Stem Cell Reports
Article

ISSCR

OPEN ACCESS

Multi-layered regulation of neuroectoderm differentiation by retinoic acid in a primitive streak-like context

Luigi Russo,[1,2] Hanna L. Sladitschek,[1,3] and Pierre A. Neveu[1,*]
[1]Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany
[2]Joint PhD Degree from EMBL and Heidelberg University, Faculty of Biosciences, 69120 Heidelberg, Germany
[3]Present address: Department of Molecular Medicine, University of Padua School of Medicine, 35126 Padua, Italy
*Correspondence: neveu@embl.de
https://doi.org/10.1016/j.stemcr.2021.12.014

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 early 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. In the absence of the RA-synthesizing enzyme Aldh1a2, a sensitive RA reporter revealed 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, even in the absence of TGFβ and Wnt antagonists. Overall, we characterized a multi-layered control of RA levels that regulates early neural differentiation in an *in vitro* PS-like system.

INTRODUCTION

During gastrulation, cells of the epiblast are allocated to the three germ layers (Tam and Behringer, 1997). Gastrulation is initiated by the 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 mediating this induction (De Robertis, 2006). Among them, antagonists of the transforming growth factor β (TGFβ) signaling pathway, and in particular of bone morphogenetic proteins, are considered pivotal for neuralization of the ectoderm (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 mammals (Levine and Brivanlou, 2007). Indeed, the distal tip of the mouse PS, the node, possesses organizer-like properties (Tam and Behringer, 1997). 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 anterior neural structures (forebrain) 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). RA signaling was detected as well in the mouse PS at E7.5 (Rossant et al., 1991). At this developmental stage, ALDH1A2 is considered to be the only enzyme synthesizing RA from retinal (Rhinn and Dollé, 2012). Both the presence of forebrain structures and an absence of expression of an RA activity reporter in *Aldh1a2*$^{-/-}$ embryos ruled out an involvement of RA in early neural induction (Niederreither et al., 1999). This contrasts with the widespread use of RA to induce neuronal fates from pluripotent cells *in vitro* (Ying et al., 2003) and the well-established role of RA in the formation of the posterior neural axis. Here, the allocation of cell types to somite and spinal cord fates from bipotent neuromesodermal progenitors (NMPs) allows the extension of the body axis (Henrique et al., 2015). It was demonstrated that RA has a critical function in NMP differentiation to the neural lineage (Diez del Corral et al., 2003). The RAR family of nuclear receptors, which acts as transcription factors regulated by RA, is the effector of the developmental functions of RA (Samarut and Rochette-Egly, 2012). While *in vivo* work established the importance of antagonizing TGFβ and Wnt signaling pathways for neural induction and ruled out a contribution of RA signaling in this process, the molecular implementation of the neuroectoderm differentiation decision is largely unexplored. *In vitro* systems based on pluripotent stem cells enable to recapitulate crucial aspects of early post-implantation mammalian development (Shahbazi et al., 2019).

We adopted an mESC-based system in which we can monitor the formation of neuroectoderm in the presence of a PS-like population and profiled by single-cell RNA sequencing the progression of the differentiating culture. In this context, we determined that RA mediates early neural induction downstream of TGFβ and Wnt inhibition.

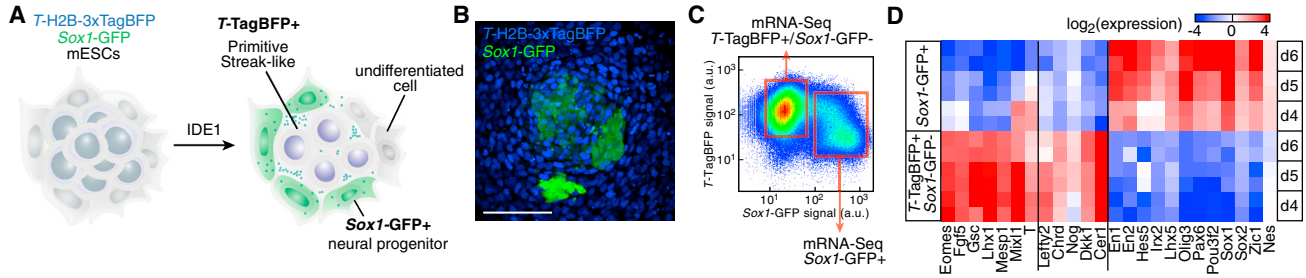

Figure 1. Characterization of neural induction by PS-like cells

(A) Experimental strategy to induce and monitor the formation of neuroectoderm by PS-like cells using a double knock-in mESC line reporting on *T* 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 populations present in the differentiating culture.

(D) Expression levels (as measured by mRNA-Seq) of PS and NP markers in FACS-purified populations after PS-like differentiation. See also Figure S1.

The formation of neural progenitors, even in the absence of the antagonists CHORDIN and NOGGIN or DKK1, was enhanced by deleting the RA-degrading enzyme CYP26A1. The development of a highly sensitive RA reporter enabled us to detect RA signaling in conditions thought to lack RA synthesis ability. Finally, the knockout of RA receptors highlighted their function as regulators of loci critical for neural induction. Altogether, our results add valuable insights into the multi-layered regulation of RA signaling in the process of early neuroectoderm formation.

RESULTS

Characterization of a system to investigate the mechanisms of neuroectoderm formation

The generation of primitive streak-like cells *in vitro* should enable the induction of a neuroectodermal fate among differentiation-competent cells (Figure 1A). To monitor the formation of PS-like cells and subsequent induction of neural progenitors (NPs), we used a double knock-in (2KI) reporter mESC line (Sladitschek and Neveu, 2019) with *Sox1* locus targeted with GFP and *T* (also known as *Brachyury*) locus targeted with H2B-3xTagBFP. While *T* is expressed in the PS (Wilkinson et al., 1990), *Sox1* marks exclusively NPs (Pevny et al., 1998). We previously showed that the small molecule IDE1, which phenocopies TGFβ pathway agonists (Borowiak et al., 2009), formed differentiation intermediates resembling mouse post-implantation epiblast and PS (Sladitschek and Neveu, 2019). Interestingly, putative NP Sox1^{GFP+} cells coexisted with T$^{TagBFP+}$ cells, the candidate PS-like cells (Figure 1B).

We monitored the composition of the differentiating culture by flow cytometry (Figure S1A). The increase in TagBFP signal and Sox1^{GFP+} cells detected by the third day matched the increase in T and Sox1 mRNA levels (Figures S1A–S1D). Increasing the number of cells seeded at the beginning of the differentiation enhanced the fraction of Sox1^{GFP+} cells (Figures S1E and S1F), indicating that cell density played a role in the formation of putative NPs.

To determine the identity of the different cell populations, we characterized the transcriptome of FACS-purified cells expressing TagBFP or GFP (Figure 1C). T$^{TagBFP+}$ cells expressed markers associated with post-implantation epiblast and PS fates such as *Fgf5*, *T*, *Mixl1*, and *Goosecoid* (Figures 1D and S1G). More importantly, the expression of secreted antagonists associated with the *in vivo* organizer was selectively higher in the T$^{TagBFP+}$ population. Among these were the TGFβ antagonists Chordin (*Chrd*) and Noggin (*Nog*) and the Wnt antagonist *Dkk1*. The neuroectodermal identity of Sox1^{GFP+} cells was confirmed by the upregulation of NP markers, including *Sox2* and *Pax6* (Figures 1D and S1G).

scRNA-seq characterization of PS-like differentiation

To better characterize the cellular heterogeneity in our PS-like differentiation, we conducted single-cell RNA sequencing (scRNA-seq) between day 2 and day 5 and obtained expression profiles for 46,700 cells (Figure 2A). Uniform manifold approximation and projection (UMAP) analysis showed minimal overlap between consecutive days (Figures 2B and S2A). Profiles of FACS-purified T$^{TagBFP+}$Sox1^{GFP-} and Sox1^{GFP+} cells projected according to the respective expression territories of *T* and *Sox1* (Figures 2C and S2B). Notably, cells with high *T* expression and *Sox1*-expressing cells formed distinct populations (Figure 2C). We could identify 35 cell subpopulations (Figures 2D and S2C) stratified by a number of markers (Figure S2D). We observed a prevalence of epiblast and PS fates till day 3,

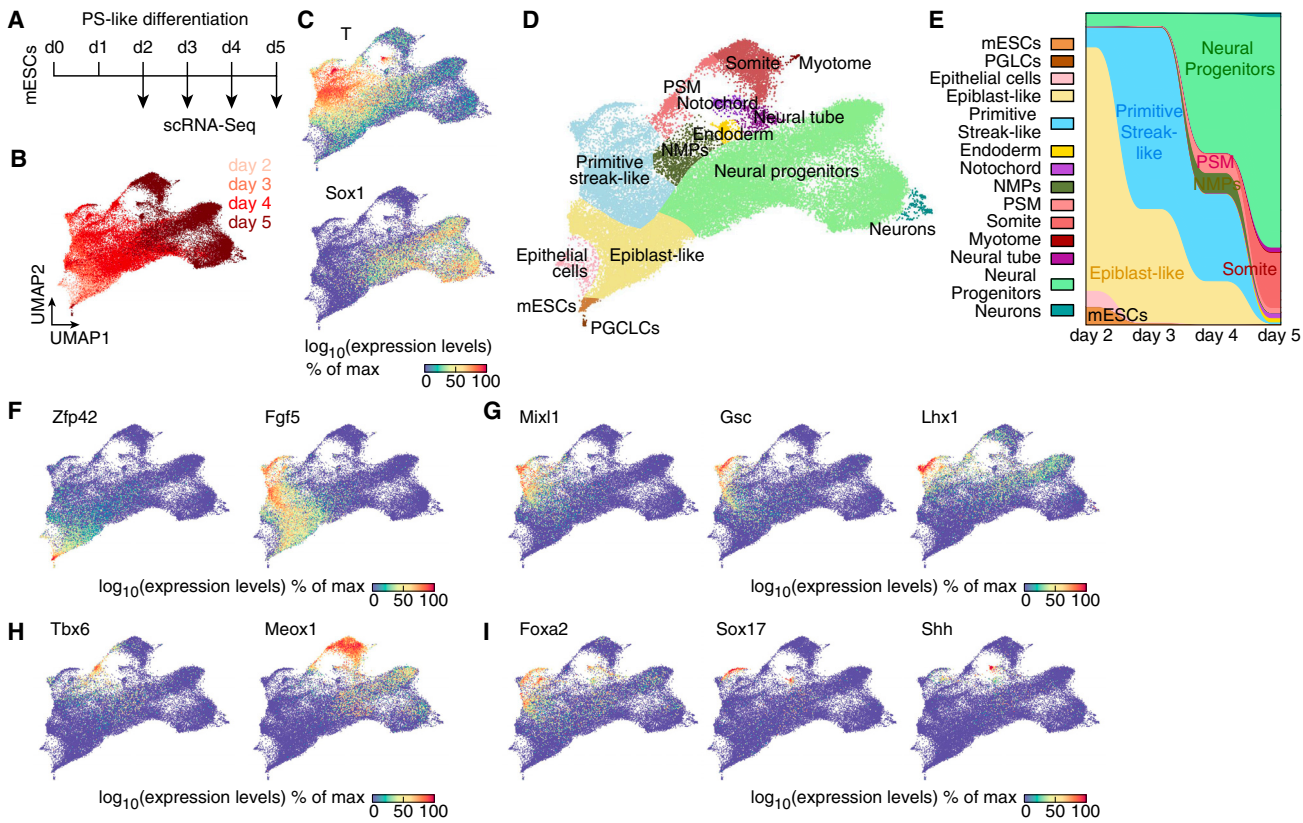

Figure 2. scRNA-seq characterization of neural induction by PS-like cells

(A) Experimental strategy to temporally monitor PS-like differentiation using scRNA-seq.

(B–D) UMAP (uniform manifold approximation and projection) of 46,700 cells colored by the collection day (B), by the scaled expression of *T* and *Sox1* (C), or according to the identified populations (D) (NMPs: neuromesodermal progenitors, PSM: presomitic mesoderm, PGCLCs: primordial germ cell-like cells).

(E) Alluvial plot showing the temporal evolution of the culture composition.

(F–I) UMAP colored by the scaled expression of the naive pluripotency marker *Zfp42* and the post-implantation epiblast marker *Fgf5* (F), of PS (G), presomitic mesoderm and somite (H), or endoderm and notochord (I) markers. See also Figure S2.

followed by the formation of NPs and PS derivatives later on (Figure 2E).

A subpopulation of cells at day 2 displayed naive pluripotency markers, while the rest initiated the expression of primed epiblast markers (Figure 2F). Pluripotency factors had distinct behaviors: while *Pou5f1* (also known as Oct4) expression was retained till day 4 in the epiblast and PS lineages, *Nanog* expression was transiently reactivated in PS-like cells (Figure S2E). *Sox2* was downregulated in PS-like cells and their derivatives, whereas its expression was maintained in NPs (Figure S2E). These findings recapitulated the expression patterns of these genes in E7.0 mouse embryos (Peng et al., 2019). PS markers were expressed in different subpopulations (Figures 2G and S2D) corresponding to different regions of the *in vivo* PS. Noteworthy, T expression encompassed both *bona fide* PS and post-implantation epiblast cells, with higher transcript levels in the former (Figures 2C and S2D). Thus, the PS-like population marked by the expression of the T^{TagBFP} reporter at day 3 comprised a mixture of these two fates.

PS derivatives were formed as differentiation proceeded. Indeed, presomitic mesoderm and distinct somite fates were found at days 4 and 5, along with a population resembling neuromesodermal progenitors (NMPs) (Figures 2E, 2H, and S2D). Moreover, endoderm and notochord fates were present by day 5 (Figure 2I). Distinct neuroectodermal cell types gradually accumulated, at the expense of epiblast and PS fates (Figures 2E, S2C, and S2F). Thus, despite the absence of defined geometrical constraints, our *in vitro* system recapitulated the fate diversification occurring in post-implantation embryos and notably the temporal evolution of the PS *in vivo*.

