nature portfolio

https://doi.org/10.1038/s41588-021-00801-6

Supplementary information

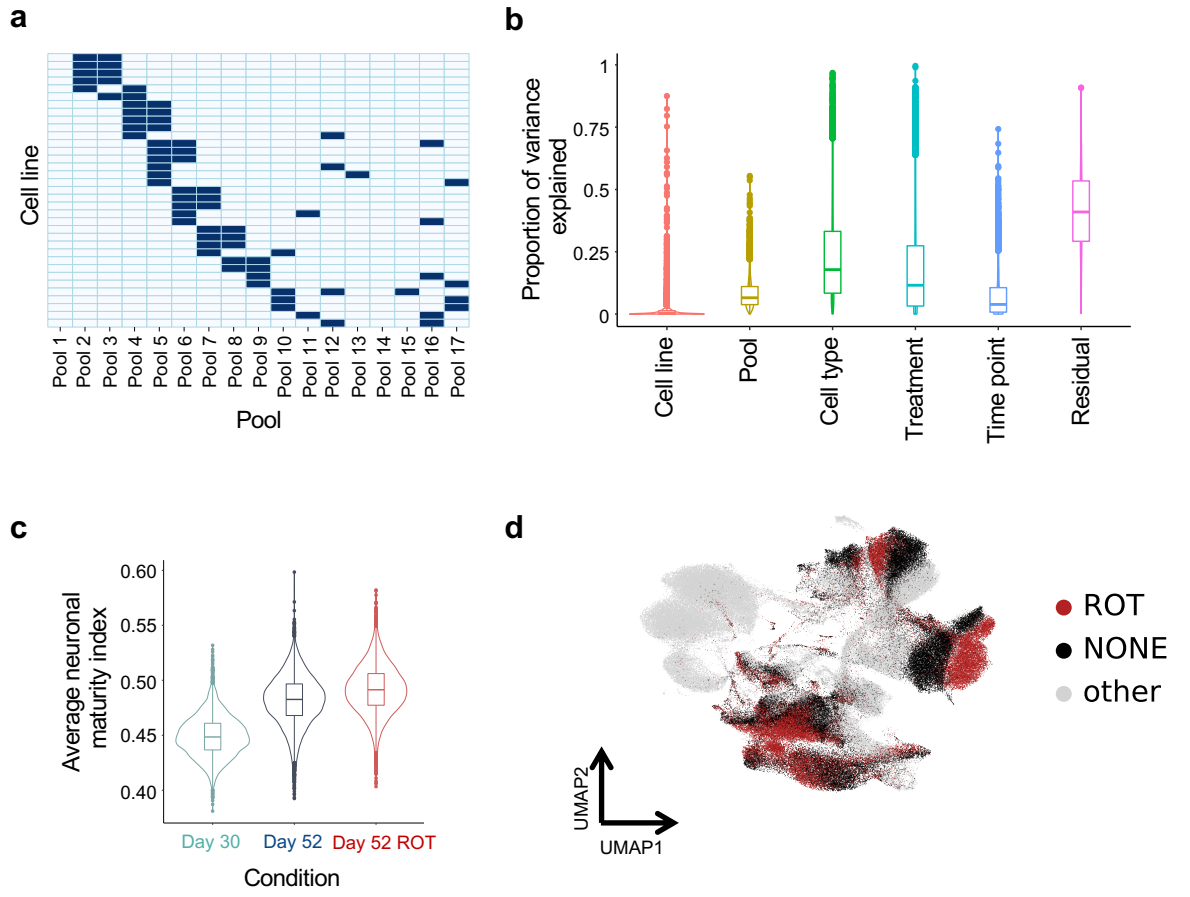
Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation

In the format provided by the authors and unedited

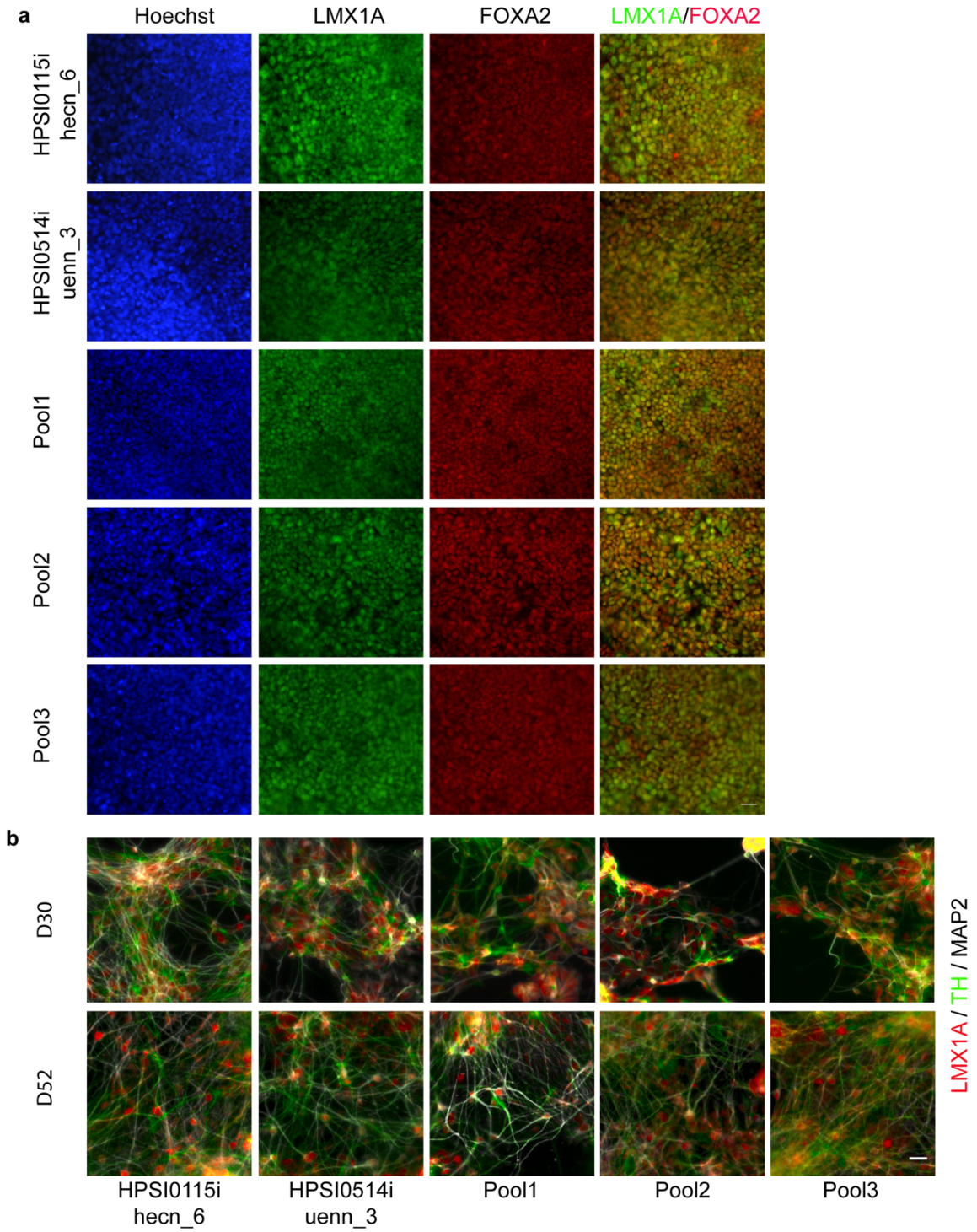
Supplementary Information for:

Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation

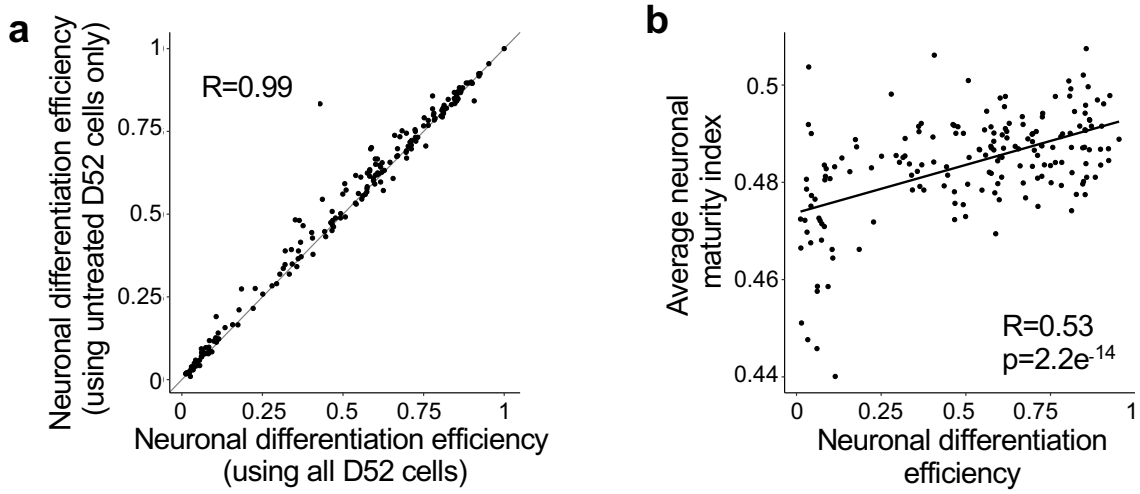
Jerber *et al.*



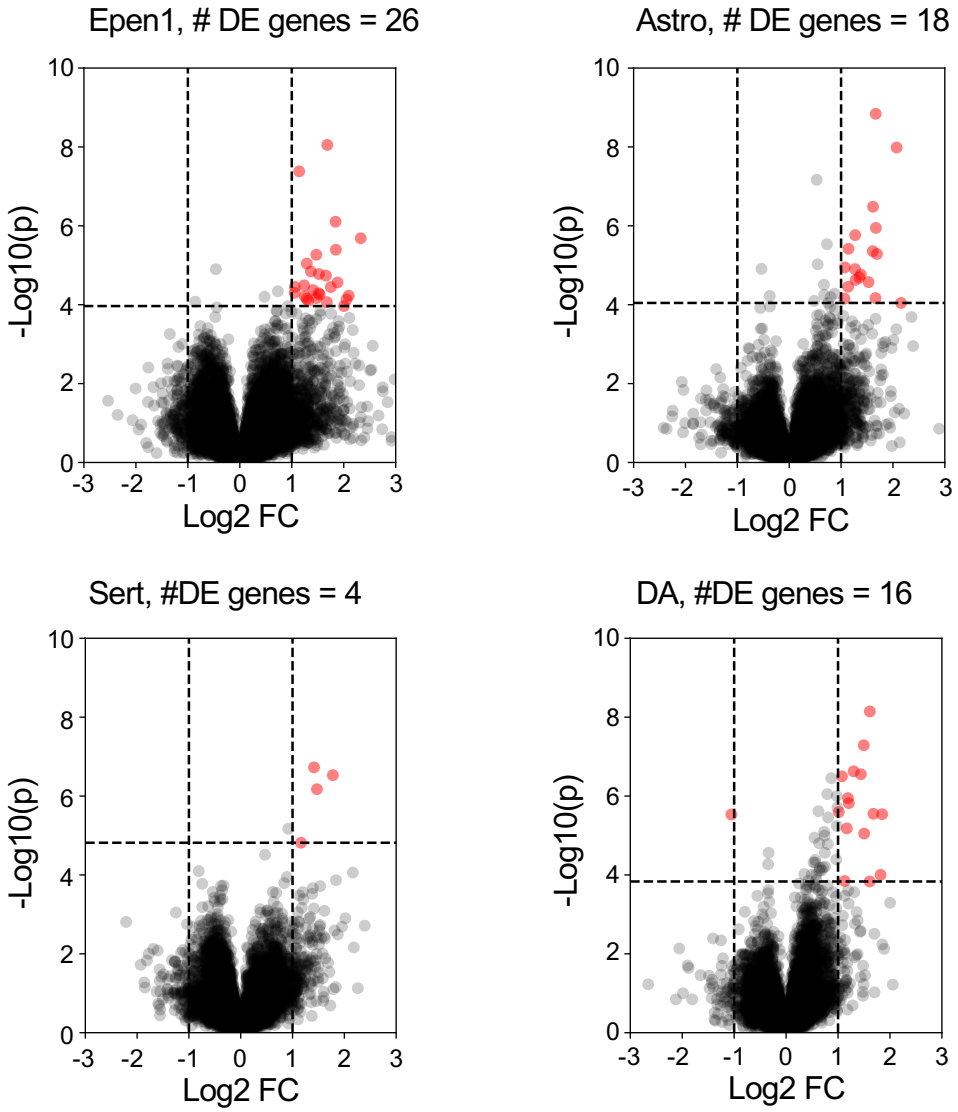
Supplementary Figure 1. Overview of experimental batches, variance components and cell line properties. (a) A subset of cell lines (n=35) are differentiated in multiple pools. Rows correspond to cell lines, columns to pools. Colour indicates that a given line (row) is contained in a pool (column). (b) Variance component analysis of gene expression for 5,899 genes with expression aggregated by donor, pool, celltype, treatment, time_point. All variables are modelled as random effect components, estimated using lme4 implemented in R. In the box plots, the middle line is the median and the lower and upper edges of the box denote the first and third quartiles. (c) Estimation of neuron maturation (using https://github.com/maplesword/neuMatIdx from He & Yu, 2018; Methods) for a subsample of 10,000 DA and Sert cells at each of day 30, day 52 untreated and day 52 rotenone-treated. Shown is the distribution of dNMI (discriminating modules only, as described in the original paper) for DA and Sert cells. In the box plots, the middle line is the median and the lower and upper edges of the box denote the first and third quartiles, while the violin plots show the distribution. (d) UMAP (as in Fig. 1) coloured by rotenone treatment. Day 52 rotenone treated cells are shown in red, day 52 untreated cells in black; grey denotes cells from day 11 and day 30.



