In vitro-differentiated neural cell cultures progress towards donor-identical brain tissue

Brooke E. Hjelm¹,²,³, Bodour Salhia⁴, Ahmet Kurdoğlu¹, Szabolcs Szelinger¹,²,³, Rebecca A. Reiman⁵, Lucia I. Sue²,⁶, Thomas G. Beach²,⁶, Matthew J. Huentelman¹,², David W. Craig¹,²,*

¹Neurogenomics Division, The Translational Genomics Research Institute (TGen), Phoenix, AZ 85004, USA
²Arizona Alzheimer's Consortium, Phoenix, AZ 85014, USA
³Molecular and Cellular Biology Interdisciplinary Graduate Program, College of Liberal Arts and Sciences, Arizona State University, Tempe, AZ 85281, USA
⁴Integrated Cancer Genomics Division, The Translational Genomics Research Institute (TGen), Phoenix, AZ 85004, USA
⁵Collaborative Sequencing Center (CSC), The Translational Genomics Research Institute (TGen), Phoenix, AZ 85004, USA
⁶Brain and Body Donation Program, Civin Laboratory for Neuropathology, Banner Sun Health Research Institute (BSHRI), Sun City, AZ 85351, USA

*Corresponding author: Address: TGen, 445 N. 5th Street, Phoenix, AZ 85004, Email: dcraig@tgen.org, Phone: 602-343-8747, Fax: 602-343-8844
Abstract

Multiple research groups have observed neuropathological phenotypes and molecular symptoms in vitro using induced pluripotent stem cell (iPSC)-derived neural cell cultures (i.e., patient-specific neurons and glia). However, the global differences/similarities that may exist between in vitro neural cells and their tissue-derived counterparts remain largely unknown. In this study, we compared temporal series of iPSC-derived in vitro neural cell cultures to endogenous brain tissue from the same autopsy donor. Specifically, we utilized RNA sequencing (RNA-Seq) to evaluate the transcriptional progression of in vitro-differentiated neural cells (over a timecourse of 0, 35, 70, 105, and 140 days), and compared this to donor-identical temporal lobe tissue. We observed in vitro progression towards the reference brain tissue, and the following three results support this conclusion: (1) there was a significant increasing monotonic correlation between the days of our timecourse and the number of actively transcribed protein-coding genes and long intergenic noncoding RNAs (lincRNAs) \( p < 0.05 \), consistent with the transcriptional complexity of the brain; (2) there was an increase in CpG methylation after neural differentiation that resembled the epigenomic signature of the endogenous tissue; and (3) there was a significant decreasing monotonic correlation between the days of our timecourse and the percent of in vitro to brain tissue differences \( p < 0.05 \) for tissue-specific protein-coding genes and all putative lincRNAs. Taken together, these results are consistent with in vitro neural development and physiological progression occurring predominantly by transcriptional activation of downregulated genes rather than deactivation of upregulated genes.
Introduction

Disease models for human genetic disorders exist in many forms, including transgenic animals (1, 2), primary or immortalized human cell lines (3, 4), or the more recently described induced pluripotent stem cells (iPSCs) (5-8). iPSCs are particularly intriguing tools for modeling human genetic disorders, because tissue-specific and disease-applicable cell types that retain the donor’s complex genetics can be generated in vitro (5-8). However, with any disease model system, there may be initial concerns about the physiological or pathological relevance of the model, and how subsequent drug screening or toxicity trials will correlate with clinical responses (7, 9-15). Potential challenges exist as to which iPSC-derived disease models will be able to produce a pathological phenotype, and how observed in vitro pathologies will correlate with in vivo disease onset, severity, progression and/or drug response (7, 10, 11, 13-15).

Neurodegenerative disorders are commonly misdiagnosed in live human patients; often, a diagnosis can only be confirmed or refuted with the additional observations of a postmortem neuropathological exam (16-20). Autopsy donors that have been subjected to these rigorous diagnostic criteria are especially useful for iPSC generation, because subsequent in vitro disease models can be produced with increased confidence that the donor was a neurological control (true negative; greater specificity) or possessed a specific neurological disease (true positive; greater sensitivity) (16). In addition, this approach enables us to compare iPSC-derived cell cultures to endogenous tissues from the same donor.

