Stem Cell Reports, Volume *16*

Supplemental Information

Defining the nature of human pluripotent stem cell-derived interneurons via single-cell analysis

Thomas Allison, Justin Langerman, Shan Sabri, Marcos Otero-Garcia, Andrew Lund, John Huang, Xiaofei Wei, Ranmal A. Samarasinghe, Damon Polioudakis, Istvan Mody, Inma Cobos, Bennett G. Novitch, Daniel H. Geschwind, Kathrin Plath, and William E. Lowry

Supplementary Material

Supplemental Table 1

Table S1: Summary of data sets used in this study

samples	cell line/sample information	RNA-sequencing method	associated publication	acession number
3i directed interneuron differentiation	hESCs, HES3:NKX2-1-GFP	single cell RNA Sequencing, Drop-Seq	this study	GEO #GSE180132
AD-directed interneuron differentiation, replicate 1	hESCs, H9	single cell RNA Sequencing, Drop-Seq	this study	GEO #GSE180132
AD-directed interneuron differentiation, replicate 2	hESCs, H9	single cell RNA Sequencing, Drop-Seq	this study	GEO #GSE180132
orgINs, day 56	hESCs, H9 - unfused mix of dorsal cortex (Cx) and ventral ganglionic eminence (GE) organoids	single cell RNA Sequencing, 10x Genomics	Samarasinghe et al, 2021, accepted	GEO #GSE165577
orgINs, day 70	hESCs, H9 - fusion of dorsal cortex (Cx) and ventral ganglionic eminence (GE) organoids	single cell RNA Sequencing, 10x Genomics	Samarasinghe et al, 2021, accepted	GEO #GSE165577
orgINs, day 100	hESCs, H9 - fusion of dorsal cortex (Cx) and ventral ganglionic eminence (GE) organoids	single cell RNA Sequencing, 10x Genomics	Samarasinghe et al, 2021, accepted	GEO #GSE165577
mINs	hESCs, H1- D24 of neural induction via Maroof et al. 2003 protocol	single cell RNA Sequencing, Smart-seq2	Close et al, 2017, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5480972/	GSE93593
adult brain #1, frozen archived cortex	Male, Age 71, Frontal Cortex	Drop-seq, NextSEQ500	Inma Cobos, unpublished	GEO #GSE181715
adult brain #2, frozen archived cortex	Female, Age 81, Frontal Cortex	Drop-seq, NextSEQ500	Inma Cobos, unpublished	GEO #GSE181715
adult brain #3, frozen archived cortex	Male, Age 83, Frontal Cortex	Drop-seq, NextSEQ500	Inma Cobos, unpublished	GEO #GSE181715
adult brain #4, frozen archived cortex	Male, Age 61, Frontal Cortex	Drop-seq, NextSEQ500	Inma Cobos, unpublished	GEO #GSE181715
adult brain #5, frozen archived cortex	Female, Age 51, Frontal Cortex	Drop-seq, NextSEQ500	Inma Cobos, unpublished	GEO #GSE181715
adult brain #6, frozen archived cortex	Female, Age 89, Frontal Cortex	Drop-seq, NextSEQ500	Inma Cobos, unpublished	GEO #GSE181715
fetal brain, 15pcw	CP, and CZ dissection	single cell RNA Sequencing, Drop-Seq	Polioudakis et al., 2019	dbGaP: phs001836
fetal brain, 17pcw	CP, and CZ dissection	single cell RNA Sequencing, Drop-Seq	Polioudakis et al., 2019	dbGaP: phs001836
Allen Brain Institute, adult brain	Frontal Cortex	sn RNA-seq, 10x Genomics	Allen Brain Institute	https://portal.brain-map.org/atlases-and-data/rnaseq