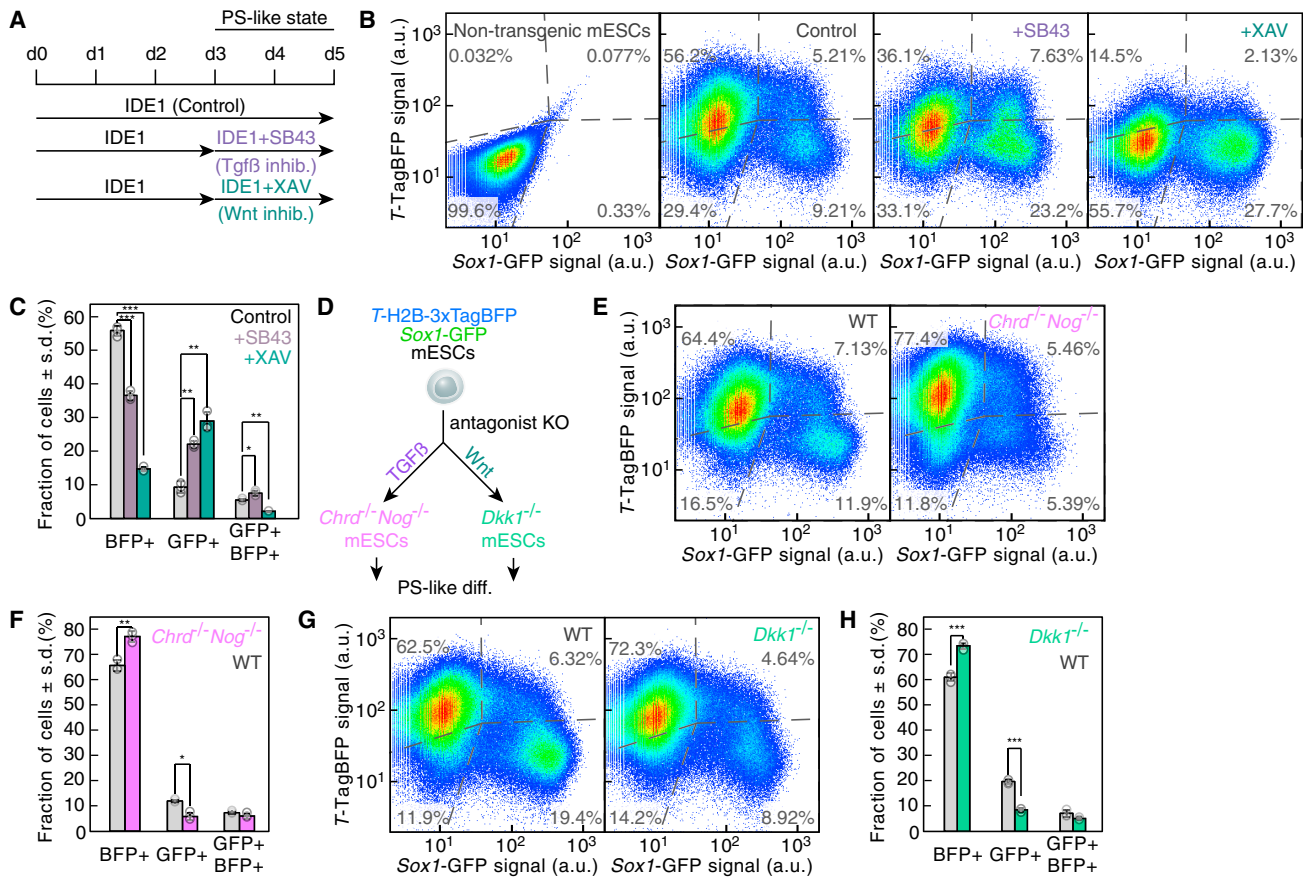

Figure 3. *In vitro* reconstitution of neural induction by Wnt and TGFβ antagonists

(A) 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.

(B) *T*-TagBFP and *Sox1*-GFP reporter expression as measured at day 5 by flow cytometry in non-transgenic mESCs and 2KI mESCs (Control), or after inhibition of the TGFβ (+SB43) or Wnt (+XAV) pathways. Dotted lines: gates fixed according to the non-transgenic mESCs negative control.

(C) Quantification of (B) data (n = 3 independent experiments; *, p < .05; **, p < .01; ***, p < .001; two-sided unpaired t test; data represented as mean ± SD).

(D) Experimental strategy to assess the impact of knockouts of antagonists of the TGFβ and Wnt signaling pathways.

(E) *T*-TagBFP and *Sox1*-GFP reporter expression as measured by flow cytometry after 5 days of PS-like differentiation of wild-type 2KI (WT) or *Chrd*$^{-/-}$*Nog*$^{-/-}$ mESCs. Dotted lines: gates fixed according to the non-transgenic mESCs negative control.

(F) Quantification of the data in (E) (n = 3 independent experiments; *, p < .05; **, p < .01; two-sided unpaired t test; data represented as mean ± SD).

(G) *T*-TagBFP and *Sox1*-GFP reporter expression as measured by flow cytometry after 5 days of PS-like differentiation of wild-type 2KI (WT) or *Dkk1*$^{-/-}$ mESCs. Dotted lines: gates fixed according to the non-transgenic mESCs negative control.

(H) Quantification of (G) data (n = 3 independent experiments; ***, p < .001; two-sided unpaired t test; data represented as mean ± SD).

See also Figure S3.

In vitro reconstitution of neural induction by Wnt and TGFβ antagonists

TGFβ or Wnt signaling inhibition are critical for neuroectoderm induction from pluripotent epiblast *in vivo* (De Robertis, 2006). We sought to recapitulate this process in our *in vitro* system by applying inhibitors once the T$^{TagBFP+}$ population was established at day 3 (Figure 3A). 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 (Figures 3B and 3C). More Sox1^{GFP+} cells could be detected as well upon addition of recombinant NOGGIN and CHORDIN or DKK1 (Figures S3A and S3B). Thus, the exogenous application of inhibitors of either signaling pathway was able to induce neural fate in our *in vitro* system. Bulk and scRNA-seq data showed that *Nog*, *Chrd*, and *Dkk1* were expressed

in PS-like cells (Figures 1D, S1G, and S2D). To test whether their endogenous expression was critical for the formation of the Sox1^{GFP+} cells, we generated *Chrd*$^{-/-}$*Nog*$^{-/-}$ 2KI and *Dkk1*$^{-/-}$ 2KI mESCs (Figures 3D, S3C, and S3D). The differentiation of the knockout cells in both cases resulted in reduced formation of Sox1^{GFP+} NPs compared to their wild-type counterparts (Figures 3E–3H). TGFβ and Wnt agonists reduced the number of Sox1^{GFP+} cells, with a nearly complete repression of neuroectoderm generation upon TGFβ signaling activation with ACTIVIN A or BMP4 (Figures S3E–S3H). These results argue that neuroectoderm formation in our system depends on the balance between the endogenous levels of agonists and inhibitors of the TGFβ and Wnt pathways.

Diverse neural progenitors emerge in the PS-like differentiation

scRNA-seq data demonstrated the heterogeneity of the NP population arising in the PS-like differentiation. Different anteroposterior identities could be assigned to NP subtypes according to the expression of markers characteristic of anterior neural tissues (*Hesx1*), anterior hindbrain (*Egr2* and *Hoxa2*), posterior hindbrain (*Hoxd4*), and spinal cord (*Hoxb9*) (Figure 4A) (Gouti et al., 2014). We hypothesized that the coexistence of multiple mechanisms of neural induction in our system could explain the formation of NPs with distinct developmental origin. To test this, we compared the transcription profile of Sox1^{GFP+} cells derived from PS-like differentiation with the ones of NPs generated by Wnt signaling inhibition, TGFβ inhibition, or the neural inducer RA (Ying et al., 2003) (Figure 4B). These NPs expressed distinct sets of transcription factors that spanned the set upregulated in the heterogeneous PS-induced NP population (Figure 4C). Wnt inhibition led to the upregulation of anterior NP markers such as *Lhx5*, *Otx2*, and *Six3* (Figure 4C), consistent with previous reports (Watanabe et al., 2005). The differentiation triggered by TGFβ inhibition increased the expression of markers of the posterior neural axis (Figure 4C). However, these two different NP-induction methods did not account for the full complexity of the expression profile of NPs obtained in the PS-like differentiation. Indeed, upregulation of markers such as *Brn1*, *Brn2*, or *Irx3* were only recapitulated by RA treatment (Figure 4C). This led us to infer that RA signaling might be in part responsible for NP formation in our system.

RA signaling mediates neural induction in the PS-like differentiation

The PS-like differentiation medium contains low concentrations of serum, which contains RA precursors that the cells could convert in RA. To test the presence of RA signaling, we stably inserted in the 2KI line a reporter construct relying on the established DR5-based RA response element (RARE) (Rossant et al., 1991) controlling the expression of the fluorescent protein Scarlet. We detected Scarlet^{+} cells in increasing amount from day 3 of the PS-like differentiation, demonstrating the activation of RA signaling and particularly in T$^{TagBFP+}$ cells (Figures 4D, S4A, and S4B). This result paralleled the identification of RA signaling in the mouse PS at E7.5 through a reporter relying on the same RARE (Rossant et al., 1991).

To identify a possible relationship between RA and NP formation, we perturbed RA signaling. Supplying the differentiation medium with additional RA precursor, vitamin A (also known as retinol), increased both RA signaling activation, as captured by the RA reporter, and the formation of Sox1^{GFP+} cells (Figures S4C–S4E). Inhibition of RA receptors (RARs) with the small molecule AGN193109 (AGN) prevented Scarlet expression and decreased the fraction of Sox1^{GFP+} cells (Figure S4F). Starting RAR inhibition at early differentiation time points further reduced NP formation (Figures S4G and S4H). This result suggested that blocking RARs could prevent the formation of new neuroectodermal cells but did not hamper the NPs already present in the culture.

The impairment of neural induction upon RAR inhibition prompted us to test the existence of a crosstalk between RA signaling and the mechanism of neural induction by TGFβ or Wnt inhibition (Figure 4E). Blocking RARs prevented the increase of the Sox1^{GFP+} population normally associated with the inhibition of either of the two pathways (Figures 4F and 4G). Moreover, we tested whether the effects of RAR antagonism were limited to the differentiation regime containing IDE1. As TGFβ pathway activation is an established cue inducing the formation of the PS *in vivo* and in hESCs (Gadue et al., 2006; Martyn et al., 2018), we turned to the Nodal/TGFβ agonist ACTIVIN A. A pulse of ACTIVIN A between day 1 and day 2 generated both T$^{TagBFP+}$ and Sox1^{GFP+} cells (Figure 4H). The early and short-term nature of the ACTIVIN A treatment was critical to avoid the repression of neural induction by Nodal/TGFβ signaling, shown in Figure S3E. As for IDE1 differentiation, RAR antagonism significantly reduced the fraction of Sox1^{GFP+} cells induced by TGFβ or Wnt inhibitors in the PS-like differentiation triggered by ACTIVIN A (Figures S4I–S4K). These results indicated that RARs controlled a step downstream of TGFβ or Wnt inhibition in the cascade of events leading to neuroectoderm formation.

Aldh1a2-independent RA signaling

We found that RA signaling mediated at least in part the neuroectoderm induction by the antagonists of the TGFβ or Wnt pathways. Given that only the RA precursor vitamin A was present in the differentiation medium, cells had to synthesize RA themselves. The oxidation of retinal in RA is performed by the retinaldehyde dehydrogenases

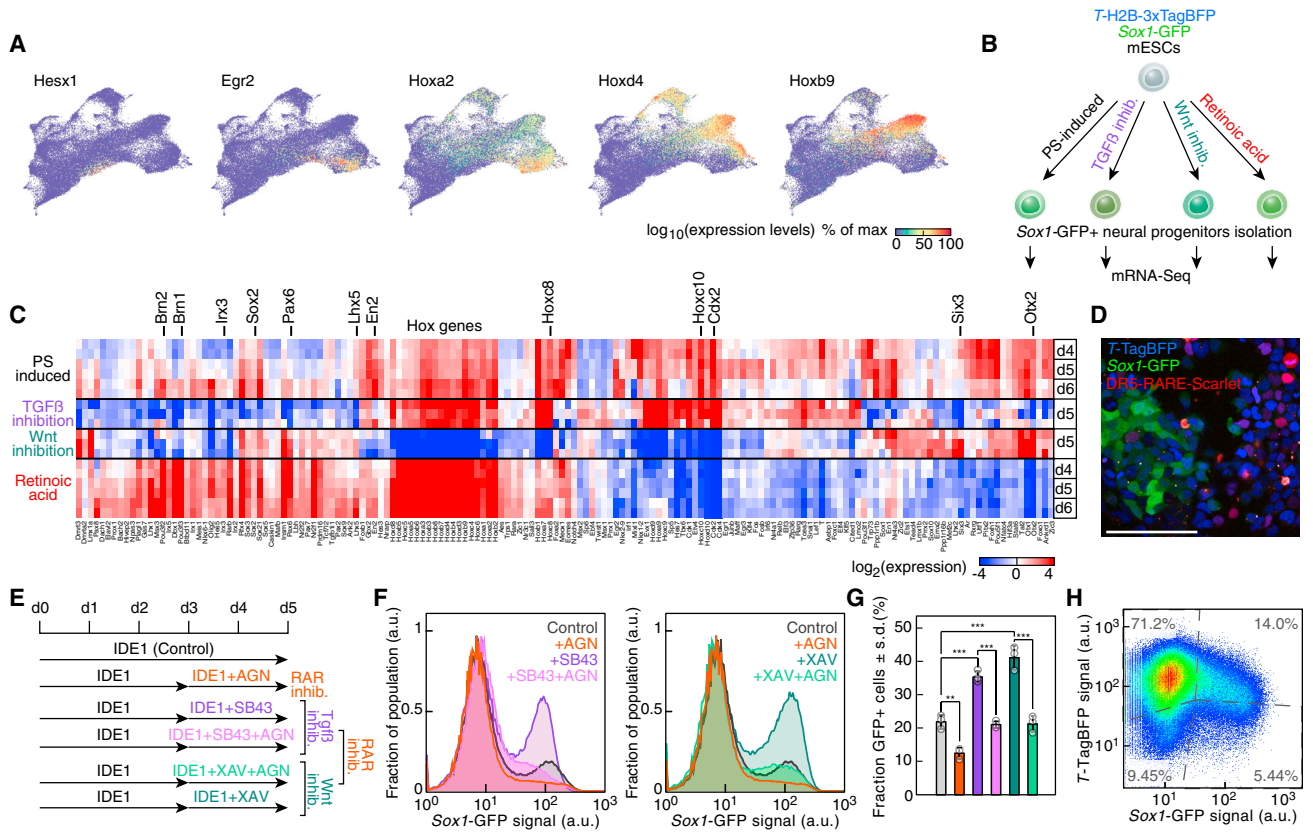

Figure 4. Retinoic acid signaling underlies NP formation in the PS-like differentiation

(A) UMAP colored by the scaled expression of neuroectodermal markers, ordered by their expression along the anteroposterior axis *in vivo*.

