



Published in final edited form as:

*Stem Cell Res.* 2021 October ; 56: 102558. doi:10.1016/j.scr.2021.102558.

## Transcriptional signatures in iPSC-derived neurons are reproducible across labs when differentiation protocols are closely matched

Xylena Reed<sup>a,\*</sup>, Melanie M. Cobb<sup>b</sup>, Gaia Skinbinski<sup>b</sup>, Dorien Roosen<sup>a</sup>, Alice Kaganovich<sup>a</sup>, Jinhui Ding<sup>a</sup>, Steve Finkbeiner<sup>b,c</sup>, Mark R. Cookson<sup>a</sup>

<sup>a</sup>Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA

<sup>b</sup>Center for Systems and Therapeutics & Taube/Koret Center for Neurodegenerative Disease, Gladstone Institutes, San Francisco, CA, 94158, USA

<sup>c</sup>Departments of Neurology and Physiology, University of California San Francisco, San Francisco, CA 94158, USA

### Abstract

Reproducibility of expression patterns in iPSC-derived cells from different labs is an important first step in ensuring replication of biochemical or functional assays that are performed in different labs. Here we show that reproducible gene expression patterns from iPSCs and iPSC-derived neurons matured and collected at two separate laboratory locations can be achieved by closely matching protocols and reagents. While there are significant differences in gene expression between iPSCs and differentiated neurons, as well as between different donor lines of the same cell type, transcriptional changes that vary with laboratory sites are relatively small. These results suggest that making great efforts to match protocols, reagents and technical methods between labs may improve the reproducibility of iPSC-derived cell models.

### 1. Introduction

Induced pluripotent stem cells (iPSC) represent an important resource for examination of human cell biology in an experimentally modifiable context. This is particularly true for cell types that are difficult to access in living tissues, such as non-dividing neurons and other cells present in the human brain. As such, iPSC-derived cells have been used to

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\*Corresponding author.

CRedit authorship contribution statement

**Xylena Reed:** Formal analysis, Investigation, Visualization, Writing – original draft. **Melanie M. Cobb:** Investigation, Writing – review & editing. **Gaia Skinbinski:** Investigation, Writing – review & editing. **Dorien Roosen:** Investigation, Writing – review & editing. **Alice Kaganovich:** Investigation, Writing – review & editing. **Jinhui Ding:** Investigation, Writing – review & editing. **Steve Finkbeiner:** Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Supervision, Writing – review & editing. **Mark R. Cookson:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Supervision, Visualization, Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

model numerous neurological (Ghaffari et al., 2018; Imaizumi and Okano, 2014; Sandor et al., 2017), psychiatric (Soliman et al., 2017) and neurodevelopmental (Telias and Ben-Yosef, 2014) conditions. However, it is unclear if the phenotypes identified in iPSC-derived neurons can be compared across studies as differentiation protocols can be complex and lengthy, leading to the potential for divergence between laboratories. Early work in human embryonic stem cells (hESCs) showed that the efficiency of both spontaneous differentiation and directed cell type induction can vary widely between lines (Osafune et al., 2008). This variability in differentiability carries over to iPSCs from different donors and has been quantified in over 700 lines using functional outputs such as RNA-seq, DNA methylation arrays, proteomics and imaging of cell morphology (Kilpinen et al., 2017).

Reproducibility in iPSC studies is vital to their advancement as a model for human disease and translational research. One of the technically simplest and most comprehensive ways to analyze the variance in iPSC-derived cells from different labs is to use bulk RNA-seq as a functional readout. A recent study from the HipSci Consortium differentiated over 100 iPSC donor lines to sensory neurons in multiple differentiations and collected transcriptomic data to compare the reproducibility of gene expression across replicates. They found that replicates from the same donor line that were maintained on mouse embryonic fibroblasts (MEFs) clustered separately from derived iPSCs that were initially grown in feeder free E8 media (Schwartzentruber et al., 2018). Another study reported, considering differentiation of two iPSC donor lines across five laboratories, the greatest source of variability in gene expression was the laboratory where the cells were differentiated and matured and not the donor line (Volpato et al., 2018). Although these results were able to be computationally normalized using an established factor-analysis based method (Risso et al., 2014), it remained unclear whether improvements in the reproducibility of neuron differentiation between labs could be made. These results indicate that apparently small changes in culturing and differentiation protocols can significantly change the resulting transcriptome making comparisons between labs difficult. A recent review by some of the same authors recommends including common iPSC lines and genetic quality control steps along with detailed differentiation protocols and technical methods to improve experimental design and reduce site-to-site variation (Volpato and Webber, 2020). The authors also suggest including transcriptomics in iPSC studies as a functional readout to help identify, model, and exploit the variance in a way that validates the rest of the functional data.