Materials and Methods

Cell Culture

TF-directed differentiation

For the TF programming method, undifferentiated cells were maintained as already described. For lentiviral infection, hESCs (H9, Wicell) were plated at a density of 1 x 104 cells/cm2 in a Matrigel (Corning #354234) coated 6 well plate in mTESR (StemCell Technologies (SCT) #85850)

and 10µm Y27632 (Tocris). After 24h, media was changed to fresh mTESR without Y27632 for a further 8h. After 8h, surviving cells were infected with pTET-O-FUW-Ascl1-puromycin (Addgene #97329) and pTet-O-FUW-Dlx2-hygromycin (Addgene #97330), FUW-M2rtTA (Addgene #20342) containing lentiviruses, kind gifts from Marius Wernig. Infected cells were maintained in mTESR and Matrigel to establish cell lines. For interneuron programming, we adopted a published method that included co-culture with murine glial cells directly isolated from brain (Yang et al., 2017). In short, infected and selected hESCs were dissociated into a single cell suspension and plated into matrigel coated T25 flasks in mTESR and 10µm Y27632. After 24h, media was replaced with mTESR without Y27632 and cells were maintained until confluency was reached. Once confluent, cells were kept in mTESR with the addition of 1µg/mL of Doxycycline (Sigma) to induce *ASCL1/DLX2* expression. After 24h of induction, puromycin (1µg/mL) and hygromycin (100µg/mL) were added to the mTESR/DOX media for a further 2 days. On day 4 of induction, puromcyin was withdrawn, but hygromycin maintained for additional 3 days. On day 7 of induction, media was changed to NBND and 10µm pluriSIn-1 (SCT) was added for 24h to kill proliferating cells and thus reduce the amount of non-neuronal cells in the culture. Induced interneurons were maintained in NBND for the remainder of the experiments.

3i directed differentiation towards interneurons (3i)

Directed differentiation was performed as described before [1]. Briefly, the hESC line HES3:NKX2-1-GFP) was maintained in mTESR on Matrigel-coated plates. Cells were passaged when confluent dishes using ReLeSR (SCT #05972). hESCs were dissociated to single cell with TrypLE and seeded into a 24-well plate coated with Matrigel at a concentration of 3×10^5 cells/cm^2 in mTESR and 10µm Y26732. After 24h, cells were washed with PBS and media was changed to NIMX. After 4 days in NIMX, media was changed to 75% NIMX, 25% N2 for 2 days. Media was then changed to 50% NIMX, 50% N2 for a further 2 days. At this stage, cells were dissociated into single cell suspension using trypLE and plated onto fresh Matrigel coated 24-well plates at a

density of 1.3 x 10^6 cells/cm^2, in 10µM Y27632 and 25% NIMX, 75% N2 media. Media was changed 24h later to remove Y27632 again with 25% NIMX, 75% N2. Neural progenitors were then ventralized using N2/B27 containing 2µM purmorphamine and 100ng/µL SHH for 8 days, with daily media changes. After 8 days, cells were cultured in N2/B27 media without Purmorphamine and SHH for a further 3 days. At this stage, cells were dissociated to a single cell suspension and FACS sorted to enrich for NKX2-1(+) interneuron progenitors. For 2D differentiation, after FACS, interneuron progenitors were plated onto fresh Matrigel coated 24-well plates at a density of 2.5 x 10^5 cells/cm^2 with N2/B27 media and 10µM Y27632. 24h later, media was changed to NBND, which was changed every 3 days for the duration of the experiment. For 3D differentiation, we followed the same protocol as the 2D differentiation, however, at the initial seeding before NIMX, we seeded dissociated undifferentiated hESCs into ultra-non-adhesive u-shaped 96 well plates at a density of 5000 cells/well in mTESR with 10µM Y27632. Once seeded, cells were centrifuged at 1000rpm for 3 mins to pellet cells. EBs routinely formed after 24h and Y27632 was removed and media was replaced with NIMX. All media changes and timings were identical as the 2D method.

