Heparan sulfate Sulfatases are essential for patterning of human stem cell-derived midbrain dopaminergic neurons

Chiara Tremolanti[#1], Carmen Abaurre[1], Yu-Chun Chien[2] Carmen Saltó[1], Conor William Mitchell[2], Emma Jurmand[2], Eneritz Agirre[1], Guochang Lyu[1], Richard Karlsson[2], Gonçalo Castelo-Branco[1], Ulrika Marklund[#1], Rebecca Louise Miller[2], Jeremy E. Turnbull[#2], Ernest Arenas[1]†.

[1]Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Solnavägen 9, Biomedicum 6C, Stockholm 17177, Sweden.

[2]Copenhagen Center for Glycomics, Department of Cellular and Molecular Medicine, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, DK-2200, Copenhagen, Denmark.

[#]Correspondence: chiara.tremolanti@ki.se (C.T.); turnbull@sund.ku.dk (J.E.T); ulrika.marklund@ki.se (U.M)

† Passed away September 15th, 2024

SUPPLEMENTARY INFORMATION

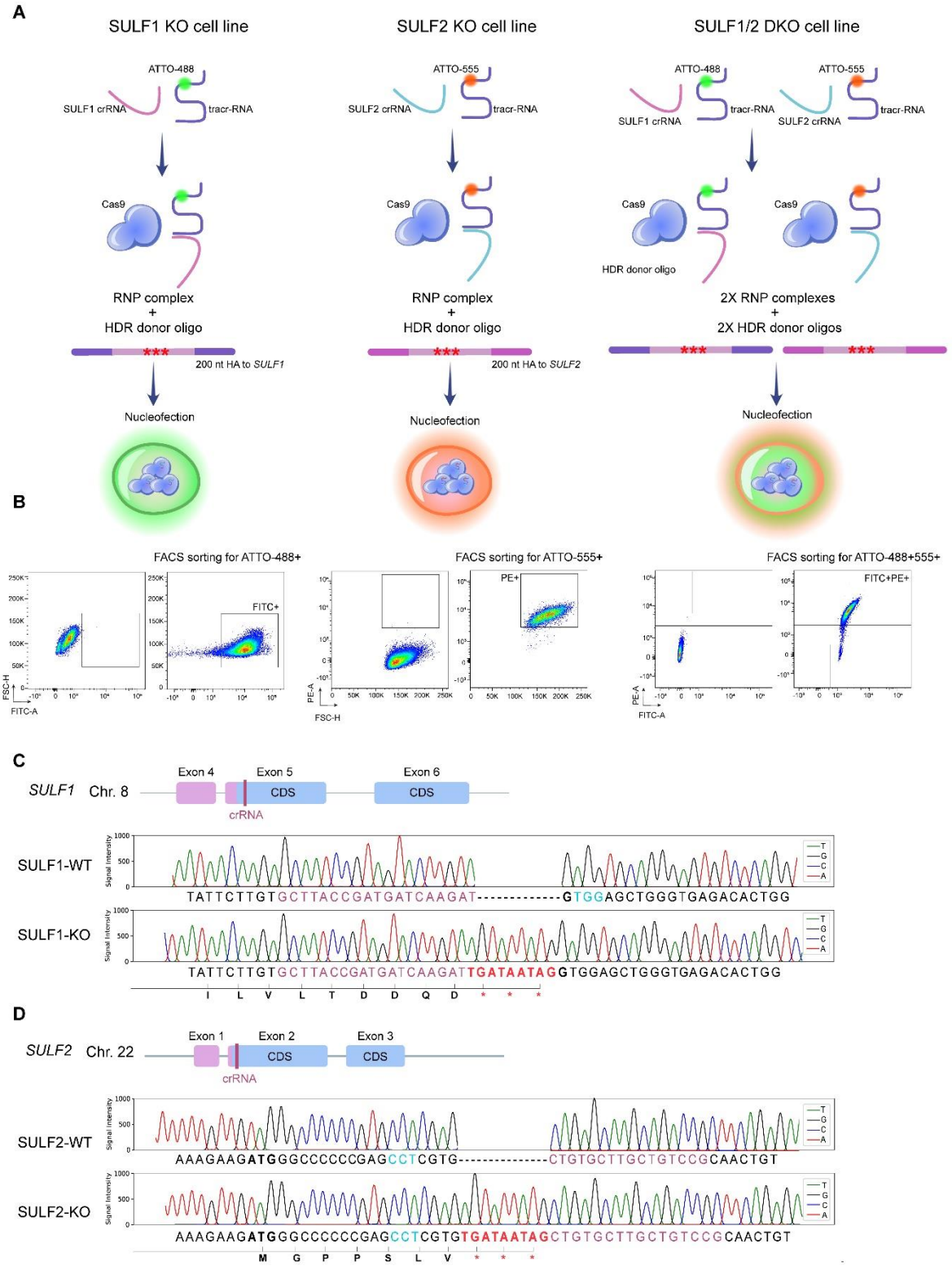


Figure S1. Generation and validation of SULF1 KO, SULF2 KO, and SULF1/2 DKO iPSC lines using a CRISPR-Cas9 triple stop codon knock-in strategy.

(**A**) Schematic representation of the CRISPR-Cas9 strategy used to generate single and double KO lines for SULF1 and SULF2. Cells were nucleofected with the ribonucleoprotein complex (RNP)

composed by: Cas9 protein, crRNAs targeting *SULF1* or *SULF2*, tracrRNA (labeled with ATTO-488 or ATTO-555 for *SULF1* and *SULF2* respectively), together with a homology-directed repair (HDR) donor oligo with 200 bp homology arms flanking the target site and bearing the 9 nucleotides knock-in (triple stop codons). For SULF1/2 double KO generation, both RNP complexes and HDR templates were co-delivered.

(**B**) FACS sorting of nucleofected cells based on fluorescence-conjugated tracr-RNAs. Left to right: gating strategy and selected sorted population for ATTO-488+ (SULF1 KO), ATTO-555+ (SULF2 KO), and ATTO-488+/ATTO-555+ (SULF1/2 DKO).

(**C**) Genomic structure of the *SULF1* locus and representative Sanger sequencing traces of WT (top) and SULF1 KO, SULF1/2 DKO clones (bottom). KO alleles show the 9 nucleotides in-frame insertion resulting in three stop codons (red asterisks).

(**D**) Genomic structure of the *SULF2* locus and representative Sanger sequencing traces of WT (top) and SULF2 KO, SULF1/2 DKO clones (bottom). KO alleles show the 9 nucleotides in-frame insertion resulting in three stop codons (red asterisks).

crRNA: CRISPR-RNA (binding the target gene); tracr-RNA: trans-activating CRISPR RNA (binding the Cas9); HA: homology arm; HDR: homology-directed recombination; CDS: coding sequence.