(B) Scheme of the experimental strategy to characterize NPs induced by PS-like cells, TGFβ, or Wnt pathway inhibition or retinoic acid (RA) treatment.

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

(D) Reporter expression after 5 days of differentiation of 2KI mESCs transgenic for a DR5-based RA signaling reporter. Bar: 100 μm.

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

(F) (left panel) *Sox1*-GFP expression levels after PS-like differentiation (black: control, orange: +AGN, purple: +SB43, pink: +SB43 + AGN). (right panel) *Sox1*-GFP expression levels after PS-like differentiation (black: control, orange: +AGN, teal blue: +XAV, light green: +XAV + AGN).

(G) Quantification of (F) data (n = 3 independent experiments; **, $p < .01$; ***, $p < .001$; one-way ANOVA followed by Tukey's post-hoc test; data represented as mean ± SD).

(H) *T*-TagBFP and *Sox1*-GFP reporter expression as measured at day 5 by flow cytometry. The differentiation was induced by a pulse of ACTIVIN A from day 1 to day 2 and without IDE1. Dotted lines: gates fixed according to the non-transgenic mESCs negative control. See also Figure S4.

ALDH1A1, ALDH1A2, or ALDH1A3 (Rhinn and Dollé, 2012). *Aldh1a2* was upregulated in mesoderm cells and $T^{TagBFP+}$ population compared with $Sox1^{GFP+}$ cells, whereas *Aldh1a1* and *Aldh1a3* expression did not exceed background levels (Figures S5A–S5C). This reproduced the expression pattern of these three genes in post-implantation mouse embryos, particularly *Aldh1a2* expression in the PS at E7.5 (Ribes et al., 2009).

The presence of forebrain structures in $Aldh1a2^{-/-}$ mouse embryos (Niederreither et al., 1999) seems to contradict our finding that RA signaling mediates early neural induction. We therefore probed whether RA signaling was completely abolished in $Aldh1a2^{-/-}$ cells. We generated $Aldh1a2^{-/-}$ mESCs (Figure S5D) bearing the RA activity reporter relying on the DR5-RARE used in the mouse model. PS-like differentiation of $Aldh1a2^{-/-}$ mESCs led to the formation of $Sox1^{GFP+}$ cells despite the absence of DR5-RARE-$Scarlet^{+}$ cells (Figures 5A and S5E), in accordance with the phenotype of $Aldh1a2^{-/-}$ mouse embryos (Niederreither et al., 1999). We then provided extra precursor for RA

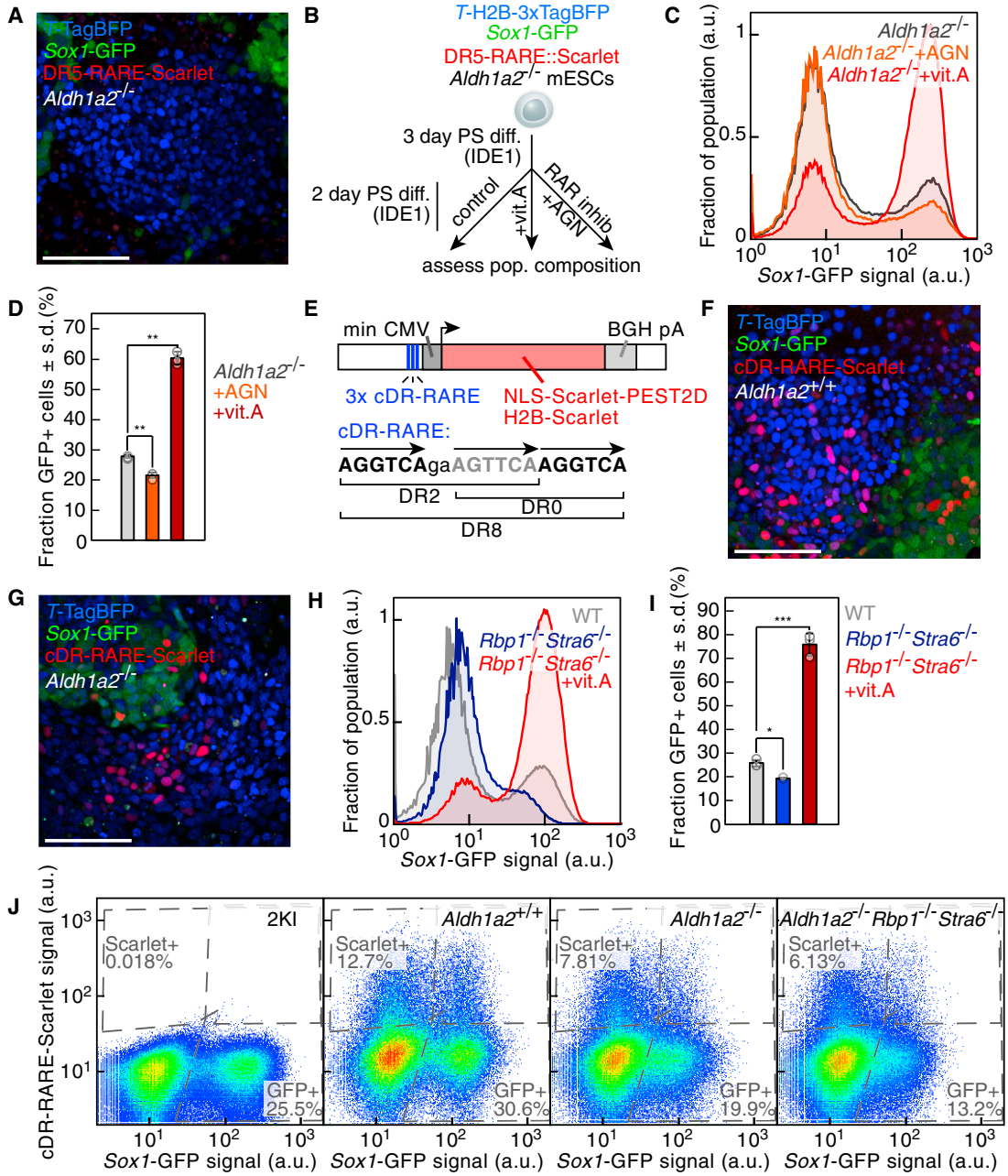

Figure 5. Aldh1a2-independent RA signaling during PS-like differentiation

(A) Reporter expression after 5 days of PS-like differentiation of 2KI *Aldh1a2*$^{-/-}$ mESCs transgenic for an 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 and D) *Sox1*-GFP reporter expression in *Aldh1a2*$^{-/-}$ cells after PS-like differentiation. (C) (black: control, orange: AGN, red: vitamin A and quantification). (D) (n = 3 independent experiments; **, p < .01; two-sided unpaired t test; data represented as mean ± SD).

(E) Scheme of an RA-responsive transcriptional reporter relying on three RA-responsive elements (RARE), each consisting of three RAR binding sites (DR, direct repeats; cDR, composite direct repeat; min CMV, minimal CMV promoter; BGH pA, bovine growth hormone poly A).

(F and G) Reporter expression after 5 days of PS-like differentiation of 2KI wild-type (*Aldh1a2*$^{+/+}$) (F) or 2KI *Aldh1a2*$^{-/-}$ (G) mESCs transgenic for the cDR RA signaling reporter. Bar: 100 μm.

(legend continued on next page)

synthesis or inhibited RARs during PS-like differentiation (Figure 5B). The fraction of Sox1^{GFP+} cells, in fact, should be insensitive to these treatments in absence of RA production. However, the RAR antagonist AGN led to a decrease of the fraction of Sox1^{GFP+} cells (Figures 5C and 5D). Upon addition of vitamin A, a majority of cells expressed Sox1^{GFP} and the DR5-RARE-Scarlet reporter could now be detected in *Aldh1a2*$^{-/-}$ differentiating cultures (Figures 5C, 5D, S5E, and S5F). Altogether, *Aldh1a2* loss did not abrogate RA signaling and this, in turn, could not be fully captured by the DR5-RARE reporter.

We wondered whether *Aldh1a2*$^{-/-}$ cells could still respond to RA at the concentration present in wild-type cultures and, vice versa, whether the response of *Aldh1a2*$^{+/+}$ cells to RA was affected by the presence of cells impaired in RA synthesis. To address both questions, we set up a co-culture experiment by mixing wild-type *Aldh1a2*$^{+/+}$ cells and mutant *Aldh1a2*$^{-/-}$ cells (Figure S5G). Under such conditions, DR5-RARE-Scarlet^{+} cells were found in the *Aldh1a2*$^{-/-}$ fraction at a rate comparable to the one in *Aldh1a2*$^{+/+}$ cells (Figures S5H and S5I). This proved that RA signaling was paracrine in our *in vitro* system. Moreover, the fraction of *Aldh1a2*$^{+/+}$ cells expressing the RA reporter was reduced in the co-culture setting (Figure S5I) compared to a pure wild-type culture (Figure S5H). This observation implied that a cell's response to RA did not depend on its own RA production but rather on the overall RA level present in the medium, that is the regulation of Scarlet expression was non-cell-autonomous.

A highly sensitive RA reporter captures Aldh1a2-independent RA signaling

Our results stressed that the DR5-based RARE might capture only a subset of conditions in which RA signaling was present. Thus, we turned to a composite RARE (cDR-RARE) consisting of three RAR binding sites (Figure 5E) that was found to have much higher affinity for RARs than the DR5-based RARE (Moutier et al., 2012). A reporter construct relying on cDR-RARE driving Scarlet expression could detect sub-nanomolar concentrations of exogenously applied RA (Figure S5J). Many more cDR-RARE-Scarlet^{+} cells could be detected during PS-like differentiation (Figure 5F) compared with the DR5-based reporter (Figures S5K and S5L). The complete abrogation of cDR-RARE-Scarlet expression upon RAR antagonism confirmed its reliance on RA signaling (Figure S5L). Crucially, the PS-like differentiation of *Aldh1a2*$^{-/-}$ cells bearing this reporter confirmed that RA signaling was reduced but not absent (Figures 5G, S5L–S5N). As for the DR5-based reporter, the cDR-RARE-Scarlet reporter was expressed particularly in T$^{TagBFP+}$ cells (Figure S5M). Altogether, a fraction of RA was produced in an Aldh1a2-independent manner at sufficient levels to be detected by the cDR-RARE reporter and to impact the formation of Sox1^{GFP+} cells.

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

The increase of the fraction of Sox1^{GFP+} cells with vitamin A levels in the medium (Figure S4E) implied that cells were sensitive to the external vitamin A concentration. Interestingly, the transcript levels of the cellular 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 during PS-like differentiation (Figures S5O and S5P). Elevated expression levels of *Rbp1* were found in the mouse PS (Ruberte et al., 1991). RBP1 binds to vitamin A and is thought to increase its intracellular concentration and to help RA synthesis (Napoli, 2016). We hypothesized that vitamin A uptake through STRA6 and intracellular storage by RBP1 could contribute to control RA levels. Therefore, we generated *Rbp1*$^{-/-}$*Stra6*$^{-/-}$ mESCs (Figure S5Q) and submitted them to PS-like differentiation. *Rbp1*$^{-/-}$*Stra6*$^{-/-}$ cells generated fewer NPs compared to their wild-type counterparts, but most cells were Sox1^{GFP+} when increasing vitamin A concentration (Figures 5H and 5I). Similarly, the fractions of both Sox1^{GFP+} and cDR-RARE-Scarlet^{+} populations were decreased after PS-like differentiation of *Aldh1a2*$^{-/-}$*Rbp1*$^{-/-}$*Stra6*$^{-/-}$ cells compared with *Aldh1a2*$^{-/-}$ cells (Figures 5J and S5R). This demonstrated that the control of intracellular vitamin A levels via the Stra6-Rbp1 axis contributed to determine RA signaling activation.

Cyp26a1 limits RA signaling during PS-like differentiation

A possible additional mechanism to define the subset of RA-responding cells is the control of RA degradation mediated by cytochrome P450 CYP26 enzymes (Rhinn and Dollé, 2012). Cyp26a1 mRNA levels were highly

(H) *Sox1*-GFP reporter expression after PS-like differentiation of 2KI cells (gray) or *Rbp1*$^{-/-}$*Stra6*$^{-/-}$ cells without (blue) or with (red) additional vitamin A.

(I) Quantification of (H) data (n = 3 independent experiments; *, p < .05; ***, p < .001; two-sided unpaired t test; data represented as mean ± SD).

(J) *Sox1*-GFP and cDR-RARE-Scarlet reporter expression after PS-like differentiation of 2KI mESCs, 2KI mESCs transgenic for the cDR RA signaling reporter (*Aldh1a2*$^{+/+}$), *Aldh1a2*$^{-/-}$, or *Aldh1a2*$^{-/-}$*Rbp1*$^{-/-}$*Stra6*$^{-/-}$ mESCs. Dotted lines: gates fixed according to the non-transgenic mESCs negative control. See also Figure S5.

