A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development

Graphical Abstract

Reverse engineered NeuroMesodermal Progenitor GRN from in vivo and in vitro single-cell RNA-seq

Reconstructed cell clusters from scRNA-seq

Cell state graph

Predicted GRN

In Brief

Neuromesodermal progenitors (NMPs) generate cells of the spinal cord and somites. Gouti et al. demonstrate that in vitro NMPs resemble in vivo counterparts at the single-cell level and define a regulatory network that balances the generation of neural and mesodermal tissue to facilitate orderly extension of the embryonic axis.

Highlights

- Single-cell RNA-seq reveals a signature of neuromesodermal progenitors
- In vitro NMPs resemble and differentiate similar to their in vivo counterparts
- Dual role for retinoic acid signaling in NMP induction and neural differentiation
- A transcriptional network regulates neural versus mesodermal allocation

Authors

Mina Gouti, Julien Delile, Despina Stamataki, ..., Jens Kleinjung, Valerie Wilson, James Briscoe

Correspondence

mina.gouti@mdc-berlin.de (M.G.), james.briscoe@crick.ac.uk (J.B.)

In Brief

Neuromesodermal progenitors (NMPs) generate cells of the spinal cord and somites. Gouti et al. demonstrate that in vitro NMPs resemble in vivo counterparts at the single-cell level and define a regulatory network that balances the generation of neural and mesodermal tissue to facilitate orderly extension of the embryonic axis.
A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development

Mina Gouti,1,3,4,* Julien Delile,1,4 Despina Stamataki,1 Filip J. Wymeersch,2 Yali Huang,2 Jens Kleinjung,1 Valerie Wilson,2 and James Briscoe1,5,*

1The Francis Crick Institute, 1 Midland Road, London, NW1 1AT, UK
2MRC Centre for Regenerative Medicine, Institute for Stem Cell Research, School of Biological Sciences, University of Edinburgh, 5 Little France Drive, Edinburgh EH16 4UU, UK
3Max Delbrück Center for Molecular Medicine, Robert-Rössle-Strasse 10, 13125 Berlin, Germany
4These authors contributed equally
5Lead Contact
*Correspondence: mina.gouti@mdc-berlin.de (M.G.), james.briscoe@crick.ac.uk (J.B.)
http://dx.doi.org/10.1016/j.devcel.2017.04.002

SUMMARY

Transcriptional networks, regulated by extracellular signals, control cell fate decisions and determine the size and composition of developing tissues. One example is the network controlling bipotent neuromesodermal progenitors (NMPs) that fuel embryo elongation by generating spinal cord and trunk mesoderm tissue. Here, we use single-cell transcriptomics to identify the molecular signature of NMPs and reverse engineer the mechanism that regulates their differentiation. Together with genetic perturbations, this reveals a transcriptional network that integrates opposing retinoic acid (RA) and Wnt signals to determine the rate at which cells enter and exit the NMP state. RA, produced by newly generated mesodermal cells, provides feedback that initiates NMP generation and induces neural differentiation, thereby coordinating the production of neural and mesodermal tissue. Together, the data define a regulatory network architecture that balances the generation of different cell types from bipotential progenitors in order to facilitate orderly axis elongation.

INTRODUCTION

Cell fate decisions in developing tissues are made by gene regulatory networks comprising transcription factors and intercellular signals. These networks determine the rate of self-renewal and differentiation to ensure the balanced generation of different cell types and the production of well-proportioned tissues (Stern et al., 2006; Davidson, 2010). The formation of the vertebrate trunk, which extends progressively during embryogenesis, is one example. Successively more posterior neural and paraxial presomatic mesodermal (PSM) cells of the trunk are generated from a bipotential population of cells (Tzouanacou et al., 2009), termed neuromesodermal progenitors (NMPs) at the posterior end of the embryo. Proliferation of NMPs fuels the elongation of axial tissues (Cambray and Wilson, 2002; 2007; Wilson et al., 2009; Henrique et al., 2015; Neijts et al., 2014; Kimelman, 2016). Hence, the rate at which NMPs are generated, self-renew, and differentiate must be carefully regulated in order to balance the production of different trunk tissues and to prevent the premature or delayed depletion of NMPs that will affect the length of the embryo.

NMPs reside in the node-streak border (NSB), caudal lateral epiblast (CLE) and the chordonuear hinge (CNH) of elongating embryos (Cambray and Wilson, 2002; 2007; Delfino-Machin et al., 2005; Kondoh and Takemoto, 2012; Olivera-Martinez et al., 2012). These regions express Wnt and FGF ligands (Wilson et al., 2009). Interfering with either signal results in the depletion of NMPs and the premature truncation of the body axis (Takada et al., 1994; Wilson et al., 2009; Yoshikawa et al., 1997; van de Ven et al., 2011). Both Wnt and FGF signaling are implicated in posteriorizing cells by inducing the Cdx transcription factors (TFs) that promote the expression of more posterior Hox genes (van den Akker et al., 2002; Nordstrom et al., 2006; van de Ven et al., 2011; Mazzoni et al., 2013; Neijts et al., 2016; Amin et al., 2016). Moreover, Wnt signaling is required for the differentiation of NMPs to mesodermal tissue (Martin and Kimelman, 2012; Garriock et al., 2015) and the loss of Wnt3a results in a depletion of mesodermal tissue in both mice and zebrafish (Yoshikawa et al., 1997; Martin and Kimelman, 2008; Garriock et al., 2015).

Primitive streak and node cells transiently express the retinoic acid (RA)-synthesizing enzyme Aldh1a2 (Ribes et al., 2009). Mouse embryos lacking Aldh1a2 are truncated, suggesting a role for RA in axis elongation (Niederreither et al., 1999; Cunningham et al., 2015; Niederreither and Dole, 2008; Duester, 2008) nevertheless, the role of RA in the establishment of NMPs remains unclear. At later stages of development, RA emanating from Aldh1a2-expressing somitic cells promotes the expression of genes characteristic of neural progenitors in the spinal cord. In addition, RA inhibits expression of both Wnt3a and Fgf8, and exposure of the tail region to increased concentrations of RA can arrest axis elongation (Olivera-Martinez et al., 2012). Excess RA concentration in the tail bud is prevented by the RA-metabolizing enzyme Cyp26a1, which is induced by Cdx genes and
T/Bra under the control of Wnt/Fgf signaling (Martin and Kimelman, 2010; Vidigal et al., 2010; Savory et al., 2009; Young et al., 2009; van Rooijen et al., 2012). However, the overlapping functions and proximity of events hindered assigning direct and indirect activities to individual signaling pathways (Kimelman, 2016; Henrique et al., 2015; Gouti et al., 2015; Neijts et al., 2014). The co-expression of the TFs T/Bra and Sox2 is characteristic of NMPs (Olvera-Martinez et al., 2012; Gouti et al., 2014; Tsakiridis et al., 2015; Wymeersch et al., 2016). Both Wnt and FGF signaling have been implicated as inducers of T/Bra in NMPs (Yamaguchi et al., 1999; Martin and Kimelman, 2008). Moreover, Cdx-binding regions have been found upstream of the T/Bra gene (Savory et al., 2009). This and subsequent analysis (van Rooijen et al., 2012) has led to the suggestion that Cdx proteins, induced by Wnt signaling, maintain T/Bra expression in NMPs, but are dispensable for its initial induction. However, Cdx proteins also appear to regulate Wnt and FGF expression in NMPs (Young et al., 2009; Savory et al., 2009; van Rooijen et al., 2012), thus the loss of T/Bra expression in the absence of Cdx might be due to the loss of these signals and the depletion of NMPs.

Neural cells differentiating from NMPs downregulate T/Bra but maintain Sox2 expression (Gouti et al., 2014, 2015; Tsakiridis and Wilson, 2015; Gouti et al., 2015). By contrast, as NMPs differentiate into mesoderm, expression of Sox2 is downregulated and Msgn1 and Tbx6 are upregulated to form nascent mesodermal progenitor cells (MPCs) (Chalamalasetty et al., 2011). Then, as cells commit to a PSM identity, expression of T/Bra is downregulated. In embryos lacking T/Bra, mesoderm induction fails and axis elongation halts (Herrmann et al., 1990). This is, at least in part, explained by a requirement for T/Bra for the induction of Msgn1 and Tbx6 (Yamaguchi et al., 1999; Yabe and Takada, 2012). Moreover, the loss of PSM tissue in the absence of Tbx6 or Msgn1 is accompanied by ectopic generation of neural tissue (Champion and Papaoannou, 1998; Yoon et al., 2000; Chalamalasetty et al., 2014), raising the question of the role that the induction of these TFs plays in balancing neural and mesodermal production from NMPs.

Together, the data suggest complex regulatory mechanisms with multiple interactions and feedbacks. It has proven challenging, however, to assemble a definitive network that explains the generation of NMPs and their balanced allocation toward mesodermal and neural tissue. These difficulties arise from the necessity of analyzing in vivo experimental perturbations in which axis elongation fails or the expression of signals is lost. To circumvent this, we have taken advantage of the in vitro directed differentiation that we and others have recently developed to generate NMPs from pluripotent stem cells (Gouti et al., 2014; Tsakiridis et al., 2015; Turner et al., 2014; Lippmann et al., 2015). This system decouples the development of NMPs and trunk cell types from the specific tissue architecture associated with axis elongation, thereby avoiding the difficulty of interpreting data from chimeric or morphologically abnormal embryos. Furthermore, it allows exogenous control of the supply and timing of signaling molecules. Thus, aspects of the gene regulatory network that are tightly linked in vivo can be separated and assayed in vitro.

Using single-cell transcriptome analysis we first established the similarity between in vivo and in vitro derived NMPs. We then reverse engineered the transcriptional network responsible for NMP induction and differentiation. This revealed a network comprising the TFs Cdx1, 2, 4, T/Bra, Sox2, Msgn1, and Tbx6, which integrate Wnt and RA signaling to regulate entry to and exit from the NMP state. Mutation of individual or multiple components validated the network. Within the network, RA plays dual roles. Initially, RA is required for the expression of Sox2 and generation of NMPs; later, increased levels of RA drive neural differentiation. Cdx genes not only posteriorize cells but also, by restraining RA signaling, maintain T/Bra expression to allow mesoderm induction. Msgn1 and Tbx6 control the timing and outcome of NMP differentiation by cell-autonomously repressing T/Bra and Sox2 to propel mesoderm differentiation and by inducing the RA-producing enzyme Aldh1a2. The latter increases RA levels to non-autonomously promote neural differentiation and thereby provides regulative feedback that balances neural and mesoderm production. A dynamic model capturing these interactions demonstrates how the network coordinates the generation of the two cell types from the bipotential progenitor and ensures the well-proportioned generation of tissues necessary to build the vertebrate trunk.

RESULTS

Single-Cell Transcriptome Analysis of NMP Cells from Mouse Embryos

NMPs are a transient population, arising early in gastrulation, located in the CLE and CNH at the anterior end of the receding primitive streak and tail bud (Cambray and Wilson, 2002; 2007; Wymeersch et al., 2016; Olvera-Martinez et al., 2012). They can be identified in vivo by the co-expression of the TFs Sox2 and T/Bra (Tsakiridis et al., 2014; Martin and Kimelman, 2012; Olvera-Martinez et al., 2012). However, defining a definitive molecular signature has proved challenging because of their scarcity and transience in vivo. To address this, we performed single-cell transcriptome analysis (RNA sequencing [RNA-seq]) of NMPs micro-dissected from mouse embryos at two developmental stages (e8.5 and e9.5) (Figure 1A). We define these as embryo NMPs (e-NMPs).

Following sequencing and mapping of transcriptomes of dissected cells, we applied quality filters (See STAR Methods) that resulted in data from a total of 128 single e-NMP transcriptomes being retained for subsequent analysis (62 cells from e8.5 and 66 cells from e9.5). We took an unbiased, data-driven approach (STAR Methods). To establish concerted patterns of expression within an initial set of ~11,000 genes we analyzed those genes whose expression level distribution showed sufficient statistical dependence with at least two other genes. We used mutual information (MI) (reviewed in Mc Mahon et al., 2014) to measure these relationships with a cut-off value MI > 0.25. Hierarchical clustering of these 136 genes produced five major groups (modules) representing distinct patterns of gene expression. We examined the genes in each module (Table S1). Module 1 contained genes expressed in e8.5 primitive streak and e8.5 NMPs, including Cdh1, Cdx1, and Fst; module 2 comprised genes associated with e9.5 NMPs including Cyp26a1 and posterior Hox genes; module 3 was enriched for genes related to PSM specification including...
Figure 1. Single-Cell Transcriptome Analysis of In Vivo NMPs Defines the Molecular Signature of e8.5 and e9.5 NMPs

(A) Strategy for single-cell transcriptional analysis of e-NMPs dissected from the CLE region of e8.5 and e9.5 mouse embryos.