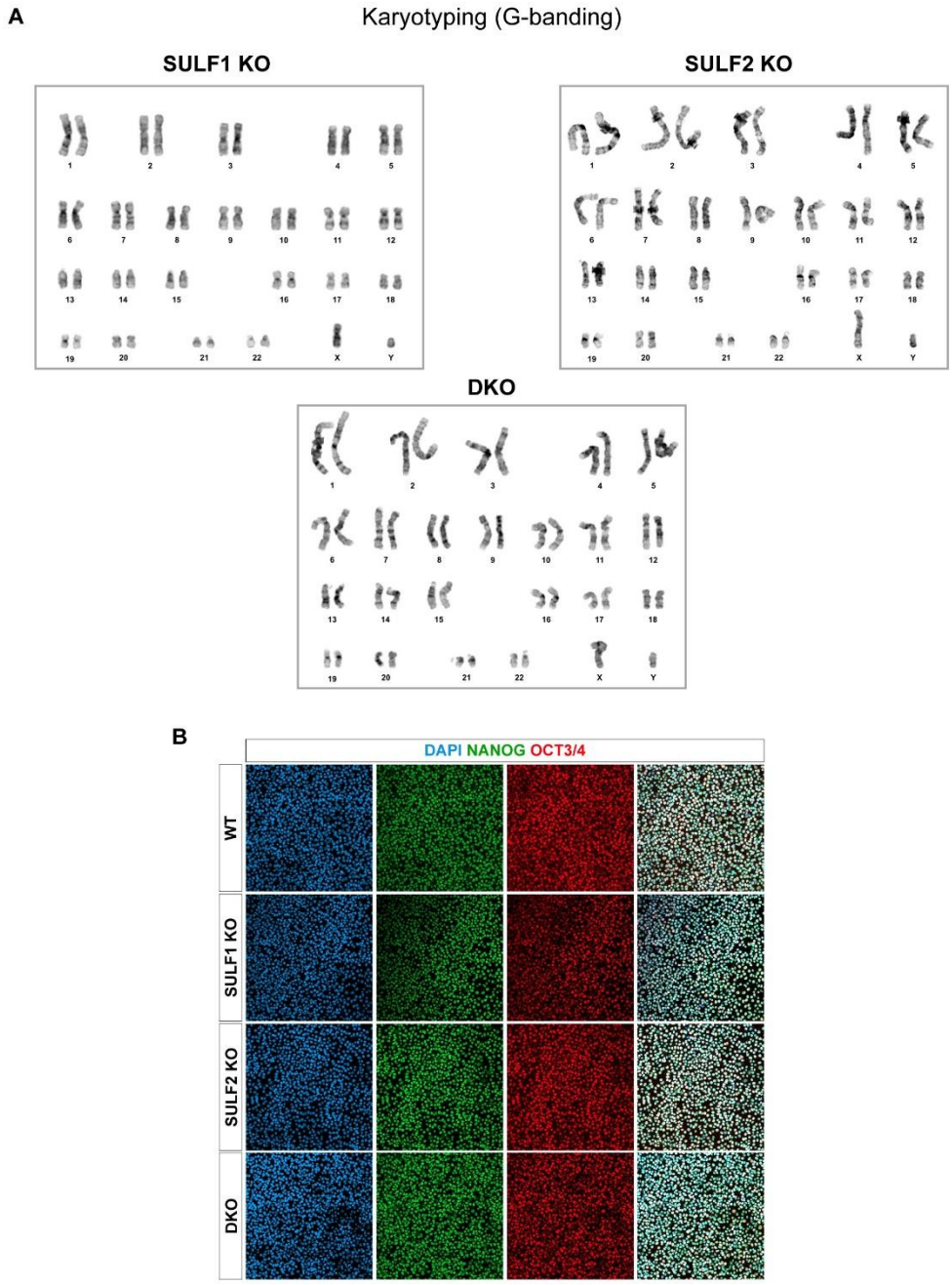


Figure S2. Karyotyping and pluripotency assessment of SULF1/2 KO iPSC lines.

(**A**) Representative G-banding karyotyping images of SULF1 KO, SULF2 KO, and SULF1/2 DKO iPSC lines.

(**B**) Immunostaining of pluripotency markers in WT, SULF1 KO, SULF2 KO, and SULF1/2 DKO iPSCs. Cells were stained for OCT3/4 (red), NANOG (green), and DAPI for the nuclei (blue).

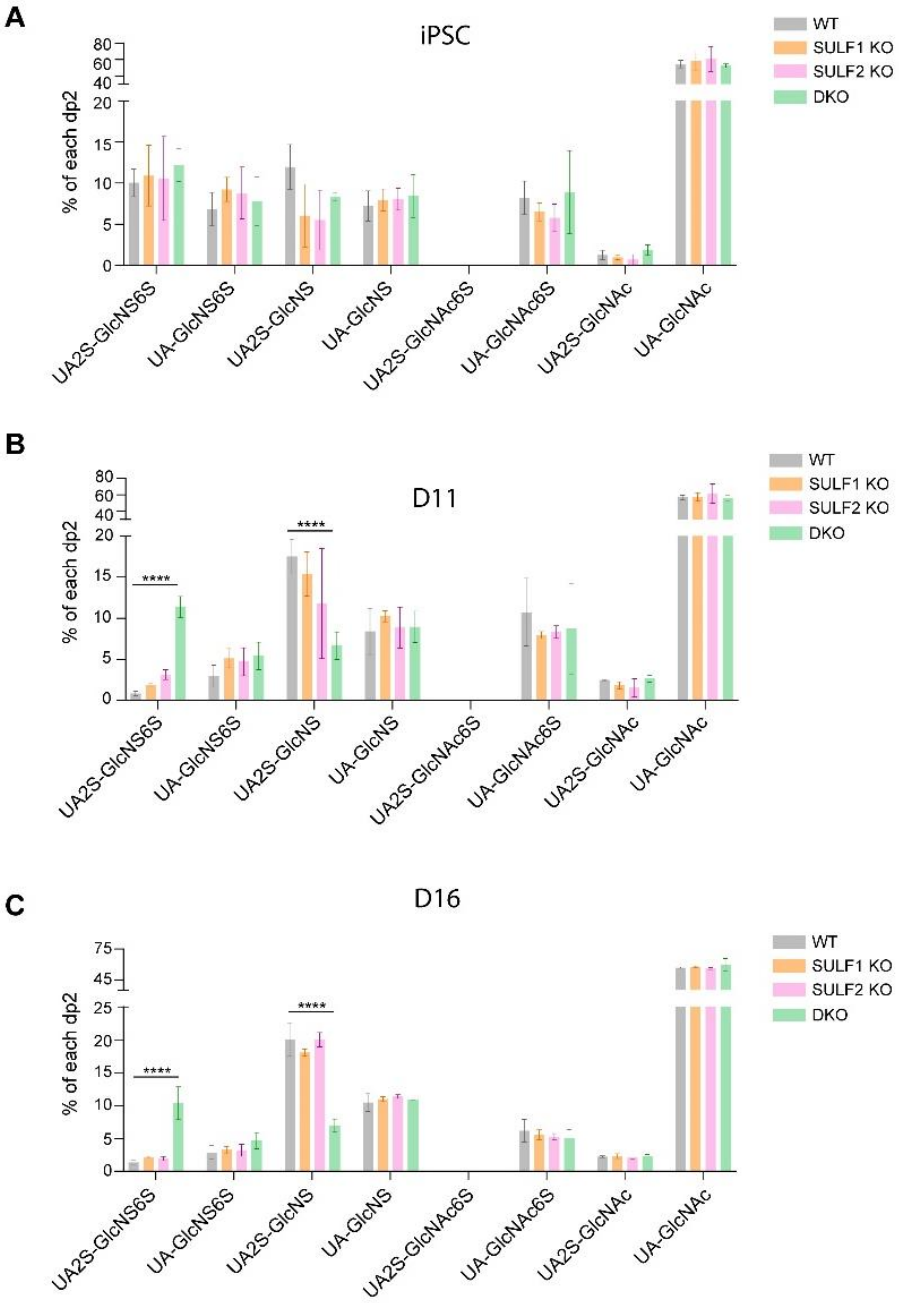


Figure S3. Analysis of disaccharide composition in WT, SULF1 KO, SULF2 KO, and SULF1/2 DKO cells during differentiation.

