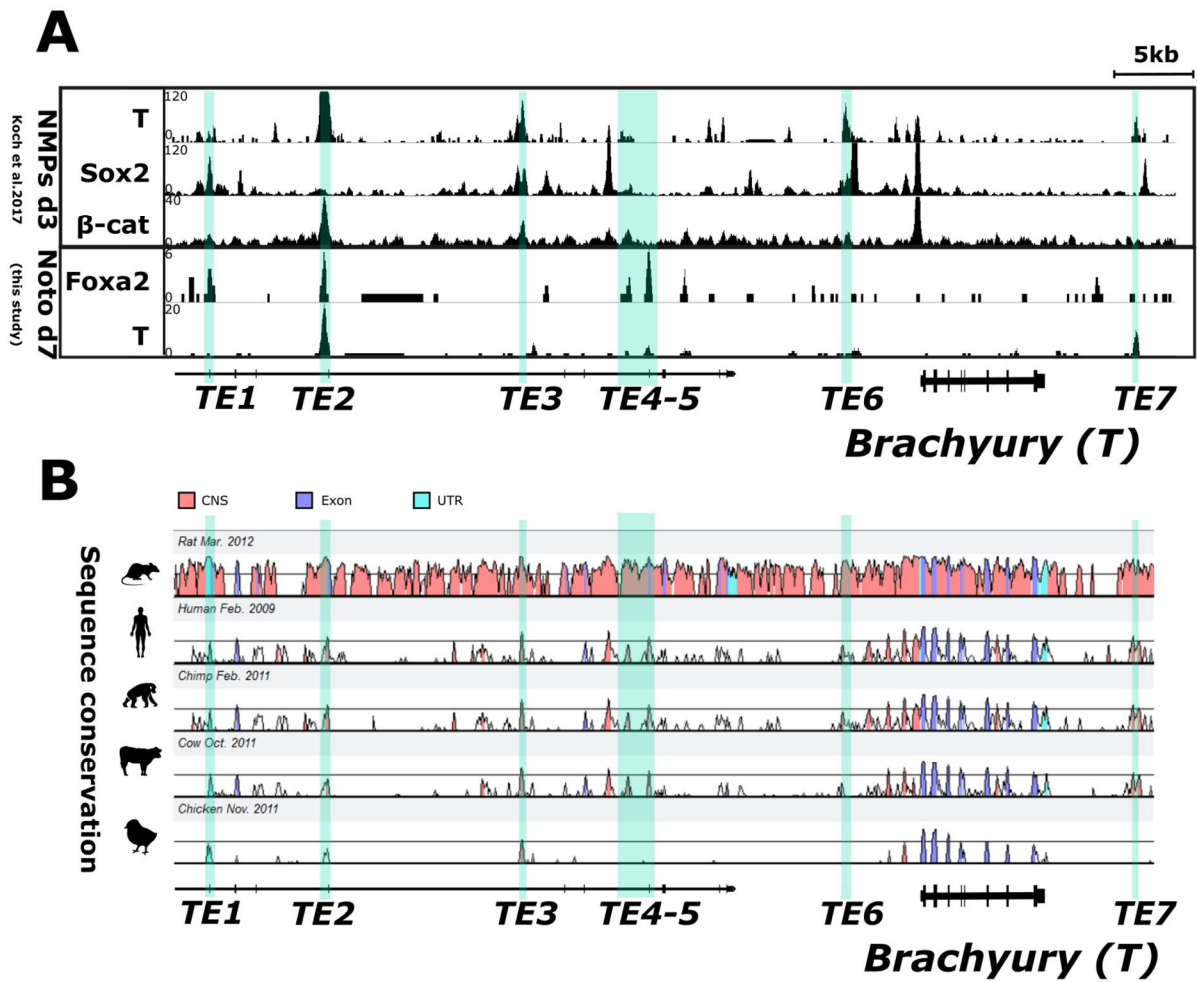
Figure 2.1.3 ChIP-Seq identifies enhancer candidates at the *T* locus

(A) ChIP-Seq tracks showing the T locus on chromosome 17 with T, Sox2 and β-Catenin binding sites in *in* vitro derived NMPs (Koch et al., 2017) and T and Foxa2 binding sites at D7 of the notochord differentiation protocol (this study). T Binding sites are annotated as *T Enhancer 1-7 (TE1-7)* highlighted in green. (B) The genomic sequences of the corresponding loci in rat, human, chimp, cow and chicken were plotted against the mouse T locus (chr17:8,386,974-8,452,208) using the VISTA browser (http://genome.lbl.gov/vista/index.shtml). Mammalian species icons were taken from a royalty free database (flaticom.com).

2.1.3   A systematic dissection of the Brachyury regulatory region using CRISPR/Cas9

The current state of knowledge about transcriptional regulation of Brachyury has mainly been established in the years after the gene was cloned (Herrmann et al., 1990). Insertion of a 23 kb genomic fragment containing the T gene and starting about 8 kb upstream of *T* can rescue the T mutant phenotype to some extent, but is not sufficient to restore notochord specification (Stott et al., 1993; Wilson and Beddington, 1997). Additionally, the T^{BoB} mutant, which has a 200 kb insertion mapped about 16 kb upstream of *T*, induces a notochord and tail bud outgrowth phenotype (Rennebeck et al., 1995). Therefore, it is reasonable that one or more notochord elements of T are embedded in the locus upstream of T and displaced or disrupted in the T^{Bob} mutant. After having identified some candidates within that region via ChIP-Seq, a systematic dissection of the T locus was carried out.

To this end, CRISPR/Cas9 technology was employed to generate a series of mutants in the Noto^{mC} mESC line (Figure 2.1.1). First, to test if the complementation assays with the genomic fragment can be recapitulated in this system, a large 35 kb deletion was introduced, which disrupts the upstream regulatory region in a way that the elements *TE1-TE5* are deleted and only 8 kb containing the extended promoter remain (hereafter referred to as T^{UD}). Then, in order to refine and narrow down the analysis, single elements were deleted in wildtype and heterozygous T mutant backgrounds (Figure 2.1.4).

The break points of the deletions were determined by PCR and Sanger sequencing (Supplementary Figures 1-3). Embryos were generated from the modified mESC clones by tetraploid complementation assays by the transgenic unit of the Max-Planck-Institute for Molecular Genetics, Berlin. Transgenic embryos were isolated between E7.5 and E13 and directly analyzed. In the following, a selection of the mutants generated is presented.
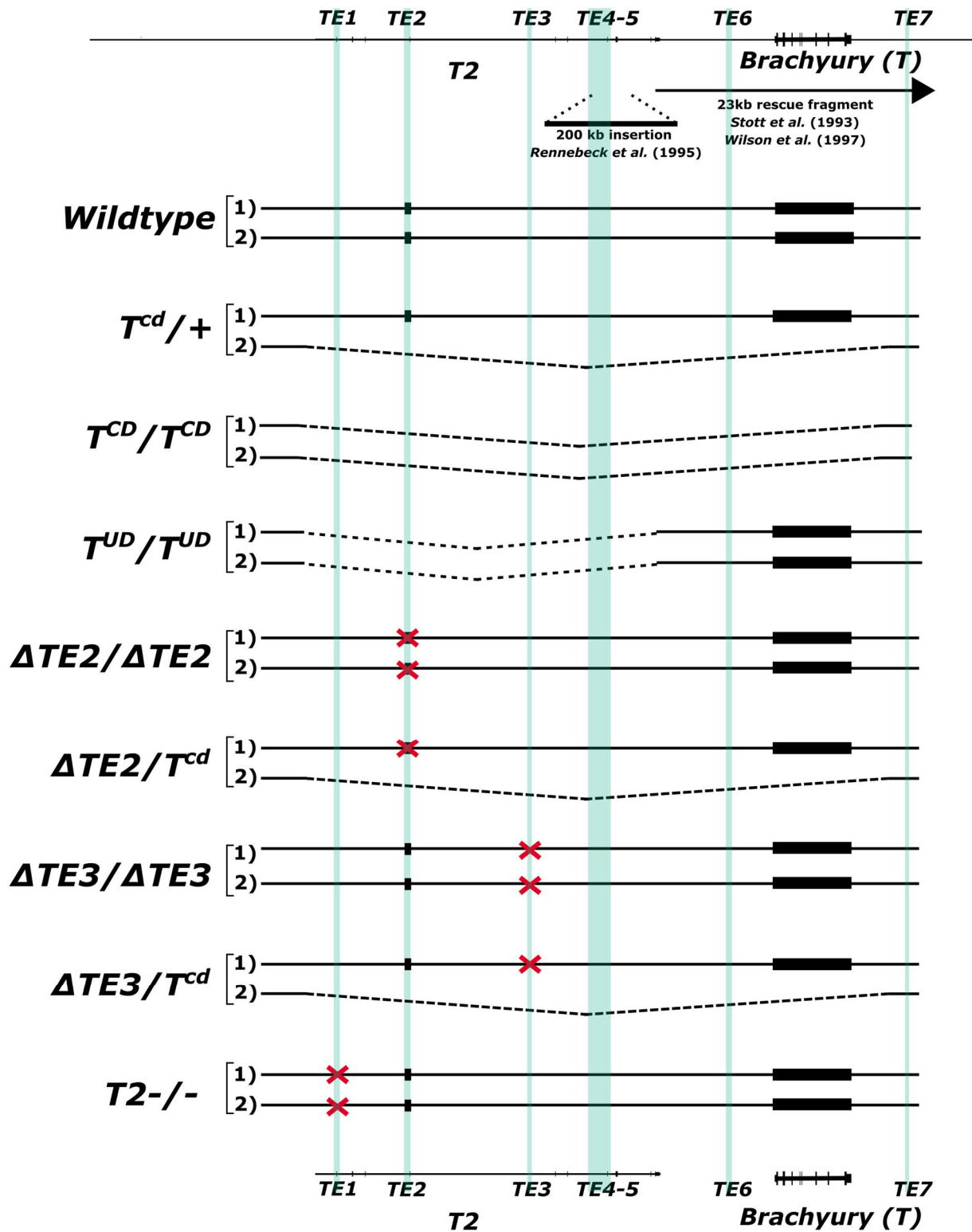
Figure 2.1.4 Genotypes of Mutants discussed in this study

Schematic showing the *T* locus (chr17:8,387,000-8,452,000; mm10) with genotypes of mutants generated in this work. Enhancer candidates highlighted in green. Large deletions indicated by dotted lines, red crosses indicate smaller deletions. T^{CD} (81 kb deletion) and T^{cd} (64 kb deletion) are both deletions of the T locus including the *T* gene, with T^{CD} starting further upstream. The exact genomic coordinates of each genotype are specified in Supplementary Table 1 and Supplementary Figures 1-3.

Table 1 A systematic dissecion of regulatory elements on the T locus using CRISPR/Cas9

List of tetraploid complementation experiments (n=number of embryos displaying the specified phenotype) of all genotypes discussed in this chapter. Proportion of embryos displaying a specific tail phenotype specified in brackets. TL = tailless. ST = short tail. NT = no tail.

Genotype	Phenotype (Noto^{mC} morphology) E9.5-10.5	Exp. clone/n E9. -10.5	Tail outgrowth E12-13	Tail E12-13	Exp. clone/n E12-13
WT	normal	#1 n =43 #1 n = 47 #1 n = 6 #1 n =33	Yes	normal	1. #1 n=4
T^{CD} / T^{CD}	Axial truncation, no paraxial mesoderm after the forelimb bud, no notochord and other defects of the T null mutant (Chesley, 1935)	#1 n= 13			
T^{cd} /+	Dispersed from hindlimb Noto^{mC}+ cells scattered in tail bud; Variable	#1 n=24 #1 n=9 #2 n=7	Yes variable tail length correlates with notochord dependent on background (BL6/SV129)	TL (17/21) ST(4/21) ST (11/11)	1. #1 n=21 2. #3 n=11
ΔT^{UD} / ΔT^{UD}	Axial truncation, Excess neuroectoderm, Notochord not maintained from forelimb bud, Noto::mC+ cells in the gut and few in the caudal end (57/58)	#1 n=29 #1 n=12 #2 n=12	No, undifferentiated tissue, neuroepithelium enlarged	TL, NT defects (6/6)	1. #1 n=6
$\Delta TE2/\Delta TE2$	Trunk notochord formed with weaker T expression Notochord dispersed from hindlimb Noto^{mC} + cells scattered in tail bud	1. #1 n=22 2. #1 n=16 3. #1 n=7 4. #2 n=33 5. #1 n=4	yes (~15 somites) Does not develop further, apoptosis -> regresses	TL(14/14)	1. #1 n=6 2. #2 n=6 3. #1 n=2
$T^{cd}/\Delta TE2$	Trunk notochord absent/disrupted, few Noto^{mC} + cells in the caudal end (26/26)	1. #1 n=15 2. #1 n=27 3. #1 n=5 4. #1 n=9	Yes (~10 somites) Does not develop further, regresses	TL (14/14)	1. #1 n=14
$\Delta TE3/\Delta TE3$	normal	1. #1 n=6 2. #1 n=8	Yes	Bifurcation at ~ somite 50 (9/9)	1. #1 n=4 2. #1 n=4 3. #2 n=1
$T^{cd}/\Delta TE3$	Dispersed from hindlimb Noto:mC+ cells scattered in tail bud Variable	1. #1 n=10 2. #1 n=5	Yes (~10 somites)	TL (8/8) bifurcations, open neural tubes, meandering neural tube	1. #1 n=4 2. #1 n=4
$\Delta TE2/\Delta TE2$; $\Delta TE3/\Delta TE3$	like $\Delta TE2/\Delta TE2$, no additional defects	1. #1 n=11 2. #2 n=14	Yes (~15 somites)	TL	1. #1 n=12
$T^{cd}/\Delta TE2$; $\Delta TE4-5$	like $T^{cd}/\Delta TE2$, no additional defects	1. #1 n=28	Yes (~10 somites)	like $T/\Delta TE2$	1. #1n=1
$T2$ -/-			Yes	normal	1. #1 n=17 2. #2 n=16
$T^{cd}/\Delta TE1$	like $T^{cd}/+$, no additional defects	1. #1 n=10			
$T^{cd}/\Delta TE7$	like $T^{cd}/+$,no additional defects	1. #1 n=10			
$T^{cd}/\Delta TE2$; $\Delta TE4$	like $T^{cd}/\Delta TE2$,no additional defects	1. #1 n=16 2. #2 n=21		like $T^{cd}/\Delta TE2$,no additional defects	1. #1 n=4 2. #2 n=7

2.1.4 A 35 kb upstream regulatory region harbors enhancers required for trunk notochord specification and tail bud outgrowth

Embryos with a homozygous deletion of the upstream regulatory region (T^{UD}/T^{UD}) displayed a distinct phenotype that became apparent at around E9.5, when the tail bud is being formed. Whereas in wild type tail buds, the neural tube is closed and flanked by paraxial mesoderm, this organization was disrupted in T^{UD}/T^{UD}. Immunostaining for T and the neuroectodermal marker Sox2 in WT, T^{UD}/T^{UD} and T^{CD}/T^{CD} revealed several defects (Figure 2.1.5 A,C,E).

Firsty, in T^{UD}/T^{UD} and T^{CD}/T^{CD}, the trunk notochord, marked by high T expression in the WT was not formed (Figure 2.1.5 A,C,E,B′,D′,F′). While T expression in the axial midline was completely ablated, the expression in the caudal end of T^{UD}/T^{UD} persisted. In addition, excess neuroectoderm was produced in both T^{UD}/T^{UD} and T^{CD}/T^{CD} which led to an enlarged neural tube that did not close and enveloped the caudal truncation as a thick Sox2+ multilayered epithelium (Figure 2.1.5 B′′,D′′,F′′).

T^{CD}/T^{CD} embryos formed the first occipitocervical somites, but did not generate paraxial mesoderm after the forelimb bud (Figure 2.1.5 E′). In T^{UD}/T^{UD} mutants, all somites of the trunk were formed, although the somites became smaller towards the hindlimb bud and sporadically showed ectopic Sox2 expression, further substantiating an imbalance of neural vs. mesodermal tissue (Figure 2.1.5 C′). Lateral plate mesoderm formation was not affected in T^{UD}/T^{UD} and T^{CD}/T^{CD} (Figure 2.1.5 B′,D′,F′).

At E12.5, tail outgrowth was severely impaired in T^{UD}/T^{UD} embryos, which had a bulge of undifferentiated tissue instead of an elongated tail (Figure 2.1.5 G-I). Also, secondary effects such as neural tube closure defects, somite irregularities and a caudal regression were visible, which are malformations known from mutants lacking signaling from the notochord. Given this severe phenotype, T^{UD}/T^{UD} is most likely embryonically lethal and would not develop much further.

Taken together, T^{UD}/T^{UD} embryos failed to specify trunk notochord and displayed a loss of paraxial mesoderm formation and a concomitant expansion of neuroectoderm leading to axial arrest at the hindlimb bud.
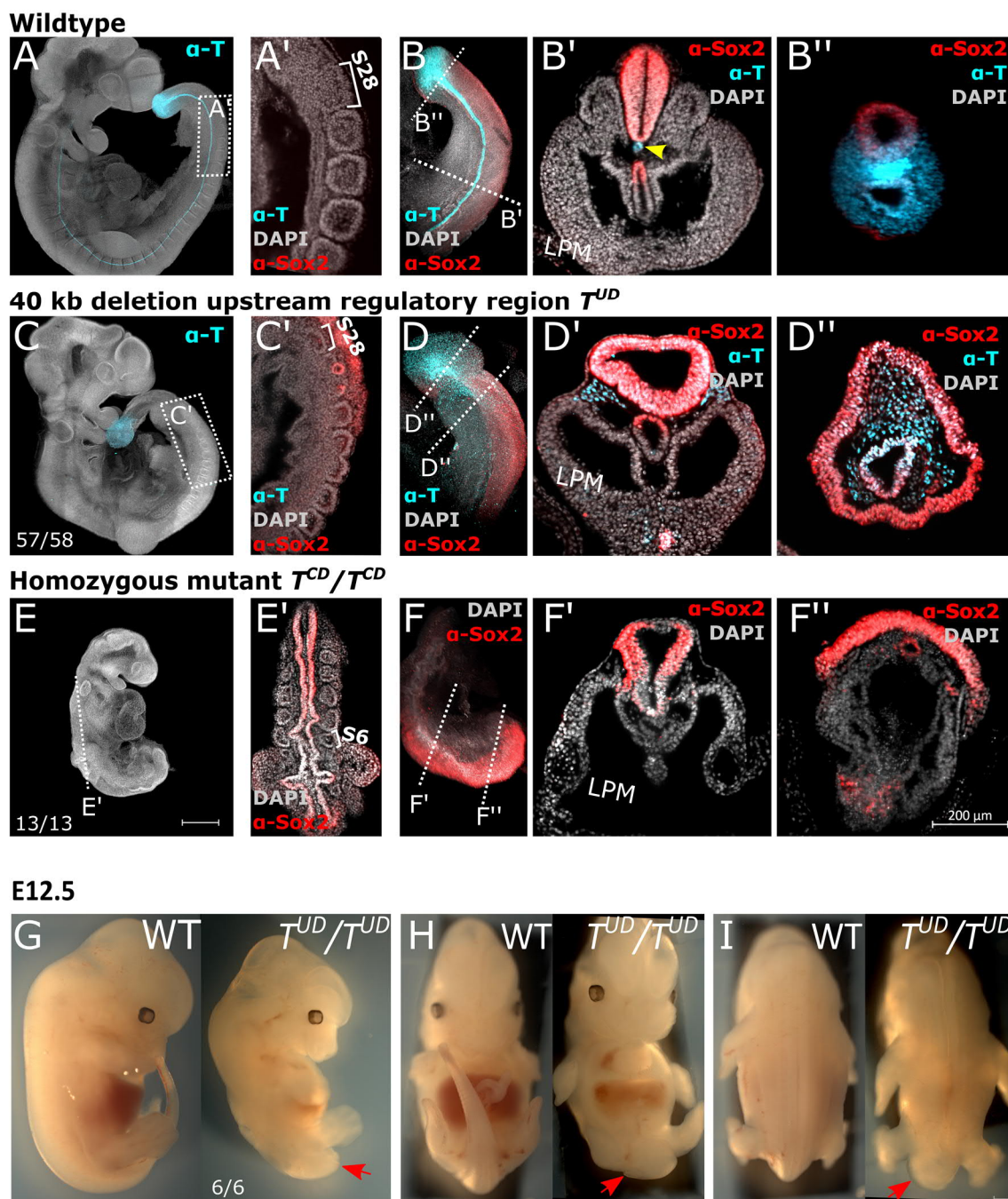
Figure 2.1.5 A 35 kb regulatory region contains elements required for notochord specification and tail outgrowth.

(A-F'') Immunostaining for T (cyan) and Sox2 (red) with nuclear DAPI (grey) staining in Wildtype, T^{UD}/T^{UD} and T^{CD}/T^{CD} E9-E9.75 embryos. Number of embryos displaying the shown genotype and total number of isolated embryos specified in the bottom left of panels depicting mutants. Maximum intensity projections of confocal stacks (A,C,E) and optical sections with light sheet microscopy (A'-B'', C'-D'', E'-F'') . Number of the last recently formed somite indicated (A', C', E'). The yellow arrow points out the position of the notochord (B'), which is not present in the mutants. LPM= Lateral Plate Mesoderm. (G-I) Lateral, ventral and dorsal views of E12.5 WT vs. T^{UD}/T^{UD} mouse embryos. Red arrows indicate the undifferentiated bulge that forms in the mutants instead of a tail.

2.1.5 *TE2* is a critical notochord enhancer of Brachyury

The deletion in T^{UD} spans multiple enhancer candidates (Figure 2.1.4) and induced a phenotype that is associated with roles of T in different sub-processes of axial elongation (Fig. 2.5). Thus, it is tempting to assume that the region contains several regulatory elements for notochord and NMP activity of Brachyury.

In order to test genomic fragments for their enhancer activity *in vivo*, an enhancer reporter assay was established. In short, the binding sites were cloned upstream of a minimal heat shock promoter and the coding sequence for Venus fluorescent protein. These donor constructs are flanked by different *loxP* sites, which allows for a targeted integration into a modified *Rosa26* locus via recombinase mediated cassette exchange (based on Vidigal et al., 2010). Transient Cre-recombinase expression facilitated recombination and successive stable selection. Single clones were genotyped by PCR spanning the *loxP* sites and expanded. Embryos were generated by diploid complementation assays and analyzed at E8.5 and E9.5.

TE2, a 1 kb element centered around the most prominent T binding peak showed strong reporter expression in the nascent notochord in the trunk and early tail bud (Figure 2.1.6). At E9.75, additional domains of T expression like the tail bud mesenchyme, the posterior neuroepithelium and the gut tube also contain single Venus+ cells in a more scattered fashion (Figure 2.1.6).

Overall, *TE2-HSP68-Venus* expression was strongest in the early notochordal lineage, where presumably the specification of notochord progenitors occurs, making *TE2* a promising notochord enhancer candidate.
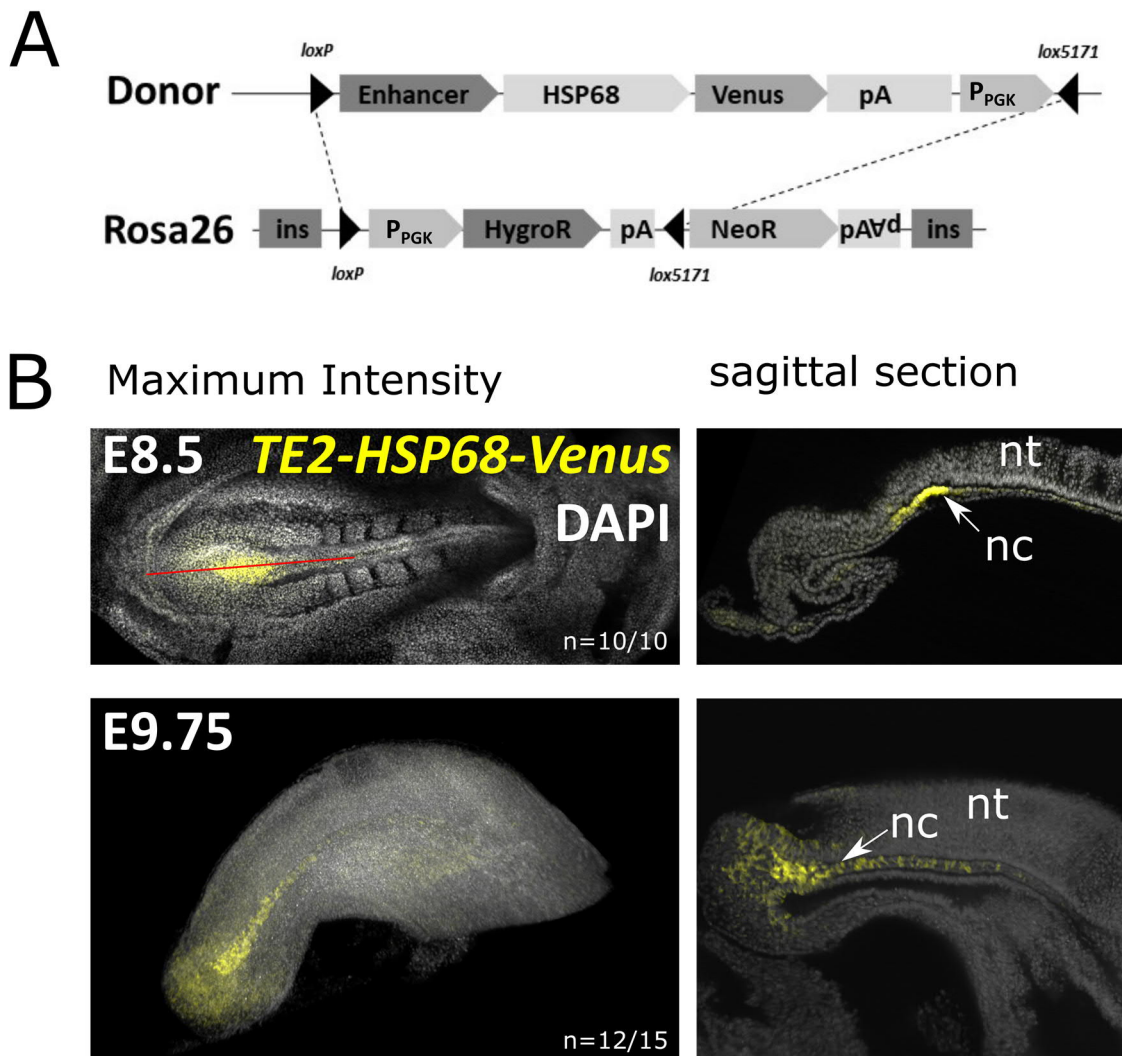
Figure 2.1.6 *TE2* is an enhancer with activity in the notochord

(A) Recombinase mediated cassette exchange cloning strategy based on Vidigal et al. (2010). Enhancer candidates were cloned upstream of the HSP68 minimal promoter. pA = polyadenylation. P_{PGK} = PGK promoter. ins = insulator. HygroR = Hygromycin resistance. NeoR = Neomycin resistance. (B) *TE2* driven Venus expression demonstrates enhancer activity in nascent axial mesoderm. Nuclei stained with DAPI (grey). Left: Maximum intensity projections of confocal microscopy. Right: sagittal optical sections; Light sheet acquisitions. nc = notochord. nt = neural tube. Number of embryos analyzed at the corresponding stage that showed the depicted expression pattern specified at the bottom right. At E9.75, no reporter signals was detected in few outliers, possibly due to chimerism in diploid complementation assays.

With the activity of the *TE2* element in early notochord demonstrated, the question whether there is a functional role for *TE2* was addressed employing CRISPR/Cas9 to generate knockouts in both wild type and $T^{cd}/+$ Noto^{mC} mESCs (Figure 2.1.4). Mouse embryos were generated by tetraploid aggregations and analyzed between E7.5-E12.5.

Again, T protein was visualized by immunostaining at trunk to tail transition (E9.75) and in the developing tail at E11.5 (Figure 2.1.7).
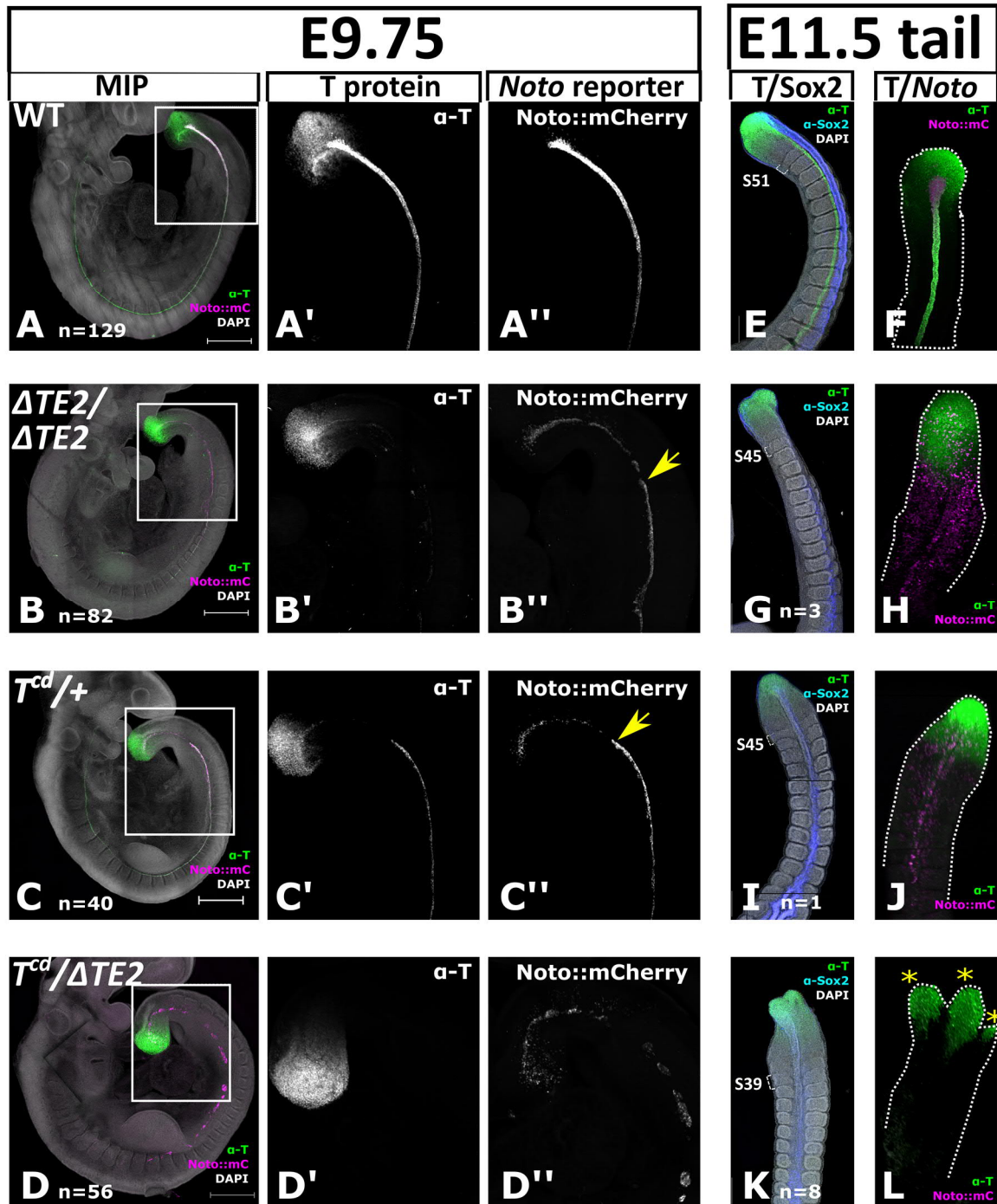
Figure 2.1.7 *TE2* is a critical notochord enhancer of *Brachyury*

(A-D) Maximum intensity projections of E9.75 embryos with Noto^mC reporter signal, immunostaining for T (green) and DAPI nuclear staining (grey). Total number of embryos isolated at E9.75 (A-D) and E11.5 (G,I,K) specified. E11.5 tails were imaged by confocal microscopy (E,G,I,K) or Light sheet (F,H,J,L) and show different specimen. Scale bar=500µm. The square indicates the area magnified in single channels (A'-D''). Yellow arrow points at the disruption Noto^mC+ notochord progenitors. (E,G,I,K) Maximum intensity projections of E11.5 tails with immunostaining for T (green), Sox2 (blue) and DAPI (grey). (F,H,J,L) Maximum intensity projections of E11.5 tails with immunostaining for T (green) and the Noto^mC reporter signal (magenta). Asterisks indicate bifurcations in the tail.
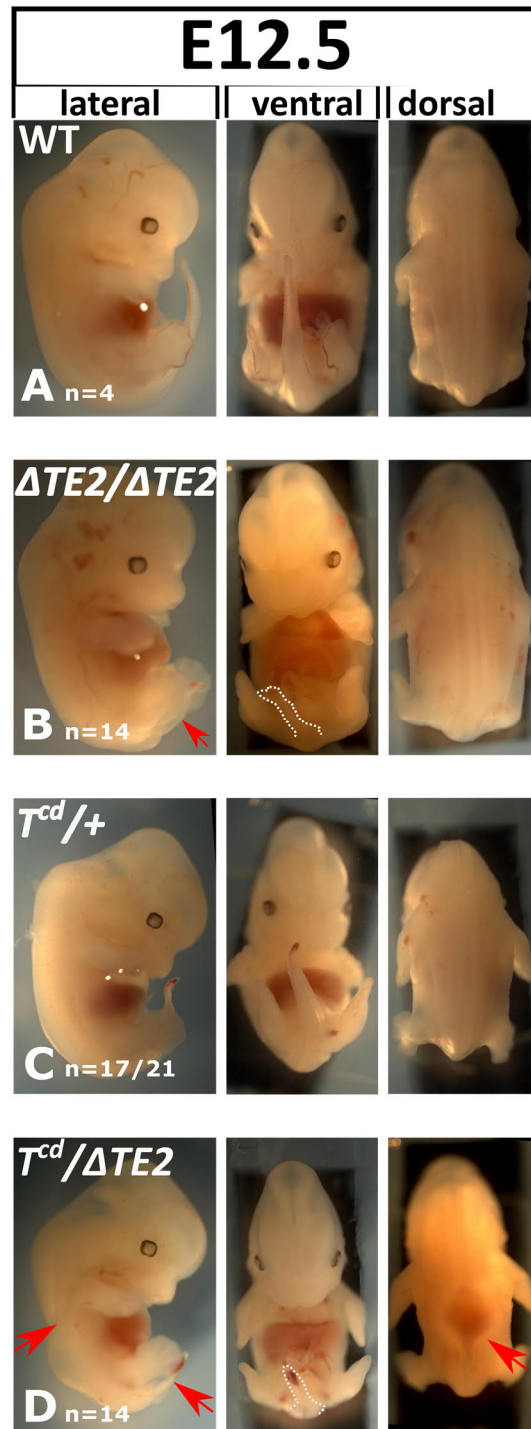
Figure 2.1.8 *TE2* is an essential notochord enhancer required for tail development

(A-D) Lateral, ventral and dorsal views of E12.5 wild type and mutant embryos. Red arrows indicate developmental defects. Number of embryos that displayed the represented phenotype shown in each panel. All mutants displayed identical morphological features, except for T^{cd}/+ which had a variable tail phenotype (17/21 tailless and 4/21 short tail).

Embryos with a homozygous deletion of the *TE2* enhancer displayed abnormalities starting at E9.75, where Noto^{mC} reporter signal in the differentiated rod-like structure of the notochord was disrupted at around hindlimb levels (Figure 2.1.7 B''). T expression in the more anterior trunk notochord was weaker than in the wild type and depleted from more posterior Noto^{mC}+ cells emerging from the caudal growth zone. In this region including the ectopically distributed Noto^{mC}+ cells, T expression was not affected by the deletion (Figure 2.1.7 B'). After trunk to tail transition, the Noto^{mC}+ cells were still generated, but not specified and therefore scattered throughout somites and neural tube (Figure 2.1.7 H). At E12.5, it was evident that the tail bud initially grew out and formed the first 15-20 post-anal somites, but did not develop further (Figure 2.1.8 B). At this stage, the tails displayed hemorrhagic lesions and were on the verge of degeneration into a filament (Figure 2.1.8 B).

To see whether *TE2* is also involved in earlier notochord development, an additional knock-out mutant was generated in the heterozygous *T^{cd}/+* background. *T+/-* mice are viable, but have a shortened tail of variable length (Gluecksohn-Schoenheimer, 1938). To generate the *T^{cd}/+* mutant in this study, a CRISPR/Cas9 approach was used to induce a heterozygous deletion of the complete *T* locus. This modification resulted in a phenotype that was largely similar to *ΔTE2/ΔTE2* (Figure 2.1.7 C). However, some differences could be observed in *T^{cd}/+* compared to *ΔTE2/ΔTE2*, such as more variability in tail length and higher T levels in Noto^{mC}+ cells (Figure 2.1.7 B', C'; 2.8 B, C). In *T^{cd}/ΔTE2* mutants, in which the only copy of the *TE2* enhancer is deleted, trunk notochord formation was entirely impaired (Figure 2.1.7 D). Even though clusters of Noto^{mC}+ cells were found in the midline, these did not express T and were mainly located in the dorsal hindgut. Because of the stability of the H2B-mCherry reporter, the signal in the endoderm was most likely a remnant from an earlier notochord progenitor state of cells that did not complete differentiation into the notochord lineage.

No Noto^{mc}+ cells could be detected in the tail (Figure 2.1.7 L), which initially grew out, but did not develop further after formation of about 10 somites (Figure 2.1.8 D). The medio-lateral organization appeared to be lost and the posterior growth zone branched into multiple T+ poles (Figure 2.1.7 L). In addition to the tail phenotype, the E12.5 *T^{cd}/ΔTE2* embryos showed several defects that were previously described in other notochord mutants and likely caused by the loss of Shh signaling from the midline: Neural tube closure defects, posterior regression syndrome and irregular somites (Chiang et al., 1996; Pennimpede et al., 2012; Zhu et al.,

2016). A significant amount of necrotic tissue in the posterior proportion of the trunk indicated that this mutant would likely be embryonically lethal.

In conclusion, these data show that *TE2* is a notochord enhancer of *Brachyury* that is active in the nascent notochord and essential for tail notochord specification and tail development.

2.1.6   Tail outgrowth is not dependent on notochord progenitors

One of the aims of this study was to assess whether the posterior notochord cells are required to maintain the progenitor cells that produce the cellular material for AP axis elongation. In the $T^{cd}/\Delta TE2$ and T^{UD}/T^{UD} mutant, *Brachyury* expression was not detected in Noto^{mC}+ cells anterior to the caudal end. Antibody staining for Foxa2 demonstrated that while the anterior head process was not affected, both trunk notochord and floor plate were truncated at around the level of the forelimb buds (Supplementary Figure 4). This showed that the remnant Noto^{mC}+ cells did not have the capacity to induce floor plate and had most likely acquired endodermal fate. Both mutants displayed different tail phenotypes: T^{UD}/T^{UD} produced an undifferentiated bulge and in $T^{cd}/\Delta TE2$, a bona fide tail with somites, gut and neural tube initially grew out.

In order to test if the different tail morphologies could be explained by a loss of Noto+ cells adjacent to the NMP domain, E9.75 tail buds were analyzed in more detail using Lighsheet microscopy (Figure 2.1.9). In WT, Noto^{mC}+ cells were those with the highest T immunofluorescence signal. This population was smaller and disrupted in $T^{cd}/+$, but those Noto^{mC}+ cells that emerged from the caudal growth zone maintained the expression of T. This was not the case in $\Delta TE2/\Delta TE2$ where T expression was reduced in the differentiated Noto^{mC}+ cells and Noto^{mC}+ were dispersed and did not form a defined structure. In $T^{cd}/\Delta TE2$ tail buds, no or very few Noto^{mC}+ cells could be detected, although the overall morphology of the tail bud was normal (Figure 2.1.9). In contrast, T^{UD}/T^{UD} tail buds, which were disorganized and not capable to generate a tail, contained few Noto^{mC}+ cells, which were dispersed and did not maintain T expression. In addition, the posterior growth zone persisted in the tails of $T^{cd}/\Delta TE2$, even though it eventually split in multiple poles with T/Sox2 co-expression (Figure 2.1.8 K), which is a signature of NMPs.

Based on these data, a dependency of NMPs on notochord progenitors cannot be conclusively demonstrated. NMPs seem to be capable to self-renew and independently give rise to neural and mesodermal primordia. These however do not mature without organizing signals from the notochord.
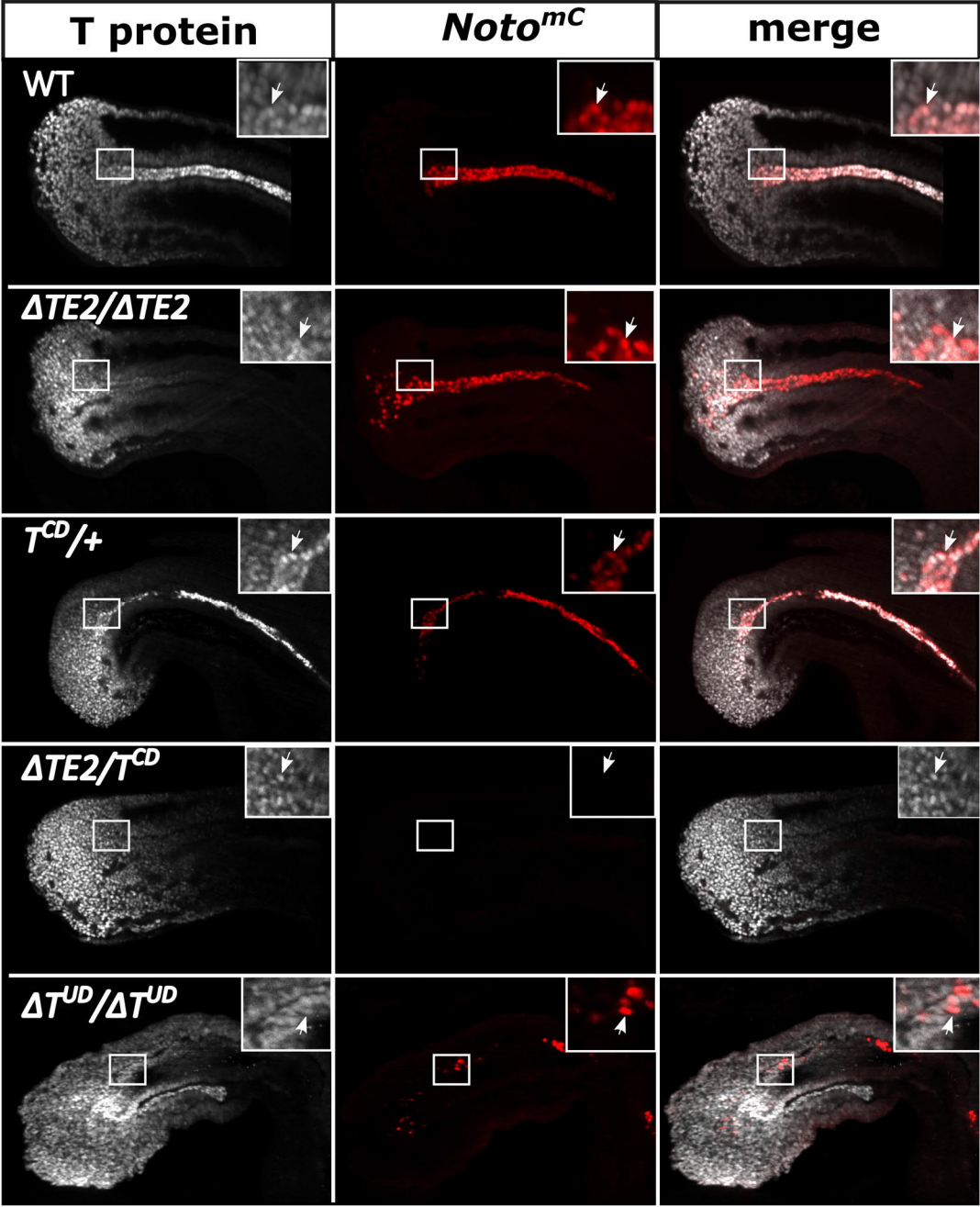
Figure 2.1.9 Notochord progenitors in the tail bud

E9.75 tail buds of different mutants. Optical sections of Light sheet acquisitions. Immunostaining for T (white) and Noto^mC+ reporter signal (red). Arrows indicate single T and Noto^mC+ cells.

2.1.7 Apoptosis in tails without notochord

After emanating from the posterior growth zone in an almost wild-type arrangement, somites and neural tube of the *TE2* enhancer mutant tails did not develop further. Reports of other tail notochord mutants have shown cell death causing regression of the tail (Pennimpede et al., 2012). To explore whether apoptosis was induced in the *ΔTE2/ΔTE2* and *T^{cd}/ΔTE2* mutants, immunostaining for cleaved Caspase 3 (Casp3) was performed (Figure 2.1.10).

Indeed, clusters of cells in neural tube and somites, as well as in the caudal growth zone displayed increased Casp3 expression (Figure 2.1.10). This suggests that without signals from the midline, the transient axial and paraxial tissues induced apoptosis shortly after their specification.
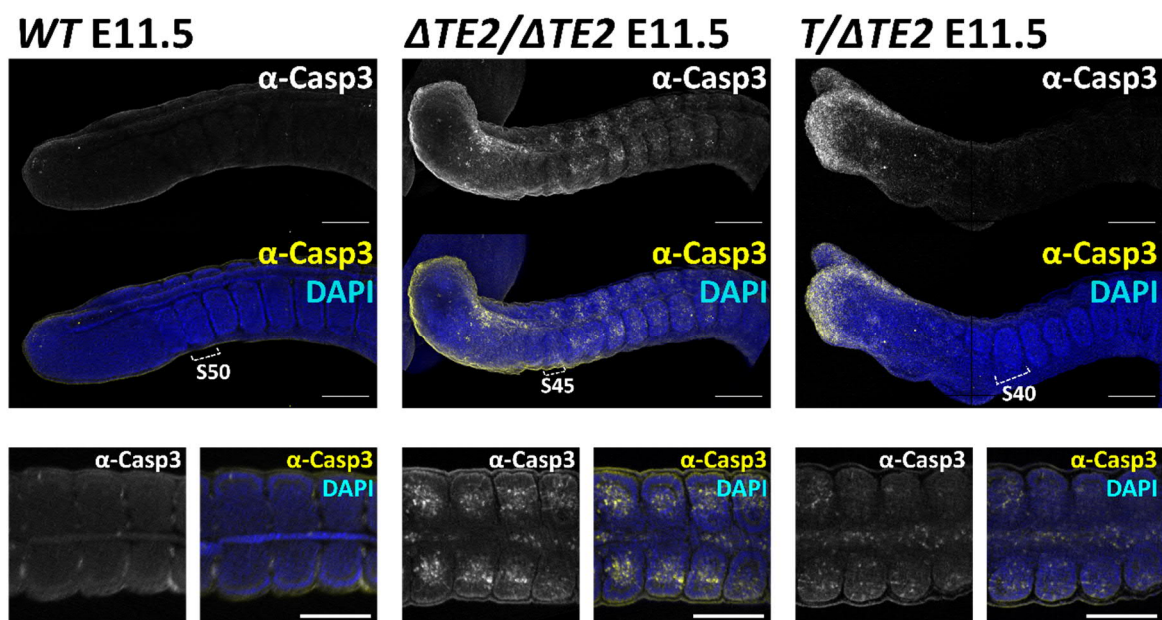
Figure 2.1.10 Apoptosis in *TE2* mutant tails

Top: Maximum intensity projections of Wild type (WT) and mutant tails. Casp3 immunofluorescence as a single channel and overlap with DAPI nuclear staining. Last formed somite number indicated. Bottom: Light sheet acquisitions of optical sections. Scale bar=500µm.

