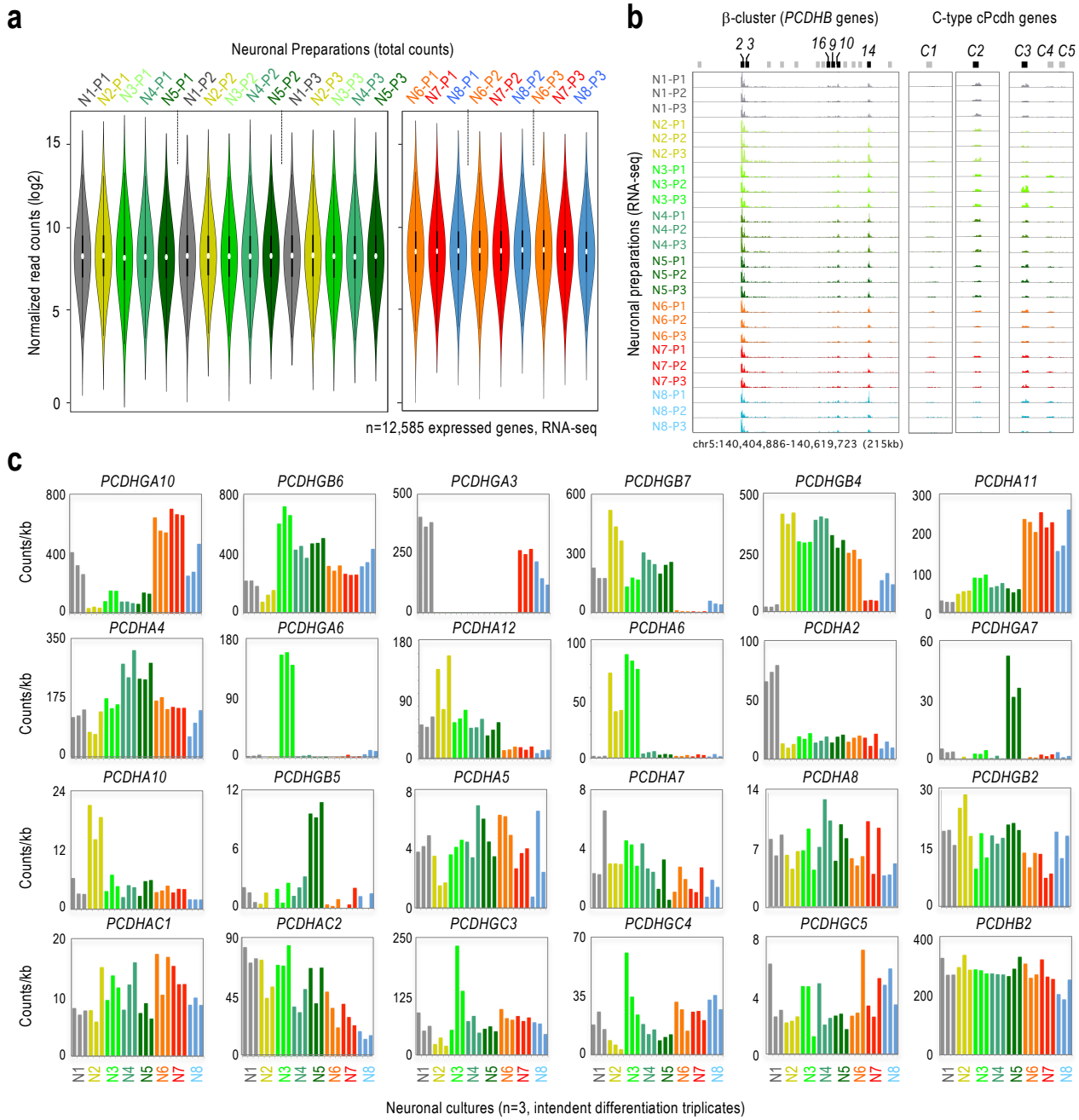
Supplementary information for:

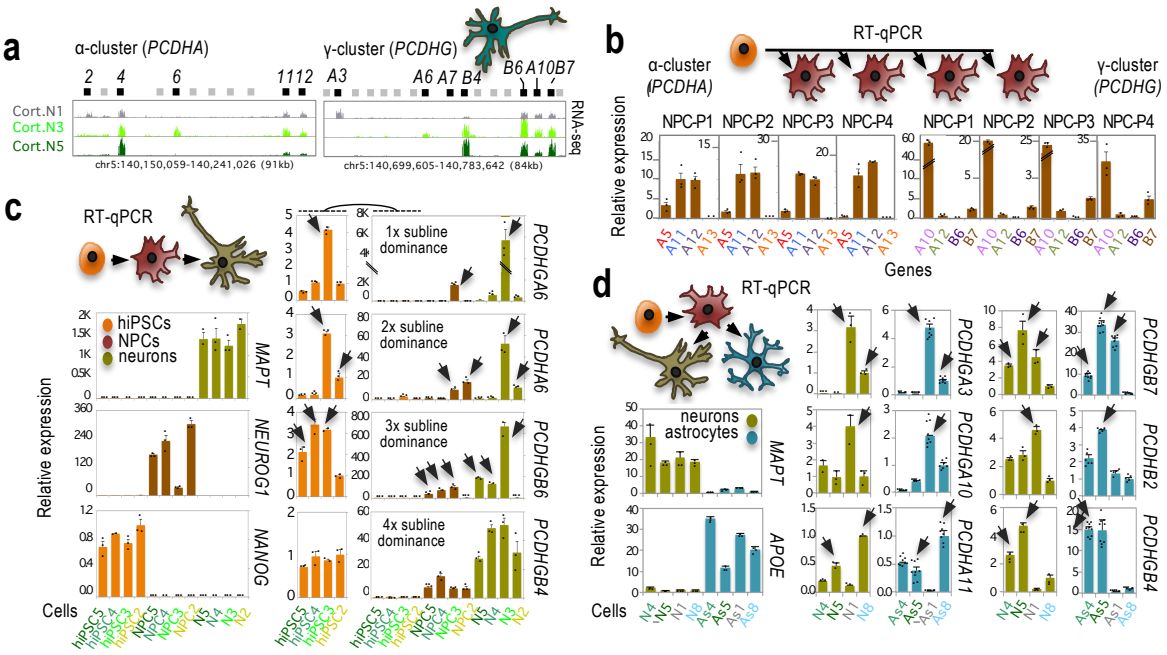
Chromatin establishes an immature version of neuronal protocadherin selection during the naive-to-primed conversion of pluripotent stem cells

September 29th, 2019

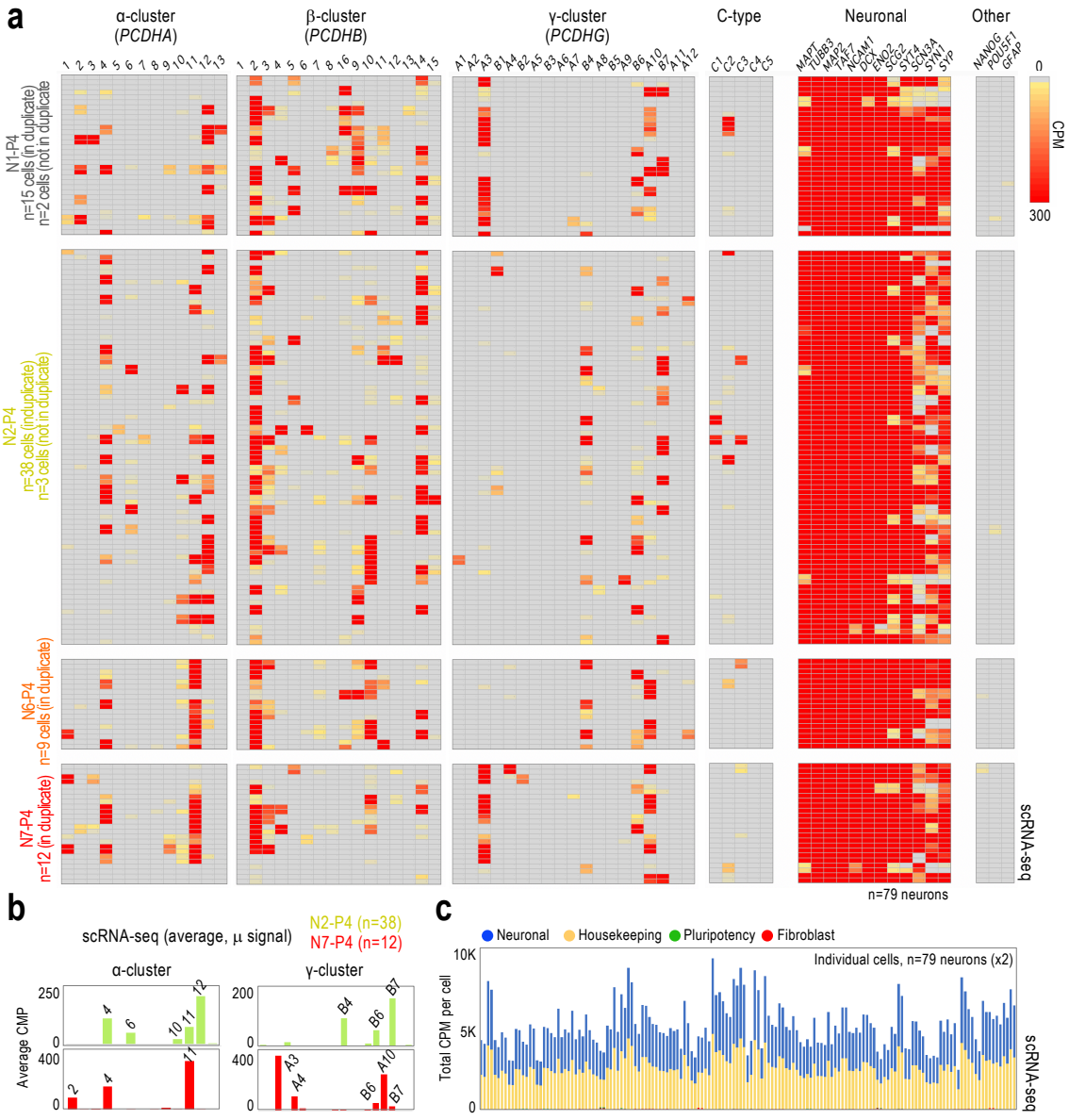
Contents



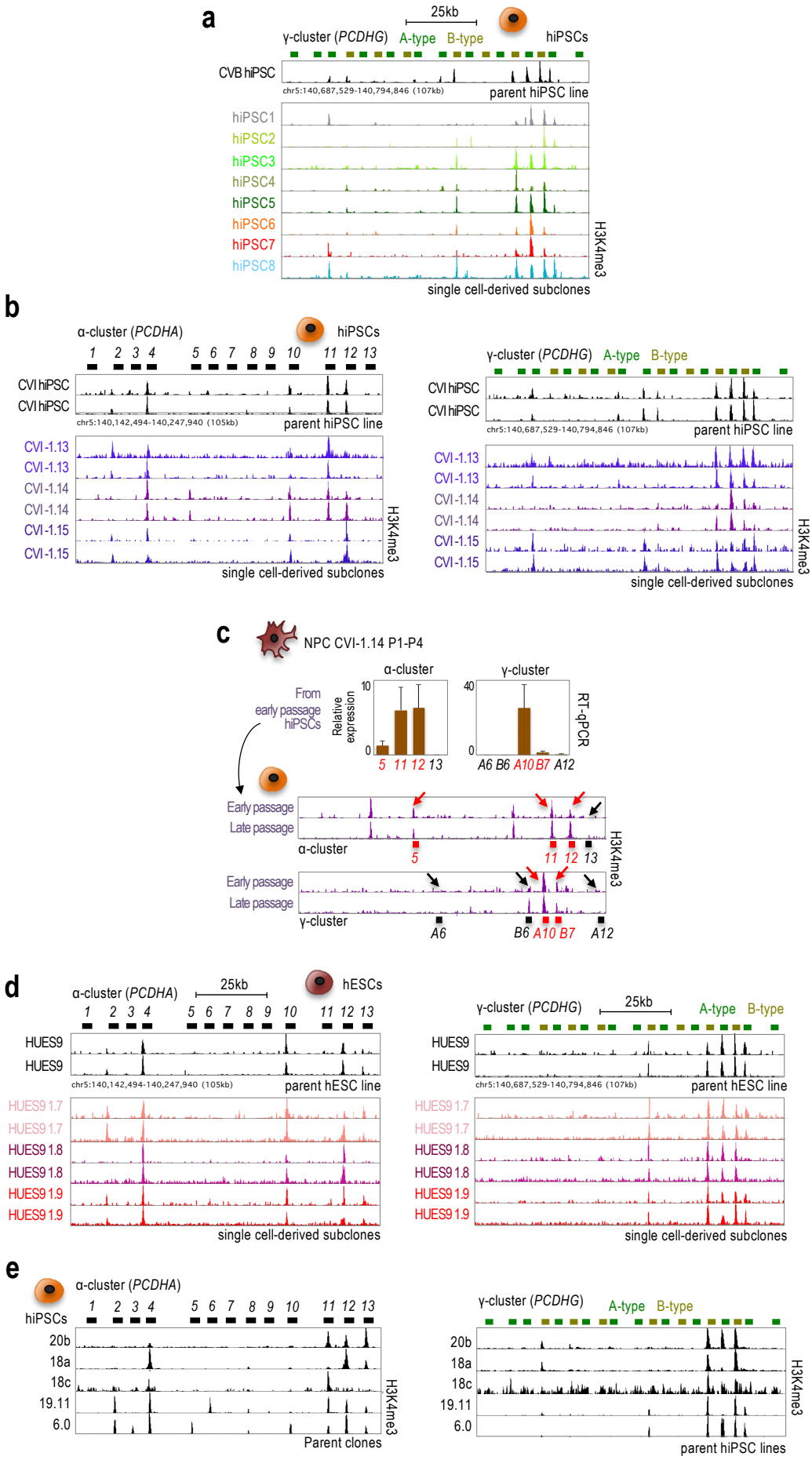
Supplementary Fig. 1. RNA-seq analysis of cPcdh expression in hiPSC-derived neurons. a, Violin plots of normalized read counts (RNA-seq data) for each independent neuronal preparation (n=3, P1-3) from each of the single-cell derived sublines (N1-8, n=8). Plots show median and interquartile range (25th and 75th). Error bars represent 95% confidence intervals. Analysis based on the subset of expressed Refseq genes (n=12,585) in at least n=3 preparations. We note that the three rounds of differentiation did not overlap in time (P1 was followed by P2 and P2 was followed by P3). Only after RNA extraction, the differentiation triplicates were processed simultaneously (i.e. during library preparation and sequencing). **b**, RNA-seq data showing expression of β-cPcdh and C-type cPcdh genes in N1-8 P1-3 (n=3 independent differentiation preparations from the n=8 independent hiPSC/NPC-derived sublines). These tracks share scale with the tracks shown in Fig. 1a for comparison purposes. Genes or 5' exons are indicated above the tracks. Genomic coordinates based on hg18. **c,** Read counts per kilobase (kb) of the most heterogeneously-expressed V-type cPcdh isoforms and others in the n=24 neuronal shown in **a** and **b** (RNA-seq, 5' exonic only data). Samples color-coded consistent with subline identity, as indicated at the bottom. C-type cPcdh genes have been included as reference but, we note, show also some expression heterogeneity among the N1-8 preparations. *PCDHB2* included as reference of relatively homogenous cPcdh expression in our cultures.

