The molecular logic of the self-organization of primitive streak and neuroectoderm fates

Luigi Russo[1,2], Hanna L. Sladitschek[1,3], Pierre A. Neveu[1,4,*]

[1]Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Meyerhofstraße 1, 69117 Heidelberg, Germany

[2]Joint PhD degree from EMBL and Heidelberg University, Faculty of Biosciences

[3]Present address: Department of Molecular Medicine, University of Padua School of Medicine, viale Colombo 3, 35126 Padua, Italy

[4]Lead Contact

[*]Correspondence: neveu@embl.de

SUMMARY

The formation of the primitive streak (PS) and the subsequent induction of neuroectoderm are hallmarks of gastrulation. Combining an *in vitro* reconstitution of this process based on mouse embryonic stem cells (mESCs) with a collection of knockouts in reporter mESC lines, we identified retinoic acid (RA) as a critical mediator of neural induction triggered by TGFβ or Wnt signaling inhibition. Single-cell RNA sequencing analysis captured the temporal unfolding of cell type diversification, up to the emergence of somite and neural fates. Genetic and pharmacological perturbations demonstrated that TGFβ and Wnt ligands induced and maintained a PS-like population and its mesendoderm derivatives at the expense of neuroectoderm formation. In conditions thought to lack RA synthesis, we discovered a hitherto unidentified residual RA signaling that specified neural fate. Genetic evidence showed that the RA-degrading enzyme Cyp26a1 protected PS-like cells from neural induction, and notably in the absence of TGFβ and Wnt antagonists.

INTRODUCTION

During gastrulation, cells of the epiblast are allocated to the three germ layers (Tam and Behringer, 1997; Arnold and Robertson, 2009). Gastrulation is initiated by formation of the primitive streak (PS) and subsequent induction of neuroectoderm. Seminal experiments by Spemann and Mangold showed that the transplantation of the dorsal blastopore in amphibians could induce a secondary axis and neural tissue in the host embryo (Spemann and Mangold, 1924). This region or "organizer" secretes a range of molecules that mediate this induction (De Robertis, 2006). Among them, antagonists of the transforming growth factor β (TGFβ) signaling pathway and in particular of bone morphogenetic proteins (BMPs) are considered pivotal for neuralization of the ectoderm (Hemmati-Brivanlou and Melton, 1997; Weinstein and Hemmati-Brivanlou, 1999). The inhibition of the Wnt signaling pathway is another potent inductive cue (Glinka et al., 1998). While most of the molecular mechanisms governing this process were determined in amphibians, they appear to be conserved in mammalian development (Levine and Brivanlou, 2007). The deletion of the two TGFβ inhibitors Chordin and Noggin (Bachiller et al., 2000) or the knockout of the Wnt inhibitor Dkk1 (Mukhopadhyay et al., 2001) lead to the absence of forebrain structures in mouse. Retinoic acid (RA) is another signaling molecule with potent neuralizing activity (Rhinn and Dollé, 2012) that was found to be produced by the Hensen's node, the chick equivalent of the organizer (Hogan et al., 1992). Furthermore, RA signaling was detected in the primitive streak region of the mouse embryo (Rossant et al., 1991). Aldh1a2 is considered to be the only enzyme that can synthesize RA from retinal at this developmental stage (Rhinn and Dollé, 2012) and its knockout leads to an absence of expression of an RA activity reporter (Niederreither et al., 1999). However, the presence of forebrain structures in *Aldh1a2*$^{-/-}$ embryos ruled out RA involvement in neuroectoderm induction in early post-implantation development (Niederreither et al., 1999). This is in stark contrast with the widespread use of RA to induce neuronal fates from pluripotent cells *in vitro* (Ying et al., 2003; Bibel et al., 2004). While classical *in vivo* work established the importance of antagonizing TGFβ and Wnt signaling pathways in the process of neural induction, the molecular implementation of the neuroectoderm differentiation decision is largely unexplored. Furthermore, how pluripotent cells in the primitive streak shield themselves from the neural induction cues they produce is an outstanding question.

In vitro systems based on pluripotent stem cells enable to recapitulate crucial aspects of early post-implantation mammalian development (Shahbazi et al., 2019). Two-dimensional colonies of human embryonic stem cells (hESCs) could form germ layer fates in a spatially ordered manner (Warmflash et al., 2014), with the colony geometry playing an important role in determining the spatial extent of territories (Etoc et al., 2016). Primitive streak-like cells could also be induced from hESCs and they remarkably possessed the

activity of an organizer when transplanted into host embryos (Martyn et al., 2018). ESC aggregates allowed to differentiate in three dimensions led to derivatives of the three germ layers and axis formation both using mouse (van den Brink et al., 2014; Beccari et al., 2018) and human (Simunovic et al., 2019; Moris et al., 2020) cells. Somitogenesis could even be recapitulated when providing the right extracellular matrix environment (van den Brink et al., 2020; Veenvliet et al., 2020).

Using an mESC-based *in vitro* reconstitution system of neuroectoderm induction by primitive streak-like cells, we determined that RA was critical for neural induction by TGFβ and Wnt antagonists. Single-cell RNA sequencing demonstrated that fate specification unfolded up to somite and neuron fates. Using genetic and pharmacological interrogations in mESCs simultaneously reporting on different fates and the activation status of signaling pathways, we found that TGFβ and Wnt ligands stabilized a PS-like fate. Genetic evidence showed that the RA degrading enzyme Cyp26a1 protected PS-like cells from neural induction, both in wild type cells and notably in the absence of the antagonists Chordin and Noggin or Dkk1. Using a highly sensitive RA reporter, we identified the presence of RA signaling in conditions commonly thought to lack RA synthesis ability. Finally, TGFβ or Wnt inhibition induced different neural progenitor gene expression signatures and this correlated with different levels of RA signaling. Altogether, our results point at RA signaling playing a major role in neuroectoderm induction.

RESULTS

In vitro reconstitution of neuroectoderm induction by primitive streak-like cells

The possibility to generate primitive streak-like cells *in vitro* should lead to the induction of a neuroectoderm fate among differentiation-competent cells (Figure 1A). To monitor the induction of neural progenitors by PS-like cells, we used a double knock-in (2KI) reporter mESC line with *Sox1* locus targeted with GFP and *T* (also known as *Brachyury*) locus targeted with H2B-3xTagBFP that we previously characterized (Sladitschek and Neveu, 2019). While *T* is a commonly used as a mesoderm marker, it is also expressed in the PS (Wilkinson et al., 1990). *Sox1* itself marks exclusively neural progenitors (Pevny et al., 1998). We previously showed that differentiation with the small molecule IDE1, which activates the TGFβ pathway (Borowiak et al., 2009), led to intermediates that resembled the primitive streak (PS) in mouse embryos (Sladitschek and Neveu, 2019). Interestingly, we could find in these cultures Sox1^{GFP+} cells along with T$^{TagBFP+}$ cells, the candidate PS-like cells (Figure 1B).

We assessed the composition of the differentiating culture every day by flow cytometry. The majority of

the population was $T^{TagBFP+}$ by the third day (Figure S1A), when T mRNA levels greatly increased as measured by mRNA-Seq (Figure S1B). In addition, we found that $Sox1^{GFP+}$ cells could be detected in increasing numbers from day 4 (Figure S1C), mirroring the gradual increase in Sox1 mRNA levels (Figure S1D). In order to identify factors that influenced the formation of $Sox1^{GFP+}$ cells, we initiated the differentiation with variable cell numbers to test the role of cell density. Increasing starting cell numbers enhanced the fraction of $Sox1^{GFP+}$ cells (Figure S1E). As secreted factors play a major role in neural induction *in vivo* (De Robertis, 2006), we wondered whether medium conditioned by differentiating cultures could induce the formation of $Sox1^{GFP+}$ cells. This was indeed the case when this conditioned medium was applied to mESCs (Figure S1F), suggesting that paracrine signaling was an important factor to induce neuroectoderm in our *in vitro* system.

To determine the identity of the different cell populations, we FACS-purified cells expressing TagBFP and GFP and measured their transcriptome by deep sequencing (Figure 1C). $T^{TagBFP+}$ cells expressed markers associated with the PS such as T, Eomes and Goosecoid (Figure 1D). More importantly, they expressed secreted antagonists that are commonly associated with the *in vivo* organizer. Among these were the TGFβ antagonists Chordin (Chrd) and Noggin (Nog) and the Wnt antagonist Dkk1. Thus, the gene expression signature of $T^{TagBFP+}$ cells identified them as a PS-like population. $Sox1^{GFP+}$ cells expressed a diverse range of neural progenitor markers, among which Sox2, Pax6 and Nestin (Figure 1D), confirming their neuroectoderm identity.

The inhibition of TGFβ or Wnt signaling was shown to be critical for the induction of the neuroectoderm fate in the embryo (De Robertis, 2006). We sought to recapitulate in our *in vitro* model system the known action of these molecules by applying inhibitors to the differentiating cultures once the $T^{TagBFP+}$ population was established at day 3 (Figure 1E). Adding a small molecule antagonist of the TGFβ pathway SB431542 or blocking Wnt signaling using the tankyrase inhibitor XAV939 led to an increase in $Sox1^{GFP+}$ cells (Figure 1F). Furthermore, induction of $Sox1^{GFP+}$ cells could also be recapitulated using recombinant versions of the secreted antagonists Noggin, DKK1 or secreted frizzled-related protein sFRP1(Figure S1G). Thus, inhibitors of either signaling pathway were potent inducers of neuroectoderm fate in our *in vitro* system.

PS-like cells induce neural progenitors expressing a diverse range of markers

The expression of markers characteristic of forebrain, hindbrain and spinal cord fates (Gouti et al., 2014) in the $Sox1^{GFP+}$ population suggested that it constituted a heterogenous population (Figure 1D) . The fact that antagonizing either TGFβ or Wnt signaling pathway can independently trigger a neuroectoderm fate might

explain the diverse markers that were expressed in neural progenitors (NPs). We therefore differentiated mESCs to neural progenitors by inhibiting these two pathways individually and in addition with the well-established neural inducer retinoic acid (RA) (Ying et al., 2003). We FACS-purified the resulting Sox1^{GFP+} populations to characterize their gene expression by mRNA-Seq (Figure 2A). The resulting expression profiles were then compared with the neural progenitors arising in the PS-like differentiation. NPs obtained after the three differentiation regimes expressed distinct sets of transcription factors that spanned the set upregulated in the heterogenous PS-induced NP population (Figure 2B). Wnt inhibition led to the expression of anterior neuroectoderm markers such as Lhx5, Otx2 and Six3. More posterior markers were found upon TGFβ inhibition. Interestingly, these two different NP-induction methods did not account for the full complexity of the expression profile of NPs induced in the PS-like differentiation. Indeed, upregulation of markers present in PS-induced NPs such as Brn1, Brn2 or Irx3, could only be recapitulated by RA treatment (Figure 2B). This led us to hypothesize that RA signaling might be in part responsible for NP-induction by PS-like cells in our self-organizing system.

Neuroectoderm induction necessitates RA signaling

To directly monitor RA signaling, we stably inserted a fluorescent reporter construct relying on the well-established DR5-based RA response element (RARE) (Rossant et al., 1991) controlling the expression of the fluorescent protein Scarlet in the 2KI reporter line. Scarlet positive cells could be detected from day 3 onwards (Figure 2C and S2A), demonstrating the presence of retinoic acid signaling during this differentiation procedure. This paralleled the finding that a reporter relying on the same RARE showed RA signaling in the primitive streak region of wild type E7.5 mouse embryos (Rossant et al., 1991). As RA is generally used as an efficient inducer of neural fate *in vitro* (Ying et al., 2003), we hypothesized that it might be involved in generating neural progenitors in our differentiation system. Adding RA to the differentiation medium at day 3 led to the conversion of the population of T$^{TagBFP+}$ cells into Sox1^{GFP+} cells (Figures S2B and S2C). This suggested that all T$^{TagBFP+}$ cells could be converted into neuroectoderm. Similarly, adding to the differentiation medium increasing amounts of vitamin A (also known as retinol), the precursor of RA, enhanced the formation of Sox1^{GFP+} cells (Figure S2D). This was accompanied by an increase in RA signaling, as captured by our RA signaling reporter line (Figure S2E).

To probe the extent of RA involvement in NP formation in our self-organizing system, we blocked RA receptors (RARs) with the small molecule inhibitor AGN193109 (AGN, Figure 2D). Surprisingly, pharmacological inhibition of RARs resulted in a strong reduction of the fraction of Sox1^{GFP+} cells (Figure 2E).

Furthermore, the vast majority of the Sox1^{GFP+} population could be abrogated by initiating RAR inhibition at early differentiation time points (Figure S2F). The magnitude of this effect led us to postulate that RA signaling might be involved in the mechanism of neural induction downstream of TGFβ or Wnt inhibition. We therefore applied simultaneously the RAR antagonist AGN with either TGFβ or Wnt inhibitors (Figure 2D). Surprisingly, AGN prevented the increase of the Sox1^{GFP+} population normally associated with TGFβ or Wnt inhibition (Figures 2F and 2G). This indicated that RA signaling mediated neuroectoderm formation elicited by TGFβ or Wnt inhibition.