2.1.8 Is the tail interaction factor (*tct*) a notochord enhancer of *Brachyury*?

The *t* haplotypes are variants of chromosome 17 of the house mouse carrying a set of alleles with severe effects on inheritance and embryonic development (Silver, 1985). The haplotypes have four large inversions, which suppress recombination with the wild-type homolog and contain genes that can hijack spermatogenesis, leading to distorted transmission ratio in their favor and therefore fairly high occurrence in wild populations (Herrmann et al., 1999; Schimenti, 2000; Silver, 1993). Numerous *t* haplotypes have been discovered both in the wild and the laboratory (Bennett, 1975).

A common feature of t haplotypes, which also lead to their discovery almost a century ago, is the interaction of the *tct* (*t complex tail interaction*) factor with *Brachyury T* mutations (Dobrovolskaia-Zavadskaia and Kobozieff, 1932). Whereas *T/+* heterozygotes have a short tail of variable length, *T/t* mice are always tailless (Chesley and Dunn, 1936). Both *t/+* and *t/t* are normal-tailed. Therefore, *tct* enhances the *T* phenotype and is genetically linked to, but not identical to the *T* gene (Justice and Bode, 1988).

In this work, the effect of *tct* on notochord development was analyzed by microscopy and transcriptome profiling. Since in today's understanding of transcriptional control, *tct* is likely to be a *cis*-regulatory element of *T*, it was explored whether *tct* is identical to the *TE2* enhancer. The *tw5* allele carries the complete set of mutations on chromosome 17 that have been described for *t* haplotypes and was therefore chosen as the genetic background for these experiments (Dunn and Suckling, 1956; Lyon and Meredith, 1964).

In order to make the *T/t* genetic background accessible to genomic modifications and for the production of *T/t* embryos via tetraploid aggregations, a *T/tw5* mESC line was established by Lars Wittler and Manuela Scholze-Wittler (MPI-MG, Berlin). The tailless phenotype was validated in embryos from tetraploid complementation assays. At E12.5, the tail regressed to a filament recapitulating the phenotype observed in *T/+* x *tw5/+* crossings (Figure 2.1.11 C; Dunn and Suckling, 1956).

To enable monitoring of notochord development, a *Noto::H2B-mCherry* reporter BAC (Figure 2.1.1) was randomly integrated into the *T/tw5* mESC line. Embryos were generated, isolated at E9.5 during trunk to tail transition and analyzed by antibody staining for T (Figure 2.1.11).

40
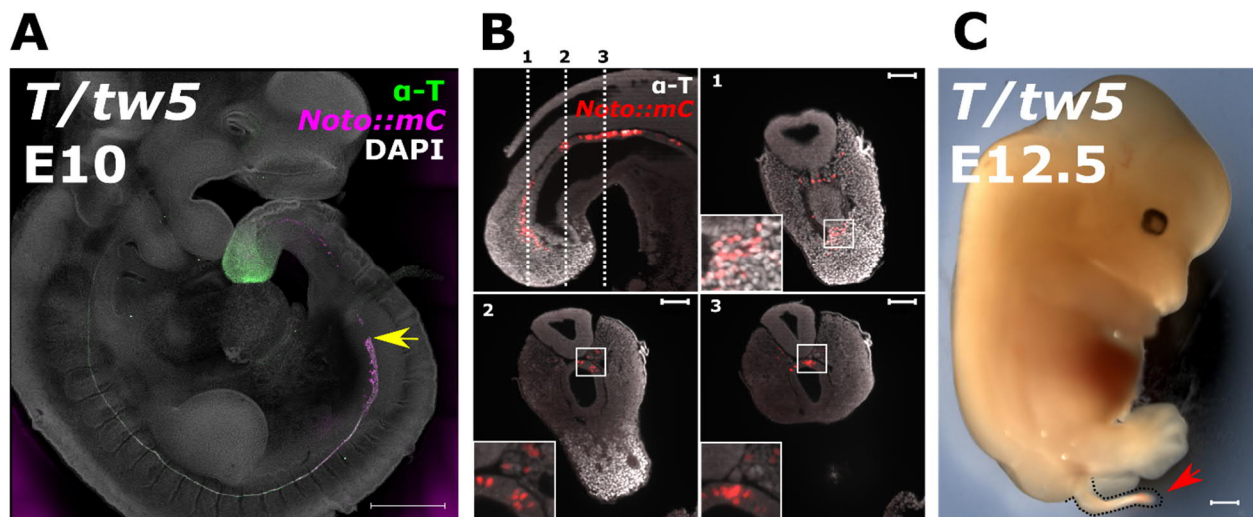
Figure 2.1.11 Disruption of tail notochord in *T/tw5* embryos

(A) E10.0 embryo. Maximum intensity projection of confocal microscopy stacks. Noto^{mC} reporter signal (magenta), T immunofluorescence (green). Nuclei stained with DAPI (grey). Scale bar=500µm. Yellow arrow indicates the disruption of trunk notochord at hindlimb level. Wild type in Figure 2.1.8 A. (B) E9.75 tail bud. Optical sections of Light sheet microscopy. Noto^{mC} reporter signal (red), T immunofluorescence (grey). Dotted bars 1-3 in top left image indicate the sectional planes 1-3. Scale bar=200µm. (C) E12.5 mouse embryo generated by tetraploid complementation assays. Red arrow indicates the tail phenotype. Wild type in Figure 2.1.8 A.

Similar to other *T* mutants, the tail phenotype was preceded by a notochord defect that becomes evident in *T/tw5* at E9.75. At the axial level where the hindlimb bud forms (somite 24-28), the notochord lost its discrete morphology and broadened (Figure 2.1.11 A). More posterior, Noto^{mC}+ cells were scattered and mislocalized in somites and dorsal hindgut (Figure 2.1.11 B). T protein expression in the caudal growth zone and the anterior trunk notochord was not affected, but not detected from Noto^{mC}+ that emerged from the caudal end posterior to the hindlimb level (Figure 2.1.11 A, B).

Since the caudal Noto^{mC}+ cells should represent the most nascent notochord and notochord precursors, their transcriptome was profiled to characterize this population and thereby assess notochord development in *T/tw5*.

To this end, WT and *T/tw5* embryos carrying the Noto^{mC} reporter were isolated at E9.5 and five caudal ends per genotype were dissected at the first visible somite border. Noto^{mC}+ cells were purified by fluorescence activated cell sorting (FACS) and the samples subjected to bulk RNA-Seq analysis (Figure 2.1.12).

Fragments per kilobase of transcript per million of mapped reads (FPKM) expression values were quantified using cufflinks (Trapnell et al., 2013) and compared, revealing 3143 genes that were dysregulated in *T/tw5* Noto[mC]+ cells by at least twofold (Figure 2.1.12).

Among these were a broad number of notochord signature genes including *Noto*, *Shh*, *Chrd*, *T*, *Foxa1*, *Bicc1*, *Foxa2* and *Sox9* (Besnard et al., 2004; Echelard et al., 1993; Klingensmith et al., 1999; Plouhinec et al., 2004; Tamplin et al., 2011; Wilkinson et al., 1990). Strikingly, even though the cells were sorted using a *Noto* reporter and no significant decrease in the proportion of Noto[mC]+ cells was detected, *Noto* was among the first eight of the most downregulated genes, indicating that the expression level of the endogenous *Noto* transcript reached only a fraction of the wild-type expression level.

Antibody staining showed that T protein was expressed within the scattered Noto[mC]+ cells in the caudal growth zone of *T/tw5* mutants, but was not maintained in the Noto[mC]+ remnants in the more differentiated tissues. Accordingly, *T* transcript was clearly detectable (FPKM=69) in the Noto[mC]+ cells isolated from *T/tw5* caudal ends, but downregulated by more than tenfold compared to the wildtype (FPKM=759).

The posterior Noto[mC]+ cells that did not form a discrete notochord structure in *T/tw5* were distributed throughout other tissues, mainly in the dorsal gut tube as well as presomitic and paraxial mesoderm (Figure 2.1.11 B). This fate transition was also reflected in the list of upregulated genes (Figure 2.1.12 B). Their transcriptome was on the one hand enriched for markers for paraxial mesoderm like *Tbx6*, *Msgn1*, *Dll1*, *Meox2*, *Mesp2* (Bettenhausen et al., 1995; Chapman et al., 1996; Chenevix-Trench et al., 1992; Saga et al., 1997; Yoon et al., 2000) and on the other for hindgut markers like *Cldn3*, *Cldn7* and *Cldn8* (Anderson et al., 2008).

Taken together, localization and expression profiling of Noto[mC]+ cells suggests that the tailless phenotype in *T/tw5* mutants is caused by defects in tail notochord specification. Notochord precursors were likely generated but failed to maintain the dosage of T required to differentiate into notochord. This further substantiates the notion that *tct* is a mutated version of a notochord enhancer of *Brachyury* rather than an independent gene.
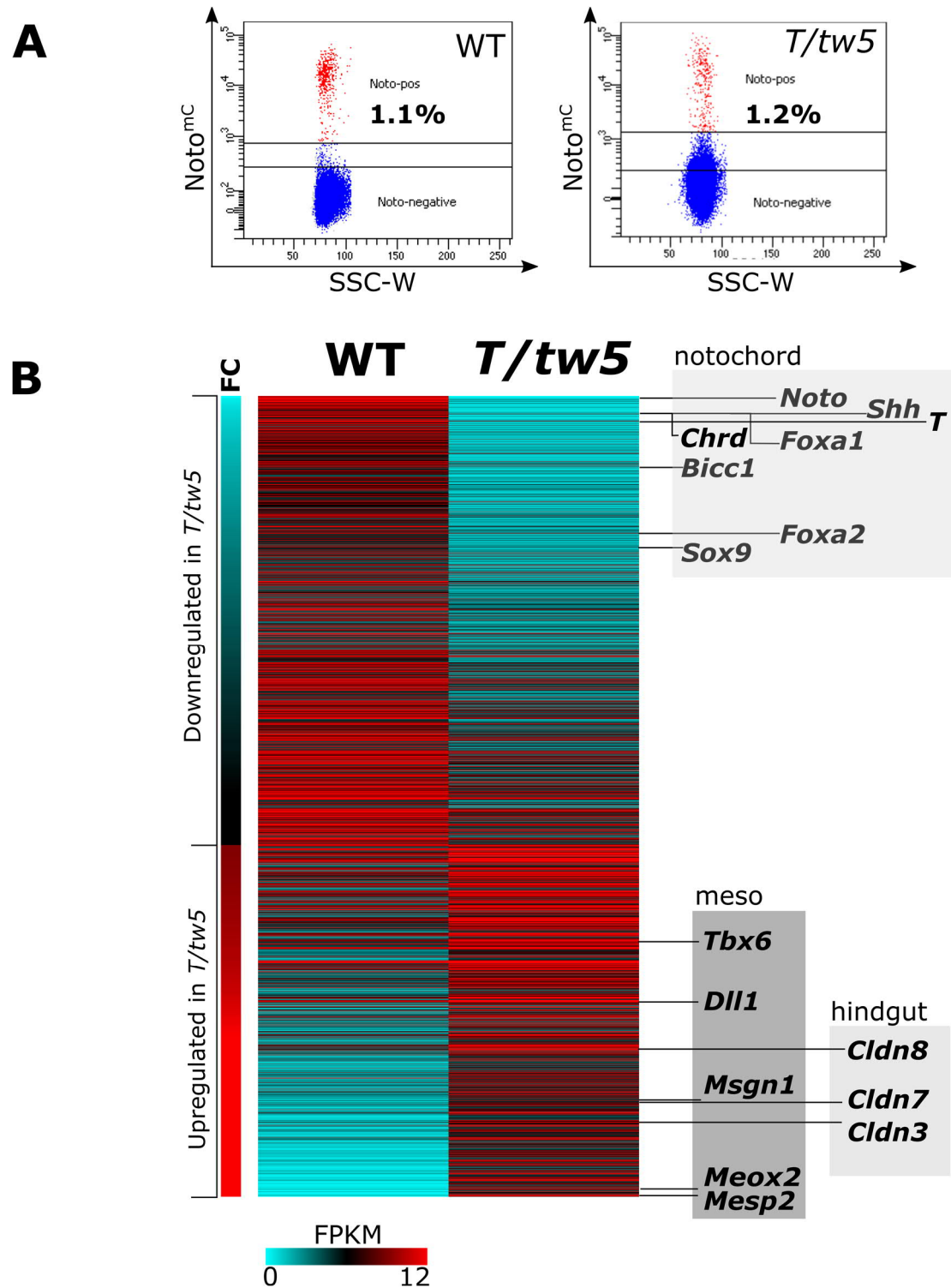
Figure 2.1.12 Differentially expressed genes in T/tw5 vs WT Noto^{mC}+ cells

(A) FACS plots of sorted cells. SSC-W = side scatter pulse width. Material from 5 embryos per genotype was sorted. Gating was slightly adapted to the higher background in the *T/tw5* sample. (B) Heatmap showing ranked FPKM expression values of 3143 dysregulated genes in *T/tw5* vs wild-type (WT) embryos. Only genes with an FPKM>2 and with at least twofold up- or downregulation included. Genes indicated on the right represent important lineage markers for the notochord (top), paraxial mesoderm (bottom left) and endoderm (bottom right).

The *T/tw5* phenotype was remarkably similar to the *ΔTE2/ΔTE2* enhancer mutant (Figure 2.1.7) which also had defects from a specific loss of T expression in the notochord posterior to the hindlimb bud. This suggests that the insights from the *T/tw5* transcriptome could be transferred to *ΔTE2/ΔTE2* and in both mutants, transient Noto^mC+ cells were not properly specified and acquired a different fate. It further implies that after trunk to tail transition, a higher dosage of T is required for notochord specification.

The exact genomic location of *tct* remains unknown, however it has been mapped to the region closer to the centromere on chromosome 17, within the t complex and not separable from the *T* locus (Fox et al., 1985). The essential notochord enhancer *TE2* presented in this work is located in this region, which makes *TE2* a candidate for *tct*. Deletion of *TE2* in the $T^{cd}/+$ heterozygous background displayed a more severe phenotype than *T/tw5* (Figure 1.7). Thus, if *TE2* and *tct* are identical, the *tw5* variant of *TE2* should not be absent but rather perturbed. In order to find mutations in the *tw5 TE2* enhancer, the element was analyzed by Sanger sequencing and mapped against the BL6 wild-type sequence. Indeed, the *tw5* haplotype *TE2* core region has 3 point mutations and one insertion which introduce a novel Myc::Max binding E-box motif (*5′-CACGTG-3′*) in proximity to the T-Box motifs (Figure 2.1.13 A). This motif as well as the nearby T motif might be bound by Mga, a T-box factor, which can form a heterodimer with Max and recruit the PRC1.6 complex, which was postulated to function as a repressive mechanism to silence T-responsive genes (Hurlin et al., 1999; Stielow et al., 2018).

Therefore, the point mutation in *tw5 TE2* suggest a scenario, in which the Mga-Max dimer can bind to the novel E-box motif and compete with T for the T-box motif (Figure 2.1.13 A). This could inhibit *TE2* activity and create a hypomorph version of this critical notochord enhancer.

To explore this hypothesis, a CRISPR/Cas9 based oligo repair strategy was persued to replace the four mutations with the BL6 wild-type sequence (Figure 2.1.13 B). The *T/tw5* mESC line carrying the Noto^mC reporter was used and transfected with a vector for guide RNA expression and a 200bp single stranded oligonucleotide as a template for the repair. For screening of positive clones, a *tw5* specific *PmlI* site was utilized for restriction digest of the PCR fragments. Clones were validated by Sanger sequencing. Only one clone that had no additional insertions or deletions was identified in 288 screened candidates. In that clone, the two 5′ mutations in the *tw5 TE2* enhancer were replaced with the BL6 sequence, removing the Myc::Max E-box motif (Figure 2.1.13).

44

To test if the repair of this site would restore T activity in the notochord and rescue the *T/tw5* tailless phenotype to a short tailed *T +/-* morphology, mouse embryos were generated from the clone with repaired 5′ mutations and isolated at E12.5. The tailless phenotype was still displayed (Supplementary Figure 5), suggesting that either *tct* and *TE2* are different alleles or that restoring the two 5′ mutations at the core of the enhancer was not sufficient for a rescue, which would imply that other mutations still play a role.
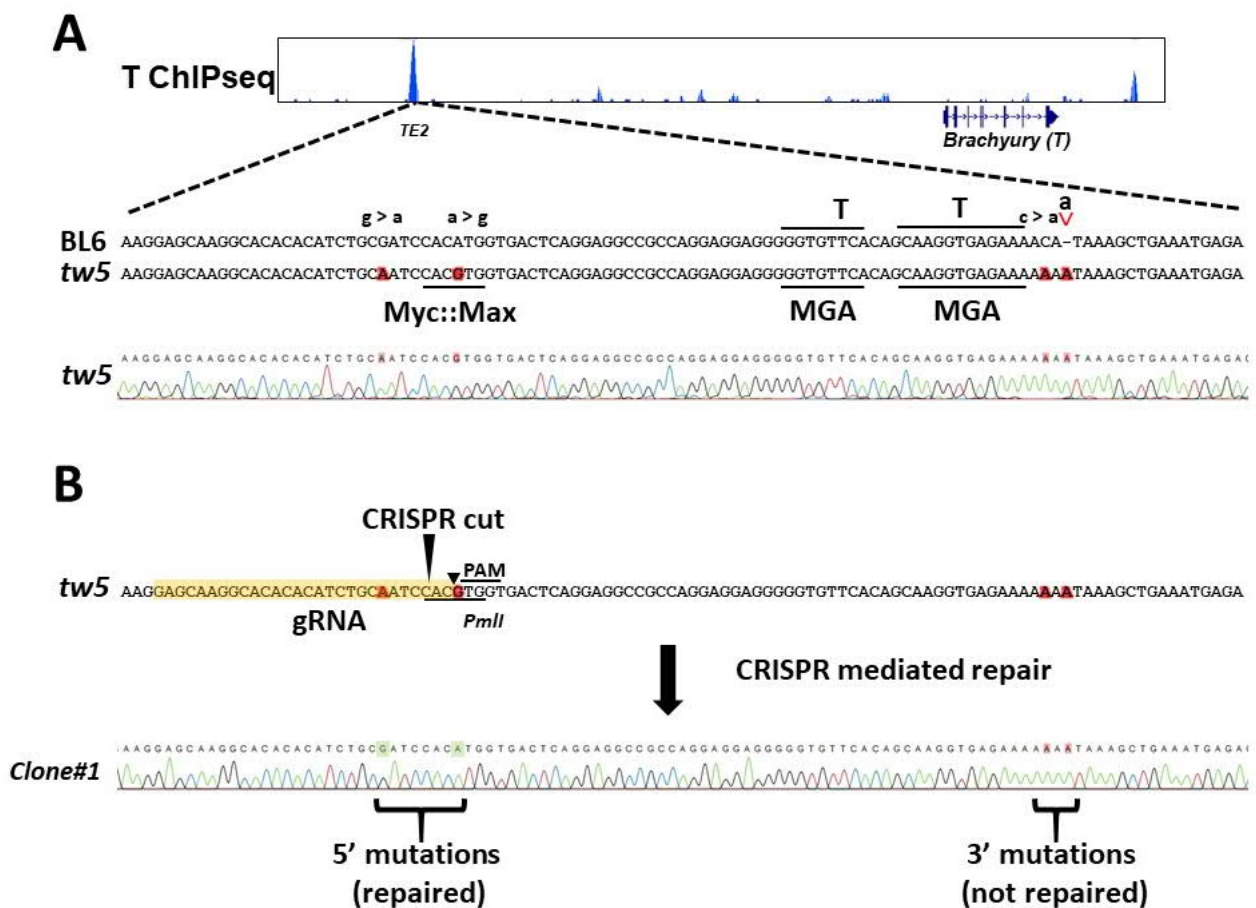
Figure 2.1.13 *TE2* is a candidate for the tail interaction factor

(A) ChIP-Seq track of T binding on the T locus to indicate relative enhancer position. Sequence at the center of the binding peak shown. Alignment of BL6 wild-type and *tw5* haplotype reveals 3 point mutations and one adenine insertion. Transcription factor binding motifs annotated: 5′ novel Myc::Max E-box motif (CACGTG), 3′ T-box motifs not affected by the mutations. (B) CRISPR-based rescue strategy. Guide RNA sequence highlighted in yellow. Large triangle indicates the Cas9 cutting site -3bp from the Protospacer Adjacent Motif (PAM). Small triangle points at the cut of the (*5′-CACGTG-3′*) *PmlI* restriction site used for screening PCR fragments for the sequence repair. Sanger sequencing alignment of the single positive clone at the bottom.

2.1.9   Disruption of the T2 open reading frame does not affect axis extension

The regulatory region upstream of *T* contains the rodent specific *T2* gene, which according to its annotation encodes 3 spliced transcript variants of 2559 bp, 1755 bp and 1080 bp (Figure 2.1.14 A; MGI-ID:104658). While the latter transcript is eliminated by nonsense mediated decay, the first two have 297 aa and 393 aa protein coding sequences (MGI-ID:104658). So far, no T2 protein has been characterized or detected.

The *T2* open reading frame is disrupted by a 200kb insertion in the *T^{Bob}* mutant, which displays a notochord development and tail outgrowth phenotype (Rennebeck et al., 1995). Therefore, when the *T2* cDNA was initially discovered, it was postulated to code for a factor with an essential function in axial development and termed *T2*, implicating a second *Brachyury* gene (Rennebeck et al., 1998). Since then, the authors have arrived at a different hypothesis and have suggested that the *T^{Bob}* mutant phenotype is induced by the disruption of a transcriptional enhancer of *Brachyury* rather than the *T2* gene (Wu et al., 2007).

Some mutants presented in this work also have deletions of *T2* exons (Figure 2.1.14 A; Figure 2.1.4) hence it is necessary for the interpretation of the results to clarify the function of *T2* in embryogenesis. To this end, CRISPR/Cas9 was employed to disrupt the *T2* protein coding sequence. Exon 2 of the *TE2.201* (ENSMUST00000163578.7) transcript was chosen, because it is present in all splice variants, it is not expected to interfere with *TE2* enhancer activity and in contrast to most exons downstream, it does not contain a methionine residue that could serve as an alternative start codon and restore the open reading frame in case of exon skipping.

The exon was targeted in the Noto^{mC} mESC line used as the parental line for all mutants presented in this study using a single CRISPR guide RNA. Single colonies were picked and screened for the deletion by PCR fragment restriction digest. PCR fragments were further subcloned into plasmids to distinguish between the different alleles and confirm homozygous frameshift mutants. Two clones were identified in which the protein coding sequence was disrupted by indels that cause a frameshift and translation running into several premature stop codons (Figure 2.1.14 B).

From these two clones, Embryos were generated by tetraploid aggregation and isolated at E12.5. In contrast to the *ΔTE2/ΔTE2* and *T^{UD}/T^{UD}* mutants, which displayed clear axis phenotypes at this stage, all *T2 -/-* specimen analyzed displayed normal development (Clone d2 n=17/17; Clone d3 n=16/16; Supplementary Figure 5). In addition, the *ΔTE1* and *ΔTE5* deletions, in which *T2* Exons 2 and 8 are affected, respectively, did not recapitulate the *ΔTE2* defects (Table 1; Supplementary Figure 2-3).

Since no T2 protein product has been identified so far and the *T2* mRNA transcript is hard to detect in most embryonic cell types, it was difficult to verify the knock out. However, the fact that neither deletions of other exons nor disruption of the *T2* open reading frame phenocopy the *ΔTE2/ΔTE2* mutant, strongly implies that *TE2* is a bona fide enhancer.
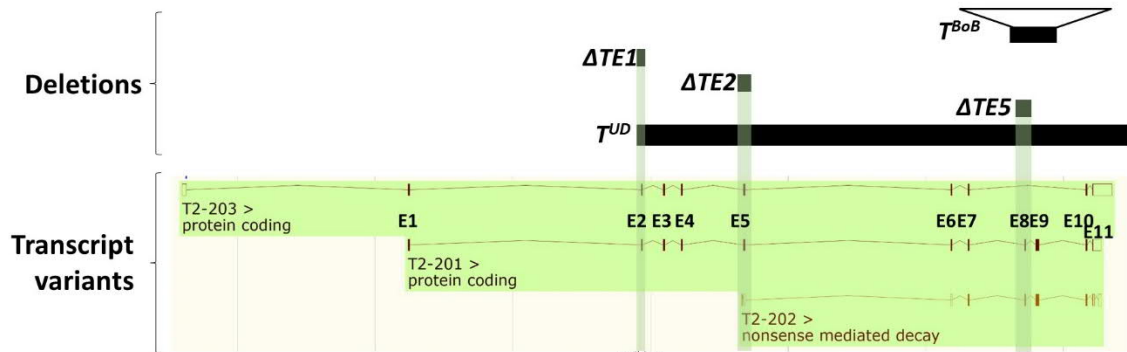
Figure 2.1.14 *T2* Open Reading Frame disruption using CRISPR/Cas9

(A) *T2* transcript variants annotated in the ensemble database. Location: Chromosome 17: 8,355,992-8,423,513 forward strand. (GRCm38:CM001010.2). Top: Mutant genotypes covering exons of the transcript. (B) Knock-out strategy using a single CRISPR guide RNA to introduce indels in Exon 2. Representative genotyping of one clone. Top: Sanger sequencing of subcloned PCR fragments. Because a hybrid cell line is used, the BL6 and SV129 alleles can be discriminated via short nucleotide polymorphisms (SNPs). Bottom: Amino acid sequence alignment. Frameshift sequence highlighted in red, stop codons indicated by red asterisks.

2.2 *TE3* is a nascent mesoderm and hindgut enhancer of *Brachyury* and required for the completion of tail development

The second most prominent T binding site revealed by ChIP-Seq is found 12 kb downstream of *TE2* and 26 kb upstream of *T* and was termed *TE3.* Binding of T at *TE3* is more pronounced in *in vitro* derived NMPs and the T peak is flanked by two Sox2 binding peaks, suggesting *TE3* could be a site of antagonistic binding found in NMPs (Figure 2.1.3).
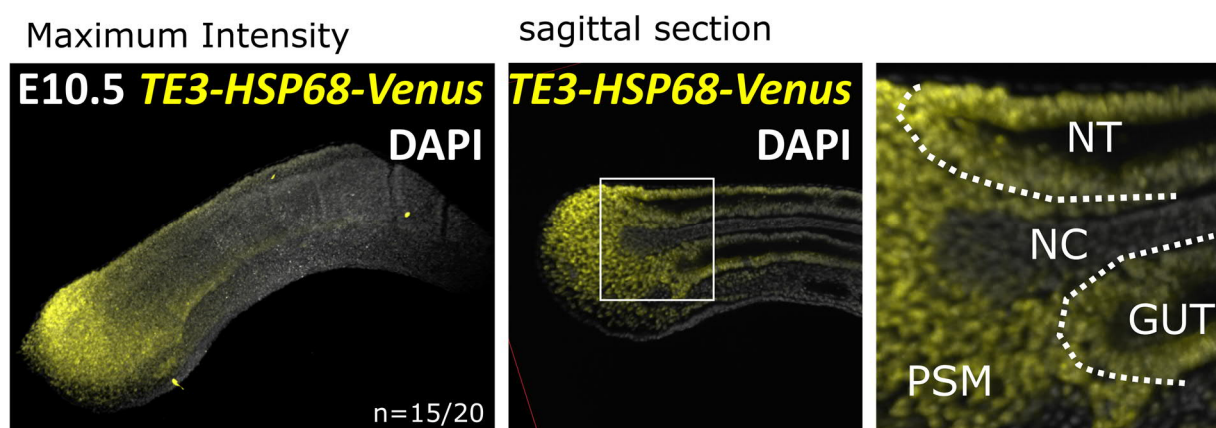
Figure 2.2.1 Reporter activity of the *TE3* enhancer.

E10.5 tail bud. Nuclei counterstained using DAPI (grey). Light sheet imaging. Left: Maximum intensity projection. Number of embryos that displayed the shown expression pattern specified in the bottom right corner. 5/20 embryos did not show any reporter signal, which could be due to chimerism in diploid aggregation assays. Right: Optical sagittal section. NT = neural tube. NC = notochord. GUT = hindgut tube. PSM= presomitic mesoderm.

To test the potential of the element to drive reporter gene expression, it was cloned and subjected to the activity assay described in Figure 1.2.6. No reporter signal was detected between E8.5 and E9.

However, after trunk to tail transition, Venus expression was detected in multiple domains of T expression in the tail bud, including the presomitic mesoderm, the posterior neural plate and the hindgut roof. As opposed to *TE2*, no signal was detected in the notochord.
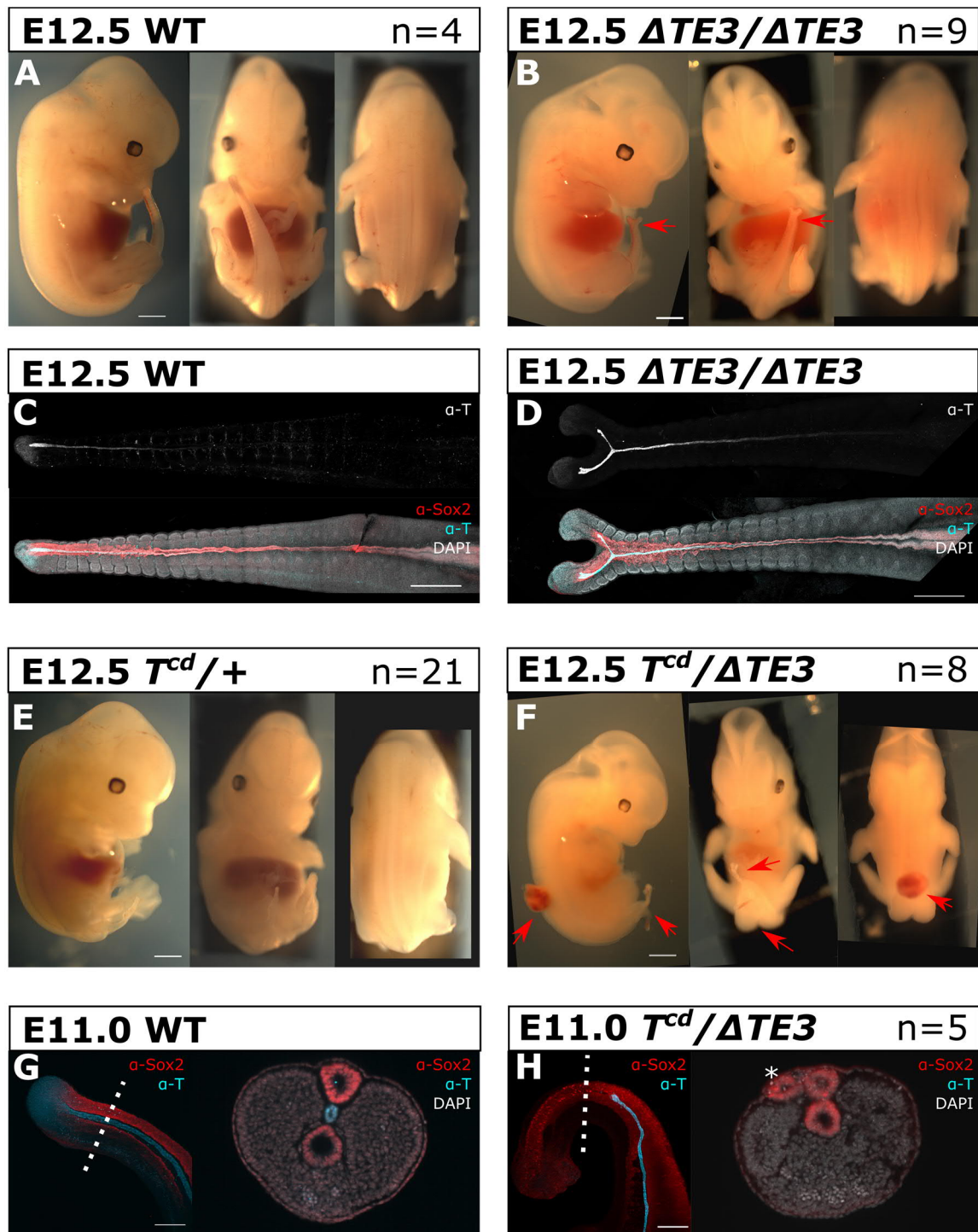
Figure 2.2.2 *TE3* enhancer deletion induces a late axis elongation phenotype

(A-F; H) E12.5 embryos generated by tetraploid complementation. Total number of isolated embryos in all experiments specified. (A) Wild-type. Scale bar=1mm. (B) *TE3* knockout in the wild-type background. Red arrows indicate caudal tail bifurcation (n=9/9). Scale bar=1mm (C-D) Maximum intensity projections of confocal stacks. E12.5 tails with immunostaining for T (light blue) and Sox2 (red). Nuclei stained using DAPI (grey). Scale bar=500μm. (E) T/+ heterozygous mutant. Scale bar=1mm. (F) *TE3* deletion in the heterozygous background. Arrows indicate additional defects like split tail or neural tube closure defects. Scale bar=1mm. (G-H) Light sheet imaging of E11.5 tails. Immunostaining for T (light blue) and Sox2 (red). Dashed line indicates optical section plane. Asterisk points out excess neural tissue. Scale bar=200μm.

To investigate a possible function of *TE3*, it was deleted in both wild-type and $T^{cd}/+$ mESCs and chimeric embryos generated by tetraploid aggregation were isolated at E9-E12.5. The gross morphology was not affected during earlier stages. However, *ΔTE3/ΔTE3* embryos displayed bifurcations and a late axis truncation in the last third of the tail at E12.5. This phenotype was more pronounced in the $T^{cd}/ΔTE3$ heterozygous background, which in addition displayed defects in the neuroectoderm, like meandering or open neural tubes (Figure 2.2.2 F).

The axis defects induced by the *TE3* deletion were rather subtle compared to the stronger and more variable phenotype of the heterozygous $T^{cd}/+$ mutant. Thus it is difficult to specifically differentiate between the enhancer knockout and $T^{cd}/+$ haploinsufficiency phenotype. For this reason, *ΔTE3/ ΔTE3* homozygous mutant with both alleles of *T* still intact was used for further experiments, even though the phenotype was not as pronounced in this background.

The first abnormality preceding the tail bifurcation became visible after trunk to tail transition, when few T/Sox2 double positive cells with lower expression for both factors accumulated at an ectopic site between the epithelium of the neural tube and the notochord (Figure 2.2.3 A-B). To detect a possible imbalance between mesodermal and neuroectodermal tissue, early tail buds at E9.75 were analyzed by immunofluorescence for T and Sox2. Using ZEN pro software (Zeiss), confocal microscopy stacks were acquired with 10μm intervals and automatically segmented into nuclei according to DAPI signal. Mean fluorescence intensity was measured for T and Sox2 protein and normalized to DAPI fluorescence intensity to minimize working distance and laser penetration depth effects.

In addition, cell types in the caudal end were manually annotated in three central sections per specimen. Plotting based on T and Sox2 expression levels showed that the different populations with T and/or Sox2 activity cluster together and therefore, the amount of notochord (NC), presomitic mesoderm (PSM), dorsal hindgut (GUT), chordo-neural-hinge (CNH), neural tube (NT) and surface ectoderm (SFE) cells can be measured. The distribution of these populations was similar between wild type and *ΔTE3/ΔTE3* mutants with two exceptions: T/Sox2 double positive cells localized in the chordo-neural hinge and dorsal hindgut showed a loss of T expression (Figure 2.2.3 C-D). This effect also became evident in the mean expression in the annotated cell types (Figure 2.2.3 E-F).
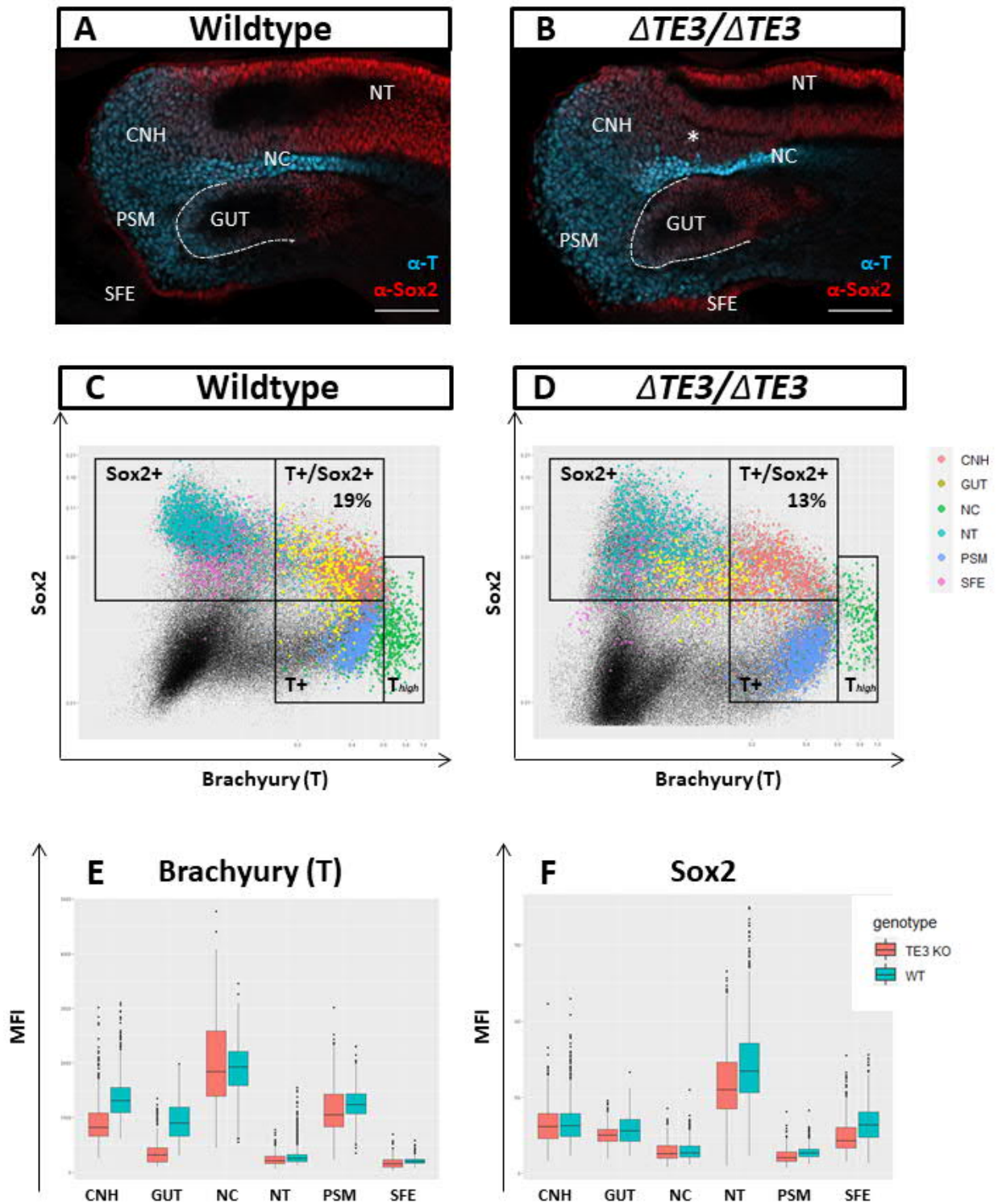
Figure 2.2.3 *TE3* controls Brachyury expression in hindgut and NMPs of the tail bud

(A-B) Sagittal optical sections of E9.75 tail buds. Immunostaining for T (blue) and Sox2 (red). Scale bar=200μm. CNH = Chordo-neural hinge. SFE = Surface ectoderm. PSM = Presomitic mesoderm. NT = Neural tube. GUT = Hindgut. NC = Notochord. Asterisk indicates ectopic T+/Sox2+ cells. (C-D) Plots for *ΔTE3/ΔTE3* and wild-type mean fluorescence intensities (MFIs) of auto-segmented cells. Manually annotated cells highlighted with color according to population. (E-F) MFI of T and Sox2 in annotated cell populations.

52

Taken together, *TE3* had enhancer activity in nascent mesoderm and T/Sox2 co-expressing cells of the chordo-neural hinge and dorsal hindgut. T expression decreased in these domains upon enhancer deletion. This led to a minor, but clear axis elongation defect and suggests a role for *TE3* in the T vs. Sox2 mutual inhibition mechanism that has been proposed previously (Koch et al., 2017).

2.3 Part II Transcriptome profiling of notochord progenitors in axis development

Nascent notochord cells are located in the center of the posterior growth zone and function as a source of secreted signals for the emerging tissue anlagen and their progenitors. Throughout trunk and tail development, the NMP domain is closely associated with the *Noto* expressing cells, suggesting that these cells form an instructive niche in the node-streak border (NSB) and later the chordo-neural hinge (CNH). The progenitor pools preserved in this niche might in addition give rise to the cells that extend the notochord, thereby moving the progenitor zone to the posterior. This function of Noto+ cells together with the progenitors would constitute an organizer for notochord and tail elongation.

In this chapter, the expression profile of Noto+ cells was analyzed to identify genes controlling notochord development as well as the signaling factors exerting *Noto*+ cell function.

2.3.1 A *Noto/T/Foxa2* triple reporter enables dissection of the axial progenitor niche

The model in which Noto+ cells interact with the progenitor domain to form the niche implies direct contact or close proximity of both populations. In order to dissect the components of the posterior growth zone, a *Noto, T* and *Foxa2* triple reporter mESC line was utilized (established by Milena Pustet, Bachelor's thesis, MPI-MG). The line is based on the Noto^{mC} clone used earlier in this work and has *T::H2B-Venus* (T^{Ve}) and *Foxa2::H2B-mTurquoise2* (Foxa2^{mT}) reporter BACs integrated in addition to Noto^{mC}.

All three transcription factors are expressed in the notochord at high levels. T^{Ve} signal was high in the pre-somitic mesoderm and highest in the NSB/CNH and hindgut containing T+/Sox2+ double positive cells. In addition to strong signal in the notochord, Foxa2^{mT} marked the hindgut pocket and could therefore be used to differentiate between T^{Ve}+ cells in the endoderm and NMP domain (Figure 2.3.1 A). In this way, it was possible to differentiate putative signal secreting Noto^{mC}+/T^{Ve}high/Foxa2^{mT}+ cells from responding T^{Ve}high/Noto^{mC}-/Foxa2^{mT}- cells by FACS (Figure 2.3.1 A).
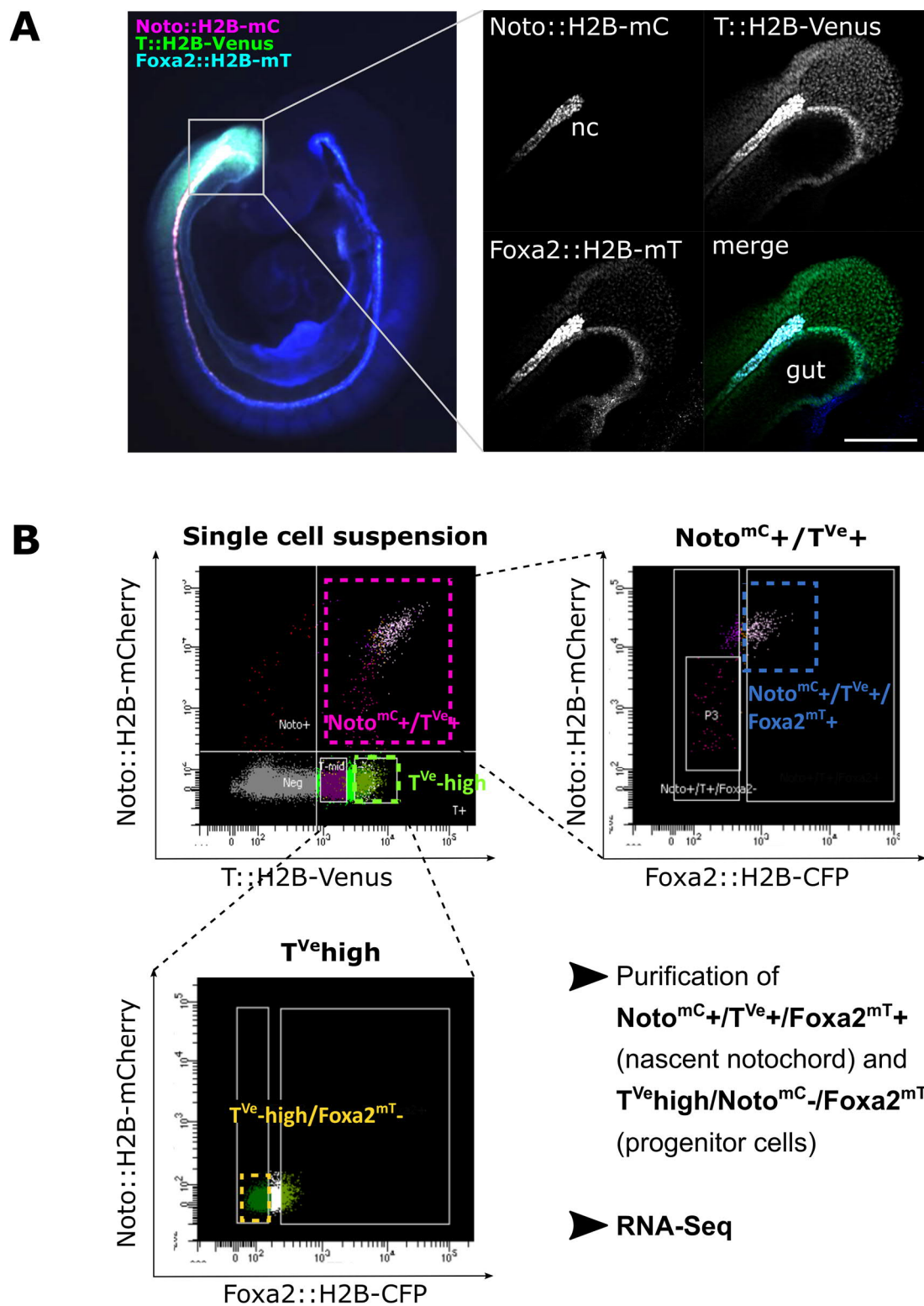
Figure 2.3.1 A *Noto/T/Foxa2* triple reporter line for the dissection of signaling in the caudal growth zone

(A) Left: E9.5 embryo with *Noto::H2B-mCherry, T::H2B-Venus, Foxa2::H2B-mTurquoise2* reporter signal. Right: Airyscan acquired optical section of the caudal growth zone. Scale bar=200μm. (B) FACS profiles of a E9.5 sample.

