


RESEARCH ARTICLE

Open Access



Insights for disease modeling from single-cell transcriptomics of iPSC-derived Ngn2-induced neurons and astrocytes across differentiation time and co-culture

D. Das^{1†}, S. Sonthalia^{2†}, G. Stein-O'Brien¹, MH. Wahbeh¹, K. Feuer¹, L. Goff^{1,3}, C. Colantuoni^{3,4,5*}, V. Mahairaki¹ and D. Avramopoulos^{1,6*} 

Abstract

Background Trans-differentiation of human-induced pluripotent stem cells into neurons via Ngn2-induction (hiPSC-N) has become an efficient system to quickly generate neurons a likely significant advance for disease modeling and in vitro assay development. Recent single-cell interrogation of Ngn2-induced neurons, however, has revealed some similarities to unexpected neuronal lineages. Similarly, a straightforward method to generate hiPSC-derived astrocytes (hiPSC-A) for the study of neuropsychiatric disorders has also been described.

Results Here, we examine the homogeneity and similarity of hiPSC-N and hiPSC-A to their in vivo counterparts, the impact of different lengths of time post Ngn2 induction on hiPSC-N (15 or 21 days), and the impact of hiPSC-N/hiPSC-A co-culture. Leveraging the wealth of existing public single-cell RNA-seq (scRNA-seq) data in Ngn2-induced neurons and in vivo data from the developing brain, we provide perspectives on the lineage origins and maturation of hiPSC-N and hiPSC-A. While induction protocols in different labs produce consistent cell type profiles, both hiPSC-N and hiPSC-A show significant heterogeneity and similarity to multiple in vivo cell fates, and both more precisely approximate their in vivo counterparts when co-cultured. Gene expression data from the hiPSC-N show enrichment of genes linked to schizophrenia (SZ) and autism spectrum disorders (ASD) as has been previously shown for neural stem cells and neurons. These overrepresentations of disease genes are strongest in our system at early times (day 15) in Ngn2-induction/maturation of neurons, when we also observe the greatest similarity to early in vivo excitatory neurons. We have assembled this new scRNA-seq data along with the public data explored here as an integrated biologist-friendly web-resource for researchers seeking to understand this system more deeply: <https://nemoanalytics.org/p?!=DasEtAINGN2&g=NES>.

Conclusions While overall we support the use of the investigated cellular models for the study of neuropsychiatric disease, we also identify important limitations. We hope that this work will contribute to understanding and optimizing cellular modeling for complex brain disorders.

[†]Das D and Sonthalia S contributed equally to this work.

*Correspondence:

C. Colantuoni

D. Avramopoulos

adimitr1@jhmi.edu

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords Schizophrenia, Autism, Alzheimer's, Transcriptome, Single cell, Induced pluripotent stem cells, Neurons, Astrocytes, Ngn2

Background

Recent advances in cell engineering have provided unprecedented tools for investigating the biology and genetics underlying psychiatric disorders [1]. For many years, our only opportunity to study the central nervous system (CNS) and create disease models was through model organisms like worms and mice or tumor-derived cell lines. These models, while valuable in understanding how the CNS functions, came with significant limitations when drawing parallels to the complex human brain. Four recent technologies have drastically widened the array of tools to model disease: the generation of human-induced pluripotent stem cells (hiPSCs) from somatic cells, techniques for differentiation to specific cell types, genome editing, and high-throughput transcriptomics including single-cell RNA sequencing (scRNA-seq). We can now generate pluripotent cells from patients or healthy controls, introduce precise genetic modifications, and generate different types of cells of interest [2–7], such as glutamatergic, GABAergic, dopaminergic neurons, or glial cells. We can then study the consequences on their transcriptome either in bulk or at single-cell resolution which allows us to detect and account for cellular heterogeneity. With all these advances cellular models are becoming a front-line tool in brain research. However, there are important limitations to consider when working with such models. Specifically, in two-dimensional (2D) cultures, different neural cell types are often grown in isolation, in the absence of the milieu of neural types and supporting cells found in vivo. While three-dimensional cultures (organoids) allow more complex cellular interactions and more advanced maturational states, 2D systems often produce more uniform cell states that are more amenable to assay development for assessing novel therapeutics. Furthermore, the differentiation technologies are far from recapitulating in vivo differentiation; although similarities to the target cell types have been shown [1], significant differences also exist. scRNA-seq provides increased resolution to answer some key questions on cell type identity and state. By acquiring transcriptomes from single cells, either from cultures or from living tissues, we can get a better-resolved picture of the component cell types/states of a culture population and perform direct comparisons between in vivo and in vitro differentiated cells.

In this study, we focus on cells differentiated in vitro from hiPSCs, specifically excitatory neurons and astrocytes (hiPSC-N and hiPSC-A). For hiPSC-N, we use a

transcription factor (Ngn2)-mediated rapid induced differentiation protocol [7], a method that is popular due to its speed and versatility of starting cell type (lymphocytes, fibroblasts, iPSCs, etc.) [7]. To generate hiPSC-A, we use a previously described protocol to generate cells similar to primary human fetal astrocytes and characteristic of a non-reactive state suggested for use in neuron-astrocyte co-cultures [8]. Studying these differentiated cells in 2D cultures can be a powerful approach to model human psychiatric disease [1]. Yet, to best interpret any observed cellular phenotyping results, it is important to test cells on four different parameters/attributes: (1) How much do these cells resemble the in vivo intended cell types? (2) How homogeneous are they in 2D cultures? (3) How does variation in the differentiation time and co-culture with human astrocytes affect the neural identity of the cells and (4) How similar are cells produced by different laboratories using the same or similar methods? To help answer these questions, we examined four conditions by scRNA-seq: (A) hiPSC-N after 15 days of differentiation (hiPSC-N15), (B) hiPSC-N after 21 days of differentiation (hiPSC-N21), (C) hiPSC-A grown alone (hiPSC-A0), and (D) hiPSC-A co-cultured with hiPSC-N21 (neurons: hiPSC-N21A, astrocytes: hiPSC-AN21). Using scRNA-seq analysis, we explore whether cells under these differing conditions can be distinguished. We perform pseudo-bulk comparisons (hiPSC-N15 vs. hiPSC-N21, hiPSC-N21 vs. hiPSC-N21A, hiPSC-A0 vs. hiPSC-AN21) to find what genes are differentially expressed and the pathways and disease genes for which they are enriched. Finally, we explore how the induced neurons and astrocytes studied here compare to in vitro and in vivo cell types in other studies.

In this study, we refer to multiple external datasets which we examine alongside ours. To provide a single location where all the diverse datasets examined in this report can be accessed and explored in an integrated environment, we have created a web resource leveraging the gEAR and NeMO Analytics platforms [9, 10]. We invite researchers to explore this resource that can visualize individual genes of interest or sets of genes simultaneously across the many datasets used here [11].

Results

Cell type marker genes and cellular heterogeneity

Following scRNA-seq of induced cell types at different times and co-culture conditions, we performed read alignment, tabulated gene level counts per million

(CPM), calculated $\log_2(\text{CPM} + 1)$, and performed principal component analysis (PCA) followed by Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction followed by K-nearest neighbor (KNN) graph construction in PC space and cell clustering with Louvain optimization of modularity [12] within the Seurat package in the R/Bioconductor environment to explore cells and homogeneity within cell types (see the “Methods” section for details). The hiPSC-N15, hiPSC-N21, and hiPSC-Ast0, which were grown in separate plates from each other, showed that, as expected, hiPSC-N and hiPSC-A clustered separately (Fig. 1A). This allowed us to distinguish hiPSC-N21A from hiPSC-AN21 based on their transcriptome profile despite them being grown in the same plate. We then calculated pairwise correlations (r^2) of pseudobulk expression profiles of all genes across the 5 conditions: hiPSC-N15, hiPSC-N21, hiPSC-N21A, hiPSC-A0, and hiPSC-AN21 (Fig. 1B). Within target cell types (i.e., within hiPCS-N or hiPCS-A), the correlations between the different conditions were strong (min $r^2 > 0.96$), while across different discrete cell types, they were significantly weaker (max $r^2 < 0.58$) highlighting the difference between hiPSC-N and hiPSC-A. To determine how faithfully each iPSC-derived group

represents in vivo excitatory neurons and astrocytes, we explored the expression of a long list of astrocytic and neuronal marker genes compiled based on what is commonly seen in the literature and our own experience. As expected, hiPSC-A0 and hiPSC-AN21 exhibited higher expression of astrocytic markers than neurons, and this was more pronounced in the hiPSC-AN21 which may be an indication of higher maturity. The fraction of the cells in each cell type expressing the genes in Table 1 and the mean expression of each gene in each cell type are shown in Additional files 1, 2, 3, 4 and 5: Sup. Figure 1, Sup. Figure 2, Sup. Figure 3, Sup. Figure 4, and Sup. Figure 5.

Most neuronal marker genes showed higher expression in hiPSC-N than hiPSC-A with one exception, *SATB2*. *SATB2* is a postmitotic determinant for upper-layer neuron specification not present in all neurons [13]. While this can explain its absence in hiPSC-N, it is not clear why we observe it expressed in hiPSC-A. Multiple calcium/calmodulin-dependent protein kinases (CaMK) were highly expressed in hiPSC-N. The only exception was *CAMK2D* whose expression was higher in hiPSC-A. This is in agreement with Vallano et al. [14] who have shown that CAMK2D is a CaM kinase type II with specific astrocyte expression. All other synaptic

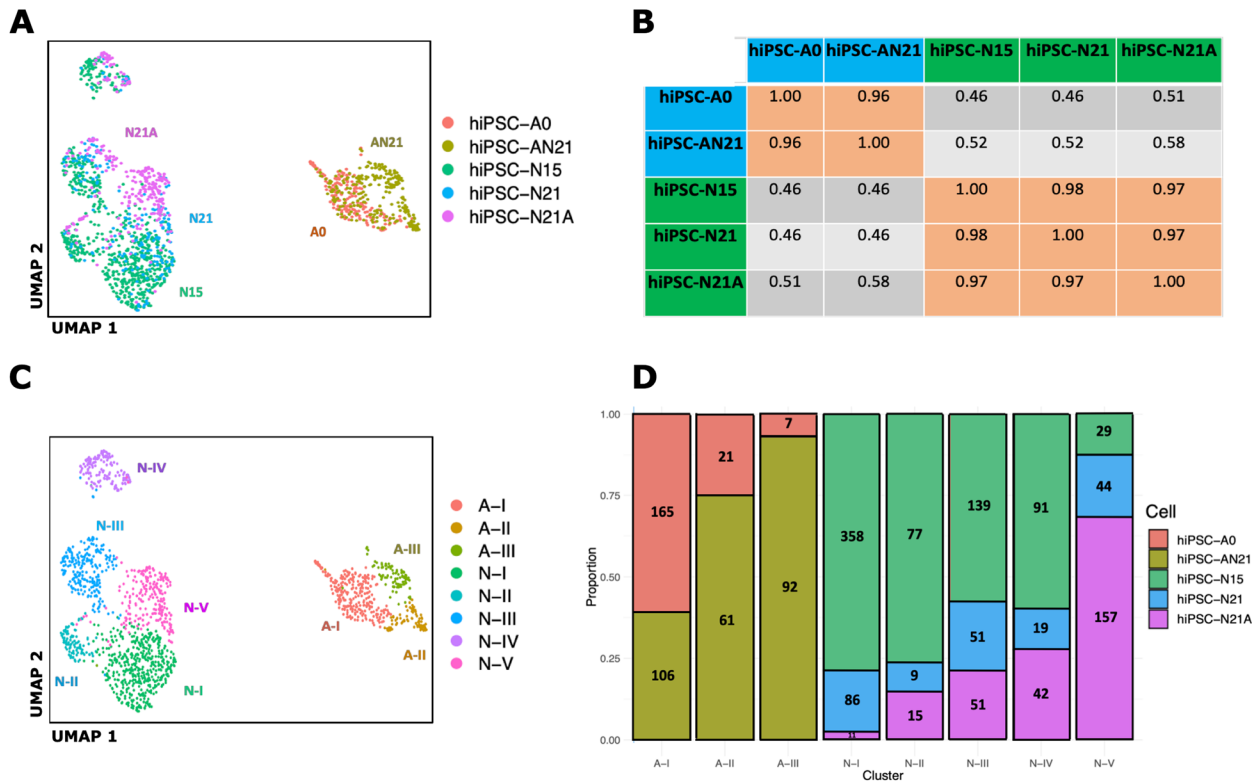


Fig. 1 UMAP plot of our data highlighting the different conditions (A), pairwise correlations of gene expression in bulk and pseudobulk gene expression for the conditions (B), the same UMAP plot highlighting cluster derived with Louvain clustering (C), and compositions of the Louvain clusters in terms of cells from the five conditions (D)

Table 1 Expression of marker genes across the different conditions. Overall expression in counts per million is shown in column 3, with darker grey shading indicating higher expression. The following columns show the expression in each condition expressed in standard deviations from the mean and proportionally highlighted from blue (negative) to red (positive). Asterisks indicate statistical significance <0.05 between adjacent columns, with the column named A-N-SIG indicating statistical significance <0.05 between all neuronal and al astrocytic cells types