(**A-C**) Quantification of individual disaccharide species (dp2) at iPSC, D11, and D16 differentiating cultures from WT, SULF1 KO, SULF2 KO, and SULF1/2 DKO lines. Each bar represents the relative abundance of a specific disaccharide normalized to the total dp2 signal. Data are presented as mean ± SEM; n=3 independent experiments; two-way anova followed by Tukey's multiple comparison test, p value: ****$p < 0.0001$.

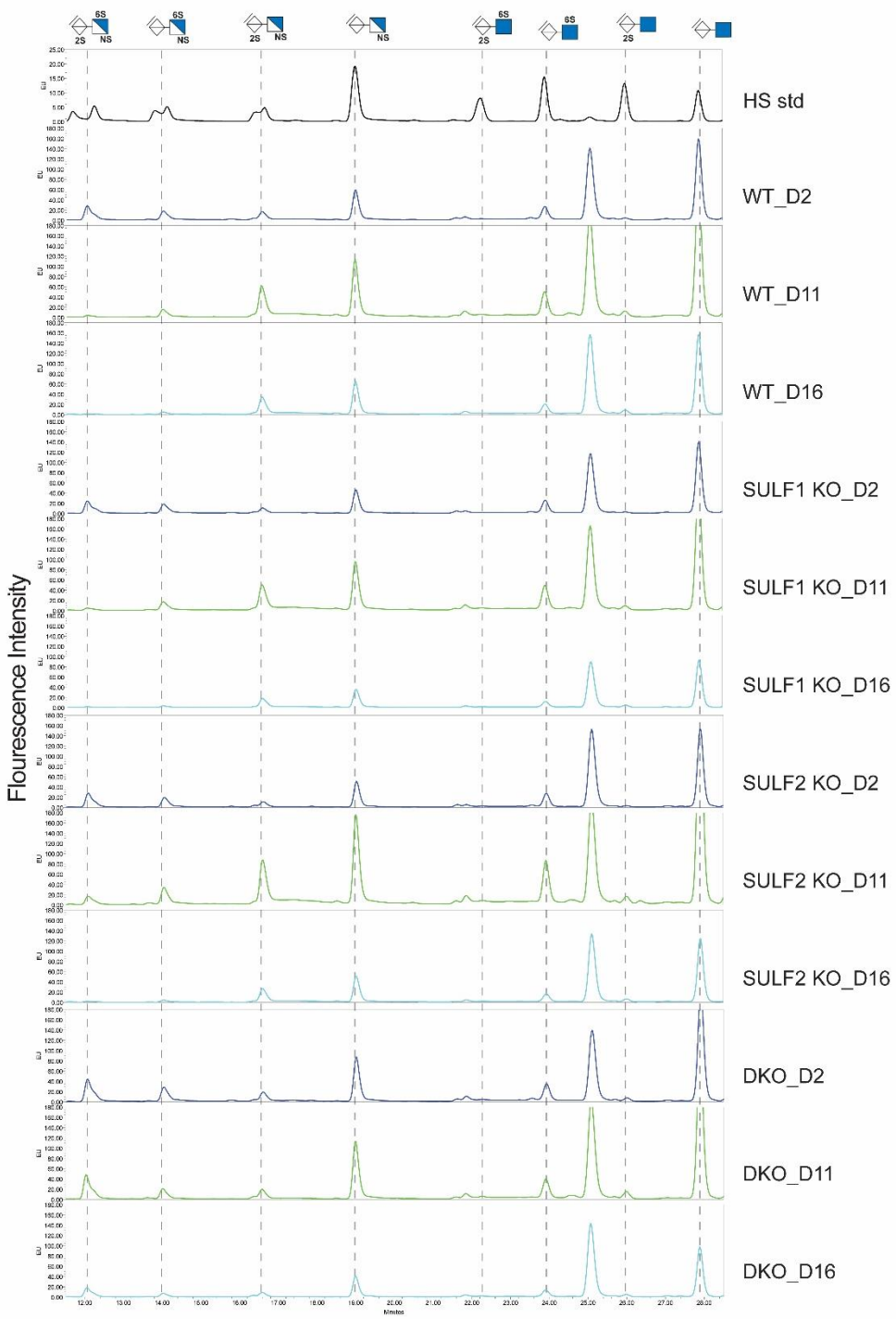


Figure S4. Representative UPLC profiles of heparan sulfate disaccharides from WT, SULF1 KO, SULF2 KO, and SULF1/2 DKO cells during differentiation.

Heparan sulfate (HS) was purified from at least three replicate batches of cells, and one representative result is shown. The purified HS was digested with heparinase, labeled with AMAC, and analyzed by ultra-performance liquid chromatography (UPLC). The peak areas of the disaccharides were further quantified and normalized to 20 pmol of each commercial HS standard. Disaccharide structures are indicated above the chromatograms.

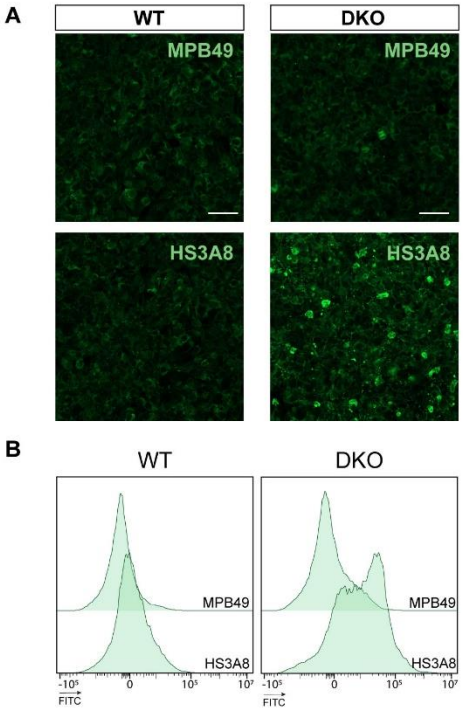


Figure S5. Assessment of binding specificity of HS3A8 scFv.