(B) Hierarchical clustering of embryo-derived single cells, using the genes in the first three modules (Table S1), columns represent cells and rows correspond to genes. This separates cells into two large groups correlating with developmental age (e8.5 and e9.5). Within the e8.5 group, three smaller clusters could be distinguished, an e8.5 NMP identity (expressing genes in module 1), MPCs (expressing genes in modules 1 and 3) and mesodermal cells (Meso) (expressing genes in module 3). The e9.5 group was subdivided into two groups, one associated with e9.5 NMP identity (module 2) the other with MPC fate (modules 2 and 3). The cluster identity of each cell from the e8.5 and e9.5 embryos is indicated in orange (e8.5 NMPs), purple (e9.5 NMPs), brown and pink (MPCs), and red (Meso).

(C) Pseudotemporal ordering of cells (right) obtained via the associated cell state graph (left) (expression levels are indicated as normalized counts per million reads). The white-to-red colors indicate NMP-to-mesodermal fractional identity of each cells defined by Sox2, Nkx1.2, Msgn1, Tbx6, and Meox1 levels.

(D) Developmental trajectories of single cells from e8.5 and e9.5 mouse embryos reveals three distinct populations, NMP (T/Bra+/Sox2+), MPC (T/Bra+/Msgn1+/Tbx6+), and PSM (T/Bra+/C0+/Msgn1+/Tbx6+).

(E) Molecular signature of e-NMP cells identified by differential expression analysis of the e8.5 (top) and e9.5 (bottom) NMPs with PSM cells. The analysis identified 68 genes that were associated with both e8.5 and e9.5 NMP identity and 31 genes associated with mesodermal differentiation. For clarity we have shown only the 31 genes most enriched in e-NMPs. The complete gene lists are given in Table S2.

(F) Differential expression analysis of e8.5 NMPs and e9.5 NMPs defines the molecular signature of e-NMPs at different developmental stages. Log CPM, logarithmic counts per million; Log f.c., logarithmic fold change; NMPs, neuromesodermal progenitors; MPC, mesodermal progenitors; PSM, presomitic mesoderm; SB, somite border.
but high levels of the PSM determinants $\text{Msgn1}$ in the $\text{e8.5}$ dataset, we asked whether it was possible to reconstruct the trajectory of NMP to PSM differentiation of NMPs to PSM, and indicates that we had successfully reconstructed the trajectory of NMP to PSM differentiation. We term these reconstructed trajectories “developmental timelines”.

The developmental timelines revealed three developmentally distinct groups of cells. First, a group of cells, present at both $\text{e8.5}$ and $\text{e9.5}$ that have the characteristics of NMPs, these expressed $\text{T/Bra}$, $\text{Sox2}$, $\text{Nkx1.2}$, $\text{Cdx2}$, and $\text{Cdx4}$. Second, a transition population that expressed $\text{T/Bra}$ and had upregulated $\text{Msgn1}$ and $\text{Tbx6}$, in which $\text{Sox2}$ and $\text{Nkx1.2}$ expression was still detectable albeit at decreasing levels. This is present at both $\text{e8.5}$ and $\text{e9.5}$ and we label these MPCs after the population defined by Chalamalasetty et al., 2014. Finally, a population of cells, present at $\text{e8.5}$ but not in the $\text{e9.5}$ that expressed $\text{Msgn1}$, $\text{Tbx6}$, and $\text{Meox1}$, in which $\text{Nkx1.2}$ and $\text{Sox2}$ had been repressed (Figures 1B–1D). This signature is characteristic of PSM cells (Chalamalasetty et al., 2014). The absence of cells with PSM identity from the $\text{e9.5}$ sample reflects the anatomical differences between the two time points and the region of the embryo included in the dissections at the two time points (Figure 1A).

The data allowed us to define a molecular signature of e-NMP cells. We performed differential expression analysis of the $\text{e8.5}$ and $\text{e9.5}$ NMP cells (labeled cerulean and purple, respectively) by comparing them with more differentiated (PSM) cells (Figure 1E). This analysis identified 99 genes differentially expressed in both e8.5 and e9.5 NMP populations (Table S2). Out of these, 31 genes were expressed at lower levels in e-NMP cells, compared with mesoderm, and were mainly associated with PSM differentiation, such as $\text{Aldh1a2}$, $\text{Meox1}$, $\text{DII}$, and $\text{Cited1}$. The other 68 differentially expressed genes were expressed more highly in e-NMPs. These included $\text{Nkx1.2}$, $\text{Sox2}$, $\text{Cdx2}$, $\text{Cdx4}$, and $\text{Sp8}$, described previously as expressed in NMPs, but also other genes potentially functioning in NMPs such as $\text{Epha2}$, $\text{Epha5}$, $\text{Sema6a}$, or $\text{Zic5}$. This set of genes thus provides an in vivo molecular signature of e-NMPs (Figure 1F).

We next identified genes differentially regulated between the $\text{e8.5}$ and $\text{e9.5}$ NMP populations, which represents change over embryonic time. Cadherin (Cdh1), Follistatin (Fst), and Pousf1 (Oct3/4) were specifically expressed in e8.5 NMPs as was, for example, Gsrf1, a gene implicated in axial elongation (Lickert et al., 2005) (Figure 1G). By contrast, Cyp26a1, an RA-degrading enzyme that functions as a negative feedback regulator of RA, was expressed at higher levels in e9.5 NMPs than in e8.5 NMPs (Figure 1G). However, the most striking difference between the $\text{e8.5}$ and $\text{e9.5}$ NMPs was the expression of Hox genes. The e9.5 NMP cells expressed significantly higher levels of posterior Hox genes (Hox 6–10), consistent with the acquisition of more posterior axial identity and a contribution to more posterior trunk regions as development progresses. The data allow the reconstruction of a developmental timeline of e-NMP to PSM differentiation and an embryonic aging timeline (associated with the axial identity of NMPs) for the transition from $\text{e8.5}$ to $\text{e9.5}$ NMPs.

**In Vitro Derived NMPs Resemble Their In Vivo Counterparts**

We have recently described the generation of NMP cells from mouse embryonic stem cells (ESCs) (Gouti et al., 2014). We generated single-cell transcriptomes of in vitro derived NMPs at day 3 (D3) of differentiation (Figure 2A) and compared these to their in vivo counterparts (e-NMPs) (Figure 1A). We took advantage of the gene signatures defined in vivo and used the three modules to characterize the in vitro NMPs (Figure 2B). A t-distributed stochastic neighbor embedding (tSNE) projection of $\text{e8.5}$, $\text{e9.5}$ and D3 NMPs, using the genes comprising the three modules, indicated that D3 NMPs were more similar to $\text{e8.5}$ NMPs than $\text{e9.5}$ NMPs (Figure 2D). Moreover, hierarchical clustering separated the in vitro NMPs into three groups (Figure 2B). Consistent with the tSNE projection, examination of the activity of each of the genes modules indicated that the majority (65%) of NMPs were similar to $\text{e8.5}$ NMPs. In addition, approximately 10% of the cells had a gene expression profile similar to $\text{e9.5}$ NMPs, and the remaining ~25% appeared to be PSM (Figure 2C).

We constructed a pseudotemporal ordering of D3 NMPs employing the 99 genes comprising the in vivo e-NMP signature (Table S2). Analysis of differentially expressed genes along the NMP trajectory recovered many of the same genes that displayed differential expression along the in vivo $\text{e8.5}$...
developmental trajectory. For example, T/Bra, Nkx1.2, Cdx1, 2, 4, and Sox2 typical of e-NMPs, were expressed in cells at the initial stages of the pathway. Sox2, Cdx2, Cdx4, and Nkx1.2 were downregulated at Msgn1 and Tbx6 were upregulated, whereas Cdx1 expression was maintained (Deschamps and van Nes, 2005), which is characteristic of the transition from NMPs to MPCs (Figures 2E and S1A). This suggests that in vitro, as in vivo, the cells are following a similar developmental pathway. Taken together therefore, the data indicate that in vitro derived D3 NMPs resemble in vivo e-NMPs and display a similar developmental timeline.

**Bifurcating Neural and Mesodermal Trajectories from NMPs**

The in vivo and in vitro developmental trajectories showed the transition of NMPs to PSM via an MPC state, but in neither case was an obvious neural lineage observed. This was expected from the experimental conditions. In vivo, the dissection avoided cells in the neural region of the embryo, whereas the Wnt signaling used in vitro favors the mesodermal lineage. To extend our analysis to neural differentiation, we allowed in vitro NMPs to differentiate to neural progenitors by removing the Wnt signaling agonist at D3 and assay cells at D4 in neurobasal (NB) medium containing vitamin A (the RA precursor) in the absence of Wnt signaling agonists (Figure 3A). Immunofluorescence analysis of T/Bra, Sox2, and Tbx6 at D4 indicated that, under these conditions, most cells acquired a neural identity while a small proportion continued to differentiate to mesoderm (Tbx6+) (Figures 3B and 3C). We therefore analyzed the transcriptome of single cells at D3 and D4, excluding the cells that were expressing mature mesoderm markers such as Meox1. Hierarchical clustering of D3 and D4 single cells (using genes comprising the first four modules obtained using the same approach as applied to e-NMPs, Table S3) separated cells into four distinct groups. These clusters appeared similar to e8.5 NMPs, NMPs transitioning to a developmentally later (e9.0) NMP expression profile (t-NMP), PSM, and neural progenitor cells (NPCs) (Figure 3E).

Combining the gene expression data from D3 and D4 we constructed pseudotemporal orderings. Two distinct routes could be clearly distinguished from this analysis (Figure 3D). Examination of the genes differentially expressed along these routes indicated that they corresponded to mesodermal and neural developmental trajectories. The mesodermal route is characterized by the transient MPCs population expressing T/Bra, Sox2, and Tbx6, as described above. By contrast, the neural lineage culminated in a population of cells that expressed Sox1, INX3, ZIC1, and ZIC2, typical of NPCs. Along this neural trajectory, genes such as T/Bra, E-Cadherin (Cdh1), and Pou5f1 (Oct3/4) associated with e8.5 NMPs were downregulated (Figures 3D and S1B).

We performed differential expression analysis of NPCs with NMPs to define genes associated with the neural trajectory. This analysis identified 143 genes differentially expressed. Out of these, 116 genes were downregulated as NMPs differentiated to neural progenitor cells, including Cdx1, T/Bra, Wnt3a, Fst, and Cdh1. The other 27 were induced as cells adopted a neural fate. These included Sox1, INX3, but also Meis1 and Meis2, which have previously been described to be expressed in neural tissue (Oulad-Abdelghani et al., 1997). Thus, this set of genes provides a molecular signature of neural induction from NMPs (Table S4).

**Induction of NMPs Requires RA Signaling**

Examination of the neural and mesodermal differentiation trajectories highlighted changes in the expression of RA signaling pathway components (Figure 3D). RA signaling is mediated by three RA receptors (RARα, β, and γ) that form heterodimers

**Figure 2. Single-Cell Analysis of In Vitro Derived NMPs Indicates that They Resemble Their In Vivo Counterparts**

(A) Schematic of the differentiation conditions used for the generation of in vitro NMPs from mouse pluripotent stem cells.

(B) Hierarchical clustering of all in vitro derived D3 cells using the genes contained in the three modules (identified from the in vivo population) reveals three distinct clusters. An e8.5 NMP identity characterized by genes in module 1, an e9.5 NMP identity characterized by genes in module 2, and a mesodermal characterized by genes in module 3.

(C) More than 60% of D3 NMPs had a profile similar to e8.5 NMP cells, and 10% had a profile similar to e9.5 NMP cells.