Time resolved scRNA-Seq characterization of PS-like differentiation

A defining feature of our PS-like differentiation protocol was the coexistence of different fates, necessitating a finer characterization of the cellular heterogeneity than bulk mRNA-Seq. We therefore conducted single-cell RNA sequencing on differentiating cultures between day 2 and day 5 (Figure 3A). In total, we obtained expression profiles for 46,700 cells. Uniform manifold approximation and projection (UMAP) analysis showed only minimal overlap between consecutive days (Figures 3B and S3A). Profiles of FACS-purified T$^{TagBFP+}$/Sox1^{GFP-} and Sox1^{GFP+} cells projected according to the respective expression territories of T and Sox1 (Figures 3C and S3B). Notably, cells with high T expression and Sox1-expressing cells appeared to form distinct populations (Figure 3C). We could identify 35 cell subpopulations (Figures 3D and S3C) that were stratified by a number of markers (Figure S3D). Naïve pluripotency markers were found in a subpopulation of day 2 cells while the other cells at this stage initiated the expression of primed epiblast markers (Figure S3E). Pluripotency factors had distinct behaviors: while Pou5f1 (also known as Oct4) expression was retained till day 4, Nanog expression was reactivated in the PS-like population (Figure S3F). Sox2 was downregulated in PS-like cells and derivatives and its expression was maintained in NPs (Figure S3F). These findings faithfully recapitulated the expression patterns of these genes observed in E7.0 mouse embryos (Peng et al., 2019). PS markers were expressed in different subpopulations (Figure 3E), corresponding to different regions of the *in vivo* PS. Interestingly, PS-derivatives were also formed in our cultures. Indeed, presomitic mesoderm and distinct subsequent somite fates could be found at days 4 and 5 (Figures 3F and S3G). Moreover, endoderm and notochord fates were present by day 5 (Figure 3G). The existence of several NP populations (Figure S3D and S3H) was consistent with our bulk mRNA-Seq results (Figure 2B). Thus, our *in vitro* system recapitulated the fate diversification occurring in post-implantation embryos and notably the temporal evolution of the primitive streak *in vivo*.

Wnt and TGFβ pathways maintain mesendoderm fates

We next sought to identify how such self-organization of the different fates arises. The TGFβ, Wnt and RA signaling pathways had crucial importance in inducing neuroectoderm. Having determined that RA could convert the PS-like state into neuroectoderm (Figure S2C), we investigated the role of the TGFβ and Wnt signaling pathways in creating, maintaining or destabilizing the PS-like state. We used pharmacological activation of one of the two pathways in conjunction with the inhibition of the other one and assessed the resulting fate as measured by our 2KI line (Figure 4A). Both Activin A or activating Wnt signaling with the GSK3 inhibitor CHIR99021 (CHIR) could drive the generation of $T^{TagBFP+}$ cells (Figure 4B). Blocking TGFβ signaling with concomitant Wnt activation did not prevent the generation of $T^{TagBFP+}$ cells, whereas blocking Wnt signaling completely ablated the $T^{TagBFP+}$ population despite the presence of the TGFβ activating molecule IDE 1 (Figure 4B). Thus, while both signaling pathways were sufficient in triggering the formation of $T^{TagBFP+}$ cells, only Wnt signaling was necessary for it.

To monitor whether Wnt signaling was active during PS-like differentiation, we introduced a fluorescent reporter of canonical Wnt signaling relying on TCF/LEF binding sites (Molenaar et al., 1996) in the 2KI line (Figure S4A). Wnt pathway activity could be detected in the $T^{TagBFP+}$ population from day 2 onwards (Figures 4C and S4B). It peaked at day 3 to gradually decrease as differentiation progressed (Figure S4B).

We next investigated how the activation of the TGFβ and Wnt pathways impacted the induction of neuroectoderm (Figure 4D). Activation of TGFβ pathway by addition of Activin A (Figure 4E) or either of the two BMP ligands BMP2 and BMP4 (Figure S4C and S4D) led to the disappearance of the $Sox1^{GFP+}$ population. Wnt activation using CHIR or WNT3A impaired the induction of $Sox1^{GFP+}$ cells (Figures 4F and S4E).

Several Wnt ligands (Wnt3, Wnt5a, Wnt5b), along with the Wnt target gene T (Arnold et al., 2000), were among the genes upregulated in the presence of CHIR in mESCs (Figure S4F). We targeted the *Wnt5a* locus with H2B-Citrine or H2B-tdTomato to characterize the Wnt5a-expressing population (Figure S4G) in relation with the status of Wnt signaling or with the PS and neuroectoderm fate markers (Figure S4H). The fraction of $Wnt5a^{Citrine+}$ cells increased as differentiation progressed (Figure S4I). Interestingly, these cells corresponded to the ones with the highest Wnt activity, as measured with the TCF/LEF-Scarlet reporter (Figure S4J). *Wnt5a*-Citrine expression could furthermore be blocked by Wnt inhibition, showing that Wnt5a expression depended on canonical Wnt signaling (Figure S4J). Ultimately, the $T^{TagBFP+}$-expressing population could be stratified by the *Wnt5a*-tdTomato reporter (Figure S4K and S4L): tdTomato was detected in cells with the highest T^{TagBFP} expression.

Endogenous TGFβ and Wnt ligands stabilize the PS-like state

As the activation of either TGFβ or Wnt signaling was able to prevent the induction of neuroectoderm, we wondered how cells would react to genetic interference with their ligand levels, by impairing the production of either their ligands or their antagonists. We turned to our scRNA-Seq dataset to determine the ligands that might be the most relevant. While Nodal stood out among the family of TGFβ and BMP ligands for its expression in PS-like cells, Wnt5a was widely expressed in PS-like subpopulations as well as their mesendoderm derivatives (Figure S4M). We therefore deleted *Nodal* or *Wnt5a* in the 2KI mESC line (Figure S4N and S4O) and differentiated the resulting knockout lines (Figure 4G). Upon PS-like differentiation, *Nodal*$^{-/-}$ cells could not form a T$^{TagBFP+}$ population and differentiated almost exclusively into neuroectoderm cells (Figure 4H). Similarly, *Wnt5a*$^{-/-}$ mESCs generated a vast majority of Sox1^{GFP+} cells (Figure 4H). Thus, the absence of either Nodal or Wnt5a was sufficient to destabilize the PS-like population and favored a neuroectoderm fate. Having shown the stabilizing effect of the expression of TGFβ or Wnt ligands on the PS-like population, we hypothesized that the loss of their antagonists might lead to a failure to induce neuroectoderm. We derived *Chrd*$^{-/-}$*Nog*$^{-/-}$ mESCs and *Dkk1*$^{-/-}$ mESCs in the 2KI background (Figures S4P and S4Q). Knockout cells were differentiated and were severely impaired in generating a Sox1^{GFP+} population (Figure 4I). Therefore, a carefully orchestrated balance between the endogenous production of ligands and antagonists of the TGFβ or Wnt signaling pathways was critical in order to either sustain the PS-like fate or destabilize it to promote neural fate.

Molecular characterization of PS-like fates induced by TGFβ and Wnt activation

TGFβ and Wnt ligands had the same overall effect of preventing the formation of neuroectoderm fate. However, our reporter line did not allow to resolve the exact fate that was stabilized (or induced) by these pathways. We thus differentiated mESCs with Activin A or CHIR and compared the resulting expression profiles with the ones of T$^{TagBFP+}$ cells induced by PS-like differentiation or with the ones of Epiblast Stem Cells (EpiSCs) whose maintenance relies on Activin A and FGF2 (Brons et al., 2007; Tesar et al., 2007) (Figure 4J). Principal component analysis revealed that TGFβ and Wnt activation led to two orthogonal gene expression signatures (Figure 4K). During differentiation, PS-like cells progressed from a state resembling the Activin A-treated cells towards a state in which Wnt signaling was more present (Figure 4K). In order to further support the interpretation of the temporal trajectory of the PS-like population, we turned to the *Wnt5a*-Citrine reporter line as a marker of relatively high Wnt signaling (Figure S4R). The expression profiles of the FACS-purified Wnt5a$^{Citrine+}$ subpopulations systematically projected to the right of the

unsorted population profiles on the principal component representing Wnt activation and were also more advanced in the differentiation trajectory (Figure S4S). We went on to assess the impact of the antagonist knockouts on the nature of the PS-like population (Figure 4L). The gene expression profiles of $T^{TagBFP+}$ cells obtained after PS-like differentiation of *Chrd*$^{-/-}$*Nog*$^{-/-}$ mESCs or *Dkk1*$^{-/-}$ mESCs projected towards the CHIR-differentiated samples (Figure 4M). For both cell lines, the Wnt signaling pathway was among the top pathways with an enrichment of differentially expressed genes as compared to wild type cells. Examining expression profiles in these samples showed a downregulation of PS and endoderm markers as well as an upregulation of presomitic mesoderm and somite markers in the $Wnt5a^{Citrine+}$, *Chrd*$^{-/-}$*Nog*$^{-/-}$ or *Dkk1*$^{-/-}$ cells as compared to the Activin A-differentiated cells (Figure 4N). Moreover, it could be inferred that Wnt signaling was particularly active in the presomitic mesoderm populations as captured by the expression of the Wnt target gene Lef1 in our scRNA-Seq data (Figure S4T). Altogether, the absence of TGFβ or Wnt antagonists led to effects comparable to the ones elicited by elevated Wnt signaling, favoring presomitic mesoderm and somite fates.

Aldh1a2-independent RA signaling

While TGFβ and Wnt signaling were critical in maintaining a mesendoderm fate among the PS-like subpopulation, we found that RA signaling was crucial in mediating neuroectoderm induction downstream of the inhibition of these two pathways (Figure 2F and 2G). Only its precursor vitamin A was present in our differentiation medium and RA therefore had to be synthesized by the cells themselves. The synthesis of RA from retinal is performed by the retinaldehyde dehydrogenases Aldh1a1, Aldh1a2 or Aldh1a3 (Rhinn and Dollé, 2012). According to our mRNA-Seq data, Aldh1a2 was upregulated during PS-like differentiation, while Aldh1a1 and Aldh1a3 expressions did not exceed background levels (Figure S5A). Aldh1a2 levels were particularly elevated in $T^{TagBFP+}$ cells compared to $Sox1^{GFP+}$ cells (Figure S5B). This mirrored the known pattern of expression of these three genes in post-implantation mouse embryos, particularly the expression of Aldh1a2 in the primitive streak at E7.5 (Ribes et al., 2009).

The presence of forebrain structures in *Aldh1a2*$^{-/-}$ mouse embryos in the absence of detectable expression of a DR5-based RA reporter at E7.5 led to the conclusion that RA signaling was not important for the generation of anterior structures *in vivo* (Niederreither et al., 1999). This result based on mouse genetics seemed at first glance in contradiction with our finding that RA signaling played a critical role in the induction of neuroectoderm by TGFβ and Wnt signaling inhibition. We therefore sought to recapitulate *in vitro* the *Aldh1a2*$^{-/-}$ phenotype and to investigate whether RA signaling was completely abolished upon *Aldh1a2*

ablation. We generated *Aldh1a2*$^{-/-}$ mESCs (Figure S5C) in the 2KI line with the RA signaling reporter based on the same DR5 used in the mouse experiments. PS-like differentiation of *Aldh1a2*$^{-/-}$ mESCs led to the formation of Sox1^{GFP+} cells despite the absence of DR5-RARE-Scarlet+ cells in the culture (Figure 5A and S5D). This result was in accordance with the phenotype of *Aldh1a2*$^{-/-}$mouse embryos (Niederreither et al., 1999). In order to assess whether RA signaling was truly absent, we interfered with RA signaling, either by providing its precursor vitamin A or antagonizing RARs, during PS-like differentiation of *Aldh1a2*$^{-/-}$ cells (Figure 5B). In the absence of RA synthesis, these treatments should have no effect, that is the fraction of Sox1^{GFP+} cells should be insensitive to perturbation to RA signaling. Surprisingly, the RAR antagonist AGN led to a decrease of the fraction of Sox1^{GFP+} population (Figure 5C). Addition of vitamin A increased greatly the fraction of Sox1^{GFP+} cells (Figure 5C) and DR5-RARE-Scarlet+ cells could now be detected in the *Aldh1a2*$^{-/-}$ differentiating cultures (Figure S5D). Altogether, this demonstrated that the genetic ablation of Aldh1a2 was not sufficient to fully abrogate RA signaling. The DR5-RARE reporter used in previous studies was however unable to capture these residual RA levels in *Aldh1a2*$^{-/-}$ cultures.

In order to determine whether *Aldh1a2*$^{-/-}$ cells were still able to respond to the RA levels produced by wild type cultures, we differentiated them together by mixing wild type *Aldh1a2*$^{+/+}$ cells and mutant *Aldh1a2*$^{-/-}$ cells (Figure S5E). The two genotypes could be distinguished by the constitutive expression of an additional fluorescent protein H2B-2xiRFP670. Under such conditions, DR5-RARE-Scarlet+ cells could be observed in the *Aldh1a2*$^{-/-}$ fraction at a rate comparable to the one in *Aldh1a2*$^{+/+}$ cells (Figures S5F and S5G). This showed that RA signaling occurred in a paracrine fashion in our *in vitro* system.

A highly sensitive RA reporter captures Aldh1a2-independent RA signaling

Our results stressed that the traditional DR5-based RARE might not be sensitive enough to RA signaling, thus capturing only a subset of conditions in which RA signaling was present. It was found that RARs had a much higher affinity for a composite RARE compared to the traditional DR5 binding site (Moutier et al., 2012). This composite RARE relied on three RAR binding sites forming direct repeats with three different spacings (Figure 5D). We developed a fluorescent reporter construct relying on this RARE (which we termed cDR-RARE) driving the expression of the fluorescent protein Scarlet and stably inserted it in the 2KI line (Figure 5D). The reporter could detect sub-nanomolar concentrations of exogenously applied RA (Figure S5H). Many more cDR-RARE-Scarlet+ cells could be detected during PS-like differentiation of wild type cells (Figure 5E) compared to the traditional DR5-based reporter (Figures S5I and S5J). As expected, Scarlet expression could be blocked by AGN (Figure S5K). Crucially, this reporter was still active

in *Aldh1a2*$^{-/-}$ cells (Figure 5F) and its expression was abrogated by AGN treatment (Figures S5L and S5M). Thus a fraction of RA in the differentiating cultures was produced in an Aldh1a2-independent manner and at sufficient levels both to be detected by a highly sensitive fluorescent reporter and to induce the formation of Sox1^{GFP+} cells.