Embryos (n=5 per stage) were generated by diploid complementation assays and isolated at the 9-11 somite stage and 24-28 somite stage. Caudal ends were obtained cutting at the somite border and populations were sorted by flow cytometry. RNA was extracted from 250 cells per sample and subjected to low input library preparation (Ovation SoLo) followed by sequencing. After index based duplicate removal and mapping, FPKM expression values were calculated using cufflinks (Trapnell et al., 2013).

High expression of markers such as *Noto*, *Shh*, *Foxa1*, *Lmx1a* (Abdelkhalek et al., 2004; Barnes et al., 1994; Besnard et al., 2004; Echelard et al., 1993) in the Noto^{mC}+/T^{Ve}high/Foxa2^{mT}+ sample confirmed that it was comprised by nascent notochord cells (Figure 2.3.2). In the T^{Ve}high/Noto^{mC}-/Foxa2^{mT}- sample, markers of pre-somitic mesoderm, like *Tbx6*, *Msgn1*, *Dll1*, *Dll3* (Bettenhausen et al., 1995; Chapman et al., 1996; Dunwoodie et al., 1997; Yoon et al., 2000) were enriched and NMP markers *Sox2*, *Cdx2*, *Cdx4* and *Nkx1.2* (Amin et al., 2016; Rodrigo Albors et al., 2018; Takemoto et al., 2011) were moderately expressed. This suggests that the sample contained a mix of few NMPs as well as NMP descendants with mesodermal fate, which both are cell types exposed to signals secreted from Noto+ cells (Figure 2.3.2). Therefore, the signaling interactions between Noto+ cells and their neighbors can be studied in the harvested populations.

Next, the FPKM list was filtered for genes with gene ontology (GO) term annotations for the major developmental signaling pathways involved in axial elongation, namely Wnt, Fgf, Bmp, Tgf-β and Notch. Literature research was conducted using the Mouse Genome Informatics (MGI) database to evaluate the candidate genes that were either specifically expressed in one group (two-fold upregulation) or highly expressed in both groups.

The ligands *Wnt3*, *Wnt3a*, *Wnt5a* and *Wnt6* and the secreted co-factor *Rspo3* (Andre et al., 2015; Liu et al., 1999; Schmidt et al., 2004; Takada et al., 1994) were detected in Noto+ cells. *Wnt6* and *Wnt3* were specifically expressed in *Noto+* cells at low levels according to the acquired data. However, expression of *Wnt3a, Wnt5a* and *Rspo3* was clearly stronger in the neighboring progenitor cells. Thus, the transcriptome did not strongly support the hypothesis that Noto+ cells function as a regionalized source of Wnt ligands. (Figure 2.3.2).

On the other hand, the Noto+ cell expression profile was enriched for components of the pathway that mediate the Wnt response. Wnt receptors *Frizzled 4/5/6*, *Lrp5/6*, *Lgr5* (Kelly et al., 2004; Morita et al., 2004; Pinson et al., 2000; Wang et al., 2016) and a number of

extracellular activators like *Cd44*, *Lypd6* and *Pkd2* (Boulter et al., 2001; Özhan et al., 2013; Schmitt et al., 2015) were upregulated. In addition, intracellular factors that promote β-catenin translocation to the nucleus (*Fermt1*) or directly interact with β-catenin to enhance the transcriptional Wnt-response, like *Sp5*, *Sp8*, *Tcf7l2 (Tcf4)* and *Sox4* (Jiao et al., 2017; Kennedy et al., 2016; Liu et al., 2017; Sinner et al., 2007) were particularly expressed. The expression profile therefore indicates that Noto+ cells may not necessarily produce high levels of Wnt ligands, but are highly capable to respond to and potentiate canonical Wnt signals.

Further, high levels of *Wnt5b* were detected and several components of the non-canonical Wnt pathway including *Vangl1*, *Prickle1*, *Celsr1*, *Cthrc1* and *Daam1/2* (Brzóska et al., 2016; Lee and Deneen, 2012; Liu et al., 2014; Nakaya et al., 2020; Yamamoto et al., 2008) were highly upregulated. This is in line with the requirement of the planar cell polarity pathway for convergent extension and notochord development (Minegishi et al., 2017).

In general, transcripts of Fgf ligands (*Fgf3*, *Fgf13* and *Fgf15*), receptors (*Fgfr1*, *Fgfr2*) and antagonists (*Spry2*, *Spry4*) were enriched in the mesodermal progenitors (del Corral and Morales, 2017). Expression of *Fgf8*, which is an essential factor for NMP maintenance (Boulet and Capecchi, 2012; Henrique et al., 2015) was clearly stronger in Noto+ cells.

Interestingly, the Apelin receptor *Aplnr* and one of its endogenous ligands, *Apela* are highly expressed in Noto+ cells. *Apela -/-* mutants display axial truncations, therefore apelinergic signaling is a potential new player in embryonic development (Freyer et al., 2017).

Bone morphogenic proteins *Bmp1*, *Bmp2*, *Bmp4* and *Bmp7* as well as *Furin*, coding for a Bmp proconvertase (Mishina, 2003; Roebroek et al., 1998) show regionalized moderate expression in Noto+ cells. Most prominent however is the notochord specific expression of *Nog*, *Chrd*, and *Dand5*, encoding secreted Bmp antagonists (Bachiller et al., 2000; Li et al., 2007; Minegishi et al., 2020). Other notable factors of the TGF-β superfamiliy that were detected in Noto+ cells included *Nodal*, *Gdf1*, *Tgfb1* and *Tgfb2* (Andersson et al., 2006; Sanford et al., 1997; Shull et al., 1992; Zhou et al., 1993). Expression of *Gdf11*, required for trunk to tail transition (Jurberg et al., 2013), was restricted to the mesodermal progenitors.

Taken together, comparative transcriptome analysis of Noto+ cells and adjacent mesodermal progenitors shows an enrichment of Wnt/PCP genes and components of the canonical Wnt response in Noto+ cells. Potential localized and secreted niche factors in Noto+ cells include Wnt3, Fgf8 and Apela. In addition, Noto+ cells are a source of patterning morphogens, especially Shh and Bmp antagonists.
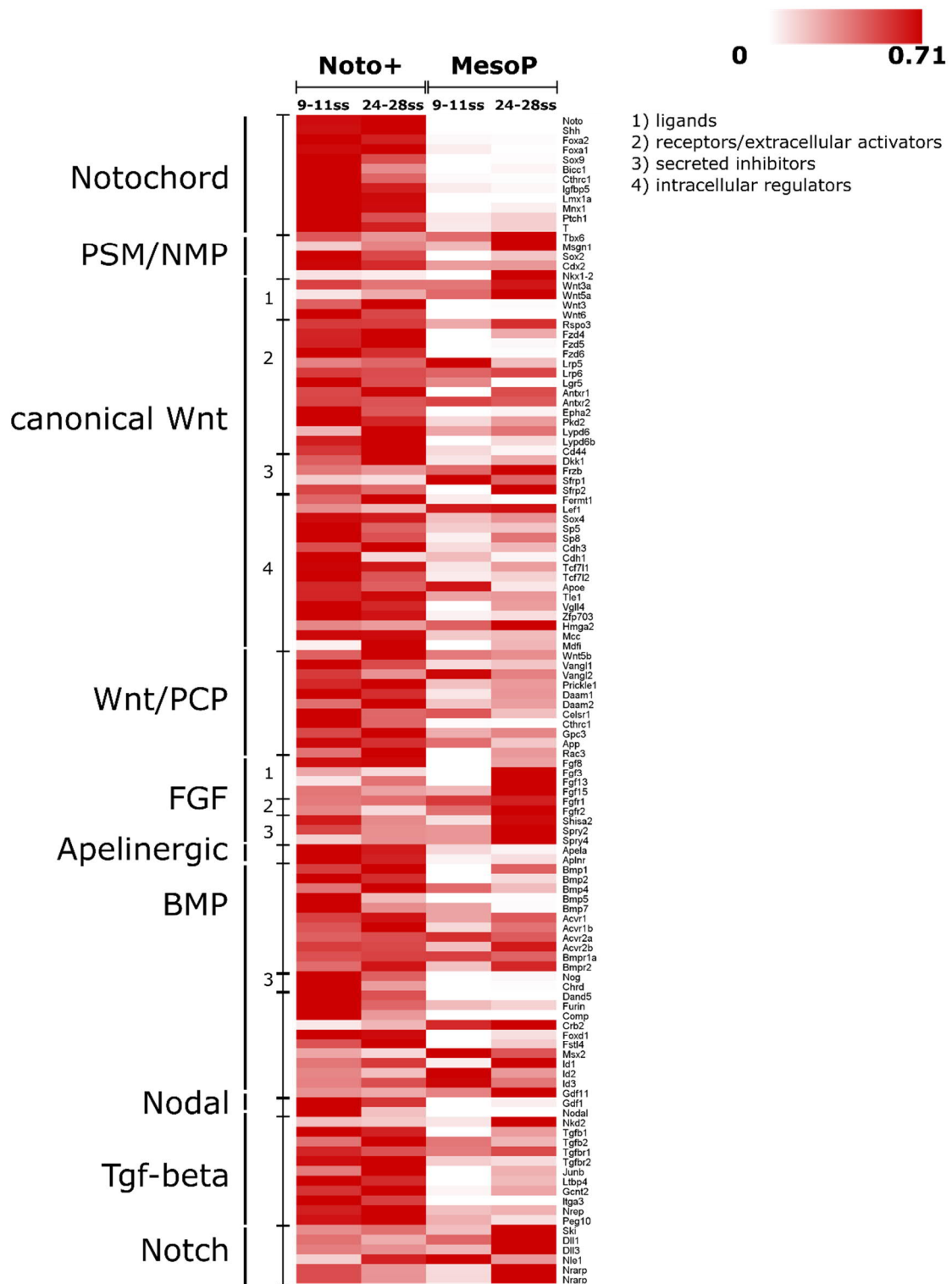
Figure 2.3.2 Differential expression of signaling pathway components in Noto+ and Mesoderm progenitors

Heatmap of per gene normalized FPKM values. Sum of the squares of values in each row equals 1. Noto+ = samples of $Noto^{mC}$+/T^{Ve}high/$Foxa2^{mT}$+ cells. MesoP = samples of T^{Ve}high/$Noto^{mC}$-/$Foxa2^m$ cells.

2.3.2 The developmental transcriptome of notochord precursors between E7.5 and E9.75

Trunk notochord is formed by cells of the node undergoing convergent extension. Notochord elongation in the tail is thought to be driven by cells recruited from posterior progenitors (Yamanaka et al., 2007). In order to determine the gene expression in these different phases of notochord formation and function, the Noto^{mC} reporter line was utilized to harvest nascent notochord cells from node formation to trunk-tail transition (Figure 2.3.3). Embryos generated by diploid aggregation assays were isolated at early headfold (EHF), late headfold (LHF), 12-15 somite, 16-18 somite and 24-28 somite stages. To determine the specificity of gene expression values in the wild-type cells several reference groups were included. First, Noto^{mC}+ cells from *T/tw5* mutants of equivalent stages, which activated the reporter but were not specified and therefore did not function like their wild-type counterparts. In addition, cells isolated from the more mature notochord anterior of the somite border were analyzed to distinguish general notochord genes from those specifically expressed in the nascent Noto^{mC}+ cells of the caudal growth zone. Finally, *in vitro* derived Noto^{mC}+ cells harvested at day 7 of the protocol were included to evaluate which *in vivo* embryonic stage they resemble the most.
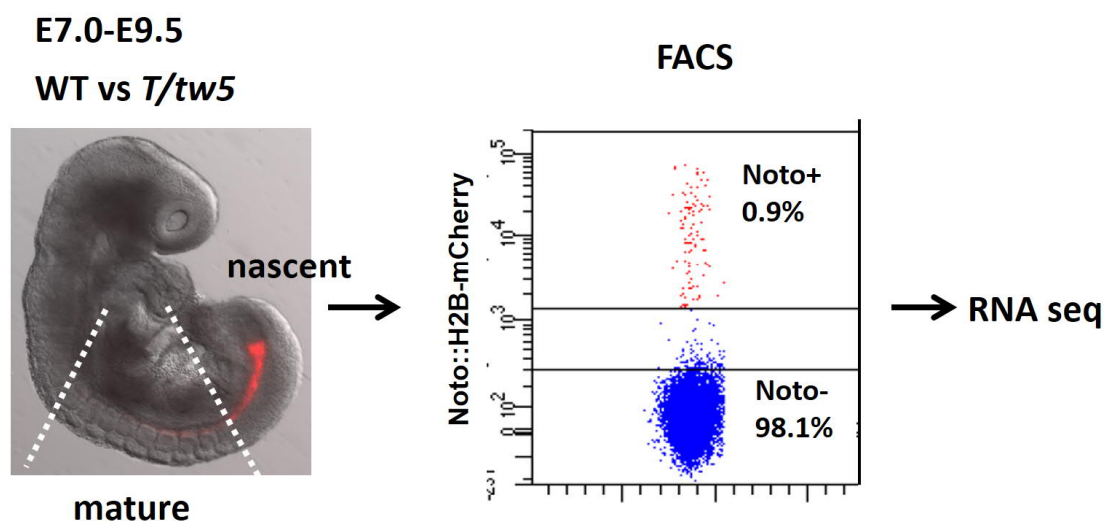
Figure 2.3.3 Experimental set up for the purification of Noto+ cells in different developmental stages

Wild-type embryos were isolated between E7.0 (Early headfold stage) and E9.5 (24-28 somite stage), regions of interest were trimmed and single cell suspensions sorted by FACS. Nascent Noto^{mC}+ cells (Typically ~1% of all cells in the caudal ends) were harvested from wild-type and *T/tw5* caudal ends cut at the somite border. Mature notochord was sorted from trunk tissue between somite border and forelimb bud using the Noto^{mC}+/T^{Ve}/Foxa2^{mT} triple reporter. *In vitro* derived Noto+ cells were isolated at d7 of the differentiation. Representative FACS plot shows an E9.0 (16-18 somite stage) wild-type sample.

Extraembryonic tissues were removed from the headfold stage embryos and caudal ends were trimmed at the somite border of older embryos. Depending on the amount of embryos available, five to ten embryos were used per sample (Supplementary table 2). Noto^{mC}+ cells, which comprised approximately one per cent of the samples, were purified from single cell suspensions using flow cytometry. As described in 2.3.1, total RNA from about 250 cells per sample was used for library preparation and RNA-Seq (Supplementary table 1). After de-duplication and mapping, FPKM values were calculated.

For a rough overview of similarities between the samples presented in this chapter, unsupervised hierarchical clustering of log-transformed FPKM values was performed. The structure of the sample tree indicates that the samples first grouped into into cell type based clusters and then clustered roughly according to developmental stage (Supplementary Figure 7). The headfold stage samples isolated at E7.0-E7.5, when Noto^{mC}+ cells form the node separated from caudal end samples isolated at E8.25-9.0 (4-15 somites) and E9.25-E9.75 (16-28 somites). The *in vitro* differentiated Noto^{mC}+ cells were most similar to the mature trunk notochord between somite border and forelimb bud at 4-11 somite stage and 24-32 somite stages, which also grouped together. This section of the notochord is derived from the node, one of the earliest domains of *Noto* expression at E7.0 (Plouhinec et al., 2004). With the first Noto^{mC}+ arising at d5, it is possible that the differentiation protocol recapitulates the endogenous sequence of notochord development. *T/tw5* mutant Noto^{mC}+ samples were closest to T^{Ve}+/Noto^{mC}- and other mesoderm precursor samples, which further substantiates that they acquired paraxial mesoderm fate.

The structure of the hierarchical clustering did not show any tendency of samples that were processed together to group together, which indicated that there were no strong batch effects from the three different rounds of library preparation and sequencing.

In conclusion, the global gene expression profile of Noto^{mC}+ cells changes throughout trunk development between specification of the node at E7.25 and the trunk to tail transition at E9.75. In order to assess the expression dynamics of genes important for notochord development and function, lineage markers and differentially regulated signaling pathway components determined in 2.3.1 were analyzed.
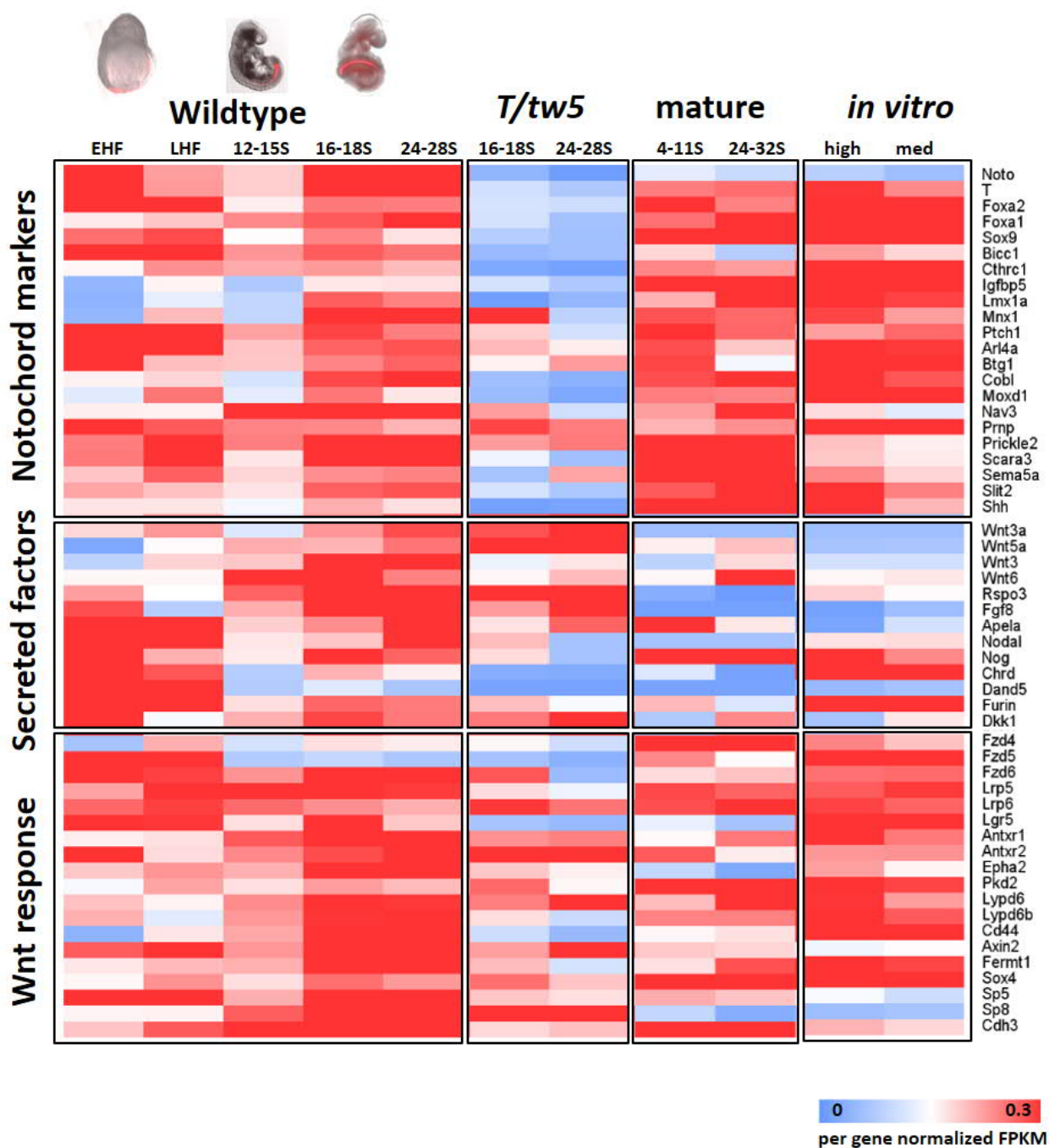
Figure 2.3.4 Developmental transcriptome of Noto+ cells

Expression of selected genes (notochord lineage markers, signaling factors and Wnt response genes) from different developmental stages. FPKM values were normalized such that the sum squares of the values in each row equal 1.

Expression analysis of notochord lineage markers validated that the *in vitro* generated Noto^{mC}+ cells resembled their *in vivo* counterparts in the node derived trunk notochord at 4-11 somite stage (Figure 2.3.4). Therefore, they also express higher levels of genes upregulated during lineage progression including *Igfbp5*, *Sox9*, *Cthrc1* or *Moxd1* and lower levels of nascent notochord markers like *Noto, Bicc1 and Nav3* (Wymeersch et al., 2019). Further, as presented in Figure 2.1.12 in more detail, the *T/tw5 Noto^{mC}*+ cells were not properly specified and therefore lack the genetic lineage signature (Figure 2.3.4).

In general, most of the notochord genes including the three key lineage regulators *T*, *Foxa2* and *Noto* were dynamically regulated in a way that starting from the early headfold stage, high transcript levels decreased reaching a minimum at 12-15 somite stage. Afterwards, these genes were upregulated between the 16 and 28 somite stage. These time points coincide with the notion that notochord formation occurs in two waves: First, the quiescent ventral node lays down most of the trunk notochord via convergent extension. The ventral node cells are thought to be quiescent at this time and notochord precursors start to proliferate again in the second phase in which the notochord is elongated by cells posterior of the node (Ukita et al., 2009; Wymeersch et al., 2019; Yamanaka et al., 2007).

Concomitantly with notochord lineage genes, the Wnt targets and activators identified in 2.3.1 including *Sp5*, *Epha2*, *Cd44*, *Lgr5*, and *Antxr1* (Abrami et al., 2008; Kennedy et al., 2016; Morita et al., 2004; Naruse-Nakajima et al., 2001; Schmitt et al., 2015) were also increasingly and specifically expressed in the Noto^{mC}+ cells towards the end of trunk generation (Figure 2.3.4). This suggests that the notochord progenitors become more sensitive to Wnt signals at later stages of axial elongation.

Notably, the expression of potential niche factors *Fgf8*, *Nodal*, *Apela* and *Wnt3* in the Noto^{mC}+ domain was upregulated at E7.0 and E9.5 at the stages when NMPs are thought to be generated in the NSB and CNH, respectively (Figure 2.3.4). The same trend could be observed for organizer factors such as *Nog*, *Chrd* and *Furin* (Bachiller et al., 2000; Roebroek et al., 1998) even though less transcript is detected in the Noto^{mC}+ cells in the caudal end compared to the node.
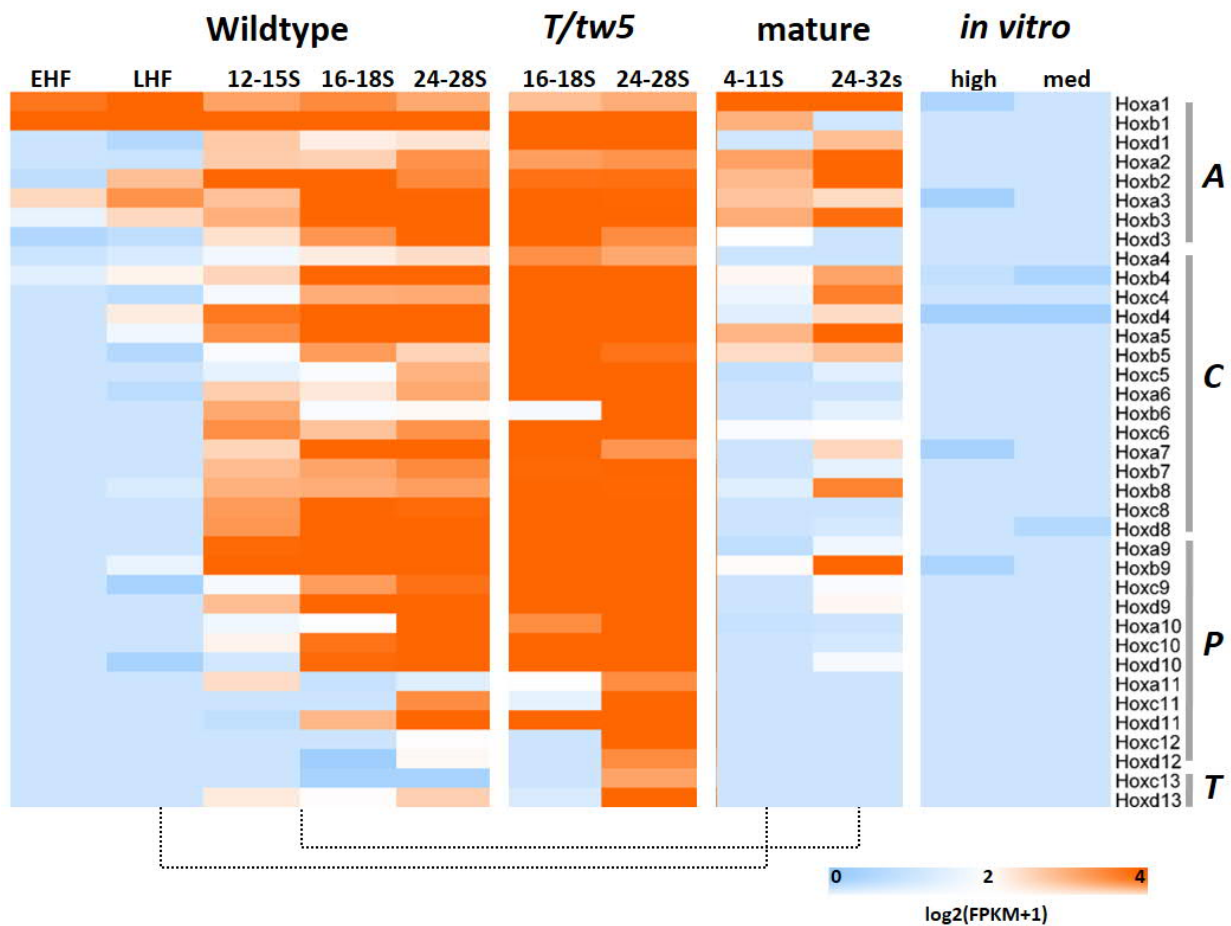
Figure 2.3.5 Collinear Hox expression in Noto+ cells

Log-transformed FPKM expression values of all expressed Hox genes with an FPKM value >2 in at least one sample. Paralogous groups clustered according to the position in the trunk. A=anterior. C=central. P=posterior. T=terminal. Dashed lines connect samples with overlapping Hox identity.

Dynamic changes in global transcription as well as stage specific marker gene activation in the $Noto^{mC}$+ cells suggests that this population undergoes maturation during development. In addition, notochord progenitors in the caudal growth zone are constantly exposed to Wnt signaling, which was shown to activate Hox expression (Denans et al., 2015). In general, sequential activation of Hox genes implements A-P positional information in tissues emerging from the progenitor zone during axial elongation. Recently, Hox collinearity has been demonstrated in NMPs and lateral plate mesoderm progenitors (Wymeersch et al., 2019) and endoderm (Nowotschin et al., 2019). Whether the same is true for notochord progenitors in the mouse is not known. Therefore, Expression of Hox genes was examined during all the analyzed stages (Figure 2.3.5).

Indeed, the transcriptome of Noto^mC+ cells isolated from E7.0-E9.5 revealed a conserved temporal sequence of Hox activation. The anterior paralogous groups (PG) 1-3 were expressed in headfold stage embryos only. Expression values of central (PG 4-8) and the first posterior (PG 9-12) groups were gradually elevated during trunk development (Figure 2.3.5). The expression within mature notochord from 4-11 somite and 24-32 somite stage reflects the profile of the Noto^mC+ precursors late headfold and 16-18 somite stage, respectively. This indicates that notochord cells acquire their Hox identity as progenitors and this identity is maintained afterwards (Figure 2.3.5).

The samples with *T/tw5* mutant Noto^mC+ cells, which were mainly converted to paraxial mesoderm and gut (Figure 2.1.12), displayed a Hox gene expression signature that was advanced compared to the wild-type notochord cells by several hours (the formation of about 5 somite pairs). This suggests that there are slight spatiotemporal differences in Hox gene activation between the early primordia of the lineages (Figure 2.3.5).

In contrast to the *in vivo* samples isolated from embryonic tissue, *in vitro* generated *Noto^mC+* cells do not express any Hox genes.

In summary, the data presented in this chapter constitute a unique resource for investigating the developmental transcriptome of the notochord, its progenitors and its direct neighboring cell populations. Transcription of genes required for notochord development, Wnt pathway response and signaling changes dynamically during trunk development, reaching the highest levels at the stages before the expansion of progenitors for trunk and tail tissues at E7.25 and E9.5, respectively.

It is possible that Noto^mC+ cells provide the regime of Nodal, Fgf, Wnt and apelinergic signaling in which progenitor pools, including the progenitors for the notochord, are amplified at these timepoints. Via Shh and BMP antagonist secretion, the notochord, a direct derivative of Noto^mC+ cells, patterns the axial tissues along the medio-lateral and dorso-ventral axis. Based on these findings, the Noto^mC+ cells could function as an essential compound of the organizer of axial elongation.

2.4   Part III Localization of Axial stem cells

The inner cell mass and epiblast of the early embryo are composed of pluripotent stem cells, which give rise to all embryonic tissues. The pluripotency is maintained by a gene regulatory network with the key transcription factors *Oct4* (*Pou5f1*), *Sox2* and *Nanog* at its core (Orkin, 2005). Pluripotency in the epiblast is thought to be a transient state that is successively lost as *Oct4* and *Nanog* are downregulated at E8.5 in the mouse (Festuccia et al., 2013).

To the contrary, axial elongation has been shown to be driven by progenitor cells in the caudal growth zone that retain germ layer plasticity and the capacity to self-renew, which are hallmarks of stem cells (Cambray and Wilson, 2007; Selleck and Stern, 1991; Tam and Tan, 1992; Tzouanacou et al., 2009; Wilson et al., 2009).

Recent evidence indicates that the pluripotency factor Oct4 is essential for trunk development by maintaining the stemness of axial progenitors. Oct4 was shown to be expressed in the node-streak border at E8.5 in a region containing multipotent (Cambray and Wilson, 2007; Downs, 2008). An induced knock-out, which bypasses the critical role of Oct4 in early embryogenesis, where Oct4 is essential, induces severe defects including posterior truncations (DeVeale et al., 2013). In addition, overexpression of Oct4 in the epiblast results in additional trunk segments and delays trunk to tail transition (Aires et al., 2016). Thus, Oct4 is a promising candidate as a marker and regulator of axial stem cells.

In this chapter, Oct4 activity was investigated during trunk and early tail development. Oct4 expressing cells were localized in the NSB and CNH and characterized by RNA-Seq.

2.4.1   Establishment of an Oct4-Venus fusion cell line

Knock-in reporters like the *Noto::H2B-mCherry* BAC transgene used in previous chapters provide faithful read-outs for the onset of gene expression (Figure 2.1.1). However, reporter stability may cause expression patterns that are substantially different from those of the endogenous protein. For stem cell populations, this approach would label not only the stem cells, but also their descendants. In addition, monitoring of asymmetric cell division events, in

which one daughter cell retains the pluripotency factor and the other daughter cell differentiates (Habib et al., 2013) would not be possible using transcriptional reporters.

To facilitate visualization of the endogenous Oct4 protein, an Oct4-Venus fusion protein expressing mESC line was engineered. A targeting vector containing a *5′-Venus-IRES-Puro-3′* cassette flanked by two homology arms was generated via recombineering (Figure 2.4.1). The reporter construct was integrated into the endogenous stop codon of *Oct4* such that a *5′-Oct4-Venus-IRES-Puro-3′* transcript was produced from which two proteins are translated: The Oct4-Venus fusion protein and a puromycin-N-acetyltransferase for selection. The homology arms were chosen in a way that they do not contain the start codon and endogenous promoter, therefore fluorescence and puromycin resistance was not likely to occur after random integration. Clones were validated by PCR, expanded and used for the generation of embryos via diploid and tetraploid complementation assays.

The best performing clone was further modified integrating the *Noto::H2B-mCherry* BAC (Figure 2.1.1) or a *T::H2B-mCherry* (T^mC^) BAC (a gift from Dr. Frederic Koch, MPIMG) to investigate the possible association of Oct4-Venus expressing cells with notochord progenitors or NMPs, respectively.
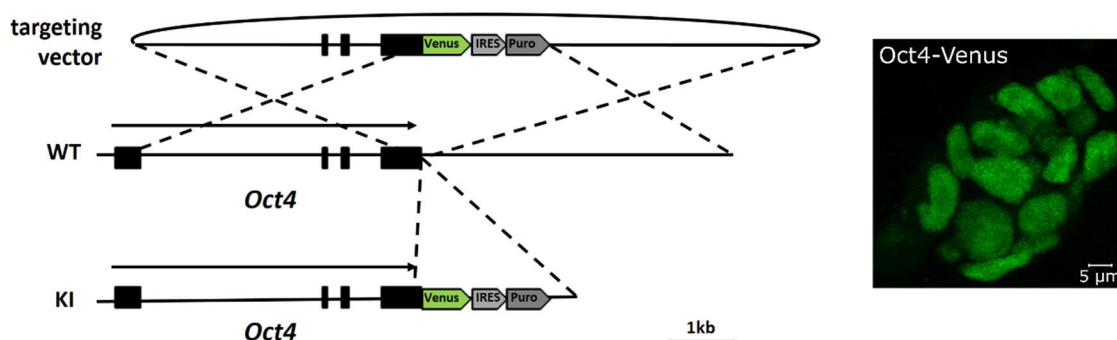
Figure 2.4.1 Oct4-Venus targeting construct for homologous recombination

Schematic showing the targeting vector with the reporter cassette flanked two homology arms (4 kb and 3.8kb). Venus = Venus fluorescent protein CDS. IRES = Internal ribosomal entry site. Puro = Puromycin resistance. WT = wild-type locus. KI = knock-in modified locus after homologous recombination. Right: Maximum intensity projection of a confocal stack showing Oct4-Venus fluorescence in a transgenic mESC colony.

2.4.2   The node streak border harbours T+/Oct4+ cells

Embryos carrying the T^{mC} reporter and Oct4-Venus tag were generated by aggregation and analyzed at early somite stage. Samples were cleared and fluorescence was acquired by Light sheet microscopy. At E8.25, endogenous Oct4-Venus signal was detected throughout the epiblast, in the allantoic bud and, to a lesser extent, in the neural epithelium of the head fold. T^{mC} activity was strongest in the node and the primitive streak, as well as the epiblast encompassing the primitive streak (Figure 2.4.2).

The subdomains of the node and NSB were analyzed in closer detail. T^{mC}, but not Oct4-Venus was strongly expressed in the ventral node (Figure 2.4.2), which is comprised by a ciliated epithelial monolayer continuous with the endoderm forming the typical indented morphology of the node (Lee and Anderson, 2008). The dorsal node, which is the epithelium continuous with the epiblast was Oct4-Venus+, except for a comb of cells in the midline, which are T^{mC}+/Oct4-Venus+ or T^{mC}+/Oct4-Venus- (Figure 2.4.2).

Strikingly, many cells in the crown, which is a group of squamous, Nodal producing cells surrounding the posterior node, co-expressed Oct4-Venus and T^{mC}. In addition, many T^{mC}+/Oct4-Venus+ cells are located in the node streak border and anterior primitive streak, which are the regions where multipotent progenitors including NMPs reside (Cambray and Wilson, 2002; Cambray and Wilson, 2007).

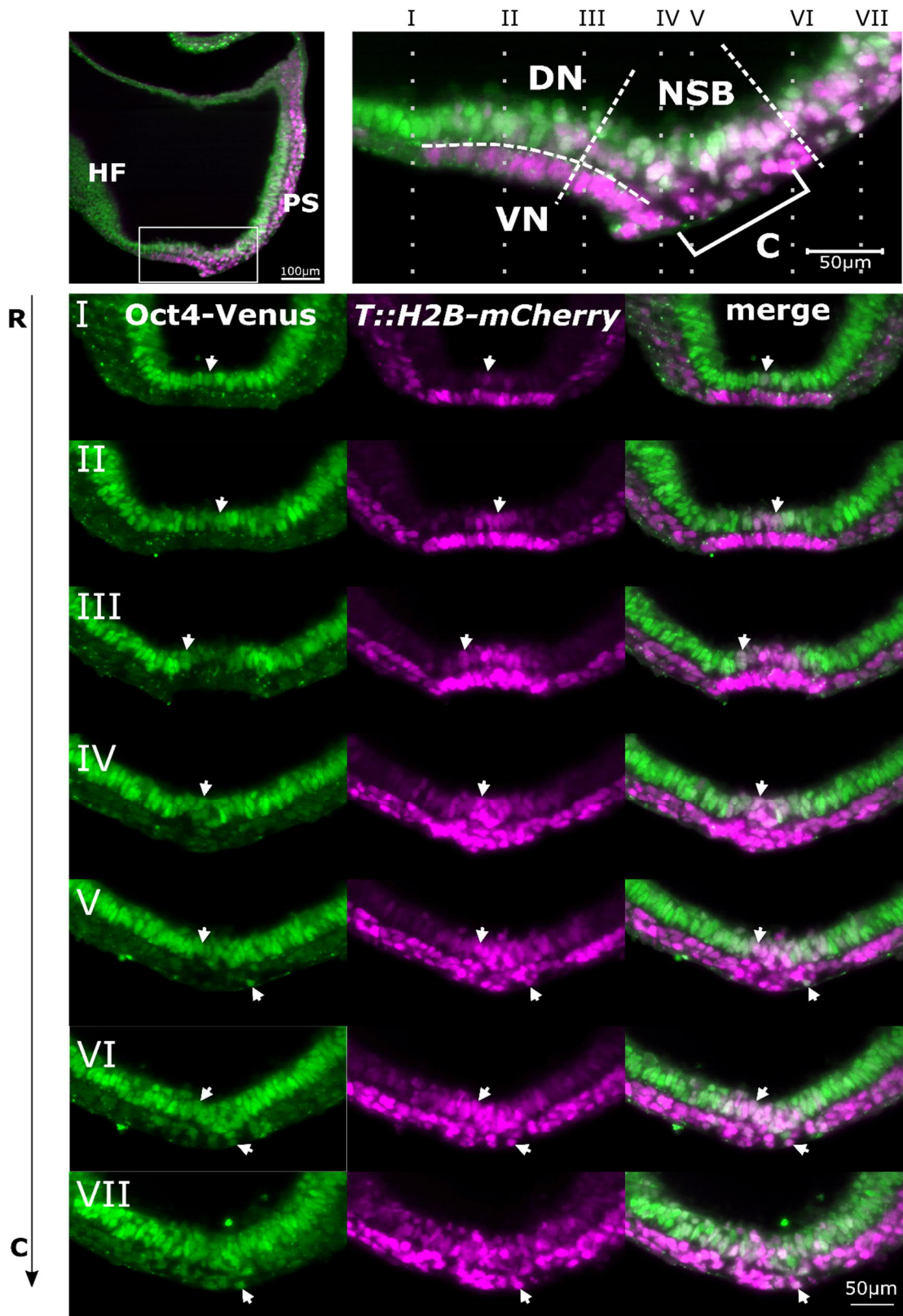Figure 2.4.2 Oct4-Venus and *T::H2B-mCherry* expression in in the E8.25 node area

Fluorescence acquired by Light sheet microscopy. Top left: Sagittal optical section of the whole embryo. HF = head fold. PS = Primitive streak. Square indicates the zoomed node region on the top right. VN = ventral node. DN = dorsal node. NSB = Node Streak Border. C = Crown. Dotted line indicates the positions of tansverse sections I-VII.

2.4.3 Oct4-Venus+/Noto^{mC}+ double positive cells in the dorsal node and crown with a notochord and floor plate progenitor signature

Because T^{mC} is active in the node, primitive streak and the NMPs, Noto^{mC} reporter activity, which is expressed exclusively in axial mesoderm was also assessed, allowing for an investigation of the relation between Oct4+ and Noto^{mC}+ cells.

Analysis of E8.0 embryos revealed several sites of Noto+ and Oct4+ double positive cells in the node region (Figure 2.4.3). In the very midline of the dorsal node, a row of single Noto^{mC}+ cells was observed, which was continuous with the epiblast and neural plate and expressed lower levels of Oct4-Venus. Based on their location, these cells could be either floor plate progenitors or notochord precursors, since some cells were detected delaminating from the dorsal node layer towards the ventral notochordal plate, which gives rise to the notochord (data not shown). The crown cells, which have been hypothesized to contain progenitors of posterior trunk and tail notochord (Wymeersch et al., 2019), co-expressed low levels of Oct4-Venus and Noto^{mC} (Figure 2.4.3). The crown was located directly adjacent to cells in the dorsal half of the NSB with strong Oct4-Venus signal.
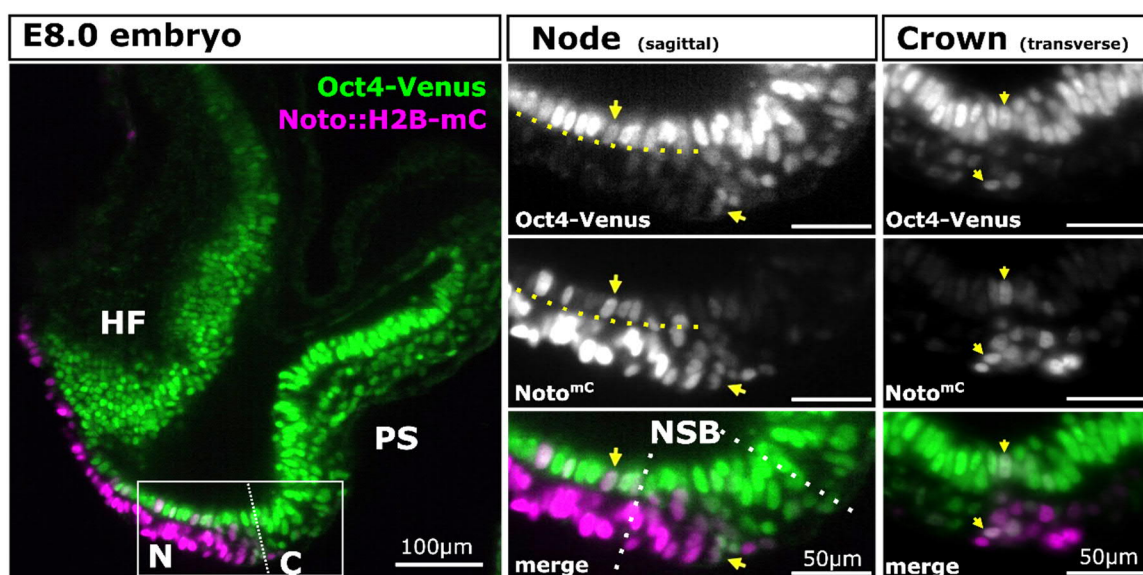
Figure 2.4.3 Oct4-Venus+ and Noto^{mC}+ double positive cells in the dorsal node and crown at E8.0

Optical sections of a cleared E8.0 embryo acquired by light sheet microscopy. Left: Sagittal section through the midline. HF = Head fold. N = Node. C = Crown. PS = Primitive streak. Box indicates the zoomed area in the middle panel. Dotted line indicates the transverse section through the crown. NSB = Node streak border. Yellow arrows point out exemplary Oct4-Venus+/Noto^{mC}+ double positive cells.

Thus, their position in the embryo would support a model in which Noto^{mC}+ cells in the NSB serve as niche for both long term NMPs and notochord progenitors. In order to further characterize the specific cell groups, RNA-Seq was employed. Embryos at early and late headfold stage were isolated and the node streak border region was microdissected (Dr. Lars Wittler, MPI-MPG). Noto^{mC}-/Oct4-Venus+, Noto^{mC} +/Oct4-Venus- single positive and Noto^{mC} +/Oct4-Venus+ double positive cells were purified using FACS (Figure 2.4.4 A). Total RNA was extracted and subjected to library preparation for ultra low input (Ovation SoLo) followed by high-throughput sequencing. FPKM values were calculated from de-duplicated and mapped reads using cufflinks (Trapnell et al., 2013).

Expression values were filtered for genes that were considered to be expressed (FPKM>2) and up- or downregulated (4 fold change) in one of the samples. This resulted in a list of 5057 differentially expressed genes, which were grouped into four clusters by k-means clustering (Figure 2.4.4 B). Cluster I is comprised by genes that are specifically and highly expressed in the Noto^{mC}+/Oct4-Venus+ double positive cells. Consistent with the localization of these cells, markers of the node and crown including *Chrd, Gdf1*, *Foxa2*, *Lmx1a*, *Ptch1* and *Furin* (Roebroek et al., 1998; Wymeersch et al., 2019) were expressed. This indicates that the Noto^{mC}+/Oct4-Venus+ sample contained the precursors of the node and notochord, which were located in the crown and dorsal node.

Cluster II is composed of genes expressed in the ventral node and notochordal plate at high levels like *Noto*, *T*, *Shh*, *Foxa1* and *Mnx1* (Tamplin et al., 2011). It is noteworthy, that *Noto* and *T* are also expressed in Cluster I at lower, but still reasonably high levels (FPKMs I=42, II=250, III=406 and I=0, II=250, III=406, respectively), consistent with the hypothesis that Cluster I contains notochord progenitors.

Cluster III is comprised by genes highly expressed in the Noto^{mC}-/Oct4-Venus+ population including the markers of the pluripotent epiblast *Nanog*, *Otx2* and *Utf1* (Acampora et al., 2013; Osorno and Chambers, 2011) as well as the nascent neuroectodermal drivers *Sox1* and *Zic5* (Furushima et al., 2000; Wood and Episkopou, 1999) .

Finally, Cluster IV consists of genes that are expressed in both the Noto^{mC}-/Oct4-Venus+ single and Noto^{mC}+/Oct4-Venus+ double positive population. These involve the stem cell markers *Oct4* and *Sox2* as well as *Wnt3a*.