(D) tSNE projection of e8.5, e9.5 and D3 NMPs using the genes comprising the three modules reveals that D3 cells appear more similar to e8.5 NMPs than e9.5 NMPs.

(E) Pseudotemporal ordering of in vitro generated NMP cells at D3 identifies a similar developmental trajectory to in vivo derived cells. The state graph was obtained using the 99-gene signature of e-NMP cells partially shown in Figure 1F. The white-to-red colors indicate NMP-to-mesodermal fractional identity of each cell.
Figure 3. Single-Cell Analysis of the Differentiation Route of NMPs toward the Neural Lineage

(A) Schematic of the differentiation conditions used for the generation of in vitro NMPs and neural progenitor cells (NPCs) from mESCs.

(B) Immunohistochemistry of cultures at D3 indicates that most cells co-express Brachyury (T/Bra) and Sox2, characteristic of NMPs, and a small percentage of cells expressed Tbx6. Removal of CHIR after D3 and culture until D4 in NB (neurobasal) conditions induces the generation of NPCs that express Sox2 in the absence of T/Bra, and a few mesodermal cells expressing Tbx6.

(C) Quantitation of T/Bra+, Sox2+, Tbx6+, or T/Bra+/Sox2+ signal area normalized to DAPI area at D3 (bFGF/CHIR) and D4 (NB) of differentiation. Error bars indicate SD of four randomly selected independent fields.

(legend continued on next page)
with the retinoid X receptors (RXRα, β, and γ). In addition, aldehyde dehydrogenases (Aldh1a1, 2, and 3) play key roles in the synthesis of all-trans-RA from vitamin A. Consistent with the known in vivo expression profiles and functions (Ribes et al., 2009; Niederreither et al., 1997; Ang and Duester, 1997), the expression of nuclear RA receptor RARβ was upregulated in NPCs, whereas RARα was increased during PSM differentiation. RARγ and its binding partner RXRγ were expressed in NMP cells. Moreover, the expression of Aldh1a2, the rate-limiting enzyme for RA synthesis, correlated strongly with the expression of Msgn1 and Tbx6 (Figure 3D).

To test the possibility that RA functions during NMP differentiation, we abolished all RA signaling by generating an Aldh1a2 mutant ESC line (Aldh1a2<sup>−/−</sup>) and culturing these cells in the absence of vitamin A (STAR Methods). Strikingly, we found that removal of all RA signaling affected the ability to form T/Bra<sup>+</sup>/Sox2<sup>+</sup> NMP cells (Figures 4B, 4D, and S3B). Sox2 expression was unaffected at D2, the time point at which differentiating ESCs display an epiblast-like identity (Figure S3B). Nevertheless, in Aldh1a2<sup>−/−</sup> cells, cultured in the absence of vitamin A, Sox2 was downregulated at both the transcript and protein level at D3 (Figures 4B, 4D, 4J, and S3B). By contrast, the expression of the MPC markers T/Bra, Msgn1, and Tbx6 was induced at higher levels than controls, suggesting that the cells more efficiently adopted a mesodermal identity in the absence of RA (Figure 4D). Strikingly, the expression of the early primitive streak/mesodermal markers Eomes and Mix1 was induced in the absence of RA (Figure S2A). These data suggest that in wild-type (WT) cells, low levels of RA are required for the induction of NMPs, whereas complete elimination of RA signaling leads to an anterior mesodermal identity (Figure 4A) (Kessel, 1992).

Previous studies in chick and mouse embryos have shown that premature exposure of the tail bud to high RA levels leads to axis truncation (Tenin et al., 2010; Olivera-Martinez et al., 2012). We tested whether this was also the case in vitro. Exposure of cells to a combination of bFGF/CHIR/RA (10 nM/100 nM) from D2 to D3 resulted in the downregulation of T/Bra and Tbx6, and the acquisition of a pre-neural tube (PNT) identity characterized by the expression of Sox2, Sox1, and Nkkx1.2 (Figures 4A, 4C, 4D, 4J, and S2A). These data agree with in vivo data in which RA treatment of e8.5 mouse embryos rapidly downregulates Wnt3a and T/Bra in the primitive streak (Figure 4D) (Shum et al., 1999; Iulianella et al., 1999).

Collectively, the data identify a novel role for RA in NMPs. In the absence of RA, Wnt/FGF signaling promotes an anterior mesodermal identity (Bra<sup>+</sup>/Tbx6<sup>+</sup>/Cdx<sup>+</sup>), and low levels of RA are necessary for the induction of NMP identity (Bra<sup>+</sup>/Sox2<sup>+</sup>), whereas high levels of RA produce pre-neural progenitors (Sox2<sup>+</sup>/Nkkx1.2<sup>+</sup>). This suggests that tightly controlled levels of RA signaling are necessary for the induction of NMP and neural identity (Figure 4A, model).

Cdx Genes Maintain the NMP State by Suppressing RA Signaling

The failure to induce NMP cells in the absence of RA was accompanied by a downregulation of Cdx expression (Figure S2B). This prompted us to address the role of Cdx genes in NMPs. Previous studies have implicated the Cdx genes in the induction of T/Bra, Wnt3a, Fgf8, posterior Hox genes, and the RA-degrading enzyme Cyp26a1. However the presence of three functionally similar Cdx genes (Cdx1, Cdx2, and Cdx4) and the failure of axis elongation when all three are removed has complicated in vivo analysis and made it difficult to define a regulatory hierarchy.

We constructed ESCs harboring nonsense mutations in all six alleles of the Cdx genes (STAR Methods). We designated these Cdx<sup>1,2,4<sup>−/−</sup>/Cdx<sup>−/−</sup> cells. We then differentiated Cdx<sup>1,2,4<sup>−/−</sup>/Cdx<sup>−/−</sup> ESCs using NMP-inducing conditions (Gouti et al., 2014) (Figure 4E). Consistent with the established role of Cdx in the activation of posterior Hox genes, there was an absence of posterior Hox gene expression in Cdx<sup>1,2,4<sup>−/−</sup>/Cdx<sup>−/−</sup> cells, although the expression of anterior Hox genes was maintained (Figure S3G) (Mazzoni et al., 2013; Neijts et al., 2016).

At D3 T/Bra was expressed in Cdx<sup>1,2,4<sup>−/−</sup>/Cdx<sup>−/−</sup> cells, albeit it at lower levels compared with controls (Figures 4F, 4I, and 4J). Tbx6 was induced in Cdx<sup>1,2,4<sup>−/−</sup>/Cdx<sup>−/−</sup> cells at similar levels, whereas the expression of Msgn1 was lower compared with control cells at D3 (Figures 4F, 4I, 4J, and S3D). The expression of T/Bra was downregulated at D4 in the Cdx<sup>1,2,4<sup>−/−</sup>/Cdx<sup>−/−</sup> cells even in the presence of Wnt signaling (Figures S3C and S3D). This suggests that, although Cdx activity is not necessary to initiate T/Bra and Tbx6 expression, they act with Wnt/FGF signaling to maintain T/Bra and enhance Msgn1 expression. Consistent with the early downregulation of T/Bra, Tbx6 expression was not maintained in Cdx<sup>1,2,4<sup>−/−</sup>/Cdx<sup>−/−</sup> cells cultured under prolonged Wnt conditions (5 μM CHIR until D5) (Figures 4G, 4J, and 4I), and the cells acquired a

(D) Pseudotemporal ordering of D3 and D4 cells identifies four different populations and two distinct differentiation trajectories that lead to either neural or mesodermal identity. Expression of Cdx1, Cdx2, Cdx4, T/Bra, and Cdx1,2,4 is high in NMP cells, the expression of mesoderm specific genes Msgn1 and Tbx6 is high in Mesoderm cells. By contrast, there is induction of Sox1, Irx3, and Zic2 along the neural trajectory. The expression of RA signaling pathway components is differentially regulated in each developmental trajectory. RXRγ and RARγ are co-expressed in D3 NMP cells (similar to e8.5 NMPs). Expression of Aldh1a2 correlated strongly with Msgn1 and Tbx6 expression. See also Figure S1.

(E) Hierarchical clustering of D3 and D4 single cells partitions the cells into four major groups. An NMP cluster (NMP, cereulean), a transitioning NMP (t-NMP, cyan) that express markers of developmentally older NMPs, a mesodermal (Meso, red), and a neural progenitor cell cluster (NPC, green). D3 cells are indicated in gray and D4 cells in black. See Tables S3 and S4.

(F) Diagram illustrating the mesoderm or neural progenitor fate choice made by NMP cells.

(G) Schematic of the posterior part of an e8.5 mouse embryo. NMPs (cereulean) expressing Bra/Sox2/Cdx genes are located in the CLE region, close to the NSP in the anterior part of the primitive streak. As cells leave the NMP zone they differentiate to MPC progenitors (cereulean/red) expressing Bra/Msgn1/Tbx6, which results in upregulation of Aldh1a2. Thus, increased levels of RA produced in close proximity to the niche promote Sox2 expression and the differentiation of NMPs to PNT cells (green/yellow) then NPCs (green). The transcriptional network that controls the cell fate decision of NMP cells toward neural or mesodermal identities is summarized adjacent to the embryo model. NMP, neuromesodermal progenitor; t-NMP, transitioning neuromesodermal progenitor; NPCs, neural progenitor cells; PSM, presomitic mesoderm; MPC, neuromesodermal progenitor cells; PNT, pre-neural tube cells; NSB, node-streak border; PS, primitive streak; NB, neurobasal conditions.
Figure 4. Induction of NMPs Requires Low Levels of RA Signaling

(A) Schematic of in vitro differentiation conditions and diagram illustrating the different cell fate choices of epiblast cells in the presence of bFGF/CHIR signaling and variable levels of RA.

(B) Immunohistochemistry at D3 of differentiation indicates that Aldh1a2<sup>−/−</sup> ESCs exposed to bFGF/CHIR downregulate Sox2 and express the mesodermal markers T/Bra and Tbx6. WT ESCs differentiated under the same conditions co-express T/Bra and Sox2.

(legend continued on next page)
neural fate characterized by the expression of Sox2, Sox1, and Tuj1 (Figures 4G, 4J, S3E, and S3F). In contrast, WT cells under the same conditions adopted predominantly a PSM identity, characterized by the expression of Tbx6 and the downregulation of T/Bra (Figures 4G and 4H).

The expression of Wnt3a and Fgf8 was significantly downregulated in Cdx1,2,4−/− cells (Figure 4I), consistent with previous reports (Savory et al., 2009). However, since the exogenous supplementation of Wnt/Fgf is not sufficient to rescue the differentiation defects in Cdx1,2,4−/− cells, there must be additional targets of Cdx function. Inspection of the developmental timeline revealed an inverse relationship between the expression of the Cdx2, Cdx4, and RA production (Figure 3D). Consistent with this, the expression of Aldh1a2 was strongly induced in Cdx1,2,4−/− cells, whereas the expression of Cyp26a1 was significantly downregulated at D3 under NMP conditions (Figure 4I). The increased expression of Aldh1a2 in Cdx1,2,4−/− cells, together with the evidence that high levels of RA antagonize NMPs prompted us to ask whether increased production of RA in Cdx1,2,4−/− cells promotes abnormal neural differentiation. To test this, we cultured D3 Cdx1,2,4−/− in NMP-inducing conditions and exposed these to an RA inhibitor (1 μM BMS) from D3 to D5 (Figure 4E). This resulted in the induction of a mesodermal population in the Cdx1,2,4−/− cells that expressed T/Bra (Figures 4H and 4J). However, Sox2 was downregulated, indicating the absence of NMPs (Figures 4H and 4J). Thus, in the absence of Cdx1,2,4−/− mesoderm induction can be partially restored by the inhibition of RA signaling, suggesting an important role for Cdx in controlling levels of RA signaling to allow mesoderm and NMP production.

Taken together these data show that Cdx genes are necessary in establishing NMP identity and subsequently for the differentiation of spinal cord and PSM cells. This is achieved by inducing the expression of posterior Hox genes and by regulating the balance of Wnt3a, Fgf8, and RA signaling.