In this study, we compared iPSC-derived neural cell cultures to donor-identical brain tissue. This particular donor was a 75-year-old male, defined by both clinical criteria and postmortem neuropathological observations as a neurological control. Data regarding the
establishment of fibroblast cell lines, iPSC generation, and initial neural differentiation tests can be found in our previous characterization paper (16). For this study, we differentiated iPSC-derived neural precursor cells (NPC) over a timecourse of 0, 35, 70, 105, and 140 days (i.e., in 5-week intervals over a period of 20 weeks) and compared this to temporal lobe tissue from the same autopsy donor. The neural differentiation protocol used in these studies was specific to the development of forebrain, cortical neurons (and glia), or what is commonly referred to as the “default” neural differentiation pathway when no additional morphogens are included in culture (21-24). As our endogenous tissue reference for these initial studies, we chose the temporal lobe because this brain region is part of the forebrain/cerebral cortex, and is pathologically relevant to multiple neurological conditions (including several late-onset diseases that are diagnostically aided by neuropathological confirmation) (17-19, 25, 26). The brain tissue used in this study was collected and frozen after a short postmortem interval (3.33hrs), thereby preserving the RNA integrity (17, 27), and allowing us to use RNA sequencing (RNA-Seq) analyses for our *in vitro* to brain tissue comparisons.

RNA-Seq is a set of methods based upon next-generation sequencing technology that allows one to evaluate the transcriptome, effectively permitting single-transcript resolution of the expressed RNA transcripts at a particular snapshot in time, regardless of the transcript’s function or protein-coding potential (28-30). This is a powerful tool because it allows us to study expression levels without any *a priori* hypotheses about which genes or regulatory features may be differentially expressed. In this study, we focused on differential expression (*in vitro* vs. brain tissue) of both well-annotated protein-coding genes, as well as long intergenic noncoding RNAs (lincRNAs), both of which have been shown to exhibit tissue-specificity and are
considered developmentally important (31-35). Unlike other types of noncoding RNAs, lincRNAs do not overlap with well-annotated protein-coding genes allowing both features to be computationally tractable (33, 35, 36).

In addition to transcriptome analyses, we also analyzed a subset of our samples for differences in genome-wide CpG methylation using an array-based platform (37). CpG methylation has been linked to differential gene expression, in both developmental and pathological contexts, and has been extensively studied in human cancer (38-41). In addition, previous studies have shown that CpG methylation can distinguish cell types in a tissue-specific manner (39, 42), and that methylation patterns vary between different regions of the brain (43). Likewise, the specific methylation states of various loci have been shown to exhibit dynamic changes in the brain during development and aging (44, 45).

This study describes the transcriptional and methylation effects of in vitro neural differentiation and prolonged neural cell culture as it relates to the physiological levels exhibited by the endogenous brain tissue. We hope that our donor-identical comparisons will provide a valuable resource for those interested in the physiological or pathological relevance of iPSC-derived neurological disease models and the specific transcriptional progression of protein-coding genes and lincRNAs we observed in vitro.

Results

Samples and Experimental Design

We selected a 75-year-old, male autopsy-donor with no neurological conditions for these initial in vitro to brain tissue comparisons. This donor was cognitively normal, and the postmortem neuropathological exam revealed he had a Braak score (Stage 1) and a CERAD
Neuritic Plaque score (0) consistent with an unaffected neurological control. A schematic diagram of our experimental design, along with samples selected for both RNA-Seq and CpG methylation analyses are shown (Fig. 1A). Briefly, early-passage (p3), iPSC-derived neural precursor cells (NPC) were used for two independent experiments, and were differentiated in vitro for a period of 35, 70, 105, or 140 days. Sample acronyms used for the neural differentiated samples in this study are listed (D35, D70, D105, D140) (Fig. 1A). An average of 33.2 ± 11.7-million quality reads per sample library (mapped to GRCh37) were used for RNA-Seq analysis, and the number of sequencing reads for each RNA-Seq sample and replicate can be found in the online supplemental data (Supplemental Table S1). The iPSC line used in this study had a normal, diploid karyotype with no major chromosomal aberrations as determined by copy number variation (CNV) analysis using whole genome sequencing (WGS) (Supplemental Fig. S2). The number of unique sequencing reads (for both the iPSC and parental fibroblast) used for CNV analysis can also be found in the online supplemental data (Supplemental Table S1).