70

A

B

C
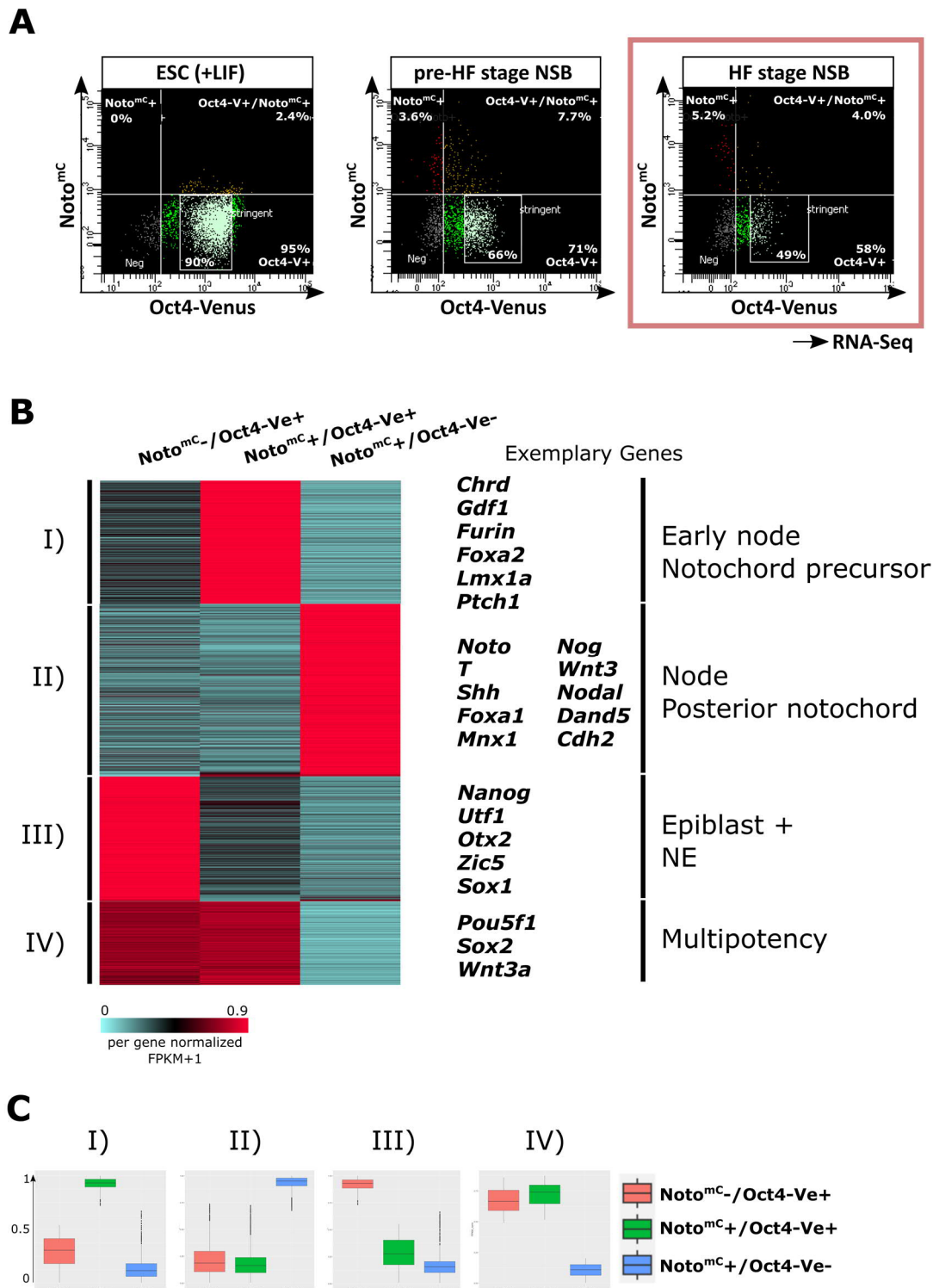
Figure 2.4.4 Transcriptome profile of Oct4 and Noto expressing cells in the NSB

(A) FACS profiles of Oct4-Venus/Noto^{mC} ESC control, pre-headfold (<E7.5) and headfold (E7.5-E8.25) stage samples. Percentages for the more stringent gating are a fraction of the total percentage of Oct4-Venus+ cells. Total Oct4-Venus+ cells were sorted. Sorted populations from the HF stages were subjected to RNA-Seq shown in (B) and (C). (B) K-means Clustering of per gene normalized FPKM+1 values. Sum of the squares of normalized values equals 1 in each row. List of genes (total=5057) filtered for FPKM>2 and up- or downregulation of 4 fold compared to other samples. (C) Boxplot showing the distribution of normalized values in Clusters I-IV.

Taken together, the node-streak border harbors cells with a multipotency signature in several compartments: On the one hand, Noto^{mC}+/Oct4-Venus+ cells in the crown and dorsal node that may give rise to notochord and floor plate, respectively. On the other, cells expressing high levels of Oct4 together with T^{mC}, located dorsally to the crown and in the anterior primitive streak, which could function as long term NMPs.

2.4.4 Oct4+/Sox2+/T+ axial stem cells in the chordo-neural-hinge of the early tail bud

After E8.5 (8 somite stage), Oct4-Venus was only detected in the primordial germ cells and sporadically in single surface ectoderm cells, which is consistent with the literature (Downs, 2008). Immunostaining for Venus was utilized to amplify the endogenous Oct4-Venus signal, but did not reveal any Oct4 positive cells in the notochord area (Supplementary Figure 8).

Towards the end of trunk development (24 somite stage), however, a new domain of Oct4 expression emerged with few dispersed low level Oct4-Venus expressing cells in the very posterior epithelium of the neural plate. During the formation of the tail bud at E9.5, this domain expanded and between the 28-32 somite stage, the number of Oct4+ cells peaked with about 30-40 single cells distributed throughout different domains of the CNH (Figure 2.4.5). Most were located in the epithelium of the CNH, others at the caudal tip of the notochord or in the tail bud mesenchyme (Figure 2.4.5). Immunofluorescence for T and Sox2 showed that all Oct4+ cells co-expressed the main NMP factors T and Sox2 at heterogeneous levels (Figure 2.4.5). This could represent a signature for a novel axial stem cell type.

In addition, embryos were generated from the Oct4-Venus/Noto::H2B-mCherry mESC line and analyzed at an equivalent stage (Figure 2.4.6). Indeed, cells at the caudal tip of the notochord co-expressed T, Oct4 and Noto, suggesting that these cells could not only contribute to NMPs, but also to the axial mesoderm lineage (Figure 2.4.6).

After the peak in Oct4 expression level and cell number at the onset of tail outgrowth (28-32 somite stage), only few cells with very low Oct4-Venus signal are detected at E10.5. At E11.5 and E12.5, Oct4-Venus was depleted from the CNH and only expressed in PGCs and single surface ectoderm cells (Supplementary Figure 8).
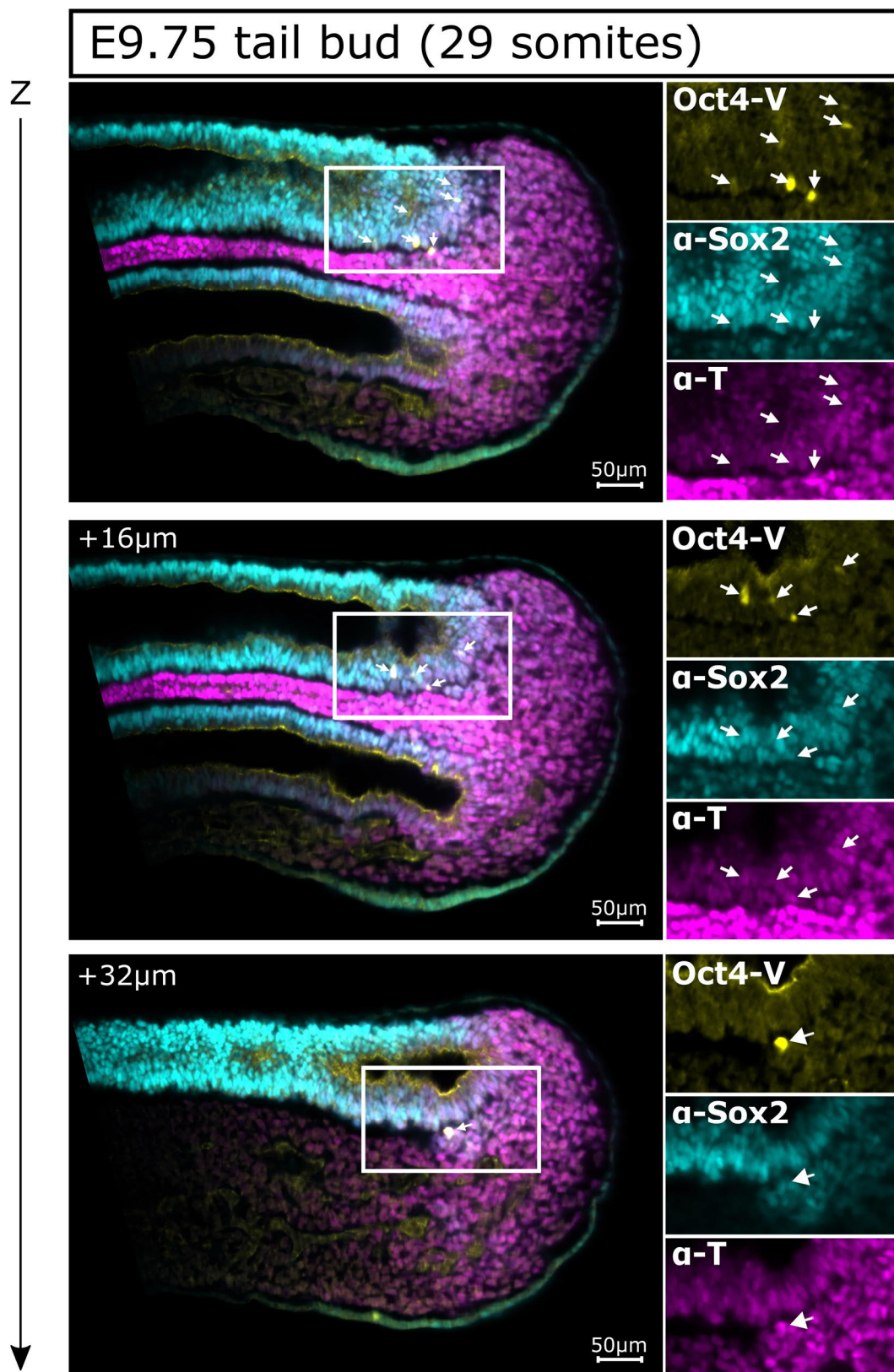
Figure 2.4.5 Oct4+/Sox2+/T+ triple positive cells in the Chordo-neural hinge

Optical sections through the tail bud of an E9.75 embryo acquired by Light sheet microscopy. Immunostaining for EGFP (yellow), T (magenta) and Sox2 (Light blue). Distance between individual planes indicated on the top left. Depth of focus = max 1.3 µm. Arrows point out single Oct4+ cells.
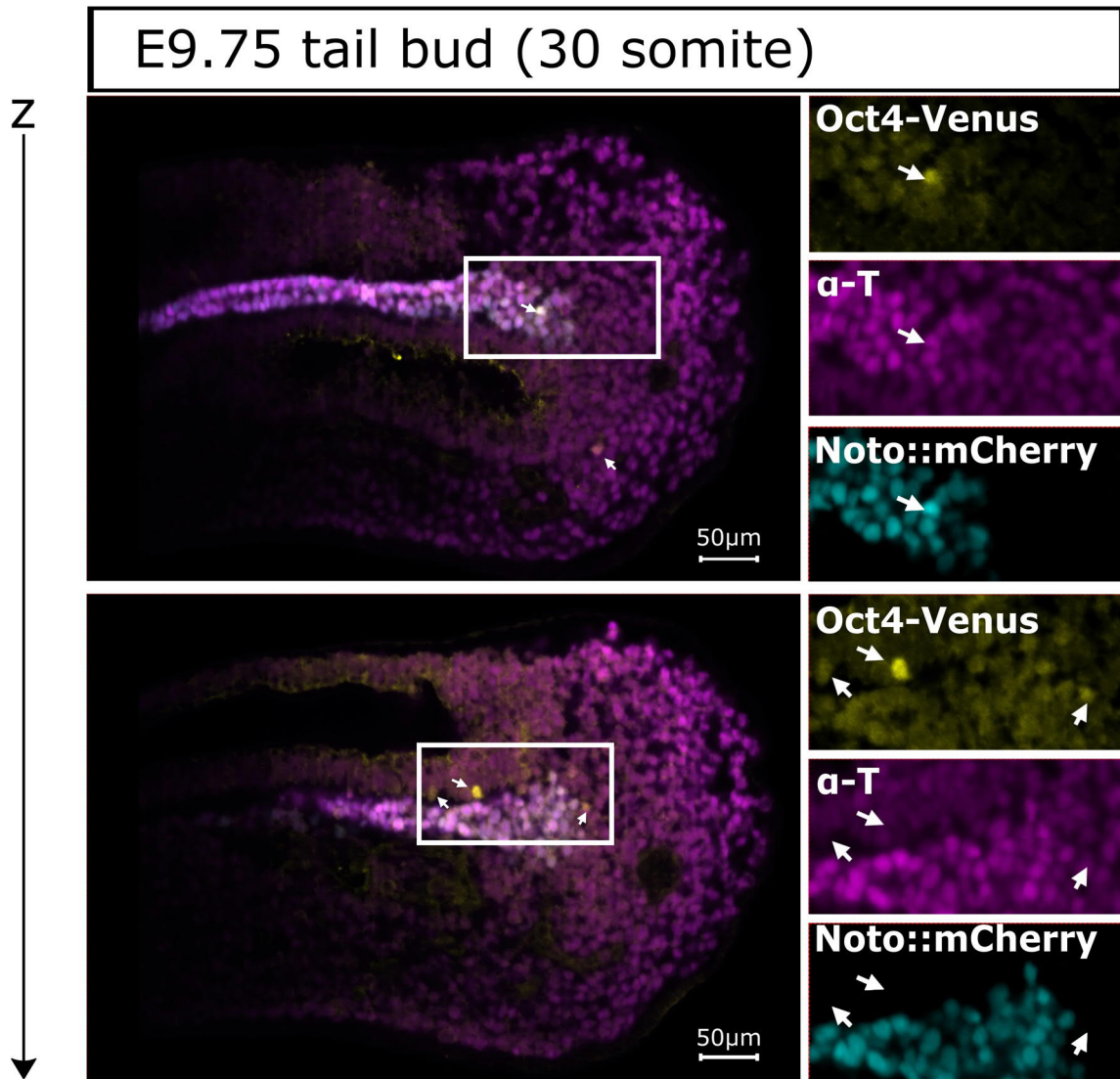
Figure 2.4.6 Oct4+/T+/Noto+ cells at the tip of the notochord

Light sheet imaging showing optical sections through the tail bud of an E9.75 embryo. Immunofluorescence for EGFP (yellow), T (magenta). *Noto::H2B-mCherry* signal in light blue. Depth of focus = max. 1.3µm. Arrows highlight individual Oct4-Venus+ cells.

In order to characterize the expression profile of these putative novel stem cells, transgenic embryos generated from the Oct4-Venus mESC line were isolated at E9.75 and tail buds were trimmed at the first visible somite border from specimen with 28-30 somites. Oct4-Venus[high] cells were purified using flow cytometry (Figure 2.4.7 A). Two reference samples were included: Oct4-Venus[high] cells from the allantoic bud, presumambly containing PGCs and Oct4-Venus[low] cells from the E9.75 tail buds, which did not really separate from the negative population and therefore may be comprised of autofluorescent cells (Figure 2.4.7 A). RNA extraction, library preparation and deep sequencing was performed as described earlier (2.4.2).

Gene expression of several known markers of NMPs, notochord, PGCs and neuroectoderm was analyzed in the samples isolated from the headfold stage NSB (Noto[mC] -/Oct4-Venus+, Noto[mC] +/Oct4-Venus-, Noto[mC] +/Oct4-Venus+), the E7.5 allantoic bud (PGCs) and the Oct4-Venus[high] and Oct4-Venus[low] (Figure 2.4.7 B). As expected, *Pou5f1 (Oct4)* transcript was not detected in the Oct4-Venus[low] sample, which can therefore be considered a reference sample comprised by more differentiated tail bud cells.

Strikingly, all analyzed previously proposed marker genes for NMPs and axial stem cells were expressed in Oct4-Venus[high] cells, including *T*, *Sox2*, *Wnt3a*, *Lin28a/b*, *Fgf8*, *Fgfr2*, *Nkx1-2*, *Gdf11*, *Cdx2*, *Sp5*, *Cyp26a*, *Snail*, *Hoxb8* and *Evx1* (Amin et al., 2016; Bell et al., 2016; Boulet and Capecchi, 2012; Dias et al., 2020; Herrmann et al., 1990; Kennedy et al., 2016; Robinton et al., 2019; Rodrigo Albors et al., 2018; Takada et al., 1994; Wymeersch et al., 2016).

In addition to *Oct4*, *Sox2* and *Lin28a/b*, Oct4-Venus[high] cells also expressed pluripotency markers like *Utf1* and *Dppa3* and *Dppa5a* (Kim et al., 2005; Okuda et al., 1998; Sato et al., 2002). This further substantiates the stem-like character of these cells, though it cannot be excluded that these transcripts were detected due to contamination with the sparse Oct4+ cells in the surface ectoderm (Supplementary Figure 8).

In summary, Oct4 activity was localized in the node streak border and chordo-neural hinge at the stages when the progenitors of trunk and tail are expanded. The activity of Oct4 in the tail bud has not been reported before. Within these domains, Oct4 is co-expressed with T and Sox2 as well as a large set of NMP markers and could be at the core of a signature for a novel class of axial stem cells.
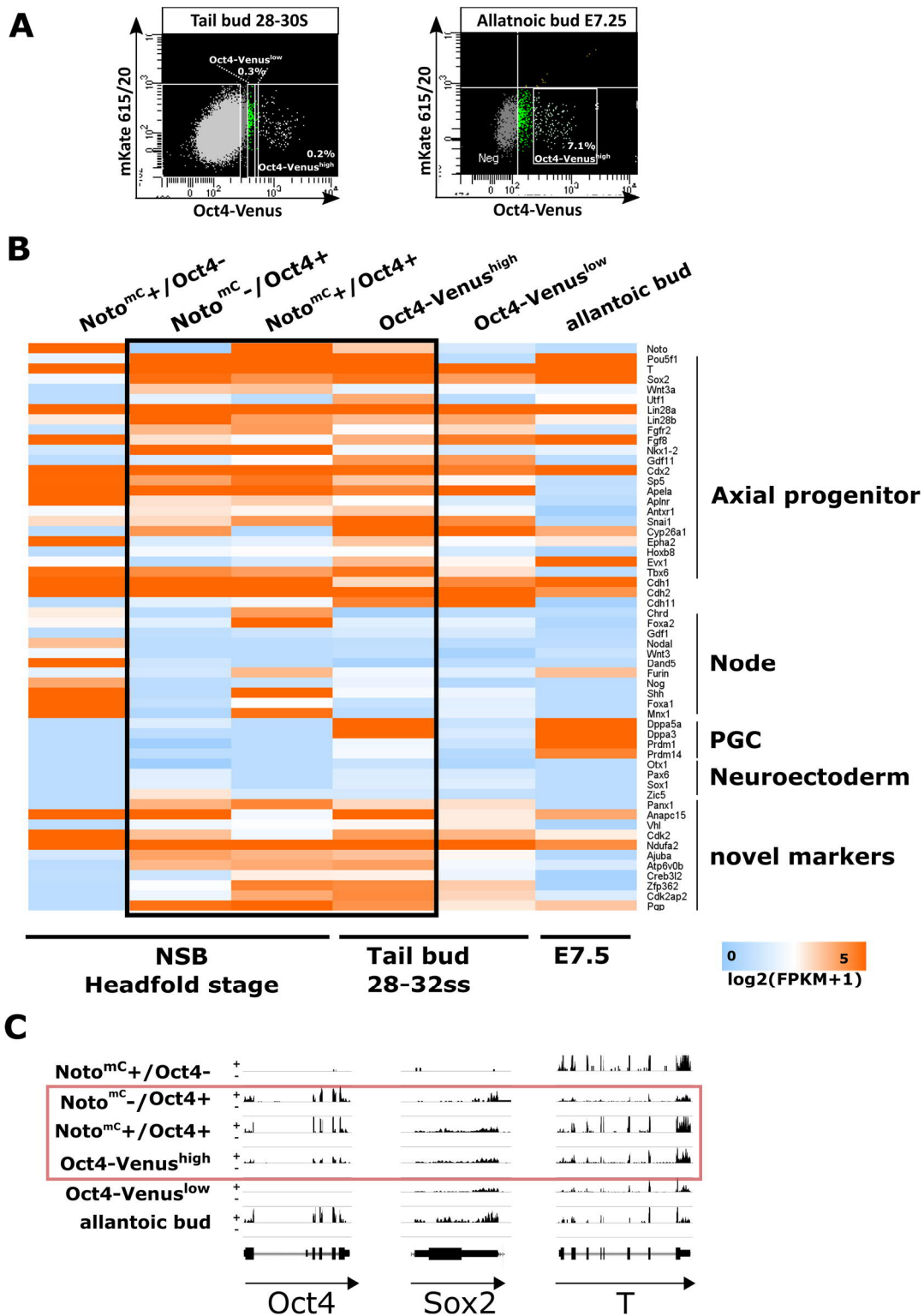
Figure 2.4.7 Transcriptome profiling of putative axial stem cells

(A) FACS profiles of the 28-30 somite stage tail buds and E7.5 allantoic bud samples from which the Oct4-Venus[high] and Oct4-Venus[low] populations and the PGC reference sample were sorted, respectively. FACS plots of headfold stages shown in Figure 2.4.4. Log-transformed FPKM expression values of selected genes grouped into proposed axial progenitor markers, genes expressed in node, primordial germ cells (PGCs), neuroectoderm and marker gene candidates with high expression values in the samples of interest. Square indicates samples containing Oct4+/T+/Sox2+ putative axial stem cells. (B) Normalized RNA-seq read density mapping to *Oct4*, *Sox2* and *T*. Wiggle tracks visualized in the Integrated Genome Browser (IGB).

2.5 Part IV – The axial mesoderm lineage

2.5.1 Contribution of notochord cells to the medial somites

Embryos carrying the *Noto:H2B-mCherry* reporter presented in this work showed contribution of Noto^{mC+} cells to the paraxial mesoderm of the trunk (Figure 2.5.1). Sectioning of micrographs revealed that these cells were localized in the medial region of the somite, which forms sclerotome (Figure 2.5.1). This observation has been reported previously in knock-in lines where the endogenous *Noto* locus was targeted with reporter (Noto$^{EGFP/+}$; Yamanaka et al., 2007) or inducible recombination driver cassette (Noto$^{nmCherry-CreERT2/+}$; Ukita et al., 2009). Therefore, it is not likely to be an artifact of the BAC transgene lacking distal regulatory elements, but indicates that some cells expressing *Noto* may not be terminally committed to axial mesoderm. The Noto^{mC+} signal in the somites decreased towards the anterior and was at all axial levels less strong than in the midline, suggesting it could be a residue of the stable H2B-mCherry reporter, which is expressed by caudal cells in an axial/paraxial mesodermal precursor state.

Figure 2.5.1 Contribution of Noto^{mC}+ cells to the medial part of the somites

Light sheet imaging of an E9.75 tail bud; Noto^{mC} signal (magenta) and DAPI staining (grey). Top left: Maximum intensity projection, lateral view. Line indicates the position of the optical section shown in the bottom panel. Right: frontal view. Arrows highlight individual cells.
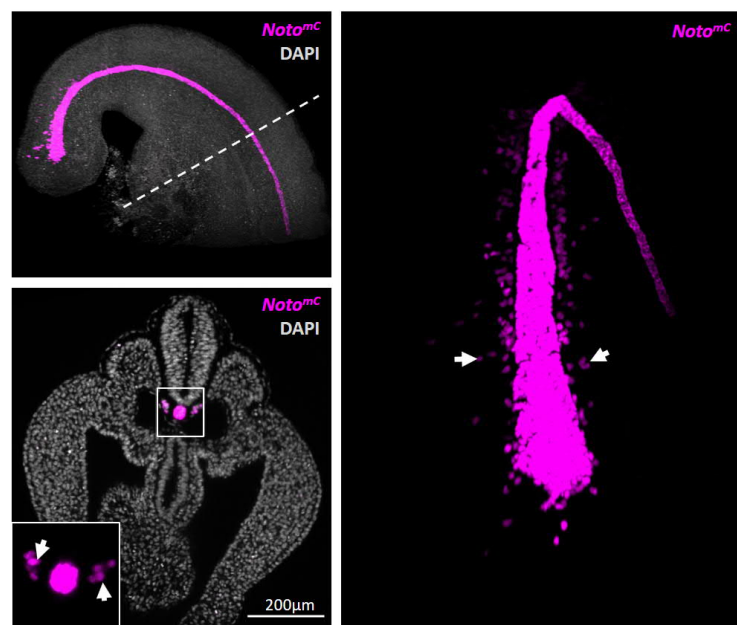
In addition to somites, Tbx6[Cre-ERT2] lineage tracing also was reported to label a small subpopulation of notochord cells (Concepcion et al., 2017) and conversely, the contribution of Noto[GFP/GFP] cells was enhanced in the *Noto* mutant (Yamanaka et al., 2007). These data are indicative of an axial vs. paraxial mesoderm fate switch downstream of *T* and could be mediated by antagonistic activities of *Noto* and *Tbx6*. To characterize the putative uncommitted cells, a *Tbx6::H2B-Venus* (Tbx6[Ve]) reporter BAC was introduced into the Noto[mC] mESC line and embryos were harvested at E9.5. Tbx6[Ve]+/Noto[mC]+ double positive cells were located at the posterior end of the notochord as well as in the medial part of the somites in the trunk (Figure 2.5.2). In order to investigate a possible role of *Tbx6* in notochord specification, a *Tbx6 -/-* null mutant was generated from Tbx6[Ve]/Noto[mC] reporter mESC line (Milena Pustet, Bachelor's thesis, University of Potsdam). To this end, exons 2 and 3 were deleted using CRISPR guide RNAs that target the *Tbx6* locus but not the Tbx6[Ve] reporter. Lightsheet microscopy of these mutants indicated that Tbx6[Ve] expression was weaker but detected also in the Noto[mC]+ population, which appeared to be expanded. FACS analysis revealed a 66% increase of Tbx6[Ve]+/Noto[mC]+ cells from 0.9% to 1.5% of the sample in *Tbx6 -/-* mutants (Figure 2.5.2 B).

In order to address a possible function of Tbx6[Ve]+/Noto[mC]+ cells, their transcriptome was analyzed and compared to single positive neighboring cells in the trunk and tail bud. Caudal ends were dissected cutting at the somite border of five E9.75 embryos per sample and Tbx6[Ve]+/Noto[mC]+ double positive cells were sorted from wild-type and *Tbx6 -/-* mutants in addition to Noto[mC]+ single positive wild-type cells (Figure). Further, Tbx6[Ve]+/Noto[mC]+ double positive cells as well as Tbx6[Ve]+ and Noto[mC]+ single positive cells were purified from single cell suspensions of trunks trimmed between the forelimb and hindlimb bud. The latter cells have downregulated endogenous *Noto* and *Tbx6* but are still labeled by fluorescent reporters. As described previously in this chapter, cells were purified by flow cytometry and RNA was extracted from 18-250 cells for library preparation and high throughput sequencing (Supplementary Table 2). After processing mapped reads to FPKM matrices, expression values were compared between all groups. Genes were filtered for expression in at least one sample (FPKM cutoff >2) and four fold up- or downregulation between at least two samples, resulting in a list of 6056 genes. After per gene normalization, k-means clustering was employed to group the differentially expressed genes into seven clusters (Figure 2.5.3).
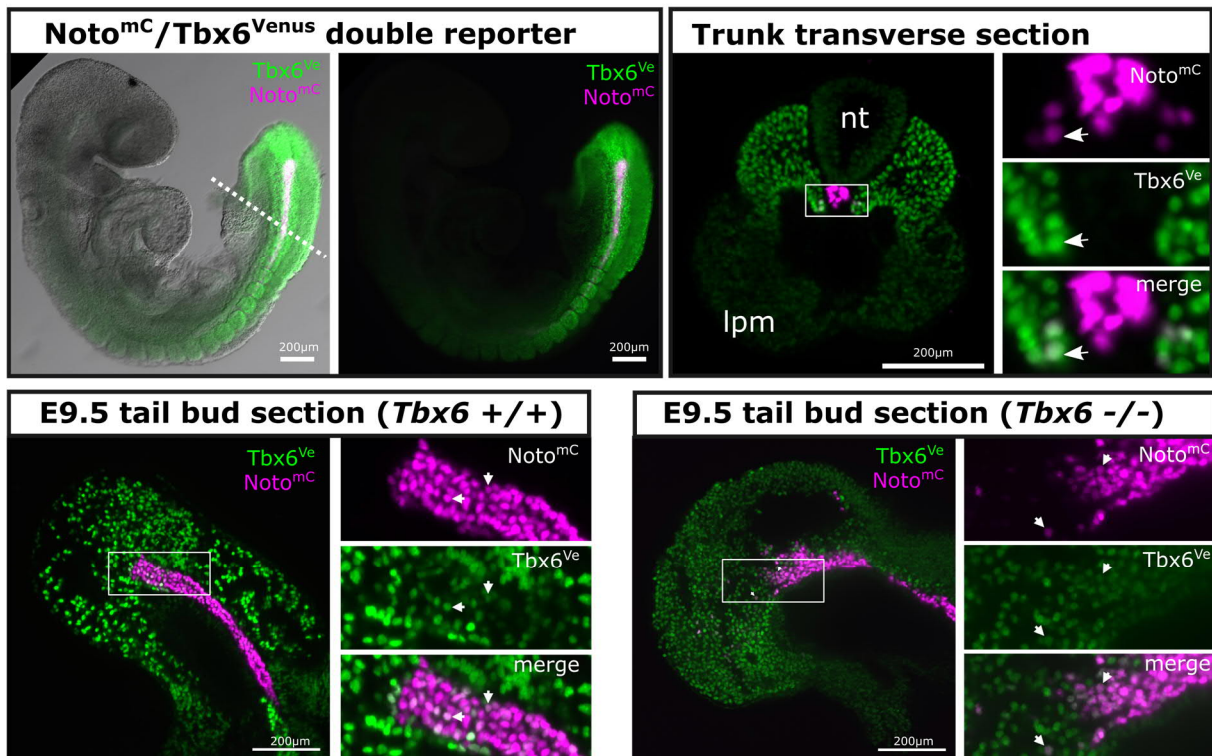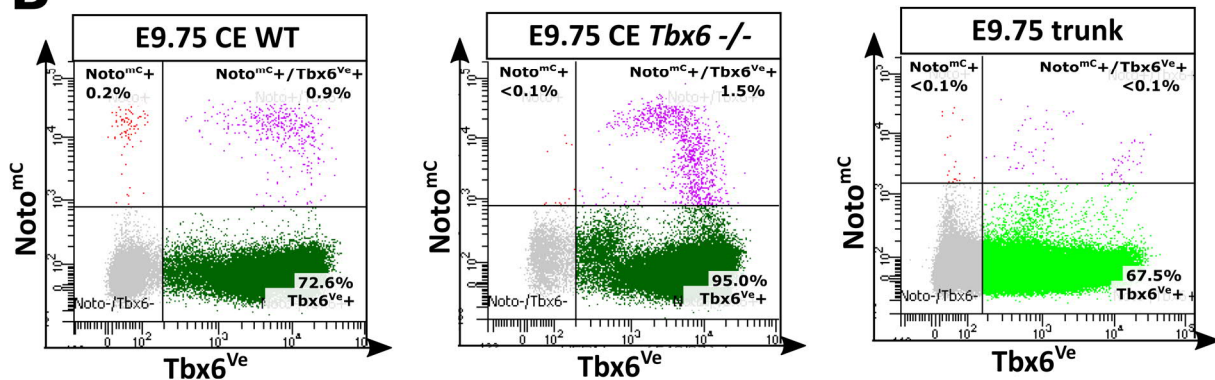
Figure 2.5.2 Coexpression of Tbx6^{Ve}+ and Noto^{mC}+ in wild-type and Tbx6 -/- mutants.

(A) Top left: E9.0 embryo with Tbx6^{Ve}+ and Noto^{mC}+ expression acquired with a stereomicroscope. Other panels: Optical sections acquired by light sheet microscopy. Depth of focus = 1.14µm. nt = neural tube. lpm = lateral plate mesoderm. Arrows highlight double positive cells. (B) FACS profiles of E9.75 WT and *Tbx6 -/-* caudal ends and trunks. Caudal end WT Noto^{mC}+, WT Noto^{mC}+/Tbx6^{Ve}+, *Tbx6 -/-* Noto^{mC}+/Tbx6^{Ve}+ populations and all trunk WT populations were subjected to RNA-Seq (Figure 2.5.3).

As expected, the Noto[mC]+ sample was predominantly represented in Cluster VI, containing a multitude of notochord markers and genes involved in notochord signaling (*Nog*, *Chrd*, *Nodal*, *Mnx1*, *Lgr5*; see chapter 2.3). Interestingly, *Tbx6* together with *Noto* and other genes that are active early in the notochord lineage (*Foxa2*, *Foxa1*, *Cdx2*, *Sp5*; see chapter 2.3) were grouped in Cluster VII and upregulated in all samples harvested from caudal ends, but most highly expressed in wild-type Noto[mC]+/Tbx6[Ve]+ cells (Figure 2.5.3). This observation is consistent with the hypothesis that the Noto[mC]+/Tbx6[Ve]+ state represents an earlier phase of notochord specification. Further, the double positive cells purified from *Tbx6 -/-* caudal ends expressed a series of NMP signature genes (*T*, *Sox2*, *Fgf8*, *Cyp26a1*, *Sp8*, *Nkx1-2*) found in Cluster IV (Koch et al.,2017; Gouti et al. 2017; Dias et al. 2020). In the absence of Tbx6, paraxial mesoderm differentiation cannot be locked (Chapman and Papaioannou, 1998), which might keep the mutant cells in a more undifferentiated, NMP-like state.

The genes upregulated in the trunk were concentrated in clusters I, II, III and V: Cluster I is enriched for marker genes of trunk notochord (*Prickle2*, *Wnt5b*, *Cthrc*, *Slit* and *Bicc1*; see chapter 2.3) expressed in the Noto[mC]+ sample. Consistent with their localization in the embryo, Noto[mC]-/Tbx6[Ve]+ expressed high levels of myotome markers like *Myf5*, *Myl4*, *Pax3*, and *Pax7* (Bober et al., 1994; Jostes et al., 1990; Lyons et al., 1990; Ott et al., 1991) found Cluster III. The Noto[mC]+/Tbx6[Ve]+ double positive cells were mainly located in the medial somite, which will give rise to the sclerotome and ultimately vertebrae and ribs. Cluster V, representing genes specifically active in trunk Noto[mC]+/Tbx6[Ve]+ cells includes chondrogenesis markers such as *Sox9* (Akiyama et al., 2002) and multiple members of collagen clusters (*Col1a2*, *Col5a1*, *Col9a1*, *Col9a3*; Kadler et al., 2007). Cluster II is comprised by genes that were highly expressed in the Noto[mC]+/Tbx6[Ve]+ and Noto[mC]+/Tbx6[Ve]- samples, including the sclerotome markers *Pax1*, *Pax9*, *Foxc2* and *Nkx3-1* (Kume et al., 2001; Peters et al., 1999; Tanaka et al., 1999). The expression of these well established paraxial markers in the notochord is not intuitive and the Noto[mC]+/Tbx6[Ve]- sample, from which only 18 cells could be purified and which clustered separately from all other samples (Supplementary Figure 7) might have been of poor quality.

Notably, Noto[mC]+/Tbx6[Ve]+ cells specifically express signaling factors that are essential for somite patterning and the induction of chondrogenesis like *Shh*, *Tgfb1* (Murtaugh et al., 1999; Wu et al., 2016). It might be intriguing to explore whether there is a specific functional role

for descendants of axial mesoderm progenitor cells in the somites, such as the initiation of cartilage formation.

In summary, these data indicate that there is a short window of time after the onset of *Noto* expression when the *Noto+/Tbx6+* cells are not yet committed and can still contribute to paraxial mesoderm. This window might be widened in the *Tbx6 -/-* mutant resulting in an accumulation of NMP-like cells differentiating towards a Noto^{mC}+ state.
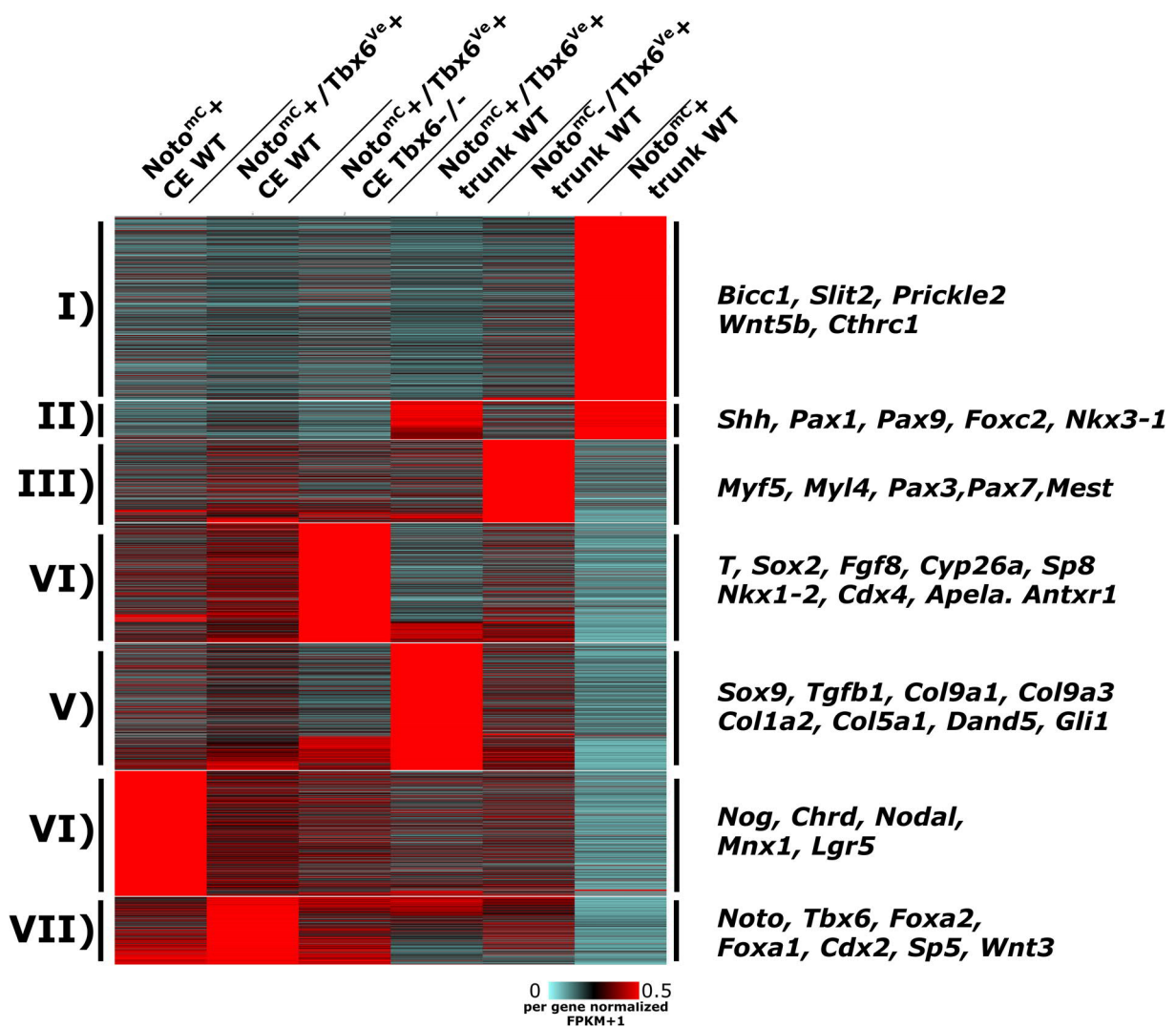
Figure 2.5.3 RNA-Seq Differential gene expression in different Tbx6^{Ve}+ and Noto^{mC}+ populations.

Per gene normalized FPKM values of differentially expressed genes (4FC; FPKM>2 in at least one sample). 6213 genes grouped into 7 clusters by k-means clustering. Exemplary genes for every cluster shown on the right. I) Mature notochord markers, II) Sclerotome, III) Myotome, IV) NMP, V) Chondrogenesis, VI) Posterior notochord VII) Notochord pecursors

2.5.2 Fate mapping of Noto+ cells throughout axial elongation

After identifying and characterizing descendants of Noto+ cells in the trunk paraxial mesoderm, the long term fate of these cells was determined by lineage tracing. To this end, a two component system was employed for genetic labelling of Noto+ cells: First, a BAC drives iCre recombinase expression under the control of the *Noto* locus (*Noto::iCre-pA-PGK-Hygro-pA*, herafter Noto^{iCre}). This transgene was randomly integrated into mESCs with modified Rosa26 locus carrying a gene trap reporter transgene (Soriano, 1999). iCre recombinase facilitates *loxP* mediated excision of a stop cassette and ubiquitous expression of β-Galactosidase under the control of a the ubiquitous Rosa26 locus (Figure 2.5.4 A; Soriano, 1999). In this way, all progeny of *Noto* expressing cells are marked by *bGal* expression, which can be visualized by X-gal conversion.

Embryos generated from this cell line by diploid aggregation were isolated between E9.0 and E13.0 and subjected to lacZ staining. In addition, microtome sectioning was performed (Figure 5.2.4). Reporter expression faithfully marked the trunk notochord at all axial levels. Consistent with *Noto* expression anterior to the primitive streak in a region containing cells that are not committed to axial mesoderm, the foregut was labelled in all specimen. By extension, derivatives of the foregut, like the epithelium of the lung were labelled (Figure 2.5.4 B). Sectioning revealed that sporadically, floor plate and hindgut cells expressed bGal, which indicates that some Noto^{mC}+ cells in the dorsal node (Figure 2.4.4) could be floor plate progenitors.

The most striking domain of Noto^{iCre}+ expression apart from the notochord however was detected in the medial part of every trunk somite pair between forelimb and hindlimb bud. As expected, Noto^{iCre}+ cells stayed regionalized in the sclerotome and ultimately the neural arch, which is the part of the vertebrae enclosing the spinal cord (Figure 2.5.4 B). This morphogenetic development becomes evident at E12.5. Notably, the highest number of bGal labelled cells was found in the cervical segments, coinciding with the area where crown cell descendants start to give rise to the notochord (Wymeersch et al., 2019).

In conclusion, these experiments validate that considerable numbers of Noto+ cell descendants in the paraxis are ultimately fated for vertebrae.
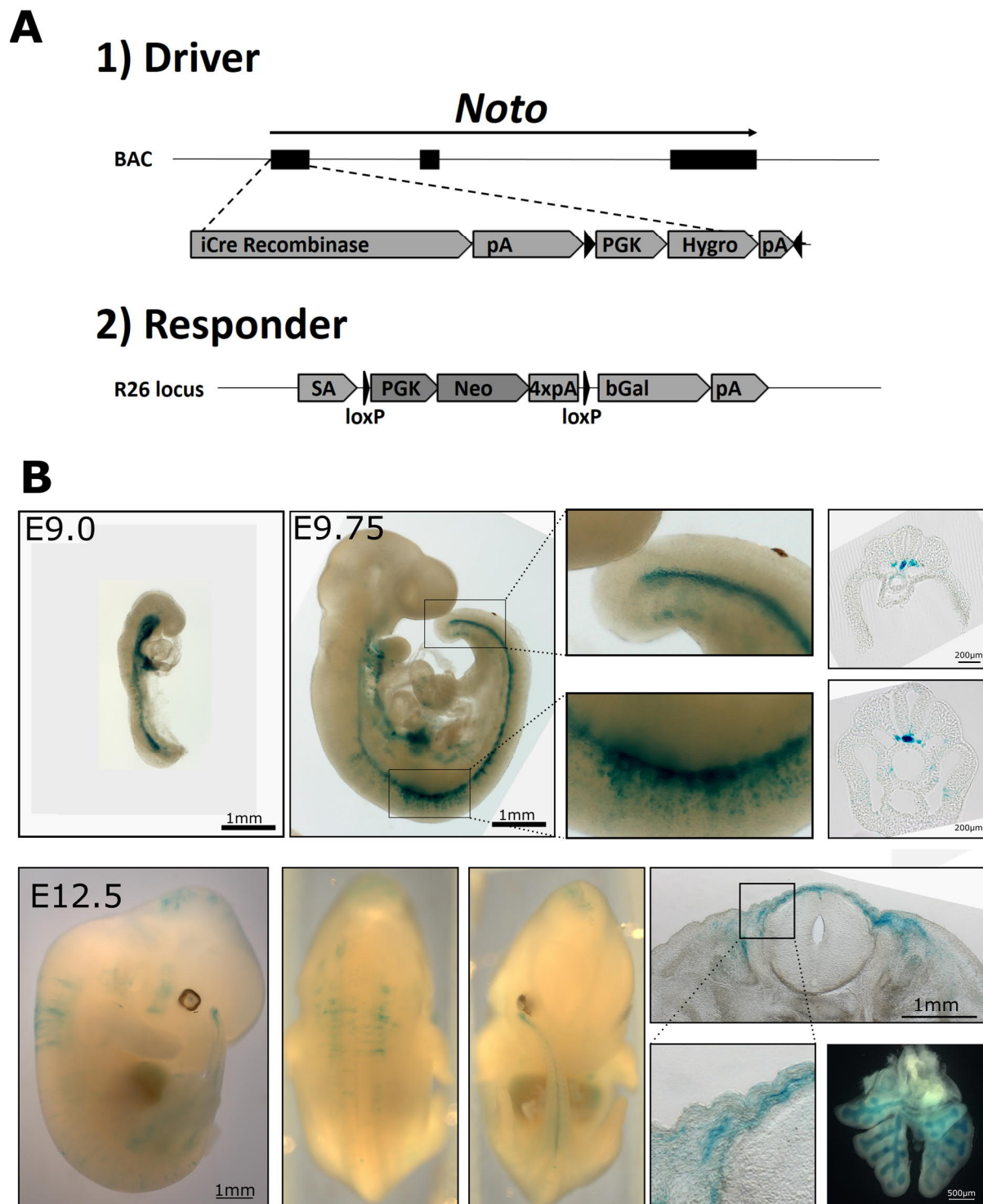
Figure 2.5.4 Fate mapping of Noto+ cells throughout embryogenesis.

(A) Lineage tracing construct consisting of a BAC driver randomly inserted into mESC with a responder transgene integrated into the Rosa26 locus. Upon recombination, the stop cassette is excised and the bGal reporter expressed under the control of the *Rosa26* locus. pA = polyadenylation signal. PGK = PGK promoter. Hygro Hygromycin resistance. SA = splice acceptor. Neo = Neomycin resistance. (B) Embryos E9.0-E12.5 after lacZ staining. Right panels: Microtome sections of 10 μm (E9.75) and 20 μm (E12.5).