Vitamin A availability regulates the levels of RA signaling during PS-like differentiation

Cells were sensitive to the external vitamin A concentration, as proven by the fact that providing increasing amounts of external vitamin A led to increasing numbers of Sox1^{GFP+} cells (Figure S2D). Interestingly during PS-like differentiation, the transcript levels of the retinol binding protein Rbp1 and the Rbp-receptor Stra6, a major mediator of the cellular uptake of vitamin A (Kawaguchi et al., 2007), were upregulated (Figures S5N and S5O). Moreover, Rbp1 expression levels were elevated in the primitive streak of E7.0 mouse embryos as determined by spatial transcriptomics (Peng et al., 2019). We hypothesized that vitamin A uptake through Stra6 and intracellular storage by Rbp1 could be an important parameter in controlling RA levels. We therefore generated double knockout *Rbp1*$^{-/-}$*Stra6*$^{-/-}$ mESCs (Figure S5P) and submitted them to PS-like differentiation. These cells were impaired in generating Sox1^{GFP+} cells compared to wild type cells (Figure 5G). However, this phenotype could be reversed by increasing the concentration of vitamin A in the differentiation medium (Figure 5G). We next asked whether disrupting the cellular supply of vitamin A would further decrease the RA signaling levels in *Aldh1a2*$^{-/-}$ cells that were impaired in RA synthesis. Upon PS-like differentiation, triple knockout *Aldh1a2*$^{-/-}$*Rbp1*$^{-/-}$*Stra6*$^{-/-}$ cells presented indeed reduced RA signaling as measured by the cDR-RARE-based fluorescent reporter and a decreased fraction of Sox1^{GFP+} cells compared to *Aldh1a2*$^{-/-}$ cells (Figures 5H and 5I). This demonstrated that the Stra6-Rbp1 axis controlling the intracellular vitamin A levels was pivotal in regulating the eventual levels of RA signaling.

Cyp26a1 limits RA signaling during PS-like differentiation

The observation that RA signaling was enabled by the upregulation of key genes during PS-like differentiation raised the question of how PS-like cells shielded themselves from the neural inducing effects of RA. One potential mechanism would be to control RA degradation, which is mediated by cytochrome P450 Cyp26 enzymes (Rhinn and Dollé, 2012). Cyp26a1 mRNA levels were highly upregulated during PS-like differentiation (Figure S6A), and particularly in the T$^{TagBFP+}$ cells as compared to Sox1^{GFP+} cells (Figure

S6B). Having shown that Wnt signaling was critical in establishing the PS-like population, we investigated whether Cyp26a1 expression could be Wnt-dependent using our *Wnt5a* KI line to purify cells with elevated Wnt signaling. There was a 10-fold upregulation of Cyp26a1 mRNA levels upon differentiating mESC with CHIR (Figure S6C) but also in the Wnt5a$^{Citrine+}$ cells isolated when the PS-like population arose at day 3 (Figure 6A). Cyp26a1 upregulation in our culture system was mirrored *in vivo* with its expression territory corresponding to the primitive streak of E7.0 mouse embryos (Fujii et al., 1997). We hypothesized that Cyp26a1 levels could hamper neural induction by limiting RA levels. We therefore deleted Cyp26a1 in the 2KI-RA reporter line allowing us to read out its effects on RA signaling and neuroectoderm formation (Figure S6D). As expected, *Cyp26a1*$^{-/-}$ cells formed more Sox1^{GFP+} cells and had elevated RA signaling as measured with the RA activity reporter (Figure 6B). Furthermore, the loss of Cyp26a1 enhanced neuroectoderm induction at the expense of the PS-like subpopulation in the presence of elevated extracellular vitamin A (Figure S6E). We wondered whether this increase in RA signaling would be limited to cell autonomous effect, that is whether RA signaling depended on a cell's own production of RA. To do so, we mixed *Cyp26a1*$^{-/-}$ cells displaying enhanced RA signaling with *Aldh1a2*$^{-/-}$ cells with impaired RA synthesis (Figure S6F). There was decreased RA signaling in *Cyp26a1*$^{-/-}$ cells in the co-culture setting as compared to a pure *Cyp26a1*$^{-/-}$ population (Figures S6G and S6H). Altogether, Cyp26a1 expression played a key role in limiting RA signaling and the acquisition of neural fate during PS-like differentiation.

Cyp26a1 expression stabilizes the PS-like state

We showed that RA mediated neuroectoderm induction and that removing the RA-degrading enzyme Cyp26a1 increased the fraction of neuroectoderm cells during PS-like differentiation at the expense of the PS-like population. On the contrary, null mutations of *Chrd* and *Nog* or *Dkk1* impaired the formation of neural progenitors. We wondered whether the absence of TGFβ or Wnt secreted inhibitors could be counteracted by an increase in RA signaling due to the loss of Cyp26a1. We thus generated mESCs lacking Cyp26a1 and either TGFβ or Wnt antagonists and subjected them to PS-like differentiation (Figures 6C, S6I and S6J). Triple knockout *Chrd*$^{-/-}$*Nog*$^{-/-}$*Cyp26a1*$^{-/-}$ and double knockout *Dkk1*$^{-/-}$*Cyp26a1*$^{-/-}$ cells generated a large fraction of Sox1^{GFP+} cells compared to *Chrd*$^{-/-}$*Nog*$^{-/-}$ cells or *Dkk1*$^{-/-}$ cells (Figures 6D, 6E and S6K). Notably, this increase could be reverted by the RAR antagonist AGN underlining that this effect was strictly dependent on RA signaling (Figures 6D and 6E). Cyp26a1-mediated dampening of RA signaling was thus critical in stabilizing the PS-like state, shielding these cells against the differentiating effects of RA they produce.

RA signaling status accounts for neural progenitor diversity

RA was a crucial stimulus triggering an NP fate. We sought to characterize the gene expression changes associated with impaired RA signaling or with active RA signaling both in $T^{TagBFP+}$ and $Sox1^{GFP+}$ subpopulations (Figure 7A). Expression changes observed upon AGN treatment were well correlated with the ones observed in *Aldh1a2*$^{-/-}$ cells (Figure 7B) and were anti-correlated with the expression changes due to active RA signaling in $T^{TagBFP+}$ cells (as captured with the cDR-RARE-Scarlet reporter) (Figure 7C). Active RA signaling could either repress (Figure S7A) or increase (Figure S7B) gene expression. Interestingly, there was no correlation between the changes in gene expression associated with RA signaling and the ones between $Sox1^{GFP+}$ and $T^{TagBFP+}$ cells (Figure S7C). This suggested that RA was involved in early steps of neuroectoderm commitment and not responsible for the final expression signature of $Sox1^{GFP+}$ cells.

We identified genes with differential regulation in the $Sox1^{GFP+}$ cells distinguished by their expression of the cDR-RARE-Scarlet reporter. We then assessed the expression of these genes in the eight different NP subpopulations identified by scRNA-Seq (Figures S3C and 7D). NP-2, NP-3, NP-4 and NP-5 resembled the $Sox1^{GFP+}$ cells with active RA signaling, whereas the NP-6, NP-7 and NP-8 classes shared a signature with cDR-RARE-Scarlet negative cells (Figure 7D). Among the differentially expressed transcription factors were NP markers associated with different antero-posterior identity such as Zic1 and Hoxb9 (Figure 7E). The NP-1 subpopulation in particular presented elevated expression levels of genes characteristic of anterior NP identity (Figure S7D and S7E). These markers were upregulated upon AGN treatment as compared with the cells in which RA signaling was not detected (Figure S7F). Altogether, differences in RA signaling activation status could account for the particular signatures of the NP subpopulations identified by scRNA-Seq, broadly dividing them into two regions (Figure 7F).

RAR knockout cells are insensitive to changes in NP-inducing signals

RA receptors (RARs) are the transcriptional effectors of RA signaling (Chambon, 1996). To obtain a condition where RA signaling cannot be transduced, we derived mESCs lacking all three RARs (RAR-null cells) (Figure S7G). As expected and contrary to wild type cells, RA failed to induce the expression of the DR5-RARE-based reporter in *Rara*$^{-/-}$ *Rarb*$^{-/-}$ *Rarg*$^{-/-}$ mESCs (Figures S7H and S7I). Nevertheless a $Sox1^{GFP+}$ population arose after PS-like differentiation of *Rara*$^{-/-}$ *Rarb*$^{-/-}$ *Rarg*$^{-/-}$ cells. We therefore sought to test whether the induction of neural fates was sensitive to alterations of RA signaling or to TGFβ and Wnt inhibition (Figure 7G). The fraction of $Sox1^{GFP+}$ cells appeared insensitive to the addition of vitamin A or a RAR antagonist (Figure 7H). Similarly, TGFβ and Wnt inhibition failed to increase the fraction of neuroectoderm

cells in RAR-null cells (Figure 7I), whereas the expected downregulation of T^{TagBFP} upon Wnt inhibition was still present (Figure S7J). Thus, neuroectoderm induction in the RAR-null cells was independent from TGFβ, Wnt and RA signaling. Functional RARs were necessary to implement the neuroectoderm-inducing action of TGFβ and Wnt inhibition.

We profiled the gene expression of the $T^{TagBFP+}$ and $Sox1^{GFP+}$ subpopulations of *Rara*$^{-/-}$ *Rarb*$^{-/-}$ *Rarg*$^{-/-}$ cells and compared them with the ones of wild type cells or of cells treated with the RAR antagonist AGN. AGN-treated $T^{TagBFP+}$ cells and $T^{TagBFP+}$ RAR-null cells presented differences in transcription factor expression that might explain the formation of $Sox1^{GFP+}$ cells in RAR-null cells (Figure 7J). In contrast to AGN-treated wild type cells, $T^{TagBFP+}$ RAR-null cells failed to express Tbx6, a gene whose loss leads to the formation of neural tissue at the expense of somites *in vivo* (Chapman and Papaioannou, 1998). Notably, these cells presented an upregulation of Sox2, whose misexpression in paraxial mesoderm leads to neural tube formation (Takemoto et al., 2011). $Sox1^{GFP+}$ RAR-null cells possessed as well a distinct expression profile (Figure S7K). Hox genes and members of the Cdx family were downregulated in these cells whereas a number of transcription factors characteristic of neural fate (such as Neurog2, Neurod4 or Nhlh1) had elevated expression. Thus, the lack of RARs had distinct effects compared to their pharmacological inhibition and could not be assimilated to an absence of RA signaling.

DISCUSSION

Combining differentiation to a primitive streak-like fate with a large collection of reporter mESC lines harboring genetic ablations of key signaling factors, we identified that retinoic acid (RA) signaling was a critical mediator of neural induction triggered by TGFβ or Wnt inhibition.

A wide array of fates co-existed in our culture differentiated in two dimensions despite the absence of defined geometrical constraints. This suggests that neither long-range order nor a three-dimensional arrangement were necessary for the self-organization of different fates. Local interactions between cells were nevertheless crucial because the entire culture was exposed to the same differentiation medium, that is cell-generated signaling microenvironments created fate diversity. However, geometrical confinement plays a critical role in bringing spatial organization to the tissue as demonstrated by the seminal study of Warmflash et al. (2014). Similarly, it was shown that the extracellular matrix provides the appropriate mechanical cues to allow proper axis elongation in a three-dimensional environment (van den Brink et al., 2020; Veenvliet et al., 2020). This uncoupling between fate specification and tissue-scale patterning/morphogenesis has

interesting implications. Indeed, complex tissue-like structures could fail to form because the appropriate cell fates are not present (for example in the absence of a conducive signaling environment) or because the mechanical environment is not appropriate (a key difference between gastruloids and trunk-like structures).

The flexibility of our *in vitro* system allowed us to manipulate the external environment in a controlled manner and study genotypes that would be difficult or impossible to generate *in vivo*. Whereas both TGFβ and Wnt signaling pathways were sufficient to induce a PS-like fate as expected from previous findings (Gadue et al., 2006), only Wnt signaling was necessary. The signaling requirements for the development of a PS-like population in mouse cells mirrored the ones used by human cells (Martyn et al., 2018). This highlights the conservation of the mechanism of PS formation between these two species despite the differences in their starting populations, mESCs being in a state of naïve pluripotency and hESCs in a state of primed pluripotency. As differentiation progressed, Wnt signaling pushed cells towards a presomitic mesoderm and somite fate while Activin signaling favored other mesendoderm derivatives (definitive endoderm and notochord).

Our *in vitro* system mirrored in multiple instances phenotypes exhibited by mutant mice. Profiling gene expression in the PS-like population obtained from *Chrd*$^{-/-}$ *Nog*$^{-/-}$ cells revealed elevated Wnt signaling compared to wild type cells. In these cells, the balance of mesendoderm fates was tipped towards acquiring a presomitic mesoderm fate at the expense of other fates like endoderm and notochord. This was in line with the relative contributions of Wnt and TGFβ signaling to determine these particular fates. Interestingly *Chrd*$^{-/-}$ *Nog*$^{-/-}$ embryos have severe defects in notochord and foregut development in addition to the spectacular forebrain truncation phenotype (Bachiller et al., 2000). The role of Wnt5a in maintaining the PS-like population revealed by the differentiation phenotype of *Wnt5a*$^{-/-}$ cells was in accordance with the defects in antero-posterior axis elongation of *Wnt5a*$^{-/-}$ mouse embryos (Yamaguchi et al., 1999). Similarly for Nodal, our *in vitro* results were coherent with the finding that *Nodal*$^{-/-}$ mice fail to form a primitive streak (Conlon et al., 1994). The impaired neuroectoderm formation in *Chrd*$^{-/-}$ *Nog*$^{-/-}$ or *Dkk1*$^{-/-}$ cells was in line with the corresponding mutant mouse embryos lacking forebrain structures (Bachiller et al., 2000; Mukhopadhyay et al., 2001). Finally, the destabilization of the PS-like population in *Cyp26a1*$^{-/-}$ cells due to excess RA signaling paralleled the axis truncation phenotype of *Cyp26a1*$^{-/-}$ embryos (Sakai et al., 2001). Therefore PS-like differentiation of mutant mESC lines largely recapitulated the phenotype of the respective mutant mouse embryos. In addition, our molecular characterization allowed to interpret these classical mouse genetic experiments in a new light.