In the current study, we employed some of these recommendations to evaluate site-to-site differences in iPSC-derived neurons to be used for downstream functional analyses. To this end, we performed RNA-Seq in iPSCs and iPSC-derived forebrain-type cortical neurons (iFBN) that were partially differentiated at one location and then shipped to the second lab for a synchronized final differentiation (Burkhardt et al., 2013). We used iPSCs from two unique donors (HD iPSC Consortium, 2017) and matured them to iFBN in two different laboratories to determine if lab-to-lab variability can be reduced by carefully controlling differentiation protocols and reagents. We collected RNA from each samples at both sites and then completed bulk RNA-seq to quantify expression genomewide. Our results indicate that careful protocol standardization and in person training of scientists can decrease inter-laboratory variance in transcriptomics to levels that are less than line-to-line variability.

## 2. Materials and methods

### 2.1. Forebrain neuron differentiation

iPSC lines CS25iCTR-18n2 (CS25i) and CS83iCTR-33n1 (33i) (HD iPSC Consortium, 2017) were grown at The Gladstone Institutes on Matrigel hESC qualified matrix (Corning, 354277) coated plates in mTeSR1 (Stemcell Technologies, 85850) until 90% confluent. iPSC cells were then differentiated as previously described (Burkhardt et al., 2013). Briefly, when iPSC cells reached 90% confluence mTeSR1, media was switched to 5 mL per well of N3 (base media is 50% DMEM/F12 (Thermo Fisher, 11320033) and 50% Neurobasal (Thermo Fisher, 21103049) with 0.5x GlutaMAX (Thermo Fisher, 25030–081), 1x Penicillin-Streptomycin (Thermo Fisher, 15140122), 0.5x B-27 minus vitamin A (Thermo Fisher, 12587010), 0.5x N2 supplement (Thermo Fisher, 17502048), 0.5x MEM Non-Essential Amino Acids (NEAA) (Thermo Fisher, 11140–050), 0.055 mM 2-mercaptoethanol (Thermo Fisher, 21985–023) and 1 µg/mL Insulin (Millipore Sigma, 91077C) plus 1.5 µM Dorsomorphin (Tocris Bioscience, 3093) and 10 µM SB431542 (Stemgent, 04–0010-05). N3 with Dorsomorphin and SB431542 was replaced every day for 11 days. Dorsomorphin and SB431542 were removed on day 12, and cells continued to be fed each day with N3. On days 16 through 20, N3 was supplemented with 0.05 µM Retinoic acid (Millipore Sigma, R2625). On day 20, cells were split 1:2 with trypsin and seeded onto plates pre-coated with 0.01% Poly-L-ornithine (Millipore Sigma, P4957–50ML) overnight, followed by coating with 2 µg/ml fibronectin (Fisher Scientific, CB40008A) and 0.2 µg/ml laminin (Millipore Sigma, L6274-.5MG) in DPBS (Thermo Fisher, 14190144) overnight. Cells were plated in N4 media (same as N3 plus 0.05 µM Retinoic acid, 2 ng/mL BDNF (R&D Systems, 248-BDB) and 2 ng/mL GDNF (R&D Systems, 212-GD)) with Y-27632 dihydrochloride (Fisher Scientific, 12–545-0). Media was changed the following day to N4 media without Y-27632 dihydrochloride. On day 24 neurons from each differentiation replicate were cryopreserved in Synth-a-Freeze (Thermo Fisher, A12542–01) and one vial from each replicate was shipped to NIH. These partially differentiated forebrain neurons were thawed by each lab on the same day, defined as day 25, and seeded in Poly-L-ornithine, laminin and fibronectin coated 12-well plates in N4 media with Y-27632 dihydrochloride. N4 media was transitioned to BrainPhys (Stemcell Technologies, 05792) with 0.05 µM Retinoic acid, 2 ng/ml BDNF and 2 ng/ml GDNF incrementally (day 26–75% N4: 25% BrainPhys; day 27–50% N4: 50% BrainPhys; day 28–25% N4: 75% BrainPhys; day 29 100% BrainPhys). Induced forebrain neurons (iFBN) were fed every other day with fresh BrainPhys media until day 37 when RNA was isolated. Although all catalog numbers were matched between labs, lot numbers were not. All growth factors were resuspended and stored as directed by their manufacturer. Complete medium was made as needed in batches of 500 mL and stored at 4°. Aliquots for daily media changes were warmed for 15–30 min in a 37° bead bath before feeding.

### 2.2. Immunocytochemistry of iFBN

Cells were plated onto 96-well plates and fixed with 4% paraformaldehyde. Cells were then blocked with 5% goat serum with 0.1% Triton-X-100 in PBS, and immunolabeled with primary antibody targeting MAP2 (Abcam, ab5392), followed by a species-specific fluorophore conjugated Alexa fluor secondary antibody (Thermo Fisher).