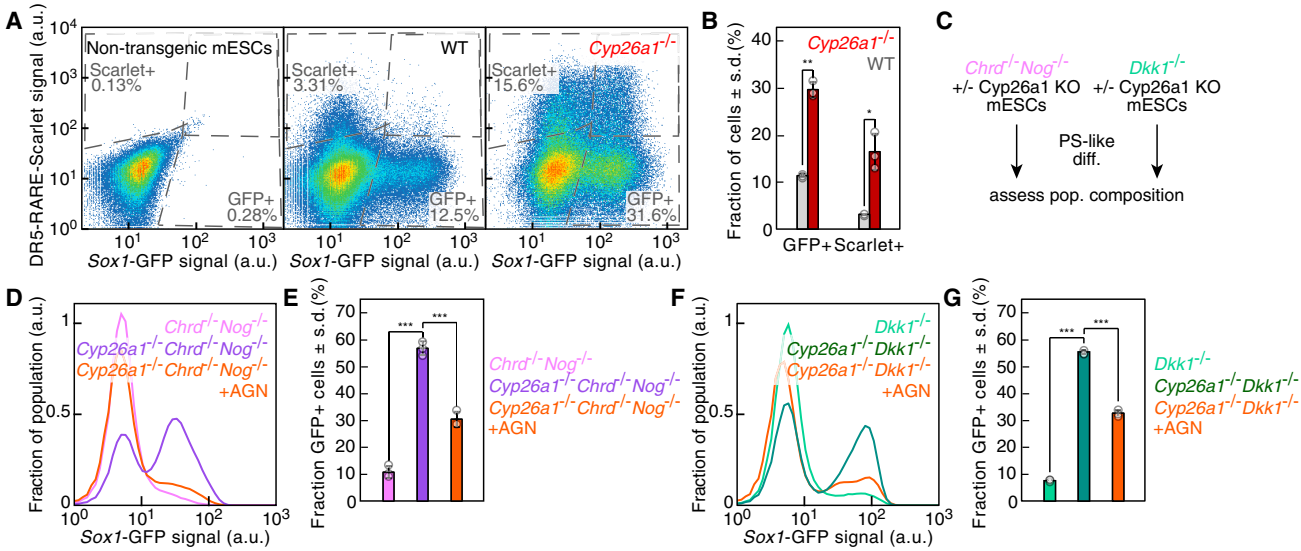

Figure 6. Cyp26a1 limits RA levels and neuroectoderm differentiation during PS-like differentiation

(A) *Sox1*-GFP and DR5-RARE-Scarlet reporter expression after PS-like differentiation of non-transgenic wild-type (WT) or *Cyp26a1*$^{-/-}$ mESCs. Dotted lines: gates fixed according to the non-transgenic mESCs negative control.

(B) Quantification of (A) data (n = 3 independent experiments; **, p < .01; ***, p < .001; two-sided unpaired t test; data represented as mean ± SD).

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

(D and E) *Sox1*-GFP reporter expression in *Chrd*$^{-/-}$*Nog*$^{-/-}$ (pink) or *Cyp26a1*$^{-/-}$*Chrd*$^{-/-}$*Nog*$^{-/-}$ (purple, orange: AGN added after day 3) cells after PS-like differentiation (D) and quantification (E) (n = 3 independent experiments; ***, p < .001; One-way ANOVA followed by Tukey's post hoc test; data represented as mean ± SD).

(F and G) *Sox1*-GFP reporter expression in *Dkk1*$^{-/-}$ (light green) or *Cyp26a1*$^{-/-}$*Dkk1*$^{-/-}$ (dark green, orange: AGN added after day 3) cells after PS-like differentiation (F) and quantification (G) (n = 3 independent experiments; ***, p < .001; one-way ANOVA followed by Tukey's post hoc test; data represented as mean ± SD). See also Figure S6.

upregulated during PS-like differentiation and particularly in T$^{TagBFP+}$ cells compared with Sox1^{GFP+} cells (Figures S6A and S6B), mirroring its expression pattern in the PS *in vivo* at E7.0 (Fujii et al., 1997). To test whether Cyp26a1 hampered neural induction by limiting RA levels, we inactivated *Cyp26a1* (Figure S6C). Compared with wild-type cells, *Cyp26a1*$^{-/-}$ cultures produced more Sox1^{GFP+} cells and displayed many more cells with active RA signaling (Figures 6A and 6B). Cyp26a1 loss further enhanced NP formation upon vitamin A addition (Figures S6D and S6E).

We investigated how the loss of Cyp26a1 impacted neural fate acquisition in response to different RA concentrations (Figure S6F). At low RA concentrations, the majority of the *Cyp26a1*$^{-/-}$ cells were Sox1^{GFP+}, in contrast to wild-type cells (Figure S6G). However, *Cyp26a1*$^{-/-}$ and wild-type cells displayed similar capacity to differentiate to neuroectoderm at higher RA concentrations (Figure S6H). These findings demonstrated that Cyp26a1 plays a key role in reducing RA levels and in the acquisition of neural fate during PS-like differentiation, particularly in response to low RA concentrations.

Impaired RA degradation increases neuroectoderm formation in *Chrd*$^{-/-}$*Nog*$^{-/-}$ and *Dkk1*$^{-/-}$ cells

We sought to test whether the impaired neuroectoderm formation due to the absence of TGFβ or Wnt inhibitors could be counteracted by Cyp26a1 loss, which increases the response to RA signaling. Therefore, we generated mESCs lacking Cyp26a1 and either TGFβ or Wnt antagonists and subjected them to PS-like differentiation (Figures 6C, S6I, and S6J). *Chrd*$^{-/-}$*Nog*$^{-/-}$*Cyp26a1*$^{-/-}$ and *Dkk1*$^{-/-}$*Cyp26a1*$^{-/-}$ cells formed more NPs compared with *Chrd*$^{-/-}$*Nog*$^{-/-}$ or *Dkk1*$^{-/-}$ cells (Figures 6D–6G). RAR inhibition reversed this increase, proving that the effect was strictly dependent on RA signaling (Figures 6D–6G). Altogether, Cyp26a1-mediated dampening of RA signaling was a critical mechanism to reduce the exposure of the PS-like population to the differentiating effects of the RA they produce.

RA signaling status accounts for NP diversity

We went on to characterize the gene expression changes associated with active (cDR-RARE-Scarlet^{+} cells) or inactive or low (AGN-treated cells and *Aldh1a2*$^{-/-}$ cells) RA

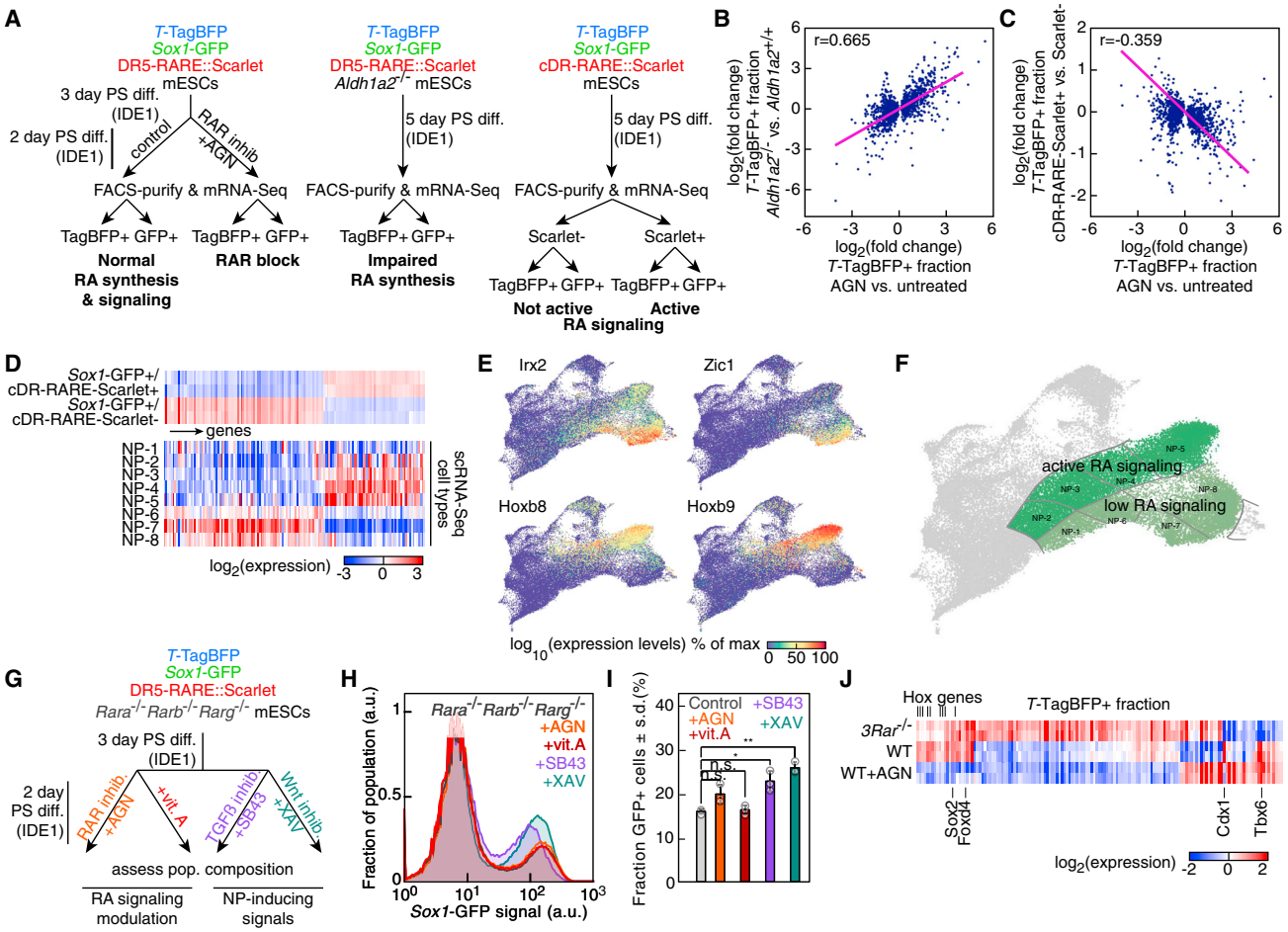

Figure 7. Role of RA signaling and RARs in neural commitment and in establishing NP diversity

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

(B and C) Comparison of differential expression in $T^{TagBFP+}$ cells after AGN treatment or in *Aldh1a2*$^{-/-}$ cells (B) (Pearson's r = 0.665, p = 10^{-181}), or in cDR-RARE-Scarlet^{+} cells with active RA signaling (C) (Pearson's r = −0.359, p = 2.10^{-44}).

(D) Expression levels of genes differentially expressed between cDR-RARE-Scarlet^{+} and Scarlet^{-} subpopulations of the $Sox1^{GFP+}$ fraction after PS-like differentiation and in the eight 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 and I) *Sox1*-GFP reporter expression in *Rara*$^{-/-}$*Rarb*$^{-/-}$*Rarg*$^{-/-}$ cells after PS-like differentiation (H) (black: control, orange: AGN, red: vitamin A, purple: SB43, teal blue: XAV) and quantification (I) (n = 3 independent experiments; *, p < .05; **, p < .01; n.s.: not significant; two-sided unpaired t test; data represented as mean ± SD).

(J) Expression levels of transcription 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.

signaling both in $T^{TagBFP+}$ and $Sox1^{GFP+}$ subpopulations (Figure 7A). The expression changes observed in the $T^{TagBFP+}$ cells upon AGN treatment were well correlated with the ones observed in *Aldh1a2*$^{-/-}$ cells (Figure 7B) and were anti-correlated with the expression changes in cDR-RARE-Scarlet^{+} $T^{TagBFP+}$ cells (Figure 7C). Active RA signaling either repressed (Figure S7A) or increased (Figure S7B) target gene expression. We next assessed genes differentially expressed in the $Sox1^{GFP+}$ cells distinguished by their RA signaling status and their expression in the NP subpopulations defined by scRNA-seq (Figure 7D). The NP-2, NP-3, NP-4, and NP-5 populations 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 transcription factors differentially expressed were NP markers associated with distinct anteroposterior identity such as *Irx2*, *Zic1*, *Hoxb8*, and *Hoxb9* (Figure 7E). Genes characteristic of anterior NP identity were upregulated in the NP-1 subpopulation (Figures S7C and S7D) and in AGN-treated cells compared with cDR-RARE-Scarlet^{-} (Figure S7E). Altogether, the activation status of RA signaling accounted for differences in the signatures of the NP subpopulations identified by scRNA-seq. We broadly distinguished two NP subgroups with high or low RA signaling, marked by different expression levels of RA target genes such as *Rarb*, *Cdx1*, and *Neurog2* (Figures 7F and S7F).

RAR knockout cells exhibit increased propensity to neuroectoderm differentiation

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). Indeed, RA failed to induce the expression of the DR5-RARE-based reporter in *Rara*$^{-/-}$*Rarb*$^{-/-}$*Rarg*$^{-/-}$ mESCs (Figure S7H). Despite the absence of RA signaling, RAR-null cells generated a Sox1^{GFP+} population after PS-like differentiation. We tested whether neural fate induction in these cells was responsive to alterations of RA signaling or to TGFβ and Wnt inhibition (Figure 7G). The formation of Sox1^{GFP+} cells was completely insensitive to the addition of vitamin A or AGN (Figures 7H and 7I), unlike the *Aldh1a2*$^{-/-}$ condition. A partial increase of the fraction of NPs was observed after inhibiting TGFβ or Wnt signaling (Figures 7H and 7I). The results show a clear functional distinction with regard to neuroectoderm formation between RAR inhibition and a complete RAR loss.

