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

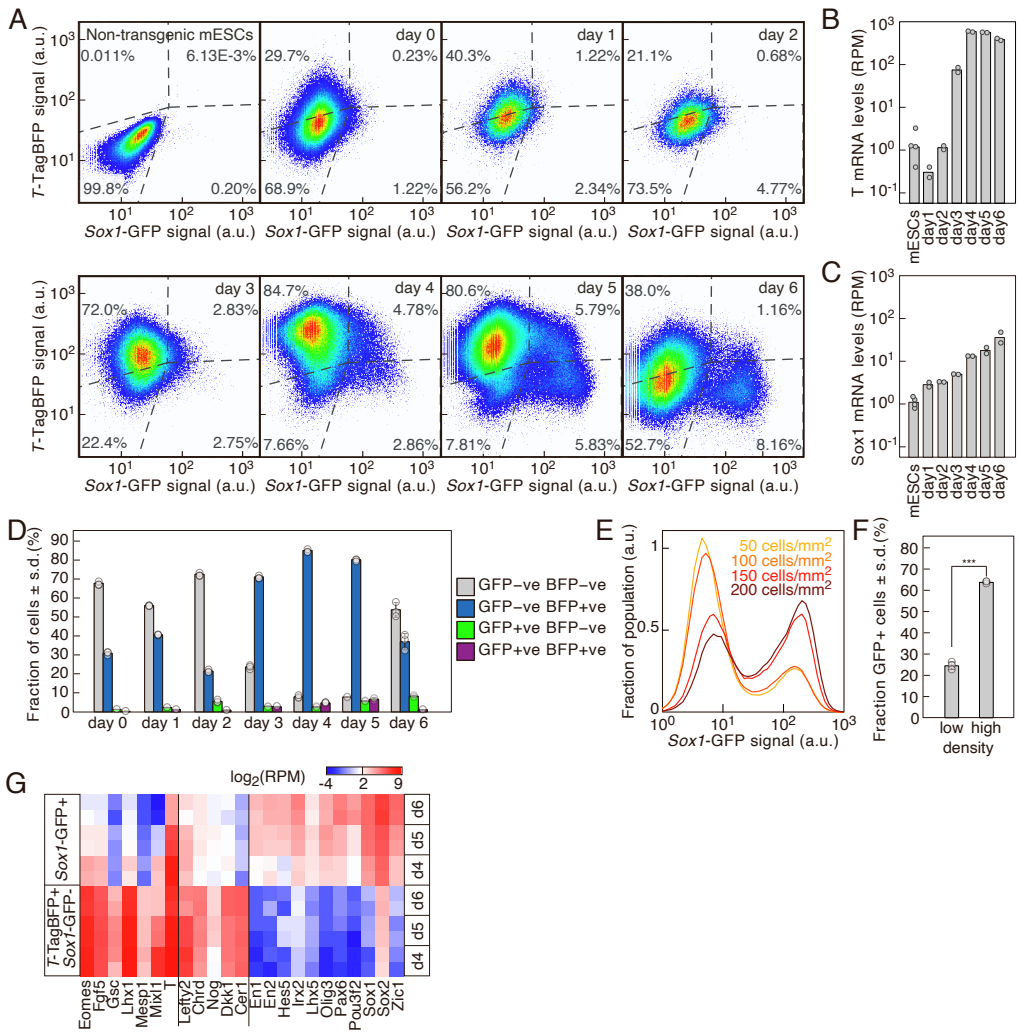

Figure S1. Related to Figure 1. Characterization of neural induction by primitive streak-like cells.
(A) Time course of *T*-TagBFP and *Sox1*-GFP reporter expression as measured by flow cytometry each day of the IDE1 PS-like differentiation. The day 0 condition corresponds to mESCs cultured in the pluripotency 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.