(**A**) Immunofluorescence images showing staining of WT and SULF1/2 DKO cultures with MPB49 and HS3A8 at D11 of differentiation. Scale bars: 100 µm.

(**B**) Representative flow cytometry histograms of WT and SULF1/2 DKO cells stained with HS3A8 or MPB49 (negative control scFv with no HS binding).

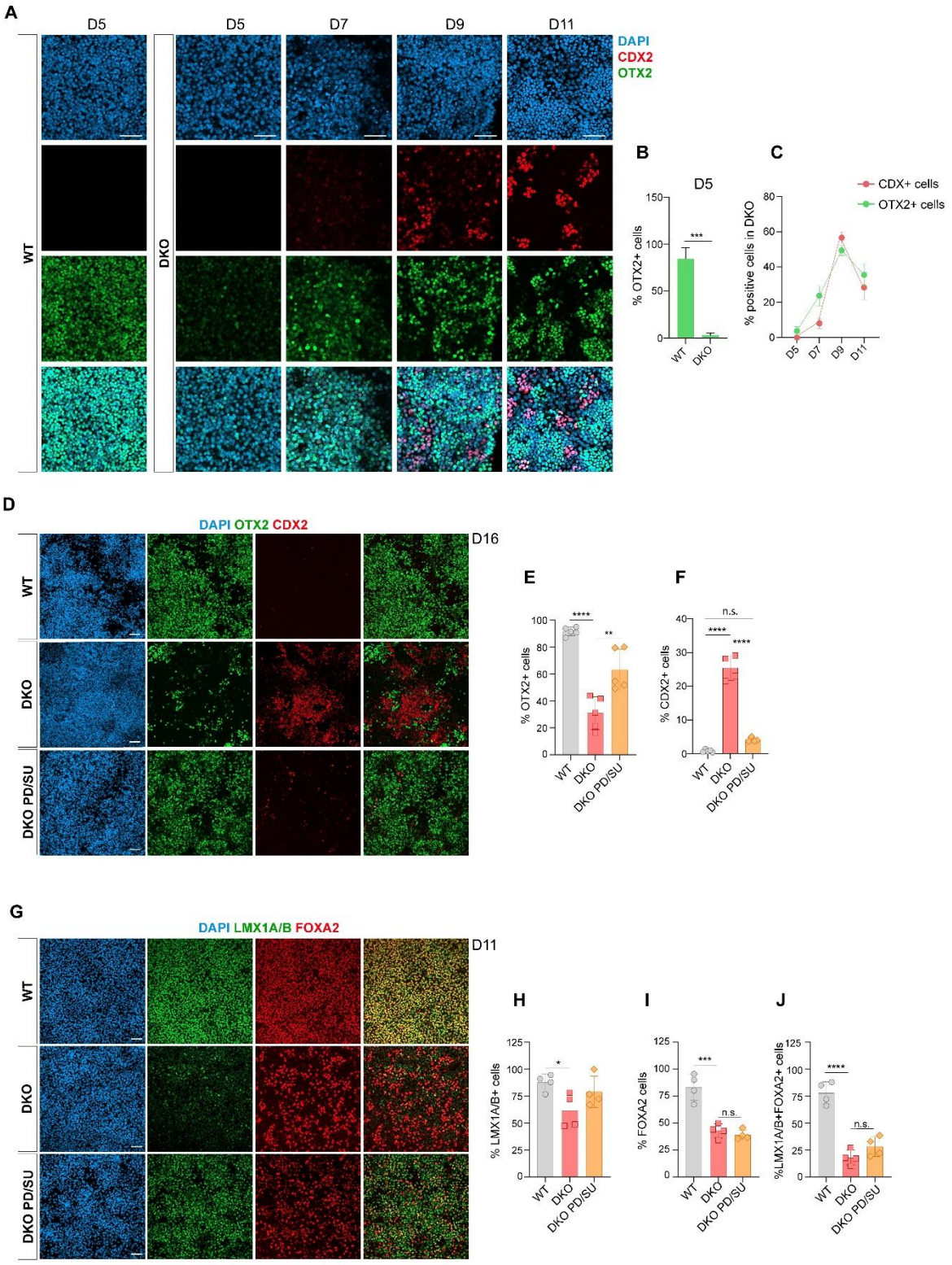


Figure S6. Temporal dynamics of CDX2+ cells emergence and CDX2/OTX2 expression pattern in PD/SU-treated DKO cells.

(**A**) Representative immunofluorescence images showing the temporal expression of OTX2 (green) and CDX2 (red) in WT and SULF1/2 KO cells from D5 to D11 of differentiation. DAPI (blue) stains nuclei.

(**B**) Quantification of OTX2+ cells at D5 in WT and SULF1/2 DKO cells. Data are presented as mean ± SEM; n=3 independent experiments; unpaired t-test; p value: ***p < 0.001.
(**C**) Quantification of CDX2+ and OTX2+ cells as a percentage of total cells at indicated time points (D5, D7, D9, D11) in SULF1/2 DKO cultures. Data represent mean ± SEM, n=3 independent differentiation experiments.
(**D**) Representative immunofluorescence images of OTX2+ (green) and CDX2+ (red) cells in WT, untreated SULF1/2 DKO, and PD/SU-treated SULF1/2 DKO cells. DAPI (blue) stains nuclei.
(**E, F**) Quantification of OTX2+ and CDX2+ cells at D16 across conditions. Data are presented as mean ± SEM; n=3 independent experiments; one-way anova followed by Tukey's multiple comparison test; p value: **p < 0.01, ****p < 0.0001; n.s., not significant. All scale bars: 100 μm.
(**G**) Representative immunofluorescence images of LMX1+ (green) and FOXA2+ (red) cells in WT, untreated SULF1/2 DKO, and PD/SU-treated SULF1/2 DKO cells. DAPI (blue) stains nuclei.
(**H, I, J**) Quantification of LMX1+, FOXA2+, and LMX1+FOXA2 cells at D11 across conditions. Data are presented as mean ± SEM; n=3 independent experiments; one-way anova followed by Tukey's multiple comparison; p value: ***p < 0.001, ****p < 0.0001; n.s., not significant. All scale bars: 100 μm.