To understand the differences at the transcriptional level, we compared the gene expression profile of the T$^{TagBFP+}$ and Sox1^{GFP+} subpopulations of *Rara*$^{-/-}$*Rarb*$^{-/-}$*Rarg*$^{-/-}$ cells and the ones of wild-type cells or of AGN-treated cells. Interestingly, the expression of *Tbx6*, whose loss *in vivo* leads to the formation of neural tissue at the expense of somites (Chapman and Papaioannou, 1998), was reduced in T$^{TagBFP+}$ RAR-null cells (Figure 7J). In addition, these cells upregulated *Sox2* (Figure 7J), whose misexpression in paraxial mesoderm causes ectopic neural tube formation (Takemoto et al., 2011). RAR-null Sox1^{GFP+} cells downregulated genes of the Hox and Cdx families (Figure S7I). Furthermore, the expression of some RA target genes, such as *Rarb*, *Hoxb1*, and *Neurog2*, was lower in AGN-treated cells compared with RAR-null ones (Figure S7I). Thus, the lack of RARs had distinct effects compared with their pharmacological inhibition and could not be assimilated to an absence of RA signaling. Interestingly, in medium devoid of RA precursors, the differentiation of RAR-null cells yielded many more Sox1^{GFP+} cells compared with wild-type cells, demonstrating the importance of RARs in the homeostasis of neuroectoderm formation (Figures S7J and S7K).

DISCUSSION

In this work, we combined a culture system reproducing the maturation of primitive streak-like cells and the formation of both anterior and posterior neuroectodermal fates with a large collection of reporter mESC lines harboring genetic ablations of key signaling factors. In such a context, we showed that RA signaling drove early neural induction downstream of Wnt or TGFβ inhibition and that multiple components of the RA pathway contribute to neuroectoderm differentiation.

The inhibition of Wnt or TGFβ pathways starts the formation of neural lineage in the anterior epiblast of the mouse conceptus. The reduced NP formation in *Chrd*$^{-/-}$*Nog*$^{-/-}$ or *Dkk1*$^{-/-}$ cells was reminiscent of the corresponding mutant mouse embryos lacking anterior neural structures (Bachiller et al., 2000; Mukhopadhyay et al., 2001). The PS-like differentiation generated a spectrum of anterior and posterior neuroectoderm. These NP subpopulations differed in their RA signaling status, with markers of anterior fates being expressed in cells with low RA signaling. This is in accordance with the proposed caudalizing effects of RA during development (Durston et al., 1989). More importantly, we showed that the mechanisms of neural induction at work in the PS-like culture were not independent, because neural induction through inhibition of Wnt or TGFβ signaling was hindered by blocking RA receptors. This suggested that the transduction of RA signaling mediates a step downstream of Wnt and TGFβ inhibition in the cascade of events leading to the acquisition of neural fate.

Aldh1a2-mediated synthesis was crucial to generate high levels of RA signaling during PS-like differentiation but did not account for all RA production. While the presence of RA signaling during the differentiation of *Aldh1a2*$^{-/-}$ cells awaits confirmation in future *in vivo* studies, its significance lies in the identification of alternative ways to respond to RA. Indeed, the different sensitivity of distinct RAREs to RA levels would enable cells to switch on different gene repertoires depending on RA concentration, thus generating positional information.

In order to safeguard their developmental capabilities, pluripotent cells such as the epiblast/PS-like population should protect themselves from the neuralizing action of RA and therefore need to carefully control the RA levels they are exposed to. We found that the PS-like cells tuned RA synthesis via the regulation of vitamin A availability

through the Rbp1-Stra6 axis. Furthermore, we determined that RA synthesis was not entirely dependent on Aldh1a2 and could not be attributed to another single aldehyde dehydrogenase by systematically knocking out all the ones expressed in PS cells. Our co-culture experiment showed that a cell's response to RA was not intrinsically determined by its own RA production. In this context, Cyp26a1-mediated RA degradation is a crucial checkpoint, limiting the differentiation toward the neural lineage of PS-like cells. 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.

The differentiation of *Rara*$^{-/-}$*Rarb*$^{-/-}$*Rarg*$^{-/-}$ mESCs underlined an even more complex involvement of the RA pathway in neuroectoderm differentiation. Unlike *Aldh1a2*$^{-/-}$ cells, RAR-null cells were completely devoid of RA signaling. NP formation in absence of RARs seems in contradiction with the reduction of neuroectoderm induction by blocking RA signaling. However, this can be explained by the binding of the receptors to their cognate RAREs in the absence of RA (Chambon, 1996). Their physical absence in RAR-null cells would remove this control mechanism and unmask binding sites, making them available to other nuclear receptors and transcription factors. Thus, besides being effectors of RA signaling, RARs might gate the expression at genomic loci important for neural specification.

In conclusion, the flexibility of our *in vitro* system allowed the manipulation of the external environment in a controlled manner. This led us to recognize the involvement of RA signaling in early neural induction. Our results highlight the potential of ESC-based systems to gain new insights about lineage specification mechanisms. Notwithstanding the strengths of this approach, it will be beneficial to exploit the tools we developed and test our findings in *in vivo* models or more complex tridimensional culture systems.

EXPERIMENTAL PROCEDURES

mESC maintenance

The parental mESC line was a *Sox1-Brachyury* double knock-in line (Sladitschek and Neveu, 2019). mESCs were maintained in "LIF + serum" as described previously (Sladitschek and Neveu, 2015b).

Generation of knockout mESC lines

RNA-guided Cas9 nucleases (Hsu et al., 2013) were used to inactivate *Aldh1a2*, *Chrd*, *Cyp26a1*, *Dkk1*, *Nog*, *Rara*, *Rarb*, *Rarg*, *Rbp1*, and *Stra6*. See supplemental experimental procedures for details.

Reporter constructs

Constructs were assembled following Sladitschek and Neveu (2015a). Transcriptional reporters relied on mScarlet (Bindels et al., 2017) and different RAREs (Moutier et al., 2012; Rossant et al., 1991). See supplemental experimental procedures for details.

Transgenic mESC lines

A list of all transgenic cell lines used in this study can be found in supplemental experimental procedures.

Primitive streak-like differentiation

Differentiation toward a primitive streak-like fate was performed using IDE1 (Sladitschek and Neveu, 2019) or a pulse of ACTIVIN A. See supplemental experimental procedures for details.

Neural progenitor differentiation

mESCs were differentiated to NPs using RA or inhibition of TGFβ or Wnt signaling. See supplemental experimental procedures for details.

Pharmacological treatments

Details of pharmacological treatments can be found in supplemental experimental procedures.

Imaging

Reporter fluorescence was assessed in live cells. Images were acquired on an inverted SP8 confocal microscope (Leica) equipped with a 40× PLApo 1.1W objective and an incubation chamber at 37°C and 5% CO_2.

Flow cytometry

Cells were FACS-purified using an Aria Fusion sorter (BD BioSciences). Samples were analyzed on an LSRFortessa flow cytometer (BD BioSciences), and data was analyzed with FlowJo. See supplemental experimental procedures for details.

RNA-seq library construction

mRNA sequencing was conducted as previously described (Sladitschek et al., 2020). See supplemental experimental procedures for details.

RNA-seq analysis

mRNA read counts were determined using Bowtie (Langmead et al., 2009). edgeR (Robinson et al., 2010) was used for differential gene expression analysis. See supplemental experimental procedures for details.

Single-cell RNA sequencing

Samples for scRNA-seq were processed with a Chromium Controller and reagents (10× Genomics). See supplemental experimental procedures for details.

scRNA-seq analysis

scRNA-seq data was pre-processed as described in supplemental experimental procedures. 46,700 cells passed quality controls. Expression levels were normalized using Seurat methods (Satija et al., 2015). Dimensionality reduction was performed using

UMAP (Becht et al., 2018). Clustering and marker determination are described in supplemental experimental procedures.

Statistical analysis

Statistical tests were computed using R or the Python SciPy module. Data is represented as mean ± SD. Two-sided unpaired Student's t test was used for pairwise comparison with a fixed control condition. For multiple pairwise comparisons with different control and treatment conditions, one-way ANOVA analysis followed by Tukey's post hoc test was used. Values with $p < 0.05$ were considered significant.

Data and code availability

Sequencing results are deposited on ArrayExpress with accession numbers ArrayExpress: E-MTAB-10242 and ArrayExpress: E-MTAB-10243. In addition, we used the datasets ArrayExpress: E-MTAB-2830, ArraxExpress: E-MTAB-3234 (Sladitschek and Neveu, 2015b), and ArrayExpress: E-MTAB-4904 (Sladitschek and Neveu, 2019).

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stemcr.2021.12.014.

AUTHOR CONTRIBUTIONS

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

CONFLICT OF INTERESTS

The authors declare no competing interests.

ACKNOWLEDGMENTS

We thank Lucia Cassella for advice on RNA-seq data analysis and 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 Program (EIPP).

Received: November 22, 2021
Revised: December 15, 2021
Accepted: December 16, 2021
Published: January 20, 2022

REFERENCES

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

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. *37*, 38–44.

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.

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.

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

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

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.

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.

Henrique, D., Abranches, E., Verrier, L., and Storey, K.G. (2015). Neuromesodermal progenitors and the making of the spinal cord. Development *142*, 2864–2875.

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.

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.

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.

Napoli, J.L. (2016). Functions of intracellular retinoid binding-proteins. Subcell. Biochem. *81*, 21–76.

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.

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.

Samarut, E., and Rochette-Egly, C. (2012). Nuclear retinoic acid receptors: conductors of the retinoic acid symphony during development. Mol. Cell. Endocrinol. *348*, 348–360.

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.

Sladitschek, H.L., Fiuza, U.-M., Pavlinic, D., Benes, V., Hufnagel, L., and Neveu, P.A. (2020). Morphoseq: full single-cell transcriptome dynamics up to gastrulation in a chordate. Cell *181*, 922–935.e21.

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.

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.

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.

Stem Cell Reports, Volume *17*

Supplemental Information

Multi-layered regulation of neuroectoderm differentiation by retinoic acid in a primitive streak-like context

Luigi Russo, Hanna L. Sladitschek, and Pierre A. Neveu

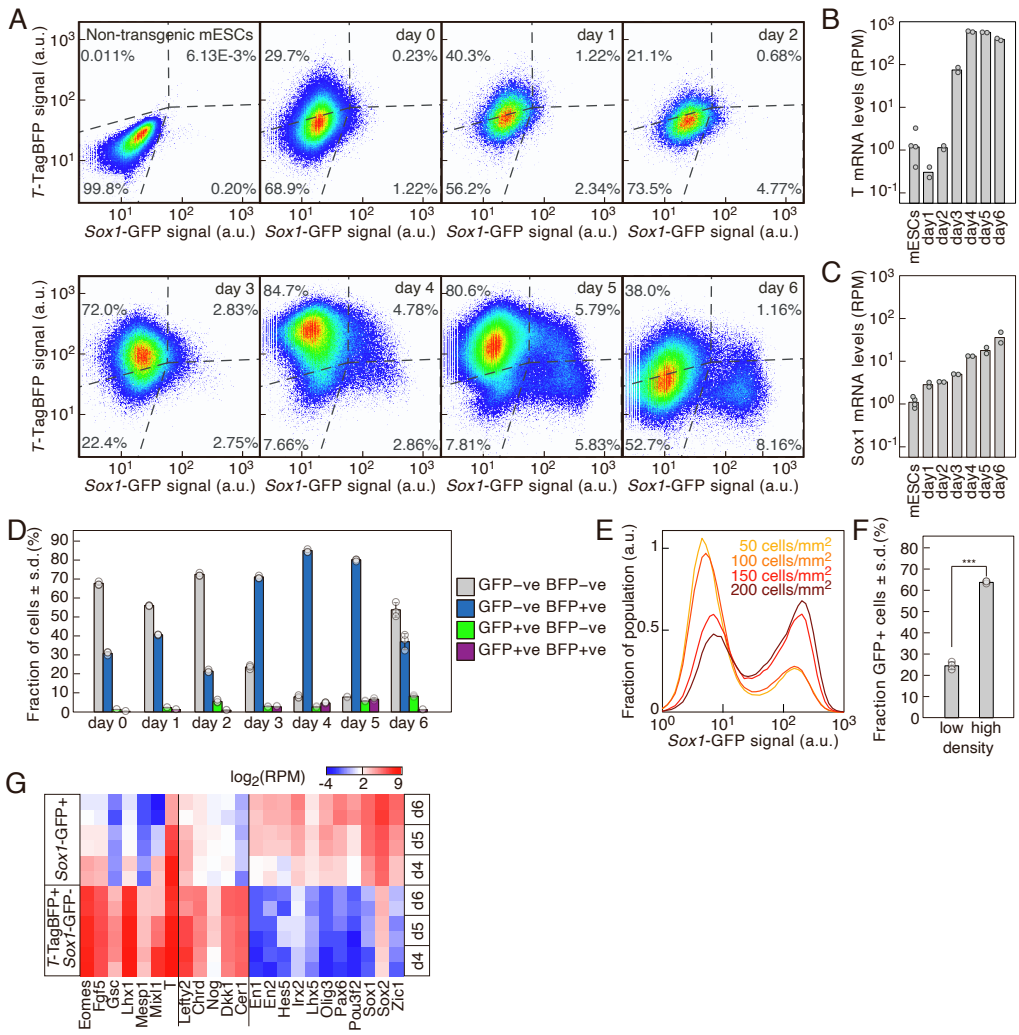

Figure S1. Related to Figure 1. Characterization of neural induction by primitive streak-like cells.
(A) Time course of *T*-TagBFP and *Sox1*-GFP reporter expression as measured by flow cytometry each day of the IDE1 PS-like differentiation. The day 0 condition corresponds to mESCs cultured in the pluripotency maintaining medium. Dotted lines: gates fixed according to the non-transgenic mESCs negative control.
(B) T mRNA levels in cultures undergoing PS-like differentiation (RPM: reads per million mapped reads, n=2 independent experiments).
(C) Sox1 mRNA levels in cultures undergoing PS-like differentiation (RPM: reads per million mapped reads, n=2 independent experiments).
(D) Quantification of Figure S1A data (n=3 independent experiments; data represented as mean±SD).
(E) *Sox1*-GFP expression levels in cultures undergoing PS-like differentiation with increasing starting cell density (yellow: 50 cells/mm^2, orange: 100 cells/mm^2, red: 150 cells/mm^2, dark red: 200 cells/mm^2).
(F) Quantification of Figure S1E data. The mean percentage of cells expressing the GFP reporter is presented in relation to the initial density at the seeding: low (50 cells/mm^2) and high (150 cells/mm^2). (n=3 independent experiments; ***, $p < .001$; two-sided unpaired t-test; data represented as mean±SD).
(G) Related to Figure 1D. Heatmap displaying the $\log_2$ of the expression levels returned as RPM counts (RPM: reads per million mapped reads) of markers of the primitive streak or neural progenitors in FACS-purified populations after 4 to 6 days of PS-like differentiation.