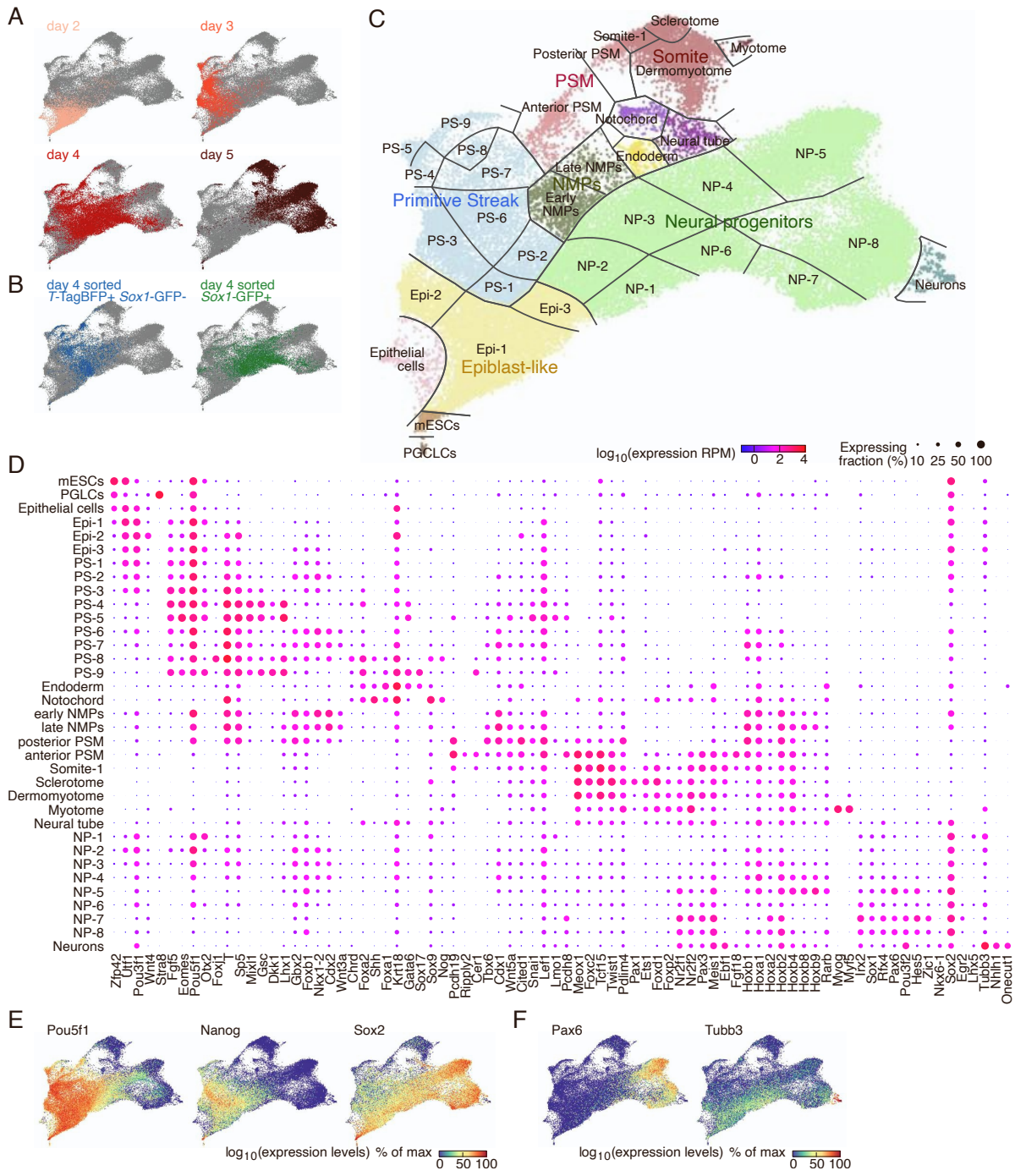

Figure S2. Related to Figure 2. Characterization of neural induction by primitive streak-like cells at the single-cell level.

(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*.