Other names and information	Gene	AV. CPM	Ast	SIG	AstN21	A-N-SIG	N15	SIG	N21	SIG	N21Ast	cell types
Glial fibrillary acidic protein	<i>GFAP</i>	1.05	-0.140	*	1.694	*	-0.528		-0.124		-0.903	ASTR
GLAST/EAAT1	<i>SLC1A3</i>	1.18	-0.172	*	1.737	*	-0.682		-0.663		-0.220	ASTR
Vimentin	<i>VIM</i>	885.83	0.404	*	1.586	*	-0.683	*	-0.683		-0.625	ASTR
Glia-Derived Nexin	<i>SERPINE2</i>	68.57	1.564	*	0.449	*	-0.683		-0.684	*	-0.646	ASTR
paraoxonase 2	<i>PON2</i>	50.92	0.433	*	1.572	*	-0.679		-0.669		-0.657	ASTR
Apolipoprotein E	<i>APOE</i>	3.44	-0.421		1.786	*	-0.514	*	-0.490		-0.362	ASTR
Neuraminidase 1	<i>NEU1</i>	3.47	0.607		1.463	*	-0.808		-0.779		-0.483	ASTR
Connexin 43	<i>GJA1</i>	14.75	1.201	*	0.983	*	-0.734		-0.736	*	-0.714	ASTR
late marker of astrocyte development - 17078026	<i>S100B</i>	1.46	-0.232	*	1.685	*	-0.802		-0.681		0.030	ASTR
Spondin 2	<i>SPON2</i>	1.18	0.062	*	1.706	*	-0.695		-0.652		-0.421	ASTR
ATPase Na+/K+ Transporting Subunit Alpha 2	<i>ATP1A2</i>	0.19	0.240		1.515	*	-1.010		0.046		-0.791	ASTR
Protein arginine N-methyltransferase 3	<i>PRMT3</i>	1.37	0.938	*	1.175	*	-0.795		-0.312		-1.005	ASTR
hevin - produced by astrocytes controls excitatory synaptogenesis	<i>SPARCL1</i>	1.11	-0.349		1.782	*	-0.528		-0.541		-0.363	ASTR
Glutamate/Aspartate Transporter II	<i>SLC1A2</i>	4.57	-1.073		-1.061	*	0.347		0.841		0.946	GLUT
Vesicular glutamate transporter 1	<i>SLC17A7</i>	1.19	-1.094	*	-1.094	*	0.745		0.780		0.664	GLUT
vesicular glutamate transporter 2	<i>SLC17A6</i>	22.37	-1.062		-1.045	*	1.052	*	0.286		0.770	GLUT
Glutaminase 2	<i>GLS2</i>	1.22	-1.134	*	-0.875	*	0.271		1.296		0.442	GLUT
Glutamine synthetase	<i>GLUL</i>	4.68	-1.361		-0.547	*	1.258		0.491		0.159	GLUT
glutamate ionotropic receptor AMPA type	<i>GRIA1</i>	5.66	-0.831	*	-0.685	*	-0.429	*	0.355	*	1.590	GLUT
glutamate ionotropic receptor AMPA type	<i>GRIA2</i>	11.25	-1.084		-1.054	*	0.361		0.935		0.842	GLUT
glutamate ionotropic receptor AMPA type	<i>GRIA3</i>	3.23	-0.908	*	-1.167	*	0.855		1.021		0.198	GLUT
glutamate ionotropic receptor AMPA type	<i>GRIA4</i>	26.95	-1.073		-1.060	*	0.701	*	0.392		1.040	GLUT
glutamate ionotropic receptor NMDA type	<i>GRIN1</i>	1.14	-1.068	*	-1.068	*	0.650		1.052		0.434	GLUT
glutamate ionotropic receptor NMDA type	<i>GRIN2A</i>	2.35	-1.330		-0.722	*	0.368		0.560		1.124	GLUT
glutamate ionotropic receptor NMDA type	<i>GRIN2B</i>	1.40	-0.779		-0.691	*	-0.289		0.079	*	1.681	GLUT
glutamate ionotropic receptor NMDA type	<i>GRIN2D</i>	2.27	-1.056		-0.932	*	0.003		0.849		1.137	GLUT
glutamate ionotropic receptor NMDA type	<i>GRIN3A</i>	2.40	-0.786	*	-0.711	*	-0.340	*	0.187	*	1.651	GLUT
HuD - expressed in glutamatergic excitatory neurons	<i>ELAVL4</i>	95.17	-1.079	*	-1.072	*	0.766	*	0.427		0.958	GLUT
Synaptophysin	<i>SYP</i>	26.51	-1.052	*	-1.022	*	0.166		0.942		0.967	SYNAP
synapsin 1	<i>SYN1</i>	4.32	-1.048		-0.977	*	0.083	*	0.807		1.135	SYNAP
synapsin 2	<i>SYN2</i>	2.87	-0.933		-0.965	*	-0.013		0.525	*	1.386	SYNAP
Synaptotagmin 1	<i>SYT1</i>	382.58	-1.062	*	-1.089	*	0.922	*	0.813		0.417	SYNAP
Neurexin 3	<i>NRXN3</i>	23.04	-0.841		-0.932	*	0.353	*	-0.093	*	1.514	SYNAP
PSD95	<i>DLG4</i>	8.26	-1.006		-0.939	*	-0.075	*	1.188		0.831	SYNAP
PSD-localized scaffolding protein, interacts with Shank3. regulates mGluR	<i>HOMER1</i>	23.83	-0.972	*	-1.046	*	0.443	*	0.275		1.300	SYNAP
Neurexin 1, excitatory Synaptic marker	<i>NRXN1</i>	699.46	-1.092	*	-1.089	*	0.869	*	0.718		0.593	SYNAP
Neurexin 2, Synaptic marker	<i>NRXN2</i>	16.98	-1.008		-0.981	*	-0.019	*	1.052		0.956	SYNAP
gephyrin - inhibitory postsynaptic scaffold protein,	<i>GPHN</i>	203.09	-1.119		-1.065	*	0.665	*	0.861		0.657	SYNAP
calcium/calmodulin dependent protein kinase 1	<i>CAMK1</i>	2.34	-1.654		0.022	*	0.053		0.874		0.705	CAMK
calcium/calmodulin dependent protein kinase 2A	<i>CAMK2A</i>	1.46	-1.045	*	-0.940	*	0.183		1.329		0.473	CAMK
calcium/calmodulin dependent protein kinase 2B	<i>CAMK2B</i>	19.21	-1.033		-0.985	*	0.072		0.790		1.156	CAMK
calcium/calmodulin dependent protein kinase 2G	<i>CAMK2G</i>	6.31	-0.928	*	-0.643	*	-0.572	*	0.843		1.299	CAMK
calcium/calmodulin dependent protein kinase 2N1	<i>CAMK2N1</i>	247.46	-0.981		-0.772	*	0.100		0.097	*	1.556	CAMK
calcium/calmodulin dependent protein kinase 2N2	<i>CAMK2N2</i>	3.71	-0.909		-0.870	*	0.123	*	0.110	*	1.546	CAMK
calcium/calmodulin dependent protein kinase 4	<i>CAMK4</i>	3.30	-0.985		-0.914	*	0.150	*	1.442		0.307	CAMK
calcium/calmodulin dependent protein kinase 2D	<i>CAMK2D</i>	59.10	0.811	*	1.337	*	-0.817	*	-0.678		-0.653	CAMK
Microtubule Associated Protein 2	<i>MAP2</i>	331.71	-1.115		-1.056	*	0.912	*	0.541		0.718	OTHER
L1 cell adhesion molecule	<i>L1CAM</i>	7.66	-1.063	*	-1.042	*	0.292		0.742		1.071	OTHER
Microtubule Associated Protein Tau	<i>MAPT</i>	133.04	-1.081		-1.075	*	0.594		0.555		1.007	OTHER
NEUN - Neuronal Splicing Regulator	<i>RBFOX3</i>	20.85	-0.983		-0.972	*	0.128		0.466		1.362	OTHER
Tubulin Beta 3 Class III (TUJ1)	<i>TUBB3</i>	4.42	-0.883	*	-0.823	*	0.140	*	1.588		-0.023	OTHER
Neurofilament Medium	<i>NEFM</i>	283.61	-1.021		-0.987	*	1.322		0.376		0.310	OTHER
Neurofilament heavy	<i>NEFH</i>	1.16	-0.847	*	-0.693	*	-0.077	*	1.671		-0.055	OTHER
Neurofilament light	<i>NEFL</i>	900.58	-1.102		-1.033	*	0.596		0.471		1.068	OTHER
Neuroigin 1 - neuronal cell surface protein	<i>NLGN1</i>	81.62	-1.088		-1.089	*	0.570	*	0.716		0.891	OTHER
Transcription factor, regulate dendritic branching, spine	<i>CUX1</i>	188.64	-1.161	*	-0.987	*	1.021	*	0.567		0.560	OTHER
SATB Family Member 2	<i>SATB2</i>	5.07	1.277	*	0.892	*	-0.753		-0.755		-0.662	OTHER
maintenance of Purkinje neuron axons	<i>ELAVL3</i>	100.33	-1.093		-1.088	*	0.612	*	0.694		0.875	OTHER
neuronal cell adhesion molecule 1	<i>NCAM1</i>	337.66	-1.095	*	-1.091	*	0.836	*	0.651		0.699	OTHER
doublecortin - neuronal migration, microtubule stabilization	<i>DCX</i>	461.32	-1.084	*	-1.075	*	0.880	*	0.442		0.837	OTHER
enolase 2, Neuron-Specific Enolase	<i>ENO2</i>	42.25	-1.073	*	-1.014	*	0.494		0.407	*	1.187	OTHER

markers showed high expression in hiPSC-N, often with a significant trend for higher expression in the direction hiPSC-N15→hiPSC-N21→hiPSC-N21A, like the glutamatergic markers *GRIA1* and *GRIN3A* and the synaptic gene *NRXN3* and *CAM2N2*. This positive correlation of glutamatergic and synaptic gene expression with time post-Ngn2 induction and co-culture with hiPSC-A may suggest increasing maturation of the neurons across time and co-culture.

We next wanted to explore whether our differentiated cells, all of which come from a single cell line, are different or similar in expression profiles with cells derived from different cell lines. We compared our hiPSC-N to others we differentiated following the same protocol in our previous work [15–18] and with an in vivo dataset [19, 20] (Additional file 6: Sup. Figure 6). Similarly, we explored whether our astrocytes are similar with astrocytes differentiated by others with the same protocol or in vivo fetal and mature astrocytes [8, 21–27] (Additional file 7: Sup. Figure 7). With the neuronal cells we found strong correlations ($r > 0.8$) with all the bulk in vitro datasets and $r > 0.7$ with the in vivo bulk dataset (Additional file 6: Sup. Figure 6). With the cells differentiated to astrocytes, we also found strong correlations ($r > 0.74$) with all the bulk in vitro datasets and $r > 0.49$ with the in vivo bulk datasets (Additional file 7: Sup. Figure 7). This suggests that the genomic differences between cells and other unintentional differences in experimental procedures do not have a major effect on the transcriptomic signature of the cells.

We next harnessed the power of single-cell sequencing to explore the cellular homogeneity of these hiPSC-derived differentiated cells. Instead of specifying cell groups based on the culture condition (or visualization in the case of hiPSC-N21A vs. hiPSC-A N21) as above, we applied Louvain community detection, a method to extract communities with shared features from large networks [12], which identified 8 clusters within our cells (Fig. 1C). Five clusters included exclusively derived neurons—N-I to N-V—and 3 included derived astrocytes—A-I, II, and III. None of them is composed of a single condition (Fig. 1D). This suggests they are influenced by but do not solely depend on the different conditions and likely reflect a property of the base differentiation methods. Cluster A-III showed a biased composition by condition, containing >90% hiPSC-AN21. Comparing gene expression levels of genes in each neuronal cluster to the remaining 4 neuronal clusters identified numerous genes, differentially expressed. These, along with the astrocytic cluster comparisons, can be found in Additional file 8: SuppTable1. According to the statistical overrepresentation test of the PANTHER bioinformatics tool [28], they were functionally enriched for many neural development

and synaptic genes compared with the whole set of genes expressed in the 5 clusters (see Additional file 9: SuppTable2 for results on each cluster). Most enrichments were observed among genes expressed lower in examined cluster, while some were in both upregulated and downregulated genes. Similarly, comparing gene expression levels of genes in each astrocytic cluster compared with the remaining 2 identified numerous genes with enrichments compared with the whole set of genes expressed in the 3 clusters, in this case only in genes expressed lower in the examined cluster. These also included, among others, neuron-related functions (Additional file: 10: SuppTable3).

In summary, our observations support that iPSC-N and iPSC-A express marker genes that broadly suggest similarity to in vivo excitatory neurons and astrocytes as previous studies of these cells in bulk have shown [7, 8]. However, we do find heterogeneity which is influenced by the specific conditions of differentiation that we tested. This suggests that the heterogeneity is in part inherent to the differentiation protocols. Heterogeneity aside, the gene marker data suggests a role of the presence of neurons in the maturation of astrocytes and vice versa. Similarly, differentiation time also seems to play an important role for the maturation of the neurons.