The involvement of RA signaling in the generation of anterior neural structures was largely ruled out

because of the phenotype of embryos lacking *Aldh1a2* and the absence of RA reporter expression in these embryos (Niederreither et al., 1999). However, RA importance for the formation of more posterior neural structures is well-established (Durston et al., 1989). In addition, RA is widely used as instructive cue in neuronal differentiation of embryonic stem cells (Ying et al., 2003; Bibel et al., 2004). In contrast to the relative inaccessibility of the post-implantation mammalian embryo, our *in vitro* system allows the easy manipulation of the cells' signaling environment. The ability to perturb RA signaling in *Aldh1a2*$^{-/-}$ cells enabled us to discover residual RA signaling that could not be captured by a classical RA reporter. Developing a highly sensitive RA activity reporter, we indeed demonstrated the presence of RA signaling during PS-like differentiation of *Aldh1a2*$^{-/-}$ cells.

The different sensitivity of distinct RA response elements (RAREs) to RA levels would enable cells to switch on different gene repertoires depending on RA concentration. The strategy of using binding sites varying in affinity is adopted by the morphogen Bicoid in fly embryos (Driever et al., 1989). RA itself forms a concentration gradient during development (Shimozono et al., 2013) and such RAREs could be used to generate positional information by translating RA concentrations into distinct gene expression profiles. While RA was an important cue to specify neuroectoderm fate, TGFβ or Wnt inhibition produced NPs with distinct expression profiles. NPs obtained through Wnt inhibition expressed anterior markers, consistent with previous reports (Watanabe et al., 2005), while TGFβ inhibition produced more posterior fates. We further determined that the various NP populations induced by PS-like cells differed in their RA signaling status with markers corresponding to anterior fates being expressed in cells with low RA signaling. This finding is in accordance with the proposed caudalizing effects of RA during development (Durston et al., 1989).

The ability of RAR-null cells to form neural progenitors contrasted with the near complete block of NP induction in presence of a RAR antagonist. However this was not incompatible with our finding that RA mediated NP formation through TGFβ or Wnt inhibition: neither modulating RA signaling nor antagonizing TGFβ or Wnt signaling pathways changed the amount of NPs obtained with RAR-null cells. Indeed, RARs are generally bound to their cognate RAREs in the absence of ligand (Chambon, 1996), repressing target gene expression. RARs might act as roadblocks positioned at genomic loci important for neural specification and RA presence would lead to a ligand-mediated release of this repression. Their physical absence in RAR-null cells would remove these roadblocks. In addition, it would unmask binding sites, making them available to other transcription factors. Possible candidates are other members of the family of nuclear receptors (to which RARs belong) because of their promiscuity regarding the DNA sequences they

recognize (Mangelsdorf et al., 1995). A number of such nuclear receptors were expressed during PS-like differentiation.

In order to safeguard their developmental capabilities, pluripotent cells should protect themselves from the neuralizing action of RA and therefore need to carefully control the RA levels they are exposed to. This can in theory be achieved at several levels: either by controlling the availability of the substrate, vitamin A, the RA synthesis steps or RA degradation itself (Rhinn and Dollé, 2012). The mRNA levels of the retinaldehyde dehydrogenase Aldh1a2 were indeed strongly upregulated during PS-like differentiation and Aldh1a2-mediated synthesis was key to generate high levels of RA signaling. However, Aldh1a2 itself did not account for all the RA production. In *Aldh1a2*$^{-/-}$ cells, other aldehyde dehydrogenases produced nonetheless RA at sufficient levels to be detected by a more sensitive RA reporter. This Aldh1a2-independent RA synthesis was also sufficient to induce NP formation. In these cells, both RA signaling and neuroectoderm formation could be enhanced by increasing the amount of vitamin A. We determined that this residual RA synthesis could not be exclusively attributed to a single aldehyde dehydrogenase by systematically knocking out all the expressed ones (data not shown). Tuning RA synthesis can also be achieved by controlling the expression of retinol dehydrogenases that oxidize vitamin A into retinal (Rhinn and Dollé, 2012). Rdh10 could be such a candidate but its expression starts to be tissue-specific only at later developmental stages (Sandell et al., 2007). Interestingly, PS-like cells control the availability of vitamin A through the specific upregulation of Rbp1, in accordance with its expression in the mouse embryo (Ruberte et al., 1991; Peng et al., 2019). Controlling RA synthesis does not however shield the producing cells from the differentiation effects elicited by RA. Our co-culture experiments showed that the presence of cells with impaired RA synthesis reduced the response to RA in an RA-producing population. Thus, a cell's response was not intrinsically determined by its own RA production. Therefore, the ability to degrade RA offers an important checkpoint to control the penetrance of a population's response to RA. Indeed Cyp26a1-mediated degradation of RA was crucial in stabilizing the PS-like population, and notably in *Chrd*$^{-/-}$*Nog*$^{-/-}$ or *Dkk1*$^{-/-}$ cells. Cyp26a1 levels were specifically elevated in the PS in mouse embryos (Fujii et al., 1997; Peng et al., 2019). Altogether, cells exploited a three-tiered control of RA levels regulating precursor availability, RA synthesis and degradation in order to induce neural differentiation only in a subpopulation of cells.

In conclusion, we could identify the molecular logic underlying the self-organization of primitive streak and neuroectoderm fates using an easily manipulatable mESC-based *in vitro* system. Retinoic acid signaling was a crucial mediator of neural induction by TGFβ or Wnt inhibition. A key enabling technology was the development of an array of reporter lines that allowed the simultaneous readout of fate marker

expression and signaling status. The impact of genetic and pharmacological perturbations could then be characterized both at the gene expression level and at the fate specification level. Our results highlight the potential of studying the self-organization of fates in *in vitro* ESC-based systems to gain new insights about fundamental aspects of mammalian development.

Author Contributions

L.R. designed experiments, performed most experiments described in the manuscript, analyzed data and contributed to drafting the manuscript. H.L.S. provided critical preliminary data. P.A.N. conceived and supervised the study, performed experiments and analyzed data. P.A.N. wrote the paper and all authors commented on the manuscript.

Acknowledgements

We thank Lucia Cassella for advice on RNA-seq data analysis. We thank Laura Villacorta for help with scRNA-Seq sample processing. This work was technically supported by the EMBL Flow Cytometry Core and Genomics Core facilities. The study was funded by EMBL. L.R. was also supported by the EMBL International PhD Programme (EIPP).

Declaration of Interests

The authors declare no competing financial interests.

Figures

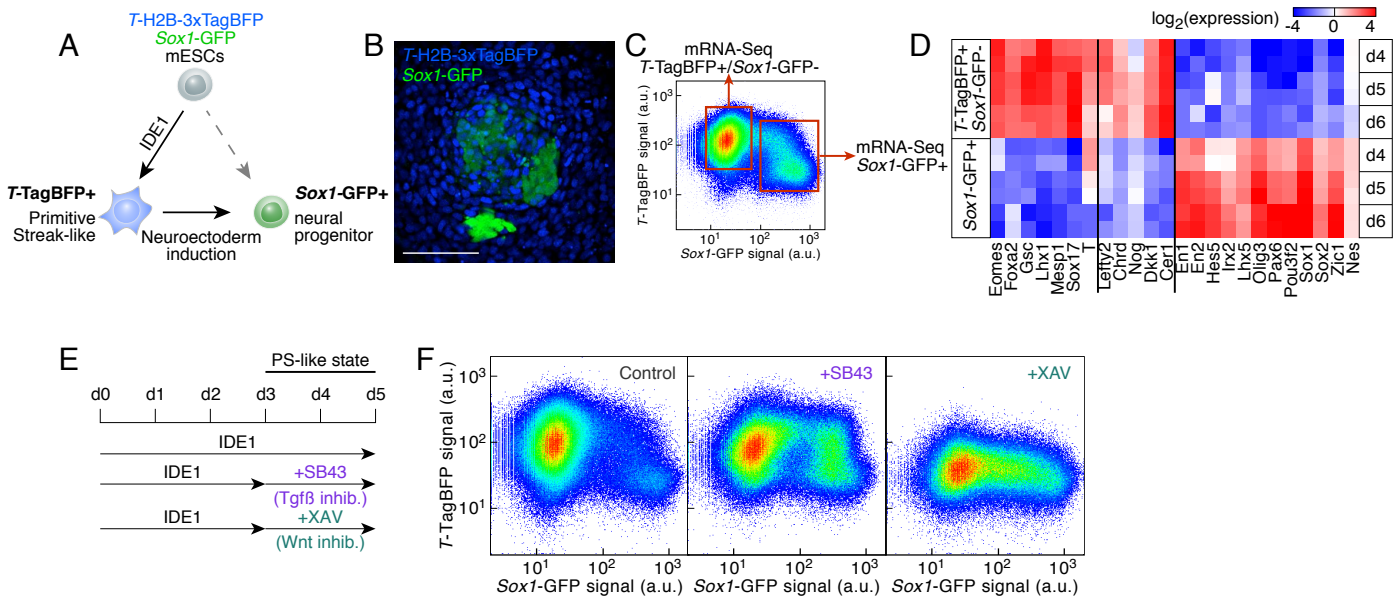

Figure 1. *In vitro* reconstitution of neuroectoderm induction by primitive streak-like cells.

(A) Scheme of the experimental principle to induce and monitor the formation of neuroectoderm by primitive streak-like cells using a double knockin mESC line reporting on *T* (also known as *Brachyury*) and *Sox1* expression.

(B) *T*-TagBFP and *Sox1*-GFP reporter expression after 5 days of PS-like differentiation. Bar: 100 μm.

(C) Experimental strategy to characterize the different populations present in the differentiating culture. Populations are FACS-purified according to the reporter fluorescence and gene expression is measured by mRNA-Seq.

(D) Expression levels (as measured by mRNA-Seq) of markers of the primitive streak or neural progenitors in FACS-purified populations after 4 to 6 days of differentiation.

(E) Experimental strategy to assess the impact of TGFβ or Wnt pathway inhibition on fate induction. SB431542 (SB43) inhibits TGFβ receptors and XAV939 (XAV) is a tankyrase inhibitor.

(F) *T*-TagBFP and *Sox1*-GFP reporter expression as measured by flow cytometry in control cultures (left panel) or in cultures treated with the TGFβ inhibitor (SB43, middle panel) or the Wnt pathway inhibitor XAV (right panel).

See also Figure S1.

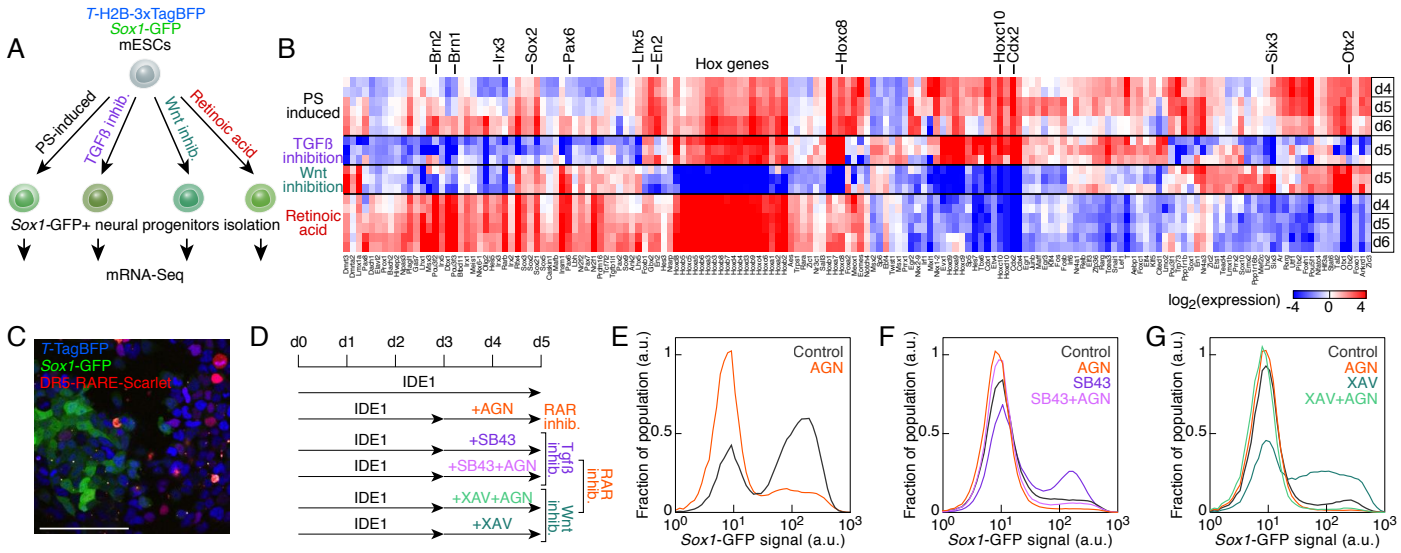

Figure 2. Functional retinoic acid signaling is critical for neuronal progenitor induction.

(A) Scheme of the experimental principle to characterize neural progenitors induced by PS-like cells, TGFβ or Wnt pathway inhibition or retinoic acid (RA) treatment.

(B) Expression levels of transcription factors and regulators differentially expressed in Sox1^{GFP+} cells .

(C) Reporter expression after 5 days of differentiation of a double knockin *T*-TagBFP and *Sox1*-GFP mESC line transgenic for a DR5-based retinoic acid signaling reporter. Bar: 100 μm.

(D) Experimental strategy to assess the crosstalk between RA signaling and TGFβ or Wnt pathway inhibition on fate induction. AGN193109 (AGN) is a retinoic acid receptor (RAR) antagonist.

(E) *Sox1*-GFP expression levels after PS-like induction with RA signaling inhibition (black: control, orange: AGN).