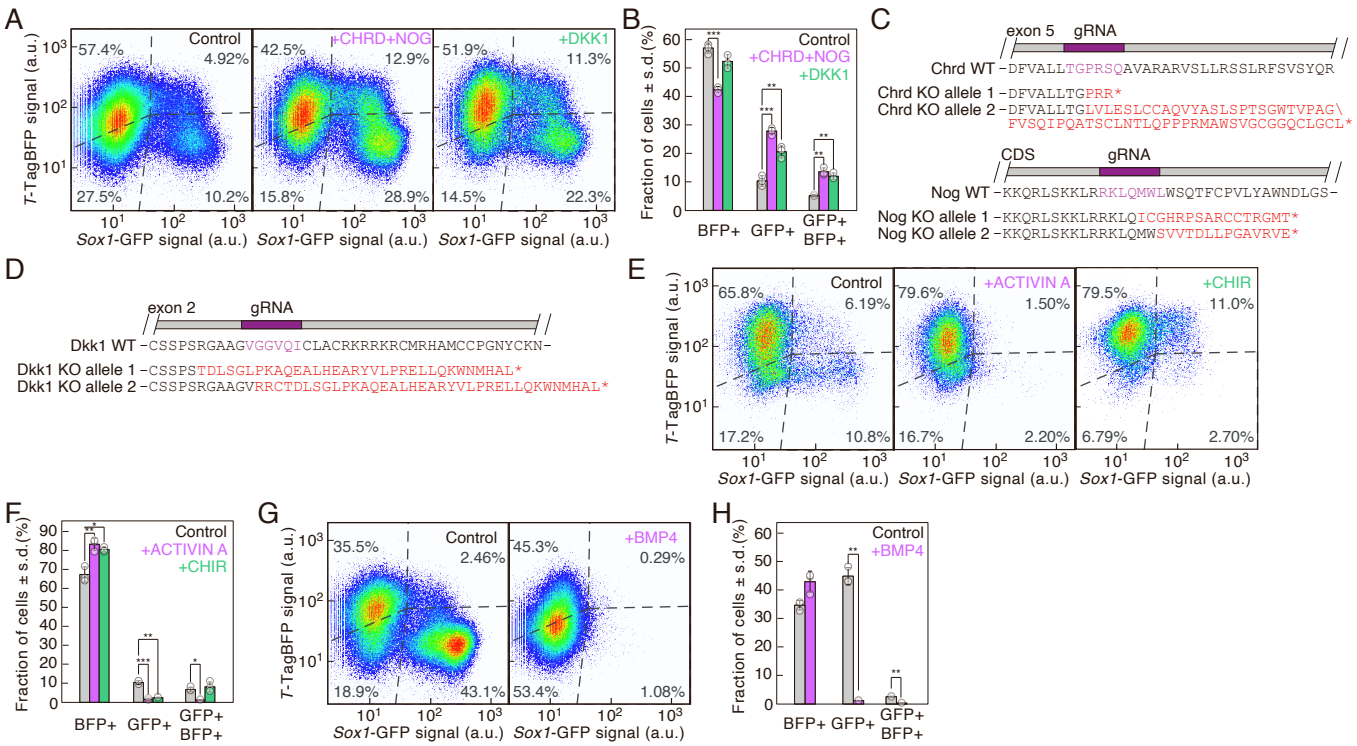

Figure S3. Related to Figure 3. A balance between agonists and inhibitors of the TGFβ and Wnt signaling pathways fine tunes the formation of neuroectodermal and primitive streak derivatives in culture.

(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).