### 2.3. RNA isolation, library preparation and sequencing

RNA was isolated using a standard TRIzol Reagent (Thermo Fisher, 15596026) protocol (Chomczynski, 1993). Briefly, after removing growth media from each well, 500ul TRIzol was added and pipetted five times to lyse cells. Lysates were transferred to a 1.5 mL tube and incubated for five minutes. After incubation 100 µl chloroform was added and tubes shaken vigorously for 15 s to mix. The mix was incubated at room temperature for three minutes and then tubes were centrifuged at 12,000xg for 15 min at 4°. The top aqueous phase was transferred to a new tube and 250 µl Isopropanol was added to precipitate the RNA. Each sample was mixed well and incubated at room temperature for 10 min. Samples were centrifuged at 12,000xg for 10 min at 4° to pellet RNA. The resulting RNA pellet was washed with 500 µl freshly made 75% ethanol then centrifuged at 7500xg for 5 min at 4°. The supernatant was discarded and the pellet was air dried at room temperature for 10 min. The final pellet was resuspended in 40 µl RNase-free water, and quality and quantity were checked using the Bioanalyzer RNA 6000 Nano kit (Agilent, 5067–1511). To construct each library, we used 1 µg total RNA in the TruSeq Stranded Total RNA Sample Prep LS Protocol with rRNA depletion using Ribo Zero Gold (Illumina, 20020598). Library size was checked using the Bioanalyzer DNA 1000 kit (Agilent, 5067–1504) and each library and subsequent pool was quantified using the ddPCR Library Quantification Kit for Illumina TruSeq (BioRad, 1863040). Six libraries were pooled per lane at 7 pM and sequenced on a HiSeq 2500 (Illumina) following cluster generation to obtain ~40 million 250 bp reads for each sample.

### 2.4. Transcriptome analysis

The standard Illumina pipeline was used to generate fastq files, Ensembl GRCh38 annotated transcript abundance was quantified using Salmon in a non-alignment-based mode, and gene level counts were estimated using tximport package (Patro et al., 2017; Sonesson et al., 2015). We used DESEQ2 to analyze differential expression and ggplot2 to plot data in Rstudio v1.1.463 (Anders and Huber, 2010; Wickham, 2016) (Rstudio team, 2016). Volcano plots were made using Enhanced Volcano with Pcutoff = 10e-6 and FCcutoff = 2 (Blighe et al., 2018).

## 3. Results

### 3.1. Transcriptomics signatures are reproducible between laboratories

In order to evaluate the reproducibility of gene expression patterns in iPSC-derived cells matured and collected in different laboratories, we first identified the sources of variance by comparing expression between differentiation states, donor cell lines, and two different collection sites, NIH and Gladstone. Two iPSC lines from separate donors, CS83iCTR-33n1 (33i) and CS25iCTR-18n2 (CS25i), were partially differentiated (day 24) to cortical forebrain neurons (iFBN) at a single site (details in Table 1), frozen down and half of the cells were shipped to the second site for a fully synchronized final maturation and RNA collection (Fig. 1a). Immunocytochemistry of iFBN differentiated at UCSF shows high protein expression of MAP2 and neuronal morphology in both 33i and CS25i iFBN by (Fig. 1b). RNA-seq analysis shows that iFBN differentiated in both labs (NIH = red, UCSF = blue) using this protocol have low expression of iPSC markers *NANOG* and *POU5F1*

and correspondingly high expression of the broad neuronal marker *MAP2*, as well as the GABAergic neuron subtype markers, *GAD1* and *GAD2* (Fig. 1c). This analysis shows that we successfully differentiated iPSC to iFBN in both labs.

Next, we completed Principal Component Analysis (PCA) of all the samples to identify the main factors contributing to variability in gene expression. We show that Principal Component 1 (PC1) accounts for 83% of the total transcriptional variance and corresponds to the difference in expression patterns between iPSCs (black) and iFBN (red) (Fig. 2a). PC2 correlates with the donor line (33i = circle, CS25i = triangle) and accounts for 11% of the total variance. Visualizing all samples on a heatmap arranged by Euclidean distance we show that samples cluster mainly based on the differential expression between iPSCs and iFBN and between donor lines, while samples from each lab are mixed within each cell type (Fig. 2b). We next identified the number of significantly differentially expressed genes by cell type, donor and lab. As expected, the highest number of differentially expressed genes are identified when we compare iPSC to iFBN (Fig. 2c). Our comparison of the two donor lines (33i vs CS25i) shows a moderate number of differentially expressed genes (Fig. 2d). However, almost no genes are identified that have a significant p-value (p-value cutoff =  $10^{-6}$ ) and fold change ( $\log_2FC$  cutoff = 2) when comparing gene expression between labs (UCSF vs NIH) (Fig. 2e). Overall, the differential expression analyses indicate that the laboratory in which the cells were collected is not responsible for a large fraction of the variability in gene expression.