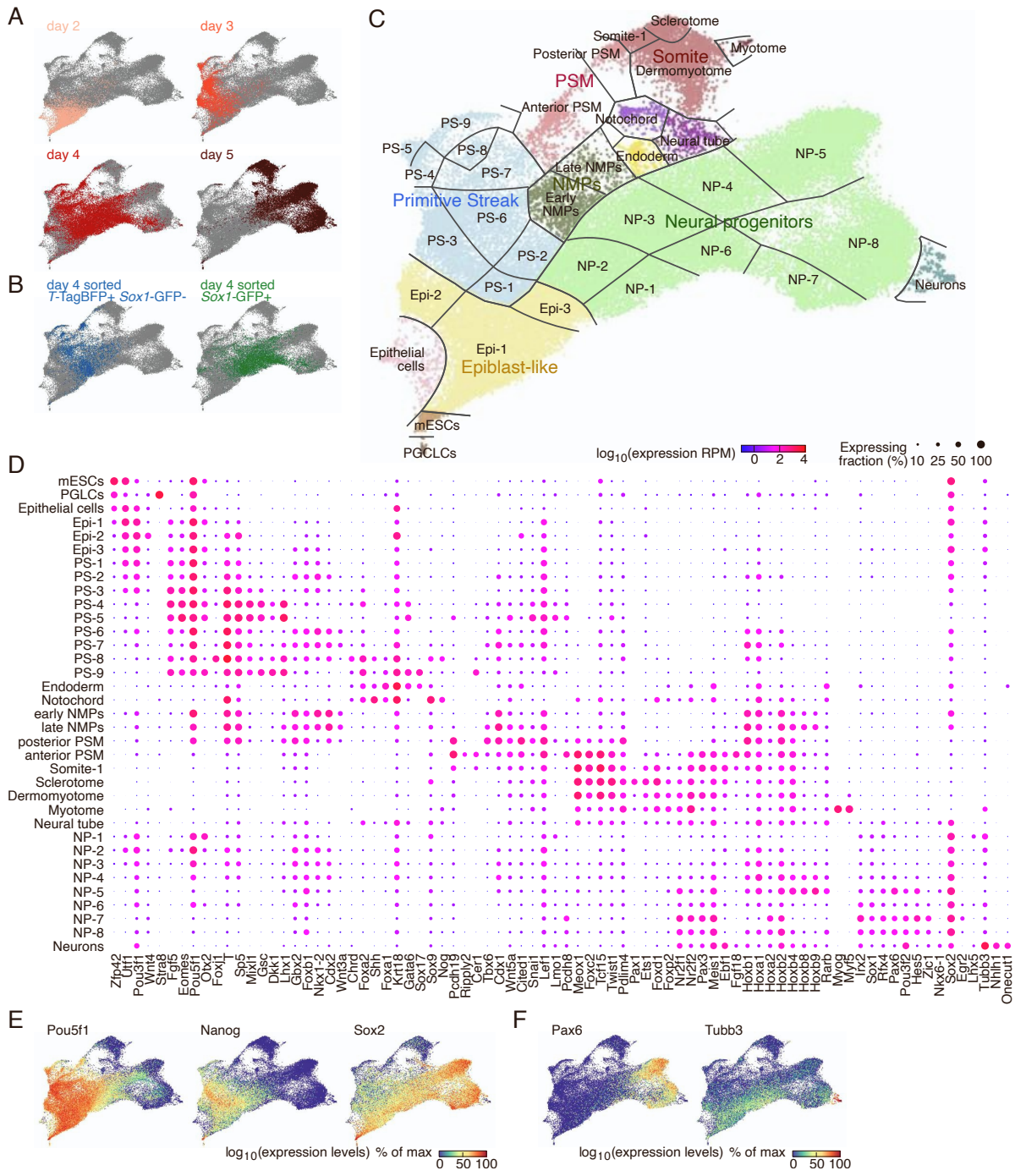

Figure S2. Related to Figure 2. Characterization of neural induction by primitive streak-like cells at the single-cell level.

(A) UMAP colored by the day of differentiation in which the corresponding cells were analyzed.

(B) UMAP location of cells FACS-purified according to their *T*-TagBFP and *Sox1*-GFP expression at day 4.

(C) UMAP colored according to the identified populations (NMPs: neuromesodermal progenitors, NP: neural progenitors, PSM: presomitic mesoderm, PGCLCs: primordial germ cell-like cells, PS: Primitive streak).

(D) Dot plots of marker expression levels in the 35 identified populations (RPM: reads per million mapped reads).

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

(F) UMAP colored by the scaled expression of the neural progenitor marker *Pax6* and the neuronal marker *Tubb3*.

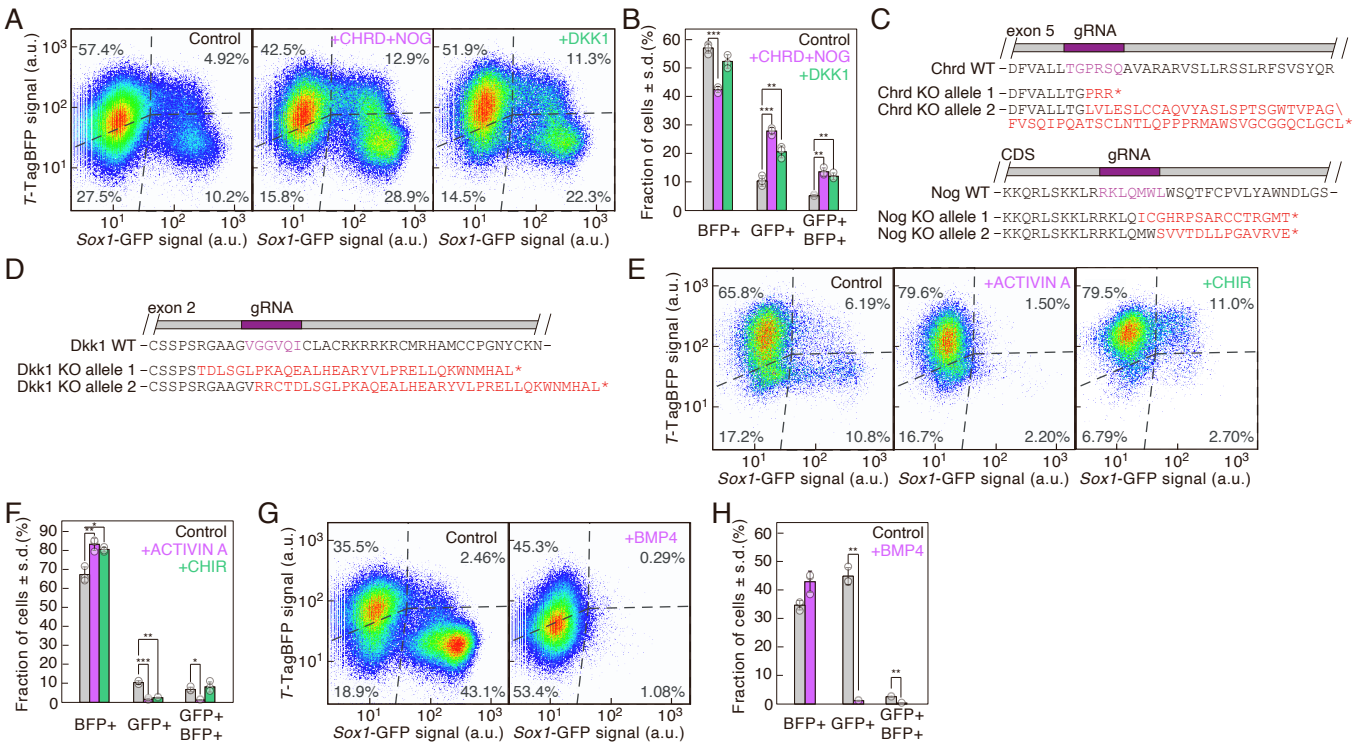

Figure S3. Related to Figure 3. A balance between agonists and inhibitors of the TGFβ and Wnt signaling pathways fine tunes the formation of neuroectodermal and primitive streak derivatives in culture.

(A) *T*-TagBFP and *Sox1*-GFP reporter expression after 5 days of PS-like differentiation in IDE1 (Control), or IDE1 supplemented with recombinant CHORDIN and NOGGIN (+CHRD+NOG) or DKK1. Dotted lines: gates fixed according to the non-transgenic mESCs negative control.

(B) Quantification of Figure S3A data (n=3 independent experiments; **, $p < .01$; ***, $p < .001$; two-sided unpaired t-test; data represented as mean$\pm$SD).

(C) Sanger sequencing-validated obtained alleles of *Chrd*$^{-/-}$*Nog*$^{-/-}$ mESCs. The relative position of the guide RNA used to target the locus is indicated in purple. $\star$: stop codon.

(D) Sanger sequencing-validated obtained alleles in *Dkk1*$^{-/-}$ mESCs. The relative position of the guide RNA used to target the locus is indicated in purple. $\star$: stop codon.

(E) *T*-TagBFP and *Sox1*-GFP reporter expression after 5 days of PS-like differentiation in IDE1 (Control), or IDE1 supplemented from day 3 with the Nodal/TGFβ agonist ACTIVIN A or the Wnt agonist CHIR. Dotted lines: gates fixed according to the non-transgenic mESCs negative control.

(F) Quantification of Figure S3E data (n=3 independent experiments; *, $p < .05$; **, $p < .01$; ***, $p < .001$; two-sided unpaired t-test; data represented as mean$\pm$SD).

(G) *T*-TagBFP and *Sox1*-GFP reporter expression after 5 days of PS-like differentiation in IDE1 (Control), or IDE1 supplemented from day 3 with the BMP/TGFβ agonist BMP4. Dotted lines: gates fixed according to the non-transgenic mESCs negative control.

(H) Quantification of Figure S3G data (n=3 independent experiments; **, $p < .01$; two-sided unpaired t-test; data represented as mean$\pm$SD).

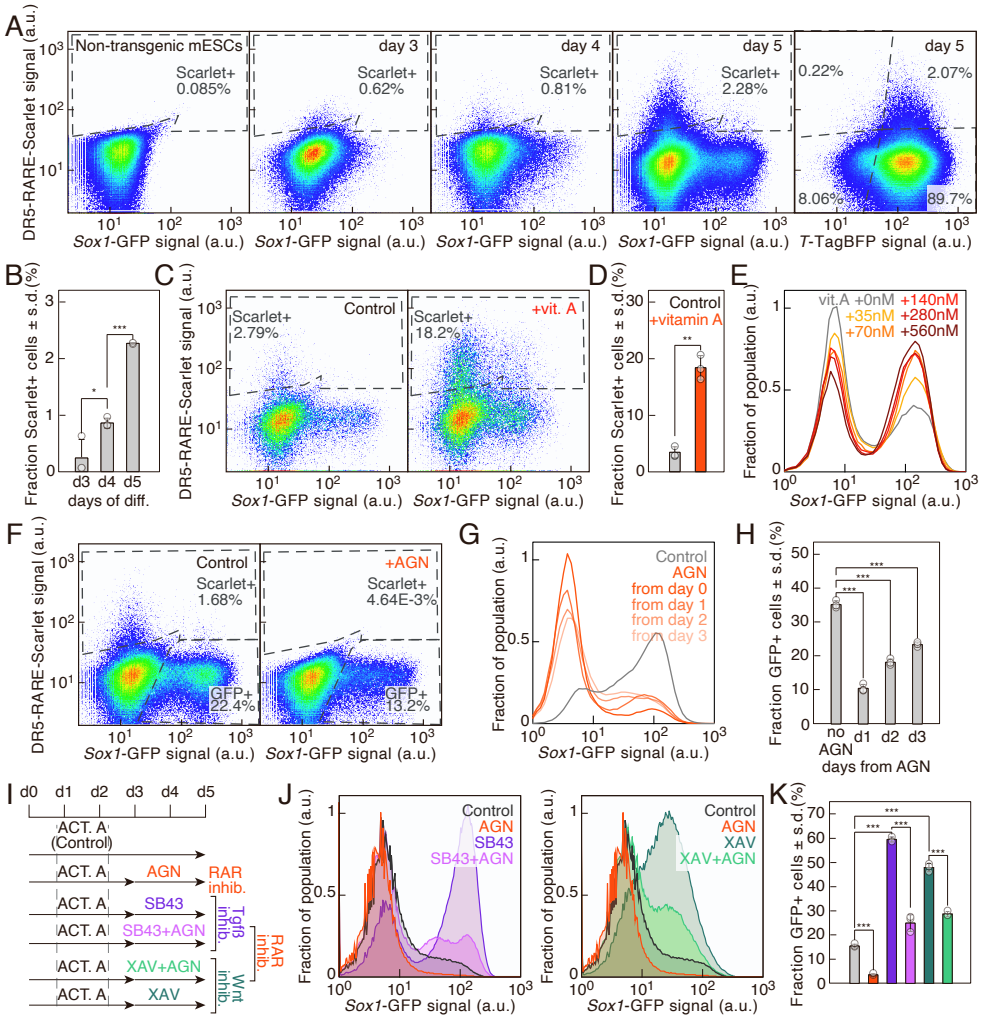

Figure S4. Related to Figure 4. Retinoic acid signaling contributes to the formation of neural progenitors during the by PS-like differentiation.

(A) Time course of the expression levels of the DR5-based RA signaling reporter (DR5-RARE::Scarlet), *Sox1*-GFP and *T*-TagBFP (right panel) in cultures undergoing PS-like differentiation. Dotted lines: gates fixed according to the non-transgenic mESCs negative control.