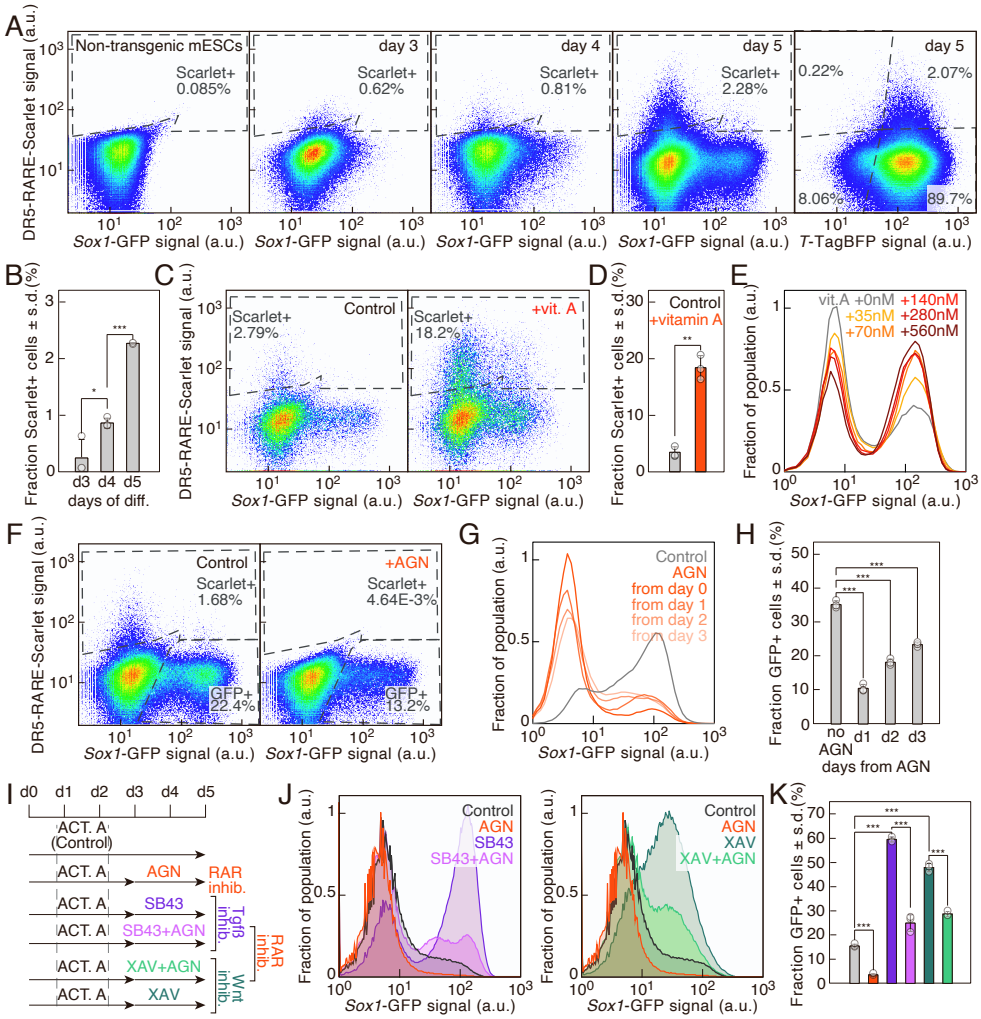

Figure S4. Related to Figure 4. Retinoic acid signaling contributes to the formation of neural progenitors during the by PS-like differentiation.

(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).