(F) *Sox1*-GFP expression levels after PS-like induction with TGFβ and/or RA signaling inhibition (black: control, orange: AGN, purple: SB43, pink: SB43 and AGN).

(G) *Sox1*-GFP expression levels after PS-like induction with Wnt and/or RA signaling inhibition (black: control, orange: AGN, teal blue: XAV, light green: XAV and AGN).

See also Figure S2.

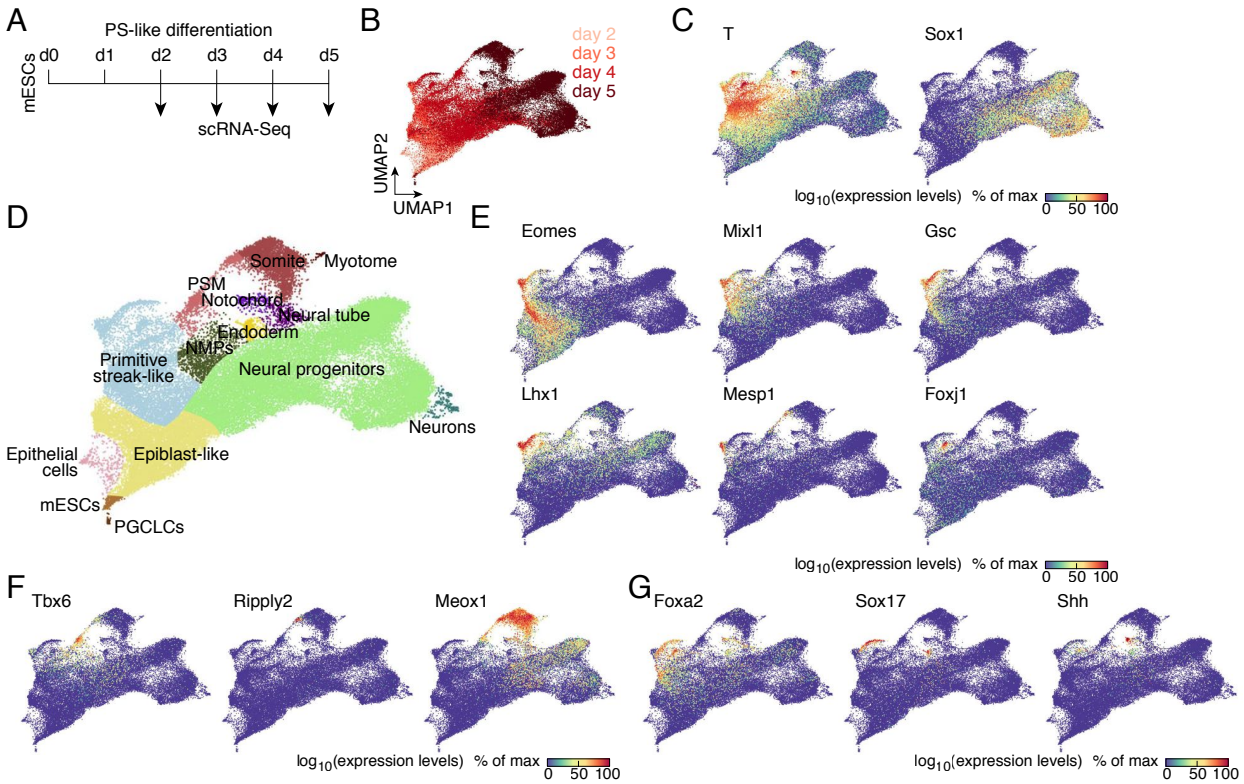

Figure 3. Time-resolved scRNA-Seq characterization of PS-like differentiation.

(A) Scheme of the experimental principle to temporally monitor PS-like differentiation using scRNA-Seq.

(B) UMAP (uniform manifold approximation and projection) of 46,700 cells colored by the day of collection.

(C) UMAP colored by the scaled expression of T and Sox1.

(D) UMAP colored according to the identified populations (NMPs: Neuromesodermal progenitors, PSM: presomitic mesoderm, PGCLCs: primordial germ cell-like cells).

(E) UMAP colored by the scaled expression of primitive streak markers.

(F) UMAP colored by the scaled expression of presomitic mesoderm and somite markers.

(G) UMAP colored by the scaled expression of endoderm and notochord markers.

See also Figure S3.

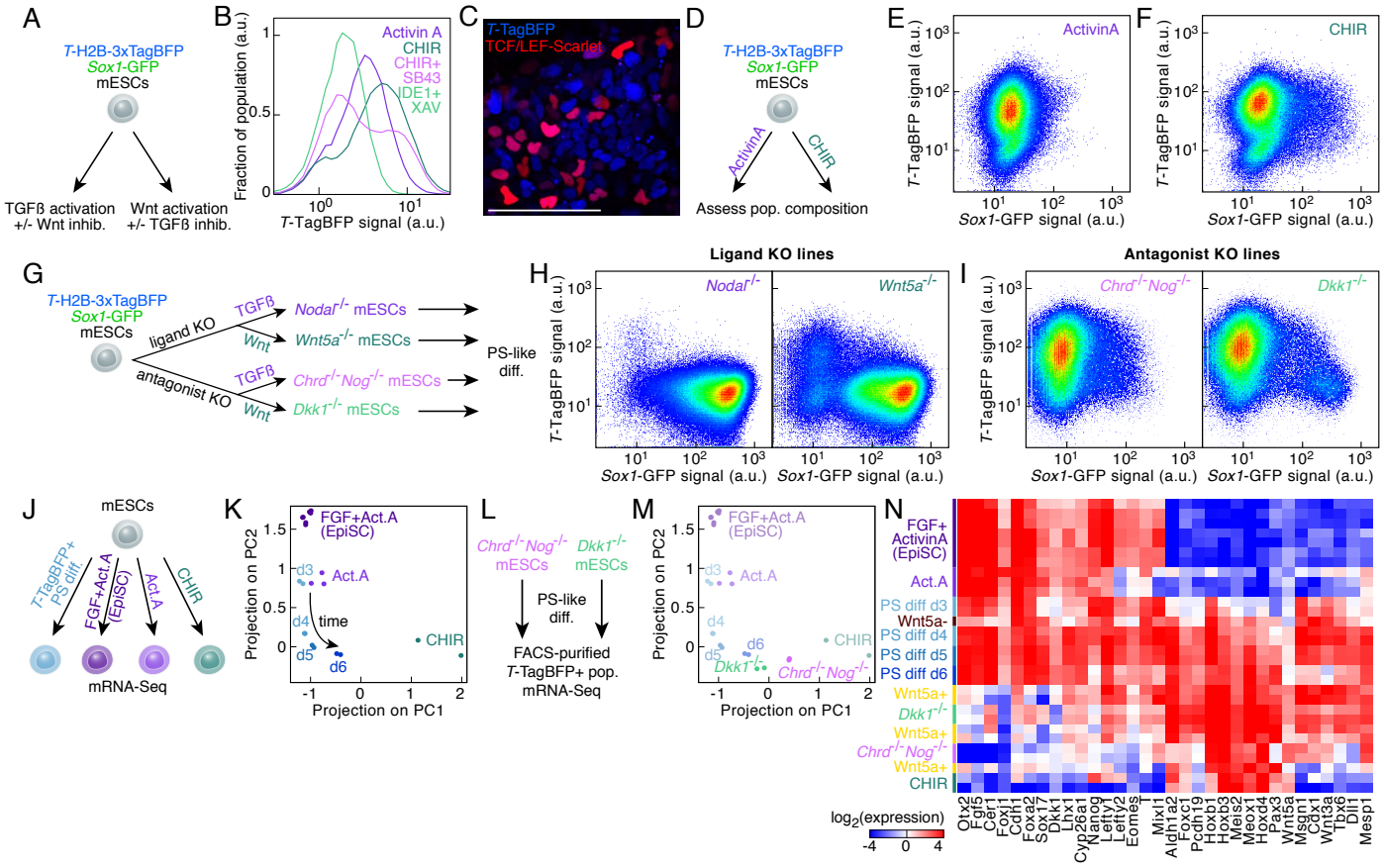

Figure 4. Role of TGFβ and Wnt signaling pathways in the induction and maintenance of the PS-like population.

(A) Scheme of the experimental principle to monitor the impact on T expression of the activation/inhibition of the TGFβ and Wnt signaling pathways. CHIR99021 (CHIR, a GSK3 inhibitor) activates the Wnt pathway.

(B) *T*-TagBFP expression levels after 2 days of pharmacological perturbations of the TGFβ and Wnt signaling pathways (purple: Activin A, teal blue: CHIR, pink: CHIR and SB43, light green: IDE1 and XAV).

(C) Reporter expression after 2 days of differentiation of a knockin *T*-TagBFP mESC line transgenic for a Wnt signaling reporter (TCF/LEF-Scarlet). Bar: 100 μm.

(D) Scheme of the experimental principle to monitor the impact on fate acquisition of the activation of the TGFβ and Wnt signaling pathways.

(E, F) *T*-TagBFP and *Sox1*-GFP reporter expression in cultures differentiated for 4 days with Activin A (E) or CHIR (F).

(G) Experimental strategy to assess the phenotype of cell lines with knockouts of ligands or antagonists of

the TGFβ and Wnt signaling pathways.

(H) *T*-TagBFP and *Sox1*-GFP reporter expression after PS-like differentiation of *Nodal*$^{-/-}$ (left panel) or *Wnt5a*$^{-/-}$ (right panel) mESCs.

(I) *T*-TagBFP and *Sox1*-GFP reporter expression after PS-like differentiation of *Chrd*$^{-/-}$*Nog*$^{-/-}$ (left panel) or *Dkk1*$^{-/-}$ (right panel) mESCs.

(J) Experimental strategy to characterize the impact of the activation of the TGFβ and Wnt signaling pathways on PS-like populations. Epiblast Stem Cells (EpiSCs, maintained with FGF and Activin A) serve as an additional comparison point.

(K) Principal component analysis of expression profiles of PS-induced T$^{TagBFP+}$ cells (blue, color according to the day of differentiation), cells after Activin A (purple) or CHIR (teal blue) differentiation and EpiSCs (dark purple).

(L) Scheme to characterize the PS-like population obtained from *Chrd*$^{-/-}$*Nog*$^{-/-}$ or *Dkk1*$^{-/-}$ mESCs.

(M) Projection on the first two principal components of expression profiles of T$^{TagBFP+}$ *Chrd*$^{-/-}$*Nog*$^{-/-}$ (pink) and *Dkk1*$^{-/-}$ (light green) cells after PS-like differentiation.

(N) Expression levels of primitive streak, endoderm, presomitic mesoderm and somite markers in EpiSCs (dark purple), cells differentiated with Activin A (purple) or CHIR (teal blue), T$^{TagBFP+}$ wild type cells (PS diff d3, d4, d5 and d6) after PS-like differentiation, T$^{TagBFP+}$ *Chrd*$^{-/-}$*Nog*$^{-/-}$ (pink) and *Dkk1*$^{-/-}$ (light green) cells after PS-like differentiation, and cells FACS-purified according to their Wnt5a levels (yellow: Wnt5a+, brown: Wnt5a-) after PS-like differentiation.

See also Figure S4.

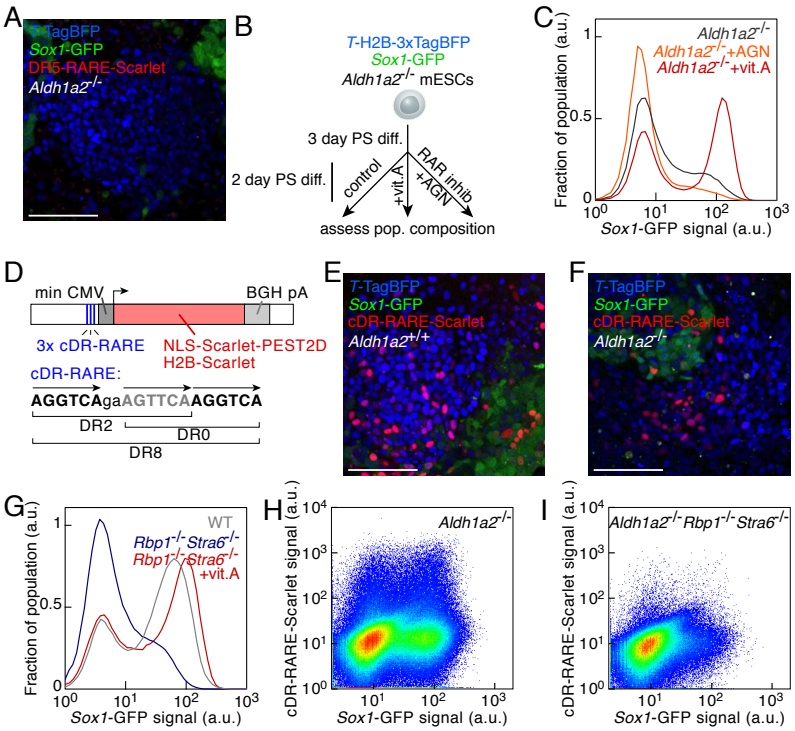

Figure 5. Aldh1a2-independent RA signaling during PS-like differentiation.

(A) Reporter expression after 5 days of PS-like differentiation of a double knockin *T*-TagBFP *Sox1*-GFP *Aldh1a2*$^{-/-}$ mESC line transgenic for a RA signaling reporter (DR5-RARE-Scarlet). Bar: 100 μm.

(B) Scheme of the experimental principle to monitor the impact of perturbing RA signaling on the PS-like differentiation of *Aldh1a2*$^{-/-}$ mESCs.

(C) *Sox1*-GFP reporter expression in *Aldh1a2*$^{-/-}$ cultures after PS-like differentiation (black: control, orange: AGN from day 3, red: vitamin A from day 3).

(D) Scheme of an RA-responsive transcriptional reporter relying on three RA responsive elements (RARE) consisting of three RAR binding sites creating direct repeats (DR) with three different spacing combinations (cDR: composite direct repeat, min CMV: minimal CMV promoter, BGH pA: bovine growth hormone poly A).