**Msgn1 Promotes the MPC-to-PSM Transition**

The production of neural cells in the absence of Cdx activity under continuous Wnt signaling led us to focus on the mechanisms that drive the exit from NMP identity. During the differentiation of NMPs to mesoderm, cells transit through an MPC state that co-expresses T/Bra, Msgn1, and Tbx6 and then downregulate T/Bra adopting a PSM identity ([*Msgn1*Tbx6*]) (Chalamalasetty et al., 2014) (Figures 1D and 2E). Mutations in *Msgn1* or *Tbx6* in vivo affect the production of PSM tissue causing defects in axis elongation (Chapman et al., 2003; Chapman and Papaioannou, 1998; Yoon and Wold, 2000).

To test how *Msgn1* affects PSM differentiation and what role it plays in the exit from the NMP state, we generated a *Msgn1* mutant ESC line (*Msgn1−/−*) (STAR Methods). We induced the differentiation of *Msgn1−/−* ESCs to NMPs (Figure 5A). At D3, NMPs were generated with high efficiency, demonstrated by the co-expression of T/Bra and Sox2 (Figures 5B and 5E). Notably, expression of T/Bra was consistently higher in the *Msgn1−/−* NMPs compared with the WT cells (Figure 5F). We then exposed NMPs to continuous Wnt signaling (CHIR) to promote differentiation into PSM. In contrast to WT cells, which have downregulated T/Bra and express the PSM markers Tbx6 by D5, *Msgn1−/−* ESCs maintained the expression of T/Bra, Sox2, Nkx1.2, and lower levels of Tbx6 (Figures 5C, 5E, and 5F). Thus, in the absence of *Msgn1*, cells are unable to progress to PSM and appear to be trapped in an NMP/MPC stage of differentiation. The consistently higher levels of T/Bra in the *Msgn1−/−* cells imply that *Msgn1* has a negative feedback on T/Bra to allow the exit from the MPC state. This is consistent with the in vivo data and suggests that *Msgn1* acts specifically to promote the MPC to PSM transition (Chalamalasetty et al., 2014; Yoon and Wold, 2000).

We next asked whether the absence of *Msgn1* affected neural differentiation. Surprisingly, although the withdrawal of the Wnt agonist allowed the neural differentiation of *Msgn1−/−* ESCs, this was substantially delayed compared with WT cells. In the *Msgn1−/−* ESCs the expression of T/Bra was maintained longer even under NB conditions, and there was a ~24 hr delay in the induction of neural identity compared with controls (Figures 5D and S4). Since *Msgn1* expression is largely confined to cells that adopt a PSM fate, these data suggested that *Msgn1* induces the production of a signal that acts non-autonomously to promote neural differentiation and block T/Bra expression. Examination of the single-cell transcriptome data revealed a strong correlation between *Msgn1* expression and Aldh1a2, the RA synthesis enzyme (Figure 3D). We hypothesized that the delay in neural differentiation in the absence of *Msgn1* might be due to low levels of *Aldh1a2*. To test this, we increased RA signaling...
by supplementing the medium with 100nM RA from D3 to D5 (Figure S4). This restored the timely downregulation of T/Bra and the upregulation of neural markers.

Taken together, the data indicate that Msgn1 is required to drive the cells from an MPC to a PSM state. This is achieved in part by the cell-autonomous enhancement of mesoderm
A Dynamical Model of NMP Induction and Differentiation

The identification of a role for RA in NMP induction and the availability of neural and mesodermal developmental timelines prompted us to explore whether this was sufficient to reverse engineer the underlying gene regulatory network. To this end, we constructed a dynamical model of linked ordinary differential equations describing gene activity. To identify relevant topologies and parameterize these models we took advantage of the gene expression dynamics revealed by the pseudotemporal orderings of single cells (Figure 3) and the results from the genetic perturbations (Figures 4, 5, and 6).

To reconstruct a minimal network we restricted our analysis to the regulatory relationships among the four core TFs T/Bra, Sox2, Msgn1, and Tbx6. Because of the tightly correlated expression and activity of Msgn1 and Tbx6, we used a single node to represent both genes. In the resulting network, we allowed all possible interactions between the three TF nodes (T/Bra, Sox2, and Msgn1/Tbx6). In total, this gave six possible interactions between the three TFs, and each could act as an activator or repressor, leading to 64 topologies (Figure 7B). In addition, the T/Bra and Sox2 nodes were connected to the two upstream signals Wnt/Fgf and RA (Figure 7C). For each topology, we identified best-fit parameters by running parallel differential evolution optimizations and comparing the resulting simulated dynamics of gene expression with a set of objectives. These objectives represented the gene expression trajectories from D2 of differentiation to Tbx6*/Msgn1* expressing mesoderm in the presence of Wnt/Fgf (Meso), or to Sox2*/Bra* neural progenitors in the presence of RA (neural). In addition, we included conditions representing T/Bra and Tbx6*/Msgn1* mutants and conditions in which in vitro NMPs are not produced (absence of Wnt signaling from D2 to D3) (Figure 7A).

The optimizations identified two topologies that generated good fits (Figure 7C). In both of these, all three TFs cross-repress each other, with the exception of the interaction of T/Bra on Tbx6, which could be either activating or weakly repressing. Examination of these topologies suggested an underlying logic in which, a toggle switch-like mechanism between a Sox2HIGH/Tbx6LOW (neural) state and a Sox2LOW/Tbx6HIGH (mesoderm) state is coupled to regulative feedback provided by RA. In this way, the circuit enforces a decision between neural and mesodermal states, with the feedback ensuring the balanced production of the two cell types: Tbx6 (mesoderm) increases RA and thereby increases Sox2 (neural) production in surrounding cells, conversely, low levels of Tbx6 increases T/Bra production and thereby Tbx6 (mesoderm) production.

We reasoned that this mechanism, together with the inherent stochasticity of gene expression, could be sufficient to explain the coordinated generation of neural and mesodermal tissue from NMPs. To test this, we performed simulations with stochastic differential equations and examined the dynamical trajectory and the proportions of each state produced. Consistent with the
Figure 6. Maintenance of NMP Cell Identity in the Absence of Tbx6

(A) Schematic of conditions to assay Tbx6–/– ESCs.

(B) Tbx6–/– cells co-expressing T/Bra and Sox2 at D3 under NMP conditions (bFGF/CHIR).

(C) Tbx6–/– cells are maintained as NMPs characterized by the co-expression of T/Bra and Sox2 at D5 (CHIR conditions), whereas WT cells mostly downregulate T/Bra and Sox2 and instead express Tbx6.

(D) At D5 in CHIR conditions, Tbx6–/– cells acquire an identity more similar to e9.5 NMPs characterized by the co-expression of T/Bra+/Sox2+ with Hoxc10+.

(E) qRT-PCR analysis of NMP, mesodermal, and neural markers at D3 (NMP conditions), D5 CHIR (mesodermal conditions), and D5 NB (neural conditions) shows that NMPs are induced at D3 in the Tbx6–/– cells and maintained during exposure to CHIR. Expression of Aldh1a2 is upregulated at D3 and D5 (CHIR) in WT ESCs, whereas Tbx6–/– cells express low levels of Aldh1a2 and higher levels of Fgf8 and Wnt3a at D5 (CHIR).

(F and G) The Tbx6–/– cells maintained NMP identity for 9 days (two passages) under bFGF/CHIR conditions (F) and progressively express more posterior Hox genes (G).

(H) tSNE projection of the in vitro WT and Tbx6–/– passaged NMPs with e8.5 and e9.5 NMPs revealed that D3 WT and Tbx6–/– in vitro NMPs are similar to e8.5 NMPs, whereas the passaged D6 Tbx6–/– in vitro NMPs closely resemble e9.5 NMPs.

(I and J) Taking the intersection of differentially expressed genes identified between in vitro D3 Tbx6–/– and D6 Tbx6–/– cells, and in vivo e8.5 and e9.5 NMPs showed similar changes in gene expression of early (I) and late NMP signature genes (J).
In neural conditions, the majority of stochastic simulations generated a Sox2\(^{\text{HIGH}}\)/Tbx6\(^{\text{LOW}}\) endpoint after transiently adopting a Sox2\(^{\text{MED}}\)/Bra\(^{\text{HIGH}}\) NMP-like state. Similarly, in mesoderm conditions an NMP-like state was transiently visited but now most stochastic simulations ended in a Sox2\(^{\text{LOW}}\)/Tbx6\(^{\text{HIGH}}\) (mesoderm) state. We next performed a series of simulations in which we systematically varied the levels of RA and Wnt in order to generate an input/output (I/O) map. This map was consistent with the regulative role of Wnt and RA feedback. Increasing the level of Wnt increased the proportion of

**Figure 7. Reverse Engineering the NMP Gene Regulatory Network**

(A) Six dynamical patterns summarize the experimental observations used as objectives to identify the best-fit network topology. Targeted time points are squares positioned at observed gene levels for Sox2 (orange), T/Bra (blue), and Msgn1/Tbx6 (red). The x axis represents the simulated time between D2 and D5. The y axis represents the protein concentration in a.u. All simulations are performed with a deterministic ordinary differential equation model from the initial conditions: Sox2\(^{\text{HIGH}}\), T/Bra\(^{\text{LOW}}\), and Msgn1/Tbx6\(^{\text{LOW}}\). In addition to WT simulations, T/Bra and Msgn1/Tbx6 mutant simulations were performed by setting the relevant protein production rate to zero. Dashed lines represent the four RA/Wnt signaling conditions and solid lines a typical best-fit solution.

(B) Overview of the 64 parallel parameter explorations evaluated. Each bar plot represents the average score of the best-fit solution cluster for one topology (a perfect score is 0). Bar colors identify the six objectives shown in (A). Underneath, circles describe the associated topology (blue circles for repression, red circles for activation) and the radius of each circle is proportional to the binding affinities in the best-fit solutions.

(C) The best-fit topology identified by the parameter exploration.

(D) Results of stochastic simulations. Colored pie charts show the ratios of cell states observed at D5 of stochastic simulations using the optimal topology under various RA/Wnt conditions between D3 and D5. Prior to D3, all trajectories are simulated with the same RA\(^{\text{LOW}}\)/Wnt\(^{\text{HIGH}}\) signaling condition.

(E) Stochastic trajectories obtained with four RA/Wnt signaling conditions that are representative of the four conditions in (D). The thick solid lines indicate the average trajectories for neural (green) and mesodermal (red) fates. Circles represent the attractors of the dynamical system.

(F) Experimental results. The ratios of different cell types obtained by co-measuring Sox2, T/Bra, and Tbx6 levels by flow cytometry. Assays were conducted using six RA/Wnt signaling conditions between D3 and D4. Grayscale panels show 2D kernel density estimations for each cell obtained from the protein fluorophore intensities. Contour plots document the highest density regions of the four cell populations identified by k-means clustering. Axes follow the “logicle” scale (Park et al., 2006). Associated bar plots indicate the ratios of cell types in each condition.

The deterministic model, in neural conditions the majority of stochastic simulations generated a Sox2\(^{\text{HIGH}}\)/Tbx6\(^{\text{LOW}}\) endpoint after transiently adopting a Sox2\(^{\text{MED}}\)/Bra\(^{\text{HIGH}}\) NMP-like state. Similarly, in mesoderm conditions an NMP-like state was transiently visited but now most stochastic simulations ended in a Sox2\(^{\text{LOW}}\)/Tbx6\(^{\text{HIGH}}\) (mesoderm) state. We next performed a series of simulations in which we systematically varied the levels of RA and Wnt in order to generate an input/output (I/O) map. This map was consistent with the regulative role of Wnt and RA feedback. Increasing the level of Wnt increased the proportion of
cells that adopted a mesoderm identity, whereas RA biased cells to neural. Combinations of both signals produced mixtures of cell types (Figure 7D).