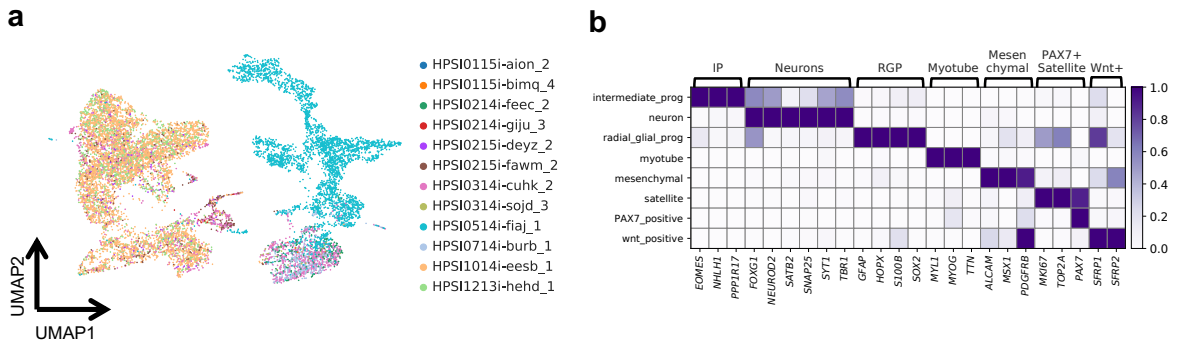
Supplementary Figure 2. Immunostaining of midbrain neural progenitors and dopaminergic neurons. (a) Immunostaining for known midbrain progenitor markers LMX1A and FOXA2 at day 11. Nuclei were counterstained with Hoechst. Scale bar: 25µm. (b) Immunostaining of differentiated dopaminergic neurons for the neuronal marker MAPT2 (white) and the dopaminergic neuronal markers TH and LMX1A. Scale bar: 25µm. Data is shown for two example individual cell lines (HPSI0155i-hecn_6 and HPSI0514i-uenn_3) as well as three entire differentiation pools (Pools 1,2,3).