### 3.2. Donor line drives the majority of variance in differentiated cells

To understand more fully how the laboratory where cells were matured influences the variability in gene expression of differentiated cells, we extracted the top 1000 differentially expressed genes from the iFBN transcriptomic data. After removing the cell type variable (iPSC vs iFBN), we now show that PC1 in differentiated iFBN accounts for 64.5% of the variance and is driven mainly by the donor line used and secondarily by the differentiation number (Fig. 3a–f). PC2 describes 14.1% of the total variance and is also correlated with the differentiation number (Fig. 3a and d). The lab in which cells were matured does not appear until PC3 and accounts for 3% of the variance in expression (Fig. 3b and e). PC4 accounts for 2% of the variance and does not correlate with donor, differentiation number or lab (Fig. 3c and f). We predict this small amount of variation is due to minor differences in technical replicates. These results show that, at least under carefully controlled differentiation protocols, transcriptional outcomes can be stable across laboratories and support the use of iPSC-derived neurons to examine gene expression in a human context.

## 4. Discussion

In this study, we show that partially differentiated iFBN can be cryopreserved, shipped across the country and successfully matured to neurons with highly reproducible gene expression by closely matching differentiation protocols and reagents. We went to great lengths to ensure that all reagents between labs were identical. Matching all reagents down to the product number may help reduce the differences in expression in cells differentiated across centers. Additionally, although an added expense, introducing mandatory in person

trainings for all collaborating scientists involved in the culturing and differentiation of iPSCs could also help to significantly reduce site-to-site variability. This type of training can minimize subjective conclusions and avoid seemingly minor methodological changes in protocols. We conclude that the initial expenses associated with in-person training would be more than paid back by achieving highly reproducible results across labs. If in-person training is not an option, then a good alternative may be to create visual multimedia protocols to go along with detailed written protocols. Improved teleconferencing technology and experience over the past year has made these tools more easily accessible to many scientists and may aid in increasing reproducibility between labs.

There are several important limitations of this study. First, we only used a relatively small number of samples and hence it is unclear how broadly applicable these results are across larger numbers of lines. Second, it might be argued that because the senior authors have published together before these highly reproducible results are due to an extremely close relationship between the labs. However, the primary authors at each site had not worked together before this project and, with the exception of a one week training at the Gladstone Institutes at the outset, communication between scientists at different sites was primarily through email and conference calls. Thus, while a prior working relationship might have been beneficial for limiting variance in this small set of experiments, we do believe that a similar closely aligned approach may be practical for small scale collaborations.

Another possible limitation is the use of cells that were partially differentiated at one site, rather than fully independent site specific differentiations. Many protocols require careful evaluation of cell density at early stages, a function of cell growth rates that are often variable, especially with different passages of cells. The use of partially differentiated cells in parallel functional assays at each site was a specific decision in the design of these experiments and we acknowledge that there may indeed be more variance between labs if each started at day 0. However, it is important to note that the initial partial differentiation replicates were completed by different individuals up to a year and a half apart and still show little variability. Additionally, although the reagents and catalog numbers were the same between all differentiation replicates, lot numbers were not. We also note that partially differentiated cells can be frozen and shipped, as we did here, and so this approach could be used widely.

Although our experiments here are therefore limited and specific, we believe that the general principles of using the same detailed protocols with matched reagents and as much shared training as is feasible, may be able to be generalized to larger collaborations. As discussed above, we note that extrapolation to a more inclusive group may require additional approaches that may include hands-on workshops and video methods.

Our transcriptomic data from two matched control iPSC lines agree with a recent review suggesting that using a common set of case and control lines in iPSC studies makes the sources of variation more easily identifiable (Volpato and Webber, 2020). However, previous studies have shown limited clustering of transcriptional signatures from the same iPSC lines collected in different laboratories even with detailed shared protocols (Volpato et al., 2018). The work reported here is a step towards improving overall experimental reproducibility

between labs and indicates that iPSCs can be partially differentiated, frozen down and shipped to collaborating labs to be used in parallel functional studies without losing cell identity. Our data also support prior studies showing that most variance in gene expression across iPSC studies is explained by the variation between individual donor lines (Kilpinen et al., 2017; Schwartzenruber et al., 2018). Further work across more labs, donor cells and cell types is needed to confirm these results.

## Acknowledgements

All transcriptomics data have been uploaded to GEO (<https://www.ncbi.nlm.nih.gov/geo/>) under Series GSE167051.