2.6 Contributions

Bernhard G. Herrmann conceived and supervised the project. Bernhard G. Herrmann and Frederic Koch identified Oct4 as a putative axial stem cell marker. Frederic Koch prepared the RNA-sequencing libraries of samples shown in Figure 2.4.7 and provided code for mapping of RNA and ChIP seq reads as well as material for cloning. Manuela Scholze-Wittler generated the Foxa2^{mT} and T^{mC} reporter BACS and provided the wild type mESC line. Manuela Scholze-Wittler and Lars Wittler established the *T/tw5* and R26::lacZ mESC lines from blastocysts. Jesse Veenvliet established the *in vitro* differentiation protocol for notochord cells and supervised FACS on for the *Noto^{mC}* transcriptome. Milena Pustet engineered the *Tbx6 -/-* knock out and Noto^{mC}/T^{Ve}/Foxa2^{mT} triple notochord reporter mESC line. Diploid and tetraploid complementation assays were performed by the transgenic unit of the MPI-MOLGEN including Lars Wittler, Karol Macura, Judith Fiedler, Dijana Mićić, Adrian Landsberger, Christin Franke and Mirjam Peetz. FACS was operated by the Flow Cytometry Facility of the MPI-MOLGEN including Uta Marchfeldner and Claudia Giesecke-Thiel. Sequencing of ChIP and RNA-seq libraries was performed in the Sequencing Core Facility of the MPI-MOLGEN including Bernd Timmermann and Norbert Mages.

3 Discussion

3.1 Dissecting the regulatory landscape of *Brachyury*

Distinct activities of *Brachyury* perpetuate axial elongation and are essential to key functionalities of the caudal growth zone including mesoderm formation, NMP state maintenance and notochord specification. In all these processes, T operates in a dosage dependent manner. T levels in the corresponding expression domains are controlled by different regulatory elements.

For instance, a 23kb fragment of the T locus comprising the *T* gene and promoter as well as the most proximal enhancers, controls the transient activity required in nascent mesoderm of the primitive streak and tail bud (Figure 2.1.2.; Stott et al 1993). This fragment can completely rescue the short tailed phenotype of the *T +/-* heterozygous mutant (Stott et al., 1993). Random integration of the transgene enables *T -/-* cells to colonize mesoderm in chimeric embryos in a way that is proportional to the integration copy numbers (Wilson and Beddington, 1997). Therefore, this minimal region of the locus is sufficient to induce mesoderm formation. In this relatively short time frame, T acts as a pioneering factor that activates mesodermal gene regulatory networks by recruitment of histone modifying enzymes (Beisaw et al., 2018; Tosic et al., 2019). As the cells differentiate further towards the mesodermal sub-lineages, T is downregulated.

In contrast, to the transient activity in cells undergoing mesodermal differentiation, T expression in notochord and the NMP reservoir needs to be persistent for the maintenance of these structures throughout axis extension from E8.0-E13.5. The promoter cannot drive reporter expression in the notochord and the 23kb region cannot functionally complement the notochord and tail phenotype in *T -/-* (Clements et al., 1996; Stott et al., 1993). In addition, notochord development and tail bud outgrowth, which is facilitated by NMPs, are impaired in the T^{BoB} mutant that has a 200 kb insertion 16 kb upstream of *T* (Rennebeck et al., 1995). Therefore, the enhancers that are required for NMP maintenance and notochord formation

are located outside the 23 kb fragment and probably upstream of T and the insertion in T^{BoB} either deletes these elements or interferes with enhancer-promoter interactions.

In the first part of my thesis, I aimed to determine the location of enhancers on the T locus. ChIP-Seq for T in *in vitro* derived NMPs (Koch et al., 2017) and notochord (Figure 2.1.2) revealed that T binds its own locus multiple times implying the importance of autoregulation. Most of these binding sites are conserved in mammals, which further substantiates a function as developmental enhancers.

An upstream regulatory region confers T activity required for notochord specification and tail outgrowth

As a starting point for a functional dissection of T activities, I deleted the complete upstream regulatory region spanning from the first prominent T binding site ($TE1$) to the 5′ end of the 23kb fragment using CRISPR/Cas9, resulting in the T^{UD}/T^{UD} mutant. In contrast to embryos carrying the T^{CD}/T^{CD} null alleles, the T^{UD}/T^{UD} mutants formed all somites of the trunk and established the chorio-allantoic fusion, which allowed them to develop at least until E12.5. However, the formation of the tail bud, as well the specification of the notochord were severely impaired in T^{UD}/T^{UD}. Both defects emphasize the relevance of T dosage in these processes.

I showed that T expression was not maintained in T^{UD}/T^{UD} notochord progenitors, which were initially generated and even gave rise to a node-like structure at E7.5, but did not separate from the endoderm and cohorts of remnant Noto^{mC}+ cells were eventually detected in the gut tube. A similar failure of notochord progenitors to become specified was reported in *Noto-/-Foxa2 +/-* double mutants, where these cells are ectopically located in paraxial mesoderm (Yamanaka et al., 2007). This suggests a mechanism for the transformation of the node to trunk notochord in which first, T is required for the detachment of axial mesoderm from the endodermal layer. In a second step, Foxa2 and Noto might cooperate to uphold convergence, epithelialization and polarity of notochord progenitors in the midline (Burtscher and Lickert, 2009). Consequently, loss of T leads to misspecification of notochord progenitors towards

endoderm as shown in this study and loss of Foxa2 and Noto allows the notochord progenitors to acquire paraxial mesoderm fate as demonstrated before (Yamanaka et al., 2007).

The other main feature of the T^{UD}/T^{UD} mutant with implications for axis development is the defect in mesoderm specification leading to an arrest of axis extension after the hindlimb bud. The trunk somites were generated in T^{UD}/T^{UD}, but were smaller and contained Sox2 expressing cells. Further, the T^{UD}/T^{UD} neural tubes were clearly enlarged towards the posterior end of the trunk, indicating the production of excess neuroectoderm at the expense of mesoderm. These abnormalities likely represent an imbalance in the neural vs. mesodermal fate choice conferred by antagonistic T and Sox2 activities (Koch et al., 2017). It is possible that due to the T deficiency in T^{UD}/T^{UD}, the NMP reservoir is gradually depleted as the trunk develops and completely exhausted at the hindlimb bud. Additionally, the excess neural tissue in T^{UD}/T^{UD} formed a thickened epithelium, which inhibited neural tube closure and completely enveloped the caudal end at E9.5. Thus, it could also be this disruption of morphology, which finally leads to the arrest of axial elongation.

The T^{UD}/T^{UD} phenotype presented in my thesis is consistent with the defects in the T^{BoB} mutant in which the T enhancer cluster is disrupted by a large insertion (Rennebeck et al., 1995). Therefore, it is likely, that with the 40 kb regulatory region, I deleted a set of specific enhancers that boost T activity to the higher levels required for notochord specification and tail bud development.

TE2 is an essential notochord enhancer of T and critical for tail development

In a systematic dissection of the T locus, I identified *TE2*, which is a conserved 1 kb sequence 35 kb upstream of T that functions as a notochord enhancer. I showed that *TE2* is active in the posterior tip of the notochord, which is the region where Noto+ cells are specified (Figure 2.13). Ablation of *TE2* led to a complete absence of tail notochord with abnormalities starting at the levels of the hindlimb bud. Transcriptome analysis revealed that Noto+ cells in the posterior trunk notochord express multiple activators of the canonical Wnt pathway, making them more sensitive to Wnt signaling (Figure 2.3.2). Both ChIP-Seq data presented in my work and published datasets confirm that *TE2* is bound by T and β-Catenin (Koch et al., 2017), as

well as Sp5 (Kennedy et al., 2016) and can therefore be considered a Wnt responsive element. Wnt and T are known to form a positive feedback loop (Martin and Kimelman, 2008; Yamaguchi et al., 1999) and it is likely that the high T dosage required for notochord specification is potentiated via *TE2*, which might integrate both T autoregulation and Wnt signaling. Accordingly, the *ΔTE2/ΔTE2* enhancer mutant phenotype shown in my work is identical to those reported for notochord specific β-Catenin ablation or Brachyury knock-down (Ukita et al., 2009; Zhu et al., 2016).

The enhancer reporter assay as well as the more severe phenotype of the heterozygous background ($T^{cd}/ΔTE2$) demonstrated that *TE2* is already active at the earlier stages of notochord development, even though the phenotype of *ΔTE2/ΔTE2* was evident only in the more posterior portion of the trunk notochord. Therefore, *TE2* functions redundantly during specification of trunk notochord and becomes essential for tail notochord.

The loss of *TE2* accounts for a significant part of the T^{UD}/T^{UD} mutant phenotype, yet there are some differences (Figure 5.1). In contrast to the impaired tail bud outgrowth in T^{UD}/T^{UD}, *TE2* knock out tails extended and developed about 15 of the 30 post-anal somites. In *ΔTE2/ΔTE2*, the paraxial mesoderm and neuroectoderm emerging from the caudal growth zone were generated, but did not develop further and quickly induced apoptosis (Figure 2.1.10). This could be a secondary effect of the loss of Shh signal which is required for somite differentiation (Chiang et al., 1996).

I further showed that at the truncated caudal end of the *TE2* mutants, the posterior growth zone branched into multiple poles of T and Sox2 co-expressing cells, indicating that the medio-lateral organization is perturbed without a signaling center in the midline. Since in most $T^{cd}/ΔTE2$ embryos, Noto^{mC}+ cells were completely absent in the tail, the NMPs might be able to elongate the axis to some extent without the putative niche provided by notochord progenitors. Axial extension without further differentiation was also reported in 'gastruloid' *in vitro* models, which also lack structured axial mesoderm, but elongate and produce somites and neuroectoderm independently for about one day (van den Brink et al., 2020; Veenvliet et al., 2020). In the enhancer mutant tails of this study, it is not possible to conclusively determine whether the NMPs are dependent on the Noto^{mC}+ cells due to the increasing secondary effects and cell death. It remains possible that the immediate NMP descendants can still give rise to a certain portion of the axis, which would lead to a delay between NMP

depletion and axial truncation. Alternatively, it has also been postulated before that the reservoir of NMPs is amplified during trunk development and gradually exhausted afterwards (Wymeersch et al., 2016). In this case, the loss of tail notochord progenitors would not affect the pool of NMPs and solely induce patterning defects.
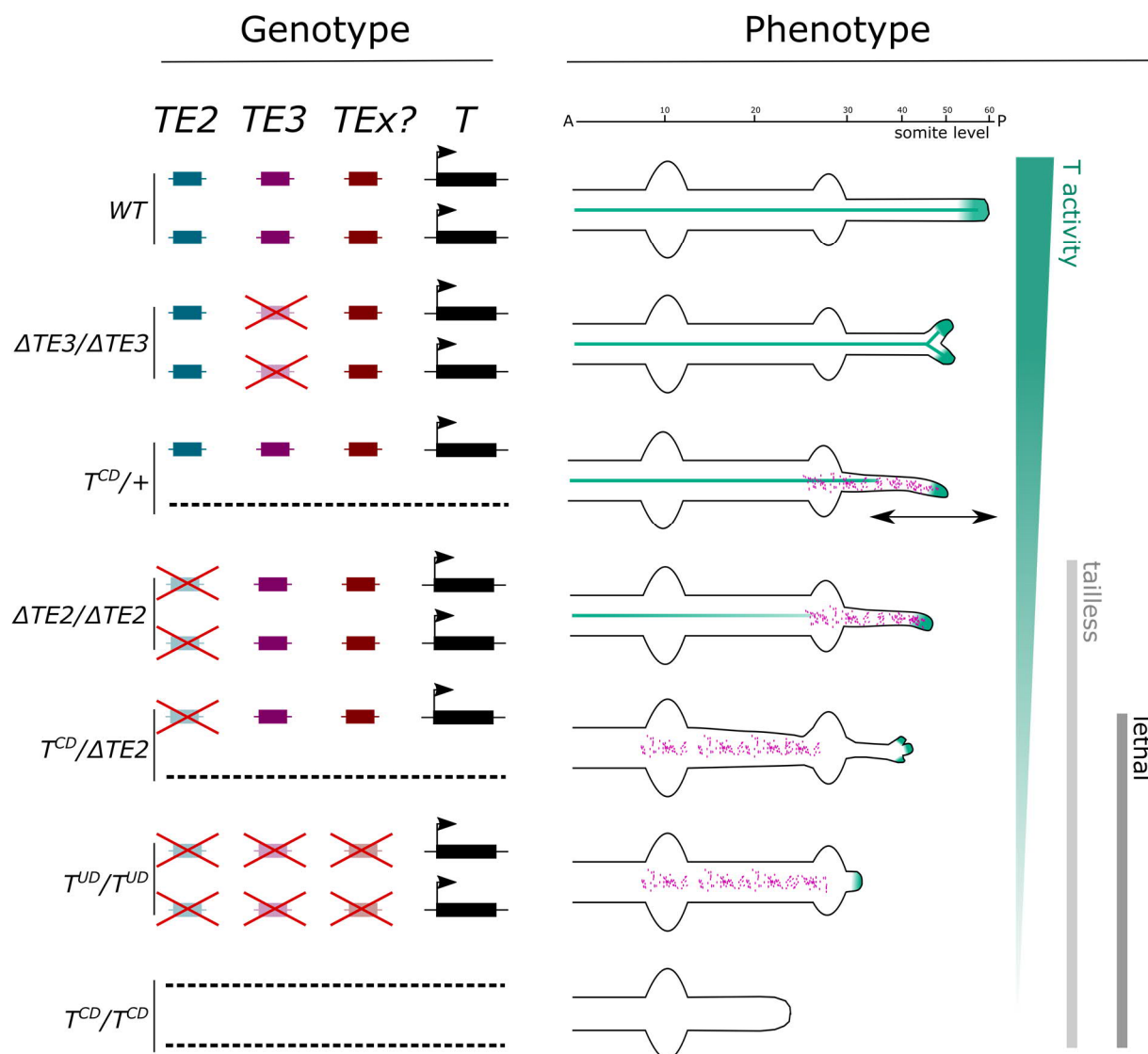
Figure 3. 1 Enhancers embedded in the upstream regulatory region of *Brachyury*

Left: Schematic of the upstream regulatory region of T. Crossed enhancer elements and dotted lines indicate CRISPR deletions. The exact genomic coordinates for each genotype are listed in Supplementary Table 1. *TEx?* represents one or more unidentified enhancer(s) within the T^{UD} region that cooperate with *TE2* in early notochord development and mesoderm formation. Right: Phenotypes corresponding to the mutants on the left side, ordered according to severity from wild type to null mutant. Colors indicate T expression (green) and the location of mis-specified Noto^{mC}+ cells (magenta). The double-headed arrow indicates the variability of notochord and tail extension in T^{CD}/+. Since the embryos were not analyzed until E13, lethal and tailless phenotypes were implied from the severe defects at this stage.

I showed that *TE2* meets all criteria of an enhancer. However, since the CRISPR deletion comprised one exon of the *T2* gene, it was necessary to exclude a role for *T2* in axis development. Neither a frameshift mutation in the *T2* open reading frame, nor the deletion of other exons did phenocopy *ΔTE2/ΔTE2* (Supplementary Figure 2; Table 2.1). My findings confirm that the *T^BoB* mutant phenotype is caused by dysregulation of Brachyury and *T2* does not play a role in axis development, which is in line with the revised interpretation of *T^BoB* mutant by the authors who cloned the *T2* gene (Wu et al., 2007).

The similarities of *ΔTE2/ΔTE2* and *T/tw5* phenotypes also raised the question whether *TE2* could be allelic with *tct*. Indeed, I identified some point mutations in the *tw5* variant of *TE2*. I used CRISPR/Cas9 and a wild type oligo template to restore the BL6 sequence in the core of the enhancer, which did not rescue the phenotype (Figure 2.1.11, Supplementary Figure 4). This indicates that either, *TE2* and *tct* are distinct regulatory elements, or that the hypomorphic activity of *TE2* in *tw5* could have other reasons, like reduced enhancer-promoter interactions. Since the *tw5* version of chromosome 17 has accumulated several large inversions and deletions that are protected from recombination (Schimenti, 2000), changes in chromatin topology might reduce *TE2* enhancer activity. In general, the T dosage threshold is barely met for tail notochord specification in *T/+* heterozygotes, leading to a variability in tail length. Therefore, a minor reduction of *TE2* activity in *T/tw5* might be just enough to produce a completely penetrant tailless phenotype.

TE3 is a regulatory element for the fine-tuning of T expression in mesoderm progenitors

The second enhancer mutant of my study that resulted in a phenotype was *ΔTE3/ΔTE3*, inducing a loss of T activity in CNH and dorsal hindgut. At approximately E11.5, the tail notochord split leading to a bifurcation of the tail with two T positive poles at E12.5 (Figure 2.2.2). This phenotype is strikingly similar to what has been reported for the *Epha2* homozygous null mutant (Naruse-Nakajima et al., 2001). In their study, the authors proposed a mechanism for the segregation of paraxial and axial mesoderm suggesting that *Epha2* receptor expressing notochord progenitors are repelled by *Efna1* ligand expressing paraxial mesoderm progenitors (Naruse-Nakajima et al., 2001). After *Epha2* ablation, this hypothesized interaction is disrupted and the notochord progenitors are no longer confined

90

to the midline, enabling them to intrude the tail bud mesenchyme and induce the bifurcation. My transcriptome data of mesoderm progenitors versus Noto^{mC}+/T^{Ve}+/Foxa2^{mT}+ nascent notochord verified a differential expression of *Efna1* and *Epha2*, respectively (Supplementary Figure 9). Further, ChIP-Seq data of in *in vitro* derived mesoderm and notochord revealed strong T binding sites proximal to both genes (Frederic Koch, unpublished data; Supplementary Figure 9). It is therefore possible, that the *Efna1* and *Epha2* loci are regulated by Brachyury and highly sensitive to T expression levels, thereby providing a mechanism for the controlled spatial allocation of different mesodermal tissues. It would be interesting to test whether *Efna1* is downregulated in *ΔTE3/ΔTE3* and if mutations in the T binding sites at the *Efna1* and *Epha2* loci result in a similar phenotype.

I further showed that deletion of *TE3* in the *T^{cd}/+* heterozygous mutant induced a less ambiguous phenotype and clearly induced the production of excess neuroectoderm in the form of additional neural tubes (Figure 2.2.2 H). Therefore, ablation of this *T* enhancer may further reduce the variable haploinsufficiency phenotype of *T^{cd}/+* and inhibit the mesodermal differentiation of NMP descendants, which ectopically adopt a neural fate.

Redundant enhancer activity in the T locus

ΔTE2/ΔTE2 did not fully recapitulate the severe genotype caused by *T^{UD}/T^{UD}*. Therefore, one or more enhancers that cooperate with *TE2* in notochord development and control T activity in NMPs are embedded in the upstream regulatory region. Neither *ΔTE2/ΔTE2; ΔTE3/ΔTE3*, nor *T^{cd}/ΔTE2; ΔTE4-5* double knock-outs had impaired tail outgrowth phenotypes (Table 2.1). Thus, the missing element or elements (Termed *TEx* in Figure 3.1) are yet to be identified. Transcriptional control by multiple, enhancers with redundant activities is common especially for powerful developmental regulators (Gorkin et al., 2020; Osterwalder et al., 2018) and it is reasonable that this is the case for *Brachyury*. Therefore, it might be difficult at some point to deconstruct the modularity of T functions completely by additive deletions of TF binding sites. Another interesting aspect of *Brachyury* regulation was reported recently in human chordoma cell lines, where T auto-regulates itself and controls the chordoma programme via transcriptional condensates (Sheppard et al., 2021). It would be compelling to explore whether a similar mechanism could play a role in the enhancer cluster I deleted in *T^{UD}*.

3.2 Investigating the organizer of axial elongation

In this study, I sought to explore whether axial elongation is driven by an organizer that functions as two-component system comprised by presumptive notochord cells that provide a niche for axial progenitors.

In mammals, the node meets the classical criteria of an organizer for trunk and tail development: Upon transplantation, it can induce an axis lacking anterior structures, it self-differentiates into notochord and it patterns neighboring tissues (Beddington, 1994; Kinder et al., 2001; Sulik et al., 1994). In addition, multi-fated, resident progenitors with stem-like properties are closely associated with the node and posterior notochord throughout axis development (Cambray and Wilson, 2007; Lawson et al., 1991; Wilson et al., 2009). Recently, single cell transplantation experiments have demonstrated that the node environment functions as an instructive niche in chick (Solovieva et al., 2020) and a similar concept has been proposed for the mouse (Wymeersch et al., 2019).

The terminology around these findings are still subject to discussion. It is not clear, whether the stem-like cells described so far are identical with long-term NMPs, parental to NMPs or a separate population. Further, they are not stem cells by the most stringent definition, because they undergo a reversible transcriptional maturation (Cambray and Wilson, 2007). In this work, I chose the broad term "axial progenitor" to refer to all progenitor cell populations in the NSB and CNH that are exposed to signaling from the node and nascent notochord, including NMPs, direct NMP descendants, notochord progenitors and putative axial stem cells.

Signaling between presumptive notochord and axial progenitors

In order to identify potential niche factors secreted by the node and nascent notochord, I analyzed the transcriptome Noto+ cells from the posterior growth zone and referenced it to mis-specified mutant cells, differentiated notochord and directly adjacent progenitor populations throughout multiple stages of trunk development (Figure 2.3.2-4). NMP maintenance and mesoderm specification *in vivo* and *in vitro* is dependent on a network of T,

Wnt and Fgfs (Ciruna and Rossant, 2001; Gouti et al., 2014; Martin and Kimelman, 2012; Takada et al., 1994; Turner et al., 2014). Because Fgf is thought to act downstream of Wnt signaling (Aulehla et al., 2003) and reporter assays show the highest Wnt response in the nascent notochord (Ukita et al., 2009), the niche was expected to be formed by secreted Wnt ligands. However, the most promising candidates which are associated with axial truncation phenotypes, like *Wnt3a*, *Wnt5a* (Andre et al., 2015; Takada et al., 1994) or the extracellular activator *Rspo3* (Aoki et al., 2007), were not expressed or expressed at significantly lower levels than in neighboring cells in my dataset (Figure 2.3.2; 2.3.4). Interestingly, I found that localized *Wnt3* expression in Noto+ cells was gradually elevated towards posterior trunk development. *Wnt3* null mutants display severe gastrulation defects and cannot maintain the primitive streak (Liu et al., 1999). A role for *Wnt3* in later stages of development has not been reported so far and it would be interesting to explore this possibility, for instance with conditional knockout experiments, bypassing early embryogenesis.

My analysis of the canonical Wnt pathway components indicate that rather than being a source of Wnt ligands, the Noto+ cells proved to be especially receptive to Wnt signaling. Noto+ cells specifically expressed a number of Wnt receptors (*Lgr5* and *Fzd6*) and activators (including *Cd44*, *Lypd6*, *Lypd6b*, *Antxr1*, *Fermt1*, *Sp5*, *Sp8*, *Tcf7l1*, *Tcf7l2*) at every level of the pathway (Figure 2.3.2). NMPs were shown to produce high levels of *Wnt3a* and *Rspo3* (Koch et al., 2017). Therefore, it is conceivable that there is reciprocal signaling between NMPs and Noto+ cells and that the Noto+ cells require Wnt signal from the progenitor cells for their specification. The high expression of Wnt response genes would explain the reported increased Wnt activation in Noto+ cells, which was diagnosed using TOPGAL reporter systems, which make use of multiplexed Tcf/Lef binding sites and are therefore a transcriptional readout of the pathway (Ukita et al., 2009). Further, this interpretation of the data is consistent with the disruption of notochord observed in *Wnt3a-/-* mutants and a *Noto* driven β-Catenin knock out (Takada et al., 1994; Ukita et al., 2009).

In addition to playing a role in notochord specification, opposing gradients of Wnt5 signals from the primitive streak and Sfrp inhibitors from the anterior were shown to control node polarization along the AP axis (Minegishi et al., 2017). Thus, both the canonical and non-canonical PCP Wnt pathways are critically involved in notochord development.

Together with Wnt and Fgf, Nodal signaling was suggested preserve a less differentiated NMP state (Edri et al., 2019). I found that Nodal and Gdf1, which could have a synergistic function, are upregulated again in Noto+ cells towards the end of trunk development. Like Wnt3, these factors play important roles early in development, and conditional knock-out experiments are required to determine their function in mid-gestational development (Andersson et al., 2006; Zhou et al., 1993).

Fgf8, which is critical for gastrulation and cooperates with Fgf4 in NMP maintenance (Boulet and Capecchi, 2012; Sun et al., 1999) was upregulated in Noto+ cells. It becomes evident from published WISH data, that the transcript in axial mesoderm is especially concentrated at the posterior end, where the Noto+ cells contact axial progenitors (Cambray and Wilson, 2007). Therefore, Fgf8 could play an important role in the putative niche, even though it functions redundantly with Fgf4 (Naiche et al., 2011).

Finally, it might be worthwhile to investigate apelinergic signaling in the context of axial elongation. Apela, which is a peptide binding the Aplnr receptor, is specifically expressed in the posterior node and mutants display low penetrance axis phenotypes (Freyer et al., 2017).

In summary, my gene expression profile of Noto+ cells suggests that they are highly responsive to canonical Wnt. Secreted niche factors in Noto+ cells included Nodal, Wnt3, Fgf8 and Apela. Their expression peaked at headfold stages and became increasingly re-activated at 16 somite stage coinciding with the amplification of axial progenitors for the trunk and tail, respectively (Figure 3.2).

Evidence for a role of Oct4 in axial elongation

In the last years, there have been publications suggesting a role for Oct4 in axial development. A induced Oct4 knock out leads to axis truncations amongst other defects (DeVeale et al., 2013) and Oct4-overexpression in the progenitor domain results in increased trunk length (Aires et al., 2016). Also, relatively stable *Oct4::H2B-CFP* fluorescent reporter signal is detectable first in the trunk notochord of E8.5-E9.5 embryos and later in the tail bud, where

the domain expands from the neural tube and to the somites (Frederic Koch, MPIMG, personal communication).

Using an Oct4-Venus fusion line to determine the exact location of stem cells, I detected Oct4 activity in groups of cells residing specifically in the NSB and CNH matching with the regions that were shown to harbor long term axial progenitors (Cambray and Wilson, 2007).

Oct4 is expressed in different cell types at the Node-Streak Border

When the node emerges at around E7.0 it is composed entirely by Oct4 expressing cells that form two epithelial layers forming a distinct structure anterior to the primitive streak (Downs, 2008). One day later, at the onset of somitogenesis, Oct4 expression is downregulated in most of the node (Downs, 2008). Consistent with the published data, I did not detect Oct4-Venus signal in the Noto+/T+/Foxa2+ pit-shaped posterior notochord (PNC) cells at E8.25, albeit Oct4+ cells were localized in different populations of the node region (Figure 2.4.2-3; Figure 3.2). For the interpretation of these results, it is important to address these distinct cell populations that are distinguished by morphology and gene expression.

I found Oct4 expressed in the epithelial, more anterior portion of the dorsal node in cells that most likely have neural fate. However, in the very midline of the dorsal node there were cells that showed low Oct4 expression or express Oct4 together with Noto or T (Figure 2.4.2-3; Figure 3.2). In the transcriptome data I identified some genes that are involved in floor plate specification upregulated in these cells (Figure 2.4.5), but fate mapping of Noto+ cells only showed irregular contribution to the floor plate (Figure 2.5.4). Imaging caught some of these cells between the dorsal and ventral node epithelium, likely representing a delamination movement. It was shown previously that labelled cells from the midline the dorsal node were located in the ventral node after culture (Beddington, 1981). These findings imply that the dorsal node is not entirely committed to neural fate at E8.25 and its midline still contains progenitors that are recruited to the PNC.

The Noto+/T+/Foxa2+ PNC is ciliated, controls the Nodal flow for left-right asymmetry and forms the more anterior portion of the trunk notochord via convergent extension (Blum et al.,

2007). I observed that while the PNC has lost Oct4, low levels of Oct4 together with Noto and T were expressed in the adjacent crown cells. Previously, proliferation assays demonstrated that in contrast to the quiescent, columnar PNC, the crown is comprised by dividing, squamous cells (Blum et al., 2007; Ukita et al., 2009). Dye labelling of the crown at E8.5 shows that it gives rise in the notochord at more posterior axial levels from somite 16 and ablation of the crown leads to almost immediate axial arrest (Wymeersch et al., 2019). Thus, the crown may contain the progenitors that elongate the posterior trunk and tail notochord after the PNC is completely stretched out in a process that was described as a posterior migration of Noto+ precursors (Yamanaka et al., 2007).

Because the maintenance of axis extension depends on these cells before they form fully specified notochord, it was hypothesized that the crown forms a stable niche anchoring long term axial progenitors that are positioned dorsally to the crown in the node streak border (Wymeersch et al., 2019). My transcriptome analysis revealed that both the Noto+/Oct4+ cells and the adjacent Noto-/Oct4+ co-express Oct4, T, Sox2 and Wnt3a, which could constitute a stem cell signature (Figure 3.2). Oct4 activity in the E8.0-E8.5 NSB could therefore preserve precursor populations for more posterior fates. The arrangement of the crown and associated stem cells could represent the mouse homolog structure of the recently described stem cell niche in the posterior portion of Hensen's node in chicken (Solovieva et al., 2020).

Oct4 activity in the Chordo-Neural Hinge – continuum or re-activation?

I observed that after E8.5, Oct4 expression in all discussed domains quickly declined and I did not detect Oct4+ cells adjacent to the notochord until E9.5. From the 24 somite stage at E9.5 however, I observed few Oct4+ cells in the caudal neuroepithelium. The number of Oct4+ then rapidly increased and at around E9.75, when the tail bud is fully formed, a cluster of Oct4+/Sox2+/T+ was distributed around the CNH. This expression domain of Oct4 has never been reported before and was only detected between the 24 and 32 somite stage.
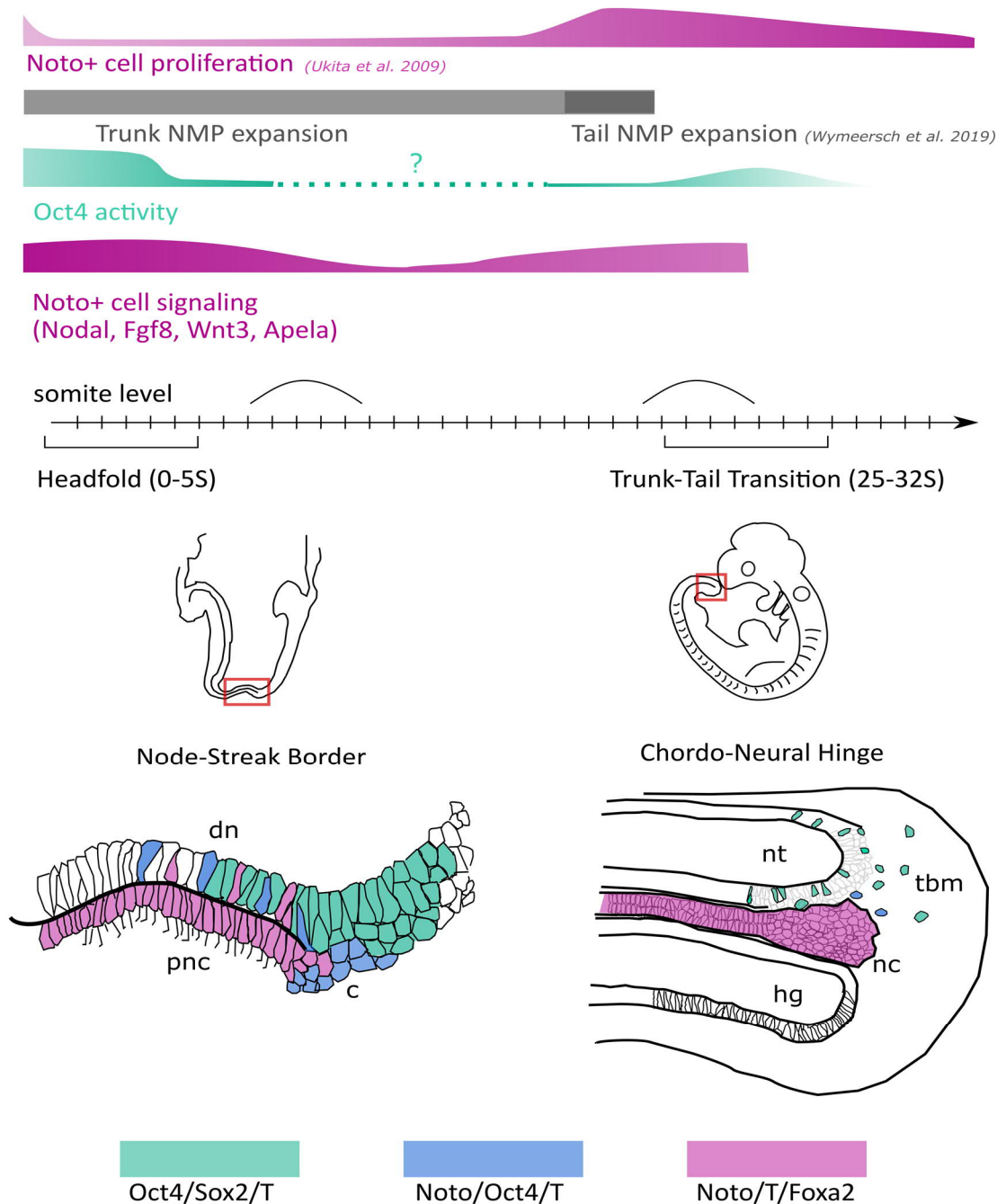
Figure 3. 2 Topological association of presumptive notochord and stem-like cells in NSB and CNH throughout axis elongation

Noto+ cell proliferation was determined by BrdU-labeling in Ukita et al 2009. NMP expansion, as the increase in cell number within the progenitor pool is discussed in Wymeersch et al 2019 on the basis of different datasets. After trunk-tail transition, NMPs in the CNH are thought to be gradually depleted until E13.5. Oct4 activity represents the amount of Oct4+ cells in the posterior growth zone (Chapter 2.4). No Oct4 expressing cells were detected in this domain between E8.5 and E9.5 and it is not clear if the population persists or Oct4 is re-activated. Noto+ signaling as analyzed in Chapter 2.3. Bottom: Close-up of conserved progenitor domains, sagittal sections. dn=dorsal node. vn=ventral node (ciliated PNC area). c=crown. nt=neural tube. tbm=tail bud mesenchyme. hg=hingut. nc=notochord

In addition to microscopy, I analyzed the transcriptome of Oct4+ cells from the CNH at these stages and found that Oct4 is co-expressed with a group of genes that were proposed as markers for "real" NMPs with stem-like properties (Figure 2.4.8). The expression pattern of the *Oct4::H2B-CFP* line, which due to the stability of the fluorescent reporter facilitates short range fate mapping shows that Oct4-expressing cells contribute first to the neural tube, then to the neural tube and somites of the tail (Frederic Koch, MPIMG, personal communication). Thus, it is possible that this Oct4+/Sox2+/T+ population represents a novel class of axial stem cell for tail development.

Both the NSB and the CNH were shown to function as niche environments and cells can be serially transplanted between these domains in successive embryos (Cambray and Wilson, 2007). Therefore, it is tempting to assume, that the Oct4+/Sox2+/T+ cells function as a persistent, self-renewing stem cell population that divides asymmetrically contributing one daughter cell per cell cycle to the primordial tissues while one cell remains in the niche and maintains Oct4 expression. This concept of a resident Oct4+ stem cell niche would imply that at any time of axial elongation, there should be at least few Oct4 expressing cells in the caudal growth zone. However, this is conflicting with my data, which locate Oct4 activity in the NSB and CNH for a relatively short period at the onset of trunk and tail development, respectively.

Two options would explain these observations. Either the Oct4+ population is continuously present and minimized to few cells during mid trunk development until the stem cell pool is expanded again in the tail bud. In that case, I would have overlooked these cells in the imaging experiments. Alternatively, Oct4 expression is re-activated in a subset of axial progenitors during trunk-to-tail transition.

While the latter possibility is not intuitive, several reasons make it worth considering. It was shown previously that in cell types which already express other Yamanaka-factors, activation of Oct4 alone is sufficient for pluripotency reprogramming (Kim et al., 2009). Since Sox2 and Myc are expressed in the CNH, the re-activation of Oct4 could induce the generation of stem cells. Further, Oct4 and Sox2 were shown to form a heterodimer, which promotes pluripotency and repress the pro-neural activity of Sox2 (Zhang and Cui, 2014). Hence, the presence of Oct4 in NMPs could inhibit differentiation and thereby amplify the NMP pool. Re-activation of the Oct4 locus was reported in somatic smooth muscle cells after injury (Cherepanova et al., 2016).

98

Very recently, it was discovered that during cranial neural crest development, neuroepithelial precursors transiently re-activate the pluripotency program and expand their potential to contribute to mesoderm (Zalc et al., 2021). Oct4 was shown to play a central role in this process (Zalc et al., 2021). Therefore, Oct4-reactivation is an available mechanism both in adult and embryonic tissues. It would be intriguing to investigate if the observed Oct4-Venus expression reflects a re-activation, which may extend the potency of the caudal neuroepithelium to contribute to mesoderm and thereby expand the tail NMP pool.

Oct4 activity as a mechanism for trunk and tail progenitor cell preservation

My data suggest that Oct4 is first present in a subset of axial progenitors at the onset of both trunk and tail development and then successively depleted. Therefore, Oct4, which has the capacity to both maintain multipotency and prime cells for differentiation (Simandi et al., 2016), might ensure the gradual release of cells from the progenitor reservoir. The duration of Oct4 expression could determine the duration of the progenitor state. This idea would be consistent with previous data showing that Oct4 overexpression in trunk progenitors delays their differentiation and results in an increased trunk length phenotype (Aires et al., 2016). It would be interesting to test whether overexpression or degradation of Oct4 at more posterior levels induces a similar effect in the tail.

In general, I found that Oct4 activity was most pronounced at the initiation of trunk and tail development, preceded by peaks in signaling activity from Noto+ cells and progenitor proliferation (Figure 3.2). It is conceivable that axial elongation, rather than being a strictly continuous process, is dependent on waves of progenitor amplification.

Trunk NMPs, lateral mesoderm and notochord progenitors are established and amplified at around headfold stage. After the occipital portion of the axis (i.e. the neck) is laid down during gastrulation (Guibentif et al., 2021), NMPs become essential for the generation of trunk tissues posterior to the forelimb bud. The notochord progenitors and NMPs for the tail are then expanded at the hindlimb bud level (Ukita et al., 2009; Wymeersch et al., 2016). The establishment and expansion of the progenitors may represent critical checkpoints of axis extension. Accordingly, mutants of key regulators in this process often induce phenotypes that

display axial truncations either shortly after the forelimb bud (e.g. *T* -/- and *Wnt3a* -/-) or after the hindlimb bud (e.g. *Noto* -/- or different *T* hypomorphs), but rarely at intermediate axial levels (Abdelkhalek et al., 2004; Herrmann et al., 1990; Pennimpede et al., 2012; Takada et al., 1994). It remains to be investigated what role pluripotency factors play in the expansion and maintenance of axial progenitors.

It would further be intriguing to study the possible linkage of Noto+ cell signaling and Oct4 activity in multiple species and to explore whether these components form a module for the evolution of vertebrate body plans.

3.3 Fate-mapping of the axial mesoderm lineage

Different parts of the notochord are generated in different morphogenetic processes. The head process notochord is given rise to via condensation of dispersed anterior cells (Yamanaka et al., 2007). The first portion of the trunk notochord is formed via convergent extension of the node, which transforms the broad PNC epithelium into the typical rod-like structure (Kinder et al., 2001; Yamanaka et al., 2007). From around mid-trunk levels, a posterior notochord progenitor population increasingly contributes to the structure (Wymeersch et al., 2019). Cell tracing experiments using dyes, as well as genetic labeling demonstrated that the original node population continues to contribute to the lineage in the tail (Ukita et al., 2009; Wymeersch et al., 2019). Therefore, the posterior trunk and tail notochord is probably elongated both by proliferation of the specified caudal notochord and by recruitment from a progenitor population. It is likely that the notochord progenitor population is established approximately at headfold stage and comprised of crown cells (Wymeersch et al., 2019). My data suggest that at this stage, notochord progenitors in the crown and dorsal node co-express Noto, T, Sox2 and Oct4, implying a higher potency compared to the other Noto+ cells (Figure 2.4.4). Imaging revealed that not all Noto+ cells are fully committed to the axial mesoderm lineage and there is a continuous contribution of Noto+ cells in the posterior growth zone to the sclerotome in the trunk (Figure 2.5.1-4).

I showed that these bi-fated cells co-express Noto and Tbx6 and are located at the posterior tip of the notochord. It is likely, that in these cells, which represent sub-sets of their lineages, the axial versus paraxial mesoderm lineage decision is mediated via mutual repression of Tbx6 and Noto. It was previously reported that *Noto* ablation in the trunk leads to increased contribution of Noto[GFP]+ cells to the somites in the mouse (Yamanaka et al., 2007) and in zebrafish, one of the main functions of the Noto orthologue *flh* is the repression of paraxial mesoderm fate (Amacher and Kimmel, 1998). Accordingly, Noto[mC]+ cell contribution to the paraxis was never observed in *Tbx6 -/-* mutants (Milena Pustet, Bachelor's thesis, MPIMG). Brachyury acts upstream of both Noto and Tbx6 (Abdelkhalek et al., 2004; Koch et al., 2017) and a higher dosage of T is required for notochord specification. Thus, the lowered T dosage in *TE2* enhancer mutants presented in this work lead to a conversion of these Noto[mC]+ notochord progenitors towards paraxial mesoderm (Figure 2.1.7).
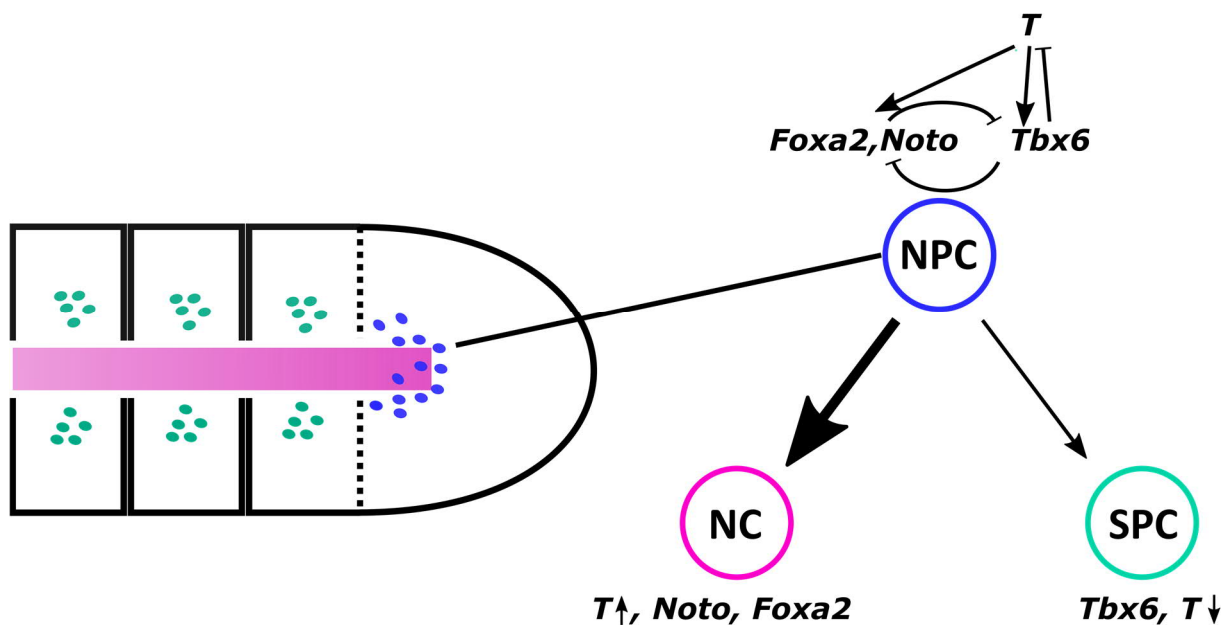
Figure 3. 3 Contribution of Notochord progenitor cells to the sclerotome

Hypothetic mechanism for axial vs. paraxial fate choice. NPC = Notochord progenitor cell. NC = notochord. SPC = sclerotome progenitor cell. T induces both Noto and Tbx6 and is repressed only by Tbx6. Most NPCs contribute to notochord, but a subset adopts paraxial fate. T expression in the paraxial mesoderm is transient.

Is there a functional role of Noto+ cells contributing to the paraxial lineage?

Expression profiling of the Noto^{mC}+/Tbx6^{Ve}+ in the caudal growth zone showed that indeed, these cells are less differentiated compared to the Noto^{mC}+/Tbx6^{Ve}-, which raises the question whether this population is a novel type of progenitor or rather reflects transcriptional plasticity during mesoderm differentiation. Strikingly, I found that the progeny of the Noto^{mC}+/Tbx6^{Ve}+ cells in the trunk sclerotome expressed secreted signals that are required for chondrogenesis, including *Shh* and *Tgfb1* (Figure 2.5.3). It is possible that these cells function as signaling centers promoting sclerotome differentiation, but more experimental data are required to substantiate this hypothesis. The transcriptome dataset acquired in this work provides a good resource for the identification of differentially expressed genes that could be used as drivers for such experiments.

In zebrafish, segmented non-notochord axial mesoderm controls the ossification of vertebrae (Wopat et al., 2018). In mice, lineage tracing of Foxa2$^{T2A-iCre}$ activity was shown to label the

102

gut, floor plate, notochord and sclerotome (Hasenöder, 2016). The cells in the sclerotome that are labelled by Foxa2[T2A-iCre] are probably identical to those labelled by Noto[iCre] in my work. A Foxa2[T2A-iCre] driven knock out of Oct4 was reported to deplete these cells entirely and induces a kinked tail phenotype amongst other defects leading to neonatal lethality (Hasenöder, 2016). A more detailed analysis of such mutants with a focus on vertebrae development could reveal a potential function for notochord progenitor derived cells in sclerotome development.

Implications for chordoma formation

Chordoma are slow-growing, malignant bone tumors occurring in the base of the skull and sacral area of the spine (Chugh et al., 2007). These neoplasms are thought to originate from embryonic notochord cells, which fail to condense and aberrantly remain in the vertebrae instead of giving rise to the nuclei pulposi of the invertebral discs (Choi et al., 2008). *Brachyury* has been identified as both a marker and driver of chordoma (Shah et al., 2017; Tirabosco et al., 2008; Vujovic et al., 2006). A SNP in the Brachyury gene and duplications of the locus have been reported to cause hereditary and sporadic chordoma (Pillay et al., 2012; Tarpey et al., 2017; Yang et al., 2009). The current concept for the genesis of chordoma implies that these mutations prevent the downregulation of Brachyury in notochord remnants within the spine during fetal development and the bone cancer develops over decades. The contribution of notochord progenitors to the sclerotome described in this chapter provides an alternative route for axial mesoderm to the vertebrae. Utilizing Noto[iCre] lineage tracing, I confirmed that these cells give rise to the neural arches of the vertebrae (Figure 2.5.4). Chordoma predominantly arise in the neck and the sacral spine, which are the axial levels where notochord progenitors proliferate during embryogenesis (Ukita et al., 2009). This may indicate that the cause of chordoma could be linked to the generation rather than to the regression of the notochord.