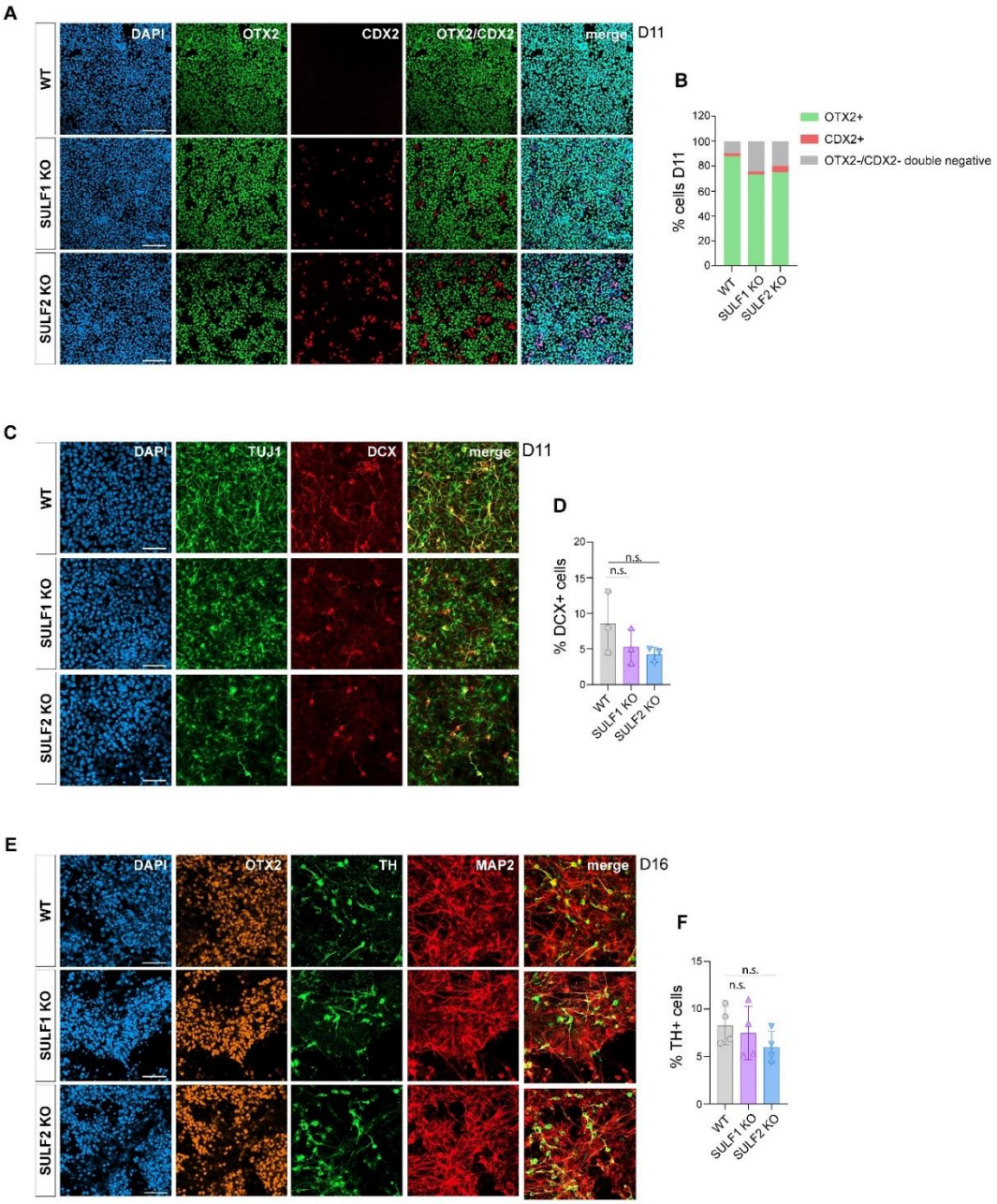


Figure S7. SULF1 or SULF2 single KO does not impair mDA neuronal differentiation.

(**A**) Representative immunofluorescence images at D11 showing OTX2 (green) and CDX2 (red) expression in WT, SULF1 KO, and SULF2 cultures. DAPI (blue) stains nuclei.

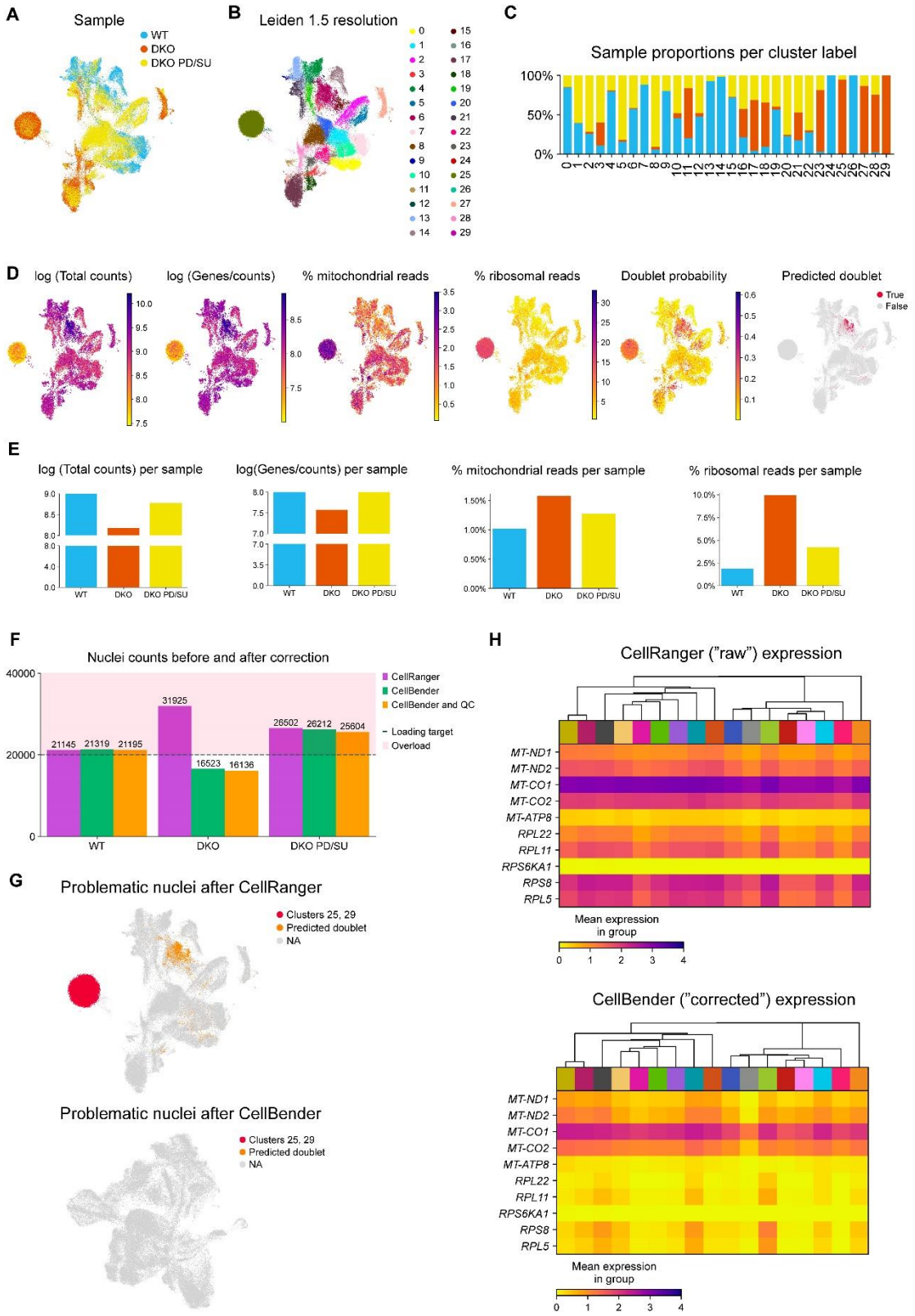
(**B**) Quantification of OTX2+ and CDX2+ cells at D11 across the three conditions. Data are presented as mean ± SEM; n=4 independent experiments.

(**C**) Representative immunofluorescence images at D11 for TUJ1 (green) and DCX (red) in WT, SULF1 KO, and SULF2 KO cultures. DAPI (blue) stains nuclei.

(**D**) Quantification of DCX+ cells at D11. Data are presented as mean ± SEM; n=3 independent experiments; one-way anova followed by Tukey's multiple comparison test; n.s.: not significant.

(**E**) Representative immunofluorescence images at D16 for OTX2 (orange), TH (green), and MAP2 (red) in WT, SULF1 KO, and SULF2 KO cultures. DAPI (blue) marks nuclei.

(**F**) Quantification of TH+ cells at D16. Data are presented as mean ± SEM; n=3 independent experiments; one-way anova followed by Tukey's multiple comparison tes; n.s.: not significant. All scale bars: 100 µm.



(**G**) Quantification of TH+ cells at D30. Data are presented as mean ± SEM; n=3 independent experiments; one-way anova followed by Tukey's multiple comparison tes; n.s.: not significant. All scale bars: 100 µm.

Figure S8. Nuclei calling and ambient RNA correction with CellRanger and CellBender.

(**A**) UMAP embedding visualization of sample of origin for each nucleus (WT: WT, DKO: SULF1/2 DKO, DKO PD/SU: PD/SU-treated SULF1/2 DKO) before CellBender nuclei calling and ambient RNA correction.

(**B**) UMAP embedding of clusters obtained with the Leiden algorithm at 1.5 resolution for the dataset before CellBender nuclei calling and ambient RNA correction.

(**C**) Bar plot representation of sample condition proportions per cluster label, using clustering results as in B, for the dataset before CellBender. The dataset was randomly downsampled to 5,000 nuclei per sample before plot generation. Color legend as in A.

(**D**) UMAP embedding of commonly used quality control metrics and Scrublet doublet detection output, applied before CellBender. The dataset was randomly downsampled to 5,000 nuclei per sample before plot generation.

(**E**) Bar plot representation of quality control metrics per sample of origin, calculated after random downsampling to 5,000 nuclei per sample condition and before CellBender correction.

(**F**) Bar plot representation of nuclei counts after CellRanger nuclei calling (first bar, purple), CellBender nuclei calling and ambient RNA correction (second bar, green), and CellBender and manual quality control (third bar, yellow) per sample condition. Nuclei counts are shown on top of each corresponding bar. Dashed line indicates the 20,000 nuclei target for loading the 10X Genomics chip G during step 1 of library preparation with the Chromium GEM-X Single Cell 3' v4 Gene Expression kit. Transparent red area indicates overloading of chip G by excessive nuclei concentration during sample preparation, which can interfere with GEM generation by increasing pressure and could therefore result in higher doublet concentrations and nuclei fracture, producing a higher fraction of low-quality GEMs and ambient RNA counts.

(**G**) UMAP embedding of two different categories of nuclei detected to be problematic after CellRanger nuclei calling and clustering pipeline. In red, clusters 24, 26, and 27 were detected to contain low-quality nuclei characterized by low transcript counts and high proportions of ribosomal reads. In orange, nuclei predicted to be doublets by the Scrublet algorithm are labelled. At the top, UMAP embedding represents the dataset obtained from the CellRanger pipeline. At the top, UMAP embedding shows the dataset after CellBender correction, which successfully removes problematic nuclei.

(**H**) Heatmap plots of mean gene expression of a selection of mitochondrial and ribosomal genes in each cell type obtained in the annotated dataset (after CellBender correction, further quality control, clustering, and manual annotation). Cell types are color labelled as in main figure 6C, and the dendrogram represents hierarchical clustering results. At the top, expression values are obtained from the raw expression matrix from CellRanger. At the bottom, expression values are CellBender-corrected.

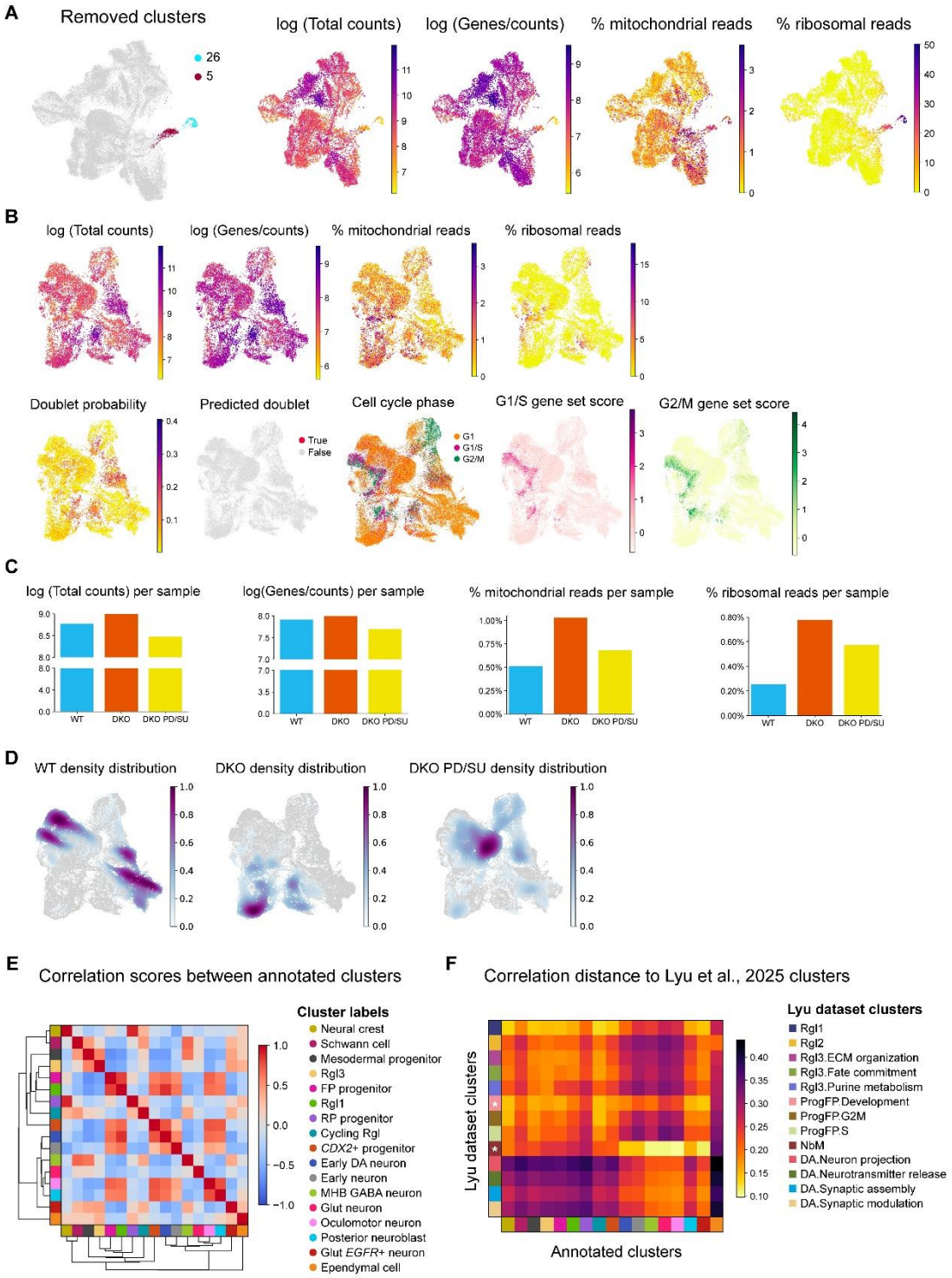


Figure S9. Quality control and characterization of CellBender-corrected dataset.

(**A**) To the left, UMAP embedding visualization of low-quality clusters that were manually removed. These clustering results were obtained with the Leiden algorithm at 1.5 resolution after CellBender nuclei calling and ambient RNA correction. To the right, UMAP embedding visualization of commonly used quality control metrics in a randomly downsampled subset of the dataset (5,000 nuclei per sample condition).

(**B**) UMAP embedding visualization of commonly used quality control metrics, Scrublet doublet detection output and cell cycle phase detection applied after CellBender correction and manual removal of the problematic nuclei shown in A. The dataset was randomly downsampled to 5,000 nuclei per sample before plot generation, with the exception of the predicted doublet and cell cycle phase plots.
(**C**) Bar plot representation of quality control metrics per sample of origin, calculated after random downsampling to 5,000 nuclei per sample condition and after CellBender correction and problematic cluster manual removal.
(**D**) Gaussian kernel density estimate plots per sample of origin in the UMAP embedding.
(**E**) Heatmap plot of correlation scores between the final annotated clusters in our dataset.
(**F**) Heatmap plot of cosine similarity scores between the Lyu and colleagues midbrain differentiation dataset and final annotated clusters. Cluster labels legend for annotated cell types as in E. Asterisks in white indicate manually selected clusters with low correlation distances (generally <0.25, higher similarity) to our annotated clusters.

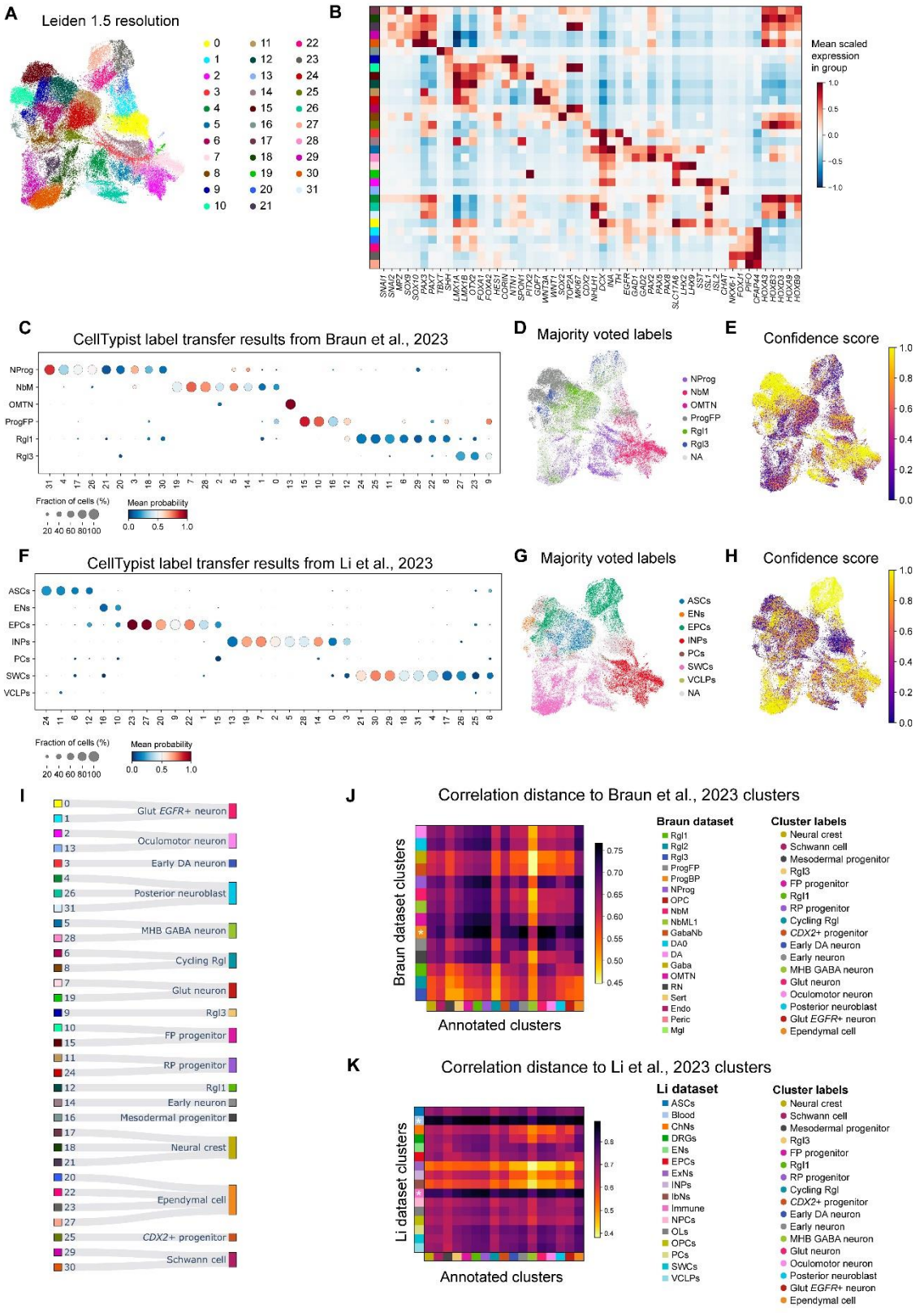


Figure S10. CellTypist label transfer from midbrain and spinal cord datasets to facilitate cluster annotation and merging from Leiden 1.5 resolution.

(**A**) UMAP embedding of clusters obtained with the Leiden algorithm at 1.5 resolution for the CellBender-corrected dataset.
(**B**) Heatmap plot of mean scaled gene expression of marker genes in each cluster. Clusters have been manually ordered to match the dendrogram results of the annotated datasets in main figure 6.
(**C**) Dot plot representation of CellTypist label transfer results from the *TH+ FOXA1+ FOXA2+* midbrain annotated subset from the Braun and colleagues dataset. Cluster numbers correspond to A and B.
(**D**) UMAP embedding of majority voting labelling results from a CellTypist model trained in the midbrain *TH+ FOXA1+ FOXA2+* annotated subset from the Braun and colleagues dataset. Unlabelled nuclei correspond to manually removed annotations due to a confidence score result of under 0.6.
(**E**) UMAP embedding of confidence score results from the CellTypist label transfering from the *TH+ FOXA1+ FOXA2+* dataset.
(**F**) Dot plot representation of CellTypist label transfer results from the human developmental spinal cord dataset from Li and colleagues after downsampling. Cluster numbers correspond to A and B.
(**G**) UMAP embedding of majority voting labelling results from a CellTypist model trained in the downsampled Li and colleagues dataset. Unlabelled nuclei correspond to manually removed annotations due to a confidence score result of under 0.6.
(**H**) UMAP embedding of confidence score results from the CellTypist label transfering from the downsampled Li and colleagues dataset.
(**I**) Sankey plot representation of the Leiden 1.5 resolution clusters that were merged to obtain the final annotated clusters shown in main figure 6.
(**J**) Heatmap plot of correlation distance scores between the midbrain *TH+ FOXA1+ FOXA2+* annotated subset from Braun and colleagues and final annotated clusters after merging. Asterisks in white indicate manually selected clusters with high correlation distances (generally >0.7, lower similarity) to our annotated clusters.
(**K**) Heatmap plot of correlation distance scores between the spinal cord Li and colleagues dataset and final annotated clusters after merging. Asterisks indicate manually selected clusters with high correlation distances (generally >0.8, lower similarity) to our annotated clusters.

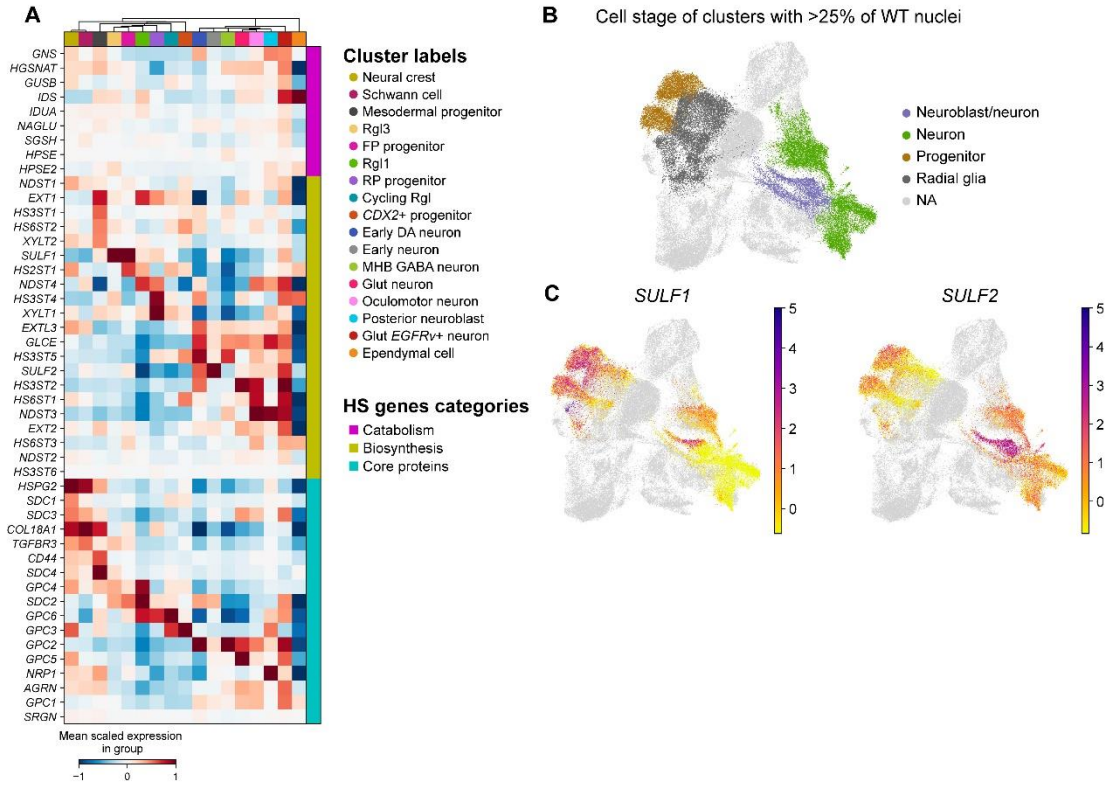


Figure S11. HS-related gene expression across clusters by snRNA-seq analysis at D16.

(**A**) Heatmap plot of mean scaled gene expression of heparan-sulfate related genes in each cell type, including dendrogram result from hierarchical clustering, which orders clusters as in C and E. Gene sets have been separated into three different categories: catabolism, biosynthesis, and core proteins. Within each category, genes were manually ordered to follow the clusters' expression patterns.

(**B**) UMAP embedding visualization of cell stage, represented in clusters which were composed of at least 25% of nuclei of WT sample origin.

(**C**) UMAP embedding visualization of *SULF1* and *SULF2* gene expression, limited to WT nuclei from the clusters shown in B.

Gene	gRNA + PAM
SULF1	GCTTACCGATGATCAAGATG TGG
SULF2	CGGACAGCAAGCACAGCACG AGG

Supplementary Table 1. CRISPR/Cas9 gRNAs used to generate *SULF1* and *SULF2* KO cell lines.

Gene	Forward primer	Reverse primer
GAPDH	TTGAGGTCAATGAAGGGGTC	GAAGGTGAAGGTCGGAGTCA
LMX1A	GATCCCTTCCGACAGGGTCTC	GGTTTCCCACTCTGGACTGC
FOXA2	TTCAGGCCCGGCTAACTCT	AGTCTCGACCCCCACTTGCT
EN1	CGTGGCTTACTCCCCATTTA	TCTCGCTGTCTCTCCCTCTC
NGN2	GCTGGGTCTGGTACACGATT	GGCCTTCAGTCTACGGGTCT
HOXA3	GTCAAACCCCTGTCAGAGTG	GCATTATAAGCGAACCCGTTG
HOXB2	TTTCACCAGTACGCTCTGTG	TTTTCCAGTAGACGGCCAAG
HOXB4	CTGGATGCGCAAAGTTCAC	TTCCTTCTCCAGCTCCAAGA
HOXB5	TCCGCAAATATTCCCCTGGA	AGGTAGCGGTTGAAGTGGAA
HOXD4	ACCCCTGGATGAAGAAGGTG	CAGTTCTAGGACTTGCTGCC
HOXC6	ATTTACCCCTGGATGCAGCG	TCCTTCTCCAGTTCCAGGGT
HOXB9	AAGCGAGGACAAAGAGAGGC	GGAGTCTGGCCACTTCGTG
FGF8	CCTACCAACTCTACAGCCGC	TCTGCTTCCAAAGGTGTCCG
CDX2	AGTCGCTACATCACCATCCG	TTCCTCTCCTTTGCTCTGCG
WNT5A	ACTGCAAGTTCCACTGGTGCTG	GTGGCACCCACTACTTGCACAC
WNT1	GAGCCACGAGTTTGGATGTT	TGCAGGGAGAAAGGAGAGAA
DKK1	GGGCGGGAATAAGTACCAGA	CAGGCGAGACAGATTTGCAC
AXIN2	TAAAAATAAGCAGCCGTTCG	AGCAATCGGCTTGGTCTCT

Supplementary Table 2. qPCR primers.