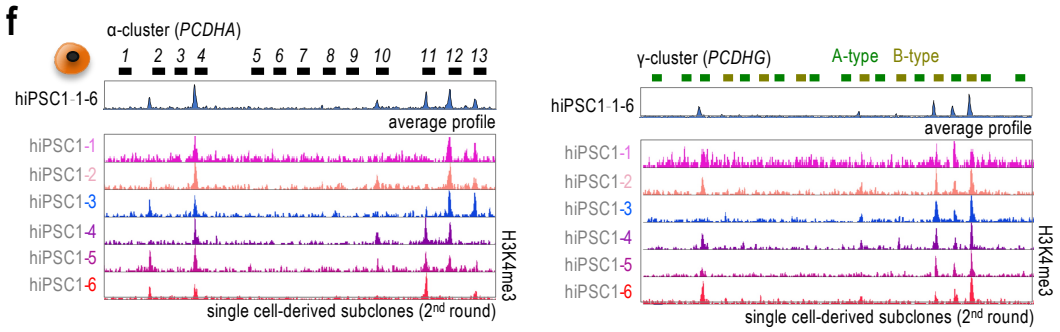


Supplementary Fig. 2. Clonal origin (not protocol of differentiation) dictates cPcdh selections in hiPSC-derived neurons. a, We tested the application of a second protocol of neuronal differentiation that generates cortical-like cells and that does not include a step of FACS-mediated enrichment; and, still, we observe similar cPcdh patterns as those observed in the 'general' neurons shown in Fig. 1a (which were not FACS sorted). Tracks show RNA-seq signal in a single culture of cortical neurons (Cort. N1/3/5) generated in each case an independent single-cell-derived hiPSC subpopulations (n=3, hiPSC1/3/5). These subpopulations were used to generate N1/3/5 in Fig. 1a. Labeling as in Fig. 1a. **b-d**, RT-qPCR analysis of cPcdh expression normalized to *RPLP27* in independent but clonally related populations from the indicated cultures (see below). In **b**, the panels show reproducibility in the expression of n=8 cPcdh genes in NPCs generated from n=4 independent differentiation preparations (NPC P1-P4) derived from the same single-cell-derived hiPSC subline (CVI1.14; this subline will be described in Supplementary Fig. 4b; it is a single-cell-derived subline generated from the CVI hiPSC line, which is a sister clone of the CVB line that never underwent a process of genome editing). In **c**, we examined the patterns of cPcdh expression in clonally related hiPSCs, NPCs, and neurons derived from hiPSC2/3/4/5 (n=4 sublines, thus in total 4+4+4 independent preparations). We examined n=4 cPcdh genes and n=3 markers: *MAPT* (neuronal), *NEUROG1* (NPC), and *NANOG* (pluripotency). The arrows indicate the most highly expressed isoforms in neurons generated from every subline, which were relatively consistent in NPCs and, surprisingly also, in barely-cPcdh-expressing hiPSCs. For instance, *PCDHGA6* was dominantly expressed in hiPSC3/NPC3/N3, while *PCDHA6* was dominantly expressed in hiPSC2/NPC2/N2 and hiPSC3/NPC3/N3. In **d**, we examined the patterns of cPcdh expression in n=4 pairs of cultures of clonally related neurons and astrocytes derived from the same NPC population in each pair (NPC1/4/5/8). We examined n=6 cPcdh genes and n=2 markers: *MAPT* (neuronal) and *APOE* (astrocytic). The arrows indicate the most highly expressed isoforms in every case, which were relatively consistent between clonally related neurons and astrocytes. In all cases, the data is normalized to the lowest value of expression for each gene. Error bars represent standard deviation of the mean of technical replicates (n≥3).

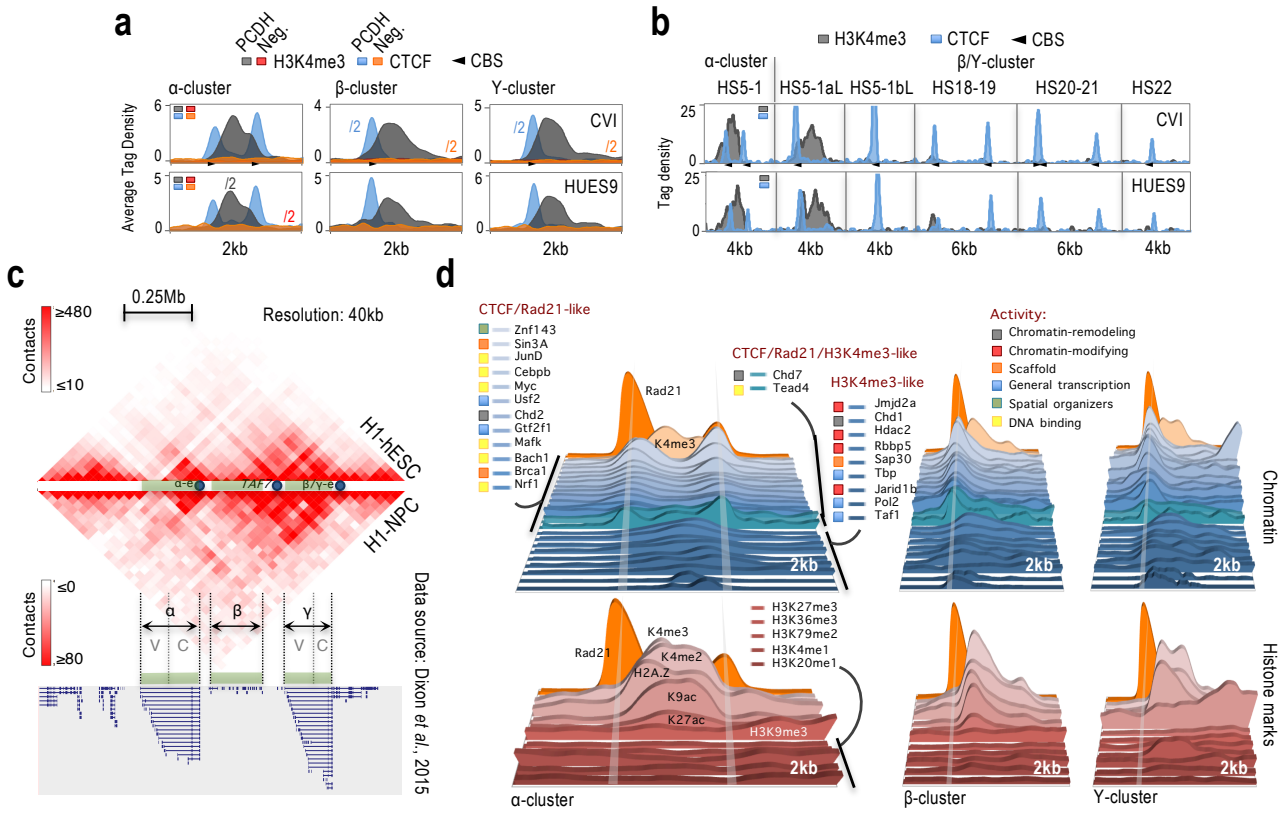


Supplementary Fig. 3. Single-cell analysis of cPcdh expression in hiPSC-derived neurons (N1/2/6/7-P4). a, In order to assess the expression of cPcdh isoforms at a single-cell scale, we performed a fourth round of neuronal differentiation (P4) using hiPSC1/2/6/7. After FACS-mediated separation of individual cells and generation of full-length cDNA, we excluded samples with signs of RNA degradation or low cDNA concentration, and processed the rest for sequencing. This heatmap shows counts per million (CPM) across 5' exonic regions of V-type and C-type cPcdh isoforms in n=79 neurons (>6 million reads/cells). In n=74 cases, we obtained a high number of reads using two different cleaning and pooling strategies of cDNA, while in n=5 cases the number of reads was too low when pooling samples before cleaning (see Methods). On average, each neuron expresses 1.44 α-isoforms and 1.29 γ-isoforms, but we observed cases expressing 0-6 and 0-3 isoforms, respectively. In total, we found 60 unique α/γ combinations (70 adding the *β*-cluster) and 14 repeated α/γ combinations in our analysis of n=74 cells. Typical neuronal genes (e.g. *MAPT* and *MAP2*) could be observed in most neurons (average expression range around six thousand CPMs), while the expression of synaptic genes was more heterogeneous (e.g. *SYT4*, *SCN3A*, *SYN1*, and *SYP*) and at lower levels. Expression of pluripotency markers (*NANOG* and *POU5F1*) and glial markers (*GFAP*) is also shown, but not observed in any neuron. **b**, Histogram representation of average CPM values for α/γ-cPcdh genes segregated by origin of single-cell-derived hiPSC subline (number of aggregated single cells indicated in the figure). These panels complement Fig. 1c. Genes represented in order of their genomic position (as in **a**). We note some value differences between the average CPM values shown here (and in Fig. 1c) and the values in the actual bulk RNA-seq analysis shown in Fig. 1a, which we explain as the average values here are based on 'only' n=9, 12, 15, and 38 cells, whereas the bulk RNA-seq values are based on millions of cells. **c**, Total CPM values per cell (n=79 neurons in duplicate based on two different cleaning and pooling steps, as described above) in exonic regions of n=72 neuronal, n=32 housekeeping, n=15 pluripotency, and n=13 fibroblast genes. This analysis shows relatively homogenous neuronal and housekeeping signal across all samples (virtually no pluripotency and fibroblast signal), which is a sign of similar coverage and neuronal identity in all of them.

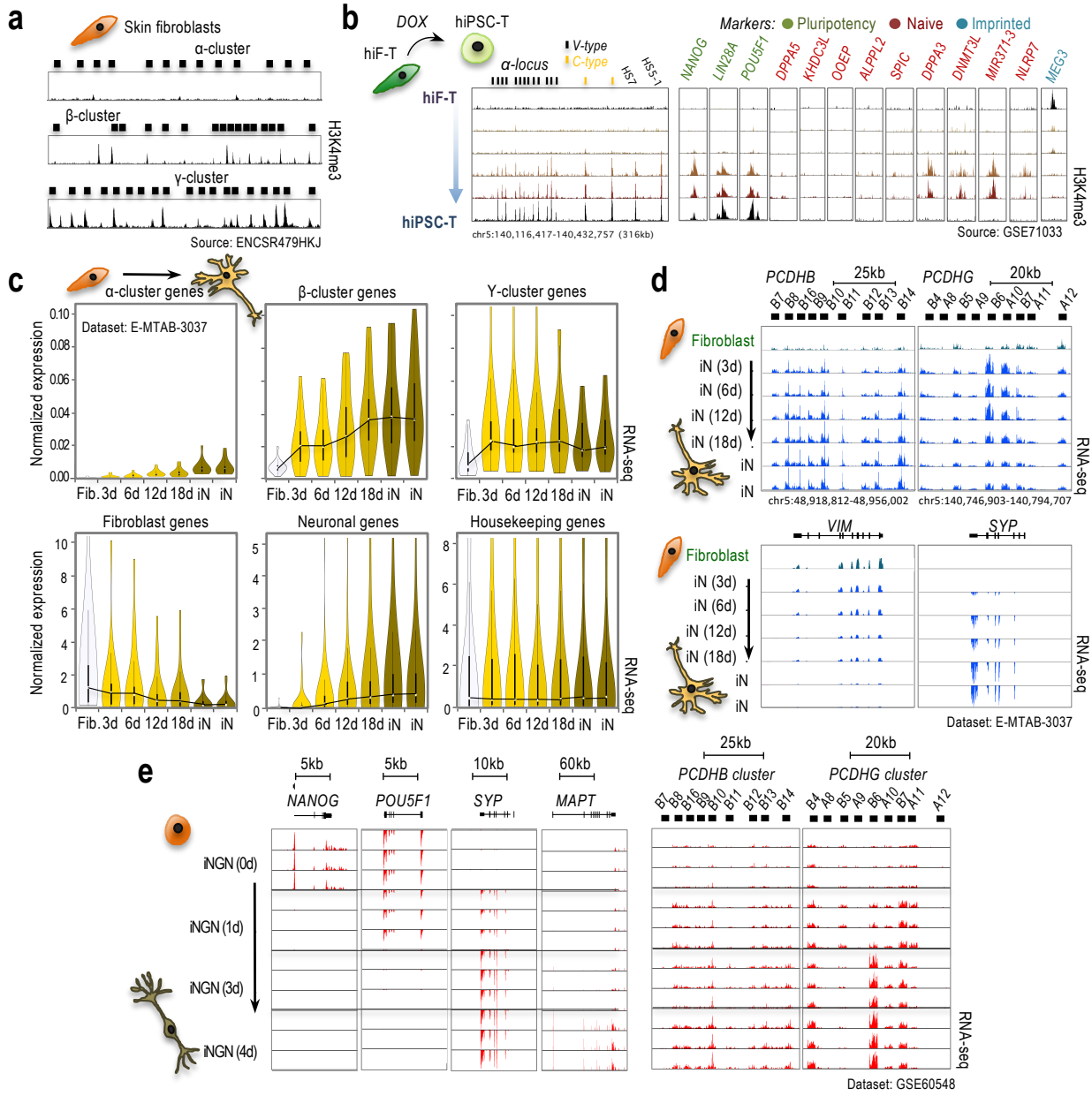

a
25kb
γ-cluster (PCDHG) A-type B-type
hiPSCs
CVB hiPSC
chr5:140,687,529-140,794,846 (107kb)
parent hiPSC line
hiPSC1
hiPSC2
hiPSC3
hiPSC4
hiPSC5
hiPSC6
hiPSC7
hiPSC8
H3K4me3
single cell-derived subclones
b
α-cluster (PCDHA)
hiPSCs
1 2 3 4 5 6 7 8 9 10 11 12 13
CVI hiPSC
CVI hiPSC
chr5:140,142,494-140,247,940 (105kb)
parent hiPSC line
CVI-1.13
CVI-1.13
CVI-1.14
CVI-1.14
CVI-1.15
CVI-1.15
H3K4me3
single cell-derived subclones
γ-cluster (PCDHG) A-type B-type
CVI hiPSC
CVI hiPSC
chr5:140,687,529-140,794,846 (107kb)
parent hiPSC line
CVI-1.13
CVI-1.13
CVI-1.14
CVI-1.14
CVI-1.15
CVI-1.15
H3K4me3
single cell-derived subclones
c
NPC CVI-1.14 P1-P4
α-cluster
γ-cluster
From early passage hiPSCs
Relative expression
10
0
40
0
RT-qPCR
5 11 12 13
A6 B6 A10 B7 A12
Early passage
Late passage
α-cluster
5
11 12 13
H3K4me3
Early passage
Late passage
γ-cluster
A6
B6 A10 B7
A12
d
α-cluster (PCDHA)
25kb
hESCs
1 2 3 4 5 6 7 8 9 10 11 12 13
HUES9
HUES9
chr5:140,142,494-140,247,940 (105kb)
parent hESC line
HUES9 1.7
HUES9 1.7
HUES9 1.8
HUES9 1.8
HUES9 1.9
HUES9 1.9
single cell-derived subclones
γ-cluster (PCDHG)
25kb
A-type B-type
HUES9
HUES9
chr5:140,687,529-140,794,846 (107kb)
parent hESC line
HUES9 1.7
HUES9 1.7
HUES9 1.8
HUES9 1.8
HUES9 1.9
HUES9 1.9
single cell-derived subclones
e
α-cluster (PCDHA)
hiPSCs
1 2 3 4 5 6 7 8 9 10 11 12 13
20b
18a
18c
19.11
6.0
H3K4me3
Parent clones
γ-cluster (PCDHG) A-type B-type
20b
18a
18c
19.11
6.0
H3K4me3
parent hiPSC lines