### Funding

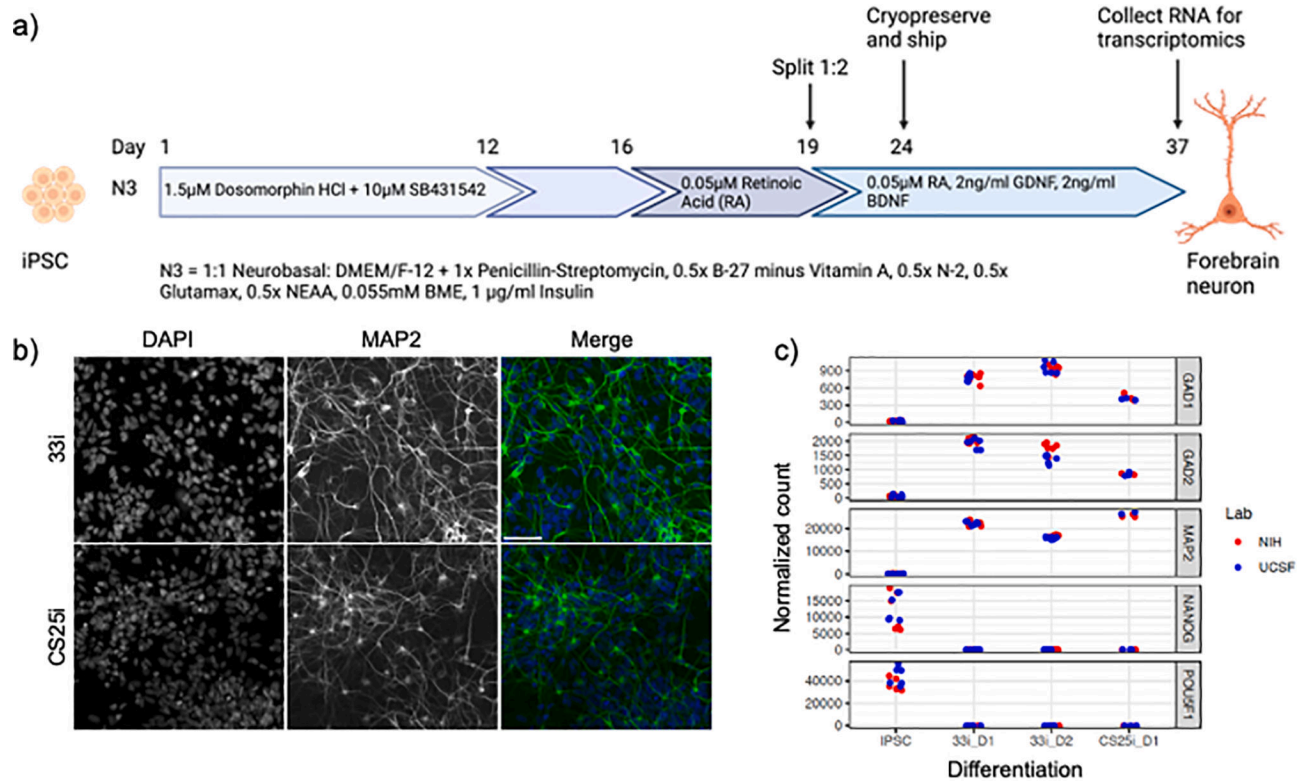
This work was supported by the Michael J Fox Foundation for Parkinson's Research under the HeadStart program (M.C., S.F.). This research was supported in part by the Intramural research Program of the NIH, National Institute on Aging and the National Institutes of Neurological Diseases and Stroke (M.C.). Additional support came from the National Institute on Aging (RF1 AG058476, P01 AG054407), the Taube/Koret Center for Neurodegenerative Disease Research and the Gladstone Institutes Center for Systems and Therapeutics (S.F.).