Electrophysiology

Neurons marked by lentiviral GFP (prior to co-culture with murine glia) and prepared for whole cell patch-clamp recordings were transferred to a submerged recording chamber at 34°C and perfused at 5 ml/min with artificial cerebrospinal fluid (ACSF) containing (in mM): 126 NaCl, 10 D-glucose, 26 $NaHCO_3$, 0.05 $MgCl_2$, 2.5 KCl, 1.25 NaH_2PO_4, 2 $CaCl_2$, 1.5 $C_3H_3NaO_3$, 1 L-Glutamine (100 % O_2, pH 7.4, 290– 300 mOsm). All cells were visualized under IR-DIC upright microscope (Olympus BX-51WI, 20x XLUMPlan FL N objective) and whole-cell recordings were obtained with borosilicate patch pipettes (4 - 6 M, King Precision Glass) containing internal solutions (ICS) (in mM): 135 K-methanesulphonate, 5 KCl, 10 HEPES, 2 MgCl2, 3 NaCl, 0.2 EGTA, 2 Na2ATP, 0.2 NaGTP. The pH of the ICS was adjusted to 7.2 with KOH and its osmolarity

was 285 - 290 mOsm. ICS were stored at -80°C in 1 ml aliquots. Before each experiment, ICS aliquots were thawed to room temperature and kept on ice during recording. All salts were purchased from Sigma-Aldrich.

Recordings were obtained using an Axopatch 200B amplifier (Molecular Devices, San Jose, CA, USA), low-pass filtered at 5 kHz (Bessel, 8-pole) and digitized at 10 kHz with a National Instruments data acquisition board (BNC 2110, National Instruments, Austin, TX, USA). The injected current was programmed with WinWCP software (Strathclyde Electrophysiology Software, UK) and generated with another National Instruments board (USB-6221 National Instruments, Austin, TX, USA). All data were acquired and analyzed with EVAN (custom-designed LabView-based software). Current clamp recording. After achieving stable whole-cell configuration in voltage-clamp, the amplifier was switched to current-clamp mode. Only recordings with series resistances < 20 MΩ were used and bridge balance compensation was applied while in current-clamp. Action potentials were evoked by current step injections (up to 1200 pA, 100 - 300 ms duration). The patch clamp data were collected from 5 batches of cultured cells prepared with the AD-method. In each batch, there were 2-11 recorded cells. In total, 29 iINs prepared by the AD method were patched. Of these, 14 cells (48%) showed either multiple action potentials firing (n=5 or 17% of total, as in the upper panel of Fig. 1d) or single action potential firing (n=9 or 31% of total, as in the lower panel of Fig. 1d) upon injections of depolarizing current pulses.

Fetal tissue dissection

Fetal tissue samples were prepared by dissecting tissue into small pieces using a razor blade in ice-cold Hank's Balanced Salt Solution (HBSS). Following dissection, the tissue was gently dissociated via enzymatic digestion with papain (Worthington) for 45m-1h in a rotation chamber

at 37C. The resulting cell suspension was then filtered firstly with a 40µm strainer, followed by an ovomucoid gradient (Worthington). Cell survival and yield were quantified with Trypan blue staining before the immediate processing of cells with Drop-seq single cell sequencing[2].

Adult brain nuclei extraction

Frozen adult cortical tissue was tested for RNA quality through assessing the RIN score and only samples with a RIN score above 7 were used. Tissue was first dissected on ice to remove the majority of white matter and chopped into small pieces using a chilled scapel and immediately put into a pre-cooled dounce grinder with 2.4mL of homogenization buffer (Tris pH8 10 nm, $MgCl_2$ 5 mM, KCl 25 mM, sucrose 250 mM, DTT 1 µM, Protease Inhibitor 0.5x (Complete, Roche #4693159001), RNase inhibitor 0.2 U/µl (NxGen, Lucigen #30281), and 0.1% v/v Triton-x100. The tissue was then homogenized with 30 strokes of the grinder and filtered through a 40µM cell strainer into two1.5mL eppendorf tubes. Success of extraction and nuclei quality was determined with Trypan Blue staining and counting in a hemocytometer. The homogenate was then centrifuged at 1000g for 8 minutes at 4C, the supernatant aspirated and nuclei gently resuspended in 200µL of homogenization buffer and pooled from both tubes. To remove debris and to attain a pure nuclei preparation, resuspended nuclei were gently mixed with a 50% iodixanol solution to a final concentration of 25% and layered onto 500µL of 29% iodixanol solution in a 1.5mL eppendorf. The tubes were then centrifuged at 13,500g for 20 minutes at 4C, pelleted nuclei were harvested, counted and quality checked with Trypan blue.