GWAS genes expressed in iPSC-derived neurons and astrocytes

To be appropriate disease models, the cells created by these differentiation methods must express many of the genes associated with diseases involving neurons and astrocytes. We compared the genes showing differential expression (DE) between hiPSC-N and hiPSC-A (hiPSC-N and hiPSC-A specific genes) to those associated with neuropsychiatric illness by large genome-wide association studies (GWAS) and sequencing studies. We focused on schizophrenia (SZ) and Alzheimer's disease (AD) due to the availability of large GWAS [29, 30] and autism spectrum disorders (ASD) where large sequencing studies from the Simons Foundation Autism Research Initiative have identified many genes [31]. In the case of ASD, the data is from sequencing studies identifying clustering of damaging variants, providing a direct link to specific genes. In the case of SZ, the genes were reported as high confidence based on co-localization with brain eQTLs and posterior probability analyses [29]. Similarly, the AD GWAS assigned genes to associated variants based on colocalization analysis, fine-mapping results, and previous literature [30]. Using DESeq2 for differential expression analysis, we used a stringent threshold of an adjusted $p < 0.001$ (see the “Methods” section) to focus on the genes with the highest confidence of DE between cell types. We also used the highest confidence genes

reported for each disorder as follows: for ASD, this was 207 genes with a score of 1 (highest confidence) in the SFARI database; for SZ, this was 130 genes reported as genome-wide significant with high confidence [29]; for AD, this was 38 genes at loci showing genome-wide significant association [30].

Of the 15,157 genes in our dataset, 5154 were higher in hiPSC-N and 4107 in hiPSC-A at $FDR < 0.001$. From the 28 AD-associated genes present in our dataset, 7 were among the 5154 genes higher in hiPSC-N and 12 among the 4107 higher in hiPSC-A. For the genes higher in hiPSC-A, this is 1.6-fold more than expected (hypergeometric $p = 0.022$), while for those higher in hiPSC-N, it was 1.4-fold less than expected and also not significant. Out of 106 SZ-associated genes in our dataset, 56 were among those higher in hiPSC-N and 26 among those expressed higher in hiPSC-A. This is 1.55-fold more than expected for genes higher in hiPSC-N (hypergeometric $p = 2.1 \times 10^{-5}$) and as expected by chance for genes higher in hiPSC-A. Finally, out of 199 ASD genes in our dataset, 109 were among the genes higher in hiPSC-N and 39 among those higher in hiPSC-A, which is a 1.6-fold excess for hiPSC-N (hypergeometric $p = 5 \times 10^{-10}$) and a significant depletion (1.4-fold) for hiPSC-A (hypergeometric $p = 5 \times 10^{-6}$). These results are consistent with what is currently believed for these disorders; SZ and ASD have been genetically linked to neuronal functions [29, 32], while astrocytes have been implicated in AD [33]. The result also supports that these hiPSC-derived cells, while not equivalent to in vivo neurons and astrocytes, may be useful for modeling disease. The complete set of genes with their expression in neurons and astrocytes and the comparison of the two is in Additional file 11: SuppTable4.

Differences between hiPSC-N15 and hiPSC-N21

To determine the importance of differentiation under specific conditions and its possible relevance to disease genes, we also performed DE analysis between conditions. The hiPSC-N15 are a deviation from the original Ngn2 induction protocol which reported mature neurons at 21 days post-induction [7]. While we have not formally measured the differences, we have empirically observed little morphological change after day 15, so we decided to use the transcriptome to explore how hiPSC-N15 differ from hiPSC-N21. Shorter differentiation time not only has practical advantages, but there is a possibility that it may resemble an earlier developmental time (as supported by Table 1) perhaps more important to some diseases. The complete DE analysis results for all genes are in Additional file 12: SuppTable5.

At adjusted $p < 0.1$, 571 of 14,095 genes were expressed higher in hiPSC-N21 and 889 higher in hiPSC-N15.

Using the multiple testing corrected statistical overrepresentation test of the PANTHER bioinformatics tool [28] which allows comparisons to user-provided reference lists (in this case the list of all 14,095 genes) and biological processes annotations from the gene ontology database [34], we found among the genes expressed *higher in hiPSC-N15* significant enrichments (adjusted $p < 0.05$) for terms including “neurogenesis,” “neuron projection morphogenesis,” “cell morphogenesis involved in neuron differentiation,” and “regulation of neuron projection development” (see Additional file 13: Supp. Figure 8). Among the genes expressed *higher in hiPSC-N21*, we found significant enrichments for the GO terms “negative regulation of neuron death,” “regulation of neurotransmitter levels,” “chemical synaptic transmission,” “neuron differentiation,” and “nervous system process” (see Additional file 14: Supp. Figure 9). This suggests that genes in early neuronal development are expressed higher in hiPSC-N15, while genes involved in neuronal function are higher in hiPSC-N21 and that hiPSC-N15 are less mature compared with hiPSC-N21. This may suggest that, despite the artificial course of differentiation, hiPSC-N15 may more closely resemble neurons earlier in their course to maturity.

To gain insight into the importance of these genes in neurodevelopmental disorders, we intersected them with the 130 genes reported by the Psychiatric Genomics Consortium (PGC3 data) [29] as associated with SZ with highest confidence. Of these, 105 were present in the expressed gene list, and of those, 18 (17%) were significantly higher (7) or lower (11) in hiPSC-N21 (Additional file 15: SuppTable6). While this is not significant for each direction separately (each (hypergeometric $p \sim 0.06$), it is significantly (1.7-fold) more DE genes than expected by chance (hypergeometric $p = 0.01$). We further compared them to the list of 207 genes reported as high confidence for ASD by SFARI. Of these, 199 were present in our expressed gene list, and 12 were expressed higher in hiPSC-N21 (hypergeometric $p = 0.06$) and 22 lower (hypergeometric $p = 0.003$). Overall, there were 34 DE genes in either direction 1.65-fold more than expected by chance (hypergeometric $p = 0.001$) (Additional file 15: SuppTable6), observing that DE genes in both directions appear to be related to risk for ASD, and SZ is complicating the choice of differentiation time for modeling disease.

Differences between hiPSC-N21 and hiPSC-N21A

We then explored how co-culture with *hiPSC-A* affects the transcriptome of *hiPSC-N*. The complete transcriptome comparison results for all genes are in Additional file 16: SuppTable7. Since hiPSC-N21A neurons were grown in the same plate as hiPSC-AN21 astrocytes, to

avoid artifacts due to possible contribution of cell-free RNA (also termed “ambient RNA”) from lysed cells in the co-culture media that could be sequenced and confound results [35], we excluded from this comparison genes that were expressed markedly higher in hiPSC-A than hiPSC-N at adjusted $p < 0.001$ (“high hiPSC-A genes”). Note that Additional file 16: SuppTable7 contains all results for all genes detected, with those excluded marked accordingly.

Overall, the expression of 359 out of 11,222 genes included in the analysis was *higher in hiPSC-N21A than hiPSC-N21* at adjusted $p < 0.1$. PANTHER bioinformatics showed a 3.3-fold enrichment for genes involved in “regulation of metal ion transport” and 1.5-fold enrichments for “cell communication,” “signal transduction,” and “signaling” (Additional file 17: Sup. Figure 10). Five hundred ten genes were expressed significantly *higher in the absence of astrocytes* (hiPSC-N21) at adjusted $p < 0.1$ out of 14,857 included in the analysis. There was a 7.1-fold enrichment for “central nervous system neuron axonogenesis,” a 3.2-fold enrichment for “axon guidance,” 2.8-fold for “axon development, 2.1-fold for neuron development, and 1.6-fold for “cell differentiation” (complete results in Additional file 18: Sup. Figure 11). The enrichments for axon development and guidance and neuron development, including genes like SEMA4D and DCC [36], the CRMP5-encoding DPYSL5 [37], and the SLIT-ROBO Rho GTPase-activating protein SRGAP1 [38], suggest an earlier stage of development for hiPSC-N21 and support the importance of the inclusion of astrocytes in maturation.

We compared these genes to the list of 130 high confidence SZ-associated genes reported by the PGC [29]. In contrast to the overlap with genes differing between hiPSC-N15 and hiPSC-N21, here, only 5 of these genes were among those higher in hiPSC-N21A (*MSI2*, *CUL9*, *CSMD1*, *OPCML*, and *DCC*) and 4 among those higher in hiPSC-N21 (*MAPK3*, *NXPH1*, *IL1RAPL1*, *GALNT17*). Similarly, when compared with the ASD genes, only 8 were among those higher in hiPSC-N21A and 11 among those higher in hiPSC-N21, not significantly more than expected. This could suggest that disease genes might not be among those impacted by the co-culture of neurons and astrocytes or could be due to reduced power.

Differences between hiPSC-A0 and hiPSC-AN21

We next explored how co-culture with *hiPSC-N* affects the transcriptome of *hiPSC-AN*. The complete transcriptome comparison results for all genes are in Additional file 19: SuppTable8. Similarly to the reverse comparison, to avoid confounding from cell-free RNA from hiPSC-AN21 lysed cells to droplets containing hiPSC-N21A, we excluded genes that were significantly higher in hiPSC-N compared with hiPSC-A (at adjusted $p < 0.001$). Note that

our Additional file 19: SuppTable8 contains all results for all genes detected, with those excluded marked accordingly.

In all, out of 10,237 genes included in the analysis after the removal of the “high hiPSC-N” genes, 583 were significantly *higher in hiPSC-AN21* than hiPSC-A0 at adjusted $p < 0.1$, and PANTHER bioinformatics showed multiple significant functional enrichments (Additional file 20: Sup. Figure 12). Notably, we observed 2.3- to ninefold enrichments for functions including “regulation of superoxide metabolic process,” “cellular oxidant detoxification,” and “response to oxidative stress,” all important functions of astrocytes in their supportive roles nervous system [39]. Due to the importance of astrocytes in AD [40], we also looked for AD GWAS genes for enrichments. Twenty-one AD-associated genes were present in the reference list of genes in our comparison, and 3 of them were among the 583 significantly higher in hiPSC-AN21 (*APOE*, *CLU*, and *CASS4*), compared with 1.2 expected by chance (hypergeometric $p = 0.02$). While this is a small number of genes, it is important to know that studying them using in vitro differentiated astrocytes might benefit from the inclusion of neurons in the cultures. This is particularly important for *APOE* which is a very widely studied AD gene.

When it comes to genes that were *higher in hiPSC-A0* than hiPSC-AN21 out of 14,912, there were 1071 at adjusted $p < 0.1$, and PANTHER bioinformatics analysis also showed multiple significant functional enrichments, shown in Additional file 21: SuppTable9, and those with more than threefold enrichments are also illustrated in Additional file 22: Sup. Figure 13. Most striking were more than eightfold enrichments for immunity related genes. Of those directly relevant to neural cells, there were multiple categories involving neuron development and neural tube closure as well as axon development and guidance.

Among 27 AD-associated genes in the tested set of genes, there were 4 AD-associated genes higher in hiPSC-A0 vs. hiPSC-AN21 (*GRN*, *PICALM*, *APH1B*, and *CD2AP*), a 2.1-fold excess from expected (hypergeometric $p = 0.04$). We observe again excess occurrence of disease genes on both sides of the distribution, suggesting that the better choice of cells for disease modeling might be gene specific.

The top 50 genes in each cell type contrast, detailed in the “Differences between hiPSC-N15 and hiPSC-N21,” “Differences between hiPSC-N21 and hiPSC-N21A,” and “Differences between hiPSC-A0 and hiPSC-AN21” sections, are shown in heatmaps in Additional file 23: Sup. Figure 14. In order to further explore the similarity of these in vitro cells to multiple in vivo lineages, we compare the expression data for the top differentially

expressed genes in our in vitro derived cells to data from additional studies of related in vivo cell types [41–46] and to directed differentiation in vitro [47, 48].

Comparisons with in vivo datasets

Having used individual a priori known marker genes and differential expression to show that hiPSC-N and hiPSC-A express many of the expected markers for their intended cell type but also show heterogeneity within cell type, we then compared these cell types and their sub-clusters to in vivo cells in two previously reported datasets by Darmanis et al. [41, 44] and Fan et al. [19, 20].