## References

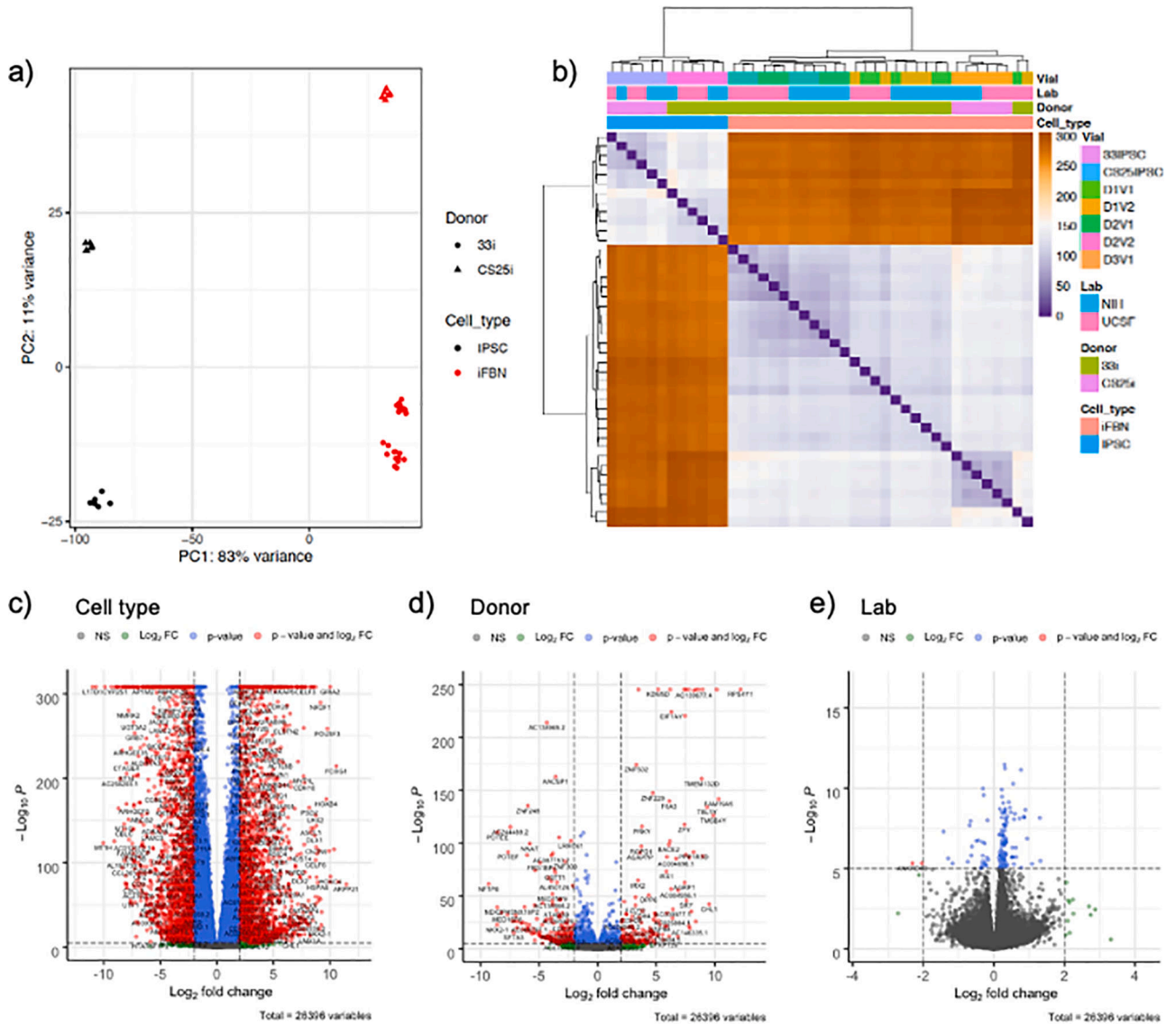
- Anders S, Huber W, 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11, R106. [PubMed: 20979621]
- Blighe K, Rana S, and Lewis M. 2018. "EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling." <https://github.com/kevinblighe/EnhancedVolcano>.
- Burkhardt MF, Martinez FJ, Wright S, Ramos C, Volfson D, Mason M, Garnes J, Dang V.u., Lievers J, Shoukat-Mumtaz U, Martinez R, Gai H, Blake R, Vaisberg E, Grskovic M, Johnson C, Irion S, Bright J, Cooper B, Nguyen L, Griswold-Prenner I, Javaherian A, 2013. A cellular model for sporadic ALS using patient-derived induced pluripotent stem cells. *Mol. Cell. Neurosci.* 56, 355–364. [PubMed: 23891805]
- Chomczynski P, 1993. A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques* 15 (532–534), 536–537.
- Ghaffari LT, Starr A, Nelson AT, Sattler R, 2018. Representing diversity in the dish: using patient-derived in vitro models to recreate the heterogeneity of neurological disease. *Front. Neurosci.* 12, 56. [PubMed: 29479303]
- HD iPSC Consortium, 2017. Developmental alterations in Huntington's disease neural cells and pharmacological rescue in cells and mice. *Nat. Neurosci.* 20, 648–660. [PubMed: 28319609]
- Imaizumi Y, Okano H, 2014. Modeling human neurological disorders with induced pluripotent stem cells. *J. Neurochem.* 129 (3), 388–399. [PubMed: 24286589]
- Kilpinen H, Goncalves A, Leha A, Afzal V, Alasoo K, Ashford S, Bala S, Bensaddek D, Casale FP, Culley OJ, Danecek P, Faulconbridge A, Harrison PW, Kathuria A, McCarthy D, McCarthy SA, Meleckyte R, Memari Y, Moens N, Soares F, Mann A, Streeter I, Agu CA, Alderton A, Nelson R, Harper S, Patel M, White A, Patel SR, Clarke L, Hali R, Kirton CM, Kolb-Kokocinski A, Beales P, Birney E, Danovi D, Lamond AI, Ouwehand WH, Vallier L, Watt FM, Durbin R, Stegle O, Gaffney DJ, 2017. Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* 546 (7658), 370–375. [PubMed: 28489815]
- Osafune K, Caron L, Borowiak M, Martinez RJ, Fitz-Gerald CS, Sato Y, Cowan CA, Chien KR, Melton DA, 2008. Marked differences in differentiation propensity among human embryonic stem cell lines. *Nat. Biotechnol.* 26 (3), 313–315. [PubMed: 18278034]
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C, 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14 (4), 417–419. [PubMed: 28263959]
- Risso D, Ngai J, Speed TP, Dudoit S, 2014. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* 32 (9), 896–902. [PubMed: 25150836]