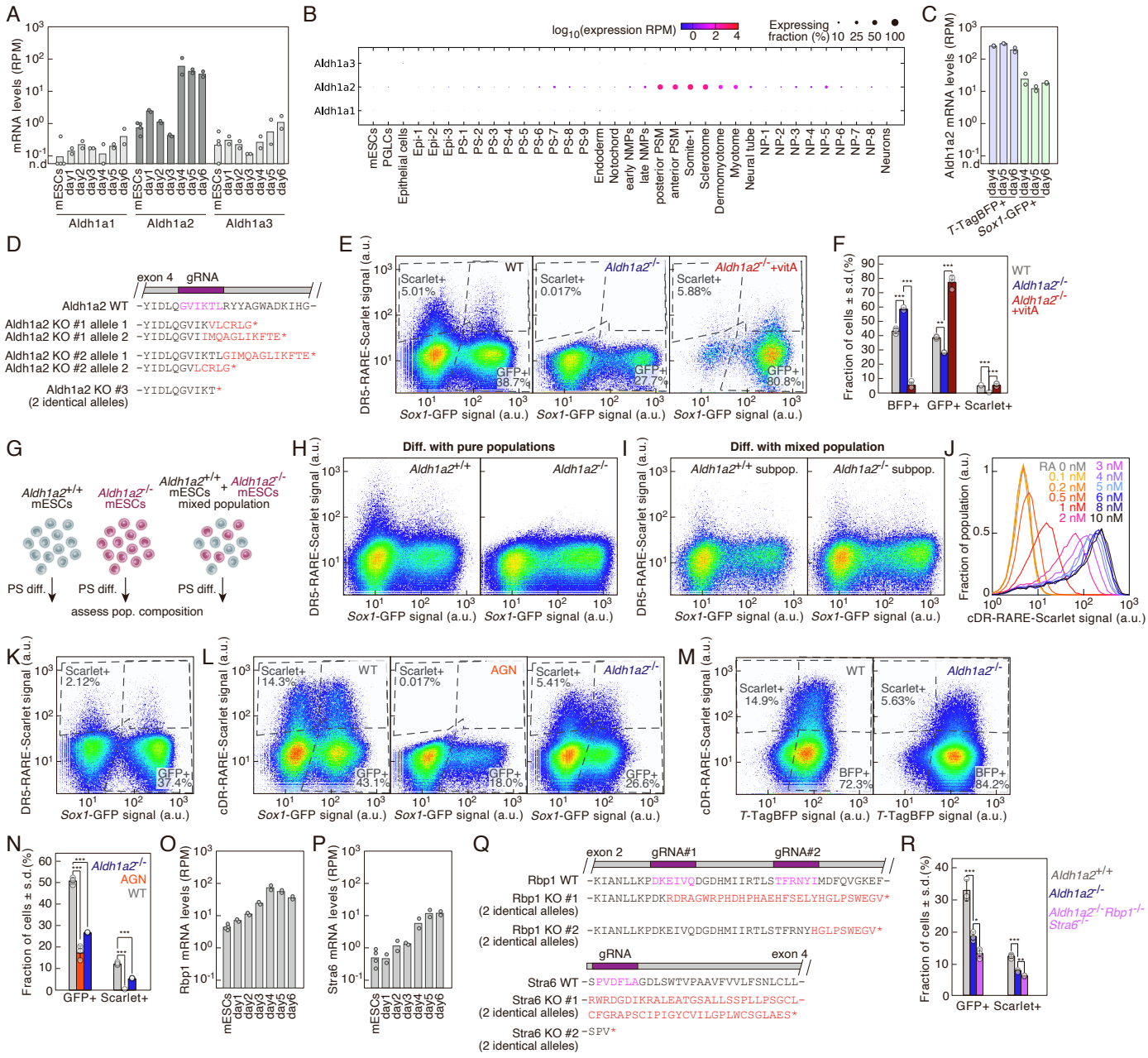

Figure S5. Related to Figure 5. Aldh1a2-independent RA signaling during PS-like differentiation.

(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).

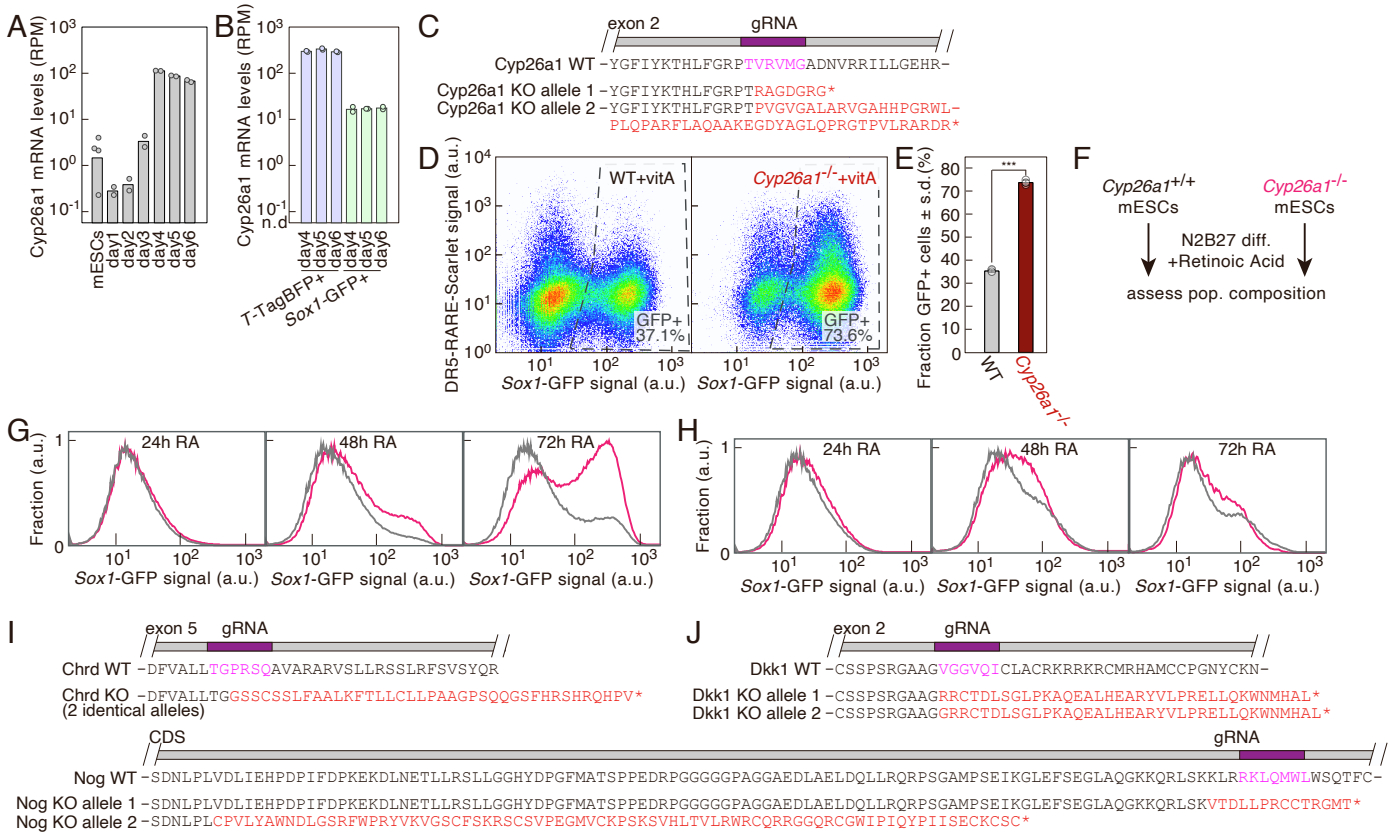

Figure S6. Related to Figure 6. Cyp26a1 is a key factor limiting RA levels and neuroectoderm differentiation during PS-like differentiation.