The simulations appeared consistent with the experimental data generated by manipulating Wnt and RA signaling levels (Figures 3 and 4). To examine this in more detail we used fluorescence-activated cell sorting to quantify the proportion of differentiating in vitro NMPs expressing Bra/Sox2/Tbx6 under defined conditions. We used the fluorescent intensity levels of the three assayed TFs to define four clusters. These correspond to NMP (Bra<sup>HIGH</sup>/Sox2<sup>LOW</sup>/Tbx6<sup>LOW</sup>), neural (Bra<sup>LOW</sup>/Sox2<sup>HIGH</sup>/Tbx6<sup>LOW</sup>), mesoderm (Bra<sup>LOW</sup>/Sox2<sup>LOW</sup>/Tbx6<sup>HIGH</sup>), and unassigned (low for all three markers). We first examined NMPs cultured from D3 to D4 in NB conditions without RA or Wnt signaling. At D4 the number of cells in the NMP compartment had markedly reduced, and approximately equal proportions of neural and mesodermal cells were observed (Figure 7F). Exposing NMP cells to 100nM RA from D3 to D4 decreased the proportion of mesodermal cells substantially, concomitantly increasing the number of neural cells. Conversely, the mesoderm compartment increased at the expense of neural cells in NMPs exposed to 5μM CHIR from D3 to D4. Exposure of cells to a combination of 10nM RA and 5μM CHIR resulted in a similar proportion of cell types to basal conditions. Moreover, maintaining 10nM RA but reducing the level of Wnt signaling to 1μM CHIR increased the proportion of neural cells and reduced mesoderm proportion (Figure 7F). Taken together these data are consistent with the I/O map and confirm the regulative activity of RA and Wnt signaling.

**DISCUSSION**

We provide a detailed molecular description of bipotential NMPs and reconstruct their differentiation into neural and mesodermal tissue. Single-cell transcriptomics define gene expression signatures characteristic of NMPs and provide evidence that in vitro and in vivo NMPs are molecularly and functionally similar. These data enabled the reverse engineering of the underlying gene regulatory network and highlighted a role for RA signaling in establishing and resolving the NMP state. This revealed a mechanism that regulates the proportion of neural and mesodermal cells generated and offered insight into the network architecture that facilitates orderly axis elongation and the elaboration of the mammalian trunk.

**A Gene Expression Signature of NMPs**

Microdissection of embryonic tissue followed by single-cell transcriptome assays established a gene expression signature shared by e8.5 and e9.5 NMPs as well as identifying the changes associated with embryonic age. As anticipated, the core NMP signature included genes such as Tbx6, Cdx2, Sp8, and Sema6a, and Zic5 were enriched in e-NMPs. The function of many of these genes will require further investigation because, for example, Pou5f1 (Oct3/4) has recently been implicated in the maintenance of axial progenitors and trunk extension (Aires et al., 2016).

Alongside the core NMP signature, stage-specific gene expression was apparent. A set of genes including Cdh1, Fst, and Grsf1 was present in e8.5 NMPs but downregulated by e9.5. The well-established spatial-temporal order of Hox gene expression was also evident. Accordingly, e8.5 NMPs expressed 3’Hox genes, whereas e9.5 NMPs expressed more posterior 5’ Hox genes, notably Hoxa9, Hoxc9, Hoxd9, and Hoxa10. Apart from differences in Hox gene expression there was a clear difference in the expression of RA signaling pathway modulators. e8.5 NMPs expressed N/6a1, a mediator of RA signaling in ESCs that is required for the exit from pluripotency (Akamatsu et al., 2009). By contrast, Cyp26a1, the RA-metabolizing enzyme, is upregulated in e9.5 NMPs. Together, the data provide a core gene expression signature of e-NMPs and identified changes that represent the embryonic age of NMPs.

During the differentiation of NMPs to PSM, cells transit through an MPC state in which T/Bra, Nkx1.2, Msdn1, and Tbx6 are co-expressed. MPCs then downregulate T/Bra and Nkx1.2, adopting a PSM identity that is Msdn1<sup>1+</sup>/Tbx6<sup>2+</sup> (Chalamalasetty et al., 2014). Cells with the molecular characteristics of MPC were evident in vivo transcriptome data at both e8.5 and e9.5. In addition, cells with a PSM identity were observed from e8.5 embryos. This is consistent with the physical proximity of these cell types in the embryo. Moreover, a pseudotemporal ordering of cells, using the gene expression data, reconstructed the progression of NMPs to PSM. Hence, in addition to the embryonic time identified by comparison of e8.5 and e9.5 data, the developmental progression of NMPs to PSM was recovered from the in vivo analysis.

Conspicuously, D3 in vitro NMPs closely resembled e8.5 NMPs. In total, of the 68 genes comprising the e-NMP signature in vivo, 49 genes showed a similar behavior in vitro. Moreover, similar to in vivo, cells with an MPC identity were observed in vitro. The developmental trajectory reconstructed from the in vitro data mirrored that of the in vivo cells, indicating that succession of gene expression during NMP to PSM differentiation was similar in vitro as in vivo. Comparison of in vitro NMPs to e-NMPs also documented differences between the two populations. These included genes implicated in various aspects of metabolism and ribosome biogenesis. It seems likely that these differences relate to the culture conditions and the composition of the medium used in vitro. Irrespective of these differences, the overall correspondence in the core gene expression signature between in vitro NMPs and e-NMPs indicates that the in vitro population represents a valid model of in vivo NMPs.

Changes in the temporal axial identity of NMPs were recapitulated in vitro. At D3, in vitro NMPs were more similar to e8.5 NMPs than e9.5 NMPs. Removing Tbx6 allowed the maintenance of in vitro NMPs for several days, and this resulted in the induction of markers characteristic of e9.5 NMPs. At D6, most Tbx6<sup>−/−</sup> cells showed Hoxc10 expression, and by D9 Hox13 paralogs were activated (Figure 6G). It was notable that the proliferation of in vitro NMPs decreased over time and it proved impossible to maintain Tbx6<sup>−/−</sup> cells for longer than ~10 days. This appears to resemble the in vivo behavior of NMPs. The number of NMPs in an embryo peaks at e9.5-e10.5, then declines, and is exhausted by e13.5, coinciding with the increased expression of Hox13 proteins in the tailbud and the termination of axis elongation (Wymeersch et al., 2016; Denans et al., 2015; Young et al., 2009). These findings suggest that the elimination of NMPs is intrinsic and not dependent on external signals. This extends
the molecular and functional similarity between in vitro and in vivo NMPs and provides an experimentally tractable system in which to investigate the mechanisms controlling the exhaustion of axial progenitors and termination of axis elongation.

**RA Signaling in the Establishment and Resolution of the NMP State**

The reconstructed developmental trajectories highlighted the importance of RA in the formation and differentiation of NMPs. In embryos, the spatial distribution of RA signaling is controlled by the RA-synthesizing and degrading enzymes, *Raldh2 (Aldh1a2)* and *Cyp26a1*, respectively. During much of axis elongation, the somites and anterior regions of the PSM express *Raldh2* (Ribes et al., 2009; Sirbu and Duester, 2006). RA from these sources is implicated in neural induction and the transition of pre-neural cells into neural progenitors of the spinal cord (Diez del Corral et al., 2003). Consistent with this, the addition of RA to in vitro NMPs at D3 promotes neural differentiation.

*Aldh1a2* is also expressed in the primitive streak, some node cells, and posterior mesoderm as early as e6.75, whereas response to RA signaling is reported from e7.5 (Sirbu et al., 2005; Ribes et al., 2009). Targeted disruption of the *Aldh1a2* gene in mouse embryos is associated with the shortening of the anterior-posterior axis (Niederreither et al., 1999; Cunningham et al., 2015; Duester, 2008; Niederreither and Dollé, 2008), whereas excess RA administration results in caudal axial truncations (Julianella et al., 1999). Embryos lacking RARγ are resistant to these RA-induced posterior defects, highlighting a unique role for this receptor (Lohnes et al., 1993; Janesick et al., 2014). These data argue for an early role of RA in axis elongation and the generation of NMPs, but definitive evidence had been missing due to the potential maternal contribution of RA in the *Aldh1a2*−/− mice. Analysis of the single-cell data showed that the RARγ receptor is expressed in NMP cells in vivo and in vitro, whereas *Cyp26a1* is expressed in e9.5, but not e8.5 NMPs. Consistent with this *Cyp26a1* is upregulated in the tailbud at e9.0 (Sakai et al., 2001) and is proposed to maintain RA signaling at levels that allow the maintenance of NMP identity, whereas the absence of RA signaling in combination with high Wnt and Fgf signaling promotes T/Bra expression and an anterior mesodermal identity. Higher levels of RA anterior to the caudal epiblast blocks T/Bra induction resulting in the generation of PNT cells and the differentiation of neural progenitors (Olivera-Martinez et al., 2012). Collectively, therefore, the data argue that tightly controlled levels of RA signaling are essential for the induction and differentiation of NMPs. This explains the importance of RA in establishing the body axis and places renewed attention on understanding the integration of RA signaling with the other signaling pathways implicated in the development of cell types in the primitive streak.

Surprisingly, RA appears dispensable for the development of the zebrafish trunk (Alexa et al., 2009). This might reflect differences in the mechanism of axis elongation in different vertebrates (Steventon et al., 2016). In this context, understanding the function of FGF signaling may provide insight. FGF has been implicated in both neural and mesodermal induction (Stern, 2005; Naiche et al., 2011; Boulet and Capecchi, 2012), with both T/Bra and Sox2 identified as direct targets (Isaacs et al., 1994; Takemoto et al., 2006). Moreover, FGF signaling upregulates the expression of genes characteristic of NMPs, such as Nkx1.2, which in turn promote Fgf8 expression (Sasai et al., 2014). This suggests that FGF signaling plays a role in establishing the bipotentiality of NMPs, while RA and Wnt signaling promote resolution into neural and mesodermal fates, respectively.

**A Regulatory Network Balances Neural and Mesodermal Differentiation**

The ability to reconstruct developmental trajectories of both spinal cord and PSM allowed us to reverse engineer the NMP regulatory network and test the function of key factors. Mutations in the *Cdx* genes block the induction of posterior Hox genes and impair axial extension in embryos (Savory et al., 2009; Young et al., 2009; van de Ven et al., 2011; van Rooijen et al., 2012; Amin et al., 2016; Neijts et al., 2016). The truncation of the body axis can be partially rescued by exposure to Wnt or FGF signaling (Young et al., 2009; van Rooijen et al., 2012; van de Ven et al., 2011). Nevertheless, although the induction of *Wnt3a, Fgf8*, and *T/Bra* by Cdx proteins have been implicated in these phenotypes, the severity of morphological defects in mutant embryos has made it difficult to establish the relative importance of target genes and to rule out others. In vitro, the elimination of all three *Cdx* genes recapitulated in vivo observations: *Wnt3a, Fgf8*, and posterior *Hox* gene expression were lost. In addition, there was increased expression of *Aldh1a2* and a downregulation of *Cyp26a1*. Inhibition of RA signaling partially restored anterior mesoderm induction but did not restore NMPs. Thus, *Cdx* genes are necessary for the generation of functional NMPs. *Cdxx* genes are required for the expression of trunk Hox genes, the expression of which can at least partially substitute for the activity of Cdx proteins (Young et al., 2009). Consequently, Hox proteins may function downstream or with *Cdx* genes in NMPs.

The TFs, *T/Bra, Tbx6,* and *Msgn1,* are also implicated in NMP development (Nowotschin et al., 2012; Chapman et al., 2003; Garrick et al., 2015). Separating the roles of *Msgn1* and *Tbx6* is complicated by the similarity in their expression and function.
In vivo, both genes are induced simultaneously and the loss of one results in reduced expression of the other (Nowotschin et al., 2012). Embryos lacking either Msgn1 or Tbx6 have defects in axis elongation, partial loss of mesodermal tissue, and enlarged tailbuds (Chapman et al., 2003; Chapman and Papaioannou, 1998; Yoon and Wold, 2000). Tbx6 is important for the repression of Sox2 during mesoderm specification (Li and Storey, 2011; Takemoto et al., 2011). Conversely, overexpression of Msgn1 results in decreased neural induction (Chalamalasetty et al., 2014). In ESCs lacking either Msgn1 or Tbx6 genes, NMP identity could be induced with high efficiency, but exit from the NMP state and the induction of mesoderm differentiation was defective. Removal of Msgn1 resulted in co-expression of T/Bra, Sox2, and Tbx6. This argues that Tbx6 is not sufficient to repress Sox2 in the absence of Msgn1. By contrast, in cells lacking Tbx6, expression of Msgn1 was initially induced but not sustained. Instead, cells reverted to an NMP gene expression profile that was maintained for several days and multiple passages. In addition to the defects in mesoderm differentiation, Aldh1a2 induction was also lost in Tbx6 or Msgn1 mutants. The reduction in RA signaling resulted in the maintenance of a progenitor identity suggesting that, once induced, RA signaling is dispensable for NMP maintenance. Collectively, the data suggest that in the presence of Wnt signaling, the loss of Msgn1 blocks mesoderm differentiation at an MPC-like stage, whereas in the absence of Tbx6, cells do not progress beyond an NMP state.