Differentiated neural samples were only selected for analysis in the case where at least 80% of the cells appeared neuronal by the first differentiation timepoint (D35) (i.e., exhibited neuritic cell-to-cell processes distinguishable by brightfield microscopy as shown in D140 sample) (Fig. 1B). Differentiated neural cell cultures that met this criterion were further evaluated by immunocytochemistry (ICC) for verification of neuronal identity and population architecture (Fig. 1C). The differentiated neural cell cultures contained a mixed population of both immature (TUBB3+/MAP2-) and mature (TUBB3+/MAP2+) neurons at the first differentiation timepoint (D35) (Fig. 1C). As a whole, the in vitro cell populations were mostly
(≥ 85%) neurons (TUBB3+), and about one third (30-35%) of these neurons were mature (MAP2+). Both glutamatergic (VGLUT1+) and GABAergic (GAD67+) neurons were identified (Fig. 1C). VGLUT1 expression was detected in the majority (70-75%) of neurons, regardless if they were mature (MAP2+) or not. However, VGLUT1 expression in immature neurons appeared as strong, punctate staining in the cell body, consistent with a vesicular localization, while mature neurons (MAP2+) expressed VGLUT1 all along the neuritic processes in addition to the cell body. GAD67 expression was only detected in a subset (15-20%) of mature (MAP2+) neurons. ICC analyses were performed across three iPSC clones from the same donor, including the clone used for these RNA-Seq studies, verifying that the neural differentiation efficiency and percent of neuronal subtype markers evaluated were both reproducible. ICC analyses, previous studies, and our currently described RNA-Seq data, demonstrate a small proportion (≤ 5%) of the cell population was astrocytic glia (16).

In Vitro Neural Cells Undergo Progressive Activation of Transcription and CpG Methylation

Studies using gene expression microarrays and RNA-Seq have demonstrated there are an increased number of actively transcribed genes in the brain compared to most other tissues, although there are exceptions (46, 47). This observation of increased transcriptional activation is presumably due to a high level of cell population complexity within the brain (46-48). In addition, functional features like postsynaptic membranes have been predicted to contain hundreds to thousands of proteins, and these numbers have been shown to increase during neuronal differentiation and subsequent equilibrium (49, 50).

Due to the possible added effects of population and functional complexity in the brain or cortex, we detected a high number of actively transcribed genes in our temporal lobe tissue
sample (i.e., BRAIN). Specifically, out of 18,062 consensus CDS (CCDS) protein-coding genes evaluated, 82.9% (14,971) had detectable transcript expression values, or FPKM (i.e., Fragments Per Kilobase of exon per Million fragments mapped) above our cutoff (FPKM > 0.01), and 63.9% (11,538) had a FPKM > 1. Previous studies using RNA-Seq or gene expression arrays have reported a similar percentage of expressed genes (e.g., 71%, 76%, 82%, 86%) within the developing or adult brain (46, 48, 51). Differences in technologies used for transcript abundance estimations, in addition to differences in the total number of sequencing reads, may affect the total number of expressed transcripts detected; as such, these published values may not be directly comparable with our data.

An initial inspection of transcript data showed an increase in the number of actively transcribed, protein-coding genes as the in vitro neural samples progressed through our differentiation timecourse. Specifically and for comparison, 73.1% (13,199), 75.6% (13,661), 78.8% (14,239), 77.5% (14,003), and 82.0% (14,811) of CCDS genes had an FPKM > 0.01, for our NPC, D35, D70, D105, and D140 samples, respectively. Likewise, 50.2% (9,060), 53.5% (9,662), 56.2% (10,146), 56.9% (10,269), and 61.3% (11,066) of CCDS genes had an FPKM > 1 in these samples (listed in order). Statistical analysis showed a significant increasing monotonic correlation between the days of our timecourse and the percent (or number) of actively transcribed genes as determined by a one-tailed non-parametric Kendall’s tau correlation test. This significant correlation was observed for both the cutoffs used to evaluate transcriptional activation (FPKM > 0.01, p < 0.05; FPKM > 1, p < 0.01).