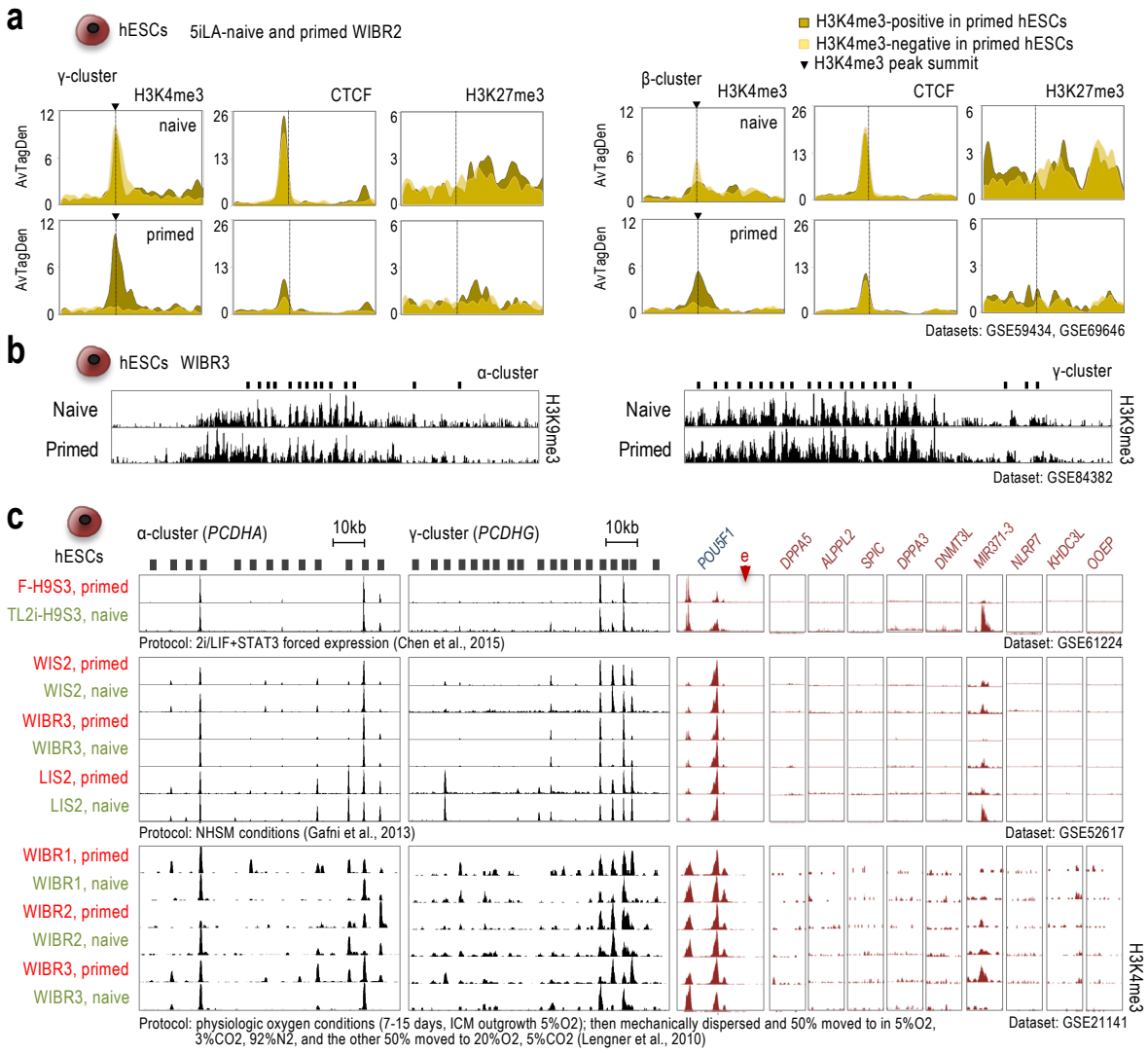
Supplementary Fig. 4. H3K4me3 accumulation across the c-Pcdh locus of single-cell-derived hiPSCs. a,b,d-f, ChIP-seq tracks of H3K4me3 read density along the γ-cluster in the parent CVB line and in n=8 single-cell-derived hiPSC1-8 sublines in **a**, in the parent CVI line and n=3 different single-cell-derived CVI sublines in duplicate (n=2, independent cultures) in **b,** in the HUES9 hiPSC line and n=3 different single-cell-derived HUES9 sublines in n=2 independent cultures in each case in **d**, in n=4 parent hiPSCs lines generated by other laboratories in **e** (Broad Institute GSM772844 for 20b, GSM773029 for 18a, and GSM537671 for 18c; and Ren lab GSM706074 for 19.11 and GSM706075 for 6.0; -we note that two hiPSC lines in **e**, 18a and 18c, derive from the same donor, 18), and, in n=6 single-cell-derived hiPSC sublines generated from a second round of single cell isolation from the already single-cell-derived hiPSC1 subline shown in **f** (in this case, we also show the average tag density calculated from aggregating individual 1-6 tracks in approximately 100bp bins, which ”simulates” a track of bulk signal that combines the 6 tracks, this track is shown on top). **c,** Matching of cPcdh expression in CVI-1.14 NPCs (top panel, average from n=4 independent differentiation rounds, P1-4, and n=3 technical replicate in each case, which are shown individually in Supplementary Fig. 2b) and H3K4me3 accumulation on cPcdh promoters in CVI-1.14 hiPSCs (tracks, n=2 independent cultures, shown also in **b**). We note that cPcdh expression was examined in NPCs derived from an early passage of the CVI 1.14 subline (top track), while H3K4me3 accumulation was examined in an early and in a later passage of this subline, both (bottom track). For clarity purposes, the tracks shown in **c** were duplicated from **b** adding arrows to the promoters whose genes were examined by RT-qPCR. Overall, we concluded that H3K4me3 promoter accumulation in hiPSCs mirrors -to a certain degree- expression in clonally derived neurons, as observed in Fig. 2a,b.



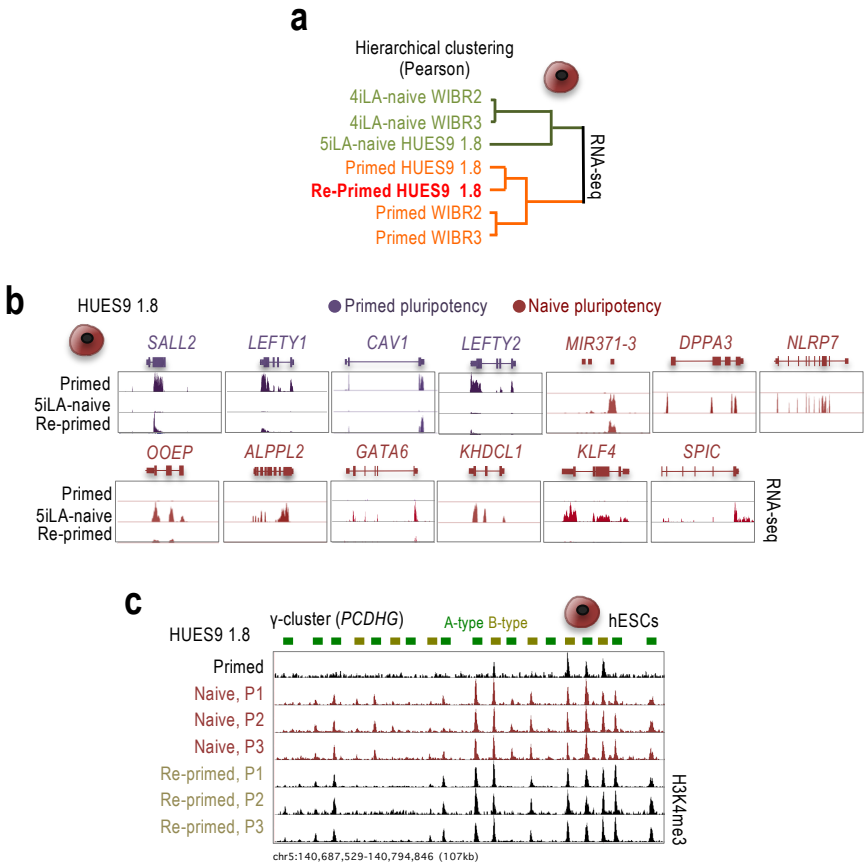
Supplementary Fig. 5. Chromatin landscape in the c-Pcdh locus of hESCs. a, 2k-, 4kb, or 6kb-wide distribution meta-profiles of H3K4me3 and CTCF ChIP-seq signal (averaged tag density, AvTagDen) along α-promoters (**a**-left panels), β-promoters (**a**-middle panels), and γ-promoters (**a**-right panels), or distal regulatory sites (in **c**, and the specific genomic elements are indicated above each panel) in CVI hiPSCs and HUES9 hESCs. Random genomic coordinates were used as negative references (Neg.). Some scale adjustments were made for comparative purposes (indicated as /2, which denotes that the values of the profile were divided by 2). The small triangles indicate CTCF binding sites and motif orientation. Window sizes are indicated at the bottom. **c**, Contact matrices at a 40kb resolution based on available HiC-seq data of H1 hESCs (top matrix) and H1-derived NPCs (bottom matrix) visualized using the 3D Genome Browser (biorxiv.org/content/early/2017/02/27/112268, Biorxiv, 2017). This analysis reveals remarkable similarities in the 3D-spatial configuration of the cPcdh locus between hESCs and NPCs. Gene annotation is shown at the bottom, as well as the region of variable isoforms (V) and constant isoforms (C). Distal sites (e) and the position of the *TAF7* gene are indicated in the matrices (blue circles). **d**, 2kb-wide distribution meta-profiles of publicly available ChIP-seq data (averaged tag density) of chromatin factors (top profiles) and histone marks (bottom profiles) on V-type promoters in the α/β/γ-clusters. We found at least n=23 datasets showing robust peaks in the cPcdh locus in hESCs, including cases of repressive chromatin. Rad21 peaks can be used as reference. Each chromatin component is listed based on similarity to CTCF/Rad21 profiles or similarity to H3K4me3 profiles; in two cases (Chd7 and Tead4), the chromatin components show a mix of the CTCF/Rad21 and H3K4me3 profiles. Color-coded the major activity of each chromatin component ("Activity"). The chromatin profiles are organized following the same order of the names of chromatin components listed in the figure.



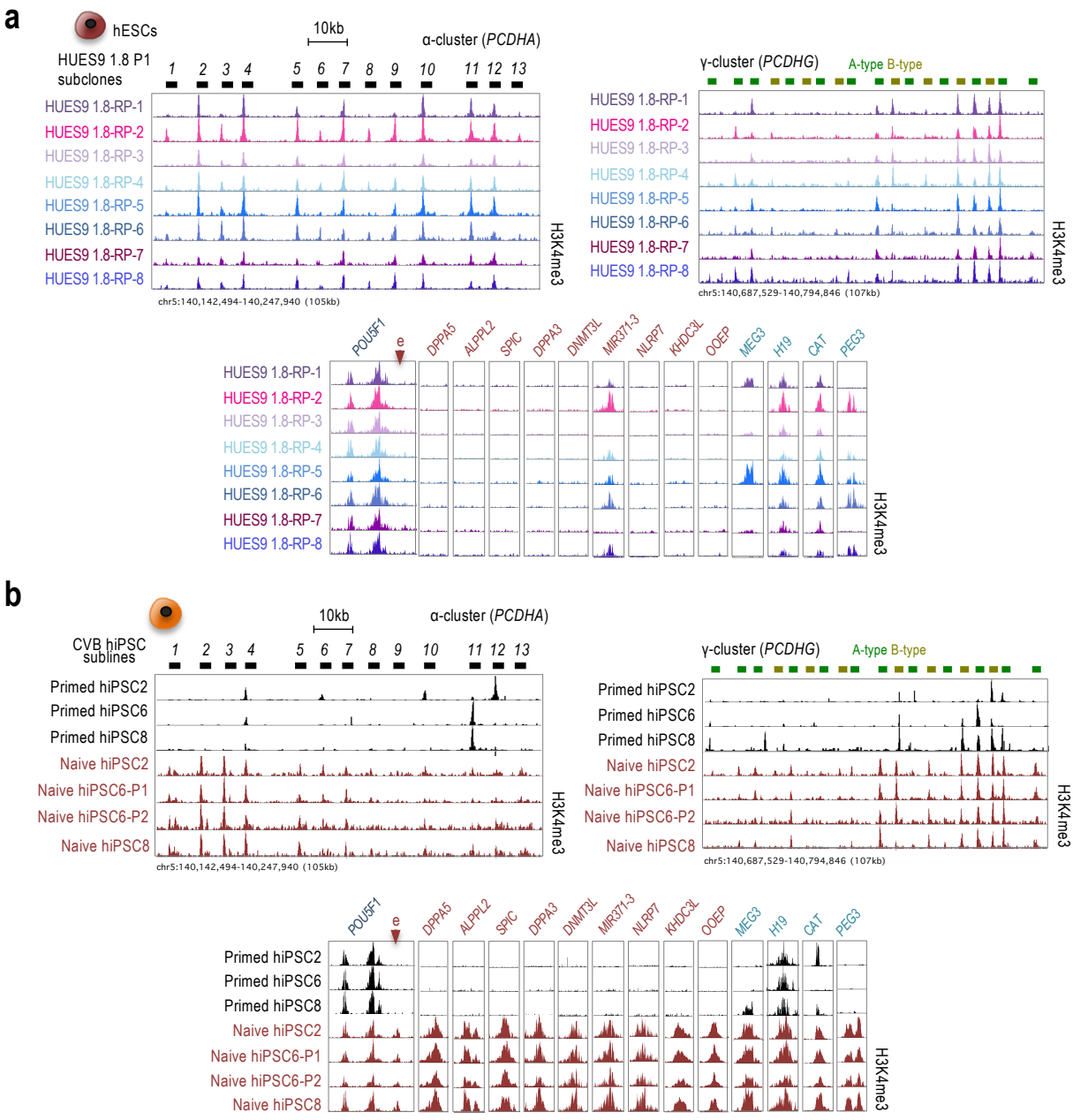
Supplementary Fig. 6. Pcdh signatures in human iNs. a, ChIP-seq signal of H3K4me3 accumulation along the cPcdh locus in arm fibroblasts. Dataset source: ENCSR479HKJ. We note that a distinct characteristic of skin fibroblasts is that there is not H3K4me3 accumulation along the α-cluster, as also observed in immortalized fibroblasts in **b**. **b**, ChIP-seq signal of H3K4me3 accumulation during the process of cell reprogramming of hTERT-immortalized human fibroblasts (hiF-T) using the DOX-inducible reprogramming factors under control of the reverse tetracycline transactivator (rtTA). Dataset source: GSE71033. Sample list (from top to bottom): hiF-T, 5dd_DOX_plus, 10dd_DOX_plus_SSEA3_pos, 24dd_TRA_pos_DOX_plus, and 24dd_TRA_pos_DOX_minus (see GSE71033 for detailed information about this labeling): the 'DOX-minus condition' represents cells reprogrammed for 20 days in DOX followed by 4 days without DOX). The figure also includes markers of the pluripotent state (*POU5F1, LIN28A,* and *NANOG* in green), the naive state (*DPPA5, ALPPL2, SPIC, DPPA3, DNMT3L, MIR371-3, NLRP7, KHDC3L*, and *OOEP* in red), and an imprinted gene (*MEG3* in turquoise). 5' exons of V-type and C-type cPcdh genes are indicated, as well as two distal sites (HS7 and HS5-1). **c,** Complementing the second column of radar plots shown in Fig. 3a, here we show violin plots of averaged cPcdh gene expression separated by cluster using these same samples (RNA-seq data combined for all V-type isoforms in the cluster and a manually curated list of fibroblast, neuronal, and housekeeping genes) at different time points of the fibroblast-iN conversion (n=1 fibroblasts and n=1 transitioning iNs for each time point, 3, 6, 12, and 18 days) and two independent cultures of iNs (n=2). The black lines connect medians in some panels. Dataset source: E-MTAB-3037. Violin plots include medians, interquartile range (25th and 75th). Error bars represent 95% confidence intervals. Dataset source: E-MTAB-3037. **d**, RNA-seq read density tracks showing expression of cPcdh genes during the fibroblast-iN conversion and in iNs, which complements **c**. Dataset source: E-MTAB-3037. Included also tracks for a fibroblast marker (*VIM*) and for a neuronal marker (*SYP*), used as reference. **e**, Complementing Fig. 3b, we show here RNA-seq read density tracks depicting the expression of cPcdh genes during the direct conversion of hiPSCs into neurons (iNGN) after 0, 1, 3, and 4 days of conversion (n=3 independent replicates, each). Also included some markers: pluripotent genes (*NANOG* and *POU5F1*) and neuronal genes (*SYP* and MAPT). Dataset source: GSE60548.