Flow cytometry of nuclei from adult human brains

Staining of nuclei was achieved by resuspending nuclei in 500µL immunostaining buffer (PBS pH 7.4, BSA 0.05%, $MgCl_2$ 5 mM, DNAse I 2U/mL, RNase inhibitor 0.2 U/µl), following a 15 min incubation at 4C. After 15 min, the primary antibody was added (anti NeuN, 1:1000; Millipore #MAB377) and the suspension was incubated for a further 40min at 4C. Nuclei were centrifuged

for 5 min at 500g/4C. This process was repeated for the secondary antibody staining (Alexa Fluor 647, 1:1000) and nuclear staining (Hoescht 33258, 5 µg/mL). Nuclei were centrifuged and washed twice in immunostaining buffer.

Once stained, nuclei were resuspended in immunostaining buffer and sorted on a BD SORP FACSAria II. FSC and SSC were utilized to identify the nuclei population and the UV Trigon 355nm- and Trigon 640nm lasers to assay the expression of Hoechst 33258 and Alexa Fluor. The Hoechst staining was used to gate FSC-W and FSC-H to ensure single nuclei were enriched. Accurate gating of samples was achieved by using unstained and immunoglobulin controls. Nuclei were sorted into 1.5mL eppendorf tubes with 200µl of collection buffer (PBS pH 7.4, RNase inhibitor 1 U/µl). BSA 10% was then added to the tube after sorting for a final concentration of 0.01% and the tube inverted several times. The samples were then processed by single-nuclei drop-seq.

Immunostaining

Cells were fixed with 4% PFA at room temperature for 15 mins, followed with two PBS washes. Cells were blocked using 10% donkey serum/PBS solution for 30 mins at room temperature. Primary antibodies were diluted in staining buffer (PBS, 10% Donkey serum, 0.1% Triton X-100) and incubated with permeabilized cells overnight at 4C. Primary antibody was washed out with x2 PBS washed and secondary antibody along with DAPI was diluted in staining buffer and applied to the cells for a further 1h at 4C. Cells subsequently cells had two PBS washed before being mounted onto microscope slides using VECTASHEILD mounting media (Vector Labs).

Single-cell and single nuclei RNA-sequencing

Drop-seq was run on single cells according to the online protocol v.3.1 (http://mccarrolllab.com/download/905/) and the methods as published in the original publication

[3]. Slight modifications were made for single nuclei sequencing, which included increasing the final concentration of Sarkosyl to 0.8% to improve nuclei membrane lysis. Flow rates were also modified such that the oil was set to 16mL/h, cells and beads to 3mL/h to decrease the size of droplets due to the smaller size of nuclei. Additionally, STAMPs were incubated for 5 mins at 72C to facilitate RNA to DNA binding before droplet breakage. Cells and nuclei were prepared in 0.01% BSA/PBS solution and diluted to a concentration of 125,000 cells or nuclei/mL. Barcoded beads were obtained from Chemgenes and the PDMS microfluidic devices were obtained from FlowJem. Libraries were prepared with the Nextera XT DNA library preparation kit (Illumina) according to the manufacter's instructions. Libraries were subsequently sequenced on the Illumina HiSeq2500 or NextSeq4000 platforms with a modified 100bp pair-end protocol, such that R1 = 25bp and R2 = 75bp to maximize mapping.