To this day, there is no mouse disease model for chordoma. The data presented in this work may be applied for the generation of such a model, for instance by establishing a conditional T overexpression system that is cooperatively induced by *Noto* and *Tbx6*.

The conserved *TE2* notochord enhancer of *Brachyury* identified in this work might be an interesting target for gene therapy. In addition, since the enhancer is essential for notochord maintenance in the mouse, a tandem duplication of the T locus spanning from the *TE2* enhancer to the gene might be sufficient to phenocopy chordoma development and represent an additional approach for a mouse model.

4 Methods

4.1 Molecular cloning

4.1.1 CRISPR plasmids

Vectors px335A_hCas9_D10A_G2P (a gift from Boris Greber) and px459-pSpCas9-2A-Puro (a gift from Feng Zhang; Addgene plasmid #48139) were used for the double nickase or conventional approach, respectively. Both vectors contain sequences coding for Cas9 enzyme controlled by a ubiquitous promoter, the guide RNA (gRNA) controlled by a human U6 promoter, a GFP-T2A-Puro cassette for selection in ESCs and a ampicillin resistance for selection in bacteria. The CRISPR/Cas9 system was used for the introduction of targeted genomic deletions. Close to the desired break points, specific targeting sites $5'$-$N_{20}NGG$-$3'$, with $5'$-NGG-$3'$ being the *Streptococcus aureus* Protospacer adjacent motif (PAM) were identified and evaluated using the CRISPOR (http://crispor.tefor.net/crispor.py) tool. In case the first nucleotide of the N_{20} targeting sequence was not a guanine, a guanine residue was added to the $5'$ end for improved expression from the U6 promoter. For cloning into *BpiI* overhangs, $5'$-$CACC$-$3'$ or $5'$-$AAAC$-$3'$ was added to the target sequence or complimentary strand sequence, respectively. 10µM of complimentary oligo pairs (purchased from Sigma Aldrich) were annealed in T4 ligation buffer *(Promega M1801)* by continuous cooling from 95°C to 25°C. Annealed oligos were ligated into expression vectors. The vectors were used for heat-shock transformation of *Escherichia coli* DH5-alpha, positive clones were screened for by *BpiI* restriction digest and verified by Sanger sequencing using primer U6_F.

4.1.2 BAC recombineering

BACS containing ~200 kb C57/BL6 genome surrounding the mouse *Noto* (RP23-289M19), *T* (RP24-530D23), *Tbx6* (RP23-421P23) and *Foxa2* (RP23-254G2) genes were obtained from BACPAC resources. In order to engineer reporter transgenes or Cre recombinase drivers, a construct containing a H2B-fused fluorescent marker (*mCherry*, *Venus*

or *mTurquoise2*) or *iCre* was inserted into the start codon of the respective gene via Red/ET recombineering (Muyrers et al., 1999), respectively. The reporter or recombinase construct is followed by a *FRT*-site flanked selection cassette. The selection cassette consists a *Pgk* promoter for expression in mESCs and an em7 promoter for selection in bacteria driving either a geneticin, puromycin or hygromycin resistance terminated by a *b-globin* polyadenylation signal. All reporter constructs were amplified from pR6K plasmid vectors supplied by Frederic Koch (MPIMG, Berlin). Reporter constructs were amplified from these plasmids by PCR (TAKARA PrimeStar Polymerase) using primers with 50bp sequences homologous to the desired insertion site. PCR fragments were purified by electrophoresis and gel extraction (QIAgen Gel extraction kit) and eluted in Millipore H20. Fresh overnight *E. coli* DH10-β cultures carrying unmodified BACS and the pSC101-BAD-gbaA plasmid were diluted (30µl bacterial culture in 1.4ml LB medium with appropriate antibiotic) and grown for another 2h at 30°C and 1000rpm before induction of recombinase expression by addition of 40 µl 10% arabinose to the reaction. After 1 h at 37°C bacteria were collected by centrifugation (30s at 10^4 rpm and 4°C) and washed twice with ice cold, tissue culture grade H_2O. Bacteria were resuspended on ice in ~70 µl tissue culture grade H20 containing ~300 ng purified PCR product. The mix was transferred to a 0.1 cm gap electroporation cuvette (*BioRad 1652093*) and immediately electroporated in a BioRad Gene Pulser at 200 Ω, 1800 V and 25 µF. After 1h recovery in LB at 37°C 1000 rpm, bacteria were plated on LB agar plates containing the correct antibiotic. Colonies were screened by restriction digest and PCR with flanking the integration site. PCR products were purified and Sanger sequenced with multiple primers covering the entire inserted cassette for verification.

4.1.3 Oct4-Venus targeting construct

First, a plasmid containing an IRES-Puro element (provided by Heinrich Schrewe) was modified with the ORF for the Venus fluorescence reporter cloned into an *EcoRI* site 5′ to the IRES-Venus. In order to engineer a targeting construct for homologous recombination targeting the mouse *Pou5f1 (Oct4)* locus with ~4kb homology arms, a region spanning the gene excluding promoter and first exon was subcloned into a PCR amplified pBlueskript SKII backbone from BAC RP23-152G18 via Red/ET recombineering (described in 4.1.2). The Venus-

IRES-Puro sub-fragment was isolated by *Nsi*I restriction digest and ligated into an *Nsi*I site on the *Oct4* targeting vector in a way that the endogenous stop codon is substituted by an in frame Venus-IRES-Puro tag. The correct insertion was screened for via restriction digest and verified by Sanger sequencing.

4.2 Mouse embryonic stem cell culture

4.2.1 Culture procedures

In this study, male mESCs of the G4 hybrid line (129S6/C57BL6) (Nagy et al., 1993) served as the parental wild type clone. For experiments investigating the *T/tw5* haplotype, a T-NMRI cell line established by Lars Wittler (MPIMG, Berlin) and Manuela Scholze-Wittler (MPIMG, Berlin) was used.

Culture procedures were described earlier in (Nagy, A., Gertsenstein, M., Vintersten, K., and Behringer, 2003). Compositions of growth media can be found in Table 4.4. In brief, handling of mESC cultures was performed under sterile conditions in a laminar flow cabinet. Cultures were incubated at 37°C and 7.5% CO_2 in humified incubators. Before seeding mESCs, dishes (*Corning 430166*) were gelatinized with 0.1% gelatin (diluted in tissue culture grade H_20 from 2% gelatin, *Sigma G1393*), air-dried and coated with a monolayer of mitotically inactivated fibroblast cells with a density of 3-4 x 10^4 cells per cm^2 seeded in mEF medium (tab). mESCs were seeded after overnight incubation of feeder cells in mESC medium (Table 4.4) containing 15% FCS and 1000 U/ml leukemia inhibitory factor (*Chemicon ESG1107*). Medium was refreshed daily and after reaching confluency, cells were passaged or frozen. For passage, the medium was aspirated and cells were washed twice with DBPS (*Lonza 17-512F*) before Trypsin/EDTA *(Gibco 25300054)* application for 10 minutes at 37°C. Trypsinization was quenched by adding the double volume of ES medium. Cells were resuspended by pipetting up and down and single cell suspensions were collected by centrifugation at 1000 rpm for 5 min. For freezing, the cell pellet was resuspended in ES medium containing 20% FBS and an equal volume of 2x freezing medium (ES containing 20% FBS and 20% DMSO). 2 ml vials containing mESCs in freezing medium were put at -80°C in isopropanol filled freezing containers (*Nalgene C1562*) and stored at ~-180°C in nitrogen tanks.

For determining cell numbers, 10 µl aliquots of single cell suspensions were counted using LUNA Automated Cell Counter (Logos Biosystems) according to the manufacturer's protocol.

In order to test cells for possible mycoplasma contamination, cell residues from collection tube and culture dish were re-plated on the culture dish with ES medium.

After ~5 days without medium refreshment, 100 µl samples from the supernatant taken for Mycoplasma testing using PCR Mycoplasma Test Kit II (*Applichem A8994*) according to the manufacturer.

4.2.2 Generation of transgenic mESC lines

4.2.2.1 *Lipofectamine transfection*

Transfection of plasmids for CRISPR/Cas9 mediated deletions was performed using Lipofectamine 2000 reagent (*Invitrogen 11668027*). On the day before transfection, $3x10^4$ cells or $1.5x10^4$ cells per well were seeded on gelatinized and wild-type feeder coated 6-well (Corning 3516) or 12-well plates (Corning 3513) respectively. After overnight incubation, transfection mixes were prepared. Plasmids used for mESC transfections were isolated and purified using Qiagen plasmid MIDI prep kit (*QIAgen 12945*), eluted in 1xTris EDTA buffer and stored at -20°C in concentrations between 0.5µg/µl to 5µg/µl. For transfections on 6-well, mixes of 125µl Opti-MEM Reduced Serum Medium (*Thermo Fisher 31985062*) and 8 µg of each vector and 110 µl Opti-MEM and 25µl Lipofectamine 2000 were prepared. 125µl of each mix were combined, mixed well and incubated at room temperature for 15 min. Subsequently, 250 µl of transfection complex mix was diluted in 1.25 ml ES + LIF, added to the cells on 6 well and incubated for 5 h. Finally, cells were trypsinized, split in 3:6, 2:6 and 1:6 ratios and seeded on 6 cm dishes coated with pyromycin resistant mEFs. 24 h post transfection, transient selection was started applying 3 ml ES+LIF containing 2µg/ml puromycin (*Gibco 10130127*) for 2 days and 3 ml ES+LIF containing 1µg/ml puromycin *(Gibco 10130127)* for 1 day. After selection, ES + LIF medium was refreshed daily until colonies were clearly visible. For stable selection after RMCE (Enhancer activity assay in Figures 2.1.6, 2.2.1; Lineage tracing in 2.5.4), cells were cultured in ES+LIF containing 350 µg/ml geneticin (*Thermo Fisher 10131027*).

4.2.2.2   Electroporation

For insertions of large linearized DNA fragments such as BACs or targeting vectors for homologous recombination, mESCs were transfected by electroporation. BACs were isolated and purified using NucleoBond BAC100 kit (*Macherey-Nagel 740579*). 5µg BAC DNA was linearized in overnight reactions with 150U *PI-SceI* (*New England Biolabs R0696S*) in 1X NEBuffer PI-SceI containing 1% BSA (*Promega R396A*) at 37°C. The next day, linearization was confirmed by gel electrophoresis on 0.7% agarose gels. The enzyme was inactivated at 65°C for 20 min and the reaction mix was directly used for electroporation.

Cells were seeded and grown to confluence for 2-3 days. For electroporation, cells were trypsinized, counted and the appropriate amount was collected by centrifugation at 1000rpm for 5 minutes (min). $3x10^6$ cells per 5µg of BAC DNA and $10x10^6$ cells per 25µg of targeting vector were used for BAC integration and homologous recombination, respectively. Cell pellets were suspended in 800µl of DPBS (*Lonza 17-512F*), mixed with the DNA and transferred to 0.4cm gap electroporation cuvettes (*BioRad 1652088*). Cells and DNA were electroporated in a BioRad GenePulser at 240V (0.240kV) and 500µF. Electroporation mixes were diluted with 10ml ES medium and collected by centrifugation at 1000rpm for 5min. $1x10^6$ cells were seeded on 6cm plates coated with feeders carrying the appropriate antibiotic resistances in ES+LIF medium. Approximately 30h after electroporation, selection was started applying 350µg/ml geniticin (*Thermo Fisher 10131027*), 150µg/ml hygromycin B (*Merck 10843555001*) or 1µg/ml puromycin (*Gibco 1013012*). Selection medium was refreshed daily until colonies were clearly visible.

4.2.2.3   Picking clones

After approximately one week of selection, cells were washed twice in DPBS and single colonies in DPBS were picked using a 10µl pipette set to 4µl, transferred to a round bottom plate (*Costar 3799*) with 30µl Trypsin/EDTA (*Gibco 25300054*) and incubated at 37°C for 10 mins. Single clones were vigorously resuspended with 60µl ES LIF and seeded on gelatinized, mEF coated 96-well (*Costar 3596*) or 48-well (*Costar 3548*) in a total volume of 150µl or 0.75

ml ES+LIF medium, respectively. Cells were expanded with media being refreshed daily until confluency. Cells on 96-well were trypsinized with 50µl Trypsin/EDTA per well. After 10 mins at 37°C the reaction was stopped adding 100µl bicarb-free DMEM/Hepes/20% FBS. Cells were split 1/2 and one half was transferred to an equal volume of 60% bicarb-free DMEM/10 mM Hepes, 20% FBS, 20% DMSO 2x freezing medium in a round bottom plate (*Costar 3799*). Plates were sealed with parafilm, wrapped in tissue, transferred to Styrofoam boxes and frozen at -80°C. 100µl ES medium was added to the remaining cells, which were grown to confluency for genomic DNA extraction. Cells on 48-well were trypsinized with 100µl Trypsin/EDTA. After trypsinization, 400µl ES containing 20% FBS was added. 4/5 were transferred to 2ml freezing vials and the same volume of 2x freezing medium and frozen in single vials as described in 4.2.1. 0.5ml ES medium was added to the remaining 100µl. Cultures for genomic DNA isolation and genotyping were incubated for another couple of days.

4.2.2.4 Genotyping

In order to isolate genomic DNA from 48-well plate cultures or embryonic tissues, cells were washed with DPBS and transferred to 1.5 ml tubes and incubated overnight in 250µl Laird's buffer (200 mM NaCl, 100 mM Tris-HCl (pH=8.3), 5 mM EDTA, 0.2 % SDS) containing 1 µg/ml Proteinase K (*Sigma Aldrich 3115887001*) at 56°C. DNA was precipitated with 250µl isopropanol and pelleted by centrifugation (13000 rpm for 15 mins at 15°C), washed with 70% EtOH and eluted in 1xTE buffer.

Cells on 96-well were washed twice with DPBS, 50µl ES lysis buffer (10mM Tris- HCl pH 7.5, 10mM EDTA pH 8.0, 10mM NaCl, 0.5% sarcosyl) containing 1µg/ml Proteinase K was added per well. Cells were incubated overnight at 56°C in a humidified box. Genomic DNA was precipitated adding 100µl ice cold 100% EtOH/75mM NaCl. After ~30 min at room temperature, DNA was washed 3x with 70°EtOH, dried and eluted in 35µl 1xTE buffer.

Positive clones were screened for by PCR using Promega GoTaq Polymerase according to manufacturer's protocol. For CRISPR/Cas9 mediated modifications, PCR was performed with primers binding ~200-300 bp from the target site. Positive clones were verified by Sanger sequencing of purified PCR products extracted from agarose gels after electrophoresis. In case

single PCR bands could not be separated by electrophoresis, fragments were cloned into pCR2.1 vectors using the reagents of TA cloning kit (*Invitrogen K202020*) according to the manufacturer's procedure. Single bacterial colonies were picked, plasmids extracted and analyzed by Sanger sequencing using primer M19_rev.

4.3   Generation of transgenic embryos

Approximately 6×10^5 ES cells were seeded on mEF coated 6cm plates in a ratio of 7:3 and incubated at 37°C and 7,5% CO_2 for 48 h, refreshing the medium every 24h. 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 in (Eakin and Hadjantonakis, 2006). All animal experiments were performed according to local animal welfare laws and approved by local authorities (covered by LaGeSo license G0243/18 and G0247/13).

4.4   Embryo isolation

Timed pregnant foster mice were euthanized by carbon dioxide application and cervical dislocation. Embryos were isolated from uteri in 4°C pre-cooled PBS. After transfer to glass vials (*Wheaton 224882*), embryos were fixed in 4% paraformaldehyde (PFA)/PBS (*Sigma Aldrich P6148*). Fixation times were adapted to embryonic stage and subsequent procedures. For immunofluorescence, E6.5-E8.5 embryos were fixed for 40min, E9.5-E10.5 for 1h and E11.5 to E12.5 for 2h. After fixation, embryos were washed 3x with PBS and stored at 4°C until further procedures.

4.5   Whole mount immunofluorescence and tissue clearing

If not specified otherwise, incubation in buffers was performed at room temperature on a roller. Embryos selected for immunofluorescence were collected in 4 ml glass vials (*Wheaton 224882*) and washed 3x10 min with PBS and 3x 10min at RT with PBST (PBS containing 0.5%

Triton X100 (*Merck 9002-93-1*). For blocking, embryos were incubated in PBSTB (PBST containing 10% FBS) at 4°C for a minimum of 24h. Primary antibody incubation was performed in PBSTB at 4°C for 48h – 96h (Used antibodies listed in Table 4.1). After incubation, remaining antibody solution was diluted rinsing the samples 3x with PBSTB followed by washing 3x 10 min with PBSTB and 3 x 10 min in PBST. After washing, the specimen were incubated in PBSTB at 4°C overnight. Secondary antibody application was performed in PBSTB at 4°C for 24h – 48h. Embryos were rinsed 3 x PBSTB and washed 2x 20 min with PBSTB + 0.02% DAPI (*Roche Diagnostics 102362760019*), 3x20min PBST + 0.02% DAPI and transferred to 8 well glass bottom slides (*Ibidi 80827*). After additional washing steps in PBS for 3 x 10 min embryos were either imaged or processed for tissue clearing.

For tissue clearing, stained embryos on 8 well glass slides were incubated in 0.02M phosphate buffer (PB, 0.005M NaH_2PO_4 and 0.015M Na_2HPO_4, pH 7.4) at room temperature for 3x 5min. Before clearing, fresh refractive index matching solution (RIMS, 133% Histodenz (*Sigma-Aldrich D2158*) in 0.02M PB was prepared and applied to the samples after careful removal of PB. Clearing was performed at 4°C on a shaking incubator for at least 24h.

4.6   Whole mount β Galactosidase staining

For lineage tracing of Noto+ cells, a *Noto::iCre* induced lacZ reporter construct in the Rosa26 Locus was used. Embryos were isolated between E8.5 and E12.5 and dissected in cold PBS. All staining procedures were performed in the dark. Embryos were fixed using 4% PFA / PBS (*Sigma Aldrich P6148*) in 4 ml glass vials (*Wheaton 224882*) for 30 min at 4°C and subsequently washed 3x for 15 minutes in Rinse Buffer (50mM EGTA, 0.1% deoxycholate, 0.2% NP-40, 20mM $MgCl_2$ in DPBS) at room temperature. After rinsing, embryos were incubated in staining solution (50mM $K_3Fe(CN)_6$, 50 mM $K_4Fe(CN)_6$, 50 mM EGTA, 0.1% deoxycholate (100x), 0.2% NP-40 (100x), 0,2M $MgCl_2$, 1 mg/mL X-gal in DPBS) at 37°C overnight. Stained embryos were washed 3x with PBS and stored in 4%PFA/PBS at 4°C for secondary fixation.

4.7 Histology

For a more precise localization of β-Galactosidase positive cells, specimen were sectioned. To this end, fixed embryos were successively sunk in 5%, 15% and 30% sucrose/DPBS at 4°C for 1 h each. After overnight incubation in 30% BSA, 20% sucrose, 0.5 % gelatine/DPBS at 4°C, samples were embedded. For embedding, 2.5 % glutaraldehyde (*Merck 111-30-8*) was added for gelatine polymerization. To allow for appropriate positioning of the embryo, the procedure was performed on ice to slow down the polymerization. Sections of 8-30 µm were prepared using the Microm HM650V (*Thermo Fisher*), transferred to Superfrost adhesion slides (*Thermo Scientific, 631-9483*) and mounted with HYDRO-MATRIX (*Micro-Tech-Lab*).

4.8 Microscopy

Embryos were imaged using a Zeiss Axiozoom V16 (wide field, HXPIllumination) system, Zeiss LSM880 laser scanning microscope with Airyscan detector or Zeiss Light sheet LS Z1 with appropriate filters for mCherry, Venus, DAPI, Alexa488. Alexa567, Alexa647. For Light sheet microscopy with, specimen were cleared as described in (4.5) and embedded in 1.5% low melting agarose (*Sigma-Aldrich A9414*)/PBS. Agarose coloumns containing the samples were inserted into the RIMS filled acquisition chamber and cleared for an additional 5 h to overnight depending on tissue volume. Post acquisition processing was performed using ZEN Blue/Black (*Zeiss*) software or Arivis Vision 4D (*Arivis*).

4.9 Fluorescence activated cell sorting

For FACS of cell cultures, cells were washed 2x PBS and dissociated by trypsinization at 37°C for 10min. Trypsin/EDTA was quenched using a double volume of 5%BSA/PBS (*Sigma Aldrich A8412*), resuspended and kept on ice until further procedure.

For FACS of embryonic material, 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. Single cell suspensions were prepared adding 100µl Trypsin/EDTA to the sample. After incubation at room temperature for 5 min, trypsin was quenched by adding 200µl PBS 5% BSA.

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

Flow cytometry data were analyzed using FlowJoV10 software.

4.10 *In vitro* differentiation of Noto+ cells

In vitro generation of notochord cells was performed following a modification of the 2011 protocol by Winzi and colleagues (Jesse Veenliet, personal communication). Embryonic stem cells were seeded on 6cm plates and passaged two times until confluence. Cells were trypsinized and resuspended in 2ml ES+LIF. mEFs were depleted from single cell suspensions by sequential plating on 0.1% gelatin (*Sigma G1393*) coated 6 well plates (*Corning 3335*) in 25min, 20min and 15min intervals. After feeder freeing, cells were resuspended in 1ml Noto-Diff medium (table), counted and seeded on 0.1% gelatin (*Sigma Aldrich G1393*) coated Nunclon Delta Surface 12-well plates (*Thermo Scientific 150628*) in a density of 5000 cells per well and ml Step 1 differentiation medium (Table 4.4).

During the 7d differentiation protocol, medium was refreshed every 24h. After 72h in Step I medium containing 1ng/ml Activin A (*R&D Systems 338-AC*), Step II differentiation medium containing 1ng/ml Activin A (*R&D Systems 338-AC*), 100ng/ml FGF2, 50ng/ml Noggin (*Peprotech 250-38*), 1µM AGN (*Santa Cruz 193109*) and 0.5µM Smoothened Agonist (*Merck 364590-63-6*) was applied for another 96h. At D7, about 10-20% of cells in the population were Noto::mCherry positive and bulk cultures were used for ChIP.

4.11 Chromatin Immunoprecipitation

For the identification of putative notochord enhancers, chromatin immunoprecipitation (ChIP) for T and Foxa2 was performed on Noto-differentiated cells at D7, following a previously published protocol (Koch et al., 2011). Formaldehyde crosslinking was performed by addition of 1/10 volume of crosslinking solution (11% formaldehyde, 100mM NaCl, 1mM EDTA pH8, 0.5mM EGTA pH8, 50mM Hepes pH7.8) to each well and incubation for 10 min on a shaker at room temperature. The reaction was quenched applying 1/10 total volume of 2.5M Glycine for 5 min at room temperature. Cells were washed twice with cold DPBS and kept on ice for further procedures. After addition of 0.005% Triton/DPBS, cells were scraped from 12 well plates, pooled in 50 ml Falcon tubes and collected by centrifugation at 1500 rpm for 5 min at 4°C. The supernatant was carefully discarded and the cell pellet resuspended in PBS containing 0.005% triton and transferred to a 1.5 ml low binding tube (*Thermo Fisher 90410*). The samples were centrifuged again at 3000 rpm for 5 min at 4°C and after removal of the supernatant, pellets of approximately 1x 10^6 cells were snap frozen on dry ice and stored at -80°C.

Cell lysis was performed in lysis buffers (LB) containing an EDTA-free protease inhibitor mix (*Roche 4693132001*). Pellets were resuspended in 2.5ml LB1 (50mM Hepes pH7.5, 140mM NaCl, 1mM EDTA pH8, 10% glycerol, 0.75% NP-40, 0.25% triton X-100), split into 2 1.5 ml low binding tubes (*Thermo Fisher 90410*) and incubated at 4°C for 20 min on a rotator. After the initial lysis, in order to get more homogenous suspensions, cells/nuclei were dounced 3x 20 times on ice with a tight pestle. Samples were pelleted by centrifugation (3000 rpm, 10 min at 4°C), resuspended in 2.5 ml LB2 (200mM NaCl, 1mM EDTA pH8, 0.5 mM EGTA pH 8, 10mM Tris pH 8) split into two 1.5 ml low binding tubes and incubated for 10 min at 4°C on a rotating wheel. For sonication, cells were centrifuged (3000 rpm, 5 min at 4°C) and resuspended in a total of 1.5ml LB3 (1 mM EDTA pH8, 0.5 mM EGTA, 10 mM Tris pH8, 100mM NaCl, 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine), which was combined in a 15 ml Falcon tube. Chromatin was sonicated in a W-450D Digital Sonifier (Branson) set to 14 cycles of 10 s on / 50 s off at 4°C with sample tubes kept in ice water. After sonication, 1/10 volume of 10% Triton was added to each sample and lysates were split into two 1.5 ml low binding tubes. Debris was pelleted by centrifugation (15000 rpm, 20 min at 4°C). Whole cell extracts of each sample were pooled and aliquoted for ChIP, keeping 50 µl per sample for the input control.

Crosslinking was reversed by addition of an equal volume of 2x elution buffer (100 mM Tris pH8, 20 mM EDTA pH8, 2% SDS) to the input followed by incubation at 65°C for 13-15h. After crosslink removal, SDS was diluted by adding an equal volume of 1x TE buffer, followed by a RNA digest with 0.2 µg/ml RNase A (*Thermo Scientific EN0531*) for 2 h at 37°C. Protein digestion was performed using 0.2 µg/ml Proteinase K (*Sigma Aldrich 71049*) for an additional 2 h at 55°C. DNA fragments were purified using Qiagen's MinElute column purification kit (*Qiagen 28004*) according to manufacturer's recommendations with an additional PE washing step and air-drying for 10 minutes before elution with 2x 15µl EB. Concentration of the input control sample was measured using an Implen P300 nanophotometer and 600 ng of sheared DNA was evaluated by gel electrophoresis on a 1.5% agarose gel. Samples with a bulk shearing size of 100-500 bp qualified as ChIP input.

For each sample, 25 µl of Dynabeads (*Invitrogen 14311D*) were washed 3x with 1 ml blocking buffer (0.5% BSA/DPBS) in low adhesion 1.5 ml tubes and resuspended in 250 µl blocking buffer. 2.5 µg of Foxa2 (*Diagenode 2683-6041*) or T antibody (*R&D AF2085*) were added to the bead solution and incubated overnight at 4°C on a rotating wheel. After antibody coupling, beads were washed 3x with 1 ml blocking buffer and resuspended in 100 µl blocking buffer. For each ChIP reaction, 2x 750 µl of chromatin lysate from approximately $5x10^6$ Noto differentiated cells (D7) was added to the antibody-coupled beads and incubated at 4°C overnight under rotation. The following day, beads were washed 8x for the Foxa2 ChIP and 9x for the T ChIP with 1 ml RIPA buffer (50mM Hepes pH7.6, 500mM LiCl, 1mM EDTA pH8, 1% NP-40, 0.7% Na-Deoxycholate) containing cOmplete protease inhibitor cocktail (*Roche 4693132001*). Finally, beads were washed 1x with 1 ml TE 50 Mm $NaCl_2$ before 110 µl elution buffer was added (50 mM Tris pH 8, 10 mM EDTA pH 8, 1% SDS). For elution, samples were incubated at 65°C under agitation for 10 min and vortexed every 2 min. Beads were removed using a magnetic device and 100 µl of the eluate were incubated at 65°C for 13-15 h under agitation for crosslink reversal. The next day, 200 µl TE buffer was added to each reaction and ChIP DNA was purified as described for the input control. Samples were quantified using the Qubit fluorometric HS DNA assay (*Thermo Fisher 12102*).

4.12 Preparation of Next Generation Sequencing libraries

4.12.1 RNAseq library preparation

For transcriptome analysis of FACS purified Noto+ populations, total RNA was isolated from 250 (or less) cells 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 350µl RW1. Therefore, reaction mixes of 10 µl DNase I (*Qiagen 79254*) and 1µl (=10U) DNase I (*Roche 4716728001*) in 70µl buffer RDD (*Qiagen 1011132*) to were applied to the spin columns for 15 min incubation at room temperature. Membranes were air dried for 10 min to remove ethanol remains and successively eluted in 2x 20 µl EB.

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

cDNA library pools (100nmol in 10µl) with 16bp barcode length (8bp barcode + 8bp UMI) were sequenced running on half a lane of the HighSeq4000 (Illumina) with 2x75bp read lengths.

4.12.2 ChIP-Seq library preparation

ChIP-Seq sequencing libraries were generated using the TrueSeq ChIP-Seq kit (*Ilumina*) following the manufacturer's intructions with minor modifications: After adapter ligation, 0.95x AmPure XP beads (*Beckman Coulter A63881*), corresponding to 41.1 µl per reaction, was used for purification. The purified DNA was eluted in 15µl Resuspension buffer and 14 µl of the clear supernatant were directly used for the first amplification. Pre-amplification reactions with 14 µl eluate, 1µl primer mix (25 µM each, Forward: 5′-AATGATACGGCGACCACCGA∗G-3′;

Reverse: 5'- CAAGCAGAAGACGGCATACGA*G-3') and 15 µl Kapa Hifi Hotstart ReadyMix (*Roche 7958935001*) were assembled and performed for 5 cycles.

98°C	*45s*	
98°C	*15s*	
63°C	*30s*	5x (Pre-Amplification)
72°C	*30s*	13 x (Amplification)
72°C	*1min*	
4°C	*∞*	

Pre-amplified DNA was purified with 0.95x (=28.5µl) AmPure XP beads and eluted in 21 µl Resuspension buffer. 20 µl of purified reaction products were mixed with 6µl 5x loading dye, loaded on an unstained 1.5% agarose gel and separated by gel electrophoresis at 120 V for 10 min and 60 V for 120 min. After running, the gel was stained with Sybr Gold (*Thermo Fisher S11494*) and fragments of 250-400bp were excised. Subsequent to MinElute column purification, DNA was eluted in 21 µl EB. 19µl of each sample were used for the main amplification with 1µl primer mix and 20 µl Kapa Hifi Hotstart ReadyMix. Samples were amplified using the same program (table) for 13 rounds. Finally, 40 µl of AmPure XP beads were used for the last purification. Concentrations of ChIP-Seq libraries were measured with the Qubit HS DNA assay (*Thermo Fisher 12102*). Library size was evaluated using DNA High Sensitivity Bioanalyzer chips (*Agilent 5067-4626*).

Library pools (100nmol in 10 µl) with 16bp barcode length (8bp barcode + 8bp UMI) were sequenced on the HighSeq4000 (Illumina) using 2x75bp read lengths.

4.13 Bioinformatics

4.13.1 RNA Seq

Prior to mapping, the first 5 nucleotides of the forward and reverse reads were trimmed using fastx_trimmer (http://hannonlab.cshl.edu/fastx_toolkit/index.html) according to manufacturer's instructions (NuGEN). The resulting reads were mapped to chromosomes 1-19, X, Y and M of the mouse mm10 genome using TopHat2 (v2.1.0) 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 (http://nugentechnologies.github.io/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 1.2.2 (Trapnell et al., 2013) with options *-u --no-effective-length-correction -b*. Based on FPKM values, generation of heatmaps, hierarchical and k-means clustering was performed in MeV (Howe et al., 2011). Barplots were generated using ggplot2 (https://ggplot2.tidyverse.org/) in RStudio Version 1.2.5033 (https://rstudio.com/products/rstudio/).

4.13.2 ChIP Seq

Reads were mapped to chromosomes 1-19, X, Y and M of the mouse mm10 genome using bowtie version 1.1.2 (Langmead et al. 2009), providing the options ' -y -m 1 -S -p 15 -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 pearl 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).

4.13.3 Image analysis

The mean fluorescence intensity (MFI) of embryos stained for T and Sox2 (2.2.3) was measured using ZEN3.2 (blue edition) software. Single cells were detected via automated segmentation based on DAPI signal with the following settings: *Gauss=Smooting; Sigma=1.2; Background Subtraction=Rolling Ball; Radius=20; Tolerance=10%; Neighborhood=13; Minimum Area=100; Minimum Hole Area=20; Separate=Watersheds; Count=5*. Regions were filtered with options *Circularity=0.5-1.0; Minimum=30M; Maximum=200M*. After analysis, cell type annotations were added to a sub-set of 3 central Z planes manually according to their position. In order to account for working distance and tissue depth variation, each MFI value was normalized to the DAPI MFI of the same cell.

Scatter plots and bar plots of these datasets were generated using ggplot2 (https://ggplot2.tidyverse.org/) in RStudio Version 1.2.5033 (https://rstudio.com/products/rstudio/).

4.14 List of antibodies

Table 4-1 Antibodies used for Immunofluorescence and ChIP

Description	Catalog number	Company	Host Organism	Concentration	Application
α-T	#81694	Cell Signal	Rabbit	1:250 in PBSTB	IF
α-T	AF2085	R&D	Goat	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-GFP	ab13970	Abcam	Chicken	1:250 in PBSTB	IF
α-Caspase-3	9662S	Cell signal	Rabbit	1:250 in PBSTB	IF
Alexa Fluor 488 **α**-Chicken	703-545-155	Jackson	Donkey	1:250 in PBSTB	IF
Alexa Fluor 488 **α**-Rabbit	ab150073	Abcam	Donkey	1:250 in PBSTB	IF
Alexa Fluor 568 **α**-Rabbit	ab175470	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	2.5µg	ChIP
α-Foxa2	2683-6041	Diagenode	Rabbit	2.5µg	ChIP

4.15 List of oligos

Table 4-2 List of oligonucleotides used for cloning and PCR

Name	5'-3' sequence	Purpose	Results section
Noto_rec_H2B_mcherry_fw	CTCCCATTGAGCTCCTTGCACAGCCTGGGAGGTCCCCTCAGGGTCGCGCAATGCCAGAGCCAGCGAAGTC	BAC reporter recombineering PCR	2.1
Noto_rec_H2B_mcherry_rv	GGGCGCAGGCTCCCGGGCTGGACCTGAGTGCCTGAGGGAGCAGGGCTGGATTCCCAGTCACGACGTTGTA	BAC reporter recombineering PCR	2.1
Noto_PCR_fw	GGCCTCAATCAGCGATGATTAAG	BAC reporter recombineering PCR	2.1
Noto_PCR_rv	CTGGACCTGAGTGCCTGAG	BAC reporter recombineering PCR	2.1
PciI-Kozak-Venus_F	tttttACATGTccgccaccATGGTGAGCAAGGGCGAG	Enhancer reporter cloning	2.1
PciI-rbpA_R	tttttACATGTACGCGTGCAGTCGAGTTCATAAGAGAAGAGGGACAGCTATGACTGGGAGTAGTCAGGAGAGGAGGAAAAATCTGGCTAGTAAAAgATGTAAGGAAAATTTTAGGGATGT	Enhancer reporter cloning	2.1
Seq_U6_F	ACTATCATATGCTTACCGTAAC	Sequencing of gRNA vectors	2.1
T-dup-mm-i_top	caccgATCAGTATTGGATCCTCGTT	T CRISPR cloning	2.1
T-dup-mm-i_bot	aaacAACGAGGATCCAATACTGATc	T CRISPR cloning	2.1
T2_upstream_C1_top	caccgATGCACGCTCTTAATCTCGG	T CRISPR cloning	2.1
T2_upstream_C1_bot	aaacCCGAGATTAAGAGCGTGCATc	T CRISPR cloning	2.1
T-dup-u_F1	CCTTTCCCCCTTGATCAC	T CRISPR genotyping PCR	2.1
T-dup-d_R1	ATGAGAGTGCCTGAGGAG	T CRISPR genotyping PCR	2.1
T-8000_R	GTCTGTCCCTGAGATGATG	T CRISPR genotyping PCR	2.1
T-8000_top	caccGCAGCGTAGAGATAGCGGCT	Tud CRISPR cloning	2.1
T-8000_bot	aaacAGCCGCTATCTCTACGCTGC	Tud CRISPR cloning	2.1
T2_upstream_F	TATTGGGATGCTGTGCTC	T CRISPR genotyping PCR	2.1
T2_201_E2_g1_top	caccgTGCAGTCGCAGCCCGAAGCT	T2 Indel CRISPR cloning	2.1
T2_201_E2_g1_bot	aaacAGCTTCGGGCTGCGACTGCAc	T2 Indel CRISPR cloning	2.1
TE1_sg1_top	CACCGGCTTCCAACTCCAAGGTAA	TE1 CRISPR cloning	2.1
TE1_sg1_bot	AAACTTACCTTGGAGTTGGAAGCC	TE1 CRISPR cloning	2.1
TE1_sg2_top	CACCGAAGAGATCACAGCATCCGA	TE1 CRISPR cloning	2.1
TE1_sg2_bot	AAACTCGGATGCTGTGATCTCTTC	TE1 CRISPR cloning	2.1
TE1_sg3_top	CACCGCTCAGGAGTACACGTCTTAC	TE1 CRISPR cloning	2.1
TE1_sg3_bot	AAACGTAAGACGTGTACTCCTGAGC	TE1 CRISPR cloning	2.1
TE1_sg4_top	CACCGAACACGCTGAGCATCCCAA	TE1 CRISPR cloning	2.1
TE2_sg1_top	CACCGGGGGACCTGACACGTCTCC	TE2 CRISPR cloning	2.1
TE2_sg1_bot	AAACGGAGACGTGTCAGGTCCCCC	TE2 CRISPR cloning	2.1
TE2_sg2_top	CACCGctggactcacagcgagttct	TE2 CRISPR cloning	2.1
TE2_sg2_bot	AAACagaactcgctgtgagtccagC	TE2 CRISPR cloning	2.1
TE2_sg3_top	CACCGACCCGTCACATCGAGCACCA	TE2 CRISPR cloning	2.1
TE2_sg3_bot	AAACTGGTGCTCGATGTGACGGGTC	TE2 CRISPR cloning	2.1
TE2_sg4_top	CACCGTGTGATTTGGAGATTCGGGT	TE2 CRISPR cloning	2.1
TE2_sg4_bot	AAACACCCGAATCTCCAAATCACAC	TE2 CRISPR cloning	2.1
TE4_sg1_top	CACCGCCATCTCCCATTTCCGGAAG	TE4 CRISPR cloning	2.1
TE4_sg1_bot	AAACCTTCCGGAAATGGGAGATGGC	TE4 CRISPR cloning	2.1
TE4_sg2_top	CACCGTGAGGCTGGCTTCGGCCTCT	TE4 CRISPR cloning	2.1
TE4_sg2_bot	AAACAGAGGCCGAAGCCAGCCTCAC	TE4 CRISPR cloning	2.1
TE4_fw	GTGGCCTGTCAGCTTTGTC	TE4 Genotyping PCR	2.1

TE5_rev	GCTGGCCTTACACTCATGGT	TBS-T-5 Genotyping PCR	2.1
TE5_sg3_top	CACCGCCCAAGGTTTTCTGCATGCT	TE5 CRISPR cloning	2.1
TE5_sg3_bot	AAACAGCATGCAGAAAACCTTGGGC	TE5 CRISPR cloning	2.1
TE5_sg4_top	CACCGtatgaacagggcatgatggc	TE5 CRISPR cloning	2.1
TE5_sg4_bot	AAACgccatcatgccctgttcataC	TE5 CRISPR cloning	2.1
TE7_sg1_top	CACCGCACTGCACTGCCAGACCCAC	TE5 CRISPR cloning	2.1
TE7_sg1_bot	AAACGTGGGTCTGGCAGTGCAGTGC	TE5 CRISPR cloning	2.1
TE7_sg2_top	CACCGCCCTGGCTTCTTGCCCCGTC	TE5 CRISPR cloning	2.1
TE7_sg2_bot	AAACGACGGGGCAAGAAGCCAGGGC	TE5 CRISPR cloning	2.1
TE7_sg3_top	CACCGCTGCATCCAAGACTGCTTCC	TE5 CRISPR cloning	2.1
TE7_sg3_bot	AAACGGAAGCAGTCTTGGATGCAGC	TE5 CRISPR cloning	2.1
TE7_sg4_top	CACCGTCTGGCTGGATAATTAAGCA	TE5 CRISPR cloning	2.1
TE7_sg4_bot	AAACTGCTTAATTATCCAGCCAGAC	TE5 CRISPR cloning	2.1
TE7_fw	CTGCTGTGGTGAGGAAGTGT	TE5 CRISPR cloning	2.1
TE7_rev	GCTCGTTTGCAGCCTCTG	TE5 CRISPR cloning	2.1
TBST2-tw5_top	CACCGCACACACATCTGCAATCCACG	T/tw5 CRISPR repair	2.1
TBST2-tw5_bot	AAACCGTGGATTGCAGATGTGTGTGC	T/tw5 CRISPR repair	2.1
TE3_sg3_top	CACCGCCTGTACATGGATGTCCCTA	TE3 CRISPR cloning	2.2
TE3_sg3_bot	AAACTAGGGACATCCATGTACAGGC	TE3 CRISPR cloning	2.2
TE3_sg4_top	CACCGTTAAACTTTGACAGTACTAC	TE3 CRISPR cloning	2.2
TE3_sg4_bot	AAACGTAGTACTGTCAAAGTTTAAC	TE3 CRISPR cloning	2.2
TE3_fw	CGCGTCCTCAGCCTTTAC	TE3 Genotyping PCR	2.2
TE3_rev	GGTTATCAGCCCTCCTCCTG	TE3 Genotyping PCR	2.2
pBS-oct4-tag_5'	TTCCCAATTCCCTTCACTGCTGCCTCTGCCAGCCCAGCCAAGT CCCTTCATAATACGACTCACTATAGGG	Oct4 targeting vector recombineering PCR	2.4
pBS-oct4-tag_3'	CATCTCTTTGTTCCCTGTCTCCCAGATTGCTTTGTATAGTCTA GTCCTAAGCAGCCTGAATGGCGAATG	Oct4 targeting vector recombineering PCR	2.4
Puro_3'_fw	ACCTCCCCTTCTACGAGC	Oct4-Venus Genotyping PCR	2.4
Oct4_3'_rv	GCAGACACCCAGGTTAGC	Oct4-Venus Genotyping PCR	2.4
Oct4-Venus-5'probe_fw	GCCCCTTTGAACCTGAAG	Oct4-Venus Southern blotting probe PCR	2.4
Oct4-Venus-5'probe_rv	GAAGTCTGAAGCCAGGTG	Oct4-Venus Southern blotting probe PCR	2.4
Oct4-Venus-3'probe_fw	CTAACCTGGGTGTCTGCT	Oct4-Venus Southern blotting probe PCR	2.4
Oct4-Venus-3'probe_rv	GTTCCTGTTCTCGTCACC	Oct4-Venus Southern blotting probe PCR	2.4
Venus-NsiI-5'_fw	tttttATGCATTCAAACATGGTGAGCAAGGGCGAG	Oct4-Venus-IRES-Puro cloning PCR	2.4
Puro-NsiI_rv	tttttATGCATGGCACCGGGCTTGCGGG	Oct4-Venus-IRES-Puro cloning PCR	2.4
Venus-EcoRI_fw	TTTTTGAATTCATGGTGAGCAAGGGCGAG	Venus ORF PCR	2.4
Venus-EcoRI_rv	TTTTTGAATTCTCACTTGTACAGCTCGTCCATGC	Venus ORF PCR	4
Tbx6-rec-Venus_3	cccagacggtagccagtccccagggaggggtacaactctcgtggatggtaTTC CCAGTCACGACGTTGTA	BAC reporter recombineering PCR	5
Noto-rec-iCre_5	CTCCCATTGAGCTCCTTGCACAGCCTGGGAGGTCCCCTCAGG GTCGCGCAATGGTGCCCAAGAAGAAGAGG	Lineage tracing recombineering PCR	5
Noto-rec-iCre_3	GGGCGCAGGCTCCCGGGCTGGACCTGAGTGCCTGAGGGAG CAGGGCTGGAATTATGTACCTGACTGATGAAGTTCC	Lineage tracing recombineering PCR	5

4.16 List of plasmids and BACs

Table 4-3 List of Plasmids and BACs generated and used in this study

Vector name	Application	Results Section
px335-TE1_sg1 (px335-hU6-TE1_sg1-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE1 CRISPR deletion, TUD 5' breaking point, T/+ 5' breaking point	2.1
px335-TE1_sg2 (px335-hU6-TE1_sg1-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE1 CRISPR deletion, 5'nick	2.1
px335-TE1_sg3 (px335-hU6-TE1_sg1-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE1 CRISPR deletion, 3'nick	2.1
px335-TE1_sg4 (px335-hU6-TE1_sg1-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE1 CRISPR deletion, 3'nick	2.1
px335-TE2_sg1 (px335-hU6-TE2_sg1-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE2 CRISPR deletion, 5'nick	2.1
px335-TE2_sg2 (px335-hU6-TE2_sg2-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE2 CRISPR deletion, 5'nick	2.1
px335-TE2_sg3 (px335-hU6-TE2_sg3-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE2 CRISPR deletion, 3'nick	2.1
px335-TE2_sg4 (px335-hU6-TE2_sg4-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE2 CRISPR deletion, 3'nick	2.1
px335-TE3_sg1 (px335-hU6-TE3_sg1-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE3 CRISPR deletion, 5'nick	2.1
px335-TE3_sg2 (px335-hU6-TE3_sg2-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE3 CRISPR deletion, 5'nick	2.1
px335-TE3_sg3 (px335-hU6-TE3_sg3-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE3 CRISPR deletion, 3'nick	2.1
px335-TE3_sg4 (px335-hU6-TE3_sg4-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE3 CRISPR deletion, 3'nick	2.1
px335-TE4_sg1 (px335-hU6-TE4_sg1-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE4 CRISPR deletion, 5'nick	2.1
px335-TE4_sg2 (px335-hU6-TE4_sg2 -CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE4 CRISPR deletion, 5'nick	2.1
px335-TE4_sg3 (px335-hU6-TE4_sg3 -CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE4 CRISPR deletion, 3'nick	2.1
px335-TE4_sg4 (px335-hU6-TE4_sg4-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE4 CRISPR deletion, 3'nick	2.1
px335-TE5_sg3 (px335-hU6-TE5_sg3-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE5 CRISPR deletion, 3'nick	2.1
px335-TE5_sg4 (px335-hU6-TE5_sg4-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE5 CRISPR deletion, 3'nick	2.1
px335-TE7_sg1 (px335-hU6-TE7_sg1-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE7 CRISPR deletion, 5'nick	2.1

px335-TE7_sg2 (px335-hU6-TE7_sg2-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE7 CRISPR deletion, 5'nick	2.1
px335-TE7_sg3 (px335-hU6-TE7_sg3-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE7 CRISPR deletion, 3'nick	2.1
px335-TE7_sg4 (px335-hU6-TE7_sg4-CAG-hCas9(D12A)-EF1-EGFP-T2A-Puro)	TE7 CRISPR deletion, 3'nick	2.1
px459-T-8000 (px459-hU6-T-8000-CAG-hCas9-EF1-EGFP-T2A-Puro)	T upstream enhancer cluster deletion, TUD 3' break point	2.1
px459-T2-upstream (px459-hU6-T2-CAG-hCas9-EF1-EGFP-T2A-Puro)	5' break point for T deletion	2.1
px459-TT2-tw5 (px459-hU6-TE2-tw5-CAG-hCas9-EF1-EGFP-T2A-Puro)	CRISPR single stranded oligo template rescue	2.1
px459-T2_E2 (px459-hU6-T2_E2-CAG-hCas9-EF1-EGFP-T2A-Puro)	T2 knockout	2.1
pDonor-TE2-HSP68-bGal (pDonor_uni_loxP_TE2-HSP68-Venusl_TKpA_PGK_lox5171)	Enhancer activity assay	2.1
pDonor-TE3-HSP68-bGal (pDonor_uni_loxP_TE3-HSP68-Venusl_TKpA_PGK_lox5171)	Enhancer activity assay	2.1
RP23-289M19 Noto::H2B-mCherry-PGK-hygro	Noto Reporter BAC	2.1-5
RP23-421P23 Tbx6::H2B-mCherry Neo	Tbx6 Reporter BAC	2.5
BAC RP24-530D23 T::H2B-Venus	T Reporter BAC	2.3
RP23-254G2 Foxa2::H2B-mTurquoise2	Foxa2 Reporter BAC	2.3
RP23-289M19 Noto::iCre-pA-PGK-hygro	Noto lineage tracing BAC	2.5
pBS_Oct4-Venus-IRES-Puro (#7)	targeting vector for homologous recombination	2.4