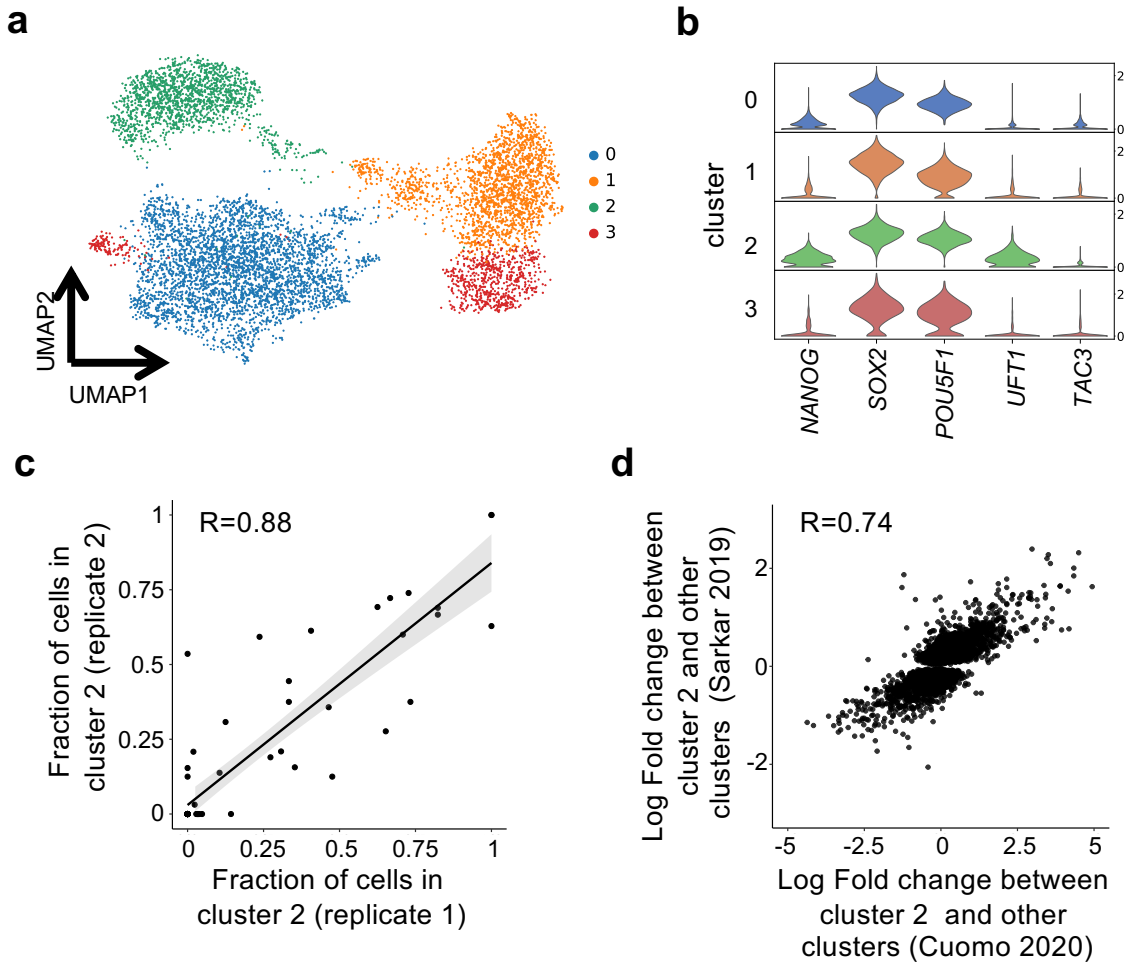
Supplementary Figure 3. Supplementary results on neuronal differentiation efficiency. (a) Scatter plot of neuronal differentiation efficiency, either considering both rotenone-treated and untreated DA and Sert cells (x-axis) versus considering only untreated DA and Sert cells (y-axis). These alternative approaches for estimating neuronal differentiation efficiency yield consent results (R=0.99, $p=2.2x10^{-16}$, two-sided t-test). (b) Scatter plot between neuronal differentiation efficiency (x-axis) versus average maturation index (y-axis) across 215 cell lines. Maturation index is estimated using https://github.com/maplesword/neuMatIdx from He & Yu, 2018, as in Supplementary Fig. 8c (R=0.53, $p=2.2x10^{-14}$, two-sided t-test).



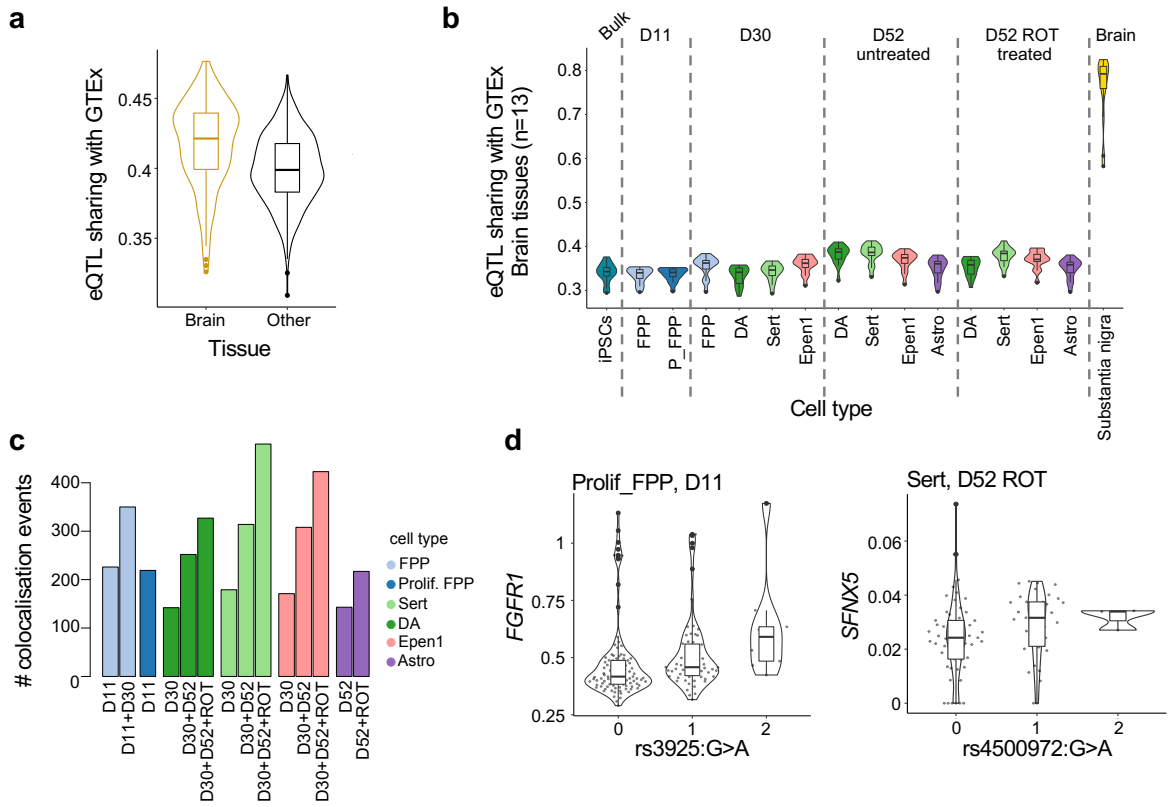
Supplementary Figure 4: Differential expression analysis, comparing the same lines in pooled versus individual differentiations. Shown are scatter plots between log fold changes (x-axis) and negative log p-values (y-axis) for differential expression analysis between individual differentiations and pooled differentiations. Shown are panels for each of the four major cell types present at day 52. Differential expression analysis was performed using a two-tailed t-test on averaged expression values, considering only cell lines for which data from both pooled and individual differentiations were available (6 cell lines, n=10 and 6 for pooled and individual differentiations, respectively). Significantly differentially expressed genes (red points) were defined at a FDR of 5% (Benjamini-Hochberg correction) with |log2FC|>1. Note that the individual and pooled differentiation experiments were conducted at different times, and therefore may have been expected to display differences in expression due to batch effects. However, only a small number of differentially expressed genes were identified between pooled and individual differentiations, indicating that pooling has no major effect. Legend: Astro: Astrocyte-like; DA: Dopaminergic neurons, Epen1: Ependymal-like1; Sert: Serotonergic-like neurons.