We formulated a minimal gene regulatory model that accounts for the experimental observations and describes NMP formation and differentiation. In this model, Wnt signaling initiates a cascade by inducing T/Bra that represses Sox2 and induces Tbx6/Msgn1. T/Bra and Sox2 are then inhibited by Tbx6/Msgn1 to establish a mesoderm state. Tbx6/Msgn1 also activates RA production (by regulating Aldh1a2), which provides feedback to activate Sox2. The induction of Sox2 inhibits T/Bra and Tbx6/Msgn1 to establish neural identity. Hence, countervailing activities of Wnt and RA signaling, acting through mutually inhibiting T/Bra and Sox2, produces a switch that resolves into either mesoderm or neural differentiation. In the deterministic formulation of the model, the levels of Wnt and RA signaling dictates the outcome. The introduction of noise to the simulations, mimicking the stochastic nature of gene expression, produces a mixture of neural and mesodermal cells. The relative levels of RA and Wnt signaling control the proportion of the two cell types. In agreement with this, altering the levels of RA and Wnt signaling in differentiating NMPs in vitro determines the ratio of Sox2+ neural and Tbx6+ mesodermal cells generated.

From this point of view, the bipotential progenitor, corresponding to Sox2 and T/Bra co-expression, represents a metastable state in which two mutually inhibitory factors are activated in response to opposing signals. The expression of conflicting lineage-associated TFs has been referred to as multilinage priming and observed in other multipotent progenitors, including hematopoietic stem cells (Hu et al., 1997; Delassus et al., 1999; Enver and Greaves, 1998; Mannson et al., 2007) and ESCs (Hayashi et al., 2008; Laslett et al., 2007). For example, Oct3/4 and Sox2 maintain pluripotency of ESCs, but also drive the mutually exclusive differentiation to endoderm and neuroectoderm, respectively (Thomson et al., 2011). Thus the activities of Oct3/4 and Sox2 appear to parallel the function of Sox2 and T/Bra in NMPs.

A salient feature of the NMP regulatory network is that RA production from Tbx6*/Msgn1* mesoderm provides regulative feedback to NMPs. The differentiation of mesoderm increases RA production that, in turn, promotes Sox2 expression in undifferentiated NMPs and thereby increases neural differentiation at the expense of mesoderm (Figure 3G). Conversely, a decrease in mesoderm reduces RA signaling, leading to a decrease in Sox2 expression and a compensatory increase in mesoderm production. This reveals a design logic to the network architecture. Opposing signals establish a bipotential state that is resolved by cross-repressive transcriptional interactions between the induced TFs (Figure 3G). Regulative feedback from the progeny then ensures appropriate proportions of each tissue are generated. Thus, a feedback-modulated noisy bistable switch allows the balanced generation of two cell types from a bipotential progenitor and facilitates orderly axis elongation.

STAR METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Dissection of NMP Cells from Mouse Embryos
  - Generation of Mouse Mutant ESC Lines Using the Crispr/Cas9 System
  - ESC Culture and Differentiation
  - Reverse Transcription – Quantitative PCR Analysis
  - Immunostaining - Antibodies
  - FACs Analysis
  - RNA-Sequencing and Data Pre-Processing
  - Dimension Reduction
  - Cell Population Clustering
  - Single-Cell State Graph
  - Pseudo-Temporal Ordering
  - Deterministic Model of Gene Regulation
  - Optimization Algorithm
  - Parameters, Initial State and External Conditions
  - Distance Function
  - Stochastic Model of Gene Regulation
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.devcel.2017.04.002.

AUTHOR CONTRIBUTIONS

M.G. and J.B. conceived and designed the project and wrote the manuscript. M.G. performed and analyzed ESC differentiation experiments and generated the single-cell RNA-seq libraries. J.D. developed the single-cell data analysis methods and performed the computational modeling. D.S. designed and generated the Crispr ESC lines and performed ESC experiments. J.K.
developed the pipeline for single-cell RNA mapping. Y.H., F.J.W., and V.W. performed the dissections of NMP cells from mouse embryos.

ACKNOWLEDGMENTS

We thank Virginia E. Papaioannou for providing us with the Tbx6 mutant ESC line, Leena Bhaw-Rosun and Abdul Sesay for processing Illumina libraries, Robert Blaisberg for assistance with FACs experiments, and Johanna Schmid and Neda Mousavy for technical assistance. This work was supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001051), the UK Medical Research Council (FC001051), and the Wellcome Trust (FC001051); by the Wellcome Trust (grant reference WT098326MA); and by the BBSRC (grant reference BB/J015539/1). M.G. was also supported by the Max Delbrück Center for Molecular Medicine in the Helmholtz Association.

Received: September 21, 2016
Revised: February 20, 2017
Accepted: April 3, 2017
Published: April 27, 2017

REFERENCES


## STAR METHODS

### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goat anti-T/Bra</td>
<td>R&amp;D</td>
<td>Cat# AF2085; RRID: AB_2200235</td>
</tr>
<tr>
<td>Goat anti-T/Bra</td>
<td>R&amp;D</td>
<td>Cat# NL001; RRID: AB_663766</td>
</tr>
<tr>
<td>Goat anti-Tbx6</td>
<td>R&amp;D</td>
<td>Cat# AF4744; RRID: AB_2200834</td>
</tr>
<tr>
<td>Rabbit anti-Sox2</td>
<td>R&amp;D</td>
<td>Cat# AB5603; RRID: AB_304980</td>
</tr>
<tr>
<td>Rabbit anti-Hoxc10</td>
<td>Abcam</td>
<td>Cat# AB153904;</td>
</tr>
<tr>
<td>Mouse anti-Tuj1</td>
<td>Covance</td>
<td>Cat# MMS-435P; RRID: AB_2313773</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chemicals, Peptides, and Recombinant Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bFGF</td>
<td>R&amp;D</td>
<td>3139-FB</td>
</tr>
<tr>
<td>CHIR99021</td>
<td>Axon</td>
<td>1386</td>
</tr>
<tr>
<td>BMS493</td>
<td>Tocris</td>
<td>3509</td>
</tr>
<tr>
<td>All-trans-RA</td>
<td>Sigma</td>
<td>2625</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Critical Commercial Assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1 Single-Cell Reagent Kit for mRNA Seq</td>
<td>Fluidigm</td>
<td>100-6201</td>
</tr>
<tr>
<td>C1 Single-Cell Auto Prep Array for mRNA-seq</td>
<td>Fluidigm</td>
<td>100-6041</td>
</tr>
<tr>
<td>Nextera XT index kit</td>
<td>Illumina</td>
<td>FC-131-1002</td>
</tr>
<tr>
<td>Nextera XT DNA Sample Preparation Kit</td>
<td>Illumina</td>
<td>FC-131-1096</td>
</tr>
<tr>
<td>SMARTer Ultra Low RNA Kit for the Fluidigm C1 System</td>
<td>Takara</td>
<td>634833</td>
</tr>
<tr>
<td>Alexa Fluor Antibody labeling kit</td>
<td>Molecular Probes</td>
<td>A20181</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Deposited Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-cell RNA-seq data</td>
<td>This study</td>
<td>ArrayExpress: E-MTAB-5208</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experimental Models: Cell Lines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HM1 mESCs</td>
<td>Thermo Scientific</td>
<td>N/A</td>
</tr>
<tr>
<td>Aldh1a2&lt;sup&gt;−/−&lt;/sup&gt; mESCs</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td>Msgn1&lt;sup&gt;−/−&lt;/sup&gt; mESCs</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td>Cdx1&lt;sup&gt;1,2,4&lt;/sup&gt;&lt;sup&gt;−/−&lt;/sup&gt; mESCs</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td>Tbx6&lt;sup&gt;−/−&lt;/sup&gt; mESCs</td>
<td>Papaioannou VE Lab</td>
<td>Mech Dev. 2003 Jul;120(7):837-47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experimental Models: Organisms/Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse: MF1</td>
<td>Biomedical Research Facility at the University of Edinburgh</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Oligonucleotides</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crispr Guide Cdx1:</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td>GGGGCGCGCAAGGGCCAGGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crispr Guide Cdx2:</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td>GGCGGTGGATTCGGCGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crispr Guide Cdx4:</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td>GGATCGAAAAACTGGGGCG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crispr Guide Msgn1:</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td>GGCTGAGACAGGGCGGCG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued on next page)
CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by, the Lead Contact, James Briscoe (James.Briscoe@crick.ac.uk).

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Animal experiments were performed under the UK Home Office project licenses PD4515DD17 and PPL60/4435, approved by the Animal Welfare and Ethical Review Panel of the Francis Crick Institute and MRC Centre for Regenerative Medicine and within the conditions of the Animals (Scientific Procedures) Act 1986.

METHOD DETAILS

Dissection of NMP Cells from Mouse Embryos

Wildtype, outbred MF1 mice were used for timed matings. The noon on the day of finding a vaginal plug was designated e0.5 and embryos were collected at 8.5d.p.c and 9.5d.p.c. The caudal lateral epiblast region (L1-L2) was manually micro-dissected from e8.5 and e9.5 mouse embryos (Figure 1A; Wymeersch et al., 2016). The dissections removed the midline and the majority of underlying paraxial mesoderm cells using hand-pulled solid glass needles. The end product was mostly CLE with some underlying paraxial mesoderm cells. From each stage, the dissected regions were pooled separately and dissociated into single cells using a combination of enzyme (0.05% trypsin/EDTA) and mechanical (mouth pipetting) dissociation.

Generation of Mouse Mutant ESC Lines Using the Crispr/Cas9 System

We generated 3 different mutant mESC lines using the Crispr/Cas9 method (Aldh1a2−/−, Msgn1−/− and Cdx1,2,4−/− line). Different guide RNAs were designed to target each gene using the CRISPR/Cas9 design tool (crispr.mit.edu.au) (STAR Methods). Each guide was cloned into the pSpCas9/(BB)-2APuro vector (Addgene 48139; Ran et al., 2013).

The wild type ESC line, HM1 (Magin et al., 1992) was maintained in ES cell medium (Evans and Kaufman, 1981) with 1000U/ml LIF (Chemicon) on mitotically inactive primary mouse embryo fibroblasts. For the electroporation, ESCs were removed from feeders by dissociation using 0.05% trypsin and then plated onto gelatinised tissue culture plates for two short successive periods (20-30mins) to remove feeder cells. Electroporation was performed using 4x10⁶ ESCs and 4 ug of CRISPR-Cas9 plasmid DNA (specific guide) using Amaxa cell line nucleofector kit (Lonza) according to the manufacturer’s instructions (program A23). After electroporation cells were plated on Cell-Bind 10cm dishes with 2i medium/LIF (Ying et al., 2008). Puromycin selection was started 24 hours post-transfection by the addition of 1.5ug/ml puromycin for 48hours. Cells were allowed to grow for 5-7 days in 2i medium (Ying et al., 2008) and
the medium was changed every other day. Resulting colonies were picked manually and expanded in 2i medium plus LIF for genotyping. Clones were screened using specific PCR primers and then sequenced to confirm the presence of indels that introduce frameshift mutations and as a result early stop codons. The cell lines were tested for the expression of pluripotent markers.

ESC Culture and Differentiation
The mouse ESC lines, HM1 (WT) (Magin et al., 1992), Aldh1a2−/−, Msgn1−/−, Cdx1,2,4−/− and Tbx6−/− (Chapman et al., 2003) were maintained in ESC medium (Evans and Kaufman, 1981) with 1000U/ml LIF (Chemicon) on mitotically inactive primary mouse embryo fibroblasts. To remove feeders, ESCs were dissociated using 0.05% trypsin and plated onto gelatinised tissue culture plates for two short successive periods (20-30mins). To induce differentiation, the cells were plated on Cell-Bind surface dishes (Corning) precoated with 0.1% gelatin (Sigma) at a density of 5·8x10⁵ cells cm⁻² in ‘N2B27’ medium (NB). N2B27 is a 1:1 medium of Advanced Dulbecco’s Modified Eagle Medium F12 supplemented with 1 x N2 (Gibco), and Neurobasal medium (Gibco) supplemented with 1 x B27 (Gibco) or B27 minus Vitamin A (Gibco), 2mM L-glutamine (Gibco), 40μg/ml BSA fraction V (Sigma), 0.1mM 2-mercaptoethanol.