(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.

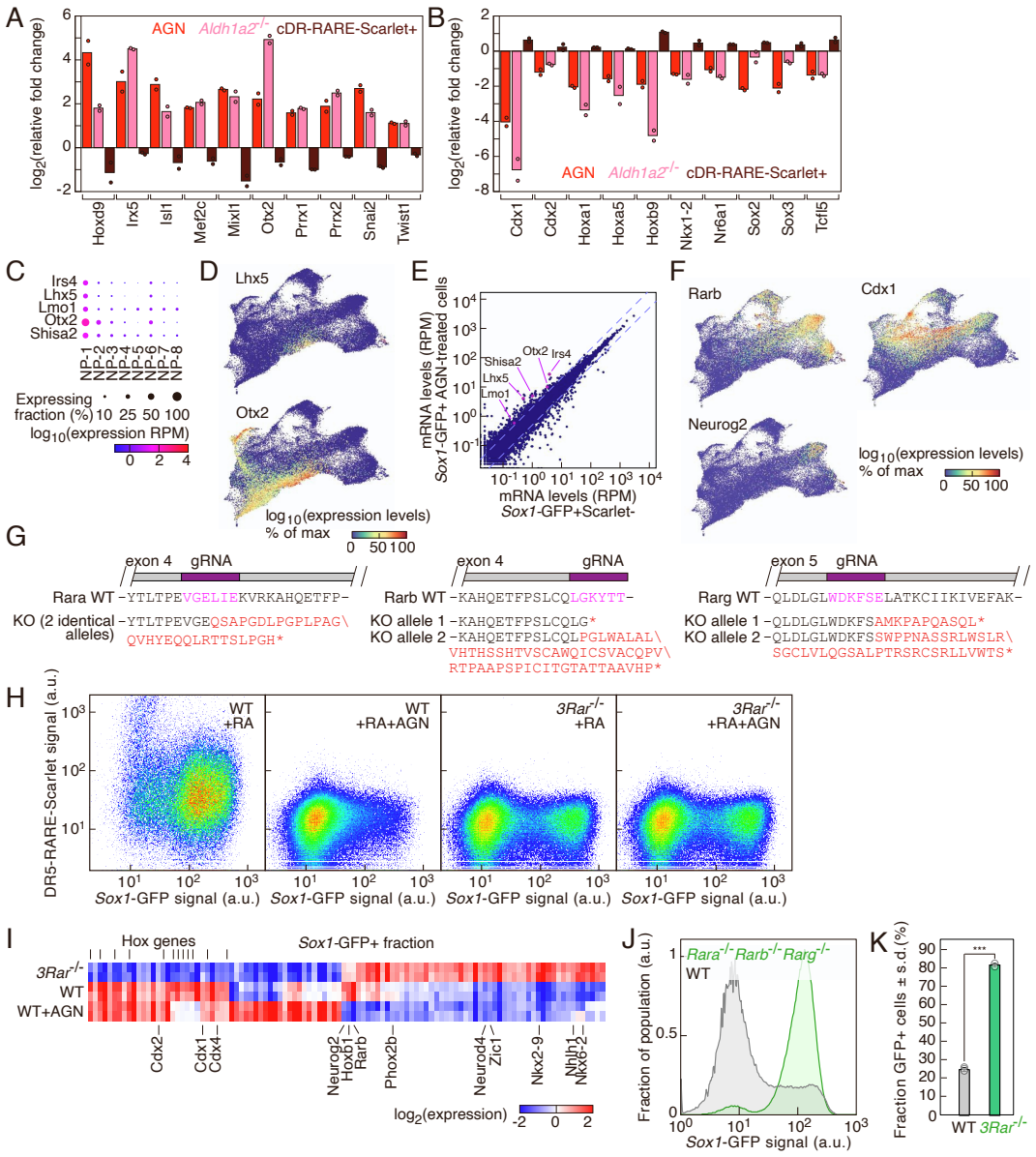

Figure S7. Related to Figure 7. Role of RA signaling and RAR receptors in neural commitment and in establishing neural progenitor diversity.

(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.