- Sandor C, Robertson P, Lang C, Heger A, Booth H, Vowles J, Witty L, Bowden R, Hu M, Cowley SA, et al. , 2017. Transcriptomic profiling of purified patient-derived dopamine neurons identifies convergent perturbations and therapeutics for Parkinson's disease. *Hum. Mol. Genet.* 26, 552–566. [PubMed: 28096185]
- Schwartzentruber J, Foskolou S, Kilpinen H, Rodrigues J, Alasoo K, Knights AJ, Patel M, Goncalves A, Ferreira R, Benn CL, Wilbrey A, Bictash M, Impey E, Cao L, Lainez S, Loucif AJ, Whiting PJ, Gutteridge A, Gaffney DJ, 2018. Molecular and functional variation in iPSC-derived sensory neurons. *Nat. Genet.* 50 (1), 54–61. [PubMed: 29229984]
- Soliman MA, Aboharb F, Zeltner N, Studer L, 2017. Pluripotent stem cells in neuropsychiatric disorders. *Mol. Psychiatry* 22 (9), 1241–1249. [PubMed: 28322279]
- Soneson C, Love MI, Robinson MD, 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* 4, 1521. [PubMed: 26925227]
- Telias M, Ben-Yosef D, 2014. Modeling neurodevelopmental disorders using human pluripotent stem cells. *Stem Cell Rev. Rep.* 10 (4), 494–511. [PubMed: 24728983]
- Volpato V, Webber C, 2020. Addressing variability in iPSC-derived models of human disease: guidelines to promote reproducibility. *Dis. Model. Mech* 13, dmm042317. [PubMed: 31953356]
- Volpato V, Smith J, Sandor C, Ried JS, Baud A, Handel A, Newey SE, Wessely F, Attar M, Whiteley E, et al. , 2018. Reproducibility of molecular phenotypes after long-term differentiation to human iPSC-derived neurons: a multisite omics study. *Stem Cell Rep.* 11, 897–911.
- Wickham H, 2016. *ggplot2: elegant graphics for data analysis*. Springer, Cham.