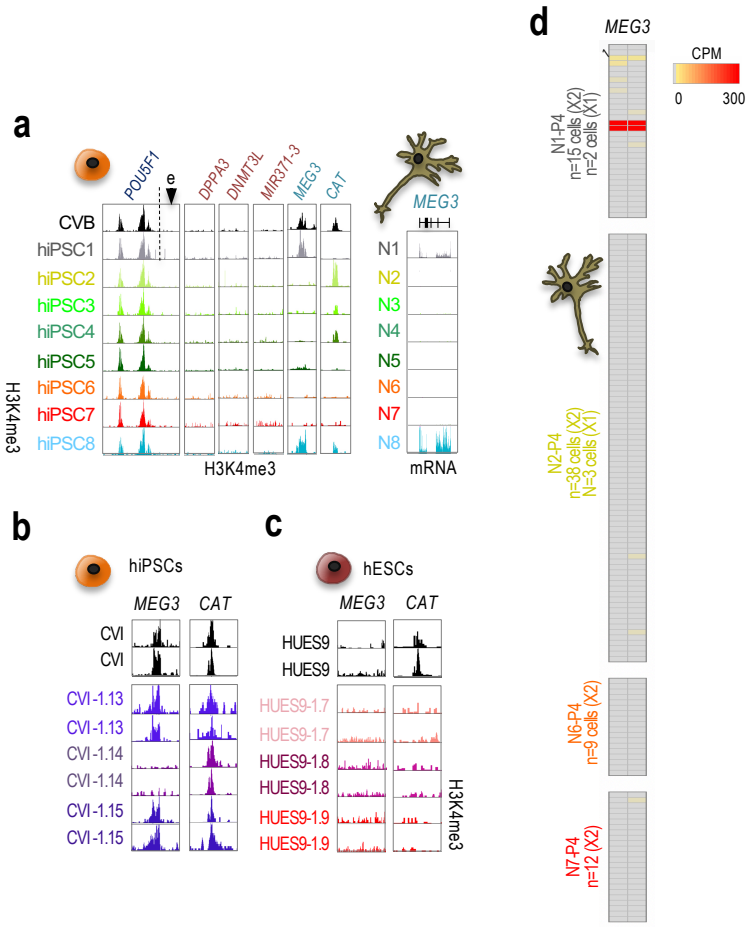
Supplementary Fig. 7. Chromatin organization of the cPcdh locus in previously generated naive hESCs. a, 9-kb distribution meta-profiles of H3K4me3, CTCF, and H3K27me3 ChIP-seq read densities on β- and γ-promoters in WIBR2 hESCs in 6iLA-naive and primed conditions. Promoters segregated based on H3K4me3-positive or -negative status in the primed state. Average Tag Density is represented. Dataset sources: GSE59434 and GSE69646. **b**, ChIP-seq tracks showing H3K9me3 accumulation across the α- and γ-clusters in a culture of primed WIBR2 hESCs and 4iLA-naïve-induced cells from the same line. 5' cPcdh exons indicated above the tracks. Dataset source: GSE84382. **c**, ChIP-seq read density tracks showing H3K4me3 accumulation in the listed hESC lines, which were cultured under conventional (primed-inducing) conditions or under three different naive-inducing conditions (represented by the three groups of panels: on top, n=1 line for the TL2i protocol; in the middle, n=3 lines for the NHSM protocol; and at the bottom, n=3 for the third protocol). Dataset sources: GSE61224, GSE52617, and GSE21141, respectively. A brief description of the naive-inducing conditions is provided underneath of each panel. Left panels show H3K4me3 accumulation across the α- and γ-clusters (the 5' exon of each variable cPcdh gene is represented above the panel). We note that the primed and naive cells shown in the bottom panels were separated soon after derivation from human embryos, while the naive cells in the top and middle panels were derived from primed cells (see information in the listed references). The right panels show H3K4me3 accumulation on genomic regions that serve as markers of the naive state (5iLA protocol). Only one of the expected markers (the *MIR371-3* promoter) slightly accumulates H3K4me3 under the three naive-inducing conditions, whereas -as shown in Fig. 4f- the 5iLA protocol induces H3K4me3 accumulation on all of these markers. These markers include the locus of the pluripotency gene *POU5F1,* in which it is highlighted the position of an enhancer (e) that is active in naive conditions; and the promoters of the following genes: *DPPA5, ALPPL2, SPIC, DPPA3, DNMT3L, MIR371-3, NLRP7, KHDC3L*, and *OOEP*.



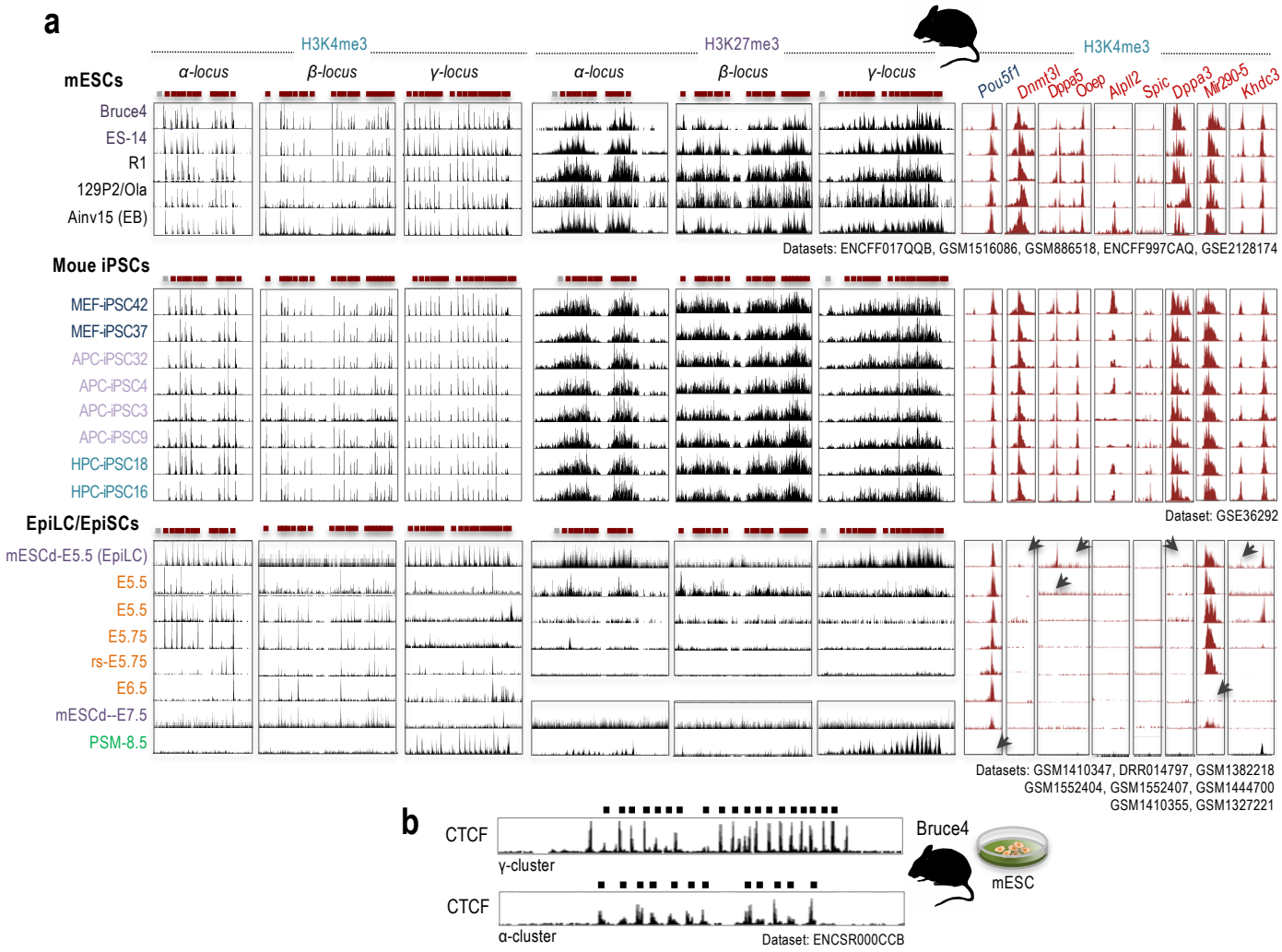
Supplementary Fig. 8. Conversion of the (primed) HUES9 1.8 hESC subline into naive cells (5iLA protocol) and returning to the primed (or re-primed) state afterwards. a, Hierarchical clustering of primed, naive, and re-primed HUES9 1.8 hESCs with primed (source: GSE85689) and naive (source: GSE83765) WIBR2 and WIBR3 hESCs (cultured under 4iLA conditions which correspond to 5iLA conditions without IM12). Analysis based on RNA-seq (n=1 for each condition). **b,** RNA-seq read density tracks of primed and naive markers in primed (n=1 culture), 5iLA-naive (n=1 preparation), and re-primed (n=1 preparation) HUES9 1.8 hESCs, as indicated. This figure complements Fig. 4e. **c**, ChIP-seq read density analysis showing H3K4me3 accumulation across the γ-cluster in n=1 culture of the primed HUES9 1.8 hESC subline, in n=3 independent 5iLA-naive preparations (P1-P3) of the same subline, and in n=3 independent re-primed preparations (P1-P3, which derive from naive P1-P3, respectively) of the HUES9 1.8 hESC subline. This figure complements Fig. 4f. 5' cPcdh exons indicated above each panel.



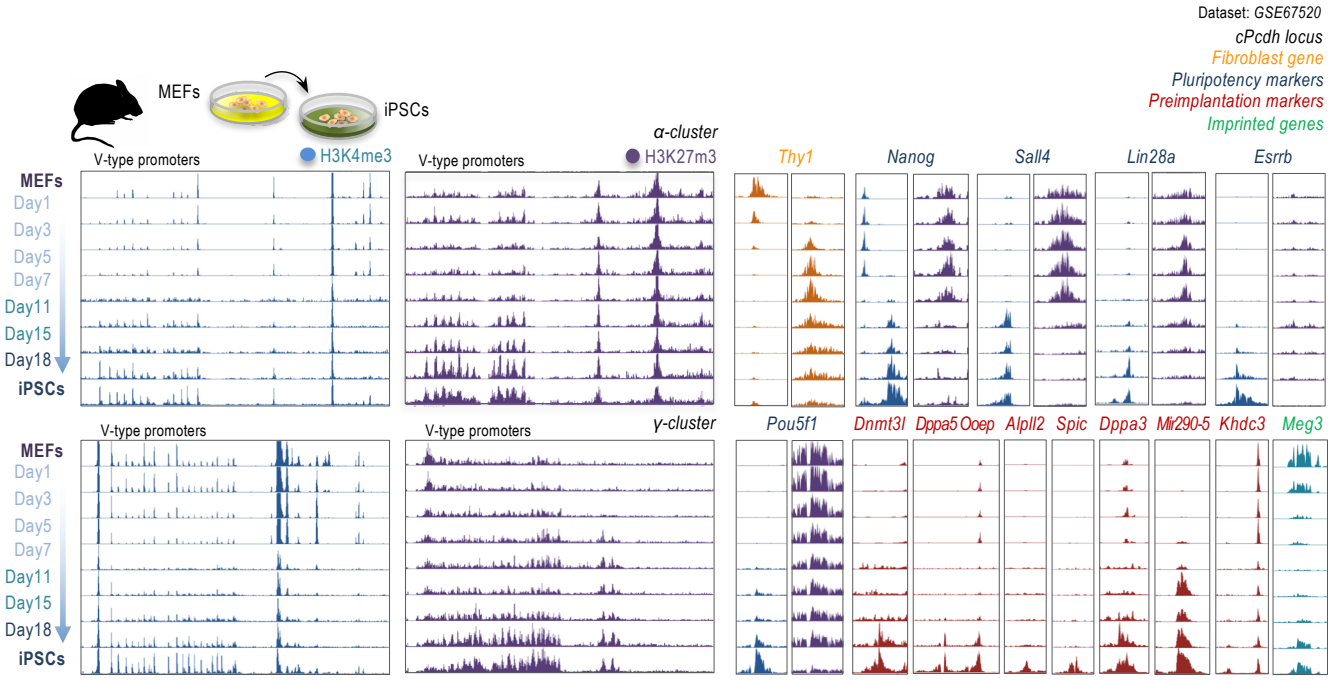
Supplementary Fig. 9. H3K4me3 accumulation across the cPcdh locus in single-cell-derived re-primed hESCs and in naive single-cell-derived hiPSCs (5iLA protocol). a, ChIP-seq read density tracks showing H3K4me3 accumulation across the cPcdh locus in n=8 independent single-cell-derived subpopulations of re-primed HUES9 1.8 cells (RP1-8) from P1 of the 5iLA-naive conversion shown in Fig. 4f. It is also shown H3K4me3 accumulation across the locus of the pluripotency gene *POU5F1* and it is highlighted the position of an enhancer (e) that is active in preimplantation-like conditions (naive-like cells). Also included the promoters of other preimplantation markers: *DPPA5, ALPPL2, SPIC, DPPA3, DNMT3L, MIR371-3, NLRP7, KHDC3L*, and *OOEP*; and the promoters of four imprinted genes: *MEG3, H19, CAT*, and *PEG3*. **b,** as in **a** but cultures of 5iLA-treated hiPSC2/6/8 cells are shown (in the case of hiPSC6, we show two, n=2, independent preparations). We note that the tracks for primed hiPSC2/6/8 are duplicates of those shown in Fig. 2a and Supplementary Fig. 4a (left and right, respectively), which have been added here as references for the naive cultures of hiPSC2/6/8 cells.



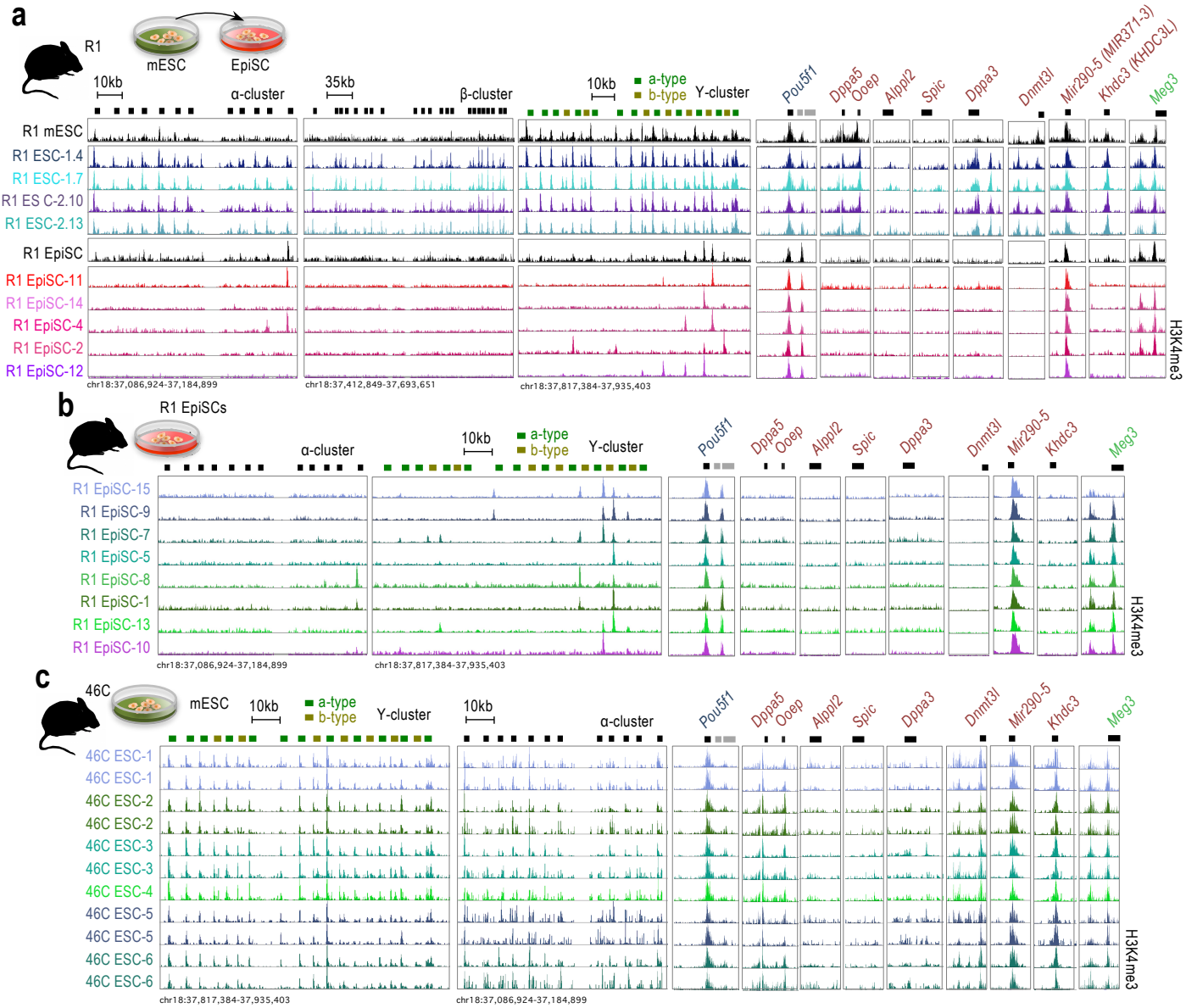
Supplementary Fig. 10. H3K4me3-enrichment on the promoters of the imprinted genes *MEG3* and *CAT* in hiPSCs and hESCs, and expression of the imprinted *MEG3* gene in single-cell hiPSC-derived neurons. a, ChIP-seq read density showing single-cell-derived hiPSC-subline-specific H3K4me3 accumulation on the *MEG3* and *CAT* promoters, complemented with RNA-seq data of hiPSC-derived neurons from the same sublines (right panel). This analysis complements the data shown in Fig. 1a. It is also shown H3K4me3 accumulation across the locus of the pluripotency gene *POU5F1* and it is highlighted the position of an enhancer (e) that is active in preimplantation-like conditions (which are not the conditions in which the hiPSC cultures shown here were grown, but analogous to postimplantation). It is also shown H3K4me3 accumulation on the promoter of preimplantation markers (*DPPA3*, *DNMT3L*, and *MIR371-3*). **b,c**, ChIP-seq read density showing single-cell-derived hESC-subline-specific H3K4me3 accumulation on the *MEG3* and *CAT* promoters. These figures complement Supplementary Fig. 4b in **b** and Supplementary Fig. 4d in **c. d**, Heatmap of counts per million (CPM) showing expression of the *MEG3* gene at the level of single cells in the n=79 single neurons analyzed in Supplemental Fig. 3a. This figure complements Supplemental Fig. 3a.



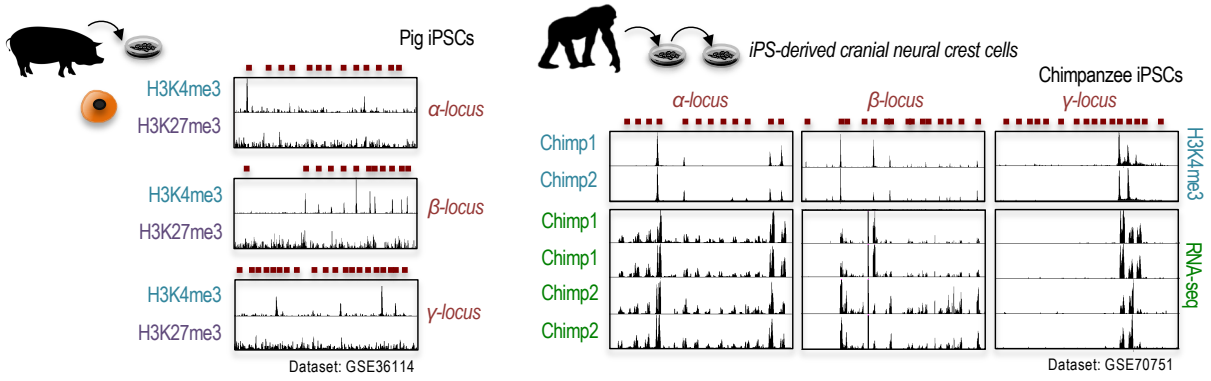
Supplementary Fig. 11. H3K4me3 and H3K27me3 accumulation across the cPcdh locus in mESCs, mouse iPSCs, and mouse EpiSCs. a, ChIP-seq data showing H3K4me3 (left panels) and H3K27me3 (middle panels) accumulation across the three cPcdh clusters in n=5 mESC lines (top panels), n=8 mouse iPSCs (middle panels, derived from three different sources of somatic cells, see below), and in n=8 mouse EpiLC/EpiSC cultures (identified as "mESCd," which stands for mESC-derived) or embryos, as indicated (bottom panels). This figure also includes H3K4me3 accumulation (right panels) on the promoters of pluripotency (*Pou5f1*) and preimplantation (*Dnmt3l, Dppa5, Ooep, Alpl2, Spic, Dppa3, Mir290-5*, and *Khdc3*) markers in the same samples. For cPcdh genes, 5' V-type exons are indicated on top. The arrows indicate relevant changes in chromatin organization in the EpiLC-EpiSC "transition", in E6.5, or in E8.5 embryos. We note that EpiLC maintain the organization observed in mESCs and iPSCs while retaining H3K4me3 enrichment in the promoter of the preimplantation markers *Dppa5* and *Mir290-5*. This enrichment is lost on the *Dppa5* promoter when the configuration in the cPcdh locus changes, but the enrichment is maintained on the *Mir290-5* promoter until late epiblast (E6.5), which would be analogous to the postimplantation-like stage of primed hESCs and hiPSCs (see Fig. 4a and 4e, top track), while reprimed would be analogous to a slightly "earlier" version (see Fig. 4f). MEF: mouse embryonic fibroblasts; APC: adipocyte progenitor cells; HPC: hepatocyte progenitor cells; rs: refers to 'region selective' EpiSCs; PSM: pre-somitic mesoderm (non-pluripotent cells). Dataset sources: indicated at the bottom of the panels. **b**, CTCF ChIP-seq read densities across the γ- and α-clusters in a culture of Bruce4 mESCs. 5' cPcdh exons indicated on top. Dataset source: ENCSR000CCB.



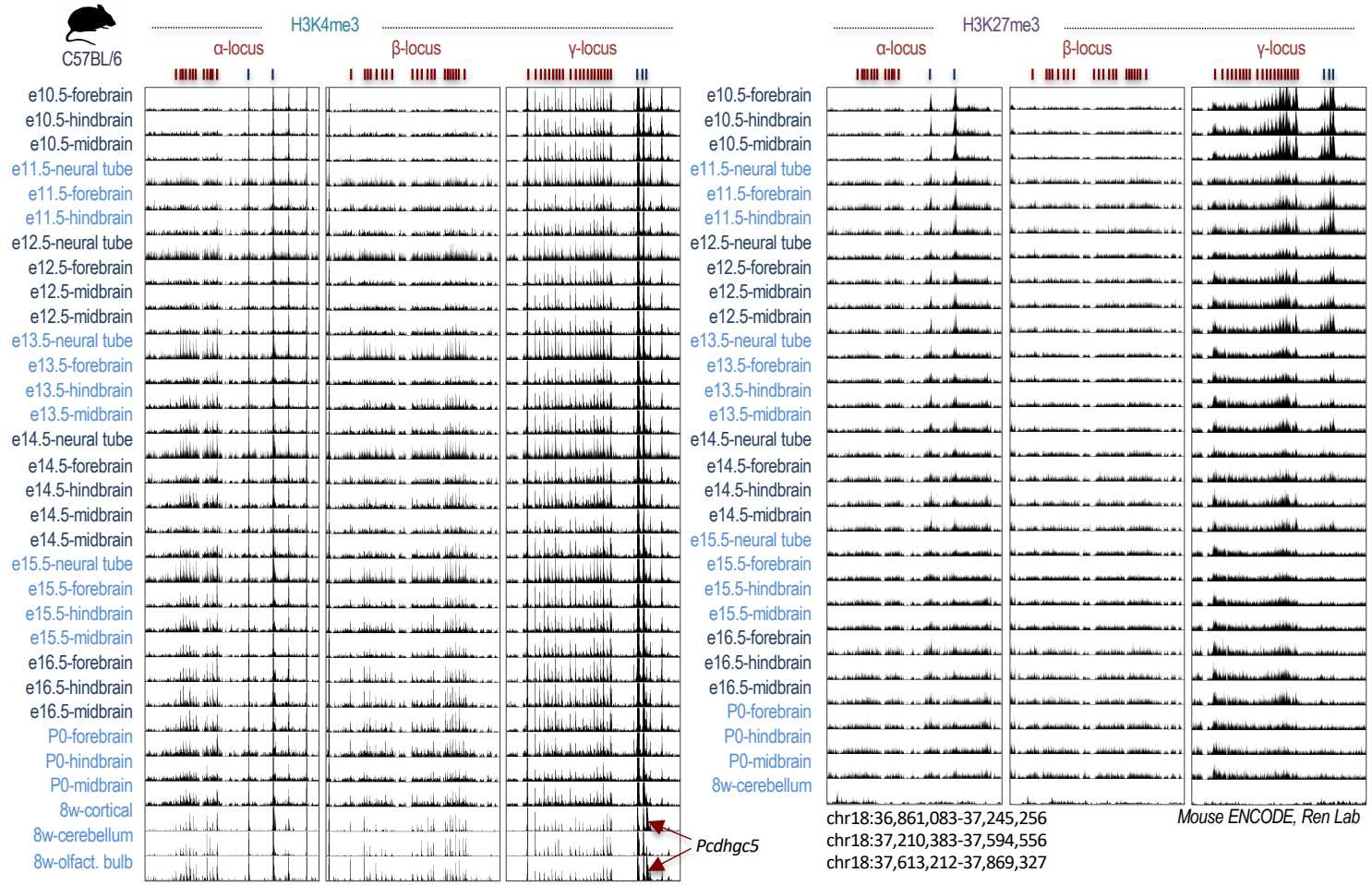
Supplementary Fig. 12. H3K4me3 and H3K27me3 enrichment across the c-Pcdh locus during the mouse embryonic fibroblast (MEF)-to-iPSC conversion. ChIP-seq data showing gradual elimination followed by gradual accumulation of H3K4me3 and H3K27me3 across the cPcdh locus during the process of cell reprogramming of a line of mouse iPSCs from a MEF culture. Days after initiated the process are indicated on the side. Included also loci used as markers and color coded, as indicated: fibroblast identity (*Thy1*), pluripotency (*Nanog, Sall4, Lin28a, Esrrb,* and *Pou5f1*), preimplantation (*Dnmt3l, Dppa5, Ooep, Alpl2, Spic, Mir290-5*, and *Khdc3*), and imprinting (*Meg3*). Dataset source: GSE67520.



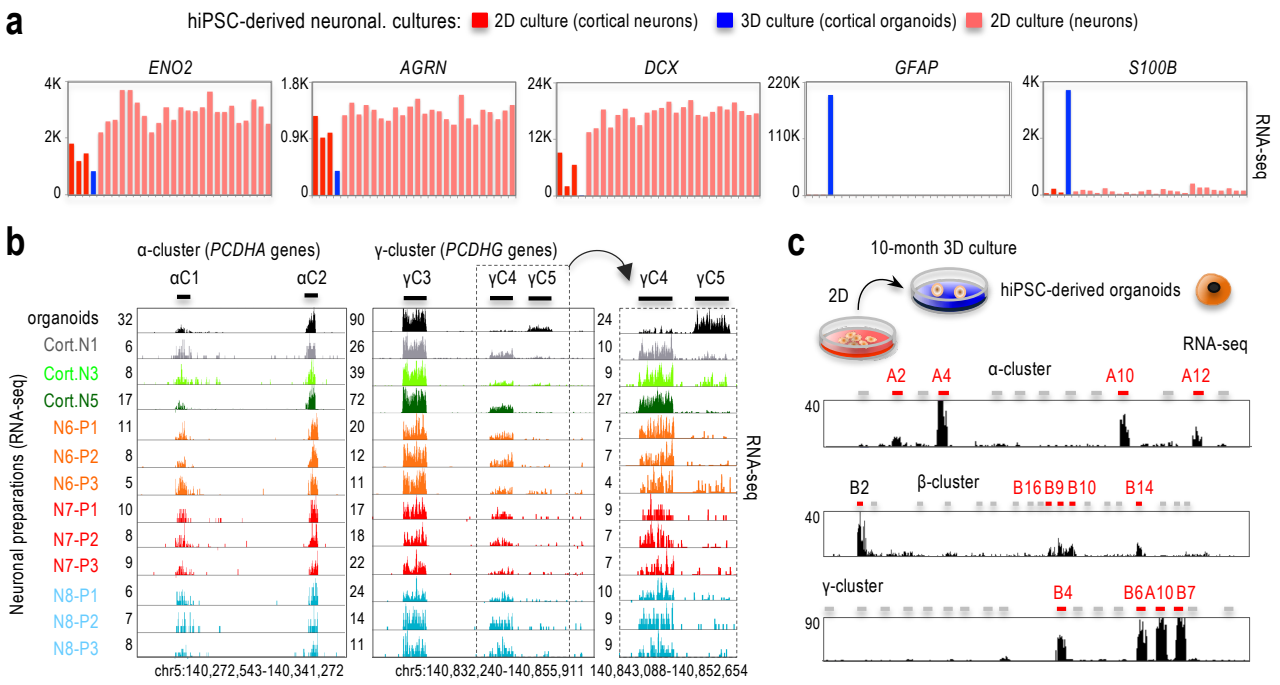
Supplementary Fig. 13. H3K4me3 enrichment across the cPcdh locus in single-cell-derived mESC and EpiSC sublines. **a-c,** H3K4me3 ChIP-seq read densities across the cPcdh clusters in n=4 independent single-cell derived sublines from the R1 mESC line (1.4, 1.7, 2.10, and 2.13) and in n=5 independent single-cell derived sublines (11, 14, 4, 2, and 12) from the R1 line after derivation into EpiSCs shown in **a** (the parental R1 mESC and EpiSC lines are also shown), in n=9 additional independent single-cell derived sublines from the R1 EpiSC line shown in **b** (15, 9, 7, 5, 8, 1, 13, and 10), and in n=6 independent single-cell derived sublines from the 46C mESC line (five of them in duplicate cultures, n=2) shown in **c**. In **a-c**, we also show (right panels) markers of pluripotency (*Pou5F1*), markers of the naive state (*Dppa3, Dppa5, Ooep*, *Alppl2, Spic, Dnmt3l,* and *Khdc3*), markers of the naïve state and early epiblast (*Dppa5* and *Mir290-5*), and the imprinting gene *Meg3* (which is not expressed in all EpiSC sublines, but it is expressed in all mESC lines). 5' cPcdh exons and the rest of genes are indicated above the tracks.



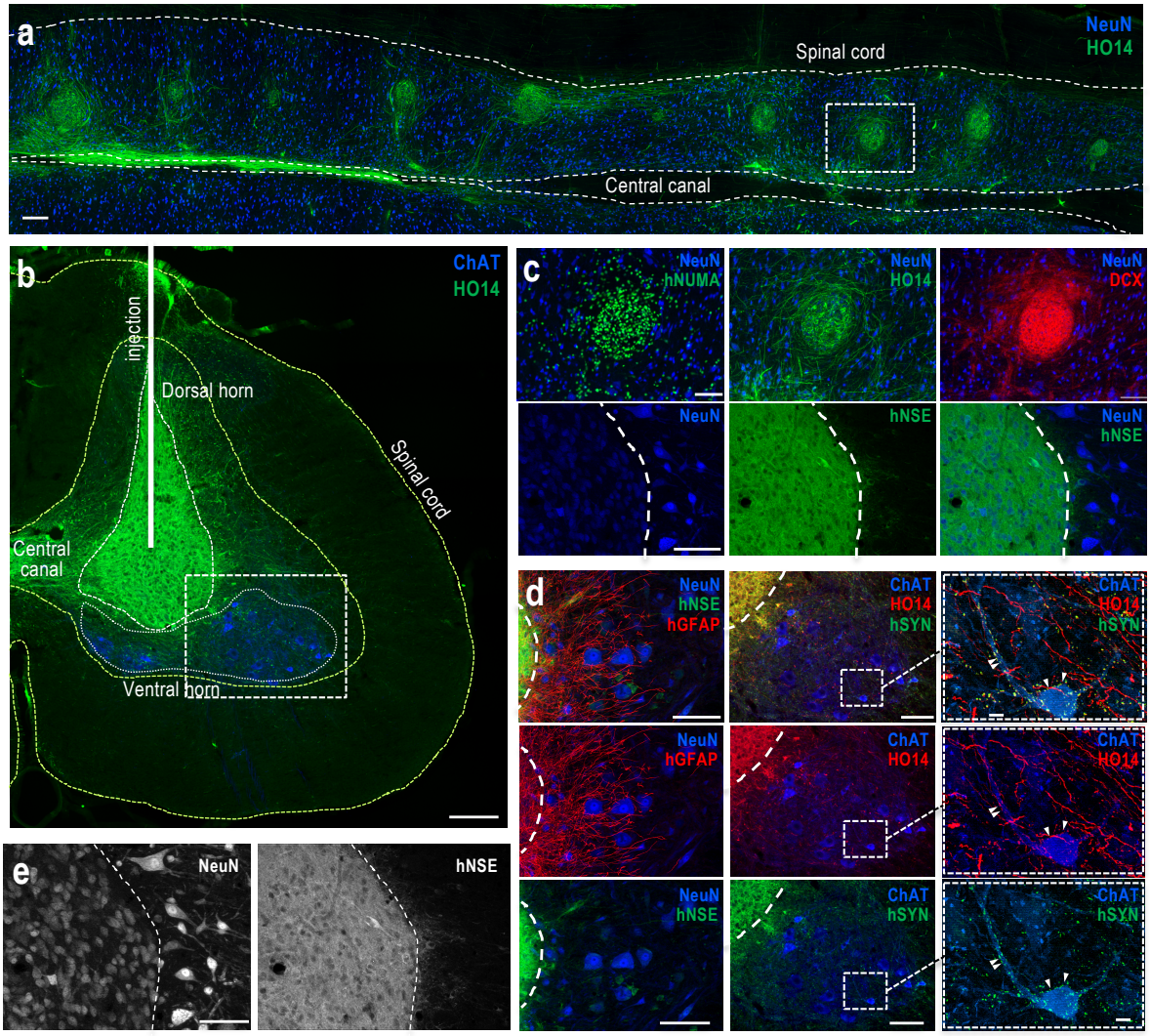
Supplementary Fig. 14. Chromatin enrichment in the cPcdh locus of pig iPSCs and chimp iPSC-derived cranial neural crest cells. H3K4me3 and H3K27me3 ChIP-seq data in a line of pig iPSCs (left panels) and H3K4me3 ChIP-seq and RNA-seq data in two independent lines of chimp iPSC-derived cranial neural crest cells showing patterns of chromatin organization and expression across the cPcdh locus as those observed in hiPSCs and derived neural cells, respectively. As in human primed cells, H3K4me3 accumulates highly selectively and in the absence of H3K27me3 in (primed) pig iPSCs and, as in hiPSC-derived neural cells, cranial neural crest cells derived from chimp iPSCs also show patterns of H3K4me3 accumulation and cPcdh expression consistent with highly selective patterns. 5' cPcdh exons represented above the tracks. Dataset sources: GSE36114 and GSE70751.



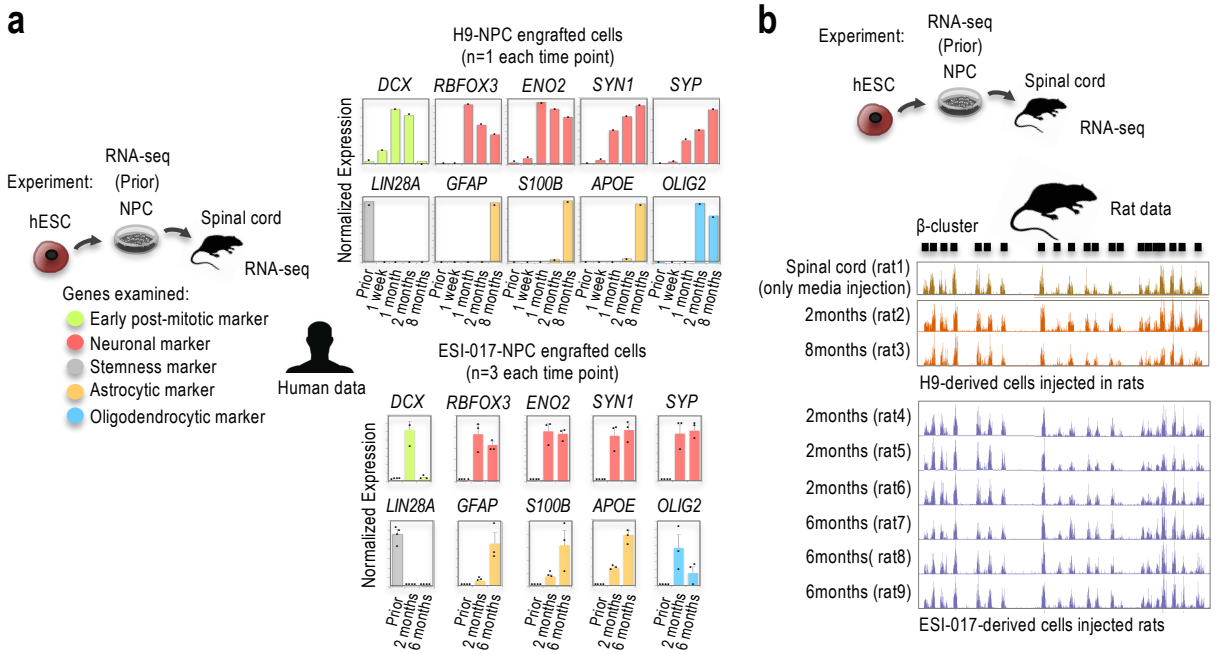
Supplementary Fig. 15. H3K4me3 signal across the cPcdh locus in mouse fetal and adult brains. ChIP-seq read density tracks (ENCODE data) showing H3K4me3 accumulation across the three clusters in the cPcdh locus of the indicated samples. We note the similarity between the tracks shown here of E10.5 and E11.5 embryos and the track shown in Supplementary Fig. 11 of E8.5 embryos (although when comparing these tracks, it should be noticed that C-type promoters are also shown here, but not in Supplementary Fig. 11). We also note that the arrows indicate H3K4me3 accumulation on the *Pcdhgc5* promoter, which is a sign of the postnatal stage in human brain not observed in the adult cerebellum (see *PCDHGC5* expression in the human brain in Fig. 6b; low *PCDHGC5* expression in human cerebellum not shown). Dataset sources: H3K4me3 - ENCFF163RNG, ENCFF251WXG, ENCFF200DJC, ENCFF638WLA, ENCFF437KKV, ENCFF876VMW, ENCFF552KFS, ENCFF471MKR, ENCFF173HJD, ENCFF014TDL, ENCFF293PAY, ENCFF119USK, ENCFF483RME, ENCFF718JOG, ENCFF229GJB, ENCFF072OAK, ENCFF657XFB, ENCFF953IAT, ENCFF253JKS, ENCFF030BTL, ENCFF911RYA, ENCFF884HYZ, ENCFF960GUP, ENCFF352BZY, ENCFF748WAP, ENCFF006GSW, ENCFF084XLB, ENCFF656DDK, ENCFF191GQW, ENCFF553HUH, ENCFF087DDW, ENCFF495YQS; H3K27me3 - ENCFF063ZLI, ENCFF898VDV, ENCFF293RMV, ENCFF217EPP, ENCFF375ONW, ENCFF485WCD, ENCFF901MQW, ENCFF597HKP, ENCFF349EPV, ENCFF612RDE, ENCFF009QET, ENCFF760VJL, ENCFF637PDW, ENCFF610HBD, ENCFF223PHS, ENCFF900BKA, ENCFF787DNX, ENCFF908XOQ, ENCFF386EYG, ENCFF848GOF, ENCFF408PKX, ENCFF376ZJO, ENCFF935NKI, ENCFF574LGA, ENCFF025FNF, ENCFF183CXT, ENCFF156SIF, ENCFF149AGA, GSM1000090.



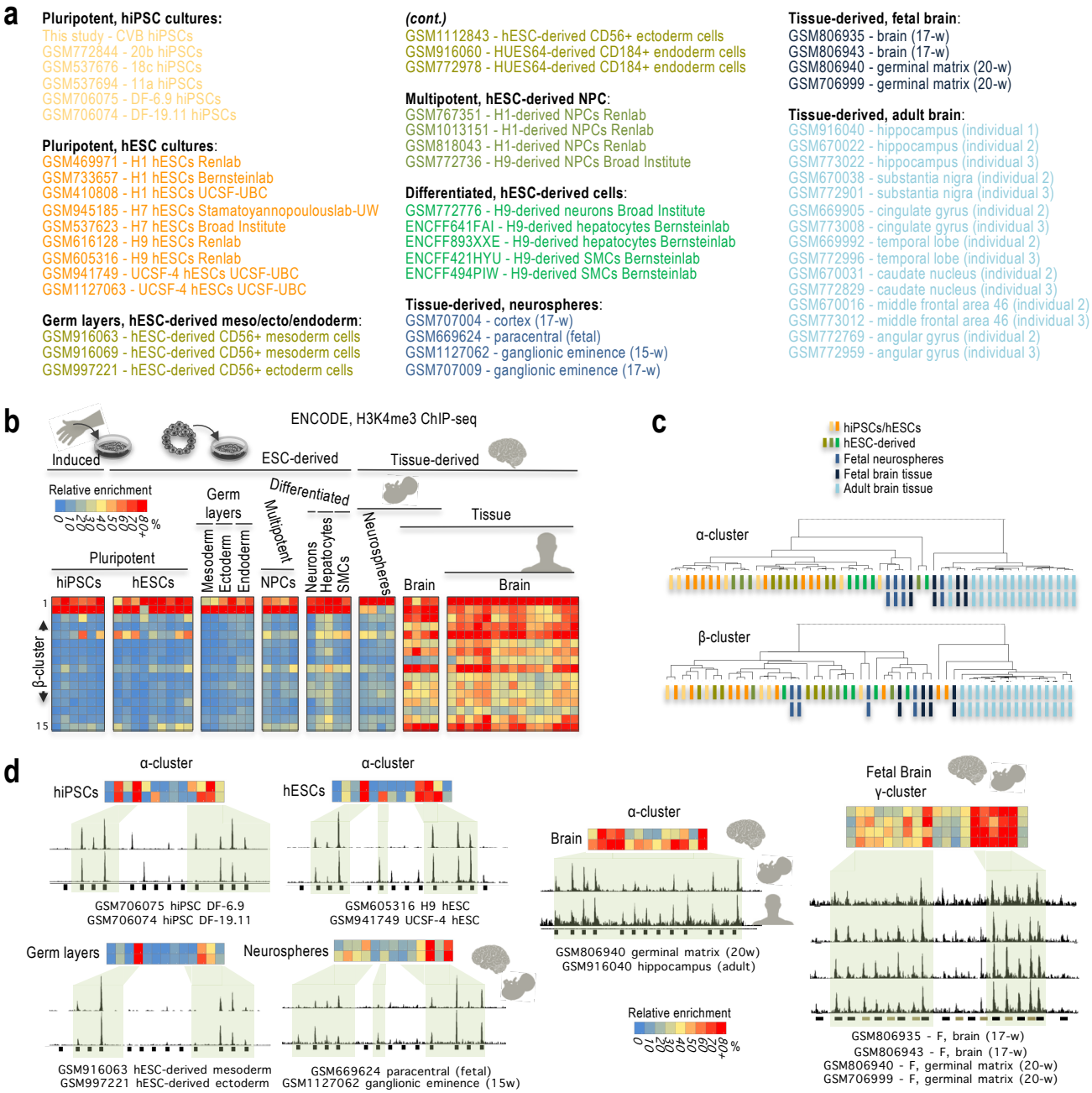
Supplementary Fig. 16. Analysis of cPcdh expression in 10-month hiPSC-derived cortical organoids. a, RNA-seq data showing expression of neuronal genes (*ENO2* and *AGRN*), postmitotic marker (*DCX*), and astrocytic markers (*GFAP* and *S100B*) in 2D cultures of hiPSC1/3/5-derived cortical neurons (n=3, same as in Supplementary Fig. 2a), 10-month hiPSC-derived organoids (n=1 pool of approximately 5-8 organoids), and 2D cultures in triplicate (n=3) independent preparations of hiPSC1-8-derived neurons (n=8, same as in Fig. 1a). **b**, RNA-seq read density tracks showing the expression of the five C-type cPcdh genes in the 10-month old hiPSC-derived cortical organoids (n=1 pool of approximately 5-8 organoids, top track), in hiPSC1/3/5-derived cortical preparations (n=3, second, third, and fourth tracks), and in triplicate independent hiPSC6/7/8-derived neuronal preparations (n=3x3), as indicated. 5' exons of the C-type genes indicated above the tracks. Scales are indicated on the side. We note a "shift" in expression of the C4 and C5 isoforms when comparing 2D and 10-month 3D cultures: *PCDHGC4* is expressed in 2D cultures (largely prenatally in the human brain, see profiles in Fig. 6b) while *PCDHGC5* is expressed in the 10-month 3D culture (largely postnatally in the human brain, see profiles in Fig. 6b). We, therefore, propose that the C4/C5-shift is a marker of neuronal maturity in human neurons that only "old" 3D cultures can recapitulate, but not 2D cultures or "younger" 3D cultures (not shown). **c**, RNA-seq read density tracks showing the expression of cPcdh genes in 10-month old cortical organoids derived from hiPSCs separated by clusters. The subset of expressed 5' cPcdh exons is indicated in red above the tracks, which follows the highly selective pattern observed also in 2D cultures of hiPSC/hESC-derived neurons (e.g. Fig. 1a).



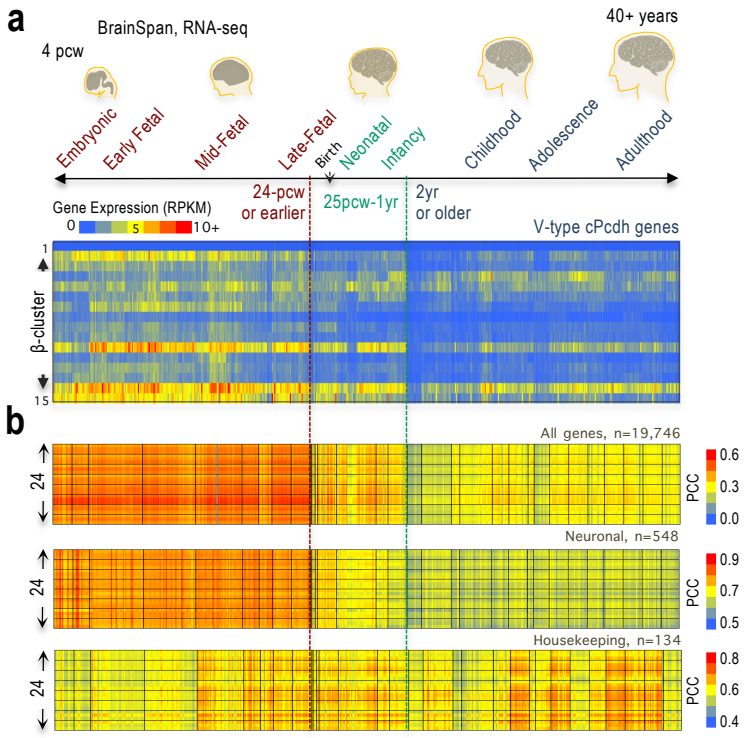
Supplementary Fig. 17. Immunostaining-based characterization of hESC-derived NPC grafted cells in rat spinal cords. **a**, Coronal section of rat spinal cord (lumbar area) taken 2 month after transplantation of ESI-017 hESC-derived cells. Section stained with NeuN (neuronal nuclear protein) and human-specific H014 antibodies. Multiple sites of injections in the central gray matter are shown. The dotted square indicates the approximate location of the top images shown in **c**. Scale bar=200µm. **b**, Transverse section of rat spinal cord (lumbar area) taken 6 month after transplantation of ESI-017 hESC-derived NPCs and stained with NeuN and choline acetyltransferase (ChAT) antibodies, which stain motor neurons. The approximate direction of the hESC-NPC injection is indicated. The dotted square indicates the approximate location of the bottom images shown in **d**. Scale bar=200µm. **c**, On top, immunofluorescence detection of postmitotic markers human-specific nuclear marker (hNUMA), human axonal neurofilament HO14, and doublecortin DCX is shown in 2-month coronal spinal sections from ESI-017-NPC-injected rats. The top-left and top-right images correspond to the same preparations stained with different antibodies. On the bottom panels, immunoreactivity of post-mitotic and early maturation marker NeuN and human human neuron specific enolase 2 (NSE), encoded by the *RBFOX3* and *ENO2* genes, respectively, is shown in 6-month transverse spinal sections from ESI-017-NPC-injected rats. **d**, Panels show astrocytic (glial fibrillary acidic protein or GFAP-positive) projections arising from the grafted area and human synaptophysin (hSYN)-positive puncta in human HO14 positive areas in the vicinity of host motor neurons (identified as ChAT-positive cells; arrowheads). The top images represent three staining together; middle and bottom images are the same images only showing two staining each time, for clarity. Panels show immunofluorescence of NeuN, human NSE, GFAP, ChAT, human HO14, and human SYN in 6-month transverse spinal sections from ESI-017-NPC-injected rats. Scale bars: 100µm, except for the bottom-right panel, in which the bar represents 10µm. **e**, Black and white images of the left and middle panels shown in **c**. Scale bar=100µm.



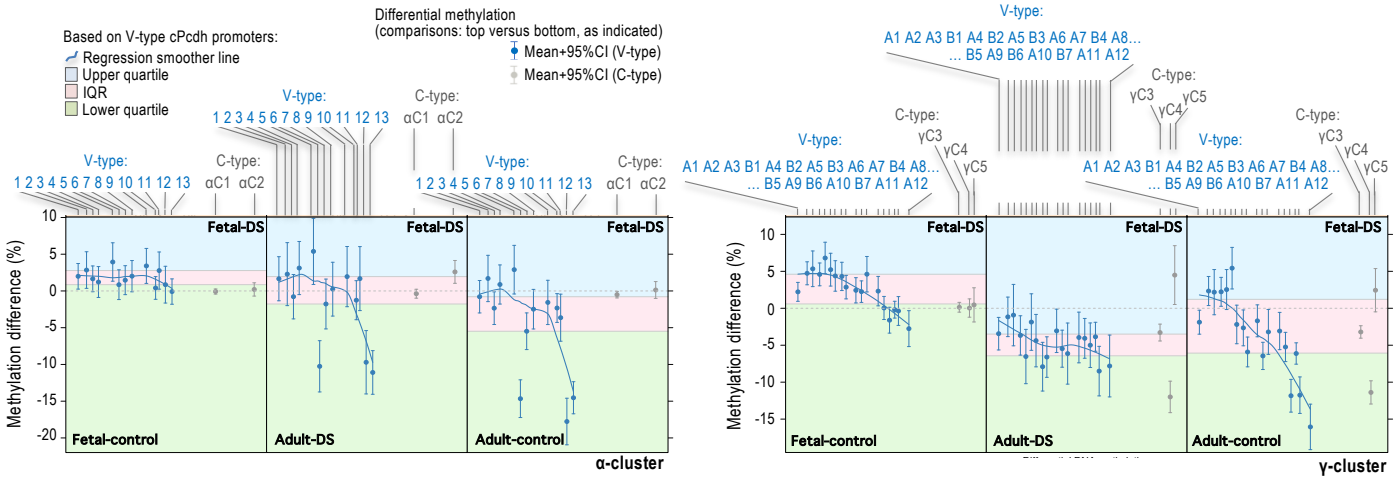
Supplementary Fig. 18. RNA-seq analysis of hESC-derived NPC grafted cells and the surrounding tissue in the rat spinal cord. **a**, Normalized expression of a selection of human genes (listed on top of each panel) in H9-derived NPCs (top panels) and ESI-017-derived NPCs (bottom set of panels) prior to transplantation (Prior) and in grafted cells after the indicated time periods (from 1 week to 8 months). The samples are biopsies from the central grey matter of the rat spinal lumbar region: n=2 independent H9 hESC-derived NPC suspension prior to injection; n=1 rat for each condition after 1 week, 1 month, 2 months, and 8 months from the time of injection of H9 hESC-derived NPC suspensions; n=4 independent ESI-017 hESC-derived NPC suspension prior to injection; and, n=3 rats each after 2 and 6 months from the time of injection of ESI-017 hESC-derived NPC suspensions. The type of selected genes analyzed in the figure is color coded: early post-mitotic marker (n=1 gene), neuronal marker (n=4 genes), stemness marker (n=1 gene), astrocytic marker (n=3 genes), and oligodendrocytic marker (n=1 gene). The plots represent mean expression values in independent cultures (NPC, prior) or in different rats (after injection) and error bars indicate s.e.m. **b**, RNA-seq data showing expression of rat β-cPcdh genes (indicated above of the panel) in rat spinal cord. The samples in **b** are also shown in **a** and in Fig. 5a-c, but the human and rat reads (shown in **a** and **b**, respectively) were computationally separated (see Methods). Scale=adjusted to the highest value in each track.



Supplementary Fig. 19. H3K4me3 accumulation on cPcdh promoters in human brain tissue (ENCODE data). a, List of cell cultures, cell preparations, or postmortem brain tissue samples analyzed in **b** and in Fig. 6a, in particular: n=6 hiPSC lines (DF stands for "derived from"), n=9 independent hESC cultures (n=4 lines cultured by different laboratories, as indicated), n=6 independent hESC/hiPSC-derived germinal layer cultures, n=4 independent hESC-derived NPC cultures, n=1 hESC-derived neurons, n=2 independent hESC-derived hepatocyte cultures, n=2 independent hESC-derived smooth muscle cell (SMC) cultures, n=4 independent fetal-derived neurosphere cultures, n=4 fetal brain tissue samples, and n=15 adult brain tissue samples. Listed in the same order as shown in **b** and in Fig. 6a. The information provided here include: GEO accession number, cell/tissue source, laboratory when relevant, brain region when relevant, age when fetal samples, individual 1/2/3 when relevant; however, more detailed information can be found in the listed GEO accession numbers. **b**, Heatmap showing relative enrichment of H3K4me3 signal on β-promoters quantified by ChIP-seq. The promoter with the highest signal in each sample represents 100% and the rest of promoters are represented relative to 100%. The list of samples is shown in **a**. **c**, Hierarchical clustering (one minus Spearman rank correlation) of the samples shown in Fig. 6a (top panel, α-promoters) and in **b** (β-promoters) as indicated. **d**, This panel complements Fig. 6a and allows an overall assessment of how accurately the quantification of relative H3K4me3 enrichment shown in Fig. 6a represents the source of H3K4me3 ChIP-seq signal used to generate Fig. 6a. It shows a side-by-side comparison of H3K4me3 peaks (tracks) and quantifications (heatmaps) for the indicated samples. The heatmaps are duplicated here from Fig. 6a.



Supplementary Fig. 20. Expression of β-cPcdh genes in the developing and adult human brains (BrainSpan data; shown in *a*), and similarity analysis of full (top panel), neuronal (middle panel), and housekeeping (bottom panel) transcriptomes between our set of three (n=3) independent differentiation preparations of n=8 hiPSC1-8-derived neuronal cultures (n=24, shown in Fig. 1a) and each of the n=524 brain tissue samples of the developing and adult human brains (BrainSpan data; shown in *b*). a, Heatmap shows RNA-seq data (RPKM, exonic, source: BrainSpan) for the n=16 β-genes (PCDHB1-16, represented in the same order as they are organized in the human genome) in the n=524 biologically independent samples of the human developing and adult brains. **b**, Heatmaps show similarity analysis based on the Pearson correlation coefficient (PCC) of RNA-seq data for the indicated gene groups above each panel: top panel, n=19,746 RefSeq genes (full transcriptome); middle panel, n=548 neuronal genes (a fraction of the neuronal transcriptome); and, bottom panel, n=134 housekeeping genes (a fraction of the housekeeping transcriptome). The compared samples are the n=24 neuronal preparations shown in Fig. 1a (rows), which correspond to three independent differentiation preparations of eight different sublines, and each of the n=524 BrainSpan tissue samples (columns). The purpose of this analysis is to show higher similarity between our n=24 hiPSC-derived neuronal preparations and the tissue samples of 24-pcw or earlier brains than between our n=24 hiPSC-derived neuronal preparations and tissue samples of 25-pcw or older brains when the comparison is based on full or neuronal transcriptomic profiles (top and middle panels) but not when the comparison is based on housekeeping genes (bottom panel). As expected, therefore, our hiPSC-derived neuronal preparations are more similar to embryonic and early/mid fetal human brain than to the late fetal, postnatal, and adult brain when the full or neuronal transcriptomes are examined. The decrease in similarity beyond the fetal stage could be explained as result of differences in neuronal maturity (prenatal versus postnatal neurons) or as result of differences in cellular composition (primarily neuronal versus a mix of neuronal and glial populations –in fact, our n=24 preparations show no apparent glial signal, see Supplementary Fi. 16a, similarly to embryonic/early/mid brain, see Fig. 6b, whereas late fetal, postnatal, and adult brains show strong glial signal, see Fig. 6b), or both.



Supplementary Fig. 21. Differential DNA methylation analysis of human fetal DS brains versus fetal control, adult DS, and adult control brains based on Illumina HumanMethylation 450K arrays. Panels show mean differential methylation values with 95% confidence intervals (CI) at promoter regions of V-type (blue circles) and C-type (grey circles) α-cPcdh genes (in the left three panels) and γ-cPcdh genes (in the right three panels). In all one-on-one comparisons, the upper side corresponds to fetal DS brains and the lower side corresponds to fetal control, adult DS, or adult control brains, as indicated (first, second, and third panels, respectively). The samples are postmortem brain tissue (see data source in Supplemental Note), in particular: fetal subjects, n=16 DS and n=27 control biologically independent samples, median gestational stage 18-pcw and 20-pcw, respectively; adult subjects, n=6 DS and n=26 control biologically independent samples, median age 45 years and 49.5 years, respectively. The promoter regions were defined as -2,200bp to +200bp from the annotated transcriptional start site (TSS). The background in each panel is colored based on the interquartile range (IQR, in pink) and upper/lower quartiles (in blue and green, respectively) calculated based on V-type promoter values (i.e. excluding C-type promoter values). The regression smoother line is also based on V-type promoter values (blue line). V-type and C-type promoter names are indicated above each individual panel.

Supplementary Table Legends

Supplementary Table S1. List of cultures/lines/sublines generated or used in this study.

Supplementary Table S2. List of sequencing runs generated in this study.

Supplementary Table S3. List of genomic coordinates for the RNA-seq analysis of cPcdh genes and controls.

Supplementary Table S4. List of genomic coordinates (hg18) for the ChIP-seq analysis of regulatory regions across the cPcdh locus.

Supplementary Note

Clonal origin (not protocol of differentiation) dictate cPcdh selections

In order to robustly confirm that hiPSCs (not the protocol of neuronal differentiation) dictate the pattern of preferential frequencies of α/γ-isoform selection in hiPSC-derived neurons, we tested a second protocol of neuronal differentiation in three of the single-cell-derived sublines (hiPSC1/3/5). This protocol induces cortical neurons and it does not include a step of FACS-mediated enrichment of the same cell population in all preparations, as opposed to the protocol applied to generate neurons in Fig. 1a (see Methods). Still, cortical neurons express virtually the same cPcdh selections as their clonally related cells generated with the first protocol (compare Cort.N1/N3/N5 in Supplementary Fig. 2a with N1/N3/N5 in Fig. 1a). We note that the new cortical preparations derived directly from hiPSC1/3/5 (not from NPC1/3/5).

Next, we tested the effect of repeatedly deriving NPCs from the same single-cell-derived hiPSC subpopulation (n=4). It results in a similar pattern of cPcdh expression in every preparation, again in support that cPcdh selections are pre-set in hiPSCs (Supplementary Fig. 2b, NPC P1-P4). Also, in line with clonal origin of hiPSCs as major determinant of neuronal cPcdh selections, clonally related neurons, NPCs, and hiPSCs express similar cPcdh choices (Supplementary Fig. 2c). We note, however, that while we observe remarkable consistency in the relative expression of cPcdh isoforms among replicates of neuronal preparations (see Fig. 1a) or among replicates of NPC preparations (Supplementary Fig. 2b), in some cases we observe substantial variations in the relative levels of cPcdh-isoforms comparing clonally related hiPSCs, NPCs, and neurons (see, for instance, *PCDHA6* or *PCDHGB6*). In a recent work published during the revision of our study that examines α-cPcdh expression in mESC-derived NPCs and neurons at a single-cell scale (ref. 38), the authors show that only 20% of cells in a mESC-derived NPC population express cPcdh isoforms, whereas 100% of neurons derived from these cells express cPcdh isoforms. If such dramatic difference in the total number of individual cells expressing cPcdh genes is also observed in hiPSC-derived NPCs and neurons, this property may alter the relative levels of cPcdh expression of some individual isoforms at a population-wide scale, which may explain our observations. Another important note is that the absolute values of expression between hiPSCs, NPCs, and neurons are -as expected- strongly associated with cell identity (i.e. hiPSCs only express marginal

levels of the cPcdh genes, and NPCs and, especially, neurons express much higher levels than hiPSCs).

Finally, we also examined the patterns of cPcdh expression in clonally related neurons and astrocytes derived from the same NPC subpopulation (Supplementary Fig. 2d). These cells also share similar cPcdh combinations although, as in the comparison of NPCs and neurons, astrocytes and neurons also show some variations in relative expression among isoforms.

H3K4me3 patterns along the cPcdh locus as a marker of hiPSC/hESC line or subline identity

One of the most surprising findings in this study is that we have found that almost every hiPSC or hESC line examined shows a distinct pattern of H3K4me3 accumulation along the α/γ-clusters, which may be used for authentication purposes. Only in a few cases we have found the same patterns in different lines or sublines (e.g. in the parent CVB and CVI lines). We highlight that these patterns are not associated to donor identity, since the 18a and 18c hiPSC lines are derived from the same donor and show different H3K4me3 patterns (in Supplementary Fig. 4e) or the primed and 'naive' versions of the WIBR1 line, the WIBR2 line, or the WIBR3 line derive from two halves, in each case, of the same blastocyst embryo (in Supplementary Fig. 7a). We have observed remarkable stability of these patterns thus, when they change, we suspect that it could be because there might have been a step of enrichment of a subpopulation (e.g. we have observed that some hESC lines cultured by different laboratories show different H3K4me3 patterns along the cPcdh locus, perhaps because the different histories of these cultures). An extreme case of enrichment is the step of single-cell isolation and clonal expansion during the application of a protocol of genome editing, regardless of successful editing or not (Fig. 1a). This aspect might be critical if hiPSCs/hESCs are edited prior to deriving neurons, since the functional consequences of comparing two subpopulations with different cPcdh signatures are unclear at this moment but could generate confounding effects with the edited genotype.

Another interesting feature is that a new round of single-cell isolation and clonal expansion from an already single-cell-derived subpopulation of hiPSCs segregates patterns that are similar but not identical to the parent population. For instance, we observe that single-cell-derived subpopulations of the hiPSC1 subline segregates patterns that combined do not exactly resemble the parent hiPSC1 or the parent CVB line. One possibility is that the number of sublines that we

have generated is small (hiPSC1 1-6), while this effect should be analyzed with a much larger number of sublines. A second possibility is that the process of single-cell isolation may trigger some limited resetting of the H3K4me3 patterns along the cPcdh locus. Two observations help us to favor this model. First, as just mentioned, the cumulative signal of the hiPSC1 1-6 sublines is not identical to the signal of hiPSC1. Second, we have observed a transient epigenetic activation of a marker of the pre-implantation stage, *MIR371-3*, after single-cell isolation of hiPSCs/hESCs (data not shown). This result suggests that the process of single-cell isolation may induce a partial and transient reversion (during the first passages) of the primed state to a new type of naive-like state. A previous study did not find evidence of a transition to a naive state after single-cell passaging of pluripotent cells[1]. However, in contrast to our study, this previous study did not interrogate the expression of the epigenetic status of *MIR371-3* or *DPPA5*, which are relevant genes to understand naive/primed conversions, as it will be shown later, nor they did examine a genuine process of single-cell isolation, but a process of single-cell passaging (i.e. passaging of cell suspensions with many single cells).

Chromatin organization of the cPcdh locus in hiPSCs, hESCs, SK-N-SH, and K562 cells

We were intrigued by the type of chromatin structure that could be pre-setting frequencies of cPcdh selection for neurons in hiPSCs. To gain insights into this organization, we performed a side-by-side comparison of H3K4me3, CTCF, and Rad21 ChIP-seq signal on cPcdh promoters and nearby regulatory regions among hiPSCs, neuroblastoma SK-*N*-SH cells, and leukemia K562 cells. This analysis reveals remarkable similarities between hiPSCs/hESCs and SK-*N*-SH cells, and profound differences between hiPSCs and K562 cells (Fig. 2c,d and Supplementary Fig. 5a,b). The major difference between hiPSCs and SK-*N*-SH cells occurs on a distal site, γ-HS5-1aL, proposed to underlie an enhancer. The major differences between hiPSCs and K562 cells include: poor H3K4me3 accumulation on α/β-promoters and α-HS5-1/γ-HS5-1aL sites, and a distinct configuration of the CTCF/Rad21 peaks at the α-HS5-1/γ-HS19 sites. We note, in any case, that these comparisons limit to H3K4me3, CTCF, and Rad21. In fact, hiPSCs/hESCs, as will show, also accumulate repressive chromatin along the cPcdh locus (Supplementary Fig. 5d), which is not observed or expected in SK-*N*-SH cells (not shown but publicly available in ENCODE datasets). In other words, the similarities between hiPSCs/hESCs and SK-*N*-SH cells are likely limited to a few activating components. The presence of repressive chromatin is what likely maintains an

inactive transcriptional state in the cPcdh locus of hiPSCs/hESCs. Intriguingly, analysis of previously generated HiC-seq data reveals remarkable similarities in the spatial configuration of the cPcdh locus between H1 hESCs and H1-derived NPCs (Supplementary Fig. 5c), which further supports that pluripotent cells display an active-like chromatin organization despite cPcdh genes being barely expressed in these cells, likely maintained repressed by complementary repressive chromatin.

Expression of cPcdh genes in iNs

If -as suggested by our analyses- hiPSC-derived neurons make cPcdh choices based on chromatin features established by progenitor (non-neuronal) cells (i.e. hiPSCs), neurons directly derived from somatic cells without undergoing an intermediate step of pluripotency (known as induced neurons, or iNs)[6] should express cPcdh genes mirroring the chromatin features of their source of somatic cells. To test this hypothesis, we exploited a chromatin singularity of skin fibroblasts, which is a typical source of somatic cells to derive iNs. This singularity refers to a complete absence of H3K4me3 deposition along the α-cPcdh cluster (Supplemental Fig. 6a). During the process of skin fibroblast-to-hiPSC conversion, this histone mark newly accumulates on V-type promoters along the α-cluster in hiPSCs, in parallel with its deposition on promoters of pluripotency regulators (e.g. *NANOG*, *LIN28A*, and *POU5F1*, Fig. 3a and Supplemental Fig. 6b)[7]. In agreement with our hypothesis, iNs show marginal expression levels of α-isoforms compared to β/γ-isoforms, which translates into a log10-scale radar plot with a 'bat-like' shape (Fig. 3a, first/second columns, skin fibroblasts/iNs, respectively; top/bottom represent n=48 V-type and n=49 housekeeping genes, respectively; other visualizations shown in Supplementary Fig. 6c,d). On the contrary, iNs generated from skin fibroblasts after adding an intermediate state of hiPSC conversion exhibit a pattern of cPcdh expression characterized by a log10-scale radar plot with a 'butterfly-like' shape (Fig. 3a, third/four columns), as in the case of hiPSC-derived neurons and hiPSC-derived iNs generated with a second protocol of iN derivation[8] (Fig. 3a, fifth column; Fig. 3b and Supplementary Fig. 6e).

Repriming effects on the chromatin organization of the cPcdh locus

The question is whether re-primed cells retain a memory of the original (primed) α/γ-cluster configuration. In the α/γ-clusters of 5iLA-naive HUES9 1.8 cells, H3K4me3 accumulates to some degree on virtually every α/γ-promoter (Fig. 4f and Supplementary Fig. 8c). In re-primed HUES9 1.8 cells, H3K4me3 is selectively eliminated from some α/γ-promoters, as a sign of a frequency pre-setting, and the re-primed configuration is completely different from the original version detected in HUES9 1.8 cells, as a sign of genuine resetting. With regard to the preimplantation markers, only the *MIR371-3* promoter retains some level of H3K4me3 enrichment in re-primed cells (Fig. 4f). Unlike the primed version, the re-primed locus exhibits a relatively broad selection of cPcdh promoters (Fig. 4f and Supplementary Fig. 8c; as similarly observed in the α-cluster of hiPSC-T cells, Supplementary Fig. 6b). We therefore postulated that if single cells were isolated and expanded from re-primed cultures, the new subpopulations would exhibit a broader subset of enhanced α/γ-promoter selections compared to single-cell-derived HUES9 sublines. To test this hypothesis, we isolated and expanded eight single cells from re-primed HUES9 1.8 cultures prior to the application of H3K4me3 ChIP-seq. In support of our hypothesis, we observed differences among some sublines (Supplementary Fig. 9a), but overall poor diversity of selections compared to single-cell-derived subpopulations of primed hiPSCs (Fig. 2a). In short, a new set of enhanced selections was generated after exiting the 5iLA-naive state, but the principles of selection were apparently different from those after ICM derivation to generate hESCs or after reprogramming to generate hiPSCs. We also interrogated the stimulation of the 5iLA-naive state in single-cell-derived hiPSCs (Supplementary Fig. 9b). As in hESCs, this process led to the formation of virtually undistinguishable α/γ-promoter selections among the different 5iLA-naive hiPSC subpopulations, although the level of H3K4me3 accumulation was not identical to that observed in naive hESCs (Supplementary Fig. 9b).

Stability of hESC/hiPSC-guided cPcdh selections in engrafted neurons in the rat spinal cord

We injected H9 and ESI-017 hESC-derived NPC suspensions in the central grey matter of the spinal lumbar region of n=10 adult rats and only media in n=4 control rats (Methods). After 1 week, 1, 2, or 8 months in H9-NPC-injected rats (n=1 rat for each time point), and after 2 or 6 months in ESI-017-NPC-injected rats (n=3 rats for each time point), grafted cells were processed for immunostaining or RNA-seq analysis (Fig. 5a and Supplementary Fig. 17a,b). In support of

cell grafting and neuronal differentiation, post-mitotic markers were immunoreactive after 2 months in the rat spinal cord (human nuclear marker, hNUMA; human axonal neurofilament, HO14; and doublecortin, or DCX, which expectedly was not detected in the surrounding rat spinal cord because doublecortin is expressed in this rat tissue only until a few days after birth[9]; top panels in Supplementary Fig. 17c). We also detected early maturation markers, such as neuronal nuclear protein (NeuN) and human neuron specific enolase 2 (NSE), encoded by the *RBFOX3* and *ENO2* genes, respectively (bottom panels in Supplementary Fig. 17c; Supplementary Fig. 17e). 'Tissue' maturation after 6 months was confirmed by the observation of astrocytic (GFAP-positive) projections arising from the grafted area and human synaptophysin (hSYN)-positive projections in close proximity with host motor neurons (choline acetyltransferase or ChAT-positive cells; Supplementary Fig. 17d). Meanwhile, RNA-seq data show that *DCX* expression rises after 1 week of *in vivo* differentiation but is virtually lost after 8 months (Fig. 5b and Supplementary Fig. 18a), while other neuronal markers remain or rise (*SYN1*, *SYP*, *ENO2*, and *RBFOX3*; Fig. 5b and Supplementary Fig. 18a; as also observed for the *MAPT*, *GAP43*, *TUBB3*, *NCAM1*, and *GRIN2C* genes, data not shown). We note that some neuronal genes slightly decline after 6/8 months, likely as an indirect effect of an abrupt emergence of glial signal, as supported by astrocytic and oligodendrocytic profiles peaking after 2 and 6/8 months, respectively (*GFAP*, *S100B*, and *APOE*, and *OLIG2*, respectively; Fig. 5b). A similar effect is also observed in organoids (Supplementary Fig. 16a). We note, however, that this effect would not similarly affect neuronal genes encoding synaptogenesis regulators, because these particular genes are robustly upregulated in parallel with glial signal, as opposed to many other neuronal genes (Fig. 5b and Supplementary Figs. 16a and 18a). We also note that, in these analyses, we could not separate human from rat signal of the *PCDHGC4* and *PCDHGC5* transcripts due to their sequence similarities. In spite of all the described indications suggesting that 6/8-month grafted cells reach substantial levels of maturation *in vivo*, we still observe hiPSC/hESC-like (restricted) patterns of cPcdh expression in hESC-NPC-derived cells (Fig. 5c). This occurs despite a much broader (unrestricted) pattern of expression of α/γ-isoforms in the surrounding rat tissue (Fig. 5c and Supplementary Fig. 18b). On a side note, we note differences in relative expression of some α/γ-isoforms between NPCs and grafted cells (Fig. 5c). These differences may partially result from cell diversification, as we have described in NPC, astrocyte, and neuron comparisons (Supplementary Fig. 2c,d). Together, these analyses

suggest that the distinct patterns of frequency selection of α/γ-isoforms detected in hESC-derived neurons are stably maintained for at least 8-10 months *in vitro* and *in vivo* regardless of protocol of neuronal differentiation.

Analysis of DNA methylation in human Down syndrome brains

We leveraged a series of DNA methylation measurements obtained using Illumina HumanMethylation 450K arrays and n=75 postmortem samples of frontal cortex brains previously generated to report dysregulation of the cPcdh locus in Down syndrome [10,11] (n=16 fetal DS and n=27 fetal control samples, median gestational stage=18-pcw and 20-pcw, respectively; n=6 adult DS and n=26 adult control samples, median age=45 years and 49.5 years, respectively). Notably, a side-by-side comparison of fetal and adult control brains reveals five major differences in the cPcdh locus. First, a subset of V-type α-promoters shows highly selective hypermethylation in adult brains (or, hypomethylation in fetal brains), fitting with our model of enhanced frequencies of selection of a subset of α-isoforms if it were to translate in distinct levels of methylation (Fig. 6c, top-left panel, V-type isoforms). Second, the *PCDHGC5* promoter shows the lowest relative hypomethylation among all isoforms in adult brains (or, hypermethylation in fetal brains), fitting with a model of postnatal-specific expression (Fig. 6c, bottom-left panel, γC5). Third, the *PCDHGC4* promoter shows one of the highest relative hypermethylation among all isoforms in adult brains (or, hypomethylation in adult brains), fitting with a model of fetal-specific expression (Fig. 6c, bottom-left panel, γC4). Fourth, V-type γ-promoters (and *PCDHGC3*), in general, show hypermethylation in adult brains (or, hypomethylation in fetal brains), in agreement with the expression decay observed after the two years of age in the BrainSpan cohort (Fig. 6c, bottom-left panel, V-type isoforms). And, fifth, V-type γ-promoters located toward the 3' end of the γ-cluster show enhanced hypermethylation in adult brains (or, selective hypomethylation in fetal brains), fitting with our model of preferential expression towards the 3' end of the γ-cluster in fetal samples if low methylation were to translate into high expression (Fig. 6c, bottom-left panel, V-type isoforms). In line with our hypothesis, a side-by-side comparison of fetal control and adult DS brains reveals that adult Down syndrome brains retain the pattern of fetal-specific hypomethylation observed toward the 3' end of the γ-cluster, but none of the other four fetal features (Fig. 6c, top/bottom-right panels). Further in support, a side-by-side comparison of adult Down syndrome and control brains resembles a side-by-side comparison of fetal and adult control brains with

regard to this fetal feature, which is aberrantly retained in adult Down syndrome tissue in spite of the median age similarity, 45 versus 49.5 years (Fig. 6c, bottom-center; compare to bottom-left). Fetal Down syndrome brains appear an extreme version of fetal control brains (Supplementary Fig. 21, bottom-right panel; see additional comparisons in Supplementary Fig. 21). Previously, we confirmed a correlation between high/low levels of DNA methylation and low/high expression, respectively, in control/Down syndrome comparisons[10]. In sum, we can ascribe the previously reported alterations in DNA methylation in the cPcdh locus of Down syndrome brains[10,11] to, more specifically, an aberrant retention of a fetal-specific (restricted) pattern characterized by a relative hypermethylation toward the 5' end and hypomethylation toward the 3' end of the γ-cluster in adult Down syndrome brains that does not, interestingly, affect the other four fetal/adult differences in the cPcdh locus. In fetal Down syndrome brains, the fetal-specific (restricted) pattern would be exacerbated.

Additional points for the Discussion

Intriguingly, we have observed that the number of H3K4me3-enriched cPcdh promoters is larger in re-primed hESCs and also in particular in the α-cluster of (primed) hiPSC-T cells in the hiF-T system. We note that the hiF-T system exhibits interesting singularities. For instance, it does not show evidence of H3K27me3 along the cPcdh locus in the naive-like state[7], and it adopts a transient naive-like stage characterized by H3K4me3 deposition on the *DPPA3*, *DNMT3L*, and *MIR371-3* promoters but not on the *DPPA5* promoter. This system, therefore, may link the pre-setting of cPcdh frequencies (at least in the α-cluster) with these singularities and with the stochastic nature of the reprogramming process, which has been artificially reduced in the hiF-T system[7]. In general, our observations suggest that the pre-setting of cPcdh-frequency selections is a generic property of primed pluripotent stem cells but, importantly, molded by the particular conditions at the time of pre- to post-implantation-like conversion. Perhaps for this reason, the patterns are similar but often not identical among most hESC/hiPSC lines, as we describe. They are, however, identical between the CVB and CVI lines, which were generated in the same process of reprogramming[12,13]. We do not know whether the 18a and 18c hiPSC lines, which we show that have different H3K4me3 patterns across the cPcdh locus despite being derived from same donor, were generated in the same process of reprogramming or in different processes. The CVB and CVI lines were generated in the same reprogramming process. On the other hand, long-term culturing

may also lead to changes in cPcdh patterns in hESCs/hiPSCs. We have observed this effect when comparing bidimensional and tridimensional cultures of the CVB line, for example, which may indicate that different subpopulations with unique cPcdh subsets dominate the final cultures in each case. We suspect that the speed of the reprograming process of hESC derivation may have an impact on the final patterns and may affect differentially each cluster in the cPcdh locus. Faster reprogramming could mean less time for naive-like hiPSCs to proliferate before becoming primed cells; thus, ultimately, less diversity of cPcdh signatures. Likewise, we have observed three different cPcdh patterns in the H9 hESC line when comparing the cells from three different laboratories, which may suggest that processes of subpopulation enrichment may occur that result in different patterns. In short, we see value on profiling cPcdh patterns in hESC and hiPSC lines and sublines to monitor line/subline identity and unperceived processes of sub-enrichment.

As we show here, another recent study suggests dynamism in the stochastic selection of the cPcdh genes. In particular, that study focuses on the stochastic selection of mouse α-cPcdh isoforms and reveals dynamic rules of interdependency[14]. Based on a Poisson binomial distribution, the authors of this study compared the expected variance in the expression of 12 V-type and 2-C-type alpha isoforms to the observed variance determined by single-cell qRT-PCR in different populations of neuronal cells (the ratio of both variances is defined as 'interdependence coefficient', or IC). That analysis reveals that while 7-day differentiated, mESC-derived neurons select cPcdh genes with a high degree of independency among isoforms (IC=1.1), Purkinje cells derived from mouse brains select these genes with a high degree of exclusivity, or anti-correlation (IC=0.34), and mESC-derived NPCs and targeted-modified neurons that overexpress one isoform select these genes with a high degree of co-occurrence (IC=1.86 and IC=1.42-1.80, respectively). In combination with our study, it therefore appears that the process of stochastic selection of cPcdh isoforms is more complex, dynamic, and subject to 'rules' than initially anticipated at the level of interdependency in the stochastic selection of isoforms dependent on the level of mouse neuronal differentiation[14] and at the level of non-uniform distribution of probabilities across the locus dependent on the level of human neuronal maturation (in this study).

Supplementary Note References

1. Bai, Q. *et al.* Temporal analysis of genome alterations induced by single-cell passaging in human embryonic stem cells. *Stem Cells Dev.* **24**, 653–662 (2015).
2. Guo, Y. *et al.* CTCF/cohesin-mediated DNA looping is required for protocadherin α promoter choice. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21081–21086 (2012).
3. Kehayova, P., Monahan, K., Chen, W. & Maniatis, T. Regulatory elements required for the activation and repression of the protocadherin-alpha gene cluster. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 17195–17200 (2011).
4. Monahan, K. *et al.* Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of Protocadherin-α gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 9125–9130 (2012).
5. Ribich, S., Tasic, B. & Maniatis, T. Identification of long-range regulatory elements in the protocadherin-alpha gene cluster. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19719–19724 (2006).
6. Mertens, J. *et al.* Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects. *Cell Stem Cell* (2015). doi:10.1016/j.stem.2015.09.001
7. Cacchiarelli, D. *et al.* Integrative Analyses of Human Reprogramming Reveal Dynamic Nature of Induced Pluripotency. *Cell* **162**, 412–424 (2015).
8. Busskamp, V. *et al.* Rapid neurogenesis through transcriptional activation in human stem cells. *Mol. Syst. Biol.* **10**, (2014).
9. Juhasova, J. *et al.* Time course of spinal doublecortin expression in developing rat and porcine spinal cord: implication in in vivo neural precursor grafting studies. *Cell. Mol. Neurobiol.* **35**, 57–70 (2015).
10. El Hajj, N. *et al.* Epigenetic dysregulation in the developing Down syndrome cortex. *Epigenetics* **11**, 563–578 (2016).
11. Horvath, S. *et al.* Accelerated epigenetic aging in Down syndrome. *Aging Cell* **14**, 491–495 (2015).
12. Woodruff, G. *et al.* Defective Transcytosis of APP and Lipoproteins in Human iPSC-Derived Neurons with Familial Alzheimer's Disease Mutations. *Cell Rep.* **17**, 759–773 (2016).
13. Woodruff, G. *et al.* The Presenilin-1 ΔE9 Mutation Results in Reduced γ-Secretase Activity, but Not Total Loss of PS1 Function, in Isogenic Human Stem Cells. *Cell Rep.* **5**, 974–985 (2013).
14. Wada, T., Wallerich, S. & Becskei, A. Stochastic Gene Choice during Cellular Differentiation. *Cell Rep.* **24**, 3503–3512 (2018).