Darmanis et al. [41] performed single-cell transcriptome analysis on adult and human fetal brain. Fan et al. [41] performed single-cell transcriptome profiling of cells from the four cortical lobes and pons during human fetal development from the 7th to the 28th gestational week (GW). We first performed a MetaNeighbor analysis [49] across all genes in the three datasets to explore similarities across cell types in the different experiments. This analysis is based on a statistical framework that quantifies the degree to which cell types replicate across datasets [49]. The complete heatmap is shown in Additional file 24: Sup. Figure 15, while Fig. 2A includes the portion comparing our in vitro derived cells to the in vivo cell types. As expected, the Ngn2-induced neurons are closest to the in vivo neuronal cell types, including both cortical and pontine. hiPSC-N21A neurons are closest to Darmanis et al. adult neurons but also close to Fan et al. late gestation pontine neurons. Interestingly, the Darmanis et al. adult neurons are also close to Fan et al. pontine neurons (Additional file 24: Sup. Figure 15). hiPSC-N21 are also closest to Darmanis et al. adult neurons, with the next closest neighbors being Fan et al. 9–12 gestational week (GW) excitatory cortical neurons and GW 9–14 pontine neurons.

hiPSC-N15 neurons are closest to Darmanis et al. fetal neurons and Fan et al. early pontine neurons with the next best neighbor being Fan et al. early cortical neurons. This similarity to early in vivo neuronal states suggests that cells at early time points in Ngn2 induction recapitulate an earlier neuronal maturation state despite forced differentiation by Ngn2 which bypasses progenitor states. This is particularly interesting in view of our observation that genes with increased expression in hiPSC-N15 contain the strongest excess of neurodevelopmental disorder risk genes (see the “Differences between hiPSC-N15 and hiPSC-N21” section).

Both hiPSC-A0 and hiPSC-AN21 were similar to the astrocytes of both Darmanis et al. and Fan et al. but were also similar to the endothelial and immune cells of both studies (Fig. 2A). While this was a surprising result, the astrocytes from both in vivo studies also showed strong

similarities to these cell types (Additional file 24: Sup. Figure 15), suggesting this might not be an irregularity of the hiPSC-A0 and hiPSC-AN21 but rather a property of these cell types in vivo as well.

To further dissect these relationships at the single-cell level, we performed Seurat CCA-based integration analysis on our cells along with cells from the Fan study. This integration analysis was carried out separately within the neuronal and then non-neuronal cell types to focus on specific lineage relationships. In the neuronal analysis (Fig. 2B, C), separation across the first UMAP dimension placed our neurons between excitatory cortical and PRPH-expressing pontine neurons in vivo, suggesting that Ngn2-induced neurons harbor elements of both cortical excitatory neuronal and more posterior or sensory neuronal identities. This is consistent with recent studies of scRNA-seq data in Ngn2-induced neurons which concluded that Ngn2-induction produces PRPH-expressing sensory neurons [50, 51]. Interestingly, the second UMAP dimension aligned with increasing maturity for both in vivo pontine and cortical neurons as well as our in vitro derived neurons (hiPSC-N15→hiPSC-N21→hiPSC-N21A). This further supports our conclusion that longer time from induction and culture with astrocytes promotes neuronal maturation in this system.

The non-neuronal integration analysis of our hiPSC-A data with data from Fan et al. (Fig. 2D, E) also indicated similarity to multiple in vivo cell types. All the hiPSC-A0 and the majority of the hiPSC-AN21 cells clustered near each other and were surrounded by three clusters of vascular and endothelial smooth muscle cells. Another cluster of hiPSC-AN21 cells (those in Louvain cluster A-II from Fig. 1D) was more proximal to in vivo astrocytes, suggesting that the specific subpopulation of hiPSC-A cells grown in co-culture with induced neurons and identified in cluster A-II achieves higher resemblance to in vivo astrocytes. To confirm this possibility, all the hiPSC-A cells were re-clustered alone (Additional file 25: Sup. Figure 16). One of the new resulting hiPSC-A clusters (A-3) is enriched in hiPSC-A cells co-cultured with neurons and in both the integrated UMAP and in a new MetaNeighbor analysis shows more similarity to in vivo astrocytes.

Comparisons of Ngn2-induced neurons with additional in vitro scRNA-seq datasets

Together with recently published scRNA-seq data in the Ngn2-induction system [50–53], these results indicate that while Ngn2-induced neurons are excitatory neurons, they share transcriptional elements with multiple in vivo neuronal lineages. To explore the reproducibility of this complex, induced neuronal phenotype, we performed a third integration analysis bringing together our

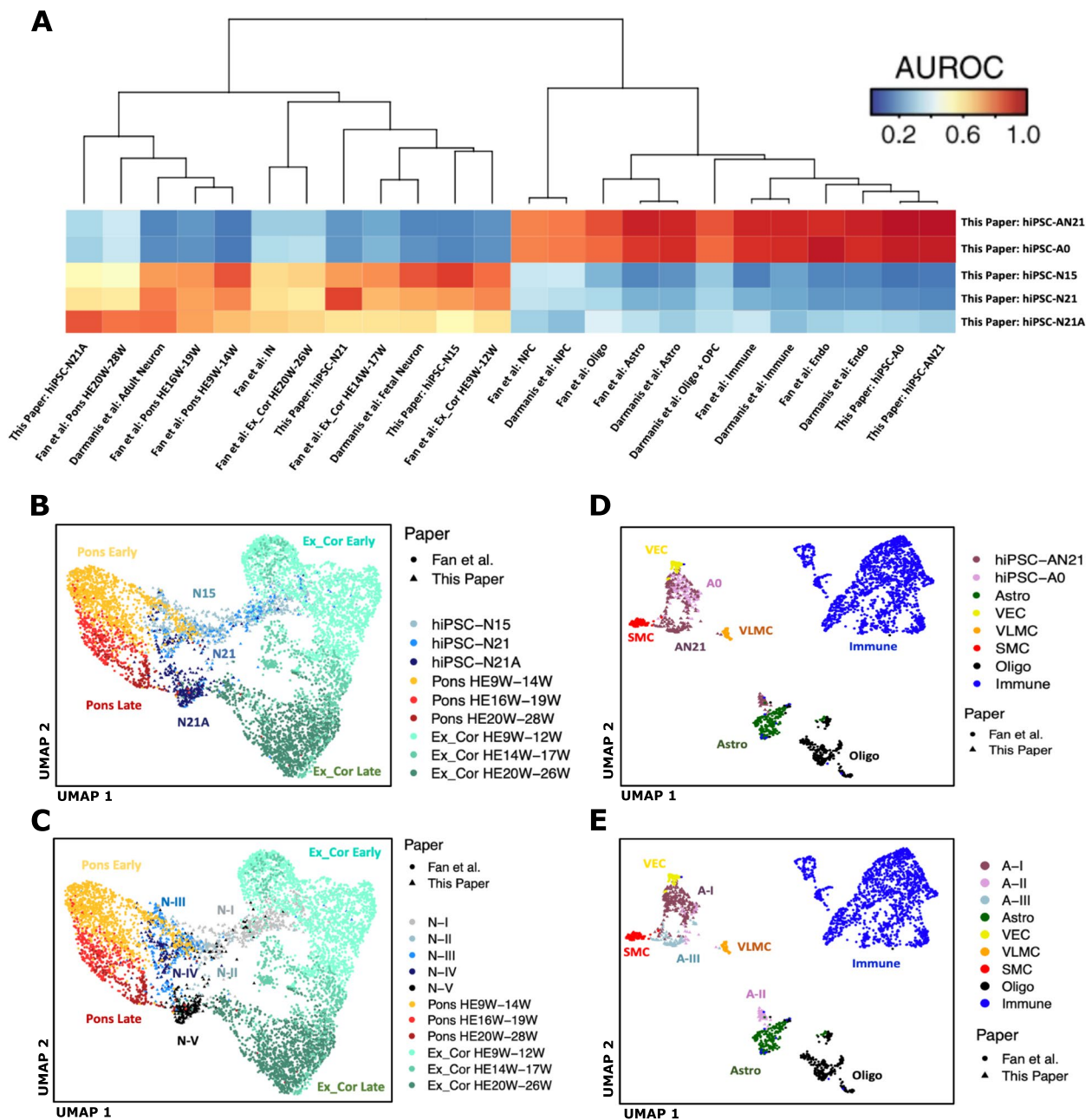


Fig. 2 **A** MetaNeighbor analysis of our bulk and pseudobulk expression data with in vivo data from two in vivo studies. Ex_Cor, excitatory cortical; HEW, human embryo week; Astro, astrocytes; Oligo, oligodendrocytes; OPC, oligodendrocyte precursor cells; Endo, endothelial. **B, C** Seurat integration analysis of our hiPSC-N cells with neuronal cells from two in vivo datasets. **B** our cells colored by condition. **C** our cells colored by Louvain cluster. Ex_Cor, excitatory cortical; HEW, human embryo week. **D, E** Seurat integration analysis of our hiPSC-A cells with non-neuronal cell from two in vivo datasets. **D** Our cells colored by condition. **E** Our cells colored by Louvain cluster. SMC, smooth muscle cells; VEC, vascular endothelial cells; VLMC, vascular leptomenigeal cells. Astro, astrocytes; Oligo, oligodendrocytes

Ngn2-induced neuronal data with recent scRNA-seq data (Fig. 3). Figure 3 depicts the integrated UMAP colored by the three data-driven cell clusters (Fig. 3A) and also colored by weeks of Ngn2 induction (Fig. 3B) across the three studies. Ngn2-induction time points showed

progressively shifting abundance across the cell clusters (Fig. 3C): ~80% of week 2 cells but only ~20% of week 5 cells are in cluster 0 and inversely <5% of week 2 cells but >50% of week 5 cells in cluster 1. Cells from all three studies were distributed across the cell clusters (Fig. 3D),

suggesting an array of reproducible neuronal end points in Ngn2-induction across laboratories. To further support this notion, Fig. 3E shows the expression of marker genes used to distinguish Ngn2-induced neuronal subpopulations in recent scRNA-seq studies of Ngn2-induced neurons [50]. GPM6A is expressed throughout the developing brain and spinal cord [54] and marks a population of neurons distinct from the PRPH+ induced neurons. The GPM6A⁺ cluster also contained cells that expressed TAC1, consistent with findings from a recent scRNA-seq study [34]. PRPH is expressed in peripheral nervous system (PNS) and CNS neurons projecting to the periphery [55], while PHOX2B, which is expressed in a subpopulation of PRPH+ cells, is expressed specifically in the posterior CNS, in the hindbrain and spinal cord. POU4F1 is expressed within the PRPH+ neurons in a population distinct from the PHOX2B+ cells. The PRPH⁺PHOX2B⁺ cluster also expressed GAL and SSTR2, while the PRPH+POU4F1+ cluster expressed GAL and PIEZO2, replicating the cell populations found in recent studies [33, 34]. These results indicate that although the Ngn2-induced neuronal state appears to include a complex combination of in vivo transcriptional programs, it is reproducible across individual induction experiments and different laboratories.

To more deeply explore the reproducible transcriptional elements of Ngn2-induced neurons, we performed another MetaNeighbor analysis to define the relationship of in vivo neurons to Ngn2-induced neurons at different points during induction and across labs (Fig. 3F). In all three Ngn2 studies, Ngn2-induced neurons at week 2 post induction more closely resemble early cortical excitatory neurons than induced neurons at later time points. This more precise recapitulation of the early in vivo cortical neuronal lineage may underlie the increased enrichment for neurodevelopmental disease gene risk that we observed above (hiPSC-N15 neurons in the “Differences between hiPSC-N15 and hiPSC-N21” and “Comparisons with in vivo datasets” sections). Also consistent across all three studies, later points in Ngn2 induction more closely resemble the transcriptional identity of pontine neurons. Since these are non-dividing cells, the differences between time points most likely represent temporal lineage dynamics in the Ngn2 induction system and

may have considerable impact when using these cells in disease modeling and therapeutic development.

Comparisons of hiPSC-A cells with additional iPSC-derived astrocyte RNA-seq data

To conduct an examination of the reproducibility of cell fates in iPSC-derived astrocytes, we compared pseudo-bulk expression from our scRNA-seq data in our hiPSC-A cells to bulk RNA-seq of additional in vitro hiPSC-derived astrocytes in addition to pseudo-bulk expression from scRNA-seq of in vivo cell types (Fig. 4). This additional bulk RNA-seq data came from the study from which we derived the astrocyte differentiation protocol used here and included RNA-seq data from iPSC-derived astrocytes as well as in vitro cultured primary astrocytes (Tcw et al. 2017 [8]). Correlation of expression data from our hiPSC-A cells to iPSC-derived astrocytes (iAstro) from the Tcw et al. study was high (Spearman correlation range 0.67–0.74)—the highest of any cross-study correlations here, suggesting that the composite signature of hiPSC-A single cells generally reflects the iPSC-derived astrocytes generated using the same protocol. Expression data in our hiPSC-A cells was also highly correlated with both in vivo astrocytes (0.52–0.67) and other non-neuronal cell types (0.46–0.67)—again indicating the broad similarity of the hiPSC-A cells to other non-neuronal lineages. This was also true for the Tcw et al. iPSC-derived astrocytes: in vivo astrocytes (0.58–0.70) and other non-neuronal cell types (0.53–0.71). Remarkably, in vitro iPSC-derived astrocytes from both this and the Tcw et al. study resemble in vivo astrocytes to nearly the same degree (0.52–0.67 and 0.58–0.70 respectively) as cultured primary astrocytes from the Tcw et al. study (0.59–0.74). This may indicate that in vitro culture itself results in a significant loss of in vivo cell identity, as has been observed in microglial cells [56]. Similar to the observation in the MetaNeighbor analysis in Additional file 24: Sup. Figure 15, correlation across in vivo non-neuronal cell types is high across study (0.48–0.75) and even higher within study (0.70–0.88 within Fan; 0.44–0.68 within Darmanis). This again indicates that some of the correlation of the hiPSC-A cells to other lineages may be in line with in vivo expression patterns. Again,

(See figure on next page.)

Fig. 3 Integration analysis of our Ngn2-induced neuronal data with other single-cell data in Ngn2-induced neurons. **A** UMAP of Seurat-integrated data from our study and two other scRNA-seq studies of Ngn2-induced neurons (Schornig et al. and Lin et al.), colored by cell cluster. **B** Same UMAP colored by weeks of Ngn2 induction. **C** Proportion of cells from different time points in each cell cluster. **D** Cells from individual studies visualized in the same UMAP as in **A** and **B**, again colored by weeks of Ngn2-induction. **E** Expression of marker genes used to delineate diversity in the lineage composition of Ngn2-induced neurons. **F** MetaNeighbor analysis of in vitro cells and in vivo cell types (both divided by time points and by study; W, weeks of Ngn2 induction or gestational week; HE, human embryo, individual cell lines used are also indicated)

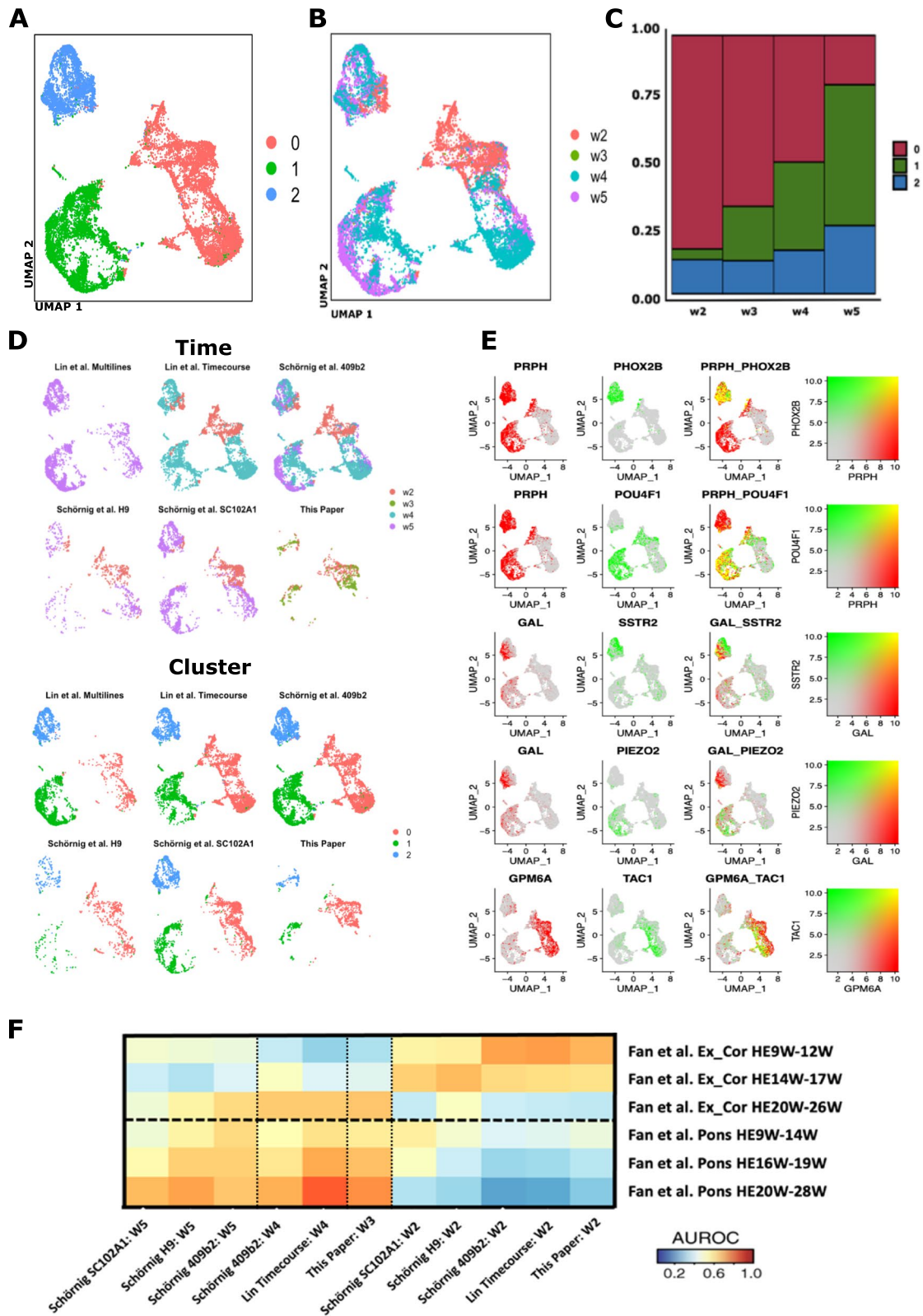


Fig. 3 (See legend on previous page.)

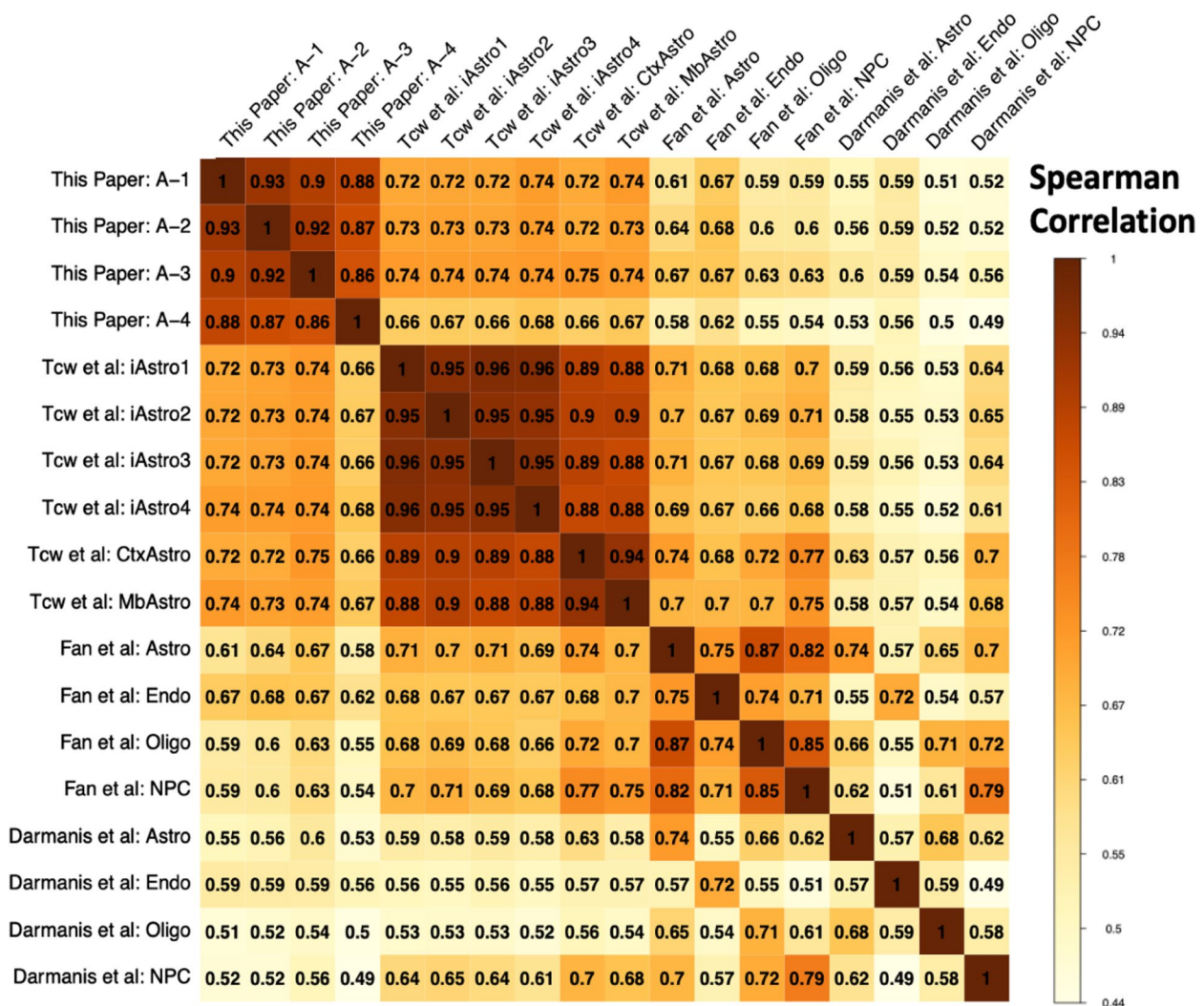


Fig. 4 Correlation matrix of genome-wide pseudo-bulk expression data from hiPSC-A cells, bulk RNA-seq data from additional in vitro iPSC-derived astrocytes (Tcw = Brennard) and in vivo cell types (Fan and Darmanis). iAstro1-4, in vitro iPSC-derived astrocytes; CtxAstro & MbAstro, cultured primary cortical and midbrain astrocytes. Only the 3000 most highly variable genes were used in the calculation of the correlation coefficients

consistent with the previous UMAP and MetaNeighbor observations, of the 4 hiPSC-A sub-clusters, cluster #3 (Additional file 25: Sup. Figure 16), which is enriched in hiPSC-AN21 cells, showed this highest correlation to in vivo astrocytes (0.67 in Fan, and 0.6 in Darmanis).

Both the differential expression analysis (the “Differences between hiPSC-N15 and hiPSC-N21,” “Differences between hiPSC-N21 and hiPSC-N21A,” and “Differences between hiPSC-A0 and hiPSC-AN21” sections) and these comparisons to in vivo data (the “Comparisons with in vivo datasets” section) indicate that the in vitro induced cell types contain transcriptional elements of both intended lineages and off-target lineages. The heterogeneity in cellular identity produced by these protocols must be considered with attention

when employing them to model human brain disease and for therapeutic assay development.

Discussion

Our single-cell analysis across post induction time and culture conditions showed that hiPSC-N and hiPSC-A could be separated based on their transcriptome (Fig. 1) and in aggregate expressed many appropriate cell type markers (Fig. 2). Both hiPSC-N and hiPSC-A however exhibit heterogeneity and could be divided to sub-clusters that were influenced but not determined by the different conditions we tested (with the near-exception of cluster A-III). Ngn2-induced neurons do not represent a singular in vivo neuronal cell type but rather express transcriptional characteristics of both

cortical excitatory neurons and more posterior neuronal fates as well as markers of other diverse neuronal subtypes. Importantly, these signatures are reflected in discrete subpopulations, whose proportions change over the course of in vitro differentiation. While this is consistent with recent in-depth single-cell analyses that concluded Ngn2-induced neurons take on specific sensory fates [57], we find that Ngn2-induction produces a type of neuronal cell (or a mix of cells) that includes transcriptional elements of multiple in vivo neuronal cell types. So, while their neuronal identity likely makes them a good model for studying brain disorders, one still needs to be cautious in reaching conclusions and further work would be helpful in refining differentiation protocols. When it comes to hiPSC-A, we find that they take on one or a mix of transcriptional states resembling several in vivo cell types (including endothelial cells, immune cells, and astrocytes). These cell types however also appear to have similarities in the vivo datasets (Additional file 24: Sup. Figure 15). Co-culture with neurons pushes a subpopulation of the in hiPSC-A toward a state more similar to bona fide in vivo astrocytes; however, even with co-culture, only some of the cells take on an astrocytic identity, so caution is necessary when studying them in bulk.

Unique aspects of our study include the examination of two post Ngn2 induction time points, the co-culture of hiPSC-N with hiPSC-A, and the integration with in vivo datasets. This allowed us to show that both the longer post-induction culture time of the Ngn2-induced neurons, and the inclusion of hiPSC-A contributed to expression profiles closer to mature neurons. However, the additional post-induction time in hiPSC-N appeared to favor more posterior fates over cortical fates. The increased neuronal maturity was observed not only in terms of expression of neuronal markers (Table 1) but also in terms of the functions of the DE genes and of similarity to in vivo neurons at different maturity states (Figs. 2 and 3). We made the same observation for hiPSC-A, where the co-culture with hiPSC-N21 also appeared to increase their maturity, and a subset of the hiPSC-AN21 distinctly co-clustered with human in vivo astrocytes. The differences in astrocytic marker-gene expression between hiPSC-A conditions were often pronounced (most astrocytic genes in Fig. 2) and included *APOE*, a very important astrocytic gene in the study of AD. In contrast to this observation though, the excess of genes associated with AD was observed among those expressed lower, not higher, in hiPSC-AN21 compared with hiPSC-A0. It is therefore unclear whether one should prefer the co-cultured neurons for modeling AD, a decision that should probably be made on a gene and experiment specific basis.

Our comparisons with in vivo datasets showed that the Ngn2-induced neurons are similar to in vivo excitatory neurons, with similarity to both cortical and pontine in vivo neurons. Assuming that the changes are not due to competition, which is unlikely in these non-dividing cells, we find it particularly interesting that the trajectory from hiPSC-N15 to hiPSC-N21 to hiPSC-N21A was along the same axis with the pontine and cortical neurons development during fetal life (Fig. 3), with the hiPSC-N15 being closer to fetal than adult neurons, which has significant implications for the use of induced neurons for the study of disease. The comparison of the hiPSC-A0 and hiPSC-AN21 with in vivo datasets confirmed their similarity to in vivo astrocytes but additionally showed strong similarities to immune and endothelial cells. However, the astrocytes in both in vivo datasets were also similar to the immune and endothelial cells. It appears that this maybe an inherent property of these cell types.

Our search for overrepresentation in GWAS genes was triggered by the main goal of this paper, to explore the use of an in vitro differentiation system for the study of psychiatric disorders, specifically neurodevelopmental (SZ, ASD) and neurodegenerative (AD) diseases. When it comes to neuron versus astrocyte-predominant genes, those expressed higher in neurons contained an excess SZ- and ASD-associated genes showing they are a useful platform to model these disorders. A suggestive excess (which could be due to lack of power) was seen in astrocytic genes for AD associations, and interestingly this was significant for genes expressed higher in hiPSC-A0 than hiPSC-AN21. Regarding finer distinctions based on days post-induction and co-culture, we found that DE genes in both directions showed high content of neurodevelopmental disorder genes, which in the case of ASD was also significant specifically for genes expressed higher in hiPSC-N15 than hiPSC-N21. The induction by Ngn2 is far from the normal course of differentiation and maturation of neurons in vivo, yet this along with the similarity of hiPSC-N15 to fetal neurons suggests that hiPSC-N15 may also be a good choice to model for SZ and ASD complementing hiPSC-N21.

An important consideration about this study is that our analyses were conducted on a single differentiation experiment with a single donor. While comparisons with our previous experiments involving different donors do not show any significant deviations as described in the comparisons we performed above, the conclusions must be considered with caution, especially when one considers individual genes, as experimental variation and/or genomic profile

differences between individuals might make such findings experiment or individual-specific.

Conclusions

We have harnessed the power of single-cell sequencing and iPSC differentiation to successfully gain important insights for disease modeling. While more studies are required, we anticipate that this study will be an important additional guide for navigating the modeling of complex brain disorders and improving differentiation protocols to achieve the optimal disease models.

Methods

Induced pluripotent stem cell (iPSC) culture and maintenance

Human BC1, an iPSC line which was obtained from Dr. Linzhao Cheng's lab at Johns Hopkins School of Medicine, was used in the study. This is an established cell line with published results [58]. Cells were cultured in StemFlex media (Gibco) on 6-well tissue culture plates coated with laminin (Biolamina). Cells were dissociated with StemPro Accutase (Gibco) into single-cell suspension and seeded in required density for the experiment. The ROCK inhibitor Y-27632 dihydrochloride (Tocris) was added on the first day of passage at a concentration of 10 μ M. Cultured cells were tested to ensure they lack mycoplasma contamination.

Ngn2 lentivirus transduction

Ngn2 and rTTA virus were procured from the University of Pennsylvania Core store. This virus has been previously reported to be successfully used for induced neural differentiation by our (Avramopoulos) laboratory [59]. It was initially reported by the Sudhof laboratory [7] who first discovered that forced expression of this single transcription factor Ngn2 can convert iPSCs into functional neurons with very high yield in 21 days. 250,000 BC1 iPSC cells were plated in each well of a 6-well plate and grown in Stem Flex media supplemented with ROCK inhibitor. Ngn2 lentiviral infection using polybrene (Santa Cruz) was done 24 h post seeding. Briefly, cells were fed with 2 ml of fresh media, and 2 μ l of polybrene (1 μ g/ml) stock was added per well. To attain a MOI of 1–10, different volumes of both the Ngn2 and rTTA virus were added per well. In 4 of the 6 wells, leaving one as a polybrene-only control, the following amount of each virus was added: 3 μ l, 5 μ l, and 10 μ l. The virus infected cells were expanded and frozen stocks made for future differentiation. We selected for the 10 μ l transduced Ngn2-BC1 cells for optimal neural differentiation.

Neuronal differentiation of Ngn2-transduced iPSCs

250,000 Ngn2-transduced BC1 cells were plated on laminin coated 6 well plates (DIV – 2). Cells were fed with fresh Stem Flex media the next day (DIV – 1). Ngn2 expression was induced by doxycycline on DIV 0 using an induction media consisting of DMEM/F12 (Thermo Fisher), N2 (Thermo Fisher), D-glucose (Thermo Fisher), 2- β ME (Life technologies), Primocin (InvivoGen), BDNF (10 ng/ml, PeproTech), NT3 (10 ng/ml, PeproTech), laminin (200 ng/ml, Millipore Sigma), and doxycycline (2 μ g/ml, Sigma). A puromycin selection was done on these cells on DIV 1, 24 h post doxycycline induction using the same induction media supplemented with puromycin (5 μ g/ml). Surviving cells were harvested on DIV 2 and plated on Matrigel-coated 24-well plates at a concentration of 100,000 cells/well in neural differentiation media consisting of neurobasal media (Thermo Fisher), B27 (Thermo Fisher), Glutamax (Thermo Fisher), Penn/Strep (Thermo Fisher), D-glucose (Thermo Fisher), BDNF (10 ng/ml), NT3 (10 ng/ml), laminin (200 ng/ml), and doxycycline (2 μ g/ml). Cells were fed with a 50% media exchange of neural differentiation media every other day till DIV 12. Cells were treated with 2 μ M cytosine β -D-arabinofuranoside hydrochloride (Ara-C) on DIV 4 to arrest proliferation and eliminate non-neuronal cells in the culture. Doxycycline induction was initiated at DIV 0 and continued till DIV 12 after which it was discontinued and cells were fed every 2 days thereafter till DIV 21 with neural maturation media consisting of neurobasal media A (Thermo Fisher), B27, Glutamax, Penn/Strep, Glucose Pyruvate mix (1:100, final conc of 5 mM glucose and 10 mM sodium pyruvate), BDNF (10 ng/ml), NT3 (10 ng/ml), and laminin (200 ng/ml). Neurons were harvested by DIV 15 or 21. Four conditions were set up for this experiment which are as follows: (i) neurons only (DIV 21), (ii) neurons only (DIV 15), (iii) astrocytes only, (iv) neurons and astrocyte co-culture. Astrocytes cultured in FBS were added on top of the differentiating cells on DIV 5 at a concentration of 50,000 cells/well in the co-culture experiment. Two micrometers of Ara-C treatment was repeated on DIV 7 for the co-culture experiment, and media changed every other day thereafter till DIV 21. For the astrocyte-only condition, astrocytes were seeded at 50,000 cells per well of a 24-well plate and fed with neural differentiation media and allowed to grow till 80% confluent (48 h) before adding 2 μ M AraC. Media was changed every other day till DIV 21 when neurons are ready to harvest. Neurons were collected using Accutase and passed through a cell strainer and counted to receive the optimal number of cells. Immunofluorescence images of the Ngn2 induced neurons can be seen in

Additional file 26: Sup. Figure 17 and our previous publications [59–61].

Neural differentiation of hiPSCs via embryoid body (EB) formation

Neural differentiation of embryoid bodies (EBs) was performed as previously described [62] with modifications. Briefly, EB formation was performed by the forced aggregation method. To this goal, PSC lines were cultured in feeder-free conditions as monolayers with E8 medium and passaged every 3 days with TrypLE. For the production of uniform-size EBs, iPSCs grown for 3–10 passages were counted and seeded at 5000 cells per well in 96-well, V-bottom uncoated plates (249,952; NUNC, Rochester, NY). For induction of neural differentiation, EBs were grown in suspension for 7–8 days followed by adherence to Matrigel-coated plates in the Neural Induction Medium (NIM) consisting of DMEM/F12 (GIBCO, 11,320,033), 2 mM l-glutamine, 0.1% bovine serum albumin (Fraction V; Sigma-Aldrich), 1% NEAA, 2% B27 without retinoic acid (GIBCO), 1% N2 supplement (GIBCO), LDN193189 (PeproTech) throughout culture, and 10 μ M SB431542 (Tocris Bioscience, Bristol, UK). Numerous rosette structures were formed 2–3 days after the adherent culture of EBs.

Isolation and culture of neural precursor cells

Neural rosettes were manually collected with stretched glass Pasteur pipettes and expanded as monolayer cultures of neural precursors (NPCs). Briefly, EB-derived neural rosettes were dissociated into single cells with Accutase for 5 min at 37 °C and plated on Matrigel or polyornithine/laminin-coated plates in the NIM complete medium supplemented with FGF2 (10 ng/mL) and epidermal growth factor (EGF) (10 ng/mL; PeproTech, Rocky Hill, NJ). Cells were expanded for several passages as a homogeneous population of NPCs.

Astrocytic differentiation

Human BC1iPSC line was differentiated into astrocytes as previously described [8]. Briefly, NPCs dissociated to single cells were seeded at 15,000 cells/cm² density on Matrigel coated plate in complete astrocytic differentiation medium (ScienCell Research Laboratories cat. No 1801), astrocyte medium (ScienCell Research Laboratories cat. No 1801-b), 2% fetal bovine serum (ScienCell Research Laboratories cat. No 0010), and astrocyte growth supplement (ScienCell Research Laboratories cat. No 1852). The cells passaged in this density for the first 30 days and fed every other day. Following this period, the astrocytes could be passaged in a 1:3 ratio and expanded for up to 120 days in the same medium. Immunofluorescence images of astrocytes are shown in

Additional file 26: Sup. Figure 17, along with co-culture with neurons.

Single-cell sequencing

Six wells in a 24-well plates of neurons were grown for 15 days post Ngn2 induction, 6 wells in a 24-well plates of neurons were grown for 21 days post Ngn2 induction, and 6 wells in a 24-well plates of neurons were grown for 21 days post Ngn2 induction with the addition of astrocytes on day5 at a density of 50,000 cells/well. After dissociation with Accutase, single-cell suspensions for 10 \times libraries were loaded onto the 10 \times Genomics Chromium Single Cell system using the v2 chemistry per manufacturer's instructions [35, 63]. Estimations of cellular concentration and live cells in suspension was made through trypan blue staining and use of the Countess II Cell Counter (Thermo Fisher). Dissociated single iPSCs were passed through a 40- μ m filter and used as input for the 10 \times chromium v2 3' gene expression kit (10 \times genomics), targeting 1000 cells per sample. Libraries were prepared according to the manufacturer's instructions and uniquely indexed. Libraries were quantified on the Nanodrop platform and sized using the Agilent 2100 Bioanalyzer RNA nano system. Barcoded libraries were pooled and sequenced on an S1 flowcells on a NovaSeq 6000 (Illumina) to an average depth of $\sim 1.33 \times 10^8$ ($\pm 3.92 \times 10^7$) paired-end reads per sample. Raw reads were pseudoaligned to the Gencode reference human transcriptome (v31; www.gencodegenes.org/human/) using kallisto (default parameters plus -t 4) and collapsed to individual UMIs using bustools correct (default parameters plus -t 4; 10 \times v2 whitelist) and bustools count (default parameters plus -t 4). Cells were filtered from empty droplets using estimated knee plot inflection point UMI cutoffs (DropletUtils) with the minimum UMI thresholds ranging between 1608 and 7616 across samples. BUS records from each sample were aggregated to a unified counts Table, used as input for the monocle3 R/Bioconductor single-cell framework (<https://cole-trapnell-lab.github.io/monocle3/>), and processed using default workflow settings.

Differential gene expression analysis in scRNA-seq data

Differential gene expression analysis across cell types in our scRNA-seq data was performed using DESeq2 [64] along with specific recommendations for its application to scRNA-seq data using additional methods in the zinbwave [65] and scran [66] packages at <http://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html#recommendations-for-single-cell-analysis>, which draws on analyses and conclusion from Van den Berge, Zhu et al., and Ahlmann and Huber [67–69]. Briefly, the computeSumFactors() function in the scran

package was used to calculate size factors that were passed to the `zinbwave()` function and then output was passed onto the DESeq2 functions `DESeqDataSet()` and `DESeq()`. The `DESeq()` differential gene expression function was implemented using `test="LRT"` rather than the Wald test for significance testing, along with these scRNAseq-specific argument values: `useT=TRUE`, `minmu=1e-6`, and `minReplicatesForReplace=Inf`. Of all the genes reported by DESeq (i.e., all genes detected in the samples), we report and perform enrichment analysis on those with >0.03 average counts per cell, considering all others too lowly expressed in the corresponding cell type and therefore unlikely to reach significance.

PANTHER bioinformatics GO term enrichment analysis

We used the statistical overrepresentation test in GO of the PANTHER bioinformatics tool [28] at <https://www.pantherdb.org/> to look for functional enrichments (GO Ontology database 10.5281/zenodo.5725227 Released 2020-11-01). The background genes were the set of genes expressed in the cell line use in each comparison with modifications in the case of comparisons between mono cultures and combined cultures as described in the “Results” section.