(E) Reporter expression after 5 days of PS-like differentiation of a double knockin *T*-TagBFP *Sox1*-GFP wild type (*Aldh1a2*$^{+/+}$) mESC line transgenic for the cDR RA signaling reporter (cDR-RARE-Scarlet). Bar: 100 μm.

(F) Reporter expression after 5 days of PS-like differentiation of a double knockin *T*-TagBFP *Sox1*-GFP *Aldh1a2*$^{-/-}$ mESC line transgenic for the cDR RA signaling reporter (cDR-RARE-Scarlet). Bar: 100 μm.

(G) *Sox1*-GFP reporter expression after PS-like differentiation of wild type (gray) or *Rbp1*$^{-/-}$ *Stra6*$^{-/-}$ cells without (blue) or with (red) additional vitamin A.

(H, I) *Sox1*-GFP and cDR-RARE-Scarlet reporter expression after PS-like differentiation of *Aldh1a2*$^{-/-}$ (H) or *Aldh1a2*$^{-/-}$ *Rbp1*$^{-/-}$ *Stra6*$^{-/-}$ (I) mESCs.

See also Figure S5.

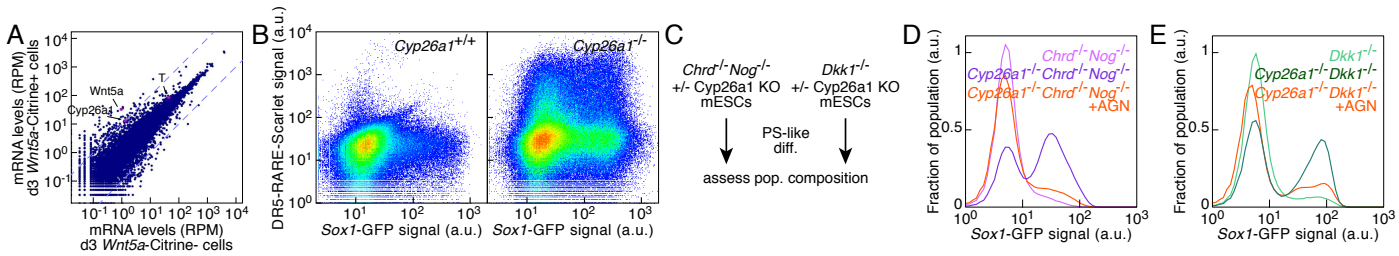

Figure 6. Cyp26a1 is a key factor limiting RA signaling during PS-like differentiation.

(A) mRNA levels in cells sorted according to their Wnt5a expression after 3 days of PS-like differentiation. Dashed lines indicate 10-fold expression changes (RPM: reads per million mapped reads).

(B) *Sox1*-GFP and DR5-RARE-Scarlet reporter expression after PS-like differentiation of wild type (left panel) or *Cyp26a1*$^{-/-}$ cells (right panel).

(C) Scheme to assess the interplay between Cyp26a1-mediated dampening of RA signaling and TGFβ or Wnt signaling.

(D) *Sox1*-GFP reporter expression in *Chrd*$^{-/-}$*Nog*$^{-/-}$ (pink) or *Cyp26a1*$^{-/-}$*Chrd*$^{-/-}$*Nog*$^{-/-}$ (purple, orange: AGN added after day 3) cultures after PS-like differentiation.

(E) *Sox1*-GFP reporter expression in *Dkk1*$^{-/-}$ (light green) or *Cyp26a1*$^{-/-}$*Dkk1*$^{-/-}$(teal blue, orange: AGN added after day 3) cultures after PS-like differentiation.

See also Figure S6.

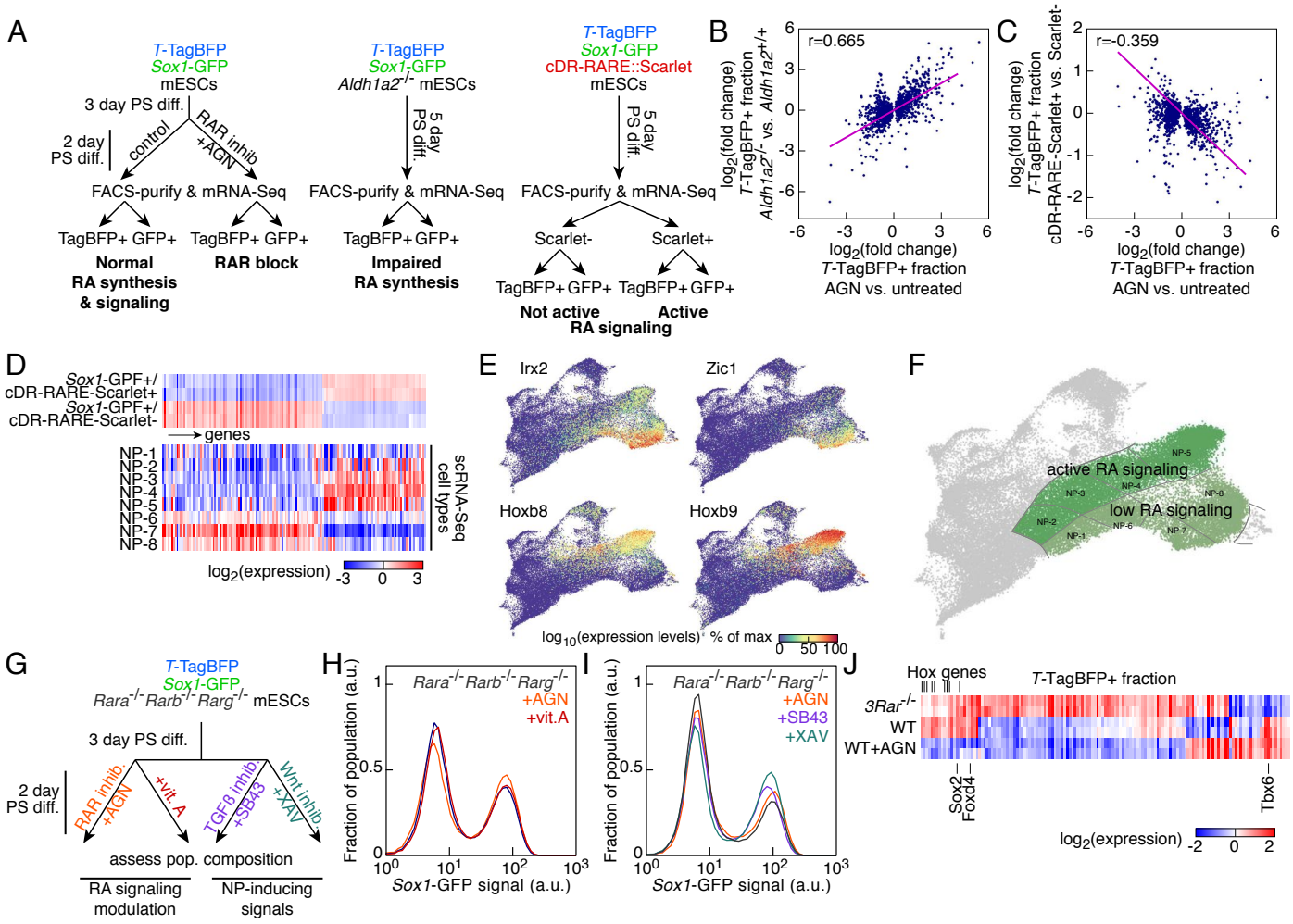

Figure 7. Role of RA signaling in establishing neural progenitor diversity.

(A) Experimental strategy to assess the impact of RA signaling on gene expression during PS-like differentiation both in the $T^{TagBFP+}$ (TagBFP+) and the $Sox1^{GFP+}$ (GFP+) populations.

(B) Comparison of differential expression in $T^{TagBFP+}$ cells after AGN treatment or in *Aldh1a2*$^{-/-}$ cells (Pearson's r=0.665, p=10^{-181}).

(C) Comparison of differential expression in $T^{TagBFP+}$ cells after AGN treatment or in cDR-RARE-Scarlet+ cells with active RA signaling (Pearson's r=-0.359, p=2.10^{-44}).

(D) Expression levels of genes with differential expression between cDR-RARE-Scarlet+ and Scarlet- subpopulations of the $Sox1^{GFP+}$ fraction after PS-like differentiation and in the 8 neural progenitor (NP) categories identified by scRNA-Seq.

(E) UMAP colored by the scaled expression of markers identifying different NP territories.

(F) UMAP with the RA signaling status highlighted in the NP territories.

(G) Experimental strategy to monitor the impact of RA signaling modulation or NP-inducing cues in *Rara*$^{-/-}$ *Rarb*$^{-/-}$ *Rarg*$^{-/-}$ mESCs.

(H) *Sox1*-GFP reporter expression in *Rara*$^{-/-}$ *Rarb*$^{-/-}$ *Rarg*$^{-/-}$ cultures after PS-like differentiation with perturbation of RA signaling (black: control, orange: AGN from day 3, red: vitamin A from day 3).

(I) *Sox1*-GFP reporter expression in *Rara*$^{-/-}$ *Rarb*$^{-/-}$ *Rarg*$^{-/-}$ cultures after PS-like differentiation with TGFβ or Wnt signaling inhibition (black: control, orange: AGN, purple: SB43, teal blue: XAV).

(J) Expression levels of transcriptional factors with differential expression in the T$^{TagBFP+}$ fraction after PS-like differentiation of *Rara*$^{-/-}$ *Rarb*$^{-/-}$ *Rarg*$^{-/-}$ (*3Rar*$^{-/-}$) cells or wild type cells without (WT) or with (WT+AGN) treatment with the RAR antagonist AGN.

See also Figure S7.

EXPERIMENTAL PROCEDURES

mESC maintenance

Parental mESC lines were AB2.2 (ATCC SCRC-1023) and a *Sox1-Brachyury* double knockin-line (Sladitschek and Neveu, 2019). mESCs were maintained in “LIF+serum” as described previously (Sladitschek and Neveu, 2015b).

Generation of *Wnt5a* mESC reporter lines

The locus of *Wnt5a* was targeted with a reporter construct consisting of H2B-Citrine or H2B-tdTomato and a Neo/Kan resistance cassette flanked by FRT sites that was compatible with the MXS-chaining strategy (Sladitschek and Neveu, 2015a). Targeting constructs were generated by standard ET recombineering (Muyrers et al., 1999; Zhang et al., 2000; Wang et al., 2006) using the BAC bMQ420-G21 (Adams et al., 2005). They were linearized and transfected using Fugene HD (Promega) according to the manufacturer’s protocol. After antibiotic selection, single colonies were expanded and screened for correct targeting. Targeting was confirmed by genomic PCR or Southern blotting carried out using ^{32}P-labelled RNA probes following Church and Gilbert (1984). The lines were subcloned after removal of the selection cassette by transient transfection of pPGKFLPobpA (Addgene plasmid 13793, a kind gift of Philippe Soriano) (Raymond and Soriano, 2007).

Generation of knockout mESC lines

RNA-guided Cas9 nucleases were used to introduce inactivating mutations in the following genes: *Aldh1a2*, *Chrd*, *Cyp26a1*, *Dkk1*, *Nodal*, *Nog*, *Rara*, *Rarb*, *Rarg*, *Rbp1*, *Stra6* and *Wnt5a*.

Guide RNA inserts targeting the fourth exon of *Aldh1a2* (with genome target sequence: 5’-AGGGAGTCATCAAAACCCTG), the fifth exon of *Chrd* (5’-GGTCCGAGTTCTTGGCGCGG), the second exon of *Cyp26a1* (5’-GCGCCCATCACCCGCACCGT), the second exon of *Dkk1* (5’-GATCTGTACACCTCCGACGC), the first exon of *Nodal* (5’-GGCTCAACATGTACGCCAGA), the coding sequence of *Nog* (5’-GGAAGTTACAGATGTGGCTG), the fourth exon of *Rara* (5’-GGTGGGCGAGCTCATTGAGA), the fourth exon of *Rarb* (5’-GCGTGGTGTATTTACCCAGC), the fifth exon of *Rarg* (5’-GTGGGACAAGTTCAGCGAGC), the second exon of *Rbp1* (5’-CACTTTTCGGAACTATATCA or 5’-TCCTGCACGATCTCTTTGTC), the fourth exon of *Stra6* (5’-TCCCCAGCCAAGAAATCCAC), the second exon of *Wnt5a* (5’-GGAGTGGCTTTGGGGACCGC), were de-

signed and cloned in pX330-U6-Chimeric-BB-CBh-hSpCas9 following Hsu et al. (2013). The resulting pX330 plasmids were transfected in the appropriate mESC lines using Fugene HD (Promega) according to the manufacturer's protocol. Successfully edited clones were validated by Sanger sequencing of genomic PCR amplicons.

Reporter constructs

All constructs were assembled using the MXS-chaining strategy (Sladitschek and Neveu, 2015a). A CAG::H2B-2xiRFP670-bGHpA cassette (combined with a PGK::NeoR-bGHpA cassette) was used as constitutive fluorescent marker for the co-culture experiments.

Transcriptional reporters consisted of binding sites upstream of a minimal CMV promoter driving the expression of NLS-Scarlet-PEST2D or H2B-Scarlet (both relying on the bright red fluorescent protein mScarlet (Bindels et al., 2017)). The plasmid contained a PGK::HygroR-bGHpA cassette to enable selection with hygromycin. Seven TCF/LEF binding sites (5'-AGATCAAAGGG) spaced by 4 to 7 nucleotides (following the TOP Flash reporter strategy (Molenaar et al., 1996)) were used for the Wnt activity reporter. Direct repeats (DR) of RAR binding sites spaced by 5 nucleotides (5'-GGTTCACCGAAAGTTCA) reported in Rossant et al. (1991) were the base of the regulatory region of the DR5-RARE-Scarlet reporter. Three DR5 spaced by 9 and 10 nucleotides were used. The regulatory sequences of the composite DR reporter cDR-RARE-Scarlet consisted of three RAR binding sites (5'-AGGTCAGAAGTTCAAGGTCA) described in Moutier et al. (2012). Three cDRs spaced by 12 nucleotides were used.