Supplementary Figure 5. Cerebral organoid analysis. (a) UMAP of cerebral organoid cells contained in one pool, as assayed by scRNA-seq after 113 days of differentiation, coloured by cell line of origin. (b) Heatmap of average expression profiles for canonical marker genes across identified cell types. Legend: IP: Intermediate progenitors, RGP: Intermediate glial progenitors, PAX7+ Satellite: Satellite cells PAX7+, Mesenchymal: Mesenchymal cells, Wnt+: Wnt_positive cells.



Supplementary Figure 6. Analysis of a single-cell RNA-seq profiles iPSC from Sarkar et al. (a) UMAP overview of the scRNA-seq data from (Sarkar et al., 2019), processed using the same workflow as employed in this study (c.f. Supp. Fig. 4), which identified 4 clusters. (b) Violin plots of gene expression for canonical pluripotency genes (*NANOG, SOX2, POU5F1*) and a subset of markers of the cluster 2 population (UTF1, TAC3). (c) Scatter plot displaying the fraction of cells assigned to cluster 2 in b between replicate experiments (n=59). LOESS curve and a 95% confidence interval is shown. (d) Scatter plot of expression log fold changes between cluster 2 and other clusters from Cuomo et al., 2020 (x-axis) versus analogous log fold change estimates between cluster 2 and other clusters from Sarkar et al., 2019 (y-axis). Shown are fold changes estimates for 5,397 genes between cluster 2 and other clusters as defined using the data from Sarkar et al. (FDR<0.05). Pearson's R is indicated in panels c,d.



Supplementary Figure 7: Supplementary results on eQTL sharing with GTEx and GWAS colocalisation. (a) Box plots of estimated sharing between all 14 eQTL maps considered in this study and either 13 GTEx brain eQTL maps (left) or 35 GTEx non-brain eQTL maps (right). Sharing was estimated using MASHR (Methods). (b) Analogous panel as displayed in Fig. 4d, however including the sharing between a selected GTEx brain tissue (substantia nigra) and all remaining GTEx tissues for reference. Shown is the sharing of eQTL signals discovered in our study for different cell types and conditions with in vivo brain eQTL maps (from GTEx). Violin plots show the extent of eQTL sharing (Methods), with each of 13 GTEx brain eQTL maps. (c) Cumulative number of colocalisation events between an eQTL identified in out study and a neurological GWAS trait for each cell type. (d) Box plots corresponding to the individual colocalisation events highlighted in Fig. 5c,d. Shown are violin plots of the expression of the corresponding eQTL genes, stratified by alleles at the eQTL lead variant (donors with <50 cells are omitted). In the box plots, the middle line is the median and the lower and upper edges of the box denote the first and third quartiles, while the violin plots show the distribution. Legend: Astro: Astrocytes-like; DA: Dopaminergic neurons, Epen1: Ependymal-like1, FPP: Floor Plate Progenitors, P_FPP: Proliferating Floor Plate Progenitors, Sert: Serotonergic-like neurons. D11 = Day 11; D30 = Day 30; D52 = Day 52; ROT = Rotenone stimulation.

Overview of Supplementary Tables:

Supplementary Table 1: Overview of the collected samples.

Supplementary Table 2: Cell distribution per cell line, pool, cell type, time point and stimulation.

Supplementary Table 3: Associations between cell lines' factors and differentiation outcomes.

Supplementary Table 4: List of genes correlated with differentiation efficiency.

Supplementary Table 5: Predicted differentiation efficiency scores for all HipSci lines.

Supplementary Table 6: DE genes between iPSC clusters at FDR<5%.

Supplementary Table 7: List of eQTL discoveries across cell types and conditions at FDR<5%.

Supplementary Table 8: List of neurological traits used for colocalisation analysis.

Supplementary Table 9: Results of the colocalisation analysis.

Supplementary Table 10: Overview of the HipSci lines used in this study.

Supplementary Table11: List of the top 20 GO terms enrichment from iPSC cluster 2.

Legend for Supplementary Tables:

All tables are supplied as external data files. For clarity, when fields are not readily interpretable we provide a small table to define them.

Supplementary Table 1: Overview of the collected samples. All collected sample information including number of lines and replicated lines, single cell technical replicates and number of cells available per pool at each time point analyzed are provided.

Supplementary Table 2: Cell distribution per cell line, pool, cell type, time point and stimulation. Numbers of cells used in this study for each cell line per cell type, time point and stimulation. Cell type: Astro: Astrocytes-like; DA: Midbrain dopaminergic neurons, Epen1: Ependymal-like1, Epen2: Ependymal-like2 FPP: Floor Plate Progenitors, NB: Neuroblasts, P_FPP: Proliferating Floor Plate Progenitors, Sert: Serotonergic-like neurons, U_Neur1: Unknown_neurons1, U_Neur1: Unknown_neurons2, U_Neur1: Unknown_neurons3.

Supplementary Table 3: Associations between cell lines' factors and differentiation efficiency. Association between factors and 1) neuronal differentiation efficiency, 2) predicted differentiation scores and 3) fraction of cells in iPSC cluster 2**.** Specified (besides the outcomes and factors considered) are the tests performed (test), the sample size (N), and the nominal p values (p). Moreover, for continuous traits, the linear regression effect size (coefficient) and squared Pearson correlation (R2) are shown, whilst for binary traits, the average difference is reported.

Supplementary Table 4: List of genes correlated with differentiation efficiency. Provides all identified significant positive and negative genes associated with differentiation efficiency at FDR <5%. Correlations were quantified using the Pearson correlation coefficient. Significance was tested using the t-test statistic, and corrected for multiple testing using the Benjamini Hochberg procedure (n=184 cell lines).