Choice of significance thresholds

A. When identifying genes to use for GO enrichment analysis, we use an FDR of 0.1. This is because this analysis makes no claims on individual genes. While allowing only few false positives, it essentially doubles the number of genes strongly increasing power for the GO enrichment analysis.

B. When reporting a result as significant, we use the conventional FDR of 0.05. We use this for example to report specific genes as significantly different between conditions (e.g., genes that show differences between cell clusters) and when we report significant functional enrichments so for specific GO terms.

C. One exception is our report on genes differing between neurons and astrocytes (the “GWAS genes expressed in iPSC-derived neurons and astrocytes” section), where the differences were so many and substantial that we could use a much higher threshold ($FDR < 0.001$) and still show significant differences in almost two thirds of the genes. This allowed us to proceed with a very high confidence groups, which we could not have done in other comparisons. This same threshold was used when excluding gene from the N21 vs. N21aAst and the Ast0 vs AstN21 comparisons as astrocyte or neuron specific genes.

Single-cell clustering and dimensionality reduction

For the original scRNA-seq data from this report, UMI count matrices were processed and analyzed using the Seurat package (v4.1.0) in R [70]. A total of 1631 cells were included in this analysis. After datasets were normalized, the top 10,000 variable genes were selected for further analysis using the variance stabilizing transformation (`vst`) method. 2D visualization of our data was accomplished using principal components analysis (PCA) followed by Uniform Manifold Approximation and Projection (UMAP) of these PCs.

Seurat integration analysis

Integration of our scRNA-seq data with public scRNA-seq UMI count data [19] (GSE120046) was carried out using canonical correlation analysis (CCA) in Seurat [71, 72]. Integrated datasets were scaled and PCA was performed. We chose the first 10 PCs for use in non-linear dimensionality reduction by identifying the elbow on a scree plot of the first 30PCs. 2D visualization of the integrated data was accomplished using the Uniform Manifold Approximation and Projection (UMAP) algorithm on these 10 PCs.

Meta Neighbor analysis

Cell-type replicability analysis across datasets was performed using the MetaNeighbor (v1.1.0; Crow et al. 2018 [49]) package in R. We used unsupervised MetaNeighbor to first determine intersecting highly variable genes across datasets and then used the Spearman correlation network as described in Crow et al. [49] to determine replicability.

Abbreviations

2D	Two-dimensional
AD	Alzheimer’s disease
ASD	Autism spectrum disorder
CNS	Central nervous system
CPM	Counts per million
DE	Differential expression
GW	Gestational week
GWAS	Genome-wide association studies
hiPSC-A	HiPSCs differentiated in vitro to astrocytes
hiPSC-A0	HiPSC-A grown alone
hiPSC-AN21	HiPSC-A co-cultured with hiPSC-N21
hiPSC-N	HiPSCs differentiated in vitro to excitatory neurons
hiPSC-N15	HiPSC-N after 15 days of differentiation
hiPSC-N21	HiPSC-N after 21 days of differentiation
hiPSC-N21A	HiPSC-N21 co-cultured with hiPSC-A
hiPSCs	Human-induced pluripotent stem cells
KNN	K-nearest neighbor
PCA	Principal component analysis
PNS	Peripheral nervous system
scRNA-seq	Single-cell RNA sequencing
SZ	Schizophrenia
UMAP	Uniform Manifold Approximation and Projection

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-024-01867-4>.

Additional file 1: Sup. Figure 1. The fraction of the cells in each cell type and the mean expression of each gene in each cell type.

Additional file 2: Sup. Figure 2. As above for glutamatergic genes.

Additional file 3: Sup. Figure 3. As above for synaptic genes.

Additional file 4: Sup. Figure 4. As above for CaM Kinase genes.

Additional file 5: Sup. Figure 5. As above for other neuronal genes.

Additional file 6: Sup. Figure 6. Gene expression correlations with bulk dataset from other cell lines similarly differentiated in vitro to neurons and in one vivo dataset.

Additional file 7: Sup. Figure 7. Gene expression correlations with bulk dataset from other cell lines similarly differentiated in vitro to astrocytes and in one vivo dataset.

Additional file 8: SupTable1. Significantly differentially expressed genes for each cluster compared with the remaining clusters of the same cell type (neuronal or astrocytic).

Additional file 9: SupTable2. Graphical representation of PANTHER bioinformatics analysis of genes showing differential expression between neuronal clusters from Table 1. All comparisons are shown together in columns. Colored cells indicate significant enrichment for that comparison and the color indicates whether it is for up-regulated genes (red), Down-regulated genes (blue) or both up and down regulated genes (yellow).

Additional file 10: SupTable3. Graphical representation of PANTHER bioinformatics analysis of genes showing differential expression between astrocytic clusters from Table 1. These are shown as in Supplementary Table 2. Note that there were no enrichments for up-regulated genes, which is why all cells are blue.

Additional file 11: SupTable4. Differential expression analysis between hiPSC-N and hiPSC-A.

Additional file 12: SupTable5. Differential expression analysis between hiPSC-N15 and hiPSC-N21.

Additional file 13: Sup. Figure 8. Supplementary Figure 8: PANTHER bioinformatics analysis of genes expressed higher in hiPSC-N15 than hiPSC-N21 shown as a bar graph. The length of each bar corresponds to fold-enrichment, the color to statistical significance and the number next to it to the number of significantly differentially expressed genes in each pathway.

Additional file 14: Sup. Figure 9. PANTHER bioinformatics analysis of genes expressed higher in hiPSC-N21 than hiPSC-N15 shown as in Supplementary Figure 8.

Additional file 15: SupTable6. GWAS genes differing in expression between different conditions.

Additional file 16: SupTable7. Differential expression analysis between hiPSC-N21 and hiPSC-N21A.

Additional file 17: Sup. Figure 10. PANTHER bioinformatics analysis of genes expressed higher in hiPSC-N21A than hiPSC-N21 shown as in Supplementary Figure 8.

Additional file 18: Sup. Figure 11. PANTHER bioinformatics analysis of genes expressed higher in hiPSC-N21 than hiPSC-N21A shown as in Supplementary Figure 8.

Additional file 19: SupTable8. Differential expression analysis between hiPSC-A0 and hiPSC-AN21.

Additional file 20: Sup. Figure 12. PANTHER bioinformatics analysis of genes expressed significantly higher in hiPSC-AN21 than hiPSC-A0 shown as in Supplementary Figure 8.

Additional file 21: SupTable9 PANTHER bioinformatics analysis of genes expressed significantly higher in hiPSC-A0 than hiPSC-AN21.

Additional file 22: Sup. Figure 13. PANTHER bioinformatics analysis of genes expressed significantly higher in hiPSC-A0 than hiPSC-AN21 shown as in Supplementary Figure 8. Only more than 3-fold enrichments are shown. All enrichments are in sip. Table 9.

Additional file 23: Sup. Figure 14. The top 50 genes in each cell type contrast, detailed in the "Differences between hiPSC-N15 and hiPSC-N21," "Differences between hiPSC-N21 and hiPSC-N21A," and "Differences between hiPSC-A0 and hiPSC-AN21" sections shown in heatmaps.

Additional file 24: Sup. Figure 15. The complete heatmap of the MetaNeighbor analysis shown in Figure 2. MetaNeighbor analysis of our bulk and pseudobulk expression data with in vivo data from two in vivo studies. Ex_Cor, excitatory cortical; HEW, Human embryo week; Astro, astrocytes; Oligo, oligodendrocytes; OPC oligodendrocyte precursor cells; Endo, endothelial.

Additional file 25: Sup. Figure 16. hiPSC-A cells re-clustered alone (A). Integrated UMAP (B) and MetaNeighbor analysis with Fan et al. show more similarity to in vivo astrocytes.

Additional file 26: Sup. Figure 17. Staining of BC1 cells differentiated into neurons and astrocyte using the same protocols described here. A. Neurons growing in isolation. blue stain = DAPI, green stain = MAP2, Red stain = NeuN. B. Culture of induced astrocytes. Red stain = GFAP. Scale in B is as in C. C. Induced neurons growing in co-culture with induced astrocytes. Red stain = MAP2, green stain = GFAP, blue stain = DAPI. More images of neurons can be in our previously published work (references are in the text).

Acknowledgements

Not applicable.

Authors' contributions

D.D. performed experiments and assisted with writing the manuscript, S.S. performed the data analysis and assisted with writing the manuscript. G.S. helped with conceptualizing the project, performed the data analysis, and assisted with writing the manuscript. M.H.W. and K.F. helped with experiments and assisted with editing the manuscript. L.G. helped with conceptualizing the project, interpreting the results, and writing the manuscript. C.C. performed and supervised data analysis and helped with interpreting the results and writing the manuscript. V.M. helped with performing and supervising experiments and writing the manuscript. D.A. conceptualized the project, secured funding, and helped with the analysis and writing the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by NIMH grants R01 MH113215 and RF1 MH122936 to DA. Data sharing and visualization via NeMO Analytics was supported by grants R24MH114815 and R01DC019370. D.A. and V.M. are also funded by Johns Hopkins Richman Family Precision Medicine Center of Excellence in Alzheimer's Disease.

Availability of data and materials

All data generated or analyzed during this study are included in this published article, its supplementary information files, and publicly available repositories. The data generated here are in the GEO database (under accession GSE260498). Other data used were from the GEO database (accessions GSE260873, GSE97904, GSE163161, GSE120046, GSE149598, GSE102956, GSE73721, GSE67835, GSE63482, GSE52564), the NCBI BioProject database under accession PRJNA596331, and the BioStudies ArrayExpress database under the accession codes E-MTAB-10632 and E-MTAB-9233.

Declarations

Ethics approval and consent to participate

The work reported here has been acknowledged by the Johns Hopkins Institutional review board as "not human subjects research" (approved application # IRB00122135).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Genetic Medicine, Johns Hopkins University School of Medicine, 733 E. Broadway, Baltimore, MD 21205, USA. ²Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, USA. ³Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, USA. ⁴Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, USA. ⁵Institute of Genome Sciences, University of Maryland School of Medicine, Baltimore, USA. ⁶Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, USA.