Cells were grown in N2B27 supplemented with 10ng/ml bFgf (R&D) for 2 days. To induce neomesodermal differentiation the cells were treated with 5μM CHIR99021 (Axon Medchem) and 10ng/ml bFgf for 24hours (D2-D3). During D3 the cells were transferred in N2B27 medium for neural differentiation or in N2B27 with 5μM CHIR99021 for mesodermal differentiation. For the maintenance of NMP identity, cells were passaged at a ratio 1:4 every 3 days on 35mm Cell-Bind dishes coated with 0.1% gelatin in N2B27 medium supplemented with 5μM CHIR99021 and 10ng/ml bFgf. The cells were dissociated into small cellular clumps mechanically using a plastic pipette tip. The medium was changed every other day.

Reverse Transcription – Quantitative PCR Analysis
Total RNA was isolated from cells using the RNeasy kit (Qiagen) according to the manufacturers instructions and digested with DNase I (Qiagen) to remove genomic DNA. First strand cDNA synthesis was performed with Superscript III system (Invitrogen) using random primers and amplified using Platinum SYBR-Green (Invitrogen). For QPCR the Applied Biosystems 7900HT Fast Real time PCR or the Light Cycler 480 SYBR Green I Master Mix (Roche) systems were used. PCR primers were designed using Primer3 software (See Table S5). All experiments were performed in biological duplicates or triplicates for each time point analysed. Expression values were normalized against the β-actin and standard calculated and plotted using Prism 6 software (GraphPad).

Immunostaining - Antibodies
For immunostaining cells were fixed for 15mins at 4°C in 4% paraformaldehyde in phosphate buffer saline (PBS), then washed in PBST (PBS with 0.1% Triton-X100). The cells were blocked for 30mins at RT using blocking buffer A (1% BSA/0.1% Triton-X100 in PBS). Cells were incubated with primary antibodies o/n at 4°C and with secondary antibodies at room temperature for 2 hours. For sequential immunostaining with two goat antibodies after the end of the 1st immunostaining, cells were incubated with blocking buffer B (1% goat serum, 1% BSA, 0.1% Triton-X100 in PBS) for 1h at RT. The second goat antibody was directly conjugated with a fluorophore (Bra–NL557). Cells were incubated with the goat-conjugated antibody for 2h at RT in blocking buffer B. After washing with blocking buffer A the cells were mounted with DAPI containing Prolong Antifade (Molecular Probes). The following primary antibodies were used: goat anti-Brachury (1:500) (R&D), goat anti- BraNL557 (1:250) (R&D), goat anti-Tbx6 (1:200) (R&D), rabbit anti-Sox2 (1:500) (Millipore), mouse anti-Tuj1 (1:1000) (Covance), mouse anti-Hoxc10 (1:200) (Abcam). The fluorescent images were captured using an inverted Leica SP5 confocal microscope.

FACS Analysis
Cells were dissociated into a single cell suspension using accutase (Gibco) for 5mins at 37°C. After washing with PBS, cells were fixed with 4% ice cold PFA on ice for 15mins. After washing with PBST (0.1% Triton-X100) the cells were incubated for 30min at RT in 100ul-200ul blocking solution (3%BSA / 0.2% Triton-X100 in PBS). The cells were treated with fluorescently conjugated primary antibodies for 1 hour at RT. Specifically, we used Sox2 conjugated to BD Horizon V450 (BD biosciences), T/Bra conjugated to NL557 (R&D) and Tbx6 (R&D) conjugated to Alexa Fluor 488. Conjugation of Tbx6 was performed using the Alexa Fluor 488 labelling kit (Thermofisher scientific) according to the manufacturer’s instructions. Conjugated antibodies were diluted 1:100 in blocking solution and incubated for 1hr at RT on a shaker. Cells were washed with PBST for 3 times. Finally, cells were filtered with 200ul PBS in a BD Facs tube. Data acquisition was performed on a BD FACSVerse flow cytometer.

To analyse gene expression levels from the FACS data, we first removed debris cells on the basis of the forward scatter-area (FSC-A) and side scatter-area signals (SSC-A). Then we removed doublets using the FSC-A and forward scatter-height (FSC-H) signals. To improve the separation of cells with minimal fluorescence, we trained and applied a “logicle” transform (Parks et al., 2006) on channels measuring Bra, Sox2 and Tbx6 levels. To do this, we used the estimateLogicle function from the flowCore R package.

To identify the populations contained in samples of our in vitro differentiation protocol, we performed a k-means clustering (k = 4) in expression space (Bra, Sox2, Tbx6) of samples from each of the D3 and D4 samples. Because the number of D4 cells were greater than D3 (128177 cells versus 16631 cells), the clustering was performed after randomly downsampling the D4 samples to match the D3 sample size. The center coordinates of the 4 resulting clusters clearly demarcate their identity: (BraHIGH/Sox2LOW) for NMPs, (BraLOW/Sox2LOW) for mesodermal cells, (BraLOW/Sox2HIGH) for neural progenitors and the fourth (BraLOW/Sox2LOW) cluster center was “unassigned” population. Cells from each individual sample were then associated to the closest cluster center.
**RNA-Sequencing and Data Pre-Processing**

The single cells were captured using the C1 Single-Cell Auto Prep Integrated Fluidic Circuit (IFC) (10-17um). RNA and cDNA was prepared from single cells using the SMARTer ultra low RNA kit (Clontech). The RNA-Seq libraries were prepared using the Nextera XT DNA library prep kit (Illumina) according to manufacturers instructions described in the protocol (PN 100-7168, http://www.fluidigm.com/). Library size, purity and concentration were determined using Agilent Technologies 2100 Bioanalyzer. The libraries from 70-80 individual cells were pooled and sequenced using the Illumina Genome Analyzer Hiseq2500.

Sequences were aligned to the Ensemble mouse genome GRCh38 using Bowtie2 (Langmead and Salzberg, 2012) and Tophat2 (Kim et al., 2013), and counted with HTSeq-count (Anders et al., 2015). Cell debris and doublets were removed from the data by inspecting microscope images of the microfluidic chips. Low-quality libraries were excluded whenever their transcript abundance was less than $10^6$ reads and the number of expressed genes was less than one thousand. The library sizes were then normalized to read counts per million (CPM). Genes with counts in less than 10 cells or annotated as pseudogenes were excluded from the analysis.

**Dimension Reduction**

The variation of gene expression contained in single cell transcriptomes is a combination of meaningful biological differences and technical noise from sampling effects due to low-numbered transcripts. We set up a dimensionality reduction strategy to identify the genes demonstrating significant variation and concerted patterns of expression.

We performed a selection of genes with concerted patterns of expression by calculating the mutual information for all pairs of gene in the dataset. Mutual information is used here as a generalized correlation measure enabling the quantification of nonlinear gene relationships. Gene levels were preliminary discretized using a nonparametric and unsupervised density estimation algorithm, Bayesian Blocks (Python module "AstroML"). Mutual information for each pair of discretized gene levels was determined by subtracting the joint entropy from the sum of both genes’ marginal entropy (R package “entropy”). Genes were retained if they shared sufficient mutual information with at least 2 other genes ($MI > 0.25$).

In order to refine gene lists and identify the dominant biological characteristics, we calculated the Spearman rank correlation matrix of gene expression for the dataset. We then grouped correlated genes into an initial set of gene modules by performing a hierarchical clustering of the correlation matrix. Because the correlated genes were more abundant in our in vitro dataset (445 genes) compared to our in vivo dataset (136 genes), we adopted a different strategy for defining the gene modules when processing the two datasets. For the in vitro dataset, the dendrogram associated with the hierarchical clustering was thresholded to obtain about 300 gene modules mostly composed of 1 or 2 genes and the larger modules (at least 4 genes) were retained. This method is similar to the one described in (Patel et al., 2014). Some modules were discarded if they did not have a recognizable function. For the in vivo dataset, the associated dendrogram was thresholded such as obtaining the 5 gene modules described in the main text.

Our pipeline does not explicitly select genes based on their variability. In order to assess whether the genes contained in the selected modules showed significant variation, we evaluated the dispersion of each gene, i.e. the ratio of variance over mean, for each gene across the single cell dataset, as described in (Satija et al., 2015). In this algorithm each gene is associated with one of 20 bins, based on their associated average expression and the dispersion measures are z-normalized within each bins. We obtained an average z-score of about 1 for our selection of genes, confirming that they indeed showed substantial variation.

**Cell Population Clustering**

Cell populations were identified by hierarchical clustering using R’s Ward.D2 clustering and the Spearman correlation of the z-scored gene levels of the genes comprising the selected modules (in vivo modules in Figure 1B, in vitro modules in Figure 3D). Hierarchical clustering was performed on the normalized in vivo gene module averages in Figure 2A: first, gene levels belonging to the same module were averaged, then z-scored by module, and eventually z-scored by cell.

**Single-Cell State Graph**

To investigate the dynamical changes of the transcriptional profile as cells differentiate, we developed a method to relate each cell to its most similar siblings. Unlike cluster analysis which aims to partition cells into groups with similar characteristics, hence breaking the continuity of cell state differentiation, graphs can connect individual cells without requiring the definition of groups. These structures most similar siblings. Unlike cluster analysis which aims to partition cells into groups with similar characteristics, hence breaking the continuity of cell state differentiation, graphs can connect individual cells without requiring the definition of groups. These structures reveal the differentiation trajectories and intermediate states that underlie the clusters of similar cells (the “clustered” populations).

Using the logarithm of normalized levels ($log_{10}$transcripts per million) in the reduced space, we first calculated the Euclidean distance matrix between each cell and hence constructed a complete weighted graph of cell similarity $D$. In (Trapnell et al., 2014, Camp et al., 2015), a minimum spanning tree algorithm (MST) was used to extract the subset of cell–cell edges which forms the backbone of differentiation branches. While MSTs ensure that all cells are connected, they are also sensitive to noise, making the local topology sensitive to small changes in the data (Zemel and Carreira-Perpinan, 2005). To improve robustness to noise of MSTs, we constructed a consensus graph which combines multiple perturbed minimum spanning trees (pMST). Each pMST is obtained by calculating a MST from the cell dissimilarity matrix $D$ with a certain ratio $j$ of its elements are set to a very large value ($j=20\%$), hence forbidding the recruitment of the associated edges. Individual pMST are merged by summing their adjacency matrix $A$ into a matrix storing the occurrences of each edge. We then exclude rare used edges by clustering the non-null edge occurrence distribution using the Fisher method (Fisher, 1958) and removing all edges belonging to the first class. This leaves edges that are used repeatedly in multiple permutations and therefore represent choices for inclusion in MST graphs. The perturb-and-merge algorithm works iteratively until convergence in the number of included edges.
Pseudo-Temporal Ordering

One of the advantages of generating a single-cell state graph is the possibility to infer a pseudotemporal ordering of the gene expression by following the gene expression implied by the spanning tree. The strategy we applied was to isolate two terminal cell populations, early and a late and identify the K-shortest paths that connect each pair of early and late cells (Martins and Pascoal, 2009). Typically, thousands of paths were generated from terminal populations consisting in 12 cells and K = 100. The resulting paths were not necessarily the same length hence, in order to average gene expression along all paths, we first rescaled them into a common pseudotime scale. Considering that all cells belonging to the early (late) population would define the initial (final) pseudotime point, and setting the number of pseudotime points to a value (typically 50), we simply rebinned the gene expression levels without interpolation according to each cell rank along the path. The averaged gene expression along the pseudotime line was then smoothed using a local polynomial regression fit (R function loess with span=0.5).

In order to compare the pseudotemporal ordering obtained from different cell state graphs, we defined the fractional identity of each cell as the score obtained by linear combination of the normalized gene levels of a given list of genes assumed to be either increasing or decreasing between the terminal populations. These fractional identity scores were then used to define the start and end coordinates of the timelines in a reference pseudotime axis, i.e. the developmental timeline. When comparing the e8.5 to the e9.5 cells (Figure 1C), the gene list was composed of Sox2, Nkx1.2, Tbx6, Msgn1 and Meox1 with the first two genes having a negative coefficient (equal to -1) and the last three a positive one (equal to 1) when linearly combined. When comparing the e8.5 to the D3 cells (Figure 2E), the gene list was composed of the 99 NMP signature genes with the 68 NMP markers having a coefficient equal to -1 and the 31 mesoderm markers’ coefficients equal to 1. In Figure 2E, the fractional identities were normalized independently for in vivo and in vitro cells, leading to equal-length timelines.