Table field	**Description**
ensembl_gene_id	Ensembl ID (Ensembl version 75)
hgnc_symbol	Hugo Gene Nomenclature Committee symbol
coef	Effect size coefficient
pval	Nominal p value
pval_adjusted	Adjusted p value (BH)

Supplementary Table 5: Predicted neuronal differentiation efficiency scores for all HipSci lines. Provides a predictive score of neuronal differentiation efficiency for the HipSci resource. Cell lines included in this prediction are classified in 3 categories: successful - cell lines that were differentiated towards midbrain fate, failed - cell lines that failed quality control at the iPSC stage and were not included in any pool, not-assessed - cell line not included.

Table field	**Description**
Donor_id	HipSci cell line ID
in_study	Cell line included or not in this study
model_score	Predicted score of neuronal differentiation efficiency (**Methods**)
diff_efficiency	Sum of the proportions of DA and Sert cells produced on day 52 (i.e. neuronal differentiation efficiency; NA for all lines not included in our study)

Supplementary Table 6: DE genes between iPSC cluster 2 and others at FDR 0.05. All differentially expressed genes between each iPSC cluster and other clusters are provided after re-analysing iPSC scRNA-seq data from (Cuomo *et al.*, 2020). Differential expression was evaluated by t-test and corrected for multiple testing using the Benjamini Hochberg procedure. The number of cells in each cluster was n=4,596, 2,093, 1,766, 989, and 217 for clusters 0, 1, 2, 3, and 4, respectively.

Table field	**Description**
ensembl_gene_id	Ensembl ID (Ensembl version 75)
hgnc_symbol	Hugo Gene Nomenclature Committee symbol
pvals	Nominal p value
pvals_adj	Adjusted p value (BH)

log_fold_changes	Log Fold Changes
scores	Z-scores
cluster_id	Cluster ID

Supplementary Table 7: List of eQTL discoveries across cell types and conditions at FDR 0.05. All lead eQTL SNP-gene pairs are provided.

Table field	**Description**
snp_id	Lead variant, SNP ID in the format [chromosome]_[position]_[reference]_[alternative allele]
p_value	Nominal p value
beta	Effect size of the eQTL
beta_se	Standard error of the effect size
empirical_feature_p_value	Gene-level corrected p value using using 1,000 permutations (**Methods**)
feature_chromosome	Gene chromosome
feature_start	Gene body start position
feature_end	Gene body end position
n_samples	Number of samples tested
snp_chromosome	Variant chromosome
snp_position	Variant position
assessed_allele	Variant allele assessed
maf	Variant minor allele frequency
hwe_p	Hardy-Weinberg equilibrium test p value
BH_fdr	FDR corrected P-value using Benjamini Hochberg (BH) procedure
q_value	Q-value, globally corrected P-value using Storey procedure (**Methods**)
label	Celltype and condition label, in the form [celltype]_[time_point](_[stimulation])
ensembl_gene_id	Ensembl ID (Ensembl version 75)

Supplementary Table 8: List of neurological traits used for colocalisation analysis. All the neurological traits analysed are reported.

Table field	**Description**
study_id	GWAS trait study ID (same as Supplementary Table 9)

pmid	PubMed unique identifier (when a publication is present and not a preprint)
pub_date	Publication date
pub_journal	Publication journal (when a publication is present)
pub_title	Publication title (when a publication is present)
pub_author	First author of the publication
trait_reported	Trait reported
ancestry_initial	Ancestry of samples in main study
ancestry_replication	Ancestry of samples in replication study (when present)
n_initial	Sample size in main study
n_replication	Sample size in replication study (when present)
traits_category	Traits category

Supplementary Table 9: Results of the colocalisation analysis. All colocalisations identified are provided.

Table field	**Description**
study_id	GWAS trait study ID (same as **Supplementary Table 8**)
celltype_tissue	Celltype and condition label, in the form [celltype]_[time_point](_[stimulation]) for our study, tissue name for GTEx tissues
Ensembl_ID	Ensembl ID (Ensembl version 75)
hgnc_symbol	Hugo Gene Nomenclature Committee symbol
n_variants	Number of variants in 1Mb window
PP0	Posterior probability 0
PP1	Posterior probability 1
PP2	Posterior probability 2
PP3	Posterior probability 3
PP4	Posterior probability 4 (of colocalisation)
chromosome_grch37	Chromosome number according to GRch37
GWAS_index_pos_grch37	Position of in index GWAS variant 1Mb window
eQTL_lead_pos_grch37	Position of lead eQTL in 1Mb window

Supplementary Table 10: Overview of the HipSci lines used in this study.

Table field	Description
Cell_Line	Cell line ID
Donor_Sex	Donor sex
Donor_Age	Age range of the donor
Avg_Donor_Age	Average value of the donor's age range
Origin_Cell	Cell type of origin used for reprogramming
Reprogramming_Method	Reprogramming method used
Mean_X_Chrom_ASE	Mean ASE fraction of all heterozygous X chromosome SNPs per sample
Pool (rep1)	Pool number for replicate 1
Passage_Number (rep1)	Cell line passage number at the pooling stage for replicate 1
Pool (rep2)	Pool number for replicate 2
Passage_Number (rep2)	Cell line passage number at the pooling stage for replicate 2
Pool (rep3)	Pool number for replicate 3
Passage_Number (rep3)	Cell line passage number at the pooling stage for replicate 3

Supplementary Table 11: List of the top 20 GO terms enrichment from cluster 2. The top 20 hits of biological process enrichment analysis using all upregulated genes with a log fold change greater or equal to 1 as an input are provided. Enrichment analysis was performed using g:Profiler (https://biit.cs.ut.ee/gprofiler/gost), with P values for enrichment obtained using the hypergeometric test and corrected for multiple testing using the Benjamini-Hochberg procedure.

Source	Gene ontology
term_name	Name of the biological pathway
Term_id	Gene ontology ID
adjusted_p_value	Adjusted p value (BH)
negative_log10_of_adjusted_p_value	Negative log10 of the adjusted p value
term_size	Number of genes associated with the biological pathway
query_size	Number of genes imputed as a query
intersection_size	Number of genes at the intersection between the number of genes associated with the pathway and the number of genes imputed as a query

Supplementary Methods

Cell type annotation

We collected a broad range of marker genes for canonical brain cell types we would expect in our data, including progenitors, glial cells and neurons. First, the targeted floorplate identity of the cells was confirmed by assessing the expression of *LMX1A* and *FOXA2*[1,2]. Floor plate progenitors were characterised by the expression of *ZEB2*, *DMBX1*, *HMGA1* and *HMGA2,* which have been previously identified as floor plate progenitor markers[3,4]. Proliferating floor plate progenitors were identified by the additional expression of *MIK67*, *TOP2A, KIAA1524.* Second, pro-neuronal genes including *NEUROG1/2*, *NEUROD1/2*, *NHLH1* and *SIM1* were used to distinguish the neuroblast population from floor plate progenitors, at day 11[5,6]. Third, between day 30 and day 52, six neuronal populations were identified, as characterised by the expression of the pan-synaptic markers *SYT1* and *SNAP25*. Of the six neuronal populations, three could not be annotated unambiguously. For the largest population of these, we note that these cells showed marked expression of multiple markers of dopaminergic neurons compared to other populations (**Extended Fig. 1d**). However, at the same time we observed highly expressed genes associated with cortical identity known as cortical hem/Cajal retzius cells[7] (**Extended Fig. 1e**). This conflicting evidence let us to characterise this cluster as Unknown_neurons1. The remaining three neuronal populations could be identified as dopaminergic neurons, serotonergic-like neurons and a cell population consisting of proliferating progenitors and serotonergic-like neurons. To characterise the dopaminergic neuronal population, we collated a total of 75 marker genes from the literature[3,4,8,9] (**Extended Fig. 1d**). Next, 15 literature-curated marker genes[3,10–12] were used to characterise serotonergic-like neurons, which were further categorised as progenitors proliferating and serotonergic-like neurons when co-expressing *MIK67*, *TOP2A, KIAA1524*. Finally, non-neuronal populations were identified as glial cells based on the expression of *TNC*, *SOX2*, *CDH2*, *HES1*[13–16] and further divided into ependymal-like and astrocyte-like cells based on the expression of *STOML3*, *CCDC153, CDHR4, FOXJ1*, *DNAH11*, *TTR*, *MLF1* and *S100B*, *AQP4*, *GFAP*, *SLC1A3*, *SOX9*, respectively[17–19].

Alignment of DA cells to published datasets

We performed joint PCA using the function “multiBatchPCA” from the Bioconductor “batchelor” package followed by batch correction using mutual nearest neighbours[20], (i.e. MNN, as implemented in the “reducedMNN” function) of log-normalised counts from our data and each of the reference *in vivo* datasets. As input we used the union of 2,000 highly variable genes (HVGs, using the trendVar function from scran) from our data and the 2,000 HVGs from the reference dataset. Next, we asked which reference cell each of our cells was most similar to (i.e. ‘mapped to', using the queryKNN function implemented in BiocNeighbors, with k=1 nearest neighbour). In particular, we mapped the DA cell population to the set of all neurons from each of three datasets. First, La Manno et al human iPSC data[3], where 99% of our DA cells are mapped to the *iDAb* population. Second, La Manno et al human[3] fetal data, where 39% of our DA cells mapped to *hDA1*, 36% to *hDA2*, and 10% to *hDA0*. Finally, Welch et al[3,21] post-mortem data, where 91% cells mapped to *NEUROdop*, 8% to *NEUROinh1*.

X chromosome inactivation (XCI) status

XCI was assessed by considering allele-specific expression (ASE) from the X chromosome, as quantified by bulk RNA-seq. Allele-specific counts were obtained for SNPs present in DBSNP using GATK ReadCounter with the command 'GenomeAnalysisTK.jar -T ASEReadCounter -U ALLOW_N_CIGAR_READS --minMappingQuality 10 --minBaseQuality 2'. Heterozygous SNPs located on the X chromosome for which the total number of overlapping reads was >20 were retained for analysis. For each SNP, the ASE fraction was defined as the fraction of reads mapping to the less expressed allele (thus the ASE fraction was ≤ 0.5 for all SNPs). For each sample, the XCI status was quantified as the mean ASE fraction of all heterozygous X chromosome SNPs in that sample.

Association of cell line features with neuronal differentiation efficiency

We tested for associations between neuronal differentiation efficiency and the cell line donor's sex (n=199, t-test) as well as XCI status (on the subset of female lines; n=97), passage number (n=195), two pluripotency scores (n=196, F-test). The pluripotency scores used were the pluritest score and the novelty score, which were generated in the course of banking the HipSci cell lines[22].

Next, we tested the same features for associations with the predicted differentiation scores we estimated for a larger set of HipSci lines as described above (**Supplementary Table 5**). Again, we tested for associations with a cell line's donor sex (n=812, t-test) as well as XCI status (n=342), two pluripotency scores (n=797, F-test). This dataset included cell lines in both feeder and feeder-free culture conditions, and therefore we tested this as an additional factor denoted 'feeder free status' (n=812, t-test).

We corrected for multiple testing separately for the two sets of tests performed using Bonferroni correction. Results from this analysis are in **Supplementary Table 3**.

Finally, we considered an exploratory analysis, mapping genome-wide common variants (MAF>5%) to the predicted differentiation outcomes across 810 iPSC lines (540 unique donors), using a linear mixed model as implemented in LIMIX[23,24]. The use of multiple iPSC lines from the same donor was accounted for by including a random effect term in the model with covariance described by a kinship matrix. This identified no associated variants at genome-wide significance ($p>5x10^{-8}$).

Inference of proportion of iPSC cluster 2 cell fractions from bulk RNA-seq

Inference of the proportion of iPSC cluster 2 cells present in bulk RNA-seq samples was performed using Decon-cell[25]. This method relies only on having expression and cell fraction data for matched samples. In this case, we have bulk RNA-seq and measured proportions of iPSC cluster 2 (from single-cell RNA-seq) for a subset of 107 cell lines. These were used to train the classifier to infer the proportion of cluster 2 cells for iPSC lines from bulk RNA-seq, as shown in **Extended Fig. 5e**.