(B) Quantification of Figure S4A data. (n=3 independent experiments; *, $p < .05$; ***, $p < .001$; One-way ANOVA followed by Tukey's post-hoc test; data represented as mean±SD).

(C) *Sox1*-GFP and DR5-RARE-Scarlet expression levels after 5 days of standard PS-like differentiation (Control), or adding 350 nM vitamin A 24 hours before the measurement (+vit.A). Dotted lines: gates fixed according to the non-transgenic mESCs negative control.

(D) Quantification of Figure S4C data (n=3 independent experiments; **, $p < .01$; two-sided unpaired t-test; data represented as mean±SD).

(E) *Sox1*-GFP expression levels in cultures undergoing PS-like differentiation in medium with additional vitamin A (vit. A), a precursor of RA (gray: no additional vit. A, line color according to the indicated concentration in nM).

(F) *Sox1*-GFP and DR5-RARE-Scarlet expression levels after 5 days of standard PS-like differentiation (Control), or the addition of the RAR antagonist AGN for the last 48 hours before the analysis (+AGN). For quantification and statistical analysis see Figure 4G. Dotted lines: gates fixed according to the non-transgenic mESCs negative control.

(G) *Sox1*-GFP expression levels in cultures undergoing PS-like differentiation in presence of the RAR antagonist AGN added at different time points (gray: control, orange line color according to starting day of AGN treatment).

(H) Quantification of Figure S4G data (n=3 independent experiments; ***, $p < .001$; two-sided unpaired t-test; data represented as mean±SD).

(I) Alternative experimental strategy to asses the crosstalk between RA signaling and TGFβ or Wnt pathway inhibition on fate induction in absence of IDE1. The formation of the PS like population was induced in this case by a pulse of the Nodal/TGFβ agonist ACTIVIN A from day 1 to day 2. The inhibitors were added from 3 days of differentiation onwards. AGN193109 (AGN) is a retinoic acid receptors (RARs) antagonist, SB431542 (SB43) inhibits TGFβ receptors and XAV939 (XAV) is a tankyrase inhibitor.
(J) (left panel) *Sox1*-GFP expression levels after PS-like differentiation induced by the ACTIVIN A pulse (black: Control), or after the addition to the medium of AGN (orange: +AGN), the TGFβ inhibitor SB43 (purple: +SB43) or both (pink: +SB43+AGN). (right panel) *Sox1*-GFP expression levels after PS-like differentiation induced by the ACTIVIN A pulse (black: Control), or after the addition to the medium of AGN (orange: +AGN), the Wnt signaling inhibitor XAV (teal blue: +XAV) or both (light green: +XAV +AGN).
(K) Quantification of Figure S4J data with FIgure S4I color code (n=3 independent experiments; ***, $p < .001$; One-way ANOVA followed by Tukey's post-hoc test; data represented as mean$\pm$SD).

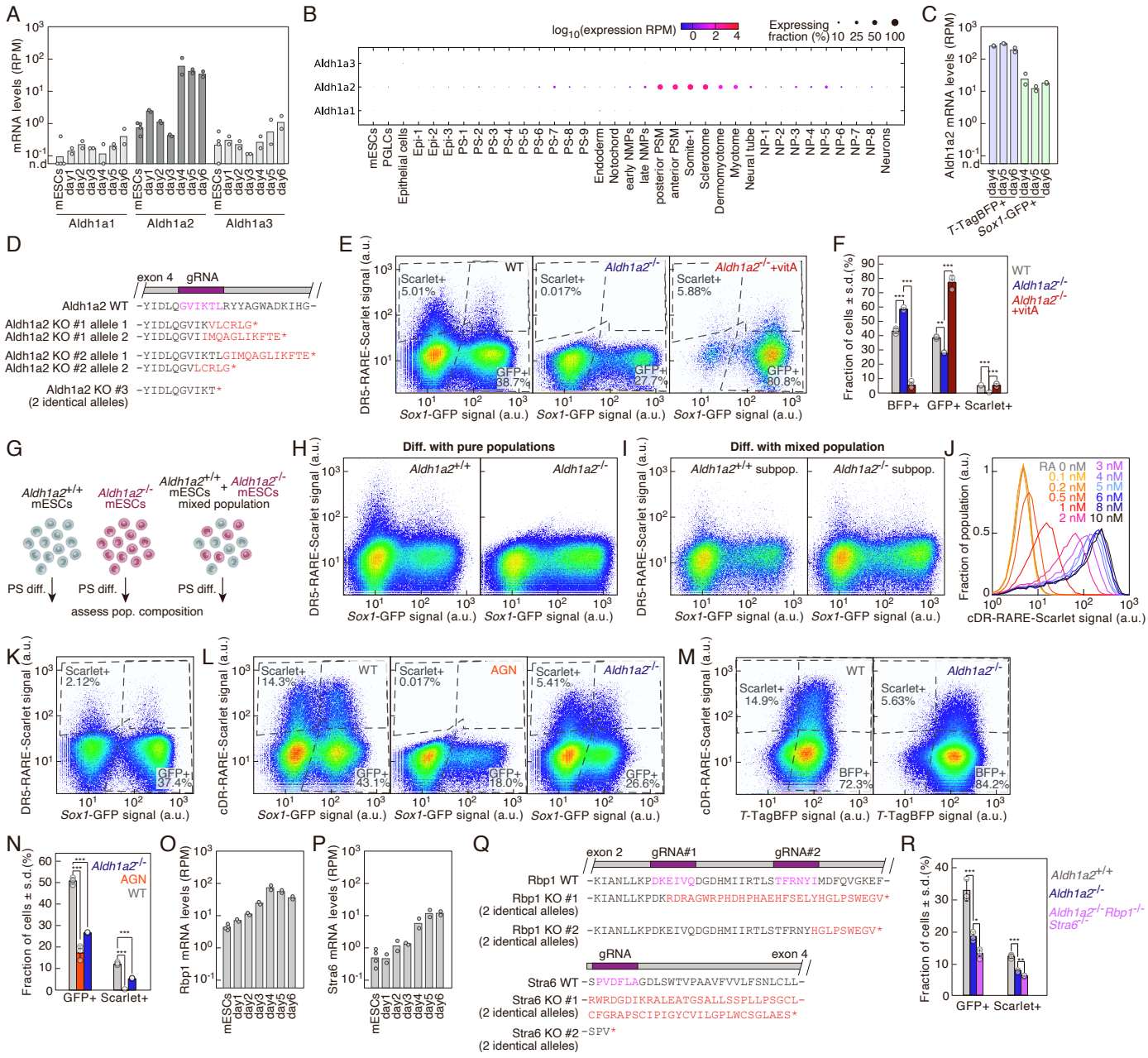

Figure S5. Related to Figure 5. Aldh1a2-independent RA signaling during PS-like differentiation.

(A) mRNA expression time course of *Aldh1a1*, *Aldh1a2*, *Aldh1a3* during PS-like differentiation as result of bulk RNA-seq (RPM: reads per million mapped reads; n=2 independent experiments).

(B) Dot plots of *Aldh1a1*, *Aldh1a2*, *Aldh1a3* expression levels in the 35 populations identified by scRNA-seq (RPM: reads per million mapped reads; n=2 independent experiments).

(C) *Aldh1a2* mRNA expression levels in $T^{TagBFP+}$ and $Sox1^{GFP+}$ cells during PS-like differentiation (RPM: reads per million mapped reads).

(D) Sanger sequencing-validated obtained alleles of *Aldh1a2*$^{-/-}$ cells. The relative position of the guide RNA used to target the locus is indicated in purple. ⋆: stop codon.

(E) *Sox1*-GFP and DR5-RARE-Scarlet reporter expression after PS-like differentiation of wild type (WT) or *Aldh1a2*$^{-/-}$ cells without (*Aldh1a2*$^{-/-}$) or with additional 140nM vitamin A (*Aldh1a2*$^{-/-}$ +vitA). Dotted lines: gates fixed according to the non-transgenic mESCs negative control.

(F) Quantification of Figure S5E data (n=3 independent experiments; **, $p < .01$; ***, $p < .001$; One-way ANOVA followed by Tukey's post-hoc test; data represented as mean±SD).

(G) Scheme of the experimental principle to assess whether RA signaling acts in a cell-autonomous manner during the PS-like differentiation by mixing wild type and *Aldh1a2*$^{-/-}$ mESCs that can be distinguished by the expression of a constitutive fluorescent marker H2B-2xiRFP670.
(H, I) *Sox1*-GFP and DR5-RARE-Scarlet reporter expression after PS-like differentiation of pure wild type (H, left panel) or *Aldh1a2*$^{-/-}$ (H, right panel) cultures or a mixed population (I) containing wild type (I, left panel) and *Aldh1a2*$^{-/-}$ (I, right panel) cells.
(J) cDR-RARE-Scarlet reporter expression after 24 hour treatment with RA (line color according to the indicated concentrations between 0 and 10 nM).
(K) *Sox1*-GFP and DR5-RARE-Scarlet reporter expression after PS-like differentiation of wild type double knockin *T*-TagBFP *Sox1*-GFP mESCs transgenic for the DR5-RARE signaling reporter. Dotted lines: gates fixed according to the non-transgenic mESCs negative control.
(L) *Sox1*-GFP and cDR-RARE-Scarlet reporter expression after PS-like differentiation of wild type double knockin *T*-TagBFP *Sox1*-GFP mESCs transgenic for the cDR-RARE signaling reporter in standard IDE1 (WT), with the addition of the RAR antagonist AGN, or of *Aldh1a2*$^{-/-}$ cells. Dotted lines: gates fixed according to the non-transgenic mESCs negative control.
(M) Related to Figure S5L. *T*-TagBFP and cDR-RARE-Scarlet reporter expression after 5 days of standard PS-like differentiation of wild type double knockin *T*-TagBFP *Sox1*-GFP mESCs transgenic for the cDR-RARE signaling reporter (WT), or of their *Aldh1a2*$^{-/-}$ counterpart. Dotted lines: gates fixed according to the non-transgenic mESCs negative control.
(N) Quantification of Figure S5L data (n=3 independent experiments; ***, $p < .001$; One-way ANOVA followed by Tukey's post-hoc test; data represented as mean$\pm$SD).
(O) *Rbp1* mRNA expression time course during PS-like differentiation according to bulk RNA-seq (RPM: reads per million mapped reads; n=2 independent experiments).
(P) *Stra6* mRNA expression time course during PS-like differentiation according to bulk RNA-seq (RPM: reads per million mapped reads; n=2 independent experiments).
(Q) Sanger sequencing-validated obtained alleles of *Rbp1*$^{-/-}$ *Stra6*$^{-/-}$ mESCs (clone $\sharp$1 in wild type cells, clone $\sharp$2 in *Aldh1a2*$^{-/-}$ cells). The relative position of the guide RNA used to target the locus is indicated in purple. $\star$: stop codon.
(R) Quantification of Figure 5J data (n=3 independent experiments; *, $p < .05$; **, $p < .01$; ***, $p < .001$; One-way ANOVA followed by Tukey's post-hoc test; data represented as mean$\pm$SD).

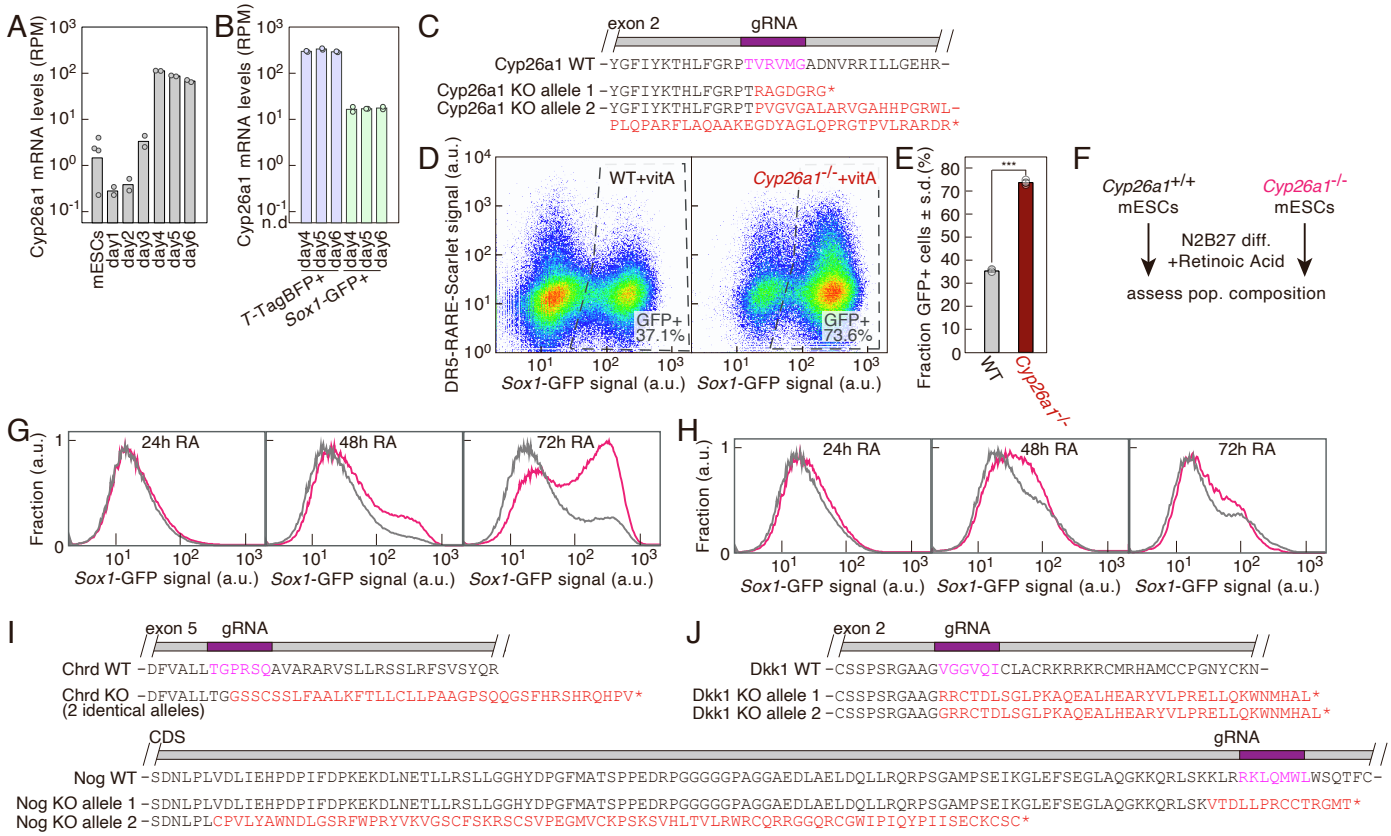

Figure S6. Related to Figure 6. Cyp26a1 is a key factor limiting RA levels and neuroectoderm differentiation during PS-like differentiation.