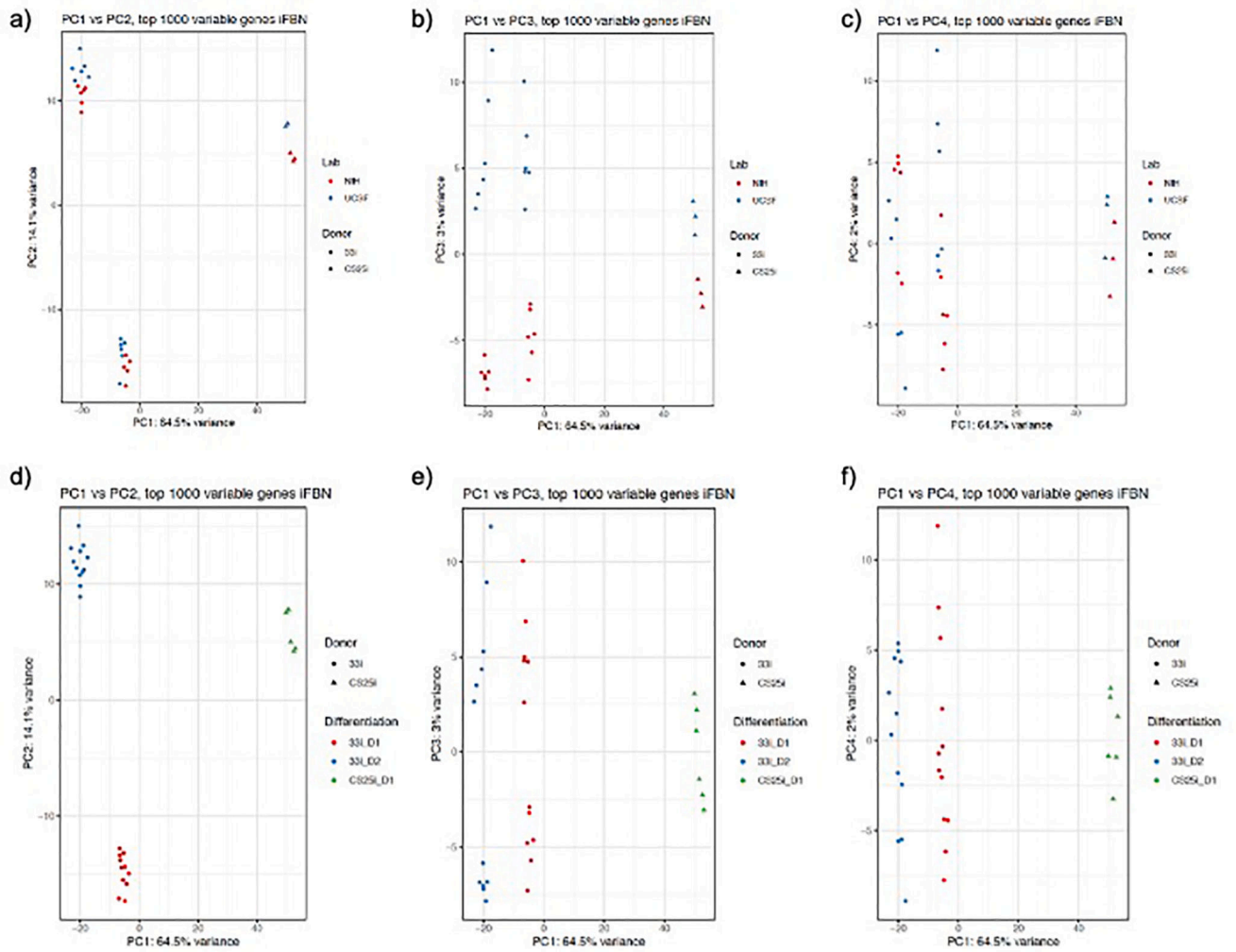




**Fig. 1.** Differentiation of iPSCs to iFBN makes cells with neuronal morphology and expression patterns. a) Differentiation strategy from iPSCs to iFBN (Created with [BioRender.com](https://www.biorender.com)). b) Representative images of iFBN from line 33i (top panels) and CS25i (bottom panels) immunolabeled with DAPI (blue) and MAP2 (green). Scale bar = 50 µm. c) RNA-seq expression of cell type markers in iPSCs and iFBN collected at NIH (red) and UCSF (blue).



**Fig. 2.** Cell type, then donor, then lab drive variation in transcriptomic data. a) PCA of IPSC (black) and differentiated iFBN (red). Circle = donor line 33i, triangles = donor line CS25i. b) Heatmap showing the Euclidean distance in gene expression between cell types collected in different labs. d) Volcano plot showing the differential expression analyzed by Cell type (IPSC vs iFBN; P value cutoff = 10e-6, log<sub>2</sub>FC cutoff = 2; Gray = Nonsignificant, Green = Significant Log<sub>2</sub>FC, Blue = Significant P value, Red = Significant P value and log<sub>2</sub>FC). d) Volcano plot showing the differential expression analyzed by Donor line (33i vs CS25i; P value cutoff = 10e-6, log<sub>2</sub>FC cutoff = 2; Gray = Nonsignificant, Green = Significant Log<sub>2</sub>FC, Blue = Significant P value, Red = Significant P value and log<sub>2</sub>FC). e) Volcano plot showing the differential expression analyzed by Lab (NIH vs UCSF; P value cutoff = 10e-6, log<sub>2</sub>FC cutoff = 2; Gray = Nonsignificant, Green = Significant Log<sub>2</sub>FC, Blue = Significant P value, Red = Significant P value and log<sub>2</sub>FC).



**Fig. 3.**

Principal component analysis of the top 1000 expressed genes in iFBN shows sources of variation in differentiated cells. a) PC1 vs PC2; Circle = donor line 33i, triangles = donor line CS25i; red = matured at NIH, blue = matured at UCSF. b) PC1 vs PC3; Circle = donor line 33i, triangles = donor line CS25i; red = matured at NIH, blue = matured at UCSF. c) PC1 vs PC4; Circle = donor line 33i, triangles = donor line CS25i; red = 33i differentiation 1, blue = 33i differentiation 2, green = CS25i differentiation 1. d) PC1 vs PC2; Circle = donor line 33i, triangles = donor line CS25i; red = 33i differentiation 1, blue = 33i differentiation 2, green = CS25i differentiation 1. e) PC1 vs PC3; Circle = donor line 33i, triangles = donor line CS25i; red = matured at NIH, blue = matured at UCSF. f) PC1 vs PC4; Circle = donor line 33i, triangles = donor line CS25i; red = 33i differentiation 1, blue = 33i differentiation 2, green = CS25i differentiation 1.

**Table 1**

iPS cell lines used and replicates per line.

Cell line	Differentiations	Vials/differentiation	Wells/Vial	Total samples per lab
33i (CS83iCTR-33n1)	2	2	3	12
CS25i (CS25iCTR-18n2)	1	1	3	3

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript