Stem Cell Reports
Article

ISSCR

OPEN ACCESS

Reproducibility of Molecular Phenotypes after Long-Term Differentiation to Human iPSC-Derived Neurons: A Multi-Site Omics Study

Viola Volpato,[1] James Smith,[2] Cynthia Sandor,[1] Janina S. Ried,[3] Anna Baud,[4] Adam Handel,[1] Sarah E. Newey,[5] Frank Wessely,[1] Moustafa Attar,[6] Emma Whiteley,[5] Satyan Chintawar,[7] An Verheyen,[8] Thomas Barta,[9] Majlinda Lako,[9] Lyle Armstrong,[9] Caroline Muschet,[10] Anna Artati,[10] Carlo Cusulin,[11] Klaus Christensen,[11] Christoph Patsch,[11] Eshita Sharma,[6] Jerome Nicod,[6] Philip Brownjohn,[2] Victoria Stubbs,[2] Wendy E. Heywood,[4] Paul Gissen,[12] Roberta De Filippis,[3] Katharina Janssen,[3] Peter Reinhardt,[3] Jerzy Adamski,[10] Ines Royaux,[8] Pieter J. Peeters,[8] Georg C. Terstappen,[3] Martin Graf,[11] Frederick J. Livesey,[2] Colin J. Akerman,[5] Kevin Mills,[4] Rory Bowden,[6] George Nicholson,[13] Caleb Webber,[1,*] M. Zameel Cader,[7,*] and Viktor Lakics[3,*]

[1]Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford OX1 3PT UK
[2]Gurdon Institute, University of Cambridge, Cambridge CB2 1QN, UK
[3]Neuroscience Discovery, Biology Department, AbbVie Deutschland GmbH & Co. KG, Ludwigshafen, Germany
[4]Centre for Translational Omics, UCL Great Ormond Street Institute of Child Health, London WC1N 1EH, UK
[5]Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT, UK
[6]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK
[7]Nuffield Department of Clinical Neuroscience, University of Oxford, Oxford OX3 9DU, UK
[8]Janssen Research and Development, Beerse 2340, Belgium
[9]Institute of Genetic Medicine, Newcastle University, Newcastle NE1 3BZ, UK
[10]Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Neuherberg 85764, Germany
[11]Roche Pharmaceutical Research and Early Development, Roche Innovation Center Basel, 4070 Basel, Switzerland
[12]MRC Laboratory for Molecular Cell Biology, University College London, London WC1E 6BT, UK
[13]Department of Statistics, University of Oxford, Oxford OX1 3LB, UK
*Correspondence: caleb.webber@dpag.ox.ac.uk (C.W.), zameel.cader@ndcn.ox.ac.uk (M.Z.C.), viktor.lakics@abbvie.com (V.L.)
https://doi.org/10.1016/j.stemcr.2018.08.013

SUMMARY

Reproducibility in molecular and cellular studies is fundamental to scientific discovery. To establish the reproducibility of a well-defined long-term neuronal differentiation protocol, we repeated the cellular and molecular comparison of the same two iPSC lines across five distinct laboratories. Despite uncovering acceptable variability within individual laboratories, we detect poor cross-site reproducibility of the differential gene expression signature between these two lines. Factor analysis identifies the laboratory as the largest source of variation along with several variation-inflating confounders such as passaging effects and progenitor storage. Single-cell transcriptomics shows substantial cellular heterogeneity underlying inter-laboratory variability and being responsible for biases in differential gene expression inference. Factor analysis-based normalization of the combined dataset can remove the nuisance technical effects, enabling the execution of robust hypothesis-generating studies. Our study shows that multi-center collaborations can expose systematic biases and identify critical factors to be standardized when publishing novel protocols, contributing to increased cross-site reproducibility.

INTRODUCTION

Reproducibility is a cornerstone of science. Yet, in recent years, a number of publications highlighted serious issues regarding this fundamental principle of scientific approach, to the extent that the expression "reproducibility crisis" was coined (Munafò et al., 2017; Baker, 2016). The more complex experimental procedures are, and the longer they are applied, the higher the possibility of introducing variability and noise during a research study. This is particularly critical for human induced pluripotent stem cells (iPSCs), which need to be differentiated using lengthy complex procedures in order to be used as novel *in vitro* models in basic science and drug discovery (Avior et al., 2016), but this increases the variability, such as well-to-well differences in cell density and cellular heterogeneity. Protocols for efficient generation of specific neuronal subtypes mimic human development both in the appearance of successive phenotypes, and also in duration, potentially taking more than 100 days *in vitro* (Shi et al., 2012a). Reproducibility is especially critical when comparing iPSC-derived cells from multiple donors to discover cellular disease phenotypes and their underlying pathways using unbiased omics experiments. While the reproducibility of transcriptomic (Li et al., 2014) and proteomic (Kim et al., 2007) approaches have been well established for simple cellular systems, no systematic studies have been performed to assess the cross-site reproducibility of these readouts after a long-term iPSC differentiation protocol, such as the derivation of human cortical neurons.

The goal of our study was to identify the extent of variability in an iPSC experiment conducted by multiple

groups of the IMI StemBANCC (Innovative Medicines Initiative, Stem cells for biological assays of novel drugs and predictive toxicology) initiative, which aims to generate and interrogate a large collection of stem cell models for disease modeling and therapeutic research. For multi-site comparative studies, large public-private partnerships offer a unique framework due to the participation of both academic and industry organizations with strong scientific background in iPSC biology, representing a "best case scenario" to assess cross-site reproducibility.

Our cross-site analysis utilized a previously published neuronal differentiation protocol (Shi et al., 2012b). In this study we set out to assess the inter- and intra-laboratory reproducibility of transcriptomic and proteomic readouts using two iPSC lines and standard laboratory practices adhered to by all participating laboratories. The differentiation protocol nevertheless enables individual laboratories to apply their laboratory-specific approaches simulating the reproduction of a published method. The key questions in this study were firstly whether a laboratory would be able to separate the two iPSC lines at the molecular level, and secondly whether the identified molecular differences between the two lines were consistent between laboratories. Three academic and two industrial organizations participated in the study to simulate this real-life reproducibility scenario. In addition to bulk omics analyses and single-cell (SC) RNA sequencing (RNA-seq) to assess cellular heterogeneity, the reproducibility of a known cellular phenotype arising from a specific mutation in one of the iPSC lines has also been evaluated.

To our knowledge, this study represents the first comprehensive experiment to assess the intra- and inter-laboratory reproducibility of multiple readouts measured in an iPSC-derived *in vitro* model system containing differentiated human neurons. Despite acceptable intra-laboratory reproducibility of omics readouts and surprisingly good cross-site reproducibility of a previously identified cellular phenotype, omics datasets from different sites have large variation that masks specific differences, rendering it impossible to distinguish these two lines from each other in a combined dataset. SC analyses demonstrate that cell-type heterogeneity is an important confounder in these comparisons, with variation undermining the detection of differentially expressed (DE) genes, proteins, and pathways. However, we show that there are identifiable sources of variation that investigators can control and thereby increase biological signals in iPSC-based molecular studies. Besides strongly recommending to disclose these identified variation-inflating confounders in published iPSC differentiation protocols, our study also shows that collaborative approaches with larger sample numbers in cross-laboratory studies are valuable to detect and remove unwanted variation (Freytag et al., 2015).

RESULTS

Experimental Design

Five laboratories (referred as A, B, C, D, and E) received the same two fibroblast-derived human iPSC lines. One line was derived from a healthy control individual while the second one originated from a patient with familial Alzheimer's disease carrying a presenilin 1 (PS1) mutation. Note that our study was not designed to examine the effects of this mutation per se but instead focusses on the reproducibility of the comparison of these two lines (see Discussion). All laboratories followed the same standard operating protocol (SOP) (see Supplemental Experimental Procedures) to differentiate the cultures into cortical projection neurons in three independent inductions (replicates) (Figure 1). Laboratory-specific variations and observations were recorded (see below and Table S1). Total RNA and cell lysates were collected at two time points during differentiation, specifically, after 25 and 55 days from the final plating (FP), respectively (representing ~50 and ~80 days *in vitro* differentiation from the iPSC state), and sent to central locations for RNA-seq and proteomic analyses (Figure 1A).

Molecular Profiles Show Strong Similarity within Laboratories and Clearly Separate by Cell Line Genotype

To assess the reproducibility of transcriptomic readouts we first examined whether each laboratory was able to demonstrate a clear segregation between the two iPSC lines at a molecular level. Detection of differential molecular profiles between the two lines might be expected due to their differing genetic and epigenetic backgrounds. It is important for molecular studies of iPSC-based models that genotypic differences between lines are identifiable.

Applying RNA-seq, the expression of variable numbers of protein-coding genes across different samples were detected (with at least one count), with about 70% (13,373) of the 19,086 protein-coding genes expressed across all 57 samples. In further analyses we considered only this set of 13,373 commonly expressed genes. Principal-component analysis (PCA) on the transcriptomic profiles from individual sites illustrated clear separation between the samples from the two cell lines in all five laboratories at both early and late time points (Figure 2A), indicating that genetic background or genotype is a clear source of variation within laboratories. Each laboratory performed three independent cortical differentiations, and the consistency within each laboratory is evident by the greater similarity in gene expression profiles of the three replicates of the same genotype compared with the gene expression profiles between genotypes (Figure 2A). The Euclidean

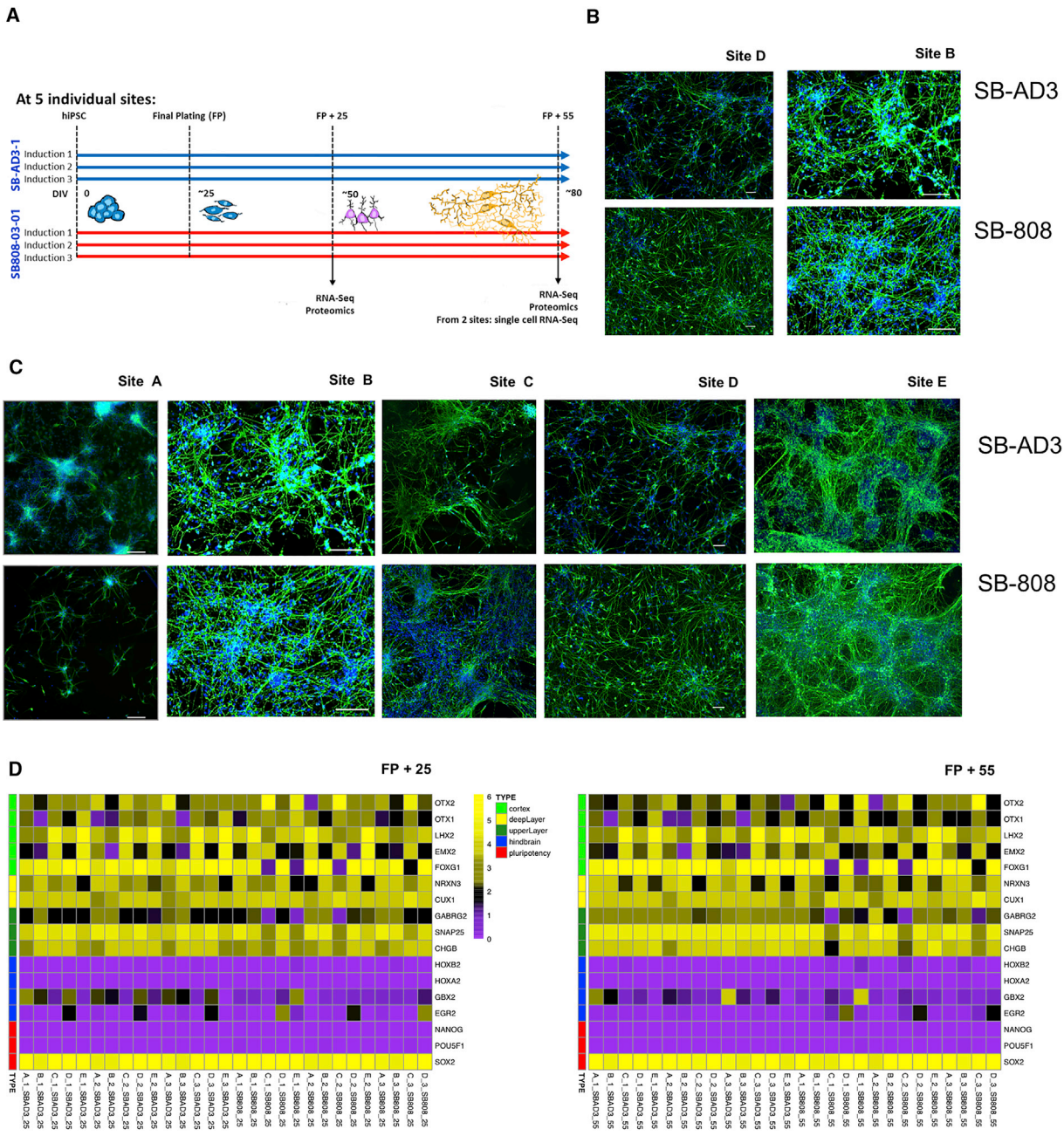


Figure 1. Experimental Outline of the Study

(A) iPSC lines from two genotypes were differentiated at five different sites with three individual inductions at each site. The given samples were taken at FP (final plating) + 25 days and FP + 55 days.

(B) Representative iPSC-induced cortical neurons at FP + 10 days in culture, immunolabeled with Tuj-1 (green) and DAPI (blue) derived from SBAD3 and AD SB808 cell lines. Neurons grown in two different laboratories are shown (sites D and B). Scale bars, 50 μm (site D), 100 μm (site B).

(C) Cortical neuronal inductions from CTR and PS1 cells 10–20 days after FP, showing presence of neuron-specific βIII-tubulin (green) for sites B, C, D, and E or MAP2 (green) for site A and nuclear marker DAPI (blue). Scale bars, 100 μm (site A); 100 μm (site B); 100 μm (site C); 50 μm (site D); 100 μm (site E).

(D) Heatmaps of gene expression (log10[fragments per kilobase of transcript per million mapped reads]) at the two time points (FP + 25 left and FP + 55 right) of cortical neuron markers, hindbrain markers, and pluripotency markers for 57 StemBANCC samples confirm the presence of expected neuronal markers and the absence of all but SOX2 non-neuronal markers.

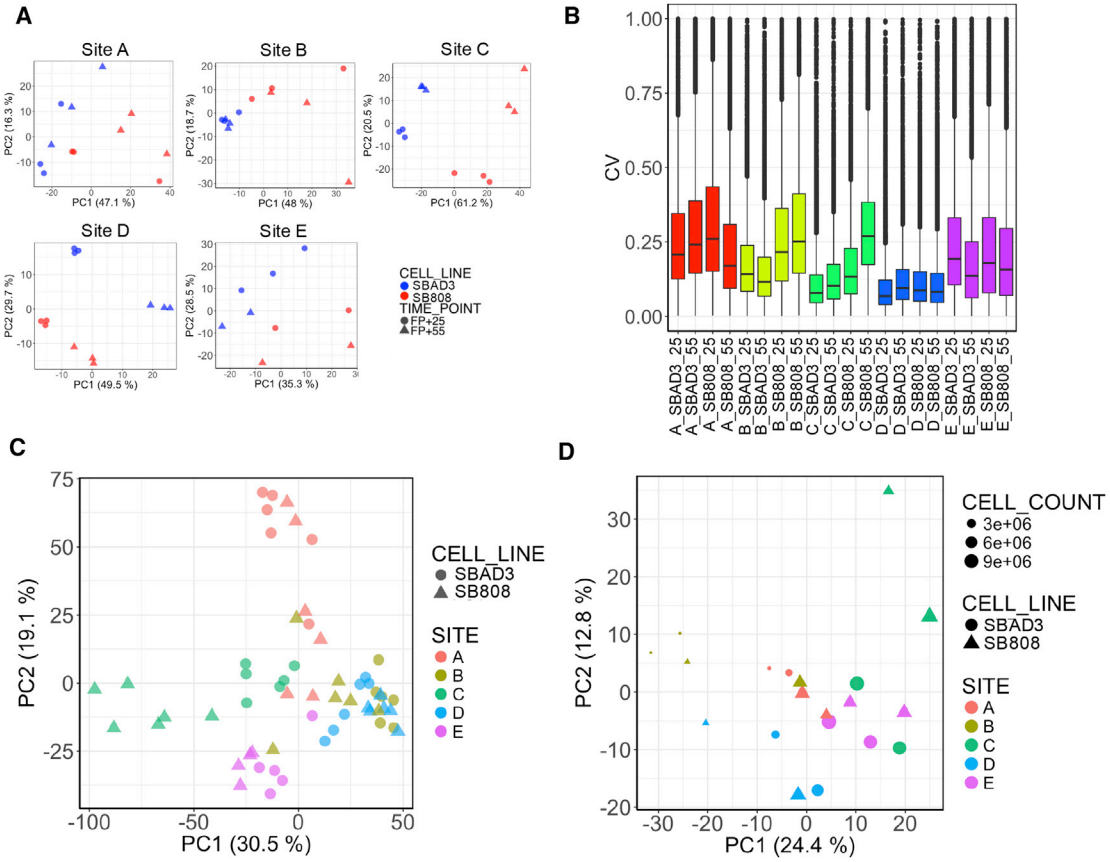


Figure 2. Detection of Laboratory as One Source of Unwanted Variability

(A) Individual laboratory experiments separate by cell line. PCA plots within laboratory (sites A–E) on 13,373 genes expressed across all samples (normalized gene counts were used) show clear separation between the samples from the two cell lines at both early and late time points.

(B) Degree of variability between replicates within the same laboratory. Boxplots showing the coefficients of variation calculated between gene expression values across replicates within each laboratory, cell line, and time point. Box-and-whisker graphs represent distributions, where the span of the box is the interquartile range (IQR) and includes the median (bold line). The ends of the upper and lower whiskers represent the data point with the maximum distance from the third and first quartiles, respectively, but no further than 1.5 × IQR. Data beyond the end of the whiskers are outliers.

(C) Samples cluster by laboratory in a combined PCA. First two principal components from a PCA on gene expression of 13,373 protein-coding genes that are expressed in all samples clearly cluster samples based on laboratory of origin.

(D) Laboratory and cell count are major confounders in protein-based PCA. First two principal components from a PCA on 1,034 proteins expressed across all samples; color coding according to laboratory; shapes correspond to cell line and sizes to averaged cell count.

See also Figure S1.

distances calculated between the gene expression profiles of each sample show that, within each laboratory, the expression profiles derived from replicates of the same line are significantly closer to each other than those between replicates of different lines for four out of five laboratories (Figure S1A).

The power to identify DE genes is strongly dependent on the experimental variance. A measure of this variance, the coefficient of variation (CV) of the transcriptomic dataset varied between laboratories. While the CV showed no clear genotype or maturation time trends (Figure 2B), the differing CV for each laboratory resulted in a large difference in the number of DE genes between genotypes controlling for time point variation (Figure S1B). Unsurprisingly, the highest number of DE genes was found in laboratory D, which shows the lowest degree of dispersion between replicates.

Cross-Site Comparison of Molecular Profiles Show Poor Reproducibility

Having demonstrated that each laboratory exhibited a clear segregation based on gene expression profiles, we asked whether the molecular differences were consistent between laboratories. Despite the use of a detailed SOP, in

a combined dataset containing data from all five partners, we found the laboratory was the dominant source of variation, masking any genotypic effects (Figure 2C). Importantly, only 15 DE genes are found in common between all laboratories indicating a remarkably low degree of cross-laboratory reproducibility (Figure S1B). The low number of overlapping genes may be a consequence of three laboratories (A, B, and E) detecting only a small number of DE genes. Certainly, sites C and D, which had the lowest CV and the highest number of DE genes, showed about ~50% overlap. At the pathway level, the similarity in enriched gene ontology (GO) terms for these five lists of DE genes is highly variable with semantic similarity comparison values ranging from 0.36 to 0.64 (Figure S1B). In summary, despite extensive efforts to replicate the same experiment, we observe significant variation that would confound any inter-laboratory comparison.

The PCA plot (Figure 2C) and the heatmap of Spearman's correlations (Figure S1C) revealed three potential outliers (SB808 line, laboratory C, here specific issues with detachment of cell monolayer were observed, see Table S1 on metadata). Nevertheless, in general, recorded variation in experimental procedures noted by individual laboratories did not explain the detected cross-laboratory sample variability (Figure S1C). The above observations suggest that much of the inter-laboratory variation arises from additional confounders that increase the within-laboratory variance.

To investigate whether cross-site variability in gene expression was also present at the proteomic level, lysates from replicate wells of the same 57 samples were pooled and analyzed (see Experimental Procedures). Similarly to the transcriptomic samples, the low number of overlapping proteins detected across all samples (only 10% of the 10,483 proteins observed in at least one sample) indicated that the abundance of various proteins is highly variable. For further analyses, we retained only those 1,037 proteins that were observed in all samples. As observed for the transcriptomics data, PCA and heatmap of Spearman's rank correlations of protein abundances did not show clustering of samples by genotype (Figures 2D and S1D). We noticed that, despite normalizing for total protein, the first principal component clearly captures a strong cell-number-related effect in addition to a laboratory-dependent effect. Taken together, the transcriptomic and proteomic profiles demonstrated a strong inter-laboratory variation that masks variation due to the genetic background of each iPSC line.

Factor Analysis Reveals the Transcriptional Axis of Maturation in iPSC-Derived Neurons and Confirms Robust Cortical Neuronal Differentiation

It is evident that cross-site comparisons can be significantly hampered by site-specific confounders, but collaborative studies that generate a large number of samples can have the power to identify nuisance technical effects. We applied a factor analysis-based method called remove unwanted variation (RUV) (Risso et al., 2014). This method can capture nuisance technical effects and RUV in the form of factors, while retaining variation associated with the biological covariate of interest. To demonstrate the utility of factor analysis in revealing biological signals, we first used the approach to determine the transcriptional determinants of *in vitro* neuronal maturation, exploiting the two time points, FP + 25 and FP + 55, in our samples. Consistent with the reported fetal nature of neurons derived from pluripotent stem cells (Handel et al., 2016), hierarchical clustering of the bulk transcriptomic profiles of 57 samples demonstrated their overall similarity to fetal postmortem brain samples from the BrainSpan Atlas of the Developing Human Brain (Sunkin et al., 2013) (Figure S2A). We performed normalization using RUV on samples from a single line (see Supplemental Experimental Procedures) to expose a clear time point variation that was not masked by any cell line-dependent effect on maturation.

PCA of the RUV-normalized gene expression profiles showed clustering of samples by time points (Figure 3A). BrainSpan samples projected onto the PCA coordinates of normalized iPSC neuron-maturity expression profiles recapitulated the direction of human neuronal maturation (Figure 3A) better than those projected onto PCA coordinates of non-normalized gene expression levels (Figure S2B). Accordingly, the post-RUV expression signature clearly separated the early and late stages of differentiation in BrainSpan fetal samples, and is in line with the observed direction of maturation in our samples. To confirm a neurodevelopmental role for genes whose expression varies in this component space, we selected those genes that maximally contributed in either direction to the identified transcriptional axis of maturation (principal component one) with a gene loading on this axis greater/less than ±0.01. To validate the biological role in neuronal maturation of the contributing genes, we used the CORTECON dataset (van de Leemput et al., 2014), which identified gene clusters representative of changes in temporal gene expression of *in vitro* cerebral cortex development from human embryonic stem cells. We observed that the genes characteristic to the less mature stage in STEMBANCC samples (with positive scores on PC1) were enriched in CORTECON gene cluster specific to the early developmental stages, namely the "cortical specification" cluster and that are active from days 10 to 20 after differentiation (van de Leemput et al., 2014). The set of genes representing the more mature stage (with negative scores on PC1) was instead significantly enriched in the "upper layer generation" cluster with an expression peak from day 60, as expected (Figure 3B). This analysis also confirmed that the laboratories were

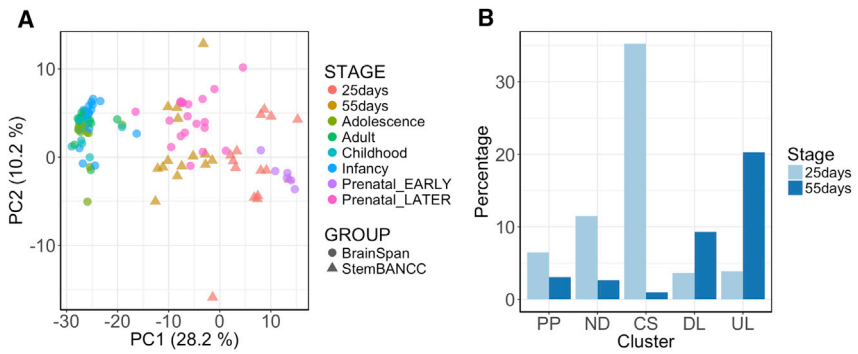


Figure 3. RUV-Corrected Gene Expression Reveals Differences of Maturation in Data

(A) Identification of a transcriptional axis of neuron maturation. BrainSpan samples are projected on the principal components calculated on the gene expression data in the present study, after RUV correction. It can be seen that the principal components of gene expression separate both STEMBANCC and BrainSpan samples based on developmental stages.

(B) Genes contributing to the identified transcriptional axis of neuron maturation are consistent with external data. The bar plot shows the percentages of time point-specific genes (selected based on gene loadings from PCA of the samples after RUV correction) falling into each CORTECON gene cluster representative of neuron developmental stages: 25-day-specific genes are enriched in pluripotency (PP), neuron development (ND), and cortical specification (CS) clusters; 55-day-specific genes are enriched in deep layer neuron generation (DL) and upper layer neuron generation (UL) clusters.

See also Figure S2.

successful with their differentiation protocol in producing cells with cortical specification at the early stage and upper layer cortical neurons as expected with this protocol at the later stages.

Factor Analyses Reveal Genotype-Related Differential Molecular Expression

To examine the sources of experimental variation, we applied RUV across all samples, retaining both cell line and time point variations. After RUV correction of gene counts on the first five estimated factors of variation, samples in the combined dataset cluster clearly by cell line and by time point (Figure 4A). Consequently, the number of DE genes detected between the 2 iPSC lines across all samples combined increased (from 1,873 before RUV correction to 3,051 after RUV correction) and between time points (from 2,186 before RUV correction to 3,868 after RUV correction) across all samples (see Table S2 for a complete list of DE genes). Examining a set of neuron-specific stage markers that are expected to be expressed in the differentiating samples (see Supplemental Experimental Procedures), the large distributional differences that were evident between samples in the non-normalized data (Figure 4C, top) were reduced upon removal of the identified variance factors (Figure 4C, bottom).

Similar to the transcriptomic analysis, when RUV correction was applied to protein abundances (available for FP + 55 time point only), good separation between the two iPSC lines was observed. After this normalization, a combined PCA shows that both transcriptomic and proteomic samples cluster together by iPSC line, indicative of a correlation between the two data types (Figure 4B). This is further supported by the increase in the number of differential abundant (DA) proteins (0 before RUV correction, 205 after RUV correction) and consequently in the percentage of overlapping DE genes and DA proteins between laboratories after RUV correction (0% before RUV correction, 14% after RUV correction).

To study further the effect of RUV on the reproducibility across laboratories, we measured the extent of homogeneity between laboratories in evaluating the same biological effect (see Supplemental Experimental Procedures). The number of genes showing high heterogeneity across all samples ($I^2 \geq 75\%$) decreased as a function of the number of RUV factors that are regressed out from the data (from 6,443 genes before RUV correction to 584 genes after RUV correction on 20 factors, Figure 5A, top). In addition, an increase in overlapping DE genes between cell lines across laboratories was also observed (from 15 before RUV to 243 after RUV correction on 20 factors, Figure 5A, middle; Table S3). The post-RUV PCA plots for each laboratory clearly reveal that the segregation by both time point and genotype is more evident than pre-RUV (Figure 5B). The I^2 measure and variance analysis at the gene expression level before RUV and after RUV (Figures 5A, top and S3) confirm that the laboratory-dependent source of variation was properly removed from the data to expose the variation of interest. Thus, given sufficient power, technical variability, including hidden laboratory-dependent variation, can be corrected and enables detection of the biological signal.

Identification of Experimental Variables Inflating Gene Expression Variance

We next examined the known and investigator-recorded covariates that correlated with the RUV factors. As expected, the variable "SITE" explained 60%, 40%, 38%, and 39% of the variance in factors 1, 2, 3, and 4, respectively (Figure 5C). The second most contributing source of variation captured by RUV was attributable to starting

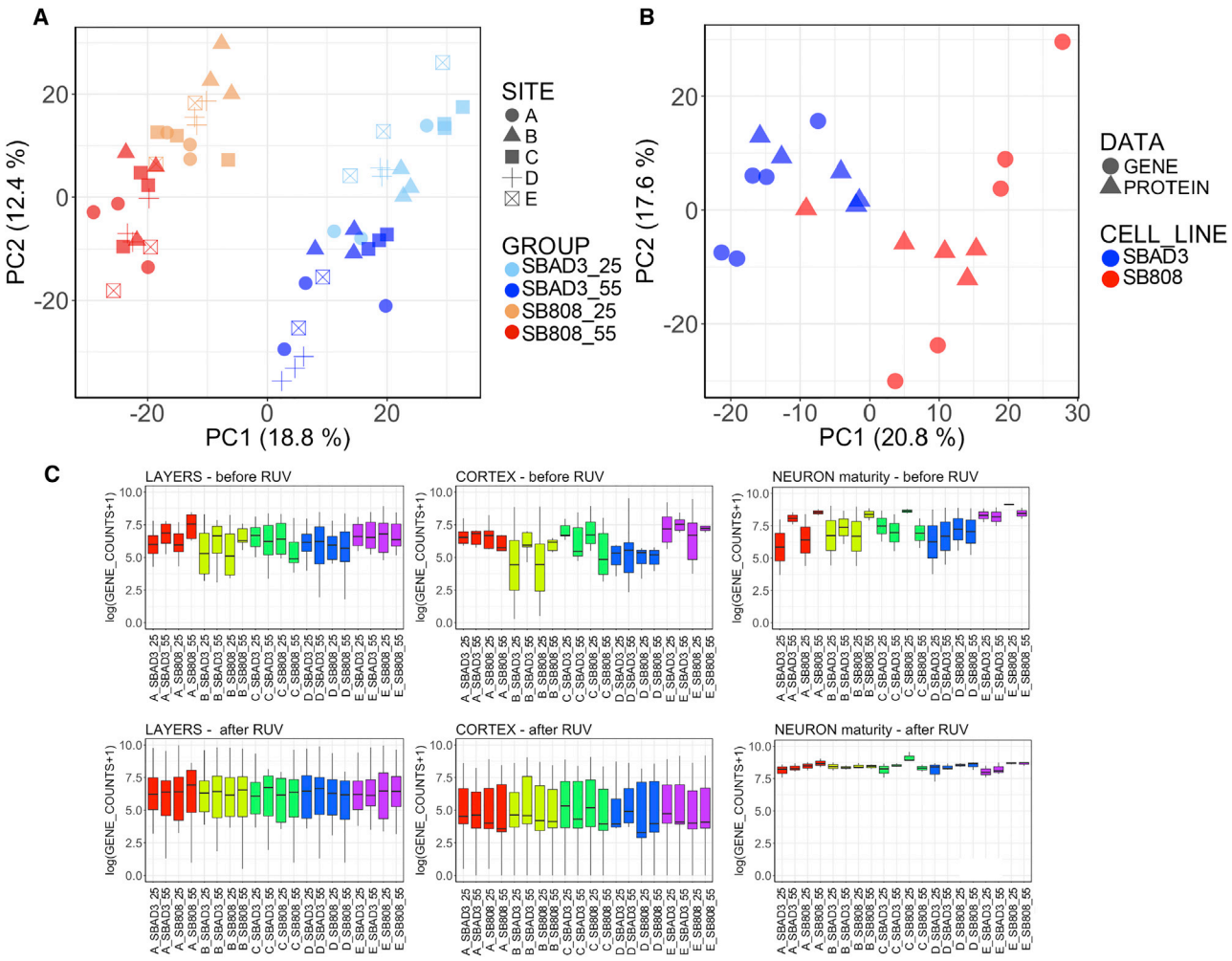


Figure 4. Impact of Unwanted Variance Removal by RUV Correction

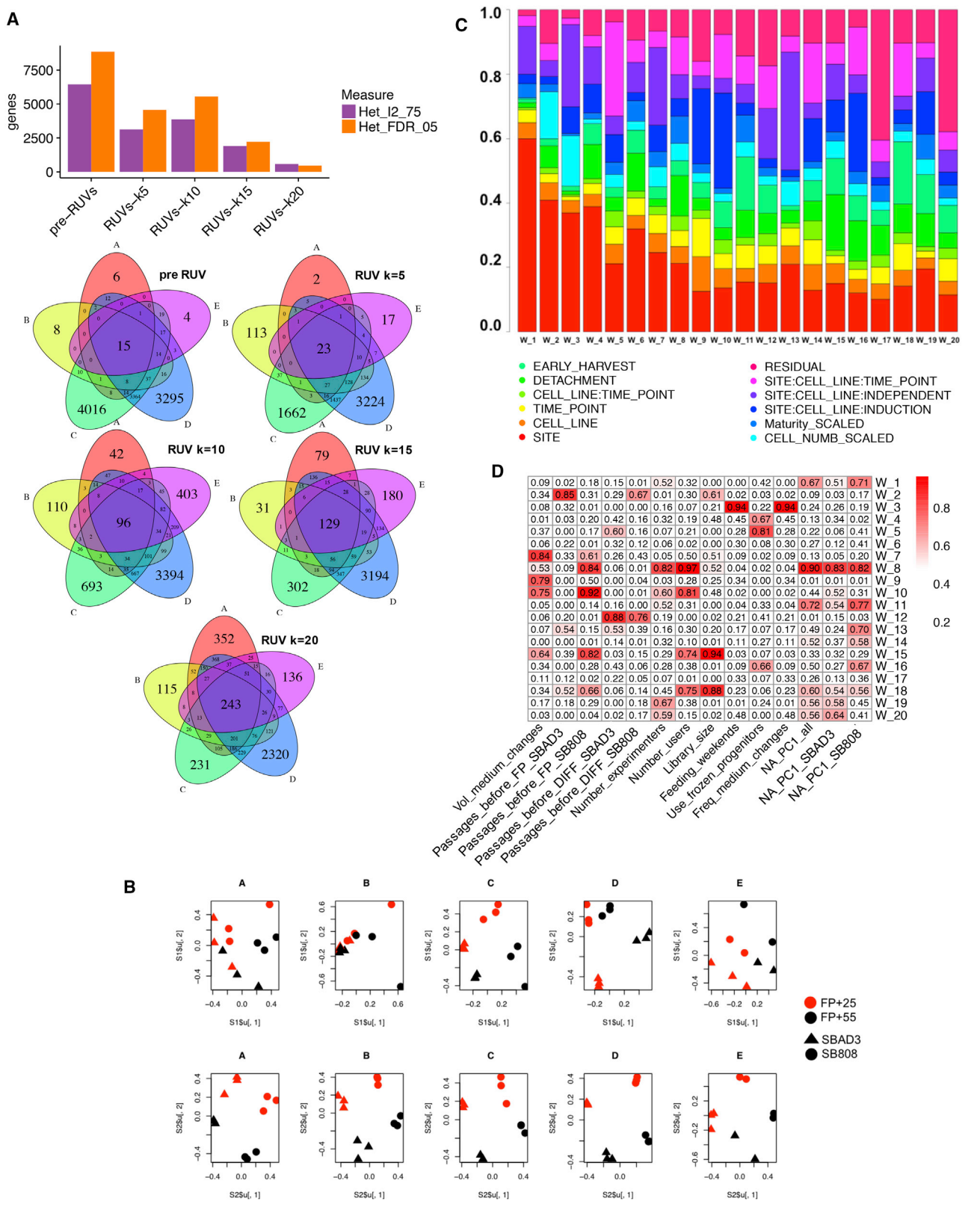
(A) RUV separates sample gene expression profiles by cell line and time point. First two principal components from a PCA on gene expression over all samples after RUV correction.

(B) RUV separates sample gene expression and protein abundance profiles by cell line. First two principal components from a "second" PCA on both pooled gene and protein expression adjusted for PC1 of "first" PCA. This first PC1 captures the differences between protein and gene expression, therefore adjustment makes the two datasets more comparable. Gene expression values and protein abundances are RUV corrected separately on the two datasets for FP + 55 time point.

(C) RUV normalizes the expression of marker genes expected to be similarly expressed across all samples. Gene expression on log scale of gene markers of three neuron-specific stages before (top row) and after RUV correction (bottom row). Box-and-whisker graphs represent distributions, where the span of the box is the interquartile range (IQR) and includes the median (bold line). The ends of the upper and lower whiskers represent the data point with the maximum distance from the third and first quartiles, respectively, but no further than 1.5 × IQR. Data beyond the end of the whiskers are outliers.

the experiment on different days for technical replicates (15%, 22%, 20%, and 38% of the variance in factor 1, 3, 7, and 13, respectively; Figure 5C). In general, the proportions of variance in RUV factors that could be explained by the remaining candidate confounders were moderate to low. Among these, 18% and 15% of variance in factors 2 and 3, respectively, were explained by differences in cell counts.

As SITE was the strongest cause of variability, we attempted to correlate SITE-specific variation in RUV factors to particular experimental effects by fitting linear models regressing site-specific RUV means outputted from the variance component analysis on each site-specific metadata variable in turn. Several covariates, namely iPSC passage number before differentiation and the number of passages before FP, media volume changes, feeding at weekends, and



(legend on next page)

use of frozen neural progenitor cells, were highly correlated with several factors (Figure 5D).

We also examined the variation underlying expression at the gene-specific level by fitting a regression model (MCMCglmm) (Hadfield, 2010) between gene counts and the known covariates. The analysis enabled the identification of genes that may underlie the covariate. The top 100 genes related to "DETACHMENT" were enriched in GO terms related to regulation of cell cycle, apoptosis, and DNA metabolism, while those similarly associated with "CELL_COUNT" are enriched in cellular respiration and lipid metabolism pathways, and those genes related to "TIME_POINT" were involved mainly in neuron differentiation processes (complete lists of GO terms in Table S4).

Cell Type Heterogeneity Is a Major Source of Inter-laboratory Variation

Cellular heterogeneity can be a major confounder in tissue and cell culture comparison (Sandor et al., 2017), and could represent an important source of inflated variance within and between laboratories. To investigate variation in the cellular composition of our iPSC-derived cell populations, we generated the individual transcriptional profiles of 1,440 fluorescence-activated cell sorting (FACS)-sorted iPSC-derived cortical neurons produced by two of five participant laboratories (sites D and E for each of the two cells lines at the FP + 55 time point; see Experimental Procedures). After discarding low-quality cell libraries (see Supplemental Experimental Procedures), 771 SC transcriptomes were available for subsequent analysis.

Using unsupervised hierarchical clustering on the expression profiles we identified four and five subpopulations of cells within the SB808 and SBAD3 cell populations, respectively (see Supplemental Experimental Procedures, Figures 6A and S5A). The cortical differentiation protocol we used has been extensively validated and efficiently produces high yields of cortical excitatory neurons as well as astrocytes. We assessed the presence of neuron-, glial-, and other cell-type-specific markers within each subpopulation (see Supplemental Experimental Procedures). For each line, we found that the largest cell subpopulation was uniquely and significantly enriched in neuron-specific markers (Figures 6A and S5A). The second largest subpopulation was also enriched for a distinct set of neuron-specific markers (cluster 2 for SB808 and SBAD3) (Figures 6A and S5A). Other subpopulations were enriched in astrocyte markers (e.g., clusters 3 and 4 for SB808 and cluster 4 for SBAD3) (Figure S5).

While Shi et al. (2012b), who described the protocol, observed astrocytes forming after day 45, we found here (1) that glial cells represented a large proportion (20.8%), (2) that the fraction of glial cells varied from site to site (15% site D versus 21% site E), and (3) that the fraction of glial cells was higher in the SB808 line than in the SBAD3 line (29% SB808 versus 12% SBAD3).

Cellular Subpopulations Can Show Opposing Differential Gene Expression and Introduce Considerable Bias in Comparative Studies

The SC analysis revealed cell culture subpopulations of differing proportions between two sites (Figure S5A) that could affect the differential gene expression analysis. Interestingly we found that subpopulations 2, 3, and 4 also expressed a small number of genes representing oligodendrocyte or microglia markers (Figure 6A). The expression of genetic markers of other cell types not intended to be induced by our differentiation may represent either a genuine developmental feature of these cells or an artifact of *in vitro* differentiation, where the epigenetic silencing of other cell type-specific genes is not fully effective—in either case the heterogeneity could significantly bias differential gene expression between lines and between sites.

After discounting technical artifacts (e.g., plate effects), we found that the DE genes and pathways varied significantly between iPSC-derived subpopulations. Most notably we observed that gene expression differences between the two lines were negatively correlated between subpopulations 1 and 6, and thus directly obscure the

Figure 5. Analyses of Factors Explaining the Unwanted Variance and Laboratory Heterogeneity

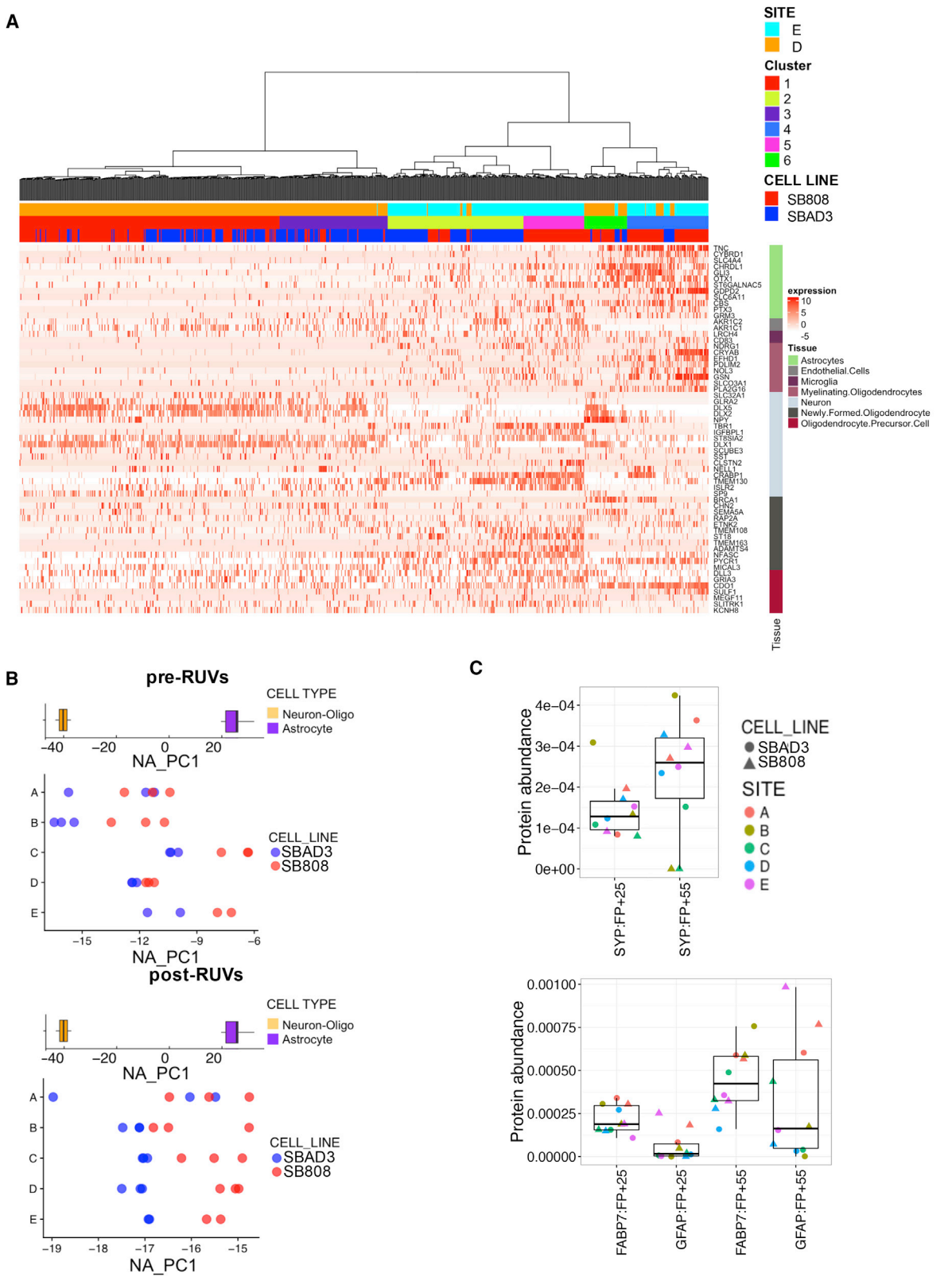
(A) Increased reproducibility of gene expression difference between lines across laboratories after RUV correction. Number of genes showing high heterogeneity across sites before and after RUV based on 5% false discovery rate (FDR) threshold (Het_FDR_05) and on 75% I^2 threshold (Het_I2_75) (Top). Overlap of DE genes between cell lines across sites before and after removal of 5, 10, 15, and 20 RUV factors (Venn diagrams, bottom).

(B) Separation between the lines in singular value decomposition plots helps explain the different number of DE genes between the two covariates of interest before and after RUV.

(C) "Laboratory" is a major confounder corrected by RUV. Each bar summarizes the proportions of variance captured by RUV factors (W_1 to W_20) and explained by known potential confounders.

(D) Laboratory variance is correlated to several experimental variations. The matrix shows the linear correlations between means of laboratory-specific RUV factors and known laboratory-specific potential confounders plus neuron-astrocyte axis scores (NA_PC1) described in Supplemental Experimental Procedures section and in Figure S4B.

See also Figures S3 and S4.



(legend on next page)

detection of DE genes (Figure S5D). To evaluate how the results of DE analyses can be biased by the observed cell heterogeneity, we randomly sampled populations of 100 cells from each line and observed the stability of DE genes and pathways. Of 192 originally observed DE genes, only 10 genes (5%) that were associated with a very low p value ($<10^{-20}$) were consistently reported with a false-negative rate <5% (Figure S5B). This clearly demonstrated that cell heterogeneity can yield a major bias in the comparison of gene expression profiles between iPSC-derived cells and that only the most significant DE genes are detectable through the heterogeneity. Indeed, upregulated and downregulated DE genes for each laboratory in the bulk transcriptomic study were significantly overrepresented among specific subpopulations of the SC study (upregulated in subpopulation numbers 1, 2, and 3, and downregulated in subpopulation numbers 1, 4, and 6, respectively, hypergeometric test; Tables S2 and S3; Figure S5C).

Cellular Composition Varies Both by Laboratory and Cell Line upon Differentiation

Given the results obtained at the SC level, we asked whether the observed heterogeneity in cell identity (neuronal and non-neuronal populations) could explain the variance in the bulk transcriptomic data due to the test site. For this, we used available RNA-seq data from purified human brain cell types to identify gene expression variation associated with cell type and extended the list of marker genes employed in the SC analyses (see Experimental Procedures). Examining the variation between genotypes across all sites, we found that the SBAD3 upregulated genes were enriched in neuronal markers ($p = 2.6 \times 10^{-18}$, after RUV correction), while the SB808 upregulated genes were enriched in non-neuronal markers ($p = 3.2 \times 10^{-11}$ for astrocytes after RUV correction). Furthermore, a clear separation by genotype on principal components reflecting the above-mentioned similarity of SBAD3 lines to neurons and of SB808 lines to non-neuronal cells was evident when the samples were projected on the PCA coordinates of the human brain cell types both before and after RUV correction (Figure 6B). However, before RUV correction, there is significant systematic variation evident in the neuronal/non-neuronal composition in the lines cultured by different laboratories following the same protocol. Indeed, the neuronal/non-neuronal composition of the lines cultured by each laboratory is well-correlated with the "laboratory" contribution to RUV factors 1, 8, and several other factors (Figure 5D). Thus, the predisposition of each line toward generating cell populations with distinct proportions of neurons and non-neuronal cells is an important driver of gene expression differences between the two cell lines at the bulk transcriptome level. This is also reflected at the protein level, whereby glial marker proteins (FABP7 and GFAP) showed increased abundance at the FP + 55 time point and that GFAP is more abundant in SB808 samples than in SBAD3 samples for all laboratories (Figure 6C). Interestingly, variation in neuronal/non-neuronal composition of the two lines did not increase from the early to the later time point when we explained RUV factors through known covariates (Figure 5C) and compared Euclidean distances between time points (Figure S1A).

DISCUSSION

In this study we examined the reproducibility of a long-term neuronal differentiation protocol undertaken in five laboratories. Further, we also intended to identify "hidden" factors important to a robust method and quantify the extent to which they contribute to experimental variation in molecular data. We therefore focused on repeated differentiations of two lines with different genetic backgrounds

Figure 6. Distinct Cellular Populations Are Identified within the iPSC-Derived Neuronal Populations That Potentially Drive Differences in the Bulk Transcriptomic Comparisons

(A) Cellular heterogeneity in individual cells across cell lines and laboratories. The heatmap of single-cell transcription data reveals six distinct cellular populations in SB808 and SBAD3 lines in two laboratories according to their expression of a set of cell identity marker genes (see Supplemental Experimental Procedures).

(B) A neuron-astrocyte axis of gene expression variation illustrates cell type is a major contributor to cell line variation across all laboratories. The projection of before (top) and after (bottom) RUV correction bulk transcriptomic expression patterns (FP + 55) onto neuronal-glia gene expression identity axis (top) shows that glia-neuronal identity contributes to the expression variation (bottom).

(C) Increase of neuron- and astrocyte-specific protein abundances at later time points in all laboratories. Protein abundance of Human Protein Atlas neuron-specific (top) and glia-specific (bottom) proteins. Protein abundances of a neuron marker at different time points (top). A significant increase of SYP is observed from FP + 25 to FP + 55 time point ($p < 0.05$). Protein abundances of two astrocyte markers at different time points (bottom). A significant increase of FABP7 and GFAP is observed from FP + 25 to FP + 55 time points ($p < 0.05$). GFAP specifically shows an increase in all SB808 lines compared with SBAD3 lines within each laboratory. Box-and-whisker graphs represent distributions, where the span of the box is the interquartile range (IQR) and includes the median (bold line). The ends of the upper and lower whiskers represent the data point with the maximum distance from the third and first quartiles, respectively, but no further than 1.5 × IQR. Data beyond the end of the whiskers are outliers.

See also Figure S5.

to assess whether it was possible to consistently distinguish two iPSC lines after neuronal differentiation, using molecular readouts. In our multi-center experiment, we deliberately chose not to investigate multiple disease and control donor lines in order to focus on reproducibility rather than identifying novel disease phenotypes. With a detailed and shared protocol applied across all partner laboratories, we controlled variability in the differentiation process to the extent it is usually disclosed in published protocols, while expecting variation between the cell lines due to their differing genotypes (the sum of all other potential genetic/epigenetic differences between the two lines), as well as originating from specific laboratory practices not harmonized across test sites.

Our approach found (1) that genotype-driven gene expression variation is detectable by a laboratory in particular when within-laboratory consistency is high; (2) that genotypic effects are masked in aggregated molecular data from multiple laboratories due to site-specific confounders; (3) that cell-type compositional heterogeneity varied both by laboratory and by cell line, and contributed significantly to the masking of genotypic effects in multi-site studies; (4) that prolonged cell culture after FP did not significantly increase inter-laboratory variance and that much of the cell-type compositional heterogeneity was likely determined during neural patterning and shortly after FP; and (5) that normalization methods were able to remove nuisance site-specific effects to reveal biological signals including genotypic effects. Our study therefore underscores the importance of recognizing, recording and reporting experimental variables, and, where possible, using appropriate statistical methods to remove unwanted variability, in order to generate more reproducible molecular studies based on differential gene and protein expression phenotypes.

The application of *in vitro* human disease models using iPSC lines is a potentially transformative approach for understanding disease mechanisms, novel target discovery, and personalized medicine. Unsurprisingly, the majority of efforts have been on monogenic forms of disease where there are strong genotype-phenotype relationships due to large effect sizes of the gene mutation. The expectation for such disorders is that, at a cellular level, highly penetrant mutations would cause easily detectable *in vitro* molecular and cellular phenotypes. Although not the focus of our study, when we examined the biochemical phenotype in line SB808, which carried a familial Alzheimer's disease mutation in the PS1 gene, we did indeed detect a highly robust change in specific β-amyloid peptide ratios (Szaruga et al., 2015) when compared with the iPSC line without the PS1 mutation (Figure S6). This strong cross-site reproducibility has likely been observed because (1) the altered production of β-amyloid peptides is proximal to the PS1 mutation, and (2) is due to the robustness of the ratio-based readout.

The molecular analysis by contrast showed very little overlap between sites despite the detailed, shared protocol, and attempts to minimize technical variability, including processing and analyzing omics samples at the same laboratories, enabling us to focus on the differentiation process-related confounders. Within a laboratory, there was much less variance, and gene expression profiles of the two lines clearly segregated even before RUV analysis. Two sites, C and D, showing low levels of dispersion between replicates, produced a large number of genes that were significantly differentially expressed. When considering all 5 sites, only 15 genes were consistently different between the two cell lines prior to normalization, compared with over 200 genes mutually detected after factor analysis-based removal of the unwanted variation. The pre-RUV overlap of the DE genes of sites C and D was quite high, but much of this overlap may have been artifactual, since for site C the number of DE genes fell from 7,524 to 3,354 after removing 5 factors, and to 1,480 genes after removing 20 factors. This suggests that a laboratory could generate "private" gene expression lists with high confidence based upon highly significant p values as in our study, but, unless sources of variance are explored, it is difficult to know whether such DE lists are biologically relevant. This is important because molecular studies by individual laboratories are often used to generate hypotheses for further investigations, and therefore our study raises significant concerns that many of the detected DE genes can be an artifactual.

Addressing this concern, our work found that, despite the numerous sources of potential confounders, it is possible to detect consistently replicated signals if there is a sufficient number of samples to power an appropriate statistical approach and due consideration is given to complexity of iPSC-differentiated cell cultures. The presence of multiple cellular subpopulations differing between two labs was confirmed by the SC transcriptome study. Strikingly, we observed that differential gene expression patterns between the two cell lines in one cellular subpopulations can have the opposite pattern in another subpopulation. We found that, in a simulated heterogeneous bulk transcriptome based on our SC data, only the most significant and strongest gene expression ($p < 10^{-20}$) differences between the two cell lines were detectable. The differing propensity in cellular fates of the two iPSC lines produced by our standardized differentiation protocols was evident in the bulk transcriptome data both before and after removing unwanted variation, demonstrating systematic variation in culture cellular composition associated with both genotype and laboratory. More rigorous quality control of cellular composition upon differentiation at a series of

intermediary time points may help improve the consistency of results between laboratories. However, we also found that different subpopulations within a culture can be characterized by aberrant expression of cell identity markers from cell types that are not present in the culture such as microglia or oligodendrocytes. Immunohistochemistry or functional studies such as calcium imaging or electrophysiology may not reveal these cell subpopulations. These cells may therefore represent a potentially important cause of variance that will be hidden to quality control measures unless these include SC profiling. Altogether our study shows that cellular heterogeneity can introduce significant bias in differential gene expression experiments and likely represents the major contributor to inflating within-laboratory variance and to inter-laboratory variability.

One of the important aspects of our study was to identify those factors which are correlated to the RUV factors, potentially explaining the increased cross-site variability. Not surprisingly, our computational analysis indicated that SITE (i.e., laboratory) is the most influential source of variation (explaining between 40% and 60% of the variance in the first 4 RUV factors), followed by the practice of starting the differentiation of progenitors on different days as opposed to plating on the same day. To further identify the sources of inter-laboratory variability, we correlated site-specific variation in RUV factors to experimental practices known to be different for the various test sites. This analysis allowed us to pinpoint experimental variables which were highly correlated to several RUV factors, and potentially hampered cross-site reproducibility. Among these were a number of factors, some of which are often not disclosed in published differentiation protocols, such as iPSC passage number before differentiation, the number of passages before FP, media volume changes, feeding at weekends, and use of frozen neural progenitor cells. Many of these factors likely alter the epigenetic and cellular programs that determine progenitor cell fate choices, including neuronal-glial balance to thereby contribute to the heterogeneity and variance. Based on our study we strongly suggest that these should be a standard part of every published differentiation protocol to increase the chance of robust reproducibility of iPSC-based studies.

Reproducibility in biomedical science is a major cause of concern and has impacted the pharmaceutical industry, where study reproducibility is a pre-requisite for target discovery, assay development, and a successful drug discovery program. The potentially underlying causes for the lack of reproducibility have been extensively scrutinized and attributed to factors such as poor study design and inappropriate statistical methods, as well as the culture of grant funding and publication biases. Moreover, as our paper illustrates, an individual laboratory conducting hypotheses generating molecular studies, without external reference, or further validation studies cannot know whether their significant differential gene findings are due to a systematic bias in their laboratory or arising from the biological condition under study. It is therefore critical that any potential hypotheses are validated including through the use of literature evidence and increasingly available complementary datasets such as human brain tissue and animal model studies. Collaborative approaches, especially large public-private partnerships involving multiple test centers, if carefully designed, offer a powerful solution to performing studies which yield reproducible mechanistic insights. These multi-center experiments also reveal important learnings for the individual laboratories by identifying experimental practices to be disclosed when publishing iPSC differentiation protocols to increase their reproducibility.

In our paper, we have shown that, while cellular heterogeneity of the iPSC cultures differentiated at various laboratories arising from site-specific practices as well as other cryptic factors can mask almost all biological effects, these confounders can be identified and overcome. The computational biology approaches employed here revealed and removed the site-specific biases, enabled access to the underlying biology, and identified publication best practices.

EXPERIMENTAL PROCEDURES

See further details in the Supplemental Experimental Procedures.

Generation and Maintenance of STEMBANCC iPSC Lines

The human iPSC lines SBAD3-1 and SB808-03-04 (the latter carried the Alzheimer's disease-related PS1 intron 4 mutation) were derived from human skin biopsy fibroblasts following signed informed consent, with approval from the UK NHS Research Ethics Committee (REC: 13/SC/0179) and were derived as part of the IMI-EU sponsored StemBANCC consortium. iPSC generation was performed using the CytoTune-iPS 2.0 Sendai Reprogramming Kit (A16517) from Thermo Fisher Scientific (Waltham, MA).

Bulk Transcriptomic Experiment

For transcriptomic analyses, 12 samples were generated in each laboratory: 3 replicates of the SBAD3 cell line and 3 replicates of the SB808 cell line at each of the 2 time points (Figure 1A). Two samples were excluded from analysis due to problems during the differentiation process and another one because of a contamination issue during RNA-seq, leading to a total of 57 samples being available for transcriptomic analysis.

SC RNA-Seq Experiment

SC Isolation

SC suspensions were generated using Accutase dissociation followed by SC filtration of iPSC-derived cortical neurons. The success

of the suspension was manually confirmed on a hemocytometer. The SC suspension was sorted into a 96-well PCR plate containing a lysis mix. Sorting gates were set to include only live (DAPI-negative) single cells. Stream alignment and sort efficiency was checked using Accudrop beads (Becton Dickinson).

SC RNA-Seq Library Preparation and Sequencing

Single iPSC-derived cortical neurons were isolated by FACS onto 96-well plates in 2 μL lysis buffer (Trombetta et al., 2014). Each plate included 4 bulks, each obtained by extracting total RNA from 4,000 cells using RNeasy Micro Kit (QIAGEN) and using 1/14th of the extracted RNA solution. Libraries were prepared following the Smart-Seq2 protocol described by Picelli et al. (2013) and Trombetta et al. (2014). Each sample was spiked with the equivalent of 1 μL of 1:10,000,000 dilution of the ERCC RNA Spike-In Mix 1 (Thermo Fisher Scientific). Libraries were pooled in 288- or 384-plexes and each pool sequenced on 1 lane of HiSeq 4000 at 75 bp paired end.

Proteomics

Proteomic Sample Processing, Measurement, and Data Analysis

Cells from three wells were detached in ice-cold PBS, pooled, and snap frozen for proteomics analysis. Prior to digestion, cell pellets were dissolved in lysis buffer, and obtained lysates were pooled from triplicate wells, replicated to increase the number of detectable proteins leading to a total of 20 samples for subsequent analysis, and spun at 1,000 × *g*. Samples were subject to in-solution proteolytic tryptic digestion and analyzed using 2D-LC-MS. Proteins were identified using Waters ProteinLynx Global server v.3.0.1 and Progenesis Bioinformatic software (non-linear dynamics) as described previously (Heywood et al., 2015).

ACCESSION NUMBERS

The accession number for the transcriptomic data reported in this paper is GEO: GSE118735.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and six tables and can be found with this article online at https://doi.org/10.1016/j.stemcr.2018.08.013.

AUTHOR CONTRIBUTIONS

V.V. led on the data analysis and data integration and interpretation. C.W. led the overall informatics team. J.S. and V.L. developed the study SOP. A.H., S.E.N., E.W., S.C., A.V., C.C., K.C., C.P., P.B., J.S., W.H., R.D.F., K.J., C.M., A.A., and F.W. performed the experimental work. M.L. and L.A. generated the iPSC lines. A.H., S.H., and S.E.N. performed the SC experimental work. A.B. performed the mass spectrometry proteomics. E.S. and J.N. performed the RNA sequencing. M.A., E.S., and J.N. performed the SC RNA sequencing. C.S. performed the SC analyses. J.S.R. contributed to the omics analysis. G.N. contributed to the statistical analyses and modeling. P.G., P.R., J.A., I.R., P.J.P., G.C.T., M.G., F.J.L., C.J.A., K.M., R.B., C.W., M.Z.C., and V.L. supervised the research teams, contributed to the experimental design and data interpretation. All authors were involved in manuscript writing and editing. C.W., M.Z.C., and V.L. developed the study concept and experimental design and coordinated the analysis.

ACKNOWLEDGMENTS

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115439, resources of which are composed of financial contribution from the European Union's Seventh Framework Program (FP7/2007-2013) and EFPIA companies' in kind contribution. A.H., S.C., and M.Z.C. were also funded by the NIHR (Oxford BRC). K.M. and A.B. were also supported by the NIHR GOSH BRC. The views expressed are those of the author(s) and not necessarily of those of the NHS, the NIHR or the Department of Health. We also thank Dr. Quin Wills and Dr. Davis McCarthy for their advice regarding the processing of our single transcriptomic dataset and David Lakics for creating the illustrations in Figure 1A. Several of the authors (R.D.F., G.C.T., P.R., J.R., K.J., V.L., A.V., C.C., K.C., C.P., I.R., P.J.P., and M.G.) are employed by the pharmaceutical industry, and therefore, as they are affiliated with a commercial entity, declare competing financial interests. J.S. is a shareholder in Talisman Therapeutics. F.J.L. is a shareholder in Talisman Therapeutics and Gen2 Neuroscience. Honoraria or research funding has been received for M.Z.C. from Orion, Daiichi Sankyo, TEVA, and Novartis; P.G. from BioMarin, Actelion, Dipharma, and SOBI; K.M. from Genzyme, Actelion, and BioMarin.

Received: April 24, 2018
Revised: August 21, 2018
Accepted: August 21, 2018
Published: September 20, 2018

REFERENCES

Avior, Y., Sagi, I., and Benvenisty, N. (2016). Pluripotent stem cells in disease modelling and drug discovery. Nat. Rev. Mol. Cell Biol. *17*, 170–182.

Baker, M. (2016). Is there a reproducibility crisis? A nature survey lifts the lid on how researchers view the 'crisis' rocking science and what they think will help. Nature *533*, 452.

Freytag, S., Gagnon-Bartsch, J., Speed, T.P., and Bahlo, M. (2015). Systematic noise degrades gene co-expression signals but can be corrected. BMC Bioinformatics *16*, 309.

Hadfield, J.D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm *R* package. J. Stat. Softw. *33*. https://doi.org/10.18637/jss.v033.i02.

Handel, A.E., Chintawar, S., Lalic, T., Whiteley, E., Vowles, J., Giustacchini, A., Argoud, K., Sopp, P., Nakanishi, M., Bowden, R., et al. (2016). Assessing similarity to primary tissue and cortical layer identity in induced pluripotent stem cell-derived cortical neurons through single-cell transcriptomics. Hum. Mol. Genet. *25*, 989–1000.

Heywood, W.E., Galimberti, D., Bliss, E., Sirka, E., Paterson, R.W., Magdalinou, N.K., Carecchio, M., Reid, E., Heslegrave, A., Fenoglio, C., et al. (2015). Identification of novel CSF biomarkers for

neurodegeneration and their validation by a high-throughput multiplexed targeted proteomic assay. Mol. Neurodegener. *10*, 64.

Kim, Y.J., Zhan, P., Feild, B., Ruben, S.M., and He, T. (2007). Reproducibility assessment of relative quantitation strategies for LC-MS based proteomics. Anal. Chem. *79*, 5651–5658.

Li, S., Łabaj, P.P., Zumbo, P., Sykacek, P., Shi, W., Shi, L., Phan, J., Wu, P.-Y., Wang, M., Wang, C., et al. (2014). Detecting and correcting systematic variation in large-scale RNA sequencing data. Nat. Biotechnol. *32*, 888–895.

Munafò, R., Nosek, B., Bishop, D., Button, K., Chambers, C., Percie du Sert, N., Simonsohn, U., Wagenmakers, E., Ware, J., and Ioannidis, J. (2017). A manifesto for reproducible science. Nat. Hum. Behav. *1*, 0021.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Risso, D., Ngai, J., Speed, T.P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. Nat. Biotechnol. *32*, 896–902.

Sandor, C., Robertson, P., Lang, C., Heger, A., Booth, H., Vowles, J., Witty, L., Bowden, R., Hu, M., Cowley, S.A., et al. (2017). Transcriptomic profiling of purified patient-derived dopamine neurons identifies convergent perturbations and therapeutics for Parkinson's disease. Hum. Mol. Genet. *26*, 552–556.

Shi, Y., Kirwan, P., Smith, J., Robinson, H.P.C., and Livesey, F.J. (2012a). Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. Nat. Neurosci. *15*, 477–486, S1.

Shi, Y., Kirwan, P., and Livesey, F.J. (2012b). Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. Nat. Protoc. *7*, 1836–1846.

Sunkin, S.M., Ng, L., Lau, C., Dolbeare, T., Gilbert, T.L., Thompson, C.L., Hawrylycz, M., and Dang, C. (2013). Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. Nucleic Acids Res. *41*, D996–D1008.

Szaruga, M., Veugelen, S., Benurwar, M., Lismont, S., Sepulveda-Falla, D., Lleo, A., Ryan, N.S., Lashley, T., Fox, N.C., Murayama, S., et al. (2015). Qualitative changes in human γ-secretase underlie familial Alzheimer's disease. J. Exp. Med. *212*, 2003–2013.

Trombetta, J.J., Gennert, D., Lu, D., Satija, R., Shalek, A.K., and Regev, A. (2014). Preparation of single-cell RNA-seq libraries for next generation sequencing. Curr. Protoc. Mol. Biol. *107*, 4.22.1–4.22.17.

van de Leemput, J., Boles, N.C., Kiehl, T.R., Corneo, B., Lederman, P., Menon, V., Lee, C., Martinez, R.A., Levi, B.P., Thompson, C.L., et al. (2014). CORTECON: a temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. Neuron *83*, 51–68.

Stem Cell Reports, Volume *11*

Supplemental Information

Reproducibility of Molecular Phenotypes after Long-Term Differentiation to Human iPSC-Derived Neurons: A Multi-Site Omics Study

Viola Volpato, James Smith, Cynthia Sandor, Janina S. Ried, Anna Baud, Adam Handel, Sarah E. Newey, Frank Wessely, Moustafa Attar, Emma Whiteley, Satyan Chintawar, An Verheyen, Thomas Barta, Majlinda Lako, Lyle Armstrong, Caroline Muschet, Anna Artati, Carlo Cusulin, Klaus Christensen, Christoph Patsch, Eshita Sharma, Jerome Nicod, Philip Brownjohn, Victoria Stubbs, Wendy E. Heywood, Paul Gissen, Roberta De Filippis, Katharina Janssen, Peter Reinhardt, Jerzy Adamski, Ines Royaux, Pieter J. Peeters, Georg C. Terstappen, Martin Graf, Frederick J. Livesey, Colin J. Akerman, Kevin Mills, Rory Bowden, George Nicholson, Caleb Webber, M. Zameel Cader, and Viktor Lakics

Supplemental Figures

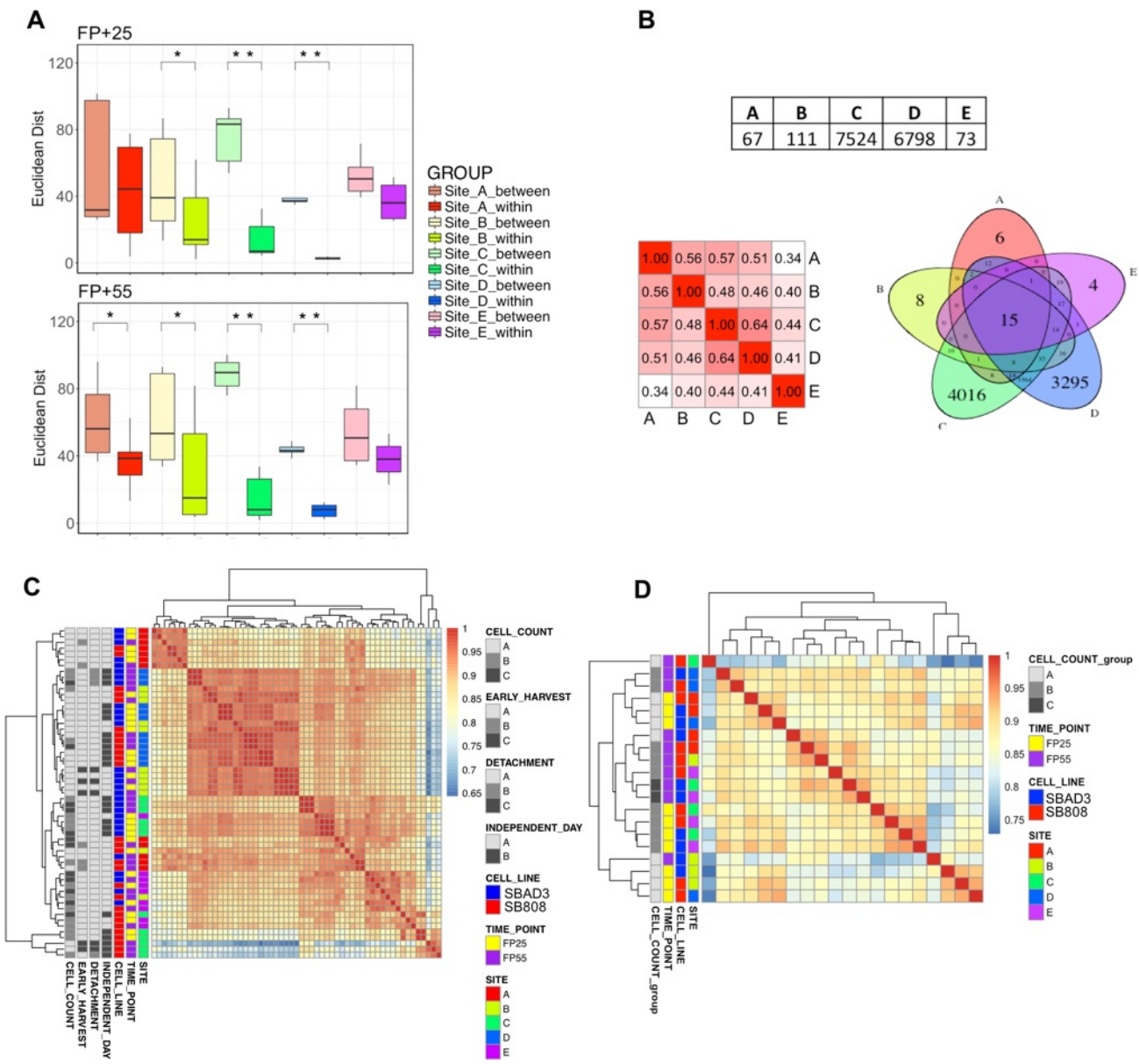


Figure S1, related to Figure 2. Variation in omics readouts reproducibility within and between laboratories.

A) Euclidean distances within cell line replicates and between all replicates of different cell lines. Euclidean distances are calculated between the gene expression profiles of each sample within each laboratory and between replicates of different lines in any laboratory by time point (FP+25 top, FP+55 bottom). Within each laboratory the expression profiles derived from replicates of the same line are significantly closer to each other than those between replicates of different lines for 4 out of 5 laboratories (** p-value<0.005, * p-value<0.05). Box-and-whisker graphs represent distributions, where the span of the box is the interquartile range (IQR) and includes the median (bold line). The ends of the upper and lower whiskers represent the data point with the maximum distance from the third and first quartiles, respectively, but no further than 1.5 * IQR.

Data beyond the end of the whiskers are outliers. **B) Results of differential expression analysis between genotype within each laboratory.** Large difference between laboratories in the number of DE genes between genotypes controlling for time point variation before RUVs correction (top). Only 15 DE genes are found in common between all laboratories indicating a remarkably low degree of cross-laboratory reproducibility (bottom right). Semantic similarity scores between the top 40 enriched GO BP terms in DE genes between genotypes controlling for time point variation before RUVs correction within laboratory (bottom left). **C) Heatmap of pair-wise Spearman's rank based correlations of gene expression between all samples.** Correlations are computed on 13,373 genes expressed across all samples (coloured based on all known covariates both biological and technical, see Methods). **D) Heatmap of Spearman's rank based correlations of protein abundances normalised for total protein amount.** 1,037 proteins observed across all samples were used. As observed for the transcriptomics data, the heatmap did not show clustering of samples by genotype.

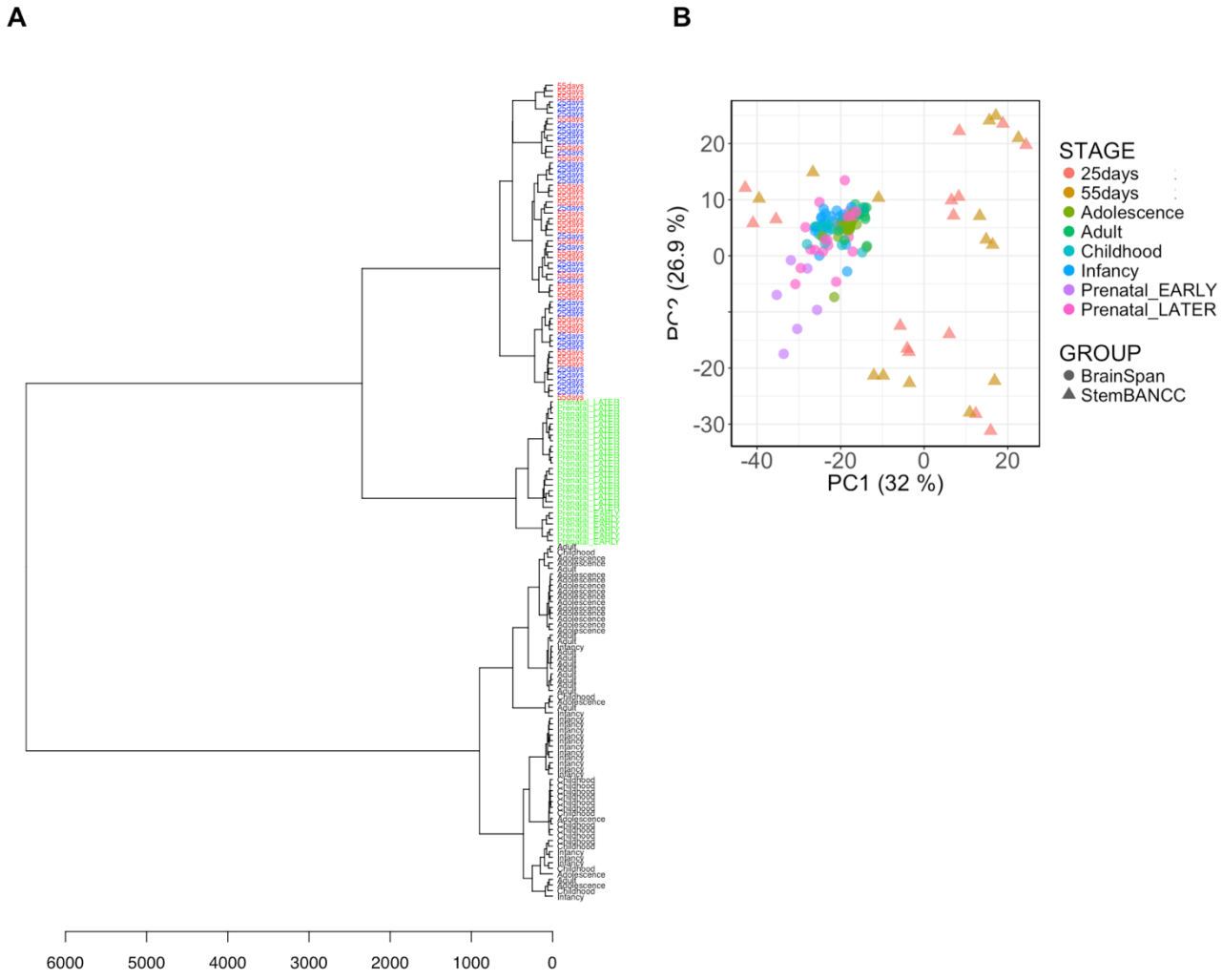


Figure S2, related to Figure 3. Comparison of temporal gene expression profiles of StemBANCC samples with post-mortem brain samples from the BrainSpan Atlas of the Developing Human Brain.

A) Hierarchical clustering of the 57 StemBANCC samples and post-mortem brain samples from the BrainSpan Atlas. StemBANCC samples cluster together with fetal post-mortem brain samples (BrainSpan fetal samples in green, BrainSpan postnatal samples in black, StemBANCC FP+25 samples in blue and StemBANCC FP+55 samples in red). **B) PCA plot of StemBANCC samples from the SBAD3 line before RUVs correction and the BrainSpan Atlas**. BrainSpan samples were projected on principal component axes of StemBANCC CTR samples before RUVs correction. The direction of human neuronal maturation was not recapitulated by non-normalized gene expression profiles.

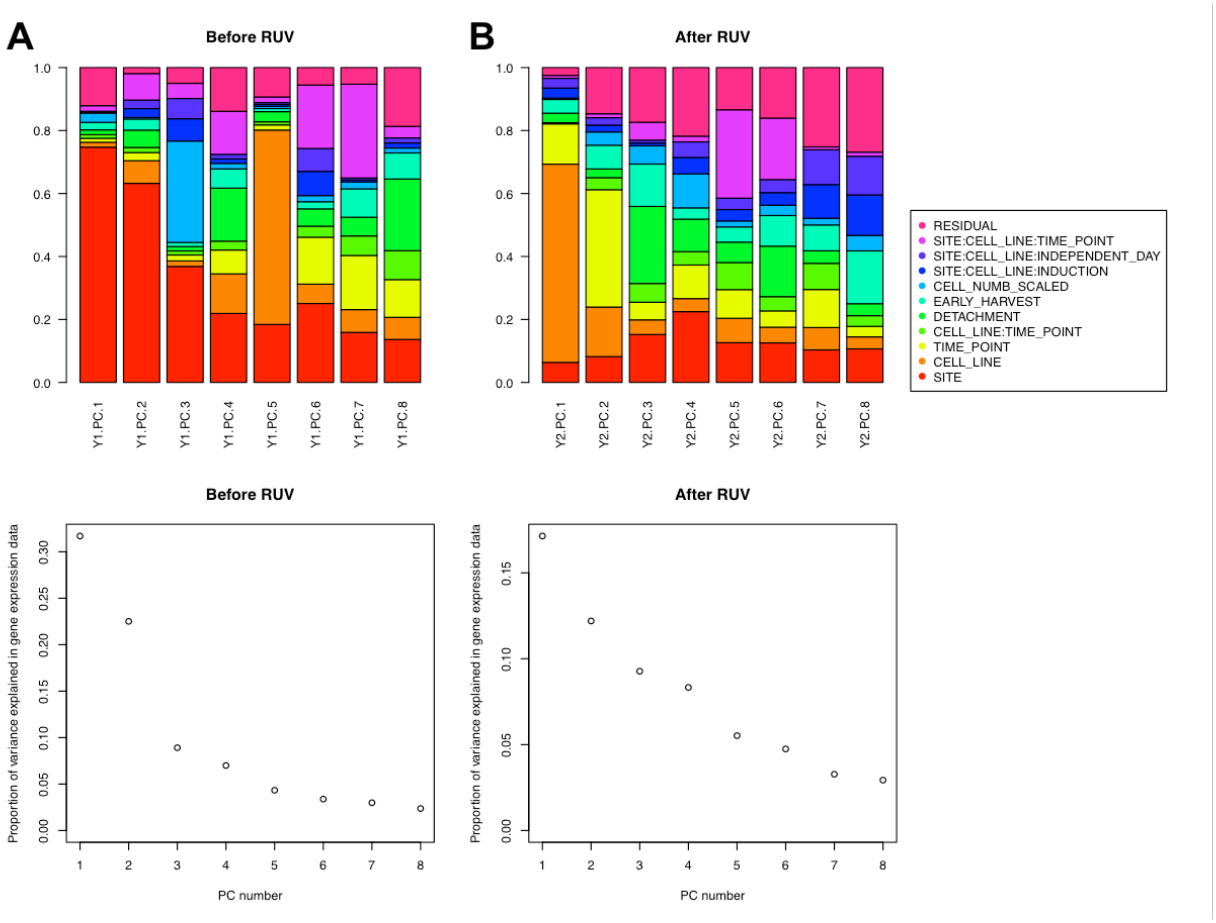


Figure S3, related to Figure 5. Variance component analysis of gene data before and after RUVs correction.

Variance captured by first 8 principal components of gene counts before (A) and after RUVs correction on 5 RUVs factors (B). Proportions of variance explained by known covariates (see Methods) for any principal component (top) and proportions of explained variance in gene expression data by any principal component (bottom) are reported.

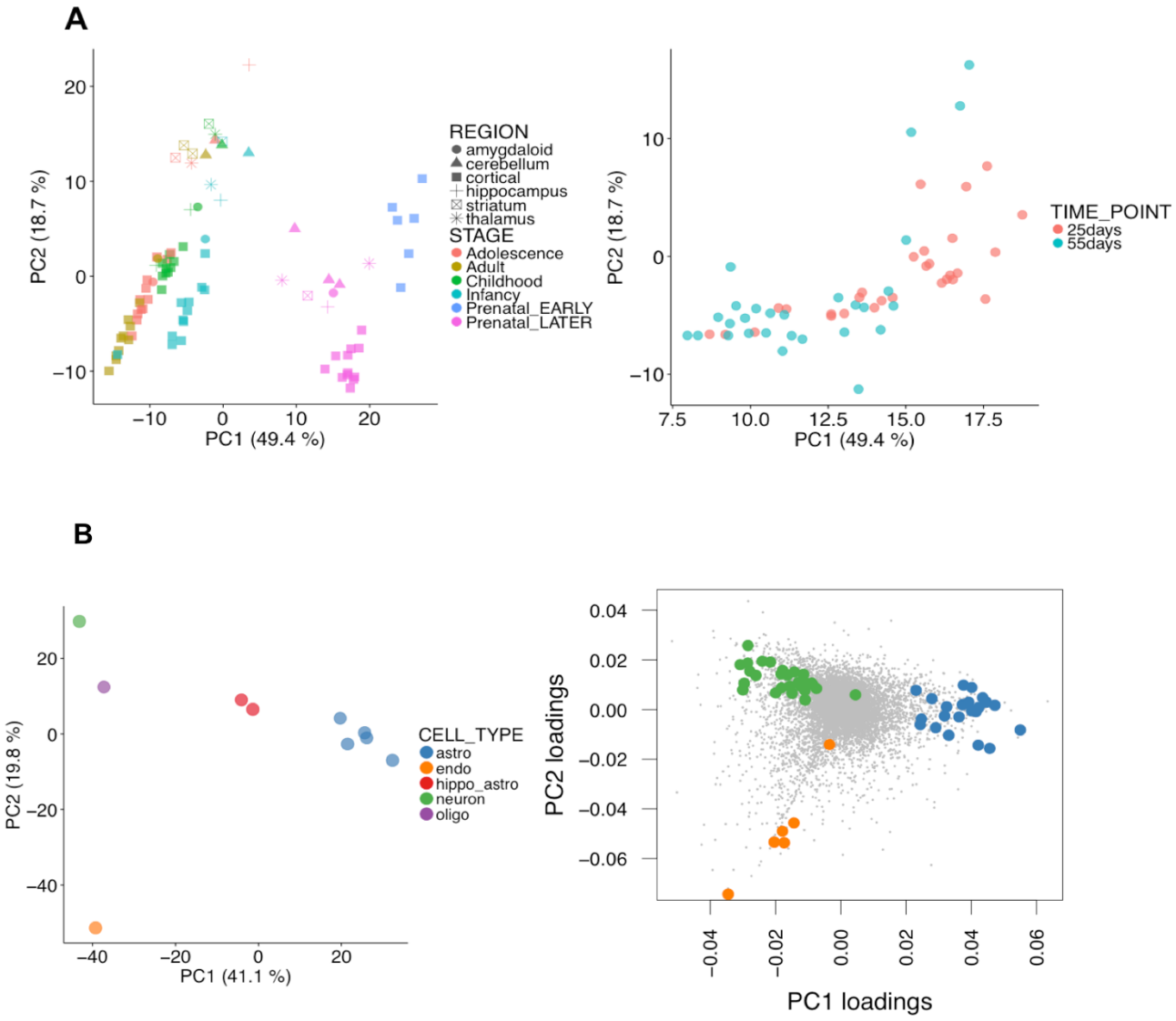


Figure S4, related to Figure 5. A neuronal maturation axis and a neuron-glia axis are derived from external gene expression data and used to explain laboratory variability in StemBANCC samples. A) Identification of a transcriptional axis of maturation.

PCA plot of BrainSpan samples (A) on a set of 787 'cortical marker genes' identified using GTEx data (see Methods) shows that BrainSpan data clearly cluster by sample age. StemBANCC samples were projected (right) on the transcriptional maturation axis (first component) identified in PCA plot (left). Except for the three FP+55 SB808 outlier samples from laboratory C that cluster with the FP+25 samples, the identified transcriptional maturation axis clearly separates samples by time point. The position of StemBANCC samples projected along this axis was used as covariate named 'MATURITY' in subsequent variance component analysis (see Methods). **B) Identification of extended lists of cell-type specific genes.** PCA plot (left) of RNAseq data from purified human brain cell types including neurons, astrocytes, oligodendrocytes and endothelial cells (Zhang et al., 2016). Extended lists of astrocyte, endothelial and neuron markers are identified based on the gene loadings (right) on first and second principal components from PCA plot (left). Reported in bold are mouse cell type specific markers identified as described in Methods (Single Cells section) and the colour code is the same for both figures.

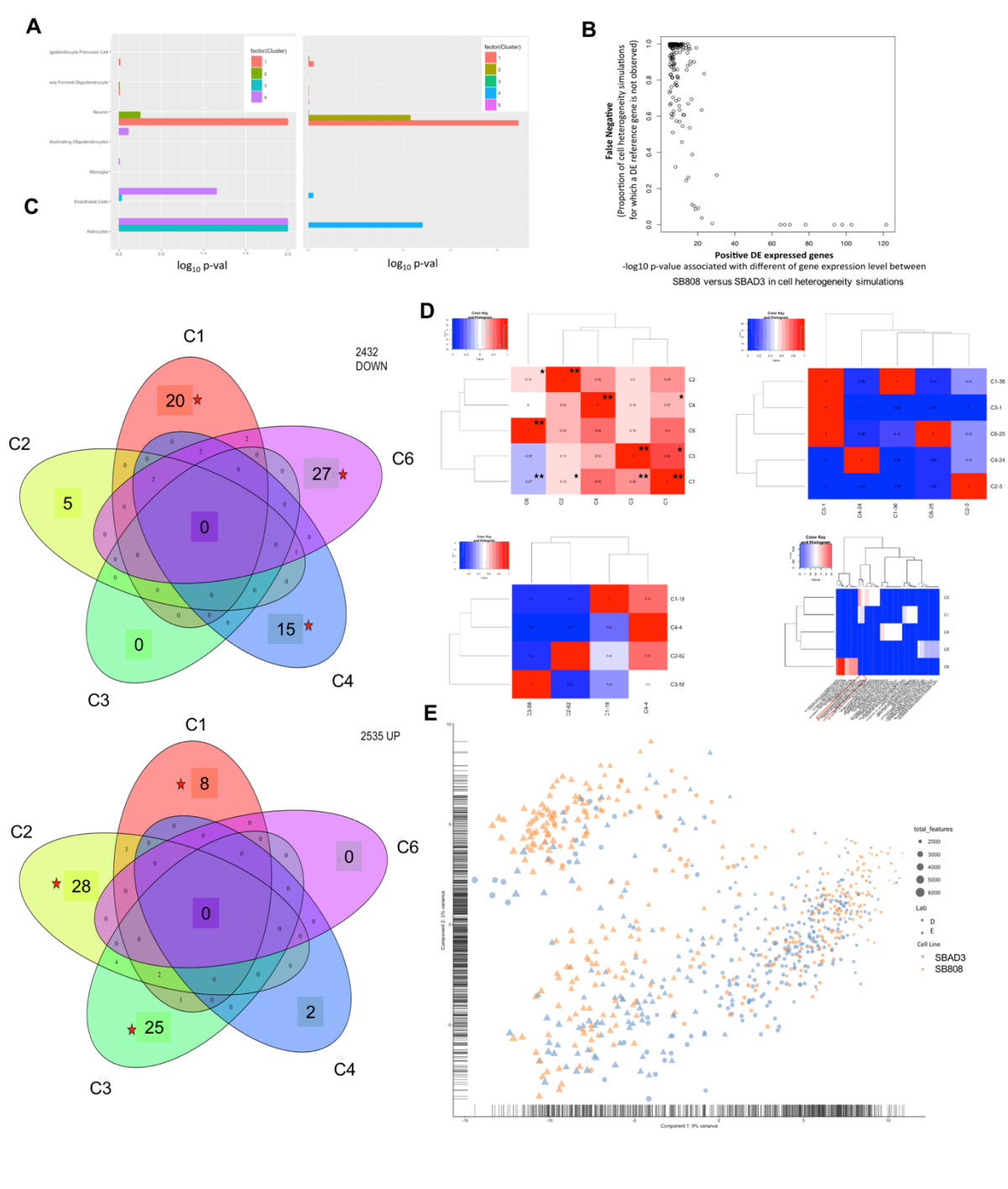


Figure S5, related to Figure 6. Identification of subpopulation in Single Cell data.

A) Cell identity of cell sub-populations in the SB808 (left) and SBAD3 line (right). To classify which brain cell type is resembled by the different subpopulations, we used a transcriptional purified cell mouse cortex catalogue, containing data for neurons, astrocytes, microglia, endothelial cells, pericytes, and various maturation states of oligodendrocytes (Zhang et al., 2014). From mouse gene expression values, we ranked the genes according to their fold change in each cell type. To evaluate the specificity of each population for a specific cell type, we compared the sum of expression values for top50 of cell specific marker gene with those of random sampling of 50 genes. In this figure, we report the log10 of the empiric p-value associated with an enrichment in markers genes of different cell type in the mouse cortex, in the cell sub-populations of the SB808 and SBAD3

line. **B) Proportion of false negative in term of DE genes due to the cell heterogeneity.** We randomly sampled populations of 100 cells from each line to simulated conditions of cell heterogeneity and performed differential expression analyses for each simulation. By considering 396 SB808 cells and 375 SBAD3 cells, we identified 192 DE genes between SB808 vs SBAD3 line with FDR adjusted p-value less than 1%. We then randomly sampled 100 cells from each line to simulated conditions of cell heterogeneity and performed differential expression analyses for each simulation. This graph represents for each 192 DE genes the fraction of simulations where a positive DE gene is not detected as DE gene (y-axis) according the log10 of p-value associated with differential expression level between SB808 vs SBAD3. We found 10/192 DE genes with proportion of false negative less than 5% that means the cell heterogeneity will affect the sensitivity to detect DE of 80% of positive DE genes. **C) Overlap between the differentially regulated genes in the single cell SB808-vs- SBAD3 subpopulations with the Bulk DE genes**. Down-regulated genes (top) and up-regulated genes (bottom) Red star indicates significance. **D) Differentially-expressed genes and pathways between genotypes vary between iPSC sub-populations.** By considering the SB808and SBAD3 cell populations together, we identified six subpopulations of cells by using unsupervised hierarchical clustering approach on the expression profiles. We detected genes DE between SB808 and SBAD3 line within each subpopulation, excepted for the cluster 5 that did not included no SBAD3 cells. **Correlation of the log fold-change of the differentially-expressed genes between subpopulations (top-left)**. For five sets of DE between SB808 and SBAD3 line in each subpopulation (row), we compared their fold-change in four others populations (column) by using a correlation test based on their fold change in both subpopulation. The value in each cell corresponds to Pearson correlation coefficient between fold-change estimation of set of genes in two subpopulations. The single and double stars indicate when the p-value and q-value are less than 5% respectively. **Overlap between down (top-right) and up (bottom-left) regulated DE genes (SB808 vs SBAD3) between different subpopulations.** For each sub-population, we identified down (lower expression level in SB808 than SBAD3 line) and up (higher expression level in SB808 than SBAD3 line) DE genes. We examined what is fraction (number in each cell) of these down/up (row) DE was detected in others sub-populations (column). The right number associated with each subpopulation gives the number of down and up regulated detected in a given subpopulation. We did not find up-regulated genes in C6. **There is little overlap in the GO Biological Processes associated with differential expressed genes between the two cells in each iPSC sub-populations (bottom-right).** We identified the top10 of Biological Process Gene Ontology pathway associated with DE genes of each iPSC subpopulation. For each topGO pathwa, we examined if they were also enriched in DE in others sub-population. The heatmap plot represents the -log10 of p-value associated with enrichment analyses by considering DE in each of five sub-population for these different topGO pathwayset. **E) Principal component analysis on normalized expression matrix (771 SC libraries x 12835 genes).**

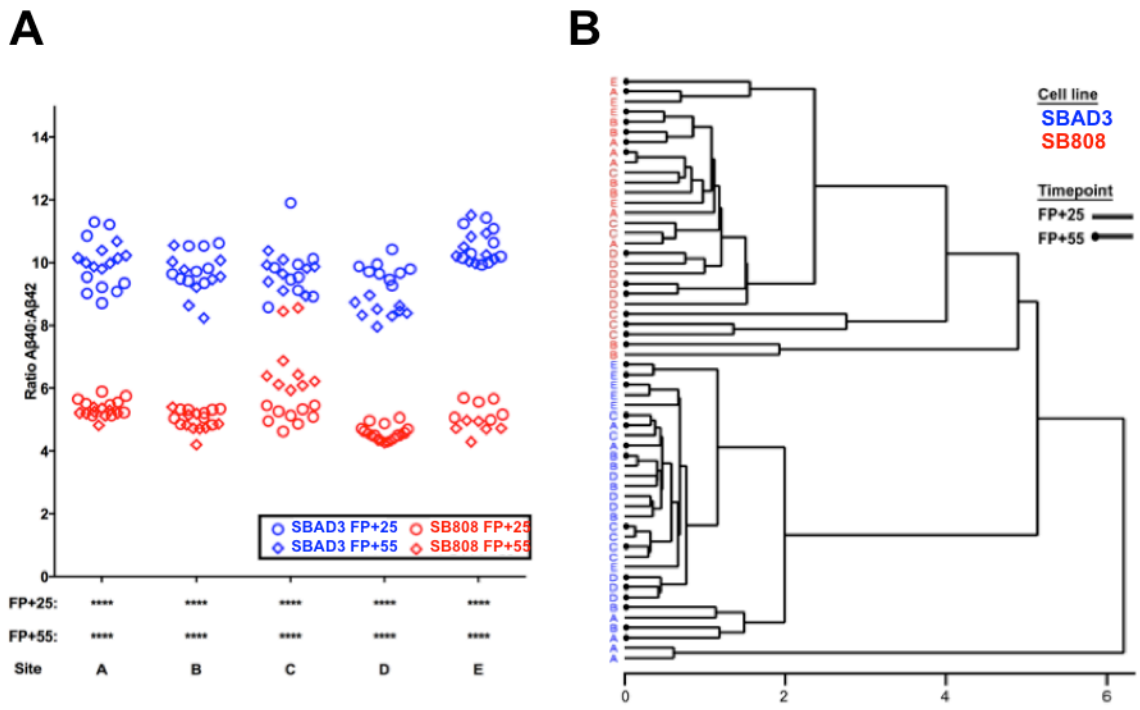


Figure S6, related to Figure 1: Robust reproducibility of a published disease phenotype in neuronally differentiated iPSC-based model system in multiple laboratories

A) Compared to SBAD3, SB808 neurons display a decreased β-amyloid 40/42 ratio at both time points across all test centers. **B)** Hierarchical clustering based on the normalized proportion of each measured β-amyloid species in the supernatants of cortical iPSC-derived neurons. With very few outliers, samples from multiple test centers cluster by cell line rather than test center, indicating the robust reproducibility of this particular disease signal (Duff et al., 1996; Sproul et al., 2014).

Supplemental Methods

Generation and maintenance of STEMBANCC iPSC lines

Fibroblasts were transduced with the reprogramming vectors KOS (a polycistronic vector encoding *KLF4*, *OCT4, SOX2*), h*c-Myc* and h*Klf4* following manufacturer's instructions. One week after transduction, fibroblasts were disaggregated and plated onto feeder layers of mitotically inactivated mouse embryonic fibroblasts in hESC culture medium (KO-DMEM, 20% Knockout™ Serum Replacement, 0.1 mM nonessential amino acids, 2 mM L-glutamine, 100 units/mL penicillin and 8ng/mL human recombinant bFGF, all from Thermo Fisher Scientific) at a density of 8,000 cells per well of a six well plate. The cultures undergoing reprogramming were maintained at 37^{o}C and 5% CO_2 in hESC medium for 2-3 weeks or until colonies with typical hESC morphology appeared. Individual colonies were mechanically dissected and plated onto fresh feeder plates for up to 8 passages before being adapted to feeder free conditions which involved plating on Matrigel-coated plates (Corning, NY) with mTeSR1 media (Stem Cell Technologies, Vancouver, British Columbia, Canada) which was also the medium used in further experiments. A detailed quality control analysis has undertaken in the lines used in the study, which included G-banding, SNP-array-based karyotyping and whole exome sequencing. Mycoplasma testing is routinely performed at each of the participating sites, at the time when they received the cells from the central repository and after passaging. To differentiate iPSCs into cortical neurons, a detailed SOP (see Supplemental Methods) has been used by all 5 partners, based on a previously published method (Shi et al., 2012b, 2012a) Briefly, confluent monolayer iPSCs were induced by dual-SMAD inhibition for 12 days followed by three weeks of progenitor expansion and differentiation. Independent inductions were seeded into 12-well culture dishes at a final plating density of 8.5 x 10^4 cells /cm^2. Samples were collected at 25 and 55 days after final plating (FP+25, FP+55s), and the morphology of the iPSC-derived neurons at 10 days after final plating is depicted in **Figure 1B**.

Assessment of β-amyloid species in culture supernatants

48-hour conditioned media collected from triplicate wells was spun at 1000 x g to remove cellular debris and supernatant stored at -80 °C until use. Samples were assayed for soluble APPβ (MesoScale Diagnostics) and soluble β-amyloid-1-38, β-amyloid-1-40 and β-amyloid-1-42 by multiplexed immunoassay (MesoScale Diagnostics).

<u>Standard Operating Procedure for cortical differentiation of human IPS cells</u>

Readapted from Shi et al 2012a, 2012b.

Product name	Supplier	Catalogue number
BD Matrigel hESC-qualified Matrix	BD Biosciences	354277

mTESR 1	Stem Cell Technlgs	05850
Geltrex (ready to use)	Life Technologies	A1569601
Y-27632 Rock inhibitor	Cell Guidance Systems	SM02-10
Essential-8	Life Technologies	A1517001
Ultrapure 0.5M EDTA	Life Technologies	15575020
Laminin	Sigma	L2020
DMEM:F12 +glutamax	Life Technologies	31331
Insulin (10mg/ml)	Sigma	I9278
2-mercaptoethanol (50mM)	Life Technologies	31350
Non essential amino acids (100x)	Life Technologies	11140
Sodium Pyruvate (100mM)	Sigma	S8636
Pens/Strep (10000 U/ul)	Life Technologies	15140
N2	Life Technologies	17502048
B27	Life Technologies	17504044
L-Glutamine (200mM)	Life Technologies	25030024
Neurobasal	Life Technologies	12348
SB431542	Tocris	1614
Dorsomorphin	Tocris	3093

Neural Maintenance media (1L)

500ml DMEM:F12 +glutamax

0.25ml Insulin

1ml 2-mercaptoethanol

5ml Non essential amino acids

5ml Sodium Pyruvate*

2.5ml Pens/Strep

5ml N2

10ml B27

5ml L-Glutamine* (or glutamax)

500ml Neurobasal

Store at 4°C and use within 3 weeks (*corrected from Shi et al)

SB431542

Supplied as a powder, 10mg, MWt (pure compound) = 384.39

Resuspend to 10mM in DMSO, prepare 50ul aliquots and freeze at -20°C

Use at 1:1000 (10uM)

Dorsomorphin

Supplied as a powder, 10mg, MWt (pure compound) = 472.41

Resuspend to 10mM in DMSO, dilute 10mM stock to 1mM with ddH2O

Prepare 50ul aliquots and freeze at -20°C

use at 1:1000 (1uM)

Neural induction media (10ml)

10ml Neural maintenance media

10ul SB431542

10ul Dorsomorphin

Store at 4°C and use within 5 days

Procedure

(**Step numbers** refer to steps in Shi et al 2012a)

1.1 **Steps 1-22:** Routine maintenance of hIPSCs. Essential-8 media & geltrex can be substituted for mTESR & Matrigel if preferred.

1.2 **Step 23:** Passage cells 2:1 in the presence of Rock Inhibitor. Starting material: 2 wells of a 6 well plate hIPSCs in mTESR (or E8) at 70-90% confluency

- Pre-coat 1 well of a 6 well plate with 1ml matrigel
- Pre-warm 0.5mM EDTA to 37°C
- Aspirate media from two nearly confluent wells of iPSc, wash each with 1ml each of PBS/well (room temp)
- Aspirate PBS and add 1ml pre-warmed EDTA/well then immediately remove
- Add 1ml EDTA, incubate 37℃ 4 – 6 min
- Remove matrigel from coated well (NB don't allow to dry)
- Check cells have start to detach from each other but not from the plastic
- Carefully aspirate EDTA and flush the loosened cells with 1ml mTeSR/10uM ROCKi, using a p1000 tip, moving around the well to ensure even flushing. Don't pipette up and down multiple times, as this will result in the patches disintegrating too much.
- Transfer all 1ml to the new coated well. Repeat with second well of iPSCs and transfer to the new well giving a total vol of 2ml.

- Transfer carefully to incubator, swirling in figure-of-8 to ensure even dispersal of cells

1.3 **Step 24:** Neural induction. (NOTE: inductions have been optimised in 35mm dish or single well of six-well plate). 24hrs after plating, check the cells have reached 100% confluence, wash the cells once with PBS and add 2ml of neural induction medium per well. This is day 0. If the cells are not 100% confluent continue to incubate in mTESR for 1 more day before switching to induction medium. Any gaps in the sheet of cells at this stage will contribute to non-specific differentiation. Refresh induction media daily.

1.4 **Steps 26-31:** On day 12 after induction, the cells should have formed a dense neuro-epithelial sheet (may well appear 'yellow' and 'lumpy'). Passage the cells with dispase as follows:

- Pre-coat 2 wells of a 6 well plate with laminin (1ml per well, 10ug/ml laminin in PBS. Coat at 37°C for 4hrs- overnight)
- Add 200ul dispase stock directly to the 2ml media in the well of the 6-well plate.
- Incubate at 37°C for 3 mins. NOTE: Dispase can be left on as long as 30 mins if sheet is not easy to detach
- Remove cells, keeping sheet as intact as possible by pipetting carefully two or three times from the edge. Clumps should be clearly visible by eye.
- Add 10 ml fresh neural induction medium to a 15ml tube and transfer the clumps into this tube. Allow the clumps to settle in the bottom, then discard the supernatant carefully. Repeat this wash.
- Remove the laminin from the wells, gently resuspend the cells, again without breaking clumps up, in 4ml of neural induction medium and transfer 2ml to each of the two pre-coated laminin wells.
- Incubate the cells overnight to allow the cells to reattach, and change the medium to neural maintenance medium +20ng/ml FGF2 the next day. If the clumps are not attached the following day, they can be transferred to a fresh laminin coated well. Media can now be refreshed at 48hr intervals.

1.5 **Step 34:** After 4 days of FGF treatment, withdraw FGF. Cells can be split 1:2 with dispase when rosettes start to meet, or if neural crest cells begin to appear. Careful dispase passaging should leave non-specific cells attached, and lift off neural rosettes.
Note: There is a possibility that control and disease-specific lines behave differently in terms of differentiation speed, so for the faster line less passages will be needed to reach a stage were the cells are ready for final plating (indicated by the appearance of a critical number of neurons). Careful observation of cell morphology (looking for the appearance of neurons) is crucial to determine how long the first period until final plating should last and this will be likely to be different for the two lines (also reflected by different numbers of dispase and accutase passages for the two lines). This also means that the final plating point (estimated to be D35 but maybe a lot less for a faster growing line) will be considered to be zero, D60 and D90 will be calculated from this time point (e.g +25 and +55 days after this time point). If you prefer to do the final plating at the same time for both lines, the faster growing line can be cryopreserved and thawed later on when the slower growing line is also ready for final plating.

1.6 **Step 42-49:** Passaging to single cells. On day 25 after induction (±1 day), cells can be dissociated with accutase at a ratio of 1:1.

- Precoat well with laminin as above. Remove the medium and wash cells once with 2ml PBS (MgCl2 and CaCl2 free)
- Add 0.5ml Accutase per 35mm well. Incubate the cells in Accutase at 37°C for 5mins.
- Pipette up and down to detach the cells and dilute into 10ml neural maintenance medium. Centrifuge cells at 400g for 5min, repeat the wash and spin, then resuspend in 2ml neural maintenance medium and transfer to laminin coated well.
- Replace the media the day after plating, and every 48hrs subsequently. Cells can be expanded 1:2 when the well reaches 90%-100% confluency (approx. every two to three days)

1.7 **Freezing/Thawing** (optimum stage for freezing is between d28 and d31)

- Following dissociation of the culture with Accutase as described, resuspend cortical stem cells in 1ml neural freezing medium (10% DMSO in neural maintenance media + 20 ng/ml FGF2) per 35mm dish of cells.
- Aliquot 1 ml of the cell suspension into each cryovial.

- Freeze in a CoolCell freezing container at − 80 °C overnight. Transfer the cryovials to liquid nitrogen for long-term storage.
- Thawing NSCs.
- Partially thaw the cells in a 37 °C water bath.
- Transfer the partially thawed NSCs to 10 volumes of room-temperature neural maintenance medium.
- Centrifuge the cells once at 400g for 3 min
- Gently resuspend the cells in 2 ml of neural maintenance medium, and plate into poly-ornithine/laminin-coated 35-mm dishes at 50,000 cells per cm2 (or 1 vial/well). Addition of 20 ng/ml FGF2 to media for the first 12-24hrs after thawing can greatly improve survival.
- Withdraw FGF the following day, and resume culturing of cells as per protocol.

1.8 **Step50, Final plating:** (As neurons are fragile, survival rate after passage is low. For this reason we routinely passage for the final time around day 35). Ideal starting material before final plating is >=60cm2 almost confluent NPs.

Remove the medium and wash cells once with 2ml PBS (MgCl2 and CaCl2 free).

- Add 0.5ml Accutase per 35mm well. Incubate the cells in Accutase at 37°C for 5mins.
- Pipette up and down gently to detach the cells and dilute into 10ml neural maintenance medium.
- Centrifuge cells at 400g for 5min, repeat the wash and spin, then re-suspend in neural maintenance medium.
- Count cells and dilute in neural maintenance media to 3x10^5 cells per ml.
- Remove laminin, and pipette 1ml re-suspended cells per well into 9 wells of each 12w plate. Thus final plating at 300K per well = 85k/cm2
- Leave enough cells to enable quality control with immunochemistry (seed cells on poly/ornithine coated dishes or coverslips to confirm cortical identity, according to your own protocol)

1.9 Replace the media the day after plating, and every 48hrs subsequently, performing full media changes each time, using 1ml media per 12w.

1.10 Add laminin (1/100 in maintenance media, final conc 0.01 mg/ml) every 10 days at d44, d54, d64, d74, d84

1.11 **FP+25:** Sample all media 48hrs after last change (total wells =18, pool into 6 tubes). FP+25 omics samples: row A collect for RNA, row B harvest for protein .

1.12 **FP+55:** Sample all media 48hrs after last change (total wells =9, pool into 3 tubes). FP+55 omics samples: row A collect for RNA, row B harvest for protein .

1.13 **Sample collection.**

- **To sample supernatants for beta-amyloid measurements** (supernatants of the "RNA" samples will be used for MSD measurements)**:**
    - 48 hrs after last media change, remove 1ml media from wells A1, A2 and A3 into three separate collection tubes, spin at 1000g for 5mins to remove cells. Collect the supernatant from each collection tube into lo-bind eppendorf tubes. Store at -80C. Unlike for the supernatants of proteomics and metabolomics wells, for the MSD measurements we keep the three "A" wells separate!
- **To sample supernatants for proteomics**:
    - 48 hrs after last media change, pool 3x1ml media from rows B into two collection tubes, spin them at 1000g for 5 mins to remove cells. Collect 1ml supernatant into lo-bind eppendorf tubes. Store at -80C. From these conditioned media samples.

- **To sample cells for transcriptomics (totalRNA):** (estimate 3ug/well)

- Pre-warm PBS wash at 37C, pre cool PBS collection buffer on ice.
- Wash well with warm PBS. Aspirate off.
- Add 1ml ice cold PBS to well and pipette to detach cells (with scraping if needed, but this is not usually necessary).
- Collect and pool PBS/cells from all three 'A' samples. Spin at >=1000g for 5mins, discard supernatant. Lyse cells in 600ul RLT buffer plus Beta-mercaptoethanol (as per Qiagen RNeasy Mini) and pipet to mix. Homogenize the lysate using QIAshredder spin column, and store homogenized lysate at -80C until needed.
- Continue to purify RNA as per RNeasy protocol, eluting in 30ul RNase free water. Quantify using Qubit RNA broad range assay.

- **To sample cells for proteomics:** (d90 estimate >500ug protein per well)
  - Prepare 37C PBS, (wet)ice cold PBS, and ethanol/dry ice bath
  - Wash well with warm PBS. Aspirate off.
  - Add 1ml ice cold PBS to 'B' wells and pipette to remove cells
  - Aspirate cells and collect in lo-bind tubes. Spin at 1600g for 5mins, discard supernatant and snap cool pellet on ethanol/dry ice bath. Store at -80C

1.14 **QC:**

For QC purposes, collect phase contrast images of each induction frequently, in particular at neuroepithelial sheet stage, on appearance of rosettes, after dissociation to single cells, and pre/post final plating.

Immunostaining for TBR1 and CTIP2 after day 45 (FP+10) to confirm cortical identity.

Example immuno protocol:

Fix with 4% PFA 20min RoomTemp

Wash with TBST (1x TBS plus 0.3% Triton) 5min RT

Repeat wash 3x

Block 4% Goat serum/TBST 1hr RT

Primary antibodies overnight 4C in blocking buffer

(Final concentration: ab31940 at 1ug/ml, ab18465 at 1ug/ml)

Wash 3x with TBST 5min RT

Secondary antibodies in TBST 1hr RT (1:500)

Wash 3x with TBST and counterstain with DAPI if required.

1.15 MSD kits used:
K15200G-1 - Aβ Peptide Panel1 (6E10) V-PLEX Plus Kit: 25ul sample

K15120E-1 - sAPPalpha/sAPPbeta Kit: 25ul sample

One neural induction refers to a single well of a 6-well plate at d0. Multiple wells may be derived from one induction but replicates should be always maintained separately after this stage.

Bulk RNAseq Analysis

Sample Collection

Pooled cells from three wells were detached in ice-cold PBS and total RNA was extracted using the RNeasy Mini Kit (Qiagen) following the manufacturer's instructions.

Sequencing, Mapping and gene count estimation:

All sequencing was carried out in single partner centre on an Illumina Hiseq4000 obtaining 75PE reads. Basic quality control screenings on unmapped reads and sequence mapping were performed through CGAT pipeline pipeline_readqc.py. The quality of the sequencing was assessed by FASTQC software (version 0.9.3), (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). RNA-seq data were mapped to the hg19 assembly via STAR version 2.2.0c (Dobin et al., 2012). Read alignments were merged in single BAM file output per sample (57 in total). Reads were filtered to remove those not uniquely mapped (mapping quality equal to 255) and all ribosomal and mitochondrial RNA reads. Gene-level read counts were obtained using FeatureCount (Liao et al., 2014). Cuffquant and Cuffnorm tools from Cufflinks program (Trapnell et al., 2010) were used to calculate fragments per kilobase per million reads (FPKM) from the merged BAM files. Only for a direct comparison between pooled proteomic data (20 samples) and transcriptomic data (57 samples), BAM files from replicate samples were merged and gene quantification process was repeated as before.

Data normalisation

Upper quartile (UQ) normalisation has been applied (in-house code) on raw gene counts to correct for library size differences. UQ normalised counts have been used throughout the paper when considering not RUV corrected gene data. UQ normalised counts were also used as input for RUVs correction and differential gene expression analysis.

Analysis of sources of variation

To minimize the impact of unwanted sources of variation we used methods implemented in the R-Project packages EDASeq(Muschet et al., 2016) and RUVseq(Heywood et al., 2015). The following steps were applied: first raw count data were normalized by upper quartile (UQ). Second, RUVs (remove unwanted variation method) was used to infer factors explaining transcriptome-wide variance components. Information about replicates' structure was given as input to RUVs in order to retain variation coming from the covariates of interest (cell line and time point). Third, these factors were regressed out from the UQ-normalised gene counts (normCounts function used from EDASeq package) and RUVs corrected gene data were used to perform further analyses. The same approach was applied to samples from the control cell line only to expose a "clean" time point signal regardless of any cell line effect.

To explain the variance captured by PCA principal components before and after RUV correction and the RUVs factors in terms of meta-data variables, an over-dispersed Gaussian response model was used. We fitted a Bayesian generalized linear multilevel model using the MCMCglmm R package (Hadfield, 2010). Eleven meta-data variables, or covariates, (SITE, CELL_LINE, TIME_POINT, DETACHMENT, EARLY_HARVEST, CELL_COUNT, MATURITY, CELL_LINE:TIME_POINT, SITE:CELL_LINE:TIME_POINT, SITE:CELL_LINE:INDUCTION, SITE:CELL_LINE:INDEPENDENT) were modelled. Posterior samples of variance proportions were obtained by standardizing the sum of posterior variances across covariates to sum to one. By fitting regression models between the first twenty RUVs factors and the covariates, we were able to estimate proportions of variance captured by RUVs factors and explained in terms of known covariates. Notably, since the factors were estimated based on genotype and time point replicates, the variation coming from these two biological covariates of interest was marginal as expected from our application of RUVs. The residual variance

was generally low in almost all RUVs factors, suggesting that most of the variation captured by RUVs is attributable to the confounding factors we modelled.

The same approach was used to explain variance attributable to each gene (before RUV correction) in terms of meta-data variables by using an over-dispersed Poisson response model. Genes were ranked based on proportion of variance explained by each meta-data variable and the top 100 were used for functional enrichment analysis (GO pathways). To dissect variation captured by any RUVs factor that we found to be explained by SITE-origin, SITE-specific means of proportions of variance were correlated to SITE-specific meta-data variables by fitting linear regression models.

Differential gene expression analysis

Within-laboratory differentially expressed (DE) genes were estimated before and after RUVs correction using Limma moderated t-statistic (Limma R package, Ritchie et al., 2015) for either cell line or time point effect correcting for the experimental structure (i.e. design formula defined as "~CELL_LINE + TIME_POINT + CELL_LINE:TIME_POINT") at FDR <= 0.01. Multi-laboratory DE genes were estimated as before but also correcting for laboratory origin before and after RUVs correction. In both cases, UQ normalized data was used as input for Limma.

Heterogeneity analysis

To quantify the effect of heterogeneity across laboratories we used the I^2 quantity(Higgins et al., 2003) that describes the percentage of total variation across experiments. I^2 is calculated based on Cochran's Q(Higgins and Thompson, 2002) test for the null hypothesis that all experiments identify the same effect (cell line and time point effects in the present study). FDR adjusted p-values (Benjamini–Hochberg) are calculated for the test. I^2 values range between 0% and 100%, where 0%, 50% and 75% correspond to no, moderate and high heterogeneity, respectively(Higgins et al., 2003). We used voom, lmFit and eBayes functions from Limma R package(Ritchie et al., 2015) to estimate the effects of interest and extract standard deviations for any effect and in any gene. Q and I^2 measures were then calculated for either cell line or time point effect correcting for the experimental structure, as described previously (Higgins and Thompson, 2002; Higgins et al., 2003).

Comparison with public data

Gene expression profiles were compared to public RNA-Seq data sets (FPKM) from the Genotype-Tissue Expression Portal (GTEx consortium, 2015) [downloaded on September 2015] and the BrainSpan Atlas of the Developing Human Brain (Miller et al., 2014). GTEx data were averaged by tissue and BrainSpan data were averaged by tissue and age to obtain six main age groups (early-prenatal, late-prenatal, infancy, childhood, adolescence and adult). Both data sets were corrected for batch effect, using ComBat function from sva R package(Johnson et al., 2007) , and log-transformed for Principal Component Analysis (PCA) (scaled and centered) and Hierarchical Clustering Analysis (scaled, Euclidean distance, Ward method). CORTECON(van de Leemput et al., 2014) gene clusters for temporal cortex development were downloaded from the database available at http://cortecon.neuralsci.org. Genes assigned uniquely to any stage-specific cluster were used in our enrichment analysis.

Identification of cortical markers and a transcriptional axis of neuronal maturation from GTEx and BrainSpan datasets

A set of 787 'cortical marker genes' was identified using GTEx data as genes showing at least five-fold higher RPKM level in three GTEx brain cortical tissues as compared to a group of at least 40 (~80%) "non-cortical" GTEx tissues. A transcriptional maturation axis was then identified on this set of cortical marker genes from PCA of BrainSpan data that clearly cluster by sample age (**Figure S4A left**). The position of STEMBANCC samples projected along this axis (first principal components) was used as covariate named 'MATURITY' in subsequent variance component analysis (**Figure S4A right**).

Identification of cell type specific markers

We used RNA-Seq data(Zhang et al., 2016) from purified human brain cell types including neurons, astrocytes, oligodendrocytes and endothelial cells available at http://www.brainrnaseq.org. In a PCA of these data (**Figure S4B left**), principal components 1 and 2 distinguish between three main groups namely neurons and oligodendrocytes, astrocytes and endothelial cells. Given that positive control cell type markers defined as in section '**Brain cell class of different iPSC subpopulations**' (Single Cell Methods) lie correctly along the identified axes of PCA gene loadings (**Figure S4B right**), we derived extended lists of cell type-specific genes contributing to PCA coordinates of the respective cell type groups. We defined different sets of cell type specific genes using different stringency thresholds on the gene loadings. These sets were compared to DE genes from our samples using a hypergeometric test for significant overlap.

GO pathway enrichment analysis

We performed a classical enrichment analysis by testing the over-representation of gene ontology biological processes (GO BP) terms within the group of differentially expressed genes using a Fisher test. Semantic similarity between lists of enriched GO terms was calculated using GOSemSim R package(Yu et al., 2010) (Wang method(Wang et al., 2007)).

Implementing RUVs

RUVs (Risso et al., 2014) assumes that the biological covariates of interest are constant across replicates. Using a set of negative control samples, or replicates, to column-center the counts RUVs estimates sources of unwanted variation on a set of control genes. In our analysis we used all genes that are expressed (count>=1) across all samples. Information about replicates' structure is given as input to RUVs. A further tuning parameter for RUVs is the number of k estimable factors whose choice should be driven by sample size, extent of technical effects and of differential expression. We estimated a maximum of 20 RUVs factors that we analysed in terms of variance decomposition, improvement of site reproducibility and increase in number of DE genes between covariates of interest. However, when referring to RUV corrected gene counts in all other analyses we intend gene counts normalised on the first five RUVs factors. This corresponds to a first clear clustering of samples on PCA plot based on the two covariates of interest.

Variance component analysis

To explain the variance captured by any RUVs factor in terms of meta-data variables we fitted a Bayesian generalized linear multilevel model using the MCMCglmm R package (Hadfield, 2010). Quantitative explanatory covariates were "CELL_NUMB_SCALED" and "Maturity_SCALED" (scaled to zero mean and unit variance).

The "Maturity" covariate consisted of scores on the first principal component of the 57 samples on a set of cortical marker genes to reflect the neuronal maturation axis (**Figure S4A, see Methods**).