(A) *Cyp26a1* mRNA levels in cultures undergoing PS-like differentiation according to bulk RNA-seq (RPM: reads per million mapped reads; n=2 independent experiments).

(B) *Cyp26a1* mRNA expression levels in $T^{TagBFP+}$ and $Sox1^{GFP+}$ cells during PS-like differentiation (RPM: reads per million mapped reads; n=2 independent experiments).

(C) Sanger sequencing-validated obtained alleles of *Cyp26a1*$^{-/-}$ mESCs. The relative position of the guide RNA used to target the locus is indicated in purple. ⋆: stop codon.

(D) *Sox1*-GFP and DR5-RARE-Scarlet reporter expression after 5 days of PS-like differentiation, with additional vitamin A from day 3, of wild type (left panel) or *Cyp26a1*$^{-/-}$ cells (right panel). Dotted lines: gates fixed according to the non-transgenic mESCs negative control.

(E) Quantification of Figure S6D data (n=3 independent experiments; ***, $p < .001$; two-sided unpaired t-test; data represented as mean±SD).

(F) Scheme of the experimental principle to assess the impact of Cyp26a1 knockout on the course of neuroectoderm differentiation in response to different RA concentrations in N2B27 medium.

(G) *Sox1*-GFP reporter expression after differentiation with 1 nM RA for 24, 48 and 72 hours of wild type double knockin cells (gray) and *Cyp26a1*$^{-/-}$ cells (magenta).

(H) *Sox1*-GFP reporter expression after differentiation with 100 nM RA for 24, 48 and 72 hours of wild type double knockin cells (gray) and *Cyp26a1*$^{-/-}$ cells (magenta).

(I) Sanger sequencing-validated obtained alleles of *Chrd*$^{-/-}$*Nog*$^{-/-}$ mESCs in a *Cyp26a1*$^{-/-}$ background. The relative position of the guide RNA used to target the locus is indicated in purple. ⋆: stop codon.

(J) Sanger sequencing-validated obtained alleles of *Dkk1*$^{-/-}$ mESCs in a *Cyp26a1*$^{-/-}$ background. The relative position of the guide RNA used to target the locus is indicated in purple. ⋆: stop codon.

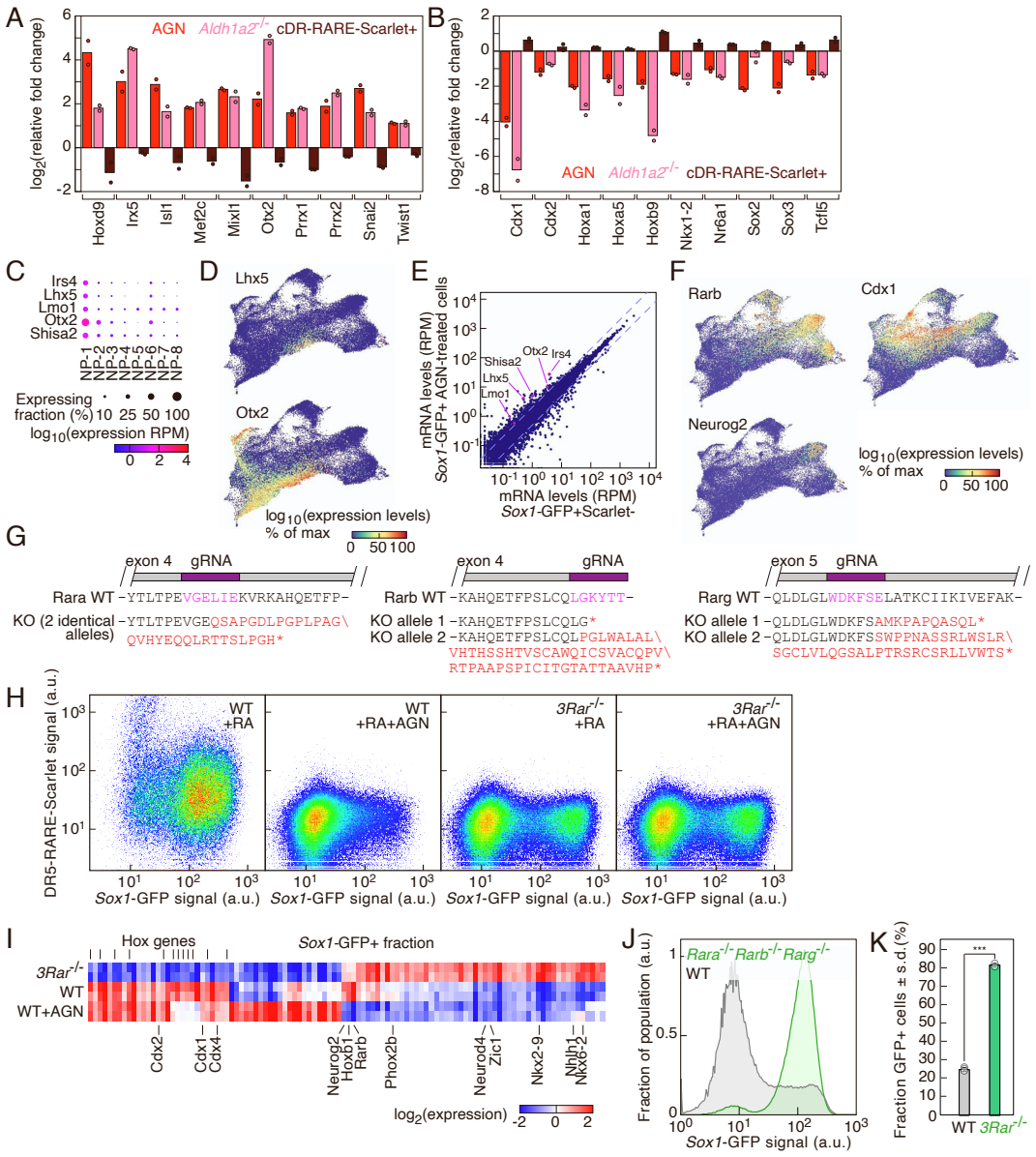

Figure S7. Related to Figure 7. Role of RA signaling and RAR receptors in neural commitment and in establishing neural progenitor diversity.

(A, B) Relative changes of expression of transcriptional factors in the $T^{TagBFP+}$ subpopulation of AGN-treated cells, *Aldh1a2*$^{-/-}$ cells or cDR-RARE-Scarlet+ cells compared to untreated wild type $T^{TagBFP+}$ cells (A: genes up-regulated by RA signaling block, B: gene downregulated by RA signaling block; n=2 independent experiments).

(C) Dot plots of markers upregulated in the NP-1 category compared to the other neural progenitor categories (RPM: reads per million mapped reads).

(D) UMAP colored by the scaled expression of *Lhx5* or *Otx2*.

(E) Differential gene expression between cDR-RARE-Scarlet- $Sox1^{GFP+}$ cells and AGN-treated $Sox1^{GFP+}$ cells. Dashed lines indicate 2-fold expression changes (RPM: reads per million mapped reads).

(F) UMAP colored by the scaled expression of established RA signaling target genes.

(G) Sanger sequencing-validated obtained alleles of *Rara*$^{-/-}$ *Rarb*$^{-/-}$ *Rarg*$^{-/-}$ mESCs. The relative position of the guide RNA used to target the locus is indicated in purple. ⋆: stop codon.

(H) *Sox1*-GFP and DR5-RARE-Scarlet reporter expression after 4-day differentiation of wild type (WT) and *Rara*$^{-/-}$ *Rarb*$^{-/-}$ *Rarg*$^{-/-}$ (*3Rar*$^{-/-}$) mESCs in N2B27 with 100 nM RA (+RA), or with RA and AGN (+RA+AGN: 100 nM RA + 1 μM AGN).

(I) Expression levels of transcriptional factors with differential expression in the $Sox1^{GFP+}$ 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.
(J) *Sox1*-GFP reporter expression after 6 days of differentiation in N2B27 (without vitamin A) of wild type cells (gray) or *Rara*$^{-/-}$ *Rarb*$^{-/-}$ *Rarg*$^{-/-}$ cells (green).
(K) Quantification and statistical analysis of the data in Figure S7J. (n=3 independent experiments; ***, $p < .001$; two-sided unpaired t-test; data represented as mean$\pm$SD).

Supplemental Experimental Procedures

mESC maintenance

The parental mESC line was a *Sox1-Brachyury* double knock-in (2KI) line[1]. mESCs were maintained in "LIF+serum" as described previously[2]. Briefly, cells were cultured at 37°C with 5% CO_2 on dishes (Nunc) coated with 0.1% gelatin (Sigma). The pluripotency maintaing medium was prepared as follows: DMEM (high glucose, no glutamine, with sodium bicarbonate)(Invitrogen) supplemented with 15% ES-qualified EmbryoMax Fetal Calf Serum (Millipore), 10 ng/ml murine LIF (EMBL Protein Expression and Purification Core Facility), 1x Non-Essential Amino Acids, 2 mM L-glutamine, 1 mM sodium pyruvate, 100 U/ml penicillin and 100 μg/ml streptomycin, 0.1 mM 2-mercaptoethanol (all Invitrogen). Medium was changed daily and cells were passaged every other day with 0.05% Trypsin-EDTA or StemPro Accutase (Invitrogen) at a passaging ratio of 1/3 - 1/12.

Generation of knockout mESC lines

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

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 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), were designed and cloned in pX330-U6-Chimeric-BB-CBh-hSpCas9 following Hsu et al.[3]. 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[4]. 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[5]). The plasmid contained a PGK::HygroR-bGHpA cassette to enable selection with hygromycin. Direct repeats (DR) of RAR binding sites spaced by 5 nucleotides (5'-GGTTCACCGAAAGTTCA) reported in Rossant et al.[6] 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.[7]. 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[1].
- *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 cDR-RARE-H2B-Scarlet-bGHpA 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 *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[1], mESCs were seeded at a density of 30-50 cells per mm^2 (unless reported otherwise) 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. For the experiments to asses the crosstalk between RA signaling and TGFβ or Wnt pathway inhibition in absence of IDE1, the following procedure was adopted. Cell seeding condition and the medium used were the same as above with the exception of removing the small molecule IDE1. The formation of the PS-like population was induced in this case with a pulse of the Nodal/TGFβ agonist ACTIVIN A, added at a concentration of 50 ng/ml from day 1 to day 2. After the pulse of 24 hours, ACTIVIN A was removed to allow the generation of neural progenitors.

Differentiation to neural progenitors for transcriptional profiling

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 the *Cyp26a1*$^{-/-}$ cells and RAR-null cells differentiation in N2B27 medium, B27 supplement without vitamin A was used instead.

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 DKK1 (150 ng/ml), CHORDIN (50 ng/ml) and NOGGIN (50 ng/ml), were obtained from Peprotech and added to the medium at the indicated concentrations from the onset of the PS-like differentiation. Recombinant ACTIVIN A (50 ng/ml) and BMP4 (10 ng/ml) were obtained from Peprotech and added from day 3 of the PS-like differentiation procedure at

the indicated concentrations. 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.

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 or Scarlet fluorescence levels using an Aria Fusion sorter (BD BioSciences). Flow cytometry data was analyzed with FlowJo. Gating strategy to quantify the percentage of cells expressing a given reporter or combination of reporters was selected according to non-transgenic mESCs negative controls.

RNA-seq library construction

RNA was extracted from pellets of trypsinized cells using the MirVana kit (Ambion) following the instructions provided by the manufacturer. Barcoded mRNA libraries were prepared using TruSeq RNA Sample Preparation (Illumina) following the manufacturer's instructions. The 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[8] 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[9]. 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.

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.

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[8] 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[10]. Dimensionality reduction was performed using uniform manifold approximation and projection (UMAP)[11]. 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.

Supplemental References

[1] Sladitschek HL, Neveu PA. A gene regulatory network controls the balance between mesendoderm and ectoderm at pluripotency exit. Mol Syst Biol. 2019;15(12):e9043.

[2] Sladitschek HL, Neveu PA. The bimodally expressed microRNA miR-142 gates exit from pluripotency. Mol Syst Biol. 2015;11(12):850.

[3] Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 2013;31(9):827-32.

[4] Sladitschek HL, Neveu PA. MXS-Chaining: A Highly Efficient Cloning Platform for Imaging and Flow Cytometry Approaches in Mammalian Systems. PloS one. 2015;10(4):e0124958.

[5] Bindels DS, Haarbosch L, van Weeren L, Postma M, Wiese KE, Mastop M, et al. mScarlet: a bright monomeric red fluorescent protein for cellular imaging. Nat Methods. 2017;14(1):53-6.

[6] Rossant J, Zirngibl R, Cado D, Shago M, Giguère V. Expression of a retinoic acid response element-hsplacZ transgene defines specific domains of transcriptional activity during mouse embryogenesis. Genes Dev. 1991;5(8):1333-44.

[7] Moutier E, Ye T, Choukrallah MA, Urban S, Osz J, Chatagnon A, et al. Retinoic acid receptors recognize the mouse genome through binding elements with diverse spacing and topology. J Biol Chem. 2012;287(31):26328-41.

[8] Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.

[9] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40.

[10] Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015;33(5):495-502.

[11] Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2018;37:38-44.