Processing, Read Alignment and Digital Gene Expression (DGE) Matrix Construction

The raw drop-seq data was processed using the Drop-seq tools v1.12 pipeline from the McCaroll lab, utilizing the standard parameters as shown in the documentation (https://github.com/broadinstitute/Drop-seq/releases/tag/v1.12). In brief, cell and molecular barcodes were extracted from raw sequencing data based on bases 1-12 for cell and 13-20 for molecular barcodes, whilst filtering out reads with poor quality bases (TagBamWithReadSequenceExtended). Subsequently reads were trimmed to remove SMART adapter sequences as well as PolyA tails (FilterBAM, TrimStartingsequence & PolyATrimmer). HiSat2 was used to align these filtered reads to the human reference genome, hg38. Aligned reads were then merged with the unaligned reads to recapture molecular/cell BAM tags and subsequently reads were tagged with 'GE' if they overlapped with gene exons (MergeBamAlignment & TagReadWithGeneExon). Bead synthesis errors were then corrected and uMIs merged (DetectBeadSynthesisErrors). Finally, digital gene expression data (DGEs) were generated using standard parameters (DigitalExpression). For published datasets from

other groups which used the 10x Genomics scRNA-seq platform (see Table S1), data was processed using the Cellranger pipeline.

Computational analysis

Downstream analysis of single cell RNA-seq data was performed using Seurat (https://satijalab.org/seurat/) in R software (https://www.r-project.org/). Seurat objects were generated with DGEs generated by the drop-seq pipeline. For the scRNA-seq data from iINs and orgINs, cells were filtered based on a 500 cut off for minimum numbers of genes, for snRNA-seq data from the adult brain, fetal cell scRNA-seq and data from the Allen Brain Atlas, based on a 2500 cut off for minimum numbers of genes. Additionally, genes were filtered based on the criteria that they must be present in a minimum of 3 cells, to reduce noise. We employed a global-scaling normalization method to normalize RNA expression measurements for each cell by the total expression, which was then multiplied by a scale factor of 10,000. Highly variable genes by cell were computed and the top 2000 features from each dataset were utilized for subsequent downstream clustering. Canonical correlation analysis was used to identify common sources of variation between datasets, to correct for batch effects. The number of correlation coefficients (CC) to use was determined using elbow plots for with shared correlation strength against CC dimensions number. CC dimensions that showed a plateau for shared correlation strength were utilized and CCA subspaces aligned for subsequent clustering. This distance matrix was then reduced to low-dimensionality using UMAP and clusters identified by a shared nearest neighbor (SNN) modularity optimization-based clustering algorithm. First k-nearest neighbors were calculated to construct the SNN graph. The SNN was then used to optimize the modularity function to determine clusters. For all subsequent analyses, meta, filtered and normalized data were exported from Seurat for integration in various R packages. All correlations were calculated using Pearson correlation values on the normalized data for each group analyzed.

TF modules

Using the genes defined to be changing across developmental timepoints in Fig. 6, we computed a Pearson gene-to-gene correlation matrix and identified genes with correlation values greater than 0.15. To build modules, we linked hub TFs to all genes if the correlation threshold was satisfied. To ensure our correlation threshold was strict, we performed 750 random permutations of the DGE matrix to empirically estimate a null distribution of correlations for every pair of genes. All permutations resulted in 0 genes that met our threshold, indicating that our correlation values were sufficiently different from the null distribution. We removed target vertices that were not annotated as transcription factors/regulators by the Animal Transcription Factor Database (bioinfo.life.hust.edu.cn/AnimalTFDB/) and removed modules that did not contain at least two other correlated genes. The average expression of all members in a given module was computed for each cluster by taking the average of all normalized expression value for all genes and cells in the network. For the heatmap, TFs with low expression were removed to reduce noise (< 0.9 average normalized transcripts).

Supplemental References

1. Close, J.L., et al., *Single-Cell Profiling of an In Vitro Model of Human Interneuron Development Reveals Temporal Dynamics of Cell Type Production and Maturation.* Neuron, 2017. **96**(4): p. 949,
2. Polioudakis, D., et al., *A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation.* Neuron, 2019. **103**(5): p. 785-801 e8.PMC6831089
3. Macosko, E.Z., et al., *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.* Cell, 2015. **161**(5): p. 1202-14.PMC4481139