Titration of the response to RA of the cDR-RARE-Scarlet reporter line was performed in N2B27 medium supplemented with all-trans retinoic acid. N2B27 medium was prepared from a 1:1 mixture of DMEM/F12 (without HEPES, with L-glutamine) and neurobasal medium with 0.5x B-27 (without vitamin A) and 0.5x N-2 supplements, 0.25 mM L-glutamine, 0.1 mM 2-mercaptoethanol (all Invitrogen), 10 μg/ml BSA fraction V and 10 μg/ml human recombinant insulin (both Sigma). Fluorescence was measured by flow cytometry 24 h after the addition of RA.

Transgenic mESC lines

We used the following transgenic cell lines in this study.

- *Sox1*-GFP, *T*-H2B-3xTagBFP mESCs (Sladitschek and Neveu, 2019).
- *Sox1*-GFP, *T*-H2B-3xTagBFP, *Eomes*-H2B-mCherry mESCs (Sladitschek and Neveu, 2019).
- *Sox1*-GFP, *T*-H2B-3xTagBFP DR5-RARE-NLS-Scarlet-PEST2D-bGHpA mESCs.

- *Sox1*-GFP, *T*-H2B-3xTagBFP cDR-RARE-NLS-Scarlet-PEST2D-bGHpA mESCs.
- *Sox1*-GFP, *T*-H2B-3xTagBFP TCF/LEF-NLS-Scarlet-PEST2D-bGHpA mESCs.
- *Wnt5a*-H2B-Citrine, TCF/LEF-NLS-Scarlet-PEST2D-bGHpA mESCs.
- *Sox1*-GFP, *T*-H2B-3xTagBFP cDR-RARE-H2B-Scarlet-bGHpA mESCs.
- *Sox1*-GFP, *T*-H2B-3xTagBFP, *Wnt5a*-H2B-tdTomato mESCs.
- *Sox1*-GFP, *T*-H2B-3xTagBFP *Aldh1a2*$^{-/-}$ mESCs.
- *Sox1*-GFP, *T*-H2B-3xTagBFP *Aldh1a2*$^{-/-}$ DR5-RARE-NLS-Scarlet-PEST2D-bGHpA CAG::H2B-2xiRFP670-bGHpA mESCs.
- *Sox1*-GFP, *T*-H2B-3xTagBFP *Aldh1a2*$^{-/-}$ cDR-RARE-H2B-Scarlet-bGHpA mESCs.
- *Sox1*-GFP, *T*-H2B-3xTagBFP *Nodal*$^{-/-}$ mESCs.
- *Sox1*-GFP, *T*-H2B-3xTagBFP *Wnt5a*$^{-/-}$ mESCs.
- *Sox1*-GFP, *T*-H2B-3xTagBFP *Dkk1*$^{-/-}$ mESCs.
- *Sox1*-GFP, *T*-H2B-3xTagBFP *Chrd*$^{-/-}$ *Nog*$^{-/-}$ mESCs.
- *Sox1*-GFP, *T*-H2B-3xTagBFP *Cyp26a1*$^{-/-}$ DR5-RARE-NLS-Scarlet-PEST2D-bGHpA mESCs.
- *Sox1*-GFP, *T*-H2B-3xTagBFP *Rbp1*$^{-/-}$ *Stra6*$^{-/-}$ mESCs.
- *Sox1*-GFP, *T*-H2B-3xTagBFP *Aldh1a2*$^{-/-}$ *Rbp1*$^{-/-}$ *Stra6*$^{-/-}$ cDR-RARE-H2B-Scarlet-bGHpA mESCs.
- *Sox1*-GFP, *T*-H2B-3xTagBFP *Cyp26a1*$^{-/-}$ *Chrd*$^{-/-}$ *Nog*$^{-/-}$ mESCs.
- *Sox1*-GFP, *T*-H2B-3xTagBFP *Cyp26a1*$^{-/-}$ *Dkk1*$^{-/-}$ mESCs.
- *Sox1*-GFP, *T*-H2B-3xTagBFP *Rara*$^{-/-}$ *Rarb*$^{-/-}$ *Rarg*$^{-/-}$ DR5-RARE-NLS-Scarlet-PEST2D-bGHpA mESCs.

Primitive streak-like differentiation

For differentiation towards a primitive streak-like fate (Sladitschek and Neveu, 2019), mESCs were seeded at a density of 30-200 cells per mm^2 onto 0.1% gelatin coated dishes one day prior to the start of the differentiation procedure. The following day, cells were washed with D-PBS and switched to Advanced RPMI 1640 (ThermoFisher) supplemented with 1 μM IDE-1 (Tocris), 0.2% (v/v) ES cell qualified fetal calf serum (Millipore), 2 mM L-glutamine (Sigma). 48 hours after the onset of differentiation, medium was replaced every day.

Differentiation to mesendoderm fates

For Activin- and CHIR-mediated differentiation, mESCs were seeded at a density of 30 cells per mm^2 onto 0.1% gelatin coated dishes one day prior to the start of the differentiation procedure. The following day,

cells were washed with D-PBS and switched to Advanced RPMI 1640 (ThermoFisher) supplemented with 0.2% (v/v) fetal calf serum (Millipore), 2 mM L-glutamine (Sigma) and 3 μM CHIR99021 (Tocris) or 50 ng/ml Activin A (Peprotech). 48 hours after the onset of differentiation, medium was replaced every day.

Differentiation to neural progenitors

For retinoic acid-mediated differentiation to neural progenitors, mESCs were seeded at a density of 100-200 cells per mm^2 onto 0.1% gelatin coated dishes one day prior to the start of the differentiation procedure. The following day, cells were washed with D-PBS and switched to N2B27 medium (N2B27 medium was prepared from a 1:1 mixture of DMEM/F12 (without HEPES, with L-glutamine) and neurobasal medium with 0.5x B-27 (with vitamin A) and 0.5x N-2 supplements, 0.25 mM L-glutamine, 0.1 mM 2-mercaptoethanol (all Invitrogen), 10 μg/ml BSA fraction V and 10 μg/ml human recombinant insulin (both Sigma)). *all-trans*-Retinoic acid (Sigma) was added at 1 μM (unless stated otherwise) to the differentiation medium 24 h after the start of the differentiation procedure. Medium was replaced every other day.

For differentiation to neural progenitors mediated by TGFβ signaling inhibition, mESCs were seeded at a density of 30 cells per mm^2 onto 0.1% gelatin coated dishes one day prior to the start of the differentiation procedure. The following day, cells were washed with D-PBS and switched to Advanced RPMI 1640 (ThermoFisher) supplemented with 0.2% (v/v) ES cell qualified fetal calf serum (Millipore), 2 mM L-glutamine (Sigma) and 10 μM SB431542 (Tocris). 3 μM CHIR99021 (Tocris) was added to the medium from day 3 onwards. 48 hours after the onset of differentiation, medium was replaced every day.

For differentiation to neural progenitors mediated by Wnt signaling inhibition, mESCs were seeded at a density of 30 cells per mm^2 onto 0.1% gelatin coated dishes one day prior to the start of the differentiation procedure. The following day, cells were washed with D-PBS and switched to Advanced RPMI 1640 (ThermoFisher) supplemented with 0.2% (v/v) ES cell qualified fetal calf serum (Millipore), 2 mM L-glutamine (Sigma) and 1 μM IDE-1 and 1 μM XAV939 (both Tocris). 48 hours after the onset of differentiation, medium was replaced every day.

Pharmacological treatments

For pharmacological interference with PS-like differentiation, compounds were added after 3 days of differentiation (unless indicated otherwise) to Advanced RPMI 1640 (ThermoFisher) supplemented with 1 μM IDE-1 (Tocris), 0.2% (v/v) fetal calf serum (Millipore), 2 mM L-glutamine (Sigma). Recombinant Activin A, BMP2, BMP4, DKK1, sFRP1, WNT3A were obtained from Peprotech. Activin A, DKK1, Noggin, sFRP1 and

WNT3A were all used at 50 ng/ml from the onset of the PS-like differentiation procedure. BMP2 and BMP4 were used at 10 ng/ml from day 2 of the PS-like differentiation procedure. All-trans retinoic acid (Sigma) and vitamin A (all-trans retinol, Sigma) were used at 1 μM and 70 nM respectively unless indicated otherwise. AGN193109 (200 nM), CHIR99021 (3 μM), SB431542 (10 μM), XAV939 (1 μM) were all obtained from Tocris and used at the indicated concentrations unless stated otherwise.

Imaging

Reporter fluorescence was assessed in live differentiating cultures. Confocal images were acquired on an inverted SP8 confocal microscope (Leica) equipped with a 40x PL Apo 1.1 W objective and an incubation chamber at 37C and 5% CO_2.

Flow cytometry and fluorescence-activated cell sorting (FACS)

Cells were trypsinized, dissociated to single-cell suspension, pelleted at 1000g for 1 min, resuspended in D-PBS and strained through a 40 μm filter. Samples were analyzed on an LSRFortessa flow cytometer (BD BioSciences). Cells were FACS-purified according to their TagBFP, GFP, Citrine, Scarlet or tdTomato fluorescence levels using an Aria Fusion sorter (BD BioSciences). Flow cytometry data was analyzed with FlowJo.

RNA-seq library construction

RNA was extracted from pellets of trypsinized cells using the MirVana kit (Ambion) following the instructions provided by the manufacturer. 67 barcoded mRNA libraries were prepared using TruSeq RNA Sample Preparation (Illumina) following the manufacturer's instructions. Barcoded libraries were sequenced on Illumina NextSeq 500 in the high density 75 bp single-end regime. Sequencing results are available on ArrayExpress with accession number E-MTAB-10242. In addition, we used mRNA expression data that we previously deposited on ArrayExpress with accession E-MTAB-2830, E-MTAB-3234 and E-MTAB-4904.

RNA-seq analysis

Ensembl cDNAs of the mouse genome release GRCm38 were masked with RepeatMasker (Smit, AFA, Hubley, R and Green, P. RepeatMasker Open-3.0. 1996-2010 http://www.repeatmasker.org) and a Bowtie index was built using these masked transcripts. Reads were aligned to this index using Bowtie (Langmead

et al., 2009) with default parameters. mRNA read counts were determined for each Ensembl ID by parsing the Bowtie output.

The differential gene expression analysis was performed using the Bioconductor package edgeR (Robinson et al., 2010). Starting from raw read counts, a normalization factor was applied taking into account differences in sequencing depth and effective library size among the libraries. Providing experimental design matrix, dispersion estimates were obtained and negative binomial generalized linear models (GLMs) were fitted to the read counts. A quasi-likelihood (QL) F-test was then applied to determine differential expression (DE) across the conditions. The KEGG pathway enrichment analysis was performed to identify over- or under-represented pathways, starting from the list of DE genes, using the edgeR built-in function kegga.

To identify transcription regulators differentially expressed in the four neural progenitors differentiation procedures, transcription regulators with FDR$<$0.05, maximal read counts $>$4 RPM and fold change $>$8 between the assessed samples were selected. Genes with differential expression between cDR-RARE-Scarlet positive and negative cells were selected with the criteria FDR$<$0.05, maximal read counts $>$4 RPM and fold change $>$2 between the assessed samples. To identify transcription regulators differentially expressed in *Rara*$^{-/-}$ *Rarb*$^{-/-}$ *Rarg*$^{-/-}$ cells, we used as criteria FDR$<$0.05, maximal read counts $>$4 RPM and fold change $>$2 between the assessed samples.

Principal component analysis was carried out as described in Neveu et al. (2010).

Single-cell RNA-Sequencing

Cultures undergoing PS-like differentiation were trypsinized at day 2 to 5. 8 different samples were collected: entire cultures at day 2, 3, 4 and 5 (day 5 was represented in biological triplicates) and FACS-purified TagBFP+/GFP- and GFP+ subpopulations from day 4. The solution was pelleted for 1 min at 1,000g and resuspended in D-PBS+0.04% Bovine Serum Albumin and strained through a 40 μm filter. For each sample, 8,000 to 10,000 cells of a single-cell suspension (of concentration 1,000 cells/μl) were loaded on a Chromium Controller (10x Genomics). Libraries were prepared using the Chromium Single Cell 3' Reagent Kit (10x Genomics) with v3 Chemistry according to the manufacturer's instructions. The eight barcoded libraries were sequenced in four runs on Illumina NextSeq 500 in the high density 40 bp paired-end regime. scRNA-Seq results are available on ArrayExpress with accession number E-MTAB-10243.

scRNA-seq analysis and quality control

For each sample, reads were demultiplexed according to the cell barcodes. mRNA reads were aligned to a mouse cDNA index using Bowtie (Langmead et al., 2009) allowing up to 3 mismatches. The Bowtie output was parsed to count the number of unique molecular identifiers (UMIs) and reads aligning to each transcript model. We kept cells with >5,000 UMIs, >2,000 expressed genes and a fraction of mitochondrial transcript <8%. For each library, cells with UMI counts greater than the average UMI count plus three standard deviations were discarded in order to remove doublets. Overall, 46,700 cells in total for the eight libraries passed quality controls. Expression levels for individual cells were normalized using Seurat methods (Satija et al., 2015). Dimensionality reduction was performed using uniform manifold approximation and projection (UMAP) (Becht et al., 2018). Clustering into subpopulations was performed using Ward distance after dimensionality reduction by UMAP. To identify potential markers of the different cell categories, we retained genes with at least 25% expressing cells and an average expression of 8 RPM in one cell category.

Statistical analysis

Statistical tests were computed using R or the Python SciPy module.

Data availability

Sequencing results are deposited on ArrayExpress with accession numbers E-MTAB-10242 and E-MTAB-10243. In addition, we used mRNA expression data that we previously deposited on ArrayExpress with accession E-MTAB-2830, E-MTAB-3234 (Sladitschek and Neveu, 2015b) and E-MTAB-4904 (Sladitschek and Neveu, 2019).