Deterministic Model of Gene Regulation

A gene regulation model based on a thermodynamic formulation (Sherman and Cohen, 2012) was used to simulate the gene expression dynamics of the Bra-Sox2-Msgn1/Tbx6 network.

This ordinary differential equation (ODE) model describes that the production rate of a gene X as function of the occupation state of its enhancer and is expressed as the product of $\rho_X^{\text{bound}}$, the probability of RNA polymerase to be bound to the promoter and $\alpha_X$ and the maximal production rate of the gene.

To calculate the polymerase binding probability, all possible states of the enhancer are enumerated, i.e. all possible combinations of upstream transcription factors (TFs) bound or not bound to the gene’s cis-regulatory modules (CRM). Let $S_X$ be the set of inputs binding to the CRM of X.

$$S_X = \{\text{Pol}\} \cup \{\text{Act}\} \cup \{\text{Rep}\}$$  \hspace{1cm} (Equation 1)

where $\{\text{Pol}\}$ represents the RNA polymerase, $\{\text{Act}\}$ is the set of activating transcription factors and $\{\text{Rep}\}$ is the set of repressing TFs upstream of X. In order to account for the situation of transcription factors being bound independently to multiple sites, multiple instances of the same TF can be stored in $\{\text{Act}\}$ and $\{\text{Rep}\}$.

Then, each possible promoter state s is represented by an element of the powerset of $S_X$, and is associated with a statistical weight $W(s)$ derived from Gibbs free energy of each binding interaction (Shea and Ackers, 1985).

$$\forall s \in 2^{S_X}, W(s) = \prod_{T \in s} K_{TX} c_{TX} x_T$$  \hspace{1cm} (Equation 2)

where $K_{TX}$ is the equilibrium binding constant of input T on X, $c_{TX}$ is the level of binding cooperativity between the input T and RNA polymerase, and $x_T$ is the concentration of T. When $c_{TX} > 1$, T acts as an activator and conversely, if $c_{TX} < 1$, T acts as a repressor. In the following, we assume that all repressors are strongly inhibiting polymerase binding, i.e. $c_{TX} = 0, \forall T \in \{\text{Rep}\}$. We denote by $S_X^+$ the set of transcriptionally active states, i.e. containing RNA polymerase and no repressor, and $S_X^-$ the other transcriptionally inactive states.

The polymerase binding probability is defined as the ratio of $Z_X^{\text{bound}}$, the sum of the statistical weights of all transcriptionally active states over the total statistical weight of all possible states.

$$\rho_X^{\text{bound}} = \frac{Z_X^{\text{bound}}}{Z_X^{\text{bound}} + Z_X^{\text{unbound}}}$$  \hspace{1cm} (Equation 3)

The factored forms of the total statistical weight of all active and inactive states read:

$$Z_X^{\text{bound}} = \sum_{s \in 2^{S_X^+}} W(s) = K_{Pol0X_Pol} \prod_{T \in \{\text{Act}\}} (1 + K_{TX} c_{TX} x_T)$$  \hspace{1cm} (Equation 4)

$$Z_X^{\text{unbound}} = \sum_{s \in 2^{S_X^+}} W(s) = \prod_{T \in \{\text{Act}\} \cup \{\text{Rep}\}} (1 + K_{TX} x_T)$$
As an example, the complete statistical weights of the genes interacting according to the best topology shown in Figure 7C read:

\[
\begin{align*}
\text{Sox2:} & \\
& Z_{\text{bound}}^S = KR_{\text{Pol}} X_{\text{Pol}} (1 + KR_{\text{BS}} C_{\text{BS}}) ^{n_{\text{BS}}} \quad (\text{Eq} \text{. 5}) \\
& Z_{\text{bound}}^B = (1 + KR_{\text{Pol}} X_{\text{R}}) ^{n_{\text{Pol}}} (1 + K_{\text{BS}} C_{\text{BS}}) ^{n_{\text{BS}}} (1 + K_{\text{TS}} C_{\text{TS}}) ^{n_{\text{TS}}} \\
& Z_{\text{bound}}^{\text{Msgn1/Tbx6}} = KR_{\text{Pol}} X_{\text{Pol}} (1 + KB_{\text{BS}} C_{\text{BS}}) ^{n_{\text{BS}}} (1 + K_{\text{SM}} C_{\text{SM}}) ^{n_{\text{SM}}} \\
& Z_{\text{bound}}^{\text{Bra}} = (1 + KB_{\text{BS}} C_{\text{BS}}) ^{n_{\text{BS}}} (1 + K_{\text{SM}} C_{\text{SM}}) ^{n_{\text{SM}}} (1 + K_{\text{TS}} C_{\text{TS}}) ^{n_{\text{TS}}} \\
& Z_{\text{bound}}^{\text{Bra}} = (1 + KB_{\text{BS}} C_{\text{BS}}) ^{n_{\text{BS}}} (1 + K_{\text{SM}} C_{\text{SM}}) ^{n_{\text{SM}}}.
\end{align*}
\]

where \( S \) stands for Sox2, \( B \) for Bra, \( M \) for Msgn1/Tbx6, \( R \) for RA, \( W \) for Wnt and \( Pol \) for RNA polymerase; and \( n_{TX} \) represent the number of binding sites for TF \( X \)’s CRM.

The rate of change of protein concentration is given by the first order rate equation:

\[
x_X = \alpha_X p_X^\text{bound} - \delta_X x_X
\]

where \( \delta_X \) is the degradation rate of protein \( X \).

Finally, to account for the external conditions to which the cells are exposed, we introduce two extrinsic concentration terms, \( x_R^{\text{ext}} \) for RA and \( x_W^{\text{ext}} \) for Wnt, which are added to the previous concentration terms when these pathways are activated.

**Optimization Algorithm**

Our optimization strategy aims at evaluating the fitness of all network topologies linking the 3 TF nodes Bra, Sox2 and Msgn1/Tbx6. Allowing each of the 6 interactions to act as an activator or as a repressor, the parameter space of \( 2^6 = 64 \) topologies was explored. We used a Python implementation of the Differential Evolution algorithm (DE) (Storn and Price, 1997) (Python function optimize.differential_evolution from the SciPy module). DE is a simple and fast evolutionary algorithm well suited for high-dimensional optimization, but it does not guarantee an optimal solution. To strengthen our confidence in the solutions identified, each topology optimization was run independently 30 times from random initial populations, hence generating a set of 30 best individuals. This population of best-fit solutions was then clustered into one or more subgroups of similar parameter signatures. The clustering was processed by fitting Gaussian Mixture Models with an expectation-maximization algorithm and identifying the optimal number of clusters according to the Bayesian information criterion for each model. Each topology run offered at least one optimal cluster of at least 4 individuals from which the average topology performance was calculated.

**Parameters, Initial State and External Conditions**

The model parameters were specified according to the following rules:

- Each topology is defined by setting the cooperativity factor \( c_{TX} \) to 10 for activators, and 0 for repressors.
- For each topology, the only optimized parameters are the 6 binding affinities \( K_{TX} \), i.e. \( K_{\text{MB}}, K_{\text{BS}}, K_{\text{BS}}, K_{\text{MS}}, K_{\text{BM}}, K_{\text{SM}} \). These parameters are constrained in the \([0.01,1000]\) interval.
- Wnt and RA binding affinities are kept constant, \( K_{\text{WB}} = K_{\text{RB}} = 10 \)
- RNA polymerase binding affinity is kept constant, \( K_{\text{PolX}} = 1 \).
- The number of binding sites is set to 2 (\( n_{TX} = 2 \)).
- The maximal production and degradation rates of each protein is set to 2 (\( \alpha_X = \delta_X = 2 \)) which constrains protein concentrations to the \([0,1]\) interval.

At the initial D2 time point, all initial protein concentrations are close to 0: \( x_B(t_{D2}) = x_M(t_{D2}) = x_R(t_{D2}) = x_W(D2) = 0.001 \) except for Sox2 concentration which is high, \( x_S(t_{D2}) = 0.8 \). This represent the “epiblast” state.

During the simulations, the cells are exposed to one of the four signaling conditions (see dashed lines in Figure 7A):

- **Mesodermal condition:** Wnt high (\( x_W^{\text{ext}} = 1.0 \)) and basal level of RA (\( x_R^{\text{ext}} = 0.1 \)) from D2 to D5.
- **Neural condition I (RAplus):** same external concentrations between D2 and D3 as above, then RA high (\( x_R^{\text{ext}} = 1.0 \)) and null Wnt (\( x_W^{\text{ext}} = 0.0 \)) between D3 and D5.
- **Neural condition II (NB):** same external concentrations between D2 and D3 as above, then basal RA (\( x_R^{\text{ext}} = 0.1 \)) and null Wnt (\( x_W^{\text{ext}} = 0.0 \)) between D3 and D5.
- **Anterior condition:** basal level of Wnt and RA (\( x_W^{\text{ext}} = x_R^{\text{ext}} = 0.1 \)) between D2 and D5.

Finally, the Brachyury mutant and Msgn1/Tbx6 mutant simulations were performed with the same parameters as the WT simulations, the only difference being the elimination of the associated protein by setting its maximum production rate to 0 (\( \alpha_B = 0 \) or \( \alpha_M = 0 \)).
Distance Function
We defined a distance function to quantitatively evaluate the differences between the simulated dynamics and reconstructed WT and mutant timepoints. Each parametric solution was simulated six times: 3 WT experiments (Mesodermal, Neural I and Anterior conditions) and 1 Bra mutant experiment (Mesodermal condition) and 2 Msgn1/Tbx6 mutant experiments. After scaling the trajectories of each gene by dividing by the maximum value for that gene over all 6 experiments, each experiment score was calculated by averaging the absolute difference between the simulated and targeted gene levels at specific time points (squares in Figure 7A). The distance function is then obtained by summing these 6 experiment scores.

Stochastic Model of Gene Regulation
To investigate how the the minimal network model behaves when challenged with the inherent stochasticity of gene regulation, we converted each gene rate equation (Equation 6) to follow the Chemical Langevin Equation (CLE) approximation:

\[
\dot{x}_X = \alpha_x p_{X}^{\text{bound}} - \delta_x x_X + \sqrt{\frac{\alpha_x p_{X}^{\text{bound}} + \delta_x x_X}{\Omega}} \xi
\]

(Equation 7)

where \(\Omega\) is the volume parameter relating concentrations with number of molecules \((N_X = x_X \Omega)\), and \(\xi\) is a random variable following a normal distribution with zero mean and unit variance.

The transformation procedure is detailed in (Perez-Carrasco et al., 2016).

QUANTIFICATION AND STATISTICAL ANALYSIS
To quantify immunofluorescence intensities in fixed cells all images were converted to 8-bit images and were segmented in classes by thresholding using methods available in FIJI. The signal area of each marker was calculated and divided by the corresponding total signal area for DAPI.

All bar plots’ error bars indicate the standard deviation of randomly selected independent fields for immunofluorescence quantifications or the standard deviation of biological replicates for qRT-PCR quantifications. Sample size are indicated in the associated figure legends.

Genes differentially expressed between distinct cell populations were assessed by performing an approximate \(\chi^2\) likelihood ratio test between a Tobit model fitted to the two populations (alternative model) and a Tobit model fitted to all cells (null model), using the differentialGeneTest function of the Monocle package (Trapnell et al., 2014). This test was performed independently for each gene and selected sets were obtained by thresholding on p-values (p-values threshold of 5e-5 for Figures 1E, 1F, 1G and 5e-7 for Figures 6I, 6J).

All other details of the single cell RNA-seq analysis are provided in the detailed methods.

DATA AND SOFTWARE AVAILABILITY
The accession number for the single cell RNA-seq data reported in this paper is ArrayExpress: [E-MTAB-5208].

All simulations and analysis were performed in R (The R Foundation) and Python (Python Software Foundation) as described in the detailed methods.