Categorical covariates were:

- SITE, modelling variation across sites,
- CELL_LINE, modelling site-homogeneous variation across cell lines,
- TIME_POINT modelling site-homogeneous variation across time points,
- CELL_LINE:TIME_POINT modelling site-homogeneous interaction between time point and cell line
- DETACHMENT,
- EARLY_HARVEST,
- SITE:CELL_LINE:INDUCTION, modelling inter-induction variation, having one level for each of 29 inductions across the whole experiment
- SITE:CELL_LINE:INDEPENDENT_DAY, modelling day-specific inter-induction variation, i.e. attributable to inductions being performed on different days (as was the case at three of the five labs), having one level for each of the 21 (site, cell line, day of induction) triples in the experiment.
- SITE:CELL_LINE:TIME_POINT, modelling site-heterogeneous variation across cell lines and time points, one level for each of the 20 (site, cell line, time point) triples in the experiment.

Each of the covariates SITE:CELL_LINE:INDUCTION, SITE:CELL_LINE:INDEPENDENT_DAY, SITE:CELL_LINE:TIME_POINT was modelled hierarchically (as a "random" effect) with its own variance component. Variance components were allocated non-informative Inverse-Gamma priors (shape = 0.01, rate = 0.01). Other covariates were treated as "fixed" effects, with parameters given non-informative priors of Gaussian distributions with zero mean and standard deviation set at 100 times the SD of the model's dependent variable. The models were fitted using Markov Chain Monte Carlo (MCMC), with samples collected for 500,000 iterations (with a thinning interval of 100) after a burn-in phase of 50,000.

At each thinned MCMC iteration, samples were saved and extracted from the posterior distributions of fixed and random effects and used to estimate the posterior distribution of variance proportions. Variances were extracted as follows: for any particular covariate in the linear model (fixed or random) encoded by the design matrix X and with parameters β (with β modelled hierarchically in the case of a random covariate), the variance attributable to the effect at MCMC iteration j was quantified as the sample variance of the fitted values at that iteration, i.e. $\mathrm{var}(X\beta^{(j)})$. Posterior samples of variance proportions were obtained by standardizing the sum of posterior variances across covariates to sum to one

Marker Genes

The neuron-specific stage markers used to illustrate the relative cell culture heterogeneity before and after RUV (**Figure 4B**) were as follows:

Cortex: FOXG1, FOXP2, LHX2, OTX1, EMX1, OTX2, LHX9, EMX2

Layers: TBR1, OTX1, CTIP2, FEZF2, ETV1/ER81, SATB2, CUX1, RORB, BRN2, FOXG1

Mature: SYP, SLC17A7, DLG4

Other analyses

Principal component analysis is performed on log-transformed gene counts through prcomp R function (center=T,scale=T). Hierarchical clustering is performed through hclust R function (on scaled data, Euclidean distance, Ward method). SVD single value decomposition analysis is performed on scaled gene counts through svd R function. Heatmaps are created using pheatmap R function.

Single Cell RNAseq Analysis

Quality control of single cell RNA sequence data

We checked the quality of RNA sequencing data using the FASTQC software (version 0.9.3) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) via the CGAT pipeline pipeline_readqc.py and reported the summary results of FASTQC by plate and by library**.**

Exons genes annotations file

We generated annotations within the ENSEMBL gene set after reconciliation with the UCSC genome assembly from human genome (hg19) by using the CGAT pipeline pipeline_annotations.py. The generated gtf file provided the information regarding exon parts of transcripts. This set includes both coding and non-coding transcripts. Coding transcripts span both the UTR and the CDS. We merged this file with ERCC-spike-ins annotations.

Alignment and quality control of alignment of single cell

We aligned our RNA sequences to the human genome (hg19) using STAR (version 2.3.0). STAR is a mapper developed for RNA-seq data and is able to ignore adapters by clipping. We generated the index required by STAR using the following options:

--runMode genomeGenerate

--genomeFastaFiles genome softmasked fasta file (hg19)

--sjdbGTFfile gtf containing all known gene models (generated with CGAT pipeline pipeline_annotations.py)

--outFilterType BySJout

We aligned reads with the CGAT pipeline pipeline_mapping.py (option: make mapping) using STAR default options and:

-- runMode alignReads

-- genomeLoad LoaDaNsdRemove

-- outStd SAM

-- outSAMstrandField intronMotif

-- outSAMunmapped Within

-- outFilterType BySJout

For the batch1, the two bam files of each of the eight libraries coming from two lanes were then merged by using samtools (version 1.8) by running the CGAT pipeline pipeline_mapping.py with the option make mergeBAMFiles. We compiled the statistic regarding the quality of mapping by using make buildBAMStats of CGAT pipeline pipeline_mapping.py.

Count read overlapping exon annotations and basic Metrics

Uniquely mapped read pairs were counted using featureCounts subread-1.5.0 by using the exons annotations generated by approach described above. To evaluate the sequencing output and the amount of usable data, we used some metrics reported by featuresCounts including sequencing depth (Number of sequenced read pairs - Count), percentage of mapped reads, numbers of mapped reads aligning to various annotated genomic features, namely non-exonic coverage (No_feature), exonic-coverage (Ambiguous_mapping, Multiple_mapping and uniquely mapped reads- ENSEMBL_Genes).

We observed a low average proportion of uniquely mapped read (38%). The 5' to 3' coverage plots show mostly uniform coverage of all samples with only a few potentially failed samples (or blanks) that show spiky coverage. The coverage plots do not suggest degradation for most of the samples.

Quality control and cells filtering

All QC metrics and plot diagnostic were computed by using the R package scater (1.0.4) (doi: http://dx.doi.org/10.1101/069633). Firstly we noted that for the batch1 including four plates, the libraries were associated with low libraries complexity measure. We determined whether RNA in each captured cell was degraded by studying the total % of mapped reads compared to the proportion of reads mapped to spike-in molecules.

We excluded the libraries with:

(1) less than 2000 expressed genes
(2) low complexity, where the % of 200 most expressed features (genes and ERCC-spike in) represented more than 50% of total number of counts.
(3) low % of endogenous RNA, for which the % of ERCC spike-in > 14%
(4) low number of mapped read for which the total counts < 10^6
(5) bulk libraries that was used as control libraries.

By using these filters we removed 669 libraries with aberrant patterns from downstream analyses. Please Note that all libraries from batch 1 (batch test) prepared with higher concentrations in ERCC spike-inns and TSO than others batch were discarded.

We performed QC diagnostic at feature level. We observed 1197 features was observed with detectable expression in 50% of libraries and that the top 20 expressed features (including 10 ERRC spike-ins) consumed ~ 25% of reads. We removed the features for which the means of counts where less than one and thus considered 12825 features. We then computed the normalized log-expression values with R package scran (1.0.4) by adding one to each count, dividing by the size factor for that cell, and log transforming.

Important explanatory variables

From normalized expression values matrix (771 single cell libraries x 12835 genes), we then identified variables that drive variation in expression data across cells by using a linear model for each cell feature and by plotting the distribution of their marginal R^2 values. As expected we found that the read depth (total features) and the library complexity (% counts top 100 features) were the two most important explanatory variables. We observed that the site origin was also an important explanatory variable (> 20% of variance) suggesting that two sites produced different cell types (**Figures S5E**). Finally, we noted that the variables batch (2,3,4) and genotype explained a low fraction of the total variance. Since we wanted to identify the causes explaining the cell population differences between sites, we did not correct for these effects.

Identification of iPSC subpopulations: clustering analyses

We performed hierarchical clustering on the Euclidean distances between cells, using Ward's criterion to minimize the total variance within each cluster. This yields a dendrogram that groups together cells with similar expression patterns across the chosen genes. Clusters are explicitly defined by applying dynamic tree cut to the dendrogram(Langfelder et al., 2008) (R package dynamicTreeCut). This approach exploits the shape of the branches in the dendrogram to refine the cluster definitions. We identified four, five and six cells subpopulations in SB808, SBAD3 and SB808 and SBAD3 together respectively (**Figure 6A**)

Gene markers between iPSC subpopulations

Markers of specific subpopulations were identified by looking at genes that are consistently differentially expressed in the largest subpopulation compared to the others (cluster 1 vs others clusters). DE analyses were performed with the R package edgeR (3.14.0)(Robinson et al., 2010) that uses negative binomial (NB) distributions to model the read counts for each sample. We estimated the NB dispersion parameter that quantifies the biological variability in expression across cells in the same cluster. Large dispersion estimates (> 0.5) were observed due to technical noise with single cell RNA-seq data (in contrast to bulk data where values of 0.05–0.2 are more typical). We then used the design matrix to fit a NB General Linear Model to the counts for each gene. Finally, the top ten sets of DE genes from each pairwise comparison were considered to be as potential marker genes separating subpopulations (**Table S5**).

GO pathway associated with markers genes between iPSC subpopulations

We performed gene set enrichment analysis with topGO R package (2.24.0) by considering top 50 marker genes identified above. We performed a classical enrichment analysis by testing the over-representation of GO BP terms within the group of differentially expressed genes by Fisher test. We then listed the top 20 significant GO BP terms identified (**Table S6**).

Brain cell class of different iPSC subpopulations

To classify which brain cell type is resembled by the different subpopulations, we used a public database containing transcriptional datasets of purified cortical cells, such as neurons, astrocytes, microglia, endothelial cells, pericytes, and various maturation states of oligodendrocytes(Zhang et al., 2014). Mouse data, instead of human data as used in bulk analysis, has been chosen for its larger number of samples available for each cell type. From Fragments Per Kilobase of transcript per Million (FPKM) mouse gene expression values (http://web.stanford.edu/group/barres_lab/brain_rnaseq.html) we ranked the genes according to their fold change in each cell type. To evaluate the specificity of each population for a specific cell type, we compared the sum of expression values for top 50 cell specific marker genes with those of random sampling of 50 genes (**Figure S5A**).To simplify the heatmap, we display only the top 20 marker genes for each cell type that were expressed in 25% single cell libraries **(Figure 6A**).

Comparison of differentially-expressed genes and pathways between genotypes between iPSC subpopulations

We performed differential expression analyses between the two line within each subpopulation. As cluster 5 included no control iPSC cells, we excluded this cluster from downstream analyses. We performed differential expression analyses with edge R with the same approach as described above to identify marker genes for the subpopulations by considering DE genes based on empiric p-value < 0.001. We then examined the overlap between up- or down-regulated genes among different iPSC subpopulations (**Figure S5D**). Further we calculated the correlation coefficients and their significance values for log fold change between DE of different iPSC subpopulations (**Figure S5D**). We reported DE genes by iPSC subpopulations in **Table S3**.

To ensure that there was no overlap between GO pathways associated with distinct sets of DE genes of each iPSC subpopulation, we compared the p-values of enrichment analyses for the top 10 BP GO terms associated to DE genes in each iPSC subpopulations. GO enrichment analysis was performed as described above using topGO R package.

Proteomics

Materials and Reagents

All materials were of analytical and mass-spectrometry grade. DL-dithiothreitol (DTT), iodoacetamide, ASB-14, Tris base and urea were all purchased from Sigma-Aldrich. UPLC-MS grade acetonitrile (ACN), Formic acid (FA) and water were obtained from Fluka, and sequencing-grade modified porcine trypsin from Promega. All buffers and solutions were prepared using ultra-pure 18 MΩ water (MilliQ) and UPLC solvents using UPLC-MS grade water.

In-solution proteolytic digestion

Lysate and supernatant were stored at -80 °C until use. The frozen cell pellets (about 2 million cells) were thawed, dissolved in lysis buffer containing 100mM Tris HCl pH 7.8, 6M Urea, 2M Thiourea, 2% ASB-14. Cells were then sonicated for 5 min to disrupt cell membranes followed by shaking for 1h in room temperature in order to solubilize proteins. Samples were reduced with the addition of 1.5 µl of DTT (30 mg/mL) for 1h at room temperature and then alkylated with 3 µL of iodoacetamide (36 mg/mL) for 30 min in the dark. To dilute the urea, 155 µL of ultra-pure water was added prior to addition of 10 µL of sequencing-grade trypsin (Promega) (0.1 µg/µL). Trypsin digestion was carried out for 12 hours at 37°C, followed by desalting and concentrated using C18 Isolute columns (Biotage).

Label-free proteomic analyses (2D-LC-MSe)

1 pmol of yeast enolase reference standard (Waters) was added to the each sample and 1 µg of the sample analysed using a 2D-LC-nanoESI-MSe using a nanoAcquity UPLC 2D-LC system and Synapt G2-Si mass spectrometer (Waters, Manchester, UK). Peptides were fractionated in 8 steps (8.7%, 11.8%, 13.6%, 15.3%, 17.1%, 19.3%, 22.5% and 50% of solvent B) under high pH conditions using XBridge BEH C18 Trap Column, 130Å, 5 µm, 180 µm x 50 mm (Waters). Solvent A was 20 mM ammonium formate in water (pH 9) and solvent B was 100% acetonitrile (ACN). Eluted peptides were trapped in the Symmetry C18, 100Å 5 µm, 300 µm x 20 mm (Waters) trap column and subsequently separated under low pH conditions on a nanoAcquity C18 column Peptide BEH C18, 130Å, 1.7 µm, 75 µm x 150 mm (Waters). Analytical chromatography was performed using a 60-min gradient starting at 97% solvent A, ramping to 40% solvent B in 40 min, then to 85% solvent B over 2 min (held for 3 min) and finally decreased to 97% solvent A in 15 min. Mobile phase A contained 95% H_2O, 5% DMSO, 0.1% FA and mobile phase B was 95% ACN, 5% DMSO and 0.1% FA.

Principal Component Analysis

Protein abundances of 1037 proteins, which are expressed across all samples, are corrected for batch effect using ComBat function from sva R package(Leek et al., 2012) (two batches done separately for the two time points). PCA is performed by prcomp R function (scale=T, center=T).

PCA of protein abundances and gene counts at FP+55 time point is performed on the same set of 1037 proteins/genes after correcting for differences between the two types of data using ComBat function.

References

Consortium, GTEx. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science *348*, 648–660.

Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* bts635 (2012). doi:10.1093/bioinformatics/bts635

Duff, K., Eckman, C., Zehr, C., Yu, X., Prada, C.M., Perez-tur, J., Hutton, M., Buee, L., Harigaya, Y., Yager, D., et al. (1996). Increased amyloid-beta42(43) in brains of mice expressing mutant presenilin 1. Nature *383*, 710–713.

Hadfield, J.D. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm *R* Package. J. Stat. Softw. *33*.

Heywood, W.E., Galimberti, D., Bliss, E., Sirka, E., Paterson, R.W., Magdalinou, N.K., Carecchio, M., Reid, E., Heslegrave, A., Fenoglio, C., et al. (2015). Identification of novel CSF biomarkers for neurodegeneration and their validation by a high-throughput multiplexed targeted proteomic assay. Mol. Neurodegener. *10*.

Higgins, J.P.T., and Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. Stat. Med. *21*, 1539–1558.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J., and Altman, D.G. (2003). Measuring inconsistency in meta-analyses. BMJ *327*, 557–560.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostat. Oxf. Engl. *8*, 118–127.

Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinforma. Oxf. Engl. *24*, 719–720.

Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics *28*, 882–883.

Liao, Y., Smyth, G. K. & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma. Oxf. Engl.* **30,** 923–930.

Miller, J.A., Ding, S-L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z., Royall, J., Aiona, K., et al. (2014). Transcriptional landscape of the prenatal human brain. Nature. *508*, 199-206

Muschet, C., Möller, G., Prehn, C., Hrabě de Angelis, M., Adamski, J., and Tokarz, J. (2016). Removing the bottlenecks of cell culture metabolomics: high-throughput normalization procedure, correlation of metabolites to cell number, and impact of the harvesting method. (Submitted).

Risso, D., Ngai, J., Speed, T.P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. Nat. Biotechnol. *32*, 896–902.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Shi, Y., Kirwan, P., Smith, J., Robinson, H.P.C., and Livesey, F.J. (2012a). Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. Nat. Neurosci. *15*, 477–86, S1.

Shi, Y., Kirwan, P., and Livesey, F.J. (2012b). Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. Nat. Protoc. *7*, 1836–46.

Sproul, A.A., Jacob, S., Pre, D., Kim, S.H., Nestor, M.W., Navarro-Sobrino, M., Santa-Maria, I., Zimmer, M., Aubry, S., Steele, J.W., et al. (2014). Characterization and Molecular Profiling of PSEN1 Familial Alzheimer's Disease iPSC-Derived Neural Progenitors. PLoS ONE *9*, e84547.

Trapnell, C. *et al.* Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nat. Biotechnol.* **28,** 511–515 (2010).

van de Leemput, J., Boles, N.C., Kiehl, T.R., Corneo, B., Lederman, P., Menon, V., Lee, C., Martinez, R.A., Levi, B.P., Thompson, C.L., et al. (2014). CORTECON: A Temporal Transcriptome Analysis of In Vitro Human Cerebral Cortex Development from Human Embryonic Stem Cells. Neuron *83*, 51–68.

Wang, J.Z., Du, Z., Payattakool, R., Yu, P.S., and Chen, C.-F. (2007). A new method to measure the semantic similarity of GO terms. Bioinformatics *23*, 1274–1281.

Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y., and Wang, S. (2010). GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinforma. Oxf. Engl. *26*, 976–978.

Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keeffe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N., et al. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J. Neurosci. Off. J. Soc. Neurosci. *34*, 11929–11947.

Zhang, Y., Sloan, S.A., Clarke, L.E., Caneda, C., Plaza, C.A., Blumenthal, P.D., Vogel, H., Steinberg, G.K., Edwards, M.S.B., Li, G., et al. (2016). Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. Neuron *89*, 37–53.