Supplemental Information

Supplemental Information includes seven figures.

References

Adams, D. J., Quail, M. A., Cox, T., van der Weyden, L., Gorick, B. D., Su, Q., Chan, W.-i., Davies, R., Bonfield, J. K., Law, F., et al. (2005). A genome-wide, end-sequenced 129sv bac library resource for targeting vector construction. Genomics, 86, 753—758.

Arnold, S., Stappert, J., Bauer, A., Kispert, A., Herrmann, B., and Kemler, R. (2000). Brachyury is a target gene of the wnt/beta-catenin signaling pathway. Mech. Dev., 91, 249—258.

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

Bachiller, D., Klingensmith, J., Kemp, C., Belo, J. A., Anderson, R. M., May, S. R., McMahon, J. A., McMa-

hon, A. P., Harland, R. M., Rossant, J., et al. (2000). The organizer factors Chordin and Noggin are required for mouse forebrain development. Nature, 403, 658–661.

Beccari, L., Moris, N., Girgin, M., Turner, D. A., Baillie-Johnson, P., Cossy, A.-C., Lutolf, M. P., Duboule, D., and Arias, A. M. (2018). Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. Nature, 562, 272—276.

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I. W. H., Ng, L. G., Ginhoux, F., and Newell, E. W. (2018). Dimensionality reduction for visualizing single-cell data using umap. Nat. Biotechnol.

Bibel, M., Richter, J., Schrenk, K., Tucker, K. L., Staiger, V., Korte, M., Goetz, M., and Barde, Y.-A. (2004). Differentiation of mouse embryonic stem cells into a defined neuronal lineage. Nat. Neurosci., 7, 1003–1009.

Bindels, D. S., Haarbosch, L., van Weeren, L., Postma, M., Wiese, K. E., Mastop, M., Aumonier, S., Gotthard, G., Royant, A., Hink, M. A., et al. (2017). mscarlet: a bright monomeric red fluorescent protein for cellular imaging. Nat. Methods, 14, 53—56.

Borowiak, M., Maehr, R., Chen, S., Chen, A. E., Tang, W., Fox, J. L., Schreiber, S. L., and Melton, D. A. (2009). Small molecules efficiently direct endodermal differentiation of mouse and human embryonic stem cells. Cell Stem Cell, 4, 348–358.

Brons, I. G. M., Smithers, L. E., Trotter, M. W. B., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S. M., Howlett, S. K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R. A., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature, 448, 191–195.

Chambon, P. (1996). A decade of molecular biology of retinoic acid receptors. FASEB J., 10, 940—954.

Chapman, D. and Papaioannou, V. (1998). Three neural tubes in mouse embryos with mutations in the t-box gene tbx6. Nature, 391, 695—697.

Church, G. and Gilbert, W. (1984). Genomic sequencing. Proc Natl Acad Sci USA, 81, 1991—1995.

Conlon, F., Lyons, K., Takaesu, N., Barth, K., Kispert, A., Herrmann, B., and Robertson, E. (1994). A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. Development, 120, 1919—1928.

De Robertis, E. M. (2006). Spemann's organizer and self-regulation in amphibian embryos. Nat. Rev. Mol. Cell Biol., 7, 296–302.

Driever, W., Thoma, G., and Nüsslein-Volhard, C. (1989). Determination of spatial domains of zygotic gene expression in the Drosophila embryo by the affinity of binding sites for the bicoid morphogen. Nature, 340, 363–367.

Durston, A., Timmermans, J., Hage, W., Hendriks, H., de Vries, N., Heideveld, M., and Nieuwkoop, P. (1989). Retinoic acid causes an anteroposterior transformation in the developing central nervous system. Nature, 340, 140—144.

Etoc, F., Metzger, J., Ruzo, A., Kirst, C., Yoney, A., Ozair, M. Z., Brivanlou, A. H., and Siggia, E. D. (2016). A balance between secreted inhibitors and edge sensing controls gastruloid self-organization. Dev. Cell, 39, 302—315.

Fujii, H., Sato, T., Kaneko, S., Gotoh, O., Fujii-Kuriyama, Y., Osawa, K., Kato, S., and Hamada, H. (1997). Metabolic inactivation of retinoic acid by a novel p450 differentially expressed in developing mouse embryos. EMBO J., 16, 4163–4173.

Gadue, P., Huber, T. L., Paddison, P. J., and Keller, G. M. (2006). Wnt and tgf-beta signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. Proc. Natl. Acad. Sci. U.S.A., 103, 16806—16811.

Glinka, A., Wu, W., Delius, H., Monaghan, A., Blumenstock, C., and Niehrs, C. (1998). Dickkopf-1 is a member of a new family of secreted proteins and functions in head induction. Nature, 391, 357—362.

Gouti, M., Tsakiridis, A., Wymeersch, F. J., Huang, Y., Kleinjung, J., Wilson, V., and Briscoe, J. (2014). In vitro generation of neuromesodermal progenitors reveals distinct roles for wnt signalling in the specification of spinal cord and paraxial mesoderm identity. PLoS Biol., 12, e1001937.

Hemmati-Brivanlou, A. and Melton, D. (1997). Vertebrate embryonic cells will become nerve cells unless told otherwise. Cell, 88, 13—17.

Hogan, B. L., Thaller, C., and Eichele, G. (1992). Evidence that Hensen's node is a site of retinoic acid synthesis. Nature, 359, 237–241.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol., 31, 827–832.

Kawaguchi, R., Yu, J., Honda, J., Hu, J., Whitelegge, J., Ping, P., Wiita, P., Bok, D., and Sun, H. (2007). A membrane receptor for retinol binding protein mediates cellular uptake of vitamin A. Science, 315, 820—825.

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

Levine, A. J. and Brivanlou, A. H. (2007). Proposal of a model of mammalian neural induction. Dev. Biol., 308, 247—256.

Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schütz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., et al. (1995). The nuclear receptor superfamily: the second decade. Cell, 83, 835–839.

Martyn, I., Kanno, T., Ruzo, A., Siggia, E., and Brivanlou, A. (2018). Self-organization of a human organizer by combined wnt and nodal signalling. Nature, 558, 132—135.

Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destrée, O., and Clevers, H. (1996). Xtcf-3 transcription factor mediates beta-catenin-induced axis formation in xenopus embryos. Cell, 86, 391—399.

Moris, N., Anlas, K., van den Brink, S. C., Alemany, A., Schröder, J., Ghimire, S., Balayo, T., van Oudenaarden, A., and Martinez Arias, A. (2020). An in vitro model of early anteroposterior organization during human development. Nature, 582, 410—415.

Moutier, E., Ye, T., Choukrallah, M.-A., Urban, S., Osz, J., Chatagnon, A., Delacroix, L., Langer, D., Rochel, N., Moras, D., et al. (2012). Retinoic acid receptors recognize the mouse genome through binding elements with diverse spacing and topology. J. Biol. Chem., 287, 26328—26341.

Mukhopadhyay, M., Shtrom, S., Rodriguez-Esteban, C., Chen, L., Tsukui, T., Gomer, L., Dorward, D., Glinka, A., Grinberg, A., Huang, S., et al. (2001). Dickkopf1 is required for embryonic head induction and limb morphogenesis in the mouse. Dev. Cell, 1, 423—434.

Muyrers, J., Zhang, Y., Testa, G., and Stewart, A. (1999). Rapid modification of bacterial artificial chromosomes by et-recombination. Nucleic Acids Res, 27, 1555—1557.

Neveu, P., Kye, M. J., Qi, S., Buchholz, D. E., Clegg, D. O., Sahin, M., Park, I.-H., Kim, K.-S., Daley, G. Q., Kornblum, H. I., et al. (2010). MicroRNA profiling reveals two distinct p53-related human pluripotent stem cell states. Cell Stem Cell, 7, 671–681.

Niederreither, K., Subbarayan, V., Dollé, P., and Chambon, P. (1999). Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. Nat. Genet., 21, 444—448.

Peng, G., Suo, S., Cui, G., Yu, F., Wang, R., Chen, J., Chen, S., Liu, Z., Chen, G., Qian, Y., et al. (2019). Molecular architecture of lineage allocation and tissue organization in early mouse embryo. Nature, 572, 528—532.

Pevny, L., Sockanathan, S., Placzek, M., and Lovell-Badge, R. (1998). A role for SOX1 in neural determination. Development, 125, 1967—1978.

Raymond, C. S. and Soriano, P. (2007). High-efficiency flp and phic31 site-specific recombination in mammalian cells. PloS One, 2, e162.

Rhinn, M. and Dollé, P. (2012). Retinoic acid signalling during development. Development, 139, 843—858.

Ribes, V., Le Roux, I., Rhinn, M., Schuhbaur, B., and Dollé, P. (2009). Early mouse caudal development relies on crosstalk between retinoic acid, shh and fgf signalling pathways. Development, 136, 665–676.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edger: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26, 139—140.

Rossant, J., Zirngibl, R., Cado, D., Shago, M., and Giguère, V. (1991). Expression of a retinoic acid response element-hsplacz transgene defines specific domains of transcriptional activity during mouse embryogenesis. Genes Dev., 5, 1333—1344.

Ruberte, E., Dolle, P., Chambon, P., and Morriss-Kay, G. (1991). Retinoic acid receptors and cellular retinoid binding proteins. ii. their differential pattern of transcription during early morphogenesis in mouse embryos. Development, 111, 45—60.

Sakai, Y., Meno, C., Fujii, H., Nishino, J., Shiratori, H., Saijoh, Y., Rossant, J., and Hamada, H. (2001). The retinoic acid-inactivating enzyme cyp26 is essential for establishing an uneven distribution of retinoic acid along the anterio-posterior axis within the mouse embryo. Genes Dev., 15, 213—225.

Sandell, L. L., Sanderson, B. W., Moiseyev, G., Johnson, T., Mushegian, A., Young, K., Rey, J.-P., Ma, J.-x., Staehling-Hampton, K., and Trainor, P. A. (2007). Rdh10 is essential for synthesis of embryonic retinoic acid and is required for limb, craniofacial, and organ development. Genes Dev., 21, 1113—1124.

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol., 33, 495—502.

Shahbazi, M. N., Siggia, E. D., and Zernicka-Goetz, M. (2019). Self-organization of stem cells into embryos: A window on early mammalian development. Science, 364, 948—951.

Shimozono, S., Iimura, T., Kitaguchi, T., Higashijima, S.-I., and Miyawaki, A. (2013). Visualization of an endogenous retinoic acid gradient across embryonic development. Nature, 496, 363—366.

Simunovic, M., Metzger, J. J., Etoc, F., Yoney, A., Ruzo, A., Martyn, I., Croft, G., You, D. S., Brivanlou, A. H., and Siggia, E. D. (2019). A 3D model of a human epiblast reveals BMP4-driven symmetry breaking. Nat. Cell Biol., 21, 900—910.

Sladitschek, H. L. and Neveu, P. A. (2015a). MXS-Chaining: A Highly Efficient Cloning Platform for Imaging and Flow Cytometry Approaches in Mammalian Systems. PloS one, 10, e0124958.

Sladitschek, H. L. and Neveu, P. A. (2015b). The bimodally expressed microRNA miR-142 gates exit from pluripotency. Mol. Syst. Biol., 11, 850.

Sladitschek, H. L. and Neveu, P. A. (2019). A gene regulatory network controls the balance between mesendoderm and ectoderm at pluripotency exit. Mol. Syst. Biol., 15, e9043.

Spemann, H. and Mangold, H. (1924). Über die Induktion von Embryonalanlagen durch Implantation artfremder Organisatoren. Arch Mikrosk Anat Entwicklungsmech, 100, 599–638.

Takemoto, T., Uchikawa, M., Yoshida, M., Bell, D. M., Lovell-Badge, R., Papaioannou, V. E., and Kondoh, H. (2011). Tbx6-dependent sox2 regulation determines neural or mesodermal fate in axial stem cells. Nature, 470, 394—398.

Tam, P. P. and Behringer, R. R. (1997). Mouse gastrulation: the formation of a mammalian body plan. Mech. Dev., 68, 3–25.

Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., Gardner, R. L., and McKay, R. D. G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature, 448, 196–199.

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

van den Brink, S. C., Baillie-Johnson, P., Balayo, T., Hadjantonakis, A.-K., Nowotschin, S., Turner, D. A., and Martinez Arias, A. (2014). Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. Development, 141, 4231—4242.

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

Wang, J., Sarov, M., Rientjes, J., Fu, J., Hollak, H., Kranz, H., Xie, W., Stewart, A. F., and Zhang, Y. (2006). An improved recombineering approach by adding reca to lambda red recombination. Mol Biotechnol, 32, 43—53.

Warmflash, A., Sorre, B., Etoc, F., Siggia, E. D., and Brivanlou, A. H. (2014). A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. Nat. Methods, 11, 847—854.

Watanabe, K., Kamiya, D., Nishiyama, A., Katayama, T., Nozaki, S., Kawasaki, H., Watanabe, Y., Mizuseki, K., and Sasai, Y. (2005). Directed differentiation of telencephalic precursors from embryonic stem cells. Nat. Neurosci., 8, 288–296.

Weinstein, D. and Hemmati-Brivanlou, A. (1999). Neural induction. Annu. Rev. Cell Dev. Biol., 15, 411—433.

Wilkinson, D., Bhatt, S., and Herrmann, B. (1990). Expression pattern of the mouse t gene and its role in mesoderm formation. Nature, 343, 657—659.

Yamaguchi, T., Bradley, A., McMahon, A., and Jones, S. (1999). A wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. Development, 126, 1211—1223.

Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M., and Smith, A. (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat Biotechnol, 21, 183–186.

Zhang, Y., Muyrers, J., Testa, G., and Stewart, A. (2000). Dna cloning by homologous recombination in escherichia coli. Nat Biotechnol, 18, 1314—1317.