Received: 18 July 2023 Accepted: 13 March 2024

Published online: 02 April 2024

References

- Das D, Feuer K, Wahbeh M, Avramopoulos D. modeling psychiatric disorder biology with stem cells. *Curr Psychiatry Rep.* 2020;22(5):24.
- Hartley BJ, Tran N, Ladrán I, Reggio K, Brennand KJ. Dopaminergic differentiation of schizophrenia hiPSCs. *Mol Psychiatry.* 2015;20(5):549–50.
- Kerr CL, Letzen BS, Hill CM, Agrawal G, Thakor NV, Sternecker JL, Gearhart JD, All AH. Efficient differentiation of human embryonic stem cells into oligodendrocyte progenitors for application in a rat contusion model of spinal cord injury. *Int J Neurosci.* 2010;120(4):305–13.
- Krencik R, Zhang SC. Directed differentiation of functional astroglial subtypes from human pluripotent stem cells. *Nat Protoc.* 2011;6(11):1710–7.
- Liu Y, Liu H, Sauvey C, Yao L, Zarnowska ED, Zhang SC. Directed differentiation of forebrain GABA interneurons from human pluripotent stem cells. *Nat Protoc.* 2013;8(9):1670–9.
- Pandya H, Shen MJ, Ichikawa DM, Sedlock AB, Choi Y, Johnson KR, Kim G, Brown MA, Elkahloum AG, Maric D, et al. Differentiation of human and murine induced pluripotent stem cells to microglia-like cells. *Nat Neurosci.* 2017;20(5):753–9.
- Zhang Y, Pak C, Han Y, Ahlenius H, Zhang Z, Chanda S, Marro S, Patzke C, Acuna C, Covy J, et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron.* 2013;78(5):785–98.
- Tcw J, Wang M, Pimenova AA, Bowles KR, Hartley BJ, Lacin E, Machlovi SI, Abdelaal R, Karch CM, Phatnani H, et al. An efficient platform for astrocyte differentiation from human induced pluripotent stem cells. *Stem Cell Reports.* 2017;9(2):600–14.
- Orvis J, Gottfried B, Kancherla J, Adkins RS, Song Y, Dror AA, Olley D, Rose K, Chrysostomou E, Kelly MC, et al. gEAR: Gene Expression Analysis Resource portal for community-driven, multi-omic data exploration. *Nat Methods.* 2021;18(8):843–4.
- Ament SA, Adkins RS, Carter R, Chrysostomou E, Colantuoni C, Crabtree J, Creasy HH, Degatano K, Felix V, Gandt P, et al. The neuroscience multi-omic archive: a BRAIN initiative resource for single-cell transcriptomic and epigenomic data from the mammalian brain. *Nucleic Acids Res.* 2023;51(D1):D1075–85.
- Neuroscience Multi-Omic Analytics (NeMO) [<https://nemoanalytics.org/?l=DasEtAINGN2&g=NES>]
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech: Theory Exp.* 2008;2008(10):P10008.
- Britanova O, de Juan RC, Cheung A, Kwan KY, Schwark M, Gyorgy A, Vogel T, Akopov S, Mitkovski M, Agoston D, et al. Satb2 is a postmitotic determinant for upper-layer neuron specification in the neocortex. *Neuron.* 2008;57(3):378–92.
- Vallano ML, Beaman-Hall CM, Mathur A, Chen Q. Astrocytes express specific variants of CaM KII delta and gamma, but not alpha and beta, that determine their cellular localizations. *Glia.* 2000;30(2):154–64.
- Lam AN, Peng X, Das D, Bader JS, Avramopoulos D. Transcriptomic data of clozapine-treated Ngn2-induced human excitatory neurons. *Data Brief.* 2021;35:106897.
- Sagar R, Azoidis I, Zivko C, Xydia A, Oh ES, Rosenberg PB, Lyketsos CG, Mahairaki V, Avramopoulos D: Excitatory neurons derived from human-induced pluripotent stem cells show transcriptomic differences in Alzheimer's patients from controls. *Cells* 2023, 12(15).
- D A: Transcriptome analysis of human induced excitatory neurons supports a strong effect of clozapine on cholesterol biosynthesis. In.: GEO: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163161>; Feb 11, 2021.
- Sagar R AI, Zivko C, Xydia A, Oh ES, Rosenberg PB, Lyketsos CG, Mahairaki V, Avramopoulos D: Excitatory neurons derived from human-induced pluripotent stem cells show transcriptomic differences in Alzheimer's patients from controls. In.: GEO: <https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE260873>; Mar 06, 2024.
- Fan X, Fu Y, Zhou X, Sun L, Yang M, Wang M, Chen R, Wu Q, Yong J, Dong J, et al. Single-cell transcriptome analysis reveals cell lineage specification in temporal-spatial patterns in human cortical development. *Sci Adv.* 2020;6(34):eaaz2978.
- Fan X FY: Regional different regulations in development of human cortex and pons revealed by single cell transcriptome. In.: GEO: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120046>; Sep 17, 2018.
- Hedegaard A, Monzon-Sandoval J, Newey SE, Whiteley ES, Webber C, Akerman CJ. Pro-maturational effects of human iPSC-derived cortical astrocytes upon iPSC-derived cortical neurons. *Stem Cell Reports.* 2020;15(1):38–51.
- Lin YT, Seo J, Gao F, Feldman HM, Wen HL, Penney J, Cam HP, Gjoneska E, Raja WK, Cheng J, et al. APOE4 causes widespread molecular and cellular alterations associated with Alzheimer's disease phenotypes in human iPSC-derived brain cell types. *Neuron.* 2018;98(6):1141–1154e1147.
- Zhang Y, Sloan SA, Clarke LE, Caneda C, Plaza CA, Blumenthal PD, Vogel H, Steinberg GK, Edwards MS, Li G, et al. Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron.* 2016;89(1):37–53.
- Hedegaard A M-SJ, Newey SE, Whiteley ES, Webber C, Akerman CJ: RNA-seq of iPSC-derived astrocytes that are capable of exerting pro-maturational effect on synaptic networks. In.: GEO: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE149598>; Apr 29, 2020.
- Lin Y SJ, Gao F, Tsai L: APOE4 causes widespread molecular and cellular alterations associated with Alzheimer's disease phenotypes in human iPSC-derived brain cell types. In.: GEO: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE102956>; May 31, 2018.
- Sloan S ZY: RNA-Seq of human astrocytes. In.: GEO: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE73721>; Dec 11, 2015.
- TCW J WM, Pimenova AA, Bowles KR, Hartley BJ, Lacin E, Machlovi S, Abdelaal R, Karch CM, Phatnani H, Slesinger PA, Zhang B, Goate AM, Brennand KJ: Functional astrocytes differentiated from hiPSCs. In.: GEO: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE97904>; Aug 18, 2017.
- Thomas PD, Ebert D, Muruganujan A, Mushayama T, Albu LP, Mi H. PANTHER: making genome-scale phylogenetics accessible to all. *Protein Sci.* 2022;31(1):8–22.
- Ripke S, Walters JTR, O'Donovan MC, Consortium TSWGotPG: Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *bioRxiv* 2020, <https://doi.org/10.1101/2020.09.12.20192922>.
- Wightman DP, Jansen IE, Savage JE, Shadrin AA, Bahrami S, Holland D, Rongve A, Borte S, Winsvold BS, Drange OK, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet.* 2021;53(9):1276–82.
- SFARI Gene database [<https://gene.sfari.org/database/human-gene/>]
- Paulsen B, Velasco S, Kedaigle AJ, Pignoni M, Quadrato G, Deo AJ, Adiconis X, Uzquiano A, Sartore R, Yang SM, et al. Autism genes converge on asynchronous development of shared neuron classes. *Nature.* 2022;602(7896):268–73.
- Deng Q, Wu C, Parker E, Liu TC, Duan R, Yang L: Microglia and astrocytes in Alzheimer's disease: significance and summary of recent advances. *Aging Dis* 2023.
- Gene Ontology Resource [<https://geneontology.org/>]
- Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun.* 2017;8:14049.
- Culotti JG, Kolodkin AL. Functions of netrins and semaphorins in axon guidance. *Curr Opin Neurobiol.* 1996;6(1):81–8.

37. Brot S, Smaoune H, Youssef-Issa M, Malleval C, Benetollo C, Besancon R, Auger C, Moradi-Ameli M, Honnorat J. Collapsin response-mediator protein 5 (CRMP5) phosphorylation at threonine 516 regulates neurite outgrowth inhibition. *Eur J Neurosci*. 2014;40(7):3010–20.
38. Brose K, Bland KS, Wang KH, Arnott D, Henzel W, Goodman CS, Tessier-Lavigne M, Kidd T. Slit proteins bind Robo receptors and have an evolutionarily conserved role in repulsive axon guidance. *Cell*. 1999;96(6):795–806.
39. Kimelberg HK, Nedergaard M. Functions of astrocytes and their potential as therapeutic targets. *Neurotherapeutics*. 2010;7(4):338–53.
40. Monterey MD, Wei H, Wu X, Wu JQ. The many faces of astrocytes in Alzheimer's disease. *Front Neurol*. 2021;12:619626.
41. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, Hayden Gephart MG, Barres BA, Quake SR. A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci U S A*. 2015;112(23):7285–90.
42. Molyneaux BJ, Goff LA, Brettler AC, Chen HH, Hrvatin S, Rinn JL, Arlotta P. DeCoN: genome-wide analysis of in vivo transcriptional dynamics during pyramidal neuron fate selection in neocortex. *Neuron*. 2015;85(2):275–88.
43. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keefe S, Phatnani HP, Guarnieri P, Caneda C, Ruderisch N, et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci*. 2014;34(36):11929–47.
44. Darmanis S EM, Quake SR, Sloan SA, Barres BA, Zhang Y, Caneda C, Hayden Gephart MG, Shuer LM: A survey of human brain transcriptome diversity at the single cell level. In: GEO: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE67835>; May 20, 2015.
45. LA G: Transcriptional profiling of sorted mouse cortical projection neuron subtypes. In: GEO: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63482>; Dec 31, 2014.
46. Zhang Y CK, Sloan SA, Scholze AR, Caneda C, Ruderisch N, Deng S, Daneman R, Barres BA, Wu JQ: An RNA-Seq transcriptome and splicing database of neurons, glia, and vascular cells of the cerebral cortex. In: GEO: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52564>; Aug 15, 2014.
47. Burke EE, Chenoweth JG, Shin JH, Collado-Torres L, Kim SK, Micali N, Wang Y, Colantuoni C, Straub RE, Hoepfner DJ, et al. Dissecting transcriptomic signatures of neuronal differentiation and maturation using iPSCs. *Nat Commun*. 2020;11(1):462.
48. Dissecting transcriptomic signatures of neuronal differentiation and maturation using iPSCs. In: NCBI: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA596331>; 17-Dec-2019.
49. Crow M, Paul A, Ballouz S, Huang ZJ, Gillis J. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat Commun*. 2018;9(1):884.
50. Lin HC, He Z, Ebert S, Schornig M, Santel M, Nikolova MT, Weigert A, Hevers W, Kasri NN, Taverna E, et al. NGN2 induces diverse neuron types from human pluripotency. *Stem Cell Reports*. 2021;16(9):2118–27.
51. Schornig M, Ju X, Fast L, Ebert S, Weigert A, Kanton S, Schaffer T, Nadif Kasri N, Treutlein B, Peter BM et al: Comparison of induced neurons reveals slower structural and functional maturation in humans than in apes. *Elife* 2021, 10.
52. Xiang-Chun Ju MS, Elena Taverna: Single-cell RNA sequencing of Ngn2-induced neurons from the pluripotent stem cells of great apes. In: ArrayExpress: <https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-9233>; 14 December 2020.
53. Zhisong He H-CL, J Gray Camp, Barbara Treutlein: Ngn2 induces diverse neuronal lineages from human pluripotency. In: ArrayExpress: <https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-10632>; 4 August 2021.
54. Diez-Roux G, Banfi S, Sultan M, Geffers L, Anand S, Rozado D, Magen A, Canidio E, Pagani M, Peluso I, et al. A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol*. 2011;9(1):e1000582.
55. Yuan A, Sasaki T, Kumar A, Peterhoff CM, Rao MV, Liem RK, Julien JP, Nixon RA. Peripherin is a subunit of peripheral nerve neurofilaments: implications for differential vulnerability of CNS and peripheral nervous system axons. *J Neurosci*. 2012;32(25):8501–8.
56. Gosselin D, Skola D, Coufal NG, Holtman IR, Schlachetzki JCM, Sajti E, Jaeger BN, O'Connor C, Fitzpatrick C, Pasillas MP, et al. An environment-dependent transcriptional network specifies human microglia identity. *Science*. 2017; 356(6344). <https://www.science.org/doi/10.1126/science.aal3222>.
57. Cho JM, Shin YJ, Park JM, Kim J, Lee MY. Characterization of nestin expression in astrocytes in the rat hippocampal CA1 region following transient forebrain ischemia. *Anat Cell Biol*. 2013;46(2):131–40.
58. Cai L, Bai H, Mahairaki V, Gao Y, He C, Wen Y, Jin YC, Wang Y, Pan RL, Qasba A, et al. A universal approach to correct various HBB gene mutations in human stem cells for gene therapy of beta-thalassemia and sickle cell disease. *Stem Cells Transl Med*. 2018;7(1):87–97.
59. Das D, Peng X, Lam AN, Bader JS, Avramopoulos D. Transcriptome analysis of human induced excitatory neurons supports a strong effect of clozapine on cholesterol biosynthesis. *Schizophr Res*. 2021;228:324–6.
60. Dobrindt K, Zhang H, Das D, Abdollahi S, Prorok T, Ghosh S, Weintraub S, Genovese G, Powell SK, Lund A, et al. Publicly available hiPSC lines with extreme polygenic risk scores for modeling schizophrenia. *Complex Psychiatry*. 2021;6(3–4):68–82.
61. Feuer KL, Peng X, Yovo CK, Avramopoulos D: DPYSL2/CRMP2 isoform B knockout in human iPSC-derived glutamatergic neurons confirms its role in mTOR signaling and neurodevelopmental disorders. *Mol Psychiatry* 2023.
62. Mahairaki V, Ryu J, Peters A, Chang Q, Li T, Park TS, Burridge PW, Talbot CC Jr, Asnaghi L, Martin LJ, et al. Induced pluripotent stem cells from familial Alzheimer's disease patients differentiate into mature neurons with amyloidogenic properties. *Stem Cells Dev*. 2014;23(24):2996–3010.
63. Clark BS, Stein-O'Brien GL, Shiao F, Cannon GH, Davis-Marcisak E, Sherman T, Santiago CP, Hoang TV, Rajaii F, James-Esposito RE, et al. Single-cell RNA-Seq analysis of retinal development identifies NFI factors as regulating mitotic exit and late-born cell specification. *Neuron*. 2019;102(6):1111–1126e1115.
64. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
65. Risso D, Perraudeau F, Gribkova S, Dudoit S, Vert JP: A general and flexible method for signal extraction from single-cell RNA-seq data. *Nat Commun*. 2018;9(1):284.
66. Lun AT, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res*. 2016;5:2122.
67. Ahlmann-Eltz C, Huber W. glmGamPoi: fitting Gamma-Poisson generalized linear models on single cell count data. *Bioinformatics*. 2021;36(24):5701–2.
68. Van den Berge K, Perraudeau F, Soneson C, Love MI, Risso D, Vert JP, Robinson MD, Dudoit S, Clement L. Observation weights unlock bulk RNA-seq tools for zero inflation and single-cell applications. *Genome Biol*. 2018;19(1):24.
69. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics*. 2019;35(12):2084–92.
70. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184(13):3573–3587e352.
71. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*. 2018;36(5):411–20.
72. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y, Stoeckius M, Smibert P, Satija R: Comprehensive integration of single-cell data. *Cell* 2019, 177(7):1888–1902 e1821.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.