4.17 Composition of cell culture media

Table 4-4 Growth media for cell culture

Feeder/Embryonic fibroblast medium

| ad 100%	Dulbecco's Modified Eagles medium w/o sodium pyruvate (*Lonza BE12-614F*)
10%	Regular fetal calf serum (*PanBioTech P30-3306*)
20Mm	Glutamine (200 Mm) (*Lonza BE17-605E*)
500 U/ml	Penicillin (5000 U/ml) / streptomycin (5000 µg/ml) (*Lonza DE17-603E*)

ES Cell medium

ad 100%	Dulbecco's Modified Eagles medium with sodium pyruvate (*Gibco 10829-018*)
75 ml	ES cell grade fetal calf serum (*PanBioTech P30-2602*)
20Mm	Glutamine (200 Mm) (*Lonza BE17-605E*)
500 U/ml	Penicillin (5000 U/ml) / streptomycin (5000 µg/ml) (*Lonza DE17-603E*)
1x	100x non-essential amino acids (*Gibco 11140-35*)
2x	1000x β-mercaptothanol (*Gibco 21985-023*)
1x	100x nucleosides (*Chemicon ES-008D*)

NotoDiff medium

Ad 100 %	Knock Out Knockout Dulbecco's Modified Eagle's Medium with sodium pyruvate (*Gibco 10829-018*)
1x	N-2 Supplement 100x (*Gibco 17502-048*)
1x	B-27 Supplement 50x w/o Vitamin A (*Gibco 12587-010*)
1x	100x MEM non-essential amino acids (*Gibco 1140-35*)
2x	1000x β-mercaptothanol (*Gibco 21985-023*)
500 U/ml	Penicillin (5000 U/ml) / streptomycin (5000 µg/ml) (*Lonza DE17-603E*)
20Mm	Glutamine (200 Mm) (*Lonza BE17-605E*)
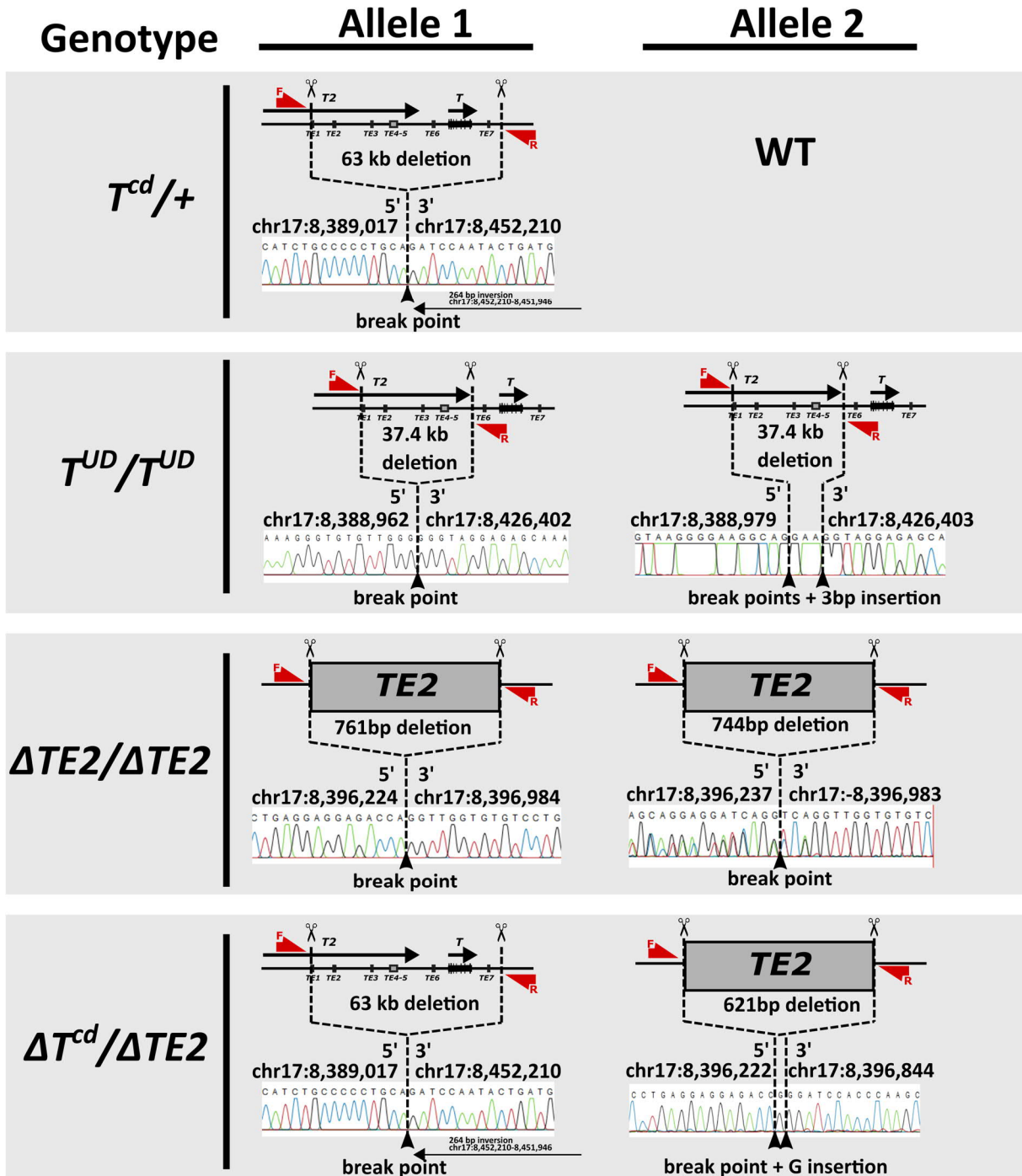
Supplementary Figure 1 Genotyping of Enhancer Deletions 1/3

Details specified in the caption of Supplementary Figure 3.

| Genotype	Allele 1	Allele 2

ΔTE3/ΔTE3
- TE3 — 241 bp deletion — 5' 3' — chr17:8,408,975 | chr17:8,409,217 — break point
- TE3 — 242 bp deletion — 5' 3' — chr17:8,408,984 | chr17:8,409,228 — break point

ΔT^{cd}/ΔTE3
- 63 kb deletion — 5' 3' 264 bp inversion chr17:8,452,210-8,451,946 — chr17:8,389,017 | chr17:8,452,209 — break point
- TE3 — 184 bp deletion — 5' 3' — chr17:8,409,017 | chr17:8,409,203 — break point

ΔTE2/ΔTE2[1] ΔTE3/ΔTE3
- TE3 — 215 bp deletion — 5' 3' — chr17:8,408,999 | chr17:8,409,215 — break point
- TE3 — 247 bp deletion — 5' 3' — chr17:8,408,967 | chr17:8,409,011-8,409,011 | chr17:8,409,215 — break points

ΔT^{cd}/ΔTE2[1]; ΔTE4-5
- 63 kb deletion — 5' 3' 264 bp inversion chr17:8,452,210-8,451,946 — chr17:8,389,017 | chr17:8,452,209 — break point
- TE4-5 — 1990 bp deletion — 5' 3' — chr17:8,415,583 | chr17:8,417,575 — break point

ΔT^{cd}/ΔTE2[1]; ΔTE4
- 63 kb deletion — 5' 3' 264 bp inversion chr17:8,452,210-8,451,946 — chr17:8,389,017 | chr17:8,452,209 — break point
- TE4 — 832 bp deletion — 5' 3' — chr17:8,415,601 | chr17:8,416,435 — break point

[1] ΔTE2 deletions
on previous page

Supplementary Figure 2 Genotyping of Enhancer Deletions 2/3

Details specified in the caption of Supplementary Figure 3.

128
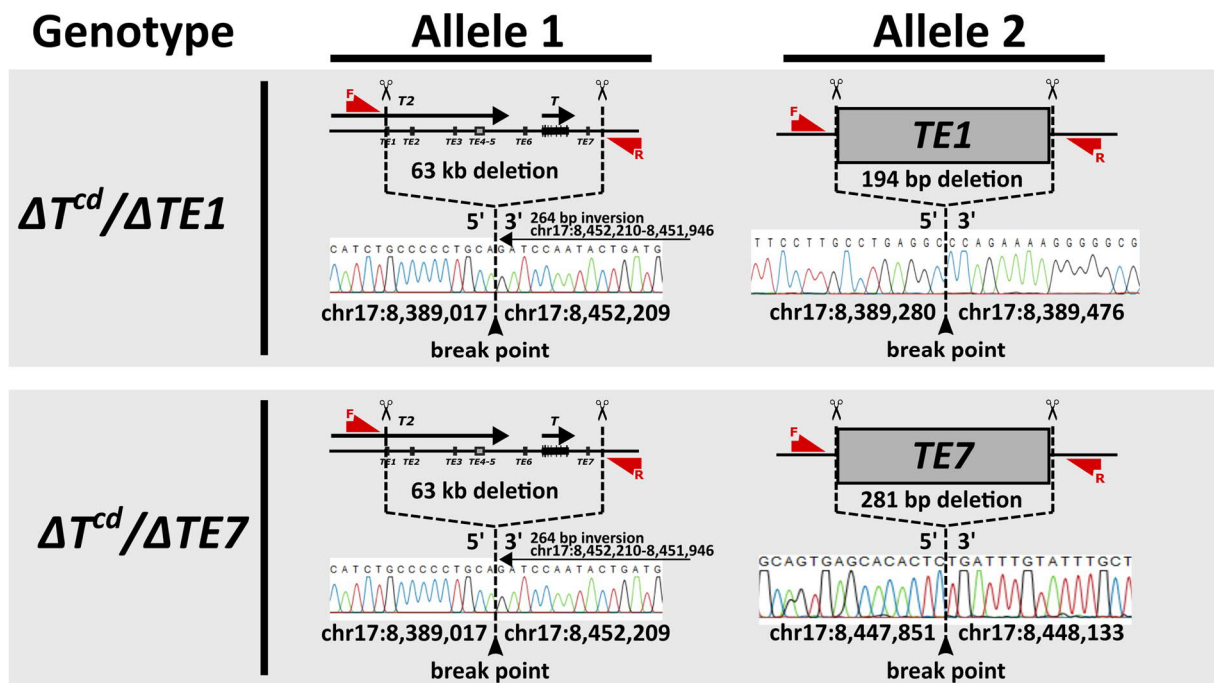
Supplementary Figure 3 Genotyping of Enhancer Deletions 3/3

Red arrows indicate the position of the forward (F) and reverse (R) primers used for genotyping by PCR, not scaled. Scissor icons indicate the approximate positions of gRNA target sites. The dotted line indicates the CRISPR/Cas9 mediated deletion and converges at the break point(s). The coordinates (mm10_Dec2011) of the bordering 5′ and 3′ edges of the deletion 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 genotypes of which more than one clone was generated, the clone #1 is shown in the panels.
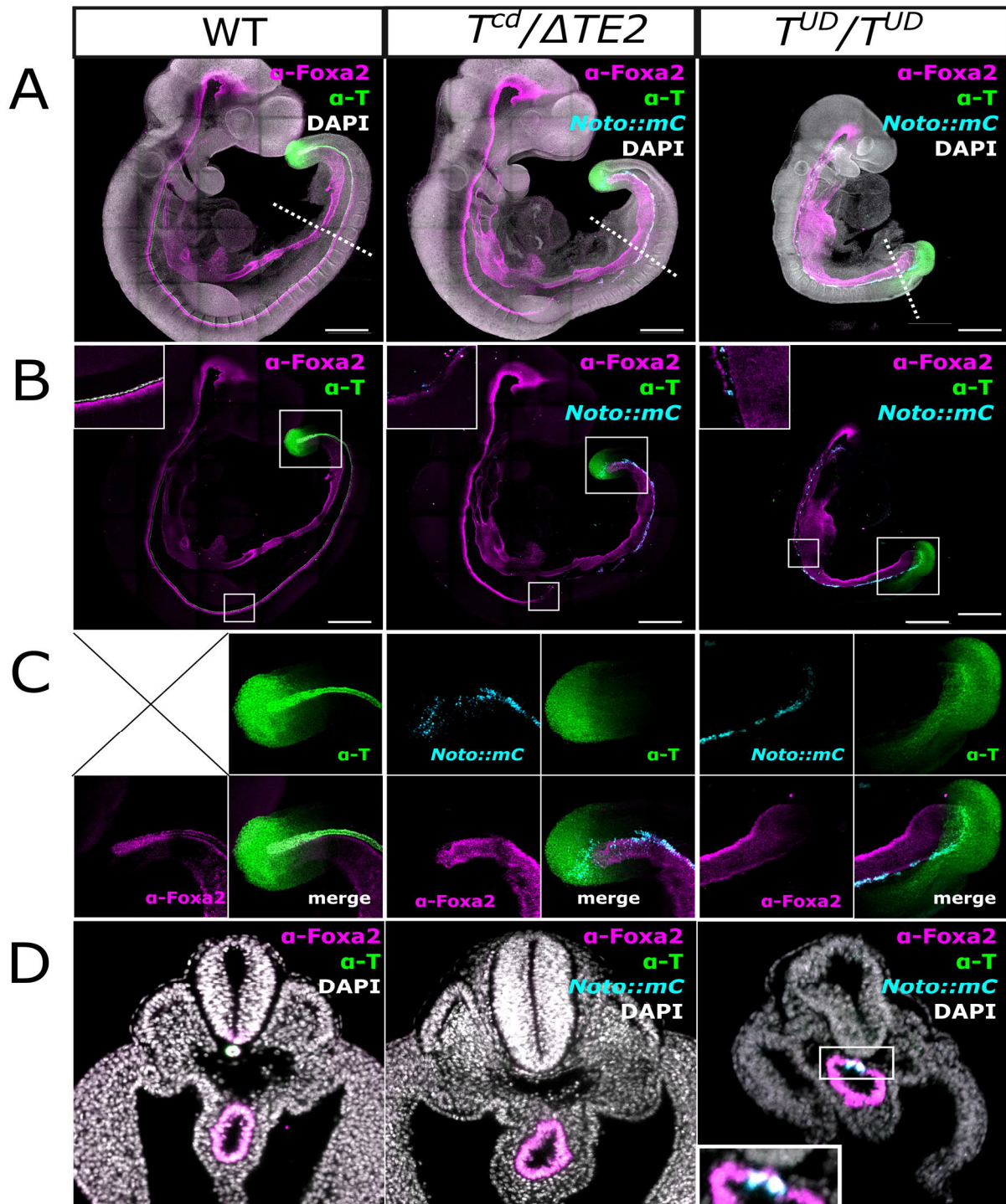
Supplementary Figure 4 Trunk notochord and floor plate are absent in $T^{cd}/\Delta TE2$ and $\Delta T^{UD}/\Delta T^{UD}$ mutants

(A) Maximum intensity projection of confocal imaging stacks with immunofluorescence for T (green) and Foxa2 (magenta). Nuclei stained with DAPI (grey). Noto^{mC} reporter signal (light blue). Dotted line indicates the plane for the optical section shown in (D). (B) The same MIP as in (A) without the DAPI channel. The smaller box indicates the enlarged area where the floor plate was disrupted in mutants. Mutants lack Noto^{mC}+ or T+ cells in the midline. The boxed region in the caudal end is enlarged in (C), showing individual channels. The wild-type control did not carry a Noto^{mC} reporter. Scale bar = 200μm
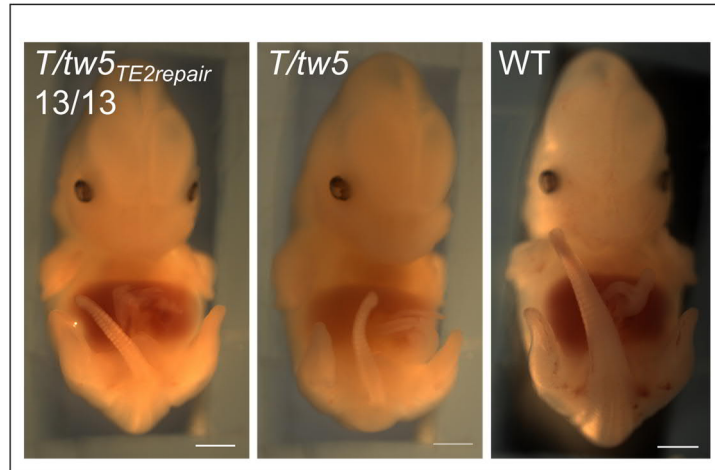
Supplementary Figure 5 No rescue of the tail phenotype in T/tw5 after repair of two point mutations in TE2

E12.5 embryos of *T/tw5*_{repair}, *T/tw5* and F1G4 wildtype clones generated by tetraploid complementation assays. The tailless phenotype in *T/tw5* was not reversed after conversion of the *TE2* core sequence to wild-type (BL6). Sequences and engineering strategy are shown in 1.2.12.

E12.5

| T2 -/- #1 17/17 | T2 -/- #2 16/16 | WT | ΔTE2/ΔTE2 |
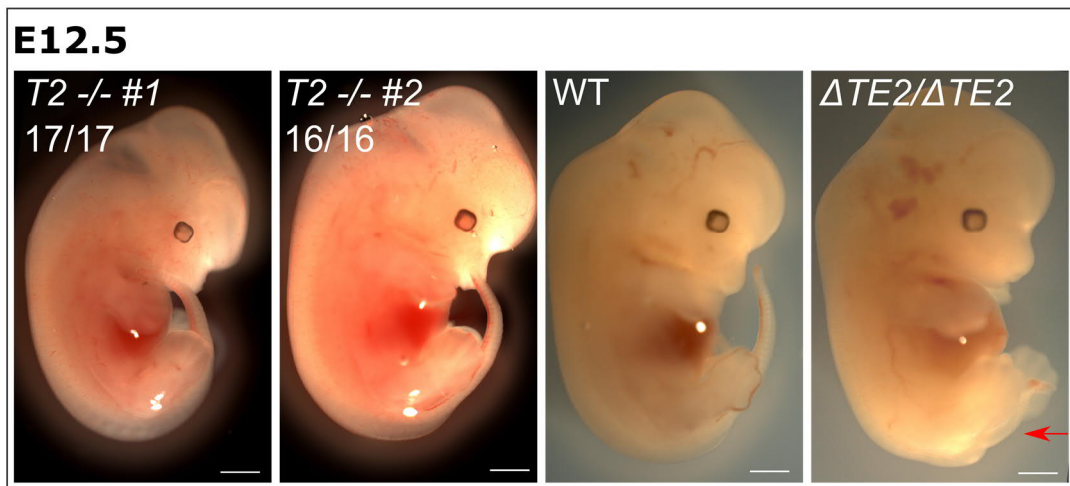
Supplementary Figure 6 Disruption of the *T2* ORF does not induce a tail phenotype

E12.5 Embryos generated via tetraploid aggregation. *T2 -/-* does not recapitulate the axial truncation phenotype of *ΔTE2/ΔTE2* pointed out by the red arrow. Number of analyzed embryos with the same morphology indicated for *T2 -/-* clone #1 (d2) and clone #2 (d3). Scale bar = 1 mm.

0.70   0.85   1.0

Noto^{mC}+ *in vitro* diff d7 (pooled)

Noto^{mC}+ *in vitro* diff d7 (high)

Noto^{mC}+/T^{Ve}+/Foxa2^{mT}+ trunk notochord 4-11S

Noto^{mC}+/T^{Ve}+/Foxa2^{mT}+ trunk notochord 24-32S

Noto^{mC}+/Tbx6^{Ve}+ trunk E9.5

T^{Ve}high/Noto^{mC}-/Foxa2^{mT}- 24-28S (MesoP)

Noto^{mC}+ *T/tw5* 24-28S

Noto^{mC}-/Tbx6^{Ve}+ trunk E9.5

Noto^{mC}+ 12-15S

Noto^{mC}+/T^{Ve}+/Foxa2^{mT}+ 24-28S

Noto^{mC}+/Tbx6^{Ve}+ CE E9.5

Noto^{mC}+ 24-28S

Noto^{mC}+ 16-18S

Noto^{mC}+/T^{Ve}+/Foxa2^{mT}+ 9-11S

Noto^{mC}+/Tbx6^{Ve}+ CE *Tbx6 -/-* E9.5

Noto^{mC}+/Tbx6^{Ve}- CE E9.5

Noto^{mC}+ *T/tw5* 16-18S

T^{Ve}high/Noto^{mC}-/Foxa2^{mT}- 9-11S (MesoP)

Noto^{mC}+ late headfold

Noto^{mC}+ early headfold

Noto^{mC}+/Tbx6^{Ve}- trunk E9.5

Supplementary Figure 7 Sample tree generated by hierarchical clustering of all samples isolated from Noto+ cells.

A list of FPKM values with all samples from Noto+ cells was generated using cufflinks (Trapnell et al., 2013). Sample trees were generated in MeV (Howe et al., 2011) based on log transformed FPKM values.
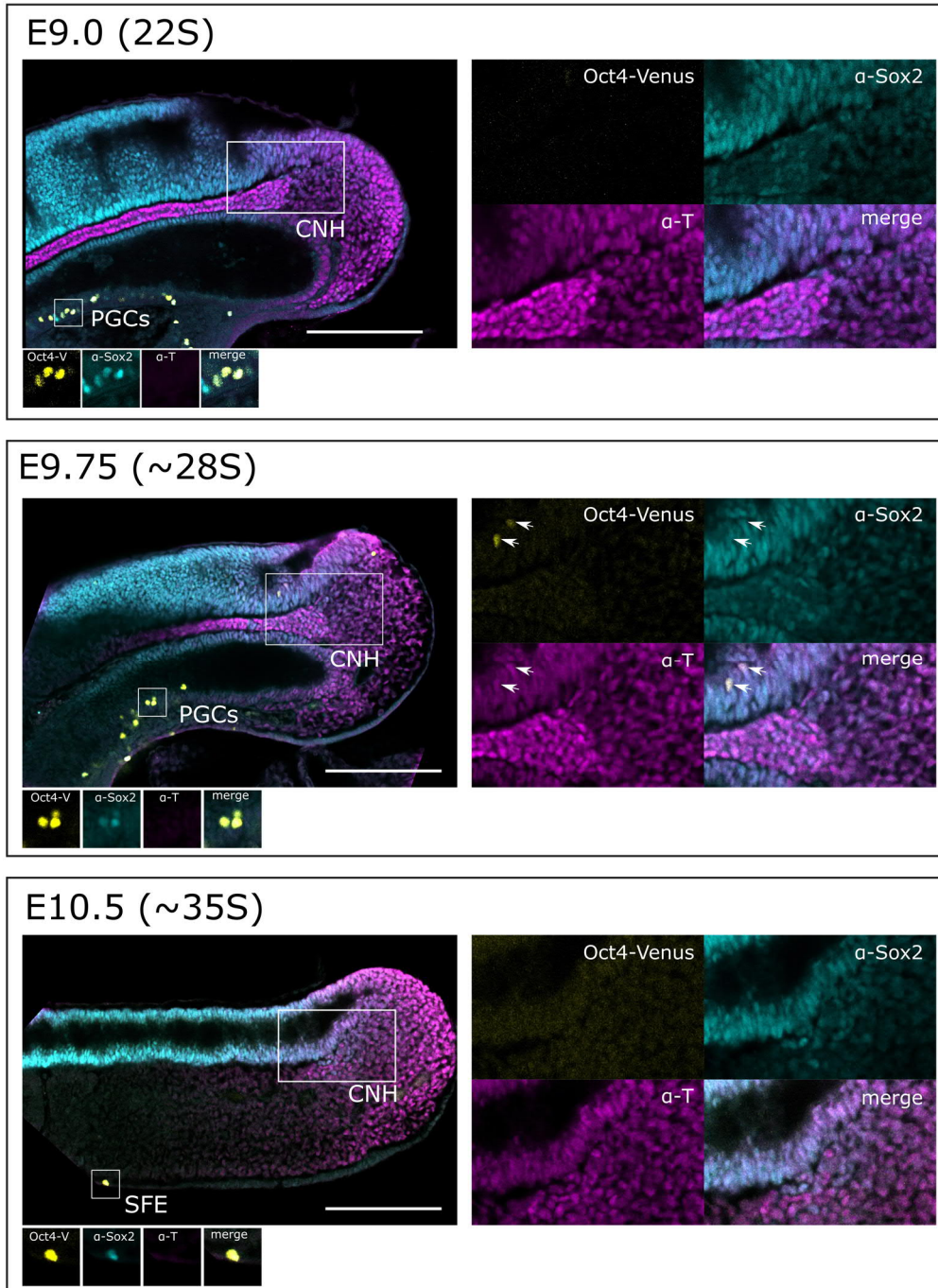
Supplementary Figure 8 Oct4-Venus expressing cells in the caudal end between E9.0 and E10.5

Optical sections acquired by confocal microscopy. Oct4-Venus signal (yellow) with immunofluorescence for T (magenta) and Sox2 (light blue). Oct4-Venus expressing cells in the Chordo-Neural-Hinge (CNH; boxed region enlarged in panels on the right) were only detected between the 25 and 32 somite stage. Arrows point out Oct4-Venus+ cells. Oct4-Venus expressing Primordial Germ Cells (PGCs) and Surface Ectoderm (SFE) cells are shown as a control. Scale bar=200µm.

A

Naruse-Nakajima et al. 2001

B
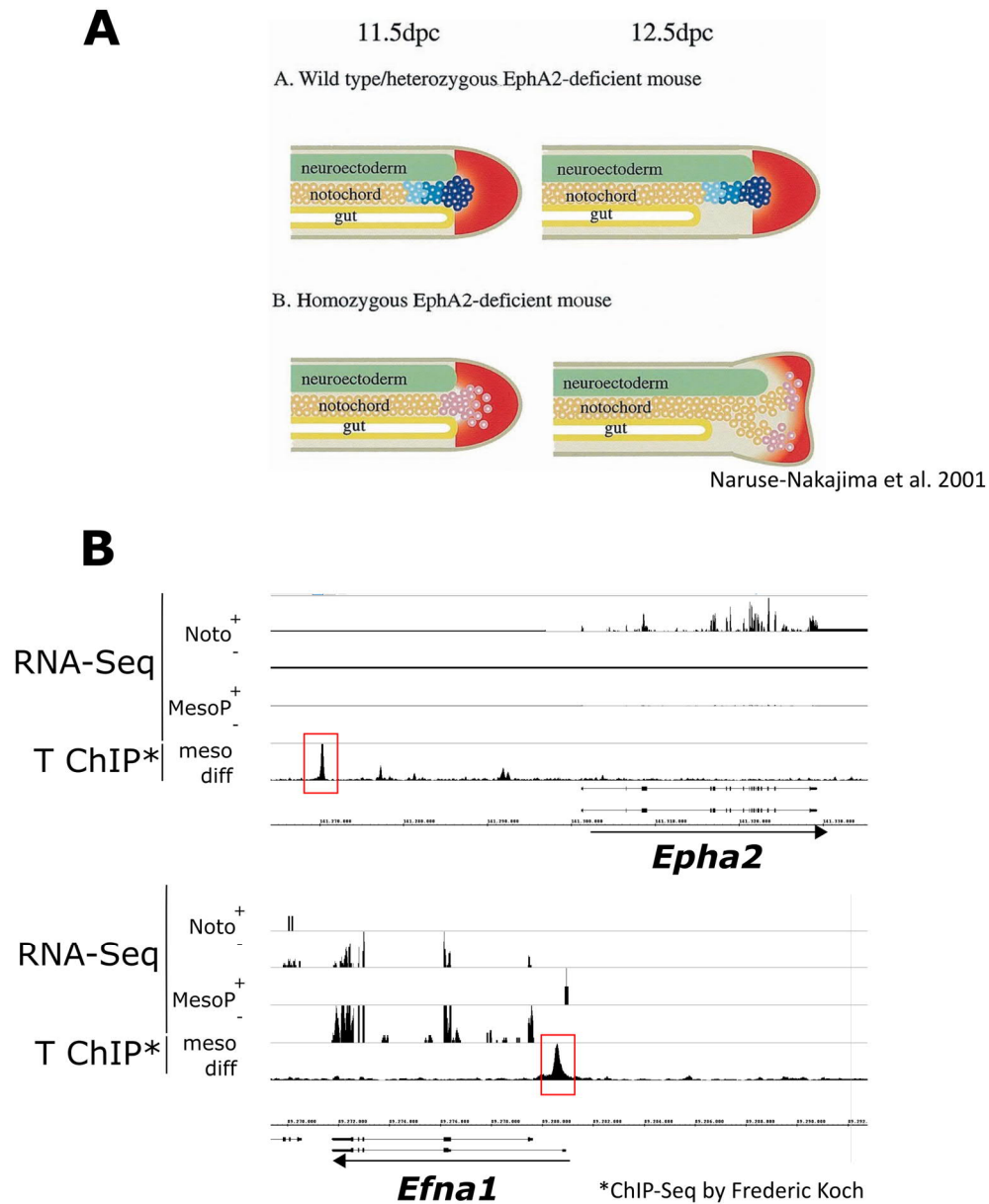
Supplementary Figure 9 A possible mechanism explaining the *ΔTE3/ ΔTE3* mutant phenotype

(A) Figure adapted from Naruse-Nakajima et al, 2001. In their study with *Epha2* deficient mice, the authors observed a split notochord and tail bifurcation phenotype that is highly similar to *ΔTE3/ ΔTE3* and proposed a mechanism for the segregation of axial and paraxial mesoderm. Notochord cells specifically express Epha2 and are thought to be repelled by Efna1 secreting mesoderm progenitors. (B) RNA-Seq data of Noto[mC]+/T[Ve]+/Foxa2[mT]+ cells and Noto[mC]-/T[Vehigh]/Foxa2[mT]- mesoderm precursors from this work confirms differential expression of *Epha2* and *Efna1* in these cell types. ChIP-Seq for T on *in vitro* differentiated mesoderm precursors (performed by Frederic Koch, MPIMG) shows that both *Epha2* and *Efna1* are regulated by T.

Genotype	Deletion mm10 – #Clone 1) allele 1 / 2) allele2)
T^{cd} /+	1) chr17:8,389,017-8,451,946 + 264 bp inversion chr17:8,452,210-8,451,946 (2) chr17:8,388,978-8,389,054)
ΔT^{UD}/ ΔT^{UD}	#1 1) chr17:8,388,962-8,426,402 #1 2) chr17:8,388,979-8,426,403 #2 1) chr17:8,388,988-8,426,402 #2 2) chr17:8,388,968-8,426,404
$\Delta TE2$/$\Delta TE2$	#1 1) chr17:8,396,238-8,396,982 #1 2) chr17:8,396,224-8,396,984
T^{cd}/$\Delta TE2$	#1 1) chr17:8,389,017-8,451,946 + 264 bp inversion #1 2) chr17:8,396,223-8,396,843 + chr17:8,396,897-8,396,967
$\Delta TE3$/$\Delta TE3$	#1 1) chr17:8,408,976-8,409,216 #1 2) chr17:8,408,985-8,409,227 #2 1) chr17:8,408,984-8,409,214 #2 2) chr17:8,408,988-8,409,214
T^{cd}/$\Delta TE3$	1) chr17:8,389,017-8,451,946 + 264 bp inversion #1 2) chr17:8,409,018-8,409,204
$\Delta TE2$/$\Delta TE2$; $\Delta TE3$/$\Delta TE3$	#1 1) chr17:8,409,000-8,409,214 #1 2) chr17:8,408,968-8,409,214 #2 1) chr17:8,408,992-8,409,214 #2 2) chr17:8,408,968-8,409,206
T^{cd}/$\Delta TE2$; $\Delta TE4$-5	#1 1) chr17:8,389,050-8,452,037 #1 2) chr17:8,415,584-8,417,574 #2 chr17:8,396,245-8,396,869 #2 chr17:8,396,897-8,396,967
T2 -/-	#1 1) chr17:8,389,335-8,389,347 #1 2) chr17:8,389,335-8,389,356 #2 1) chr17:8,389,347 C insertion #2 2) chr17:8,389,340-8,389,458
T^{cd}/$\Delta TE1$	#1 1) chr17:8,389,017-8,451,946 + 264 bp inversion #1 2) chr17:8,389,281-8,389,475 + chr17:8,388,978-8,389,054
T^{cd}/$\Delta TE7$	#1 chr17:8,389,017-8,451,946 + 264 bp inversion #1 2) chr17:8,447,850-8,448,132
T^{cd}/$\Delta TE2$; $\Delta TE4$	#1 1) chr17:8,389,050-8,452,037 #1 2) chr17:8,396,245-8,396,869; chr17:8,415,602-8,416,434 #2 1) chr17:8,389,050-8,452,037 #2 2) chr17:8,396,245-8,396,869; chr17:8,415,594-8,416,428

Supplementary Table 1 Coordinates of deletions (mm10_Dec2011) in mutants generated in this work

sample	embryos	cells	Lib. Type	Kit	mapped reads (final)
Noto^{mC}+/T^{Ve}+/Foxa2^{mT}+ E8.5 9-11S	5	250/308	RNA-seq	Ovation SoLo	3343606
T^{Ve}_high/Noto^{mC}-/Foxa2^{mT}- 9-11S	5	24	RNA-seq	Ovation SoLo	1198064
Noto^{mC}+/T^{Ve}+/Foxa2^{mT} _high E9.5 24-28 somites	5	250/294	RNA-seq	Ovation SoLo	4788324
T^{Ve}_high/Noto^{mC}-/Foxa2^{mT} - 24-28S	5	250/854	RNA-seq	Ovation SoLo	6246472
Noto^{mC}+ early headfold	10	153	RNA-seq	Ovation SoLo	3211886
Noto^{mC}+ late headfold	10	242	RNA-seq	Ovation SoLo	7408828
12-15S Noto^{mC}+	8	234	RNA-seq	Ovation SoLo	4205712
16-18S Noto^{mC}+	7	113	RNA-seq	Ovation SoLo	11010782
24-28S Noto^{mC}+	3	250/712	RNA-seq	Ovation SoLo	10447696
Noto^{mC}+ *T/tw5* 16-18S	5	250/320	RNA-seq	Ovation SoLo	2349740
Noto^{mC}+ *T/tw5* 24-28S	5	250/378	RNA-seq	Ovation SoLo	2460036
Noto^{mC}+/T^{Ve}+/Foxa2^{mT} _high 4-11S	8	73	RNA-seq	Ovation SoLo	1664032
Noto^{mC}+/T^{Ve}+/Foxa2^{mT} _high 24-32 S	5	90	RNA-seq	Ovation SoLo	1581548
Noto^{mC} *in vitro* d7 (high)	5	250/10K	RNA-seq	Ovation SoLo	7243676
Noto^{mC} *in vitro* diff d7 (pooled)	5	250/20K	RNA-seq	Ovation SoLo	8090474
EHF/LHF Oct4-Venus+	5	250/418	RNA-seq	Ovation SoLo	5281278
EHF/LHF Noto^{mC}+/Oct4-Venus+	5	26	RNA-seq	Ovation SoLo	1858086
EHF/LHF Noto^{mC}+	5	37	RNA-seq	Ovation SoLo	812152
28-32S Oct4-Venus high	15	144	RNA-seq	Ovation SoLo	4598610
28-32S Oct4-Venus low	15	250/369	RNA-seq	Ovation SoLo	5026586
allantois Oct4-Venus+ stringent	5	54	RNA-seq	Ovation SoLo	2074570
CE Noto^{mC}+	5	78	RNA-seq	Ovation SoLo	5198926
CE Noto^{mC}+/ Tbx6^{Ve} +	5	250/346	RNA-seq	Ovation SoLo	11167966
Noto^{mC}+/ Tbx6^{Ve} + *Tbx6* -/-	5	250/911	RNA-seq	Ovation SoLo	9062172
trunk Noto^{mC}+/Tbx6^{Ve}+	5	75	RNA-seq	Ovation SoLo	5140352
trunk Noto^{mC}-/Tbx6^{Ve}+ trunk somite	5	250/5000	RNA-seq	Ovation SoLo	9245464
Noto^{mC}+/Tbx6^{Ve}- trunk notochord	5	18	RNA-seq	Ovation SoLo	5913248

Supplementary Table 2Mapped reads in RNA-Seq experiments

Acknowledgements

I want to express my gratitude to Bernhard Herrmann for his guidance, discussion of concepts and the chance to pursue this exciting project.

I wish to thank Frederic Koch, who once upon a time accepted me as a Master's student and whose mentorship and insight was instrumental throughout my doctoral research.

I would like to express my great appreciation to Manuela Scholze-Wittler, who taught me everything I know about cloning and cell culture and has been an amazing and supportive bench neighbor over the years.

I also want to thank Jesse Veenvliet for great scientific discussions and suggestions for my thesis manuscript.

I sincerely thank Peter Robin Hiesinger for kindly agreeing to review my thesis.

Finally, I am grateful to my partner, friends and family for providing the support needed to pull through the project.

Bibliography

Abdelkhalek, H. Ben, 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.

Abrami, L., Kunz, B., Deuquet, J., Bafico, A., Davidson, G. and van der Goot, F. G. (2008). Functional interactions between anthrax toxin receptors and the WNT signalling protein LRP6. *Cell. Microbiol.* 10, 2509–2519.

Acampora, D., Di Giovannantonio, L. G. and Simeone, A. (2013). Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. *Development* 140, 43–55.

Aires, R., Jurberg, A. D., Leal, F., N??voa, A., Cohn, M. J. and Mallo, M. (2016). Oct4 Is a Key Regulator of Vertebrate Trunk Length Diversity. *Dev. Cell* 38, 262–274.

Aires, R., de Lemos, L., Nóvoa, A., Jurberg, A. D., Mascrez, B., Duboule, D. and Mallo, M. (2019). Tail Bud Progenitor Activity Relies on a Network Comprising Gdf11, Lin28, and Hox13 Genes. *Dev. Cell* 48, 383-395.e8.

Akiyama, H., Chaboissier, M.-C., Martin, J. F., Schedl, A. and de Crombrugghe, B. (2002). The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev.* 16, 2813–2828.

Albazerchi, A. and Stern, C. D. (2007). A role for the hypoblast (AVE) in the initiation of neural induction, independent of its ability to position the primitive streak. *Dev. Biol.* 301, 489–503.

Amacher, S. L. and Kimmel, C. B. (1998). Promoting notochord fate and repressing muscle development in zebrafish axial mesoderm. *Development* 125, 1397–1406.

Amin, S., Neijts, R., Simmini, S., van Rooijen, C., Tan, S. C., Kester, L., van Oudenaarden, A., Creyghton, M. P. and Deschamps, J. (2016). Cdx and T Brachyury Co-activate Growth Signaling in the Embryonic Axial Progenitor Niche. *Cell Rep.* 17, 3165–3177.

Anderson, W. J., Zhou, Q., Alcalde, V., Kaneko, O. F., Blank, L. J., Sherwood, R. I., Guseh, J. S., Rajagopal, J. and Melton, D. A. (2008). Genetic targeting of the endoderm with claudin-6CreER. *Dev. Dyn.* 237, 504–512.

Andersson, O., Reissmann, E., Jörnvall, H. and Ibáñez, C. F. (2006). Synergistic interaction between Gdf1 and Nodal during anterior axis development. *Dev. Biol.* 293, 370–381.

Andre, P., Song, H., Kim, W., Kispert, A. and Yang, Y. (2015). Wnt5a and Wnt11 regulate mammalian anterior-posterior axis elongation. *Development* 142, 1516–1527.

Ang, S.-L. and Rossant, J. (1994). HNF-3beta Is Essential for Node and Notochord Formation in Mouse Development. *Cell* 76, 561–574.

Antonica, F., Carlo, L., Lester, R. and Zernicka-goetz, M. (2019). Concerted cell divisions in

embryonic visceral endoderm guide anterior visceral endoderm migration. *Dev. Biol.* 450, 132–140.

Aoki, M., Mieda, M., Ikeda, T., Hamada, Y., Nakamura, H. and Okamoto, H. (2007). R-spondin3 is required for mouse placental development. *Dev. Biol.* 301, 218–226.

Arkell, R. and Tam, P. (2012). Initiating head development in mouse embryos: Integrating signalling and transcriptional activity. *Open Biol.* 2, 120030.

Arnold, S. J. and Robertson, E. J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* 10, 91–103.

Aulehla, A., Wehrle, C., Brand-Saberi, B., Kemler, R., Gossler, A., Kanzler, B. and Herrmann, B. G. (2003). Wnt3a plays a major role in the segmentation clock controlling somitogenesis. *Dev. Cell* 4, 395–406.

Aulehla, A., Wiegraebe, W., Baubet, V., Wahl, M. B., Deng, C., Taketo, M., Lewandoski, M. and Pourquié, O. (2008). A beta-catenin gradient links the clock and wavefront systems in mouse embryo segmentation. *Nat. Cell Biol.* 10, 186–193.

Bachiller, D., Klingensmith, J., Kemp, C., Belo, J. A., Anderson, R. M., May, S. R., McMahon, J. A., McMahon, A. P., Harland, R. M., Rossant, J., et al. (2000). The organizer factors Chordin and Noggin are required for mouse forebrain development. *Nature* 403, 658–661.

Balmer, S., Nowotschin, S. and Hadjantonakis, A. K. (2016). Notochord morphogenesis in mice: Current understanding and open questions. *Dev. Dyn.* 245, 547–557.

Barnes, J. D., Crosby, J. L., Jones, C. M., Wright, C. V. E. and Hogan, B. L. M. (1994). Embryonic Expression of Lim-1, the Mouse Homolog of Xenopus XLim-1, Suggests a Role in Lateral Mesoderm Differentiation and Neurogenesis. *Dev. Biol.* 161, 168–178.

Beddington, S. P. (1981). An autoradiographic analysis of the potency of embryonic ectoderm in the 8th day postimplantation mouse embryo. *J. Embryol. Exp. Morphol.* 64, 87–104.

Beddington, R. S. (1994). Induction of a second neural axis by the mouse node. *Development* 120, 613–620.

Beisaw, A., Tsaytler, P., Koch, F., Schmitz, S. U., Melissari, M., Senft, A. D., Wittler, L., Pennimpede, T., Macura, K., Herrmann, B. G., et al. (2018). BRACHYURY directs histone acetylation to target loci during mesoderm development. *EMBO Rep.* 19, 118–134.

Bell, C. C., Amaral, P. P., Kalsbeek, A., Magor, G. W., Gillinder, K. R., Tangermann, P., di Lisio, L., Cheetham, S. W., Gruhl, F., Frith, J., et al. (2016). The Evx1/Evx1as gene locus regulates anterior-posterior patterning during gastrulation. *Sci. Rep.* 6, 26657.

Ben-Haim, N., Lu, C., Guzman-Ayala, M., Pescatore, L., Mesnard, D., Bischofberger, M., Naef, F., Robertson, E. J. and Constam, D. B. (2006). The nodal precursor acting via activin receptors induces mesoderm by maintaining a source of its convertases and BMP4. *Dev. Cell* 11, 313–323.

Bénazéraf, B. and Pourquié, O. (2013). Formation and Segmentation of the Vertebrate Body Axis. *Annu. Rev. Cell Dev. Biol* 29, 1–26.

Bennett, D. (1975). The T-locus of the mouse. *Cell* 6, 441–454.

Besnard, V., Wert, S. E., Hull, W. M. and Whitsett, J. A. (2004). Immunohistochemical localization of Foxa1 and Foxa2 in mouse embryos and adult tissues. *Gene Expr. Patterns* 5, 193–208.

Bettenhausen, B., Hrabě de Angelis, M., Simon, D., Guénet, J. L. and Gossler, A. (1995). Transient and restricted expression during mouse embryogenesis of Dll1, a murine gene closely related to Drosophila Delta. *Development* 121, 2407–2418.

Blum, M., Andre, P., Muders, K., Schweickert, A., Fischer, A., Bitzer, E., Bogusch, S., Beyer, T., Van Straaten, H. W. M. and Viebahn, C. (2007). Ciliation and gene expression distinguish between node and posterior notochord in the mammalian embryo. *Differentiation* 75, 133–146.

Bober, E., Franz, T., Arnold, H. H., Gruss, P. and Tremblay, P. (1994). Pax-3 is required for the development of limb muscles: a possible role for the migration of dermomyotomal muscle progenitor cells. *Development* 120, 603–612.

Boija, A., Klein, I. A., Sabari, B. R., Dall'Agnese, A., Coffey, E. L., Zamudio, A. V, Li, C. H., Shrinivas, K., Manteiga, J. C., Hannett, N. M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* 175, 1842-1855.e16.

Boulet, A. M. and Capecchi, M. R. (2012). Signaling by FGF4 and FGF8 is required for axial elongation of the mouse embryo. *Dev. Biol.* 371, 235–245.

Boulter, C., Mulroy, S., Webb, S., Fleming, S., Brindle, K. and Sandford, R. (2001). Cardiovascular, skeletal, and renal defects in mice with a targeted disruption of the Pkd1 gene. *Proc. Natl. Acad. Sci. U. S. A.* 98, 12174–12179.

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. (Berl).* 188, 239–245.

Brzóska, H. Ł., d'Esposito, A. M., Kolatsi-Joannou, M., Patel, V., Igarashi, P., Lei, Y., Finnell, R. H., Lythgoe, M. F., Woolf, A. S., Papakrivopoulou, E., et al. (2016). Planar cell polarity genes Celsr1 and Vangl2 are necessary for kidney growth, differentiation, and rostrocaudal patterning. *Kidney Int.* 90, 1274–1284.

Burtscher, I. and Lickert, H. (2009). Foxa2 regulates polarity and epithelialization in the endoderm germ layer of the mouse embryo. *Development* 136, 1029–1038.

Cambray, N. and Wilson, V. (2002). Axial progenitors with extensive potency are localised to the mouse chordoneural hinge. *Development* 129, 4855–4866.

Cambray, N. and Wilson, V. (2007). Two distinct sources for a population of maturing axial progenitors. *Development* 134, 2829–2840.

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.

Chapman, D. L., Agulnik, I., Hancock, S., Silver, L. M. and Papaioannou, V. E. (1996). Tbx6,a Mouse T-Box Gene Implicated in Paraxial Mesoderm Formation at Gastrulation. *Dev. Biol.* 180, 534–542.

Chenevix-Trench, G., Jones, K., Green, A. C., Duffy, D. L. and Martin, N. G. (1992). Cleft lip with or without cleft palate: associations with transforming growth factor alpha and retinoic acid receptor loci. *Am. J. Hum. Genet.* 51, 1377–1385.

Cherepanova, O. A., Gomez, D., Shankman, L. S., Swiatlowska, P., Williams, J., Sarmento, O. F., Alencar, G. F., Hess, D. L., Bevard, M. H., Greene, E. S., et al. (2016). Activation of the pluripotency factor OCT4 in smooth muscle cells is atheroprotective. *Nat. Med.* 22, 657–665.

Chesley, P. (1935). Development of the Short-tailed Mutant in the House Mouse. *J. Exp. Zool.*

Chesley, P. and Dunn, L. C. (1936). The Inheritance of Taillessness (Anury) in the House Mouse. *Genetics* 21, 525–536.

Chiang, C., Litingtung, Y., Lee, E., Young, K. E., Corden, J. L., Westphal, H. and Beachy, P. a (1996). Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* 383, 407–413.

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. an Off. Publ. Am. Assoc. Anat.* 237, 3953–3958.

Chugh, R., Tawbi, H., Lucas, D. R., Biermann, J. S., Schuetze, S. M. and Baker, L. H. (2007). Chordoma: the nonsarcoma primary bone tumor. *Oncologist* 12, 1344–1350.

Ciruna, B. and Rossant, J. (2001). Mesoderm Cell Fate Specification and Morphogenetic Movement at the Primitive Streak. 1, 37–49.

Cleaver, O. and Krieg, P. A. (2001). Notochord Patterning of the Endoderm. *Dev. Biol.* 234,.

Clements, D., Taylor, H. C., Herrmann, B. G. and Stott, D. (1996). Distinct regulatory control of the Brachyury gene in axial and non-axial mesoderm suggests separation of mesoderm lineages early in mouse gastrulation. *Mech. Dev.* 56, 139–149.

Concepcion, D., Washkowitz, A. J., Desantis, A., Ogea, P., Yang, J. I., Douglas, N. C. and Papaioannou, V. E. (2017). Cell lineage of timed cohorts of Tbx6 -expressing cells in wild-type and Tbx6 mutant embryos. *Biol. Open* 1065–1073.

Davidson, B. P., Kinder, S. J., Steiner, K., Schoenwolf, G. C. and Tam, P. P. (1999). Impact of node ablation on the morphogenesis of the body axis and the lateral asymmetry of the mouse embryo during early organogenesis. *Dev. Biol.* 211, 11–26.

De Robertis, E. M. and Kuroda, H. (2004). DORSAL-VENTRAL PATTERNING AND NEURAL INDUCTION IN XENOPUS EMBRYOS. *Annu. Rev. Cell Dev. Biol.* 20, 285–308.

del Corral, R. and Morales, A. V (2017). The Multiple Roles of FGF Signaling in the Developing Spinal Cord. *Front. Cell Dev. Biol.* 5, 58.

Delfino-Machín, M., Lunn, J. S., Breitkreuz, D. N., Akai, J. and Storey, K. G. (2005). Specification and maintenance of the spinal cord stem zone. *Development* 132, 4273–4283.

Denans, N., Iimura, T. and Pourquié, O. (2015). Hox genes control vertebrate body elongation by collinear Wnt repression. *Elife* 4,.

DeVeale, B., Brokhman, I., Mohseni, P., Babak, T., Yoon, C., Lin, A., Onishi, K., Tomilin, A., Pevny, L., Zandstra, P. W., et al. (2013). Oct4 is required ~E7.5 for proliferation in the primitive streak. *PLoS Genet.* 9, e1003957.

Dias, A., Lozovska, A., Wymeersch, F. J., Nóvoa, A., Binagui-Casas, A., Sobral, D., Martins, G. G., Wilson, V. and Mallo, M. (2020). A Tgfbr1/Snai1-dependent developmental module at the core of vertebrate axial elongation. *Elife* 9, e56615.

Diez del Corral, R., Olivera-Martinez, I., Goriely, A., Gale, E., Maden, M. and Storey, K. (2003). Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. *Neuron* 40, 65–79.

Dobrovolskaia-Zavadskaia, N. and Kobozieff, N. (1932). Les souris anoures et à queue filiforme qui se reproduisent entre elles sans disjonction. *CR des séances la société Biol.* 60, 882–784.

Downs, K. M. (2008). Systematic localization of Oct-3/4 to the gastrulating mouse conceptus suggests manifold roles in mammalian development. *Dev. Dyn. an Off. Publ. Am. Assoc. Anat.* 237, 464–475.

Dunn, L. C. and Suckling, J. (1956). STUDIES OF THE GENETIC VARIABILITY IN WILD POPULATIONS OF HOUSE MICE. I. ANALYSIS OF SEVEN ALLELES AT LOCUS T. *Genetics* 41, 344–352.

Dunwoodie, S., Henrique, D., Harrison, S. and Beddington, R. S. P. (1997). Mouse Dll3: A novel divergent Delta gene which may complement the function of other Delta homologues during early pattern formation in the mouse embryo. *Development* 124, 3065–3076.

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.

Echelard, Y., Epstein, D. J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J. A. and McMahon, A. P. (1993). Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell* 75, 1417–1430.

Edri, S., Hayward, P., Jawaid, W. and Martinez Arias, A. (2019). Neuro-mesodermal progenitors (NMPs): a comparative study between pluripotent stem cells and embryo-derived populations. *Development* 146,.

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.

Festuccia, N., Osorno, R., Wilson, V. and Chambers, I. (2013). The role of pluripotency gene regulatory network components in mediating transitions between pluripotent cell states. *Curr. Opin. Genet. Dev.* 23, 504–511.

Filosa, S., Rivera-Pérez, J. A., Gómez, A. P., Gansmuller, A., Sasaki, H., Behringer, R. R. and Ang, S. L. (1997). Goosecoid and HNF-3beta genetically interact to regulate neural tube patterning during mouse embryogenesis. *Development* 124, 2843–2854.

Fox, H. S., Martin, G. R., Lyon, M. F., Herrmann, B., Frischauf, A. M., Lehrach, H. and Silver, L. M. (1985). Molecular probes define different regions of the mouse t complex. *Cell* 40,

63–69.

Frank, S., Ahuja, G. and Bartsch, D. (2019). Short Article yyIncT Defines a Class of Divergently Transcribed lncRNAs and Safeguards the T-mediated Mesodermal Commitment of Human PSCs Short Article yyIncT Defines a Class of Divergently Transcribed lncRNAs and Safeguards the T-mediated Mesodermal Comm. *Stem Cell* 1–10.

Freese, N. H., Norris, D. C. and Loraine, A. E. (2016). Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* 32, 2089–2095.

Freyer, L., Hsu, C.-W., Nowotschin, S., Pauli, A., Ishida, J., Kuba, K., Fukamizu, A., Schier, A. F., Hoodless, P. A., Dickinson, M. E., et al. (2017). Loss of Apela Peptide in Mice Causes Low Penetrance Embryonic Lethality and Defects in Early Mesodermal Derivatives. *Cell Rep.* 20, 2116–2130.

Fujimoto, H. and Yanagisawa, K. (1983). Defects in the archenteron of mouse embryos homozygous for the T-mutation. *Differentiation* 25, 44–47.

Furlong, E. E. M. and Levine, M. (2018). Developmental enhancers and chromosome topology. *Science (80-.).* 361, 1341–1345.

Furushima, K., Murata, T., Matsuo, I. and Aizawa, S. (2000). A new murine zinc finger gene, Opr. *Mech. Dev.* 98, 161–164.

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.

Glinka, A., Wu, W., Onichtchouk, D., Blumenstock, C. and Niehrs, C. (1997). Head induction by simultaneous repression of Bmp and Wnt signalling in Xenopus. *Nature* 389, 517–519.

Gluecksohn-Schoenheimer, S. (1938). THE DEVELOPMENT OF TWO TAILLESS MUTANTS IN THE HOUSE MOUSE. *Genetics* 23, 573–584.

Gluecksohn-Schoenheimer, S. (1944). The Development of Normal and Homozygous Brachy (T/T) Mouse Embryos in the Extraembryonic Coelom of the Chick. *Proc. Natl. Acad. Sci. U. S. A.* 30, 134–140.

Gorkin, D. U., Barozzi, I., Zhao, Y., Zhang, Y., Huang, H., Lee, A. Y., Li, B., Chiou, J., Wildberg, A., Ding, B., et al. (2020). An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature* 583, 744–751.

Goto, H., Kimmey, S. C., Row, R. H., Matus, D. Q. and Martin, B. L. (2017). FGF and canonical Wnt signaling cooperate to induce paraxial mesoderm from tailbud neuromesodermal progenitors through regulation of a two-step epithelial to mesenchymal transition. *Development* 144, 1412–1424.

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., Stamataki, D., Kleinjung, J., Wilson, V., Briscoe, J., Bra, T., Gouti, M.,

144

Delile, J., Stamataki, D., et al. (2017). Article A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Article A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. *Dev. Cell* 41, 243-261.e7.

Guibentif, C., Griffiths, J. A., Imaz-Rosshandler, I., Ghazanfar, S., Nichols, J., Wilson, V., Göttgens, B. and Marioni, J. C. (2021). Diverse Routes toward Early Somites in the Mouse Embryo. *Dev. Cell* 56, 141-153.e6.

Habib, S. J., Chen, B.-C., Tsai, F.-C., Anastassiadis, K., Meyer, T., Betzig, E. and Nusse, R. (2013). A localized Wnt signal orients asymmetric stem cell division in vitro. *Science* 339, 1445–8.

Harland, R. (2008). Induction into the Hall of Fame: tracing the lineage of Spemann{\textquoteright}s organizer. *Development* 135, 3321–3323.

Harland, R. and Gerhart, J. (1997). Formation and function of Spemann's organizer. *Annu. Rev. Cell Dev. Biol.* 13, 611–667.

Hasenöder, S. (2016). The role of Oct4 in murine endoderm development. *Dissertation, Technische Universität München* URL:http://d-nb.info/1084212978/34.

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.

Henrique, D., Abranches, 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 Brachyury 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.

Herrmann, B. G., Koschorz, B., Wertz, K., McLaughlin, K. J. and Kispert, A. (1999). A protein kinase encoded by the t complex responder gene causes non-mendelian inheritance. *Nature* 402, 141–146.

Hirokawa, N., Tanaka, Y., Okada, Y. and Takeda, S. (2006). Nodal flow and the generation of left-right asymmetry. *Cell* 125, 33–45.

Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. and Sharp, P. A. (2017). A Phase Separation Model for Transcriptional Control. *Cell* 169, 13–23.

Holtfreter, J. F. (1988). A new look at Spemann's organizer. *Dev. Biol. (N. Y. 1985)* 5, 127–150.

Howe, E. A., Sinha, R., Schlauch, D. and Quackenbush, J. (2011). RNA-Seq analysis in MeV. *Bioinformatics* 27, 3209–3210.

Hurlin, P. J., Steingrìmsson, E., Copeland, N. G., Jenkins, N. A. and Eisenman, R. N. (1999). Mga, a dual-specificity transcription factor that interacts with Max and contains a T-domain DNA-binding motif. *EMBO J.* 18, 7019–7028.

Huxley, J. and De Beer, G. (1934). *The elements of experimental embryology,*. Cambridge [England: The University Press.

Inman, K. E. and Downs, K. M. (2006). Brachyury is required for elongation and vasculogenesis in the murine allantois. *Development.*

Jiao, S., Li, C., Hao, Q., Miao, H., Zhang, L., Li, L. and Zhou, Z. (2017). VGLL4 targets a TCF4–TEAD4 complex to coregulate Wnt and Hippo signalling in colorectal cancer. *Nat. Commun.* 8, 14058.

Jostes, B., Walther, C. and Gruss, P. (1990). The murine paired box gene, Pax7, is expressed specifically during the development of the nervous and muscular system. *Mech. Dev.* 33, 27–37.

Joubin, K. and Stern, C. D. (2001). Formation and maintenance of the organizer among the vertebrates. *Int. J. Dev. Biol.* 45, 165–175.

Jurberg, A. D., Aires, R., Varela-Lasheras, I., Nóvoa, A. and Mallo, M. (2013). Switching Axial Progenitors from Producing Trunk to Tail Tissues in Vertebrate Embryos. *Dev. Cell* 25, 451–462.

Justice, M. J. and Bode, V. C. (1988). New evidence supporting the allelism of T and tct. *Mouse News Lett.* 168–170.

Kadler, K. E., Baldock, C., Bella, J. and Boot-Handford, R. P. (2007). Collagens at a glance. *J. Cell Sci.* 120, 1955–1958.

Keller, R., Davidson, L. A. and Shook, D. R. (2003). How we are shaped : The biomechanics of gastrulation. *Differentiation* 71, 171–205.

Kelly, O. G., Pinson, K. I. and Skarnes, W. C. (2004). The Wnt co-receptors Lrp5 and Lrp6 are essential for gastrulation in mice. *Development* 131, 2803 LP – 2815.

Kennedy, M. W., Chalamalasetty, R. B., Thomas, S., Garriock, R. J., Jailwala, P. and Yamaguchi, T. P. (2016). Sp5 and Sp8 recruit β-catenin and Tcf1-Lef1 to select enhancers to activate Wnt target gene transcription. *Proc. Natl. Acad. Sci.* 113, 3545–3550.

Kim, S.-K., Suh, M. R., Yoon, H. S., Lee, J. B., Oh, S. K., Moon, S. Y., Moon, S.-H., Lee, J. Y., Hwang, J. H., Cho, W. J., et al. (2005). Identification of developmental pluripotency associated 5 expression in human pluripotent stem cells. *Stem Cells* 23, 458–462.

Kim, J. B., Greber, B., Araúzo-Bravo, M. J., Meyer, J., Park, K. I., Zaehres, H. and Schöler, H. R. (2009). Direct reprogramming of human neural stem cells by OCT4. *Nature* 461, 643–649.

Kim, T.-K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., Harmin, D. A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187.

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.

Kinder, S. J., Tsang, T. E., Wakamiya, M., Sasaki, H., Behringer, R. R., Nagy, A. and Tam, P. P. (2001). The organizer of the mouse gastrula is composed of a dynamic population of

progenitor cells for the axial mesoderm. *Development* 128, 3623–3634.

Kinney, B. A., Row, R. H., Tseng, Y.-J., Weidmann, M. D., Knaut, H. and Martin, B. L. (2020). Sox2 and canonical Wnt signaling interact to activate a developmental checkpoint coordinating morphogenesis with mesodermal fate acquisition. *bioRxiv* 2020.01.29.924050.

Kispert, A. and Herrmann, B. G. (1994). Immunohistochemical analysis of the Brachyury protein in wild-type and mutant mouse embryos. *Dev. Biol.* 161, 179–193.

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.

Klingensmith, J., Ang, S.-L., Bachiller, D. and Rossant, J. (1999). Neural Induction and Patterning in the Mouse in the Absence of the Node and Its Derivatives. *Dev. Biol.* 216, 535–549.

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.

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.*

Kume, T., Jiang, H., Topczewska, J. M. and Hogan, B. L. (2001). The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis. *Genes Dev.* 15, 2470–2482.

Kvon, E. Z. (2015). Using transgenic reporter assays to functionally characterize enhancers in animals. *Genomics* 106, 185–192.

Kvon, E. Z., Waymack, R., Elabd, M. G. and Wunderlich, Z. (2021). Enhancer redundancy in development and disease. *Nat. Rev. Genet.*

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. 10,.

Lawson, K. A., Meneses, J. J. and Pedersen, R. A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* 113, 891–911.

Lee, J. D. and Anderson, K. V (2008). Morphogenesis of the node and notochord: the cellular basis for the establishment and maintenance of left-right asymmetry in the mouse. *Dev. Dyn. an Off. Publ. Am. Assoc. Anat.* 237, 3464–3476.

Lee, H. K. and Deneen, B. (2012). Daam2 Is Required for Dorsal Patterning via Modulation of Canonical Wnt Signaling in the Developing Spinal Cord. *Dev. Cell* 22, 183–196.

Levine, M., Cattoglio, C. and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. *Cell* 157, 13–25.

Li, Y., Litingtung, Y., Ten Dijke, P. and Chiang, C. (2007). Aberrant Bmp signaling and notochord delamination in the pathogenesis of esophageal atresia. *Dev. Dyn.*

Liu, P., Wakamiya, M., Shea, M. J., Albrecht, U., Behringer, R. R. and Bradley, A. (1999). Requirement for Wnt3 in vertebrate axis formation. *Nat. Genet.* 22, 361–365.

Liu, C., Lin, C., Gao, C., May-Simera, H., Swaroop, A. and Li, T. (2014). Null and hypomorph Prickle1 alleles in mice phenocopy human Robinow syndrome and disrupt signaling downstream of Wnt5a. *Biol. Open* 3, 861–870.

Liu, C.-C., Cai, D.-L., Sun, F., Wu, Z.-H., Yue, B., Zhao, S.-L., Wu, X.-S., Zhang, M., Zhu, X.-W., Peng, Z.-H., et al. (2017). FERMT1 mediates epithelial–mesenchymal transition to promote colon cancer metastasis via modulation of β-catenin transcriptional activity. *Oncogene* 36, 1779–1792.

Lopes, R., Korkmaz, G. and Agami, R. (2016). Applying CRISPR-Cas9 tools to identify and characterize transcriptional enhancers. *Nat. Rev. Mol. Cell Biol.* 17, 597–604.

Lyon, M. F. and Meredith, R. (1964). Investigations of the nature of t-alleles in the mouse. *Heredity (Edinb).* 19, 313–325.

Lyons, G. E., Ontell, M., Cox, R., Sassoon, D. and Buckingham, M. (1990). The expression of myosin genes in developing skeletal muscle in the mouse embryo. *J. Cell Biol.* 111, 1465–1476.

Martin, B. L. and Kimelman, D. (2008). Regulation of canonical Wnt signaling by Brachyury is essential for posterior mesoderm formation. *Dev. Cell* 15, 121–133.

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.

Martinez Arias, A. and Steventon, B. (2018). On the nature and function of organizers. *Development.*

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.*

Minegishi, K., Hashimoto, M., Ajima, R., Takaoka, K., Shinohara, K., Ikawa, Y., Nishimura, H., McMahon, A. P., Willert, K., Okada, Y., et al. (2017). A Wnt5 Activity Asymmetry and Intercellular Signaling via PCP Proteins Polarize Node Cells for Left-Right Symmetry Breaking. *Dev. Cell* 40, 439-452.e4.

Minegishi, K., Rothé, B., Komatsu, K. R., Ono, H., Ikawa, Y., Nishimura, H., Miyashita, E., Takaoka, K., Bando, K., Kiyonari, H., et al. (2020). Fluid flow-induced left-right asymmetric decay of Dand5 mRNA in the mouse embryo requires Bicc1-Ccr4 RNA degradation complex. *bioRxiv.*

Mishina, Y. (2003). Function of bone morphogenetic protein signaling during mouse development. *Front. Biosci.* 8, d855-69.

Morita, H., Mazerbourg, S., Bouley, D. M., Luo, C.-W., Kawamura, K., Kuwabara, Y., Baribault, H., Tian, H. and Hsueh, A. J. W. (2004). Neonatal lethality of LGR5 null mice is associated with ankyloglossia and gastrointestinal distension. *Mol. Cell. Biol.* 24, 9736–

148

9743.

Morris, S. A., Grewal, S., Barrios, F., Patankar, S. N., Strauss, B., Buttery, L., Alexander, M., Shakesheff, K. M. and Zernicka-goetz, M. (2012). in the developing mouse embryo. *Nat. Commun.* 1–10.

Muller, C. W. and Herrmann, B. G. (1997). Crystallographic structure of the T domain-DNA complex of the Brachyury transcription factor. *Nature* 389, 884–888.

Murtaugh, L. C., Chyung, J. H. and Lassar, A. B. (1999). Sonic hedgehog promotes somitic chondrogenesis by altering the cellular response to BMP signaling. *Genes Dev.* 13, 225–237.

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.

Nagy, A., Gertsenstein, M., Vintersten, K., and Behringer, R. R. (2003). *Manipulating the Mouse Embryo : A Laboratory Manual.* Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory.

Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. and Roder, J. C. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 90, 8424–8428.

Naiche, L. A., Holder, N. and Lewandoski, M. (2011). FGF4 and FGF8 comprise the wavefront activity that controls somitogenesis. *Proc. Natl. Acad. Sci.* 108, 4018 LP – 4023.

Nakaya, M.-A., Gudmundsson, K. O., Komiya, Y., Keller, J. R., Habas, R., Yamaguchi, T. P. and Ajima, R. (2020). Placental defects lead to embryonic lethality in mice lacking the Formin and PCP proteins Daam1 and Daam2. *PLoS One* 15, e0232025–e0232025.

Naruse-Nakajima, C., Asano, M. and Iwakura, Y. (2001). Involvement of EphA2 in the formation of the tail notochord via interaction with ephrinA1. *Mech. Dev.* 102, 95–105.

Niehrs, C. (1999). Head in the WNT: the molecular nature of Spemann's head organizer. *Trends Genet.* 15, 314–319.

Nieuwkoop, P. D. (1952). Activation and organization of the central nervous system in amphibians. Part III. Synthesis of a new working hypothesis. *J. Exp. Zool.* 120, 83–108.

Nonaka, S., Shiratori, H., Saijoh, Y. and Hamada, H. (2002). Determination of left-right patterning of the mouse embryo by artificial nodal flow. *Nature* 418, 96–99.

Nowotschin, S., Setty, M., Kuo, Y.-Y., Liu, V., Garg, V., Sharma, R., Simon, C. S., Saiz, N., Gardner, R., Boutet, S. C., et al. (2019). The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature.*

Okuda, A., Fukushima, A., Nishimoto, M., Orimo, A., Yamagishi, T., Nabeshima, Y., Kuro-o, M., Nabeshima, Y. i, Boon, K., Keaveney, M., et al. (1998). UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. *EMBO J.* 17, 2019–2032.

Orkin, S. H. (2005). Chipping away at the embryonic stem cell network. *Cell* 122, 828–830.

Osorno, R. and Chambers, I. (2011). Transcription factor heterogeneity and epiblast

pluripotency. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 366, 2230–2237.

Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B. J., Afzal, S. Y., Lee, E. A., Zhu, Y., Plajzer-Frick, I., Pickle, C. S., et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* 554, 239–243.

Ott, M. O., Bober, E., Lyons, G., Arnold, H. and Buckingham, M. (1991). Early expression of the myogenic regulatory gene, myf-5, in precursor cells of skeletal muscle in the mouse embryo. *Development* 111, 1097–1107.

Özhan, G., Sezgin, E., Wehner, D., Pfister, A. S., Kühl, S. J., Kagermeier-Schenk, B., Kühl, M., Schwille, P. and Weidinger, G. (2013). Lypd6 Enhances Wnt/β-Catenin Signaling by Promoting Lrp6 Phosphorylation in Raft Plasma Membrane Domains. *Dev. Cell* 26, 331–345.

Pennimpede, T., Proske, J., König, A., Vidigal, J. a., Morkel, M., Bramsen, J. B., Herrmann, B. G. and Wittler, L. (2012). In vivo knockdown of Brachyury results in skeletal defects and urorectal malformations resembling caudal regression syndrome. *Dev. Biol.* 372, 55–67.

Perea-Gomez, A., Rhinn, M. and Ang, S. L. (2001). Role of the anterior visceral endoderm in restricting posterior signals in the mouse embryo. *Int. J. Dev. Biol.* 45, 311–320.

Perea-Gomez, A., Vella, F. D. J., Shawlot, W., Oulad-Abdelghani, M., Chazaud, C., Meno, C., Pfister, V., Chen, L., Robertson, E., Hamada, H., et al. (2002). Nodal Antagonists in the Anterior Visceral Endoderm Prevent the Formation of Multiple Primitive Streaks. *Dev. Cell* 3, 745–756.

Peters, H., Wilm, B., Sakai, N., Imai, K., Maas, R. and Balling, R. (1999). Pax1 and Pax9 synergistically regulate vertebral column development. *Development* 126, 5399–5408.

Pillay, N., Plagnol, V., Tarpey, P. S., Lobo, S. B., Presneau, N., Szuhai, K., Halai, D., Berisha, F., Cannon, S. R., Mead, S., et al. (2012). A common single-nucleotide variant in T is strongly associated with chordoma. *Nat. Genet.* 44, 1185–1187.

Pinson, K. I., Brennan, J., Monkley, S., Avery, B. J. and Skarnes, W. C. (2000). An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* 407, 535–538.

Placzek, M. and Briscoe, J. (2018). Sonic hedgehog in vertebrate neural tube development. *Int. J. Dev. Biol.* 62, 225–234.

Plouhinec, J.-L., Granier, C., Le Mentec, C., Lawson, K. A., Sabéran-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.

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.

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.

Rennebeck, G. M., Lader, E., Chen, Q., Bohm, R. A., Cai, Z. S., Faust, C., Magnuson, T., Pease, L. R. and Artzt, K. (1995). Is there a Brachyury the Second? Analysis of a transgenic mutation involved in notochord maintenance in mice. *Dev. Biol.* 172, 206–217.

Rennebeck, G., Lader, E., Fujimoto, A., Lei, E. P. and Artzt, K. (1998). Mouse Brachyury the Second (T2) is a gene next to classical T and a candidate gene for tct. *Genetics* 150, 1125–1131.

Robinton, D. A., Chal, J., Lummertz da Rocha, E., Han, A., Yermalovich, A. V, Oginuma, M., Schlaeger, T. M., Sousa, P., Rodriguez, A., Urbach, A., et al. (2019). The Lin28/let-7 Pathway Regulates the Mammalian Caudal Body Axis Elongation Program. *Dev. Cell* 48, 396-405.e3.

Rodrigo Albors, A., Halley, P. A. and Storey, K. G. (2018). Lineage tracing of axial progenitors using Nkx1-2CreERT2 mice defines their trunk and tail contributions. *Development* 145, dev164319.

Roebroek, A. J., Umans, L., Pauli, I. G., Robertson, E. J., van Leuven, F., Van de Ven, W. J. and Constam, D. B. (1998). Failure of ventral closure and axial rotation in embryos lacking the proprotein convertase Furin. *Development* 125, 4863–4876.

Romanos, M., Allio, G., Combres, L., Médevielle, F., Escalas, N., Soula, C., Steventon, B., Trescases, A. and Bénazéraf, B. (2020). Cell-to-cell heterogeneity in Sox2 and Brachyury expression guides progenitor destiny by controlling their movements. *bioRxiv.*

Saga, Y., Hata, N., Koseki, H. and Taketo, M. M. (1997). Mesp2: a novel mouse gene expressed in the presegmented mesoderm and essential for segmentation initiation. *Genes Dev.* 11, 1827–1839.

Sambasivan, R. and Steventon, B. (2020). Neuromesodermal Progenitors: A Basis for Robust Axial Patterning in Development and Evolution. *Front. cell Dev. Biol.* 8, 607516.

Sanford, L. P., Ormsby, I., Gittenberger-de Groot, A. C., Sariola, H., Friedman, R., Boivin, G. P., Cardell, E. L. and Doetschman, T. (1997). TGFbeta2 knockout mice have multiple developmental defects that are non-overlapping with other TGFbeta knockout phenotypes. *Development* 124, 2659–2670.

Sato, M., Kimura, T., Kurokawa, K., Fujita, Y., Abe, K., Masuhara, M., Yasunaga, T., Ryo, A., Yamamoto, M. and Nakano, T. (2002). Identification of PGC7, a new gene expressed specifically in preimplantation embryos and germ cells. *Mech. Dev.* 113, 91–94.

Savory, J. G. A., Bouchard, N., Pierre, V., Rijli, F. M., De Repentigny, Y., Kothary, R. and Lohnes, D. (2009). Cdx2 regulation of posterior development through non-Hox targets. *Development* 136, 4099–4110.

Schimenti, J. (2000). Segregation distortion of mouse t haplotypes the molecular basis emerges. *Trends Genet.* 16, 240–243.

Schmidt, C., Wilson, V., Stott, D. and Beddington, R. S. P. (1997). T Promoter Activity in the Absence of Functional T Protein during Axis Formation and Elongation in the Mouse. *Dev. Biol.* 189, 161–173.

Schmidt, C., Stoeckelhuber, M., McKinnell, I., Putz, R., Christ, B. and Patel, K. (2004). Wnt 6 regulates the epithelialisation process of the segmental plate mesoderm leading to somite formation. *Dev. Biol.* 271, 198–209.

Schmitt, M., Metzger, M., Gradl, D., Davidson, G. and Orian-Rousseau, V. (2015). CD44 functions in Wnt signaling by regulating LRP6 localization and activation. *Cell Death*

Differ. 22, 677–689.

Selleck, M. A. and Stern, C. D. (1991). Fate mapping and cell lineage analysis of Hensen's node in the chick embryo. *Development* 112, 615–626.

Shah, S. R., David, J. M., Tippens, N. D., Mohyeldin, A., Martinez-Gutierrez, J. C., Ganaha, S., Schiapparelli, P., Hamilton, D. H., Palena, C., Levchenko, A., et al. (2017). Brachyury-YAP Regulatory Axis Drives Stemness and Growth in Cancer. *Cell Rep.* 21, 495–507.

Sharifnia, T., Wawer, M. J., Chen, T., Huang, Q.-Y., Weir, B. A., Sizemore, A., Lawlor, M. A., Goodale, A., Cowley, G. S., Vazquez, F., et al. (2019). Small-molecule targeting of brachyury transcription factor addiction in chordoma. *Nat. Med.* 25, 292–300.

Sheppard, H. E., Dall'Agnese, A., Park, W. D., Shamim, M. H., Dubrulle, J., Johnson, H. L., Stossi, F., Cogswell, P., Sommer, J., Levy, J., et al. (2021). Targeted brachyury degradation disrupts a highly specific autoregulatory program controlling chordoma cell identity. *Cell Reports Med.* 2, 100188.

Shlyueva, D., Stampfel, G. and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* 15, 272–286.

Shull, M. M., Ormsby, I., Kier, A. B., Pawlowski, S., Diebold, R. J., Yin, M., Allen, R., Sidman, C., Proetzel, G., Calvin, D., et al. (1992). Targeted disruption of the mouse transforming growth factor-β1 gene results in multifocal inflammatory disease. *Nature* 359, 693–699.

Silver, L. M. (1985). Mouse t haplotypes. *Annu. Rev. Genet.* 19, 179–208.

Silver, L. M. (1993). The peculiar journey of a selfish chromosome: mouse t haplotypes and meiotic drive. *Trends Genet.* 9, 250–254.

Simandi, Z., Horvath, A., Wright, L. C., Cuaranta-Monroy, I., De Luca, I., Karolyi, K., Sauer, S., Deleuze, J.-F., Gudas, L. J., Cowley, S. M., et al. (2016). OCT4 Acts as an Integrator of Pluripotency and Signal-Induced Differentiation. *Mol. Cell.*

Sinner, D., Kordich, J. J., Spence, J. R., Opoka, R., Rankin, S., Lin, S.-C. J., Jonatan, D., Zorn, A. M. and Wells, J. M. (2007). Sox17 and Sox4 Differentially Regulate β-Catenin/T-Cell Factor Activity and Proliferation of Colon Carcinoma Cells. *Mol. Cell. Biol.* 27, 7802–7815.

Smith, J. C. and Slack, J. M. (1983). Dorsalization and neural induction: properties of the organizer in Xenopus laevis. *J. Embryol. Exp. Morphol.* 78, 299–317.

Solnica-Krezel, L. and Sepich, D. S. (2012). Gastrulation: Making and Shaping Germ Layers. *Annu. Rev. Cell Dev. Biol.* 28, 687–717.

Solovieva, T., Lu, H.-C., Moverley, A., Plachta, N. and Stern, C. D. (2020). The embryonic node functions as an instructive stem cell niche. *bioRxiv.*

Song, H., Hu, J., Chen, W., Elliott, G., Andre, P., Gao, B. and Yang, Y. (2010). Planar cell polarity breaks bilateral symmetry by controlling ciliary positioning. *Nature* 466, 378–382.

Sonnen, K. F., Lauschke, V. M., Uraji, J., Falk, H. J., Petersen, Y., Funk, M. C., Beaupeux, M., François, P., Merten, C. A. and Aulehla, A. (2018). Modulation of Phase Shift between Wnt and Notch Signaling Oscillations Controls Mesoderm Segmentation. *Cell* 172, 1079-1090.e12.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat. Genet.* 21, 70–71.

Spemann, H. and Mangold, H. (1924). über Induktion von Embryonalanlagen durch Implantation artfremder Organisatoren. *Arch. für mikroskopische Anat. und Entwicklungsmechanik* 101, 458.

Spitz, F. and Furlong, E. E. M. (2012). Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* 13, 613–626.

Stemple, D. L. (2005). Structure and function of the notochord: an essential organ for chordate development. *Development* 132, 2503–2512.

Stern, C. D. (2001). Initial patterning of the central nervous system: how many organizers? *Nat. Rev. Neurosci.* 2, 92–98.

Stielow, B., Finkernagel, F., Stiewe, T., Nist, A. and Suske, G. (2018). MGA, L3MBTL2 and E2F6 determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *PLoS Genet.* 14, e1007193.

Stott, D., Kispert, L. a. and Herrmann, B. G. (1993). Rescue of the tail defect of Brachyury mice. *Genes Dev.* 7, 197–203.

Sulik, K., Dehart, D. B., Iangaki, T., Carson, J. L., Vrablic, T., Gesteland, K. and Schoenwolf, G. C. (1994). Morphogenesis of the murine node and notochordal plate. *Dev. Dyn. an Off. Publ. Am. Assoc. Anat.* 201, 260–278.

Sun, X., Meyers, E. N., Lewandoski, M. and Martin, G. R. (1999). Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. *Genes Dev.* 13, 1834–1846.

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.

Takaoka, K., Yamamoto, M. and Hamada, H. (2011). Origin and role of distal visceral endoderm, a group of cells that determines anterior-posterior polarity of the mouse embryo. *Nat. Cell Biol.* 13, 743–752.

Takaoka, K., Nishimura, H. and Hamada, H. (2017). Both Nodal signalling and stochasticity select for prospective distal visceral endoderm in mouse embryos. *Nat. Commun.* 8, 1492.

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.

Tam, P. P. (1981). The control of somitogenesis in mouse embryos. *J. Embryol. Exp. Morphol.* 65 Suppl, 103–128.

Tam, P. P. L. and Loebel, D. A. F. (2007). Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.* 8, 368–381.

Tam, P. P. and Steiner, K. A. (1999). Anterior patterning by synergistic activity of the early gastrula organizer and the anterior germ layer tissues of the mouse embryo. *Development* 126, 5171–5179.

Tam, P. P. and Tan, S. S. (1992). The somitogenetic potential of cells in the primitive streak and the tail bud of the organogenesis-stage mouse embryo. *Development* 115, 703–715.

Tam, P. P., Steiner, K. A., Zhou, S. X. and Quinlan, G. A. (1997). Lineage and functional analyses of the mouse organizer. *Cold Spring Harb. Symp. Quant. Biol.* 62, 135–144.

Tamplin, O. J., Cox, B. J. and Rossant, J. (2011). Integrated microarray and ChIP analysis identi fi es multiple Foxa2 dependent target genes in the notochord. *Dev. Biol.* 360, 415–425.

Tanaka, M., Lyons, G. E. and Izumo, S. (1999). Expression of the Nkx3.1 homobox gene during pre and postnatal development. *Mech. Dev.* 85, 179–182.

Tarpey, P. S., Behjati, S., Young, M. D., Martincorena, I., Alexandrov, L. B., Farndon, S. J., Guzzo, C., Hardy, C., Latimer, C., Butler, A. P., et al. (2017). The driver landscape of sporadic chordoma. *Nat. Commun.* 8, 890.

Tirabosco, R., Mangham, D. C., Rosenberg, A. E., Vujovic, S., Bousdras, K., Pizzolitto, S., De Maglio, G., den Bakker, M. A., Di Francesco, L., Kalil, R. K., et al. (2008). Brachyury expression in extra-axial skeletal and soft tissue chordomas: a marker that distinguishes chordoma from mixed tumor/myoepithelioma/parachordoma in soft tissue. *Am. J. Surg. Pathol.* 32, 572–580.

Torres-Padilla, M.-E., Parfitt, D.-E., Kouzarides, T. and Zernicka-Goetz, M. (2007). Histone arginine methylation regulates pluripotency in the early mouse embryo. *Nature* 445, 214–218.

Tosic, J., Kim, G.-J., Pavlovic, M., Schroder, C. M., Mersiowsky, S.-L., Barg, M., Hofherr, A., Probst, S., Kottgen, M., Hein, L., et al. (2019). Eomes and Brachyury control pluripotency exit and germ-layer segregation by changing the chromatin state. *Nat. Cell Biol.* 21, 1518–1531.

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L. and Pachter, L. (2013). Articles Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31,.

Tsakiridis, A. and Wilson, V. (2015). Assessing the bipotency of in vitro-derived neuromesodermal progenitors. *F1000Research* 4, 100.

Turner, D. A., Hayward, P. C., Baillie-Johnson, P., Rué, P., Broome, R., Faunes, F. and Martinez Arias, A. (2014). Wnt/β-catenin and FGF signalling direct the specification and maintenance of a neuromesodermal axial progenitor in ensembles of mouse embryonic stem cells. *Development* 141, 4243–4253.

Tzouanacou, E., Wegener, A., Wymeersch, 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.*

van den Brink, S. C., Baillie-Johnson, P., Balayo, T., Hadjantonakis, A.-K., Nowotschin, S., Turner, D. A. and Martinez Arias, A. (2014). Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Development* 141,

4231–4242.

van den Brink, S. C., Alemany, A., van Batenburg, V., Moris, N., Blotenburg, M., Vivié, J., Baillie-Johnson, P., Nichols, J., Sonnen, K. F., Martinez Arias, A., et al. (2020). Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature* 582, 405–409.

Veenvliet, J. V, Bolondi, A., Kretzmer, H., Haut, L., Scholze-Wittler, M., Schifferl, D., Koch, F., Guignard, L., Kumar, A. S., Pustet, M., et al. (2020). Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. *Science (80-.).* 370,.

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.*

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.

Vujovic, S., Henderson, S., Presneau, N., Odell, E., Jacques, T. S., Tirabosco, R., Boshoff, C. and Flanagan, A. M. (2006). Brachyury, a crucial regulator of notochordal development, is a novel biomarker for chordomas. *J. Pathol.* 209, 157–165.

Wang, Y., Chang, H., Rattner, A. and Nathans, J. (2016). Frizzled Receptors in Development and Disease. *Curr. Top. Dev. Biol.* 117, 113–139.

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.

Weinstein, D. C., Ruiz i Altaba, A., Chen, W. S., Hoodless, P., Prezioso, V. R., Jessell, T. M. and Darnell, J. E. (1994). The winged-helix transcription factor HNF-3β is required for notochord development in the mouse embryo. *Cell* 78, 575–588.

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I. and Young, R. A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell* 153, 307–319.

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. and Beddington, R. (1997). Expression of T Protein in the Primitive Streak Is Necessary and Sufficient for Posterior Mesoderm Movement and Somite Differentiation. *Dev. Biol.* 192, 45–58.

Wilson, V., Olivera-Martinez, I. and Storey, K. G. (2009). Stem cells, signals and vertebrate body axis extension. *Development* 136, 1591–1604.

Wood, H. B. and Episkopou, V. (1999). Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. *Mech. Dev.* 86, 197–201.

Wopat, S., Bagwell, J., Sumigray, K. D., Poss, K. D., Schulte-merker, S., Bagnat, M., Wopat, S., Bagwell, J., Sumigray, K. D., Dickson, A. L., et al. (2018). Spine Patterning Is Guided by Segmentation of the Notochord Sheath Spine Patterning Is Guided by Segmentation of the Notochord Sheath. *CellReports* 22, 2026–2038.

Wu, J. I., Centilli, M. a., Vasquez, G., Young, S., Scolnick, J., Durfee, L. a., Spearow, J. L.,

Schwantz, S. D., Rennebeck, G. and Artzt, K. (2007). tint maps to mouse chromosome 6 and may interact with a notochordal enhancer of Brachyury. *Genetics* 177, 1151–1161.

Wu, M., Chen, G. and Li, Y.-P. (2016). TGF-β and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. *Bone Res.* 4, 16009.

Wymeersch, F. J., Huang, Y., Blin, G., Cambray, N., Wilkie, R., Wong, F. C. K. and Wilson, V. (2016). Position-dependent plasticity of distinct progenitor types in the primitive streak. *Elife* 5, e10042–e10042.

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,.

Wymeersch, F. J., Wilson, V. and Tsakiridis, A. (2021). Understanding axial progenitor biology in vivo and in vitro. *Development* 148,.

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.

Yamamoto, S., Nishimura, O., Misaki, K., Nishita, M., Minami, Y., Yonemura, S., Tarui, H. and Sasaki, H. (2008). Cthrc1 Selectively Activates the Planar Cell Polarity Pathway of Wnt Signaling by Stabilizing the Wnt-Receptor Complex. *Dev. Cell* 15, 23–36.

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.*

Yang, X., Dormann, D., Mu, A. E. and Weijer, C. J. (2002). Cell Movement Patterns during Gastrulation in the Chick Are Controlled by Positive and Negative Chemotaxis Mediated by FGF4 and FGF8. 3, 425–437.

Yang, X. R., Ng, D., Alcorta, D. A., Liebsch, N. J., Sheridan, E., Li, S., Goldstein, A. M., Parry, D. M. and Kelley, M. J. (2009). T (brachyury) gene duplication confers major susceptibility to familial chordoma. *Nat. Genet.* 41, 1176–1178.

Ybot-Gonzalez, P., Savery, D., Gerrelli, D., Signore, M., Mitchell, C. E., Faux, C. H., Greene, N. D. E. and Copp, A. J. (2007). Convergent extension, planar-cell-polarity signalling and initiation of mouse neural tube closure. *Development* 134, 789–799.

Yoon, J. K., Moon, R. T. and Wold, B. (2000). The bHLH class protein pMesogenin1 can specify paraxial mesoderm phenotypes. *Dev. Biol.* 222, 376–391.

Young, T., Rowland, J. E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M., van Nes, J., de Graaff, W., Duluc, I., Freund, J.-N., et al. (2009). Cdx and Hox Genes Differentially Regulate Posterior Axial Growth in Mammalian Embryos. *Dev. Cell* 17, 516–526.

Zalc, A., Sinha, R., Gulati, G. S., Wesche, D. J., Daszczuk, P., Swigut, T., Weissman, I. L. and Wysocka, J. (2021). Reactivation of the pluripotency program precedes formation of the cranial neural crest. *Science (80-.).* 371,.

Zernicka-Goetz, M. (2002). Patterning of the embryo: the first spatial decisions in the life of a mouse. *Development* 129, 815–829.

Zernicka-Goetz, M., Morris, S. a and Bruce, A. W. (2009). Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nat. Rev. Genet.* 10, 467–477.

Zhang, S. and Cui, W. (2014). Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World J. Stem Cells* 6, 305–311.

Zhou, X., Sasaki, H., Lowe, L., Hogan, B. L. and Kuehn, M. R. (1993). Nodal is a novel TGF-beta-like gene expressed in the mouse node during gastrulation. *Nature* 361, 543–547.

Zhu, J., Kwan, K. M. and Mackem, S. (2016). Putative oncogene Brachyury (T) is essential to specify cell fate but dispensable for notochord progenitor proliferation and EMT. *Proc. Natl. Acad. Sci. U. S. A.* 113, 3820–3825.

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.

Zoltewicz, J. S. and Gerhart, J. C. (1997). The Spemann organizer of Xenopus is patterned along its anteroposterior axis at the earliest gastrula stage. *Dev. Biol.* 192, 482–491.

List of Abbreviations

2D	Two dimensional
3D	Three dimensional
AP	Anterior-posterior
AVE	Anterior visceral endoderm
BAC	Bacterial artificial chromosome
BL6	Black 6
BMP	Bone morphogenic protein
bp	base pair
cDNA	Complementary deoxyribonucleic acid
CDS	Coding sequence
ChIP seq	Chromatin immunoprecipitation DNA sequencing
CNH	Chordo-neural hinge
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
d	day
d.p.c.	Days post coitum
DAPI	4′,6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
DPBS	Dulbecco's phosphate-buffered saline
DV	Dorsoventral
DVE	Distal visceral endoderm
E	Embryonic day
EGO	Early gastrula organizer
EHF	Early head fold
EMT	Epithelial to mesenchymal transition
et al.	et alia
ExE	Extraembryonic ectoderm
FACS	Fluorescence activated cell sorting
FBS	Fetal Bovine Serum
FGF	Fibroblast growth factor
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
gRNA	guide RNA
HF	head fold
H2B	Histone2B
HSP	Heat shock promoter
i.e.	id est
IRES	Internal ribosomal entry site
kb	kilo base pairs
LB	lysogeny broth
LGO	Late gastrula organizer
LHF	Late head fold
LR	Left-right
mEF	mouse embyronic fibroblast
mESC	Mouse embryonic stem cell
MGI	Mouse Genome Informatic
MGO	Mid gastrula organizer

ML	Medio-lateral
MPIMG	Max-Planck-Institute for Molecular Genetics
NE	neuroectoderm
NMP	Neuromesodermal progenitor
NSB	Node-streak border
ORF	Open reading frame
pA	poly-adenylation
PAM	Protospacer adjacent motif
PCP	Planar cell polarity
PCR	Polymerase Chain Reaction
PD	Proximal-Distal
PFA	paraformaldehyde
PG	Paralogous group
PGC	Primordial germ cell
PNC	Posterior notochord
PS	Primitive streak
PSM	pre-somitic mesoderm
Puro	Puromycin
RIMS	refractive index matching solution
RIPA	Radioimmunoprecipitation assay buffer
RMCE	Recombinase mediated cassette exchange
RNA-seq	Ribo nucleic acid sequencing
rpm	rotations per minute
SFE	surface ectoderm
SNP	Single nucleotide polymorphism
TAD	Topologically associated domain
tct	*t complex tail interaction (factor)*
TE	T enhancer
TF	Transcription factor
TSS	Transcription start site
vs.	versus
WISH	Whole mount in situ hybridization
Wnt	Wingless Int1
WT	wild type

List of Figures and Tables

Main Figures

Supplementary Figures

Tables

Supplementary Tables