Evaluation of alternative cis eQTL mapping approaches

We performed cis eQTL mapping when omitting the second random effect component accounting for the number of cells (i.e. 1/n in Model 0, below), resulting in a substantially smaller number of discoveries (**Extended Fig. 6b**). In addition, for a selected eQTL map (untreated day 52 DA cells), we compared a wider range of alternative eQTL mapping approaches. We considered the following methods:

Model 0: y = PC1:15 + SNP + 1/n + noise (baseline, 1,024 eGenes at FDR<5%)
Model 1: y = pool + sex + SNP + 1/n + noise (608 eGenes, 574 of which also in Model 0)
Model 2: y = pool + sex + SNP + K + noise (320 eGenes, 312 of which shared with Model 0)
Model 3: y = PC1:15 + K + noise (471 eGenes, 457 of which shared with Model 0)
Model 4: y = pool + SNP + K + 1/n + noise (856 eGenes, 734 of which shared with Model 0)

Briefly, we considered models including pool as well as other factors such as sex directly as covariates, instead of relying on principal component-based correction (Models 1, 2 and 4). We also considered models where we computed average expression at the pool and line level, considering measurements from the same lines across pools as replicates, and accounting for such repeated structure using a random effect term ("K" in Models 2, 3 and 4), similar to traditional approaches to account for population structure. All the models considered resulted in largely overlapping sets of eGenes (85%-98% of eGenes identified with each of the alternative models are also identified in our baseline, see above), highlighting the robustness of this model, with the model considered in this study (Model 0) identifying the largest number of eGenes, at FDR<5% (1,024 eGenes, see **Extended Fig. 7c**).

References

1. Ferri A. L. M. *et al.* Foxa1 and Foxa2 regulate multiple phases of midbrain dopaminergic neuron development in a dosage-dependent manner. *Development* **134,** 2761-2769 (2007).
2. Andersson, E. *et al.* Identification of intrinsic determinants of midbrain dopamine neurons. *Cell* **124**, 393–405 (2006).
3. La Manno, G. *et al.* Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell* **167**, 566–580.e19 (2016).
4. Park, C.-H. *et al.* Acquisition of in vitro and in vivo functionality of Nurr1-induced dopamine neurons. *FASEB J.* **20**, 2553–2555 (2006).
5. Bertrand, N., Castro, D. S. & Guillemot, F. Proneural genes and the specification of neural cell types. *Nat. Rev. Neurosci.* **3**, 517–530 (2002).

6. Lacomme, M., Liaubet, L., Pituello, F. & Bel-Vialar, S. NEUROG2 drives cell cycle exit of neuronal precursors by specifically repressing a subset of cyclins acting at the G1 and S phases of the cell cycle. *Mol. Cell. Biol.* **32**, 2596–2607 (2012).
7. Loo, L. *et al.* Single-cell transcriptomic analysis of mouse neocortical development. *Nat. Commun.* **10**, 134 (2019).
8. Arenas, E., Denham, M. & Villaescusa, J. C. How to make a midbrain dopaminergic neuron. *Development* **142**, 1918–1936 (2015).
9. Ramonet, D. *et al.* PARK9-associated ATP13A2 localizes to intracellular acidic vesicles and regulates cation homeostasis and neuronal integrity. *Hum. Mol. Genet.* **21**, 1725–1743 (2012).
10. Ren, J. *et al.* Single-cell transcriptomes and whole-brain projections of serotonin neurons in the mouse dorsal and median raphe nuclei. *Elife* **8**, (2019).
11. Huang, K. W. *et al.* Molecular and anatomical organization of the dorsal raphe nucleus. *Elife* **8**, (2019).
12. Okaty, B. W. *et al.* A single-cell transcriptomic and anatomic atlas of mouse dorsal raphe neurons. *Elife* **9**, (2020).
13. Mercurio, S., Serra, L. & Nicolis, S. K. More than just Stem Cells: Functional Roles of the Transcription Factor Sox2 in Differentiated Glia and Neurons. *Int. J. Mol. Sci.* **20**, (2019).
14. Wu, Y., Liu, Y., Levine, E. M. & Rao, M. S. Hes1 but not Hes5 regulates an astrocyte versus oligodendrocyte fate choice in glial restricted precursors. *Dev. Dyn.* **226**, 675–689 (2003).
15. Wiese, S., Karus, M. & Faissner, A. Astrocytes as a source for extracellular matrix molecules and cytokines. *Front. Pharmacol.* **3**, 120 (2012).
16. Redies, C. Cadherins in the central nervous system. *Prog. Neurobiol.* **61**, 611–648 (2000).
17. Campbell, J. N. *et al.* A molecular census of arcuate hypothalamus and median eminence cell types. *Nat. Neurosci.* **20**, 484–496 (2017).

18. Sloan, S. A. *et al.* Human Astrocyte Maturation Captured in 3D Cerebral Cortical Spheroids Derived from Pluripotent Stem Cells. *Neuron* **95**, 779–790.e6 (2017).
19. Zhang, Y. *et al.* Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* **89**, 37–53 (2016).
20. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).
21. Welch, J. D. *et al.* Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell* **177**, 1873–1887.e17 (2019).
22. Müller, F.-J. *et al.* A bioinformatic assay for pluripotency in human cells. *Nat. Methods* **8**, 315–317 (2011).
23. Lippert, C., Casale, F. P., Rakitsch, B. & Stegle, O. LIMIX: genetic analysis of multiple traits. *bioRxiv* (2014).
24. Casale, F. P., Rakitsch, B., Lippert, C. & Stegle, O. Efficient set tests for the genetic analysis of correlated traits. *Nat. Methods* **12** 755–758 (2015).
25. Aguirre-Gamboa, R. *et al.* Deconvolution of bulk blood eQTL effects into immune cell subpopulations. *BMC Bioinformatics* **21**, 243 (2020).