In order to better characterize the transcript diversity, we conducted a density analysis of all transcripts under varying FPKM values. Density plots can be used for meta-analysis and
visualization of the FPKM distributions, and have been previously described by Trapnell et al. in their published methods regarding RNA-Seq analysis and the CummeRbund analysis package (29). For clarification, a pseudo-count of 0.01 was added to all FPKM values so that density plots could be displayed on a log$_{10}$ scale. As such, density plots exhibit a bimodal distribution, whereby the first peak shows the density of inactive, non-expressed or ‘off’ genes, while the second peak shows the density and distribution of actively transcribed, expressed or ‘on’ genes in that sample. Density plots show a decrease in inactive genes (shrinkage in first peak) and an increase in actively transcribed genes (augmentation in second peak) as our differentiation timecourse progressed (Fig. 2A). The temporal lobe also showed a higher density of actively transcribed genes and is plotted alongside the in vitro samples as a reference (Fig. 2A). In summation, the density plots and rank-ordered scatter plots show a systematic increase of area under the second curve, consistent with transcriptional activation of CCDS genes (Fig. 2A,D).

In addition to the progressive increase in the number of actively transcribed CCDS genes, we also observed a progressive increase in the number of lincRNA transcripts expressed, although this trend was less linear than that of protein-coding genes ($R^2$=0.75 vs. $R^2$=0.91) (data not shown). Out of 8,262 putative lincRNAs previously described by Cabili et al. (33), the temporal lobe sample (BRAIN) had 31.7% (2,623) of lincRNA actively expressed at an FPKM > 0.01, and 31.2% (2,581) had an FPKM > 1 in this sample. Comparatively, the in vitro neural samples showed 16.0% (1,318), 18.7% (1,544), 25.3% (2,087), 22.5% (1,855), and 25.4% (2,097) of lincRNAs had an FPKM > 0.01, for our NPC, D35, D70, D105, and D140 samples, respectively. Likewise, 15.5% (1,280), 18.1% (1,496), 24.0% (1,986), 21.5% (1,775) and 24.4% (2,020) of lincRNAs had an FPKM > 1 (listed in order). Statistical analysis showed a significant increasing...
monotonic correlation between the days of our timecourse and the percent (or number) of actively transcribed lincRNAs as determined by a one-tailed non-parametric Kendall’s tau correlation test. This significant correlation was observed for both the cutoffs used to evaluate transcriptional activation (FPKM > 0.01, p < 0.05; FPKM > 1, p < 0.05).

There was a smaller percentage of lincRNAs expressed in both the in vitro neural samples and the endogenous brain tissue than there were protein-coding genes. These observations have been previously reported by others and are likely, in part, due to the high level of tissue-specificity in lincRNAs and a lower level that are constitutively active compared to protein-coding genes (33). On average, 23.2% to 22.5% of lincRNAs were expressed in our neural samples, for FPKM > 0.01 and > 1, respectively. In contrast, 78.3% to 57% of protein-coding genes were expressed (on average) in our neural samples, for FPKM > 0.01 and > 1. Density plots and rank-ordered scatter plots demonstrate this increase in transcriptional activation of lincRNAs, as previously described for CCDS genes (Fig. 2B,E). Interestingly, for both protein-coding genes (18,062) and putative lincRNAs (8,262), we saw an activation rate or slope of ≈2.2% per timepoint, or ≈0.4% per week, as determined by linear equation tests for our in vitro samples (data not shown). We also observed increased variation between our replicates (for both CCDS genes and lincRNAs) as our timecourse progressed, and more variability in progressive transcriptional activation at the middle stages of our timecourse (Supplemental Fig. S3 and S4).

As a final measure of the global differences between our in vitro samples and the endogenous brain tissue, we performed density analysis on methylation sites (485,261) with varying levels of methylation (i.e., beta values) obtained from a genome-wide methylation
array. We observed an increased number of methylated CpG sites in the last neural
differentiation timepoint (D140) compared to the methylated sites at the beginning or our
timecourse (NPC), and this correlated with increased methylation in the tissue (BRAIN) (Fig. 2C).
Beta value bins of < 0.2 (unmethylated) and > 0.8 (methylated) have been previously described
by Illumina publications describing these arrays (37). From the 485,261 methylation sites
examined, the temporal lobe tissue sample (BRAIN) had 30.8% (149,561) of these sites with a
beta value > 0.8. In comparison, the two in vitro samples examined had 17.6% (85,259) and
29.3% (142,285) of sites with a beta value > 0.8, for the NPC and D140 samples respectively.
Conversely, 44.3% (215,021), 33.3% (161,552) and 35.8% (173,711) of these sites had a beta
value < 0.2, for the NPC, D140, and BRAIN samples, respectively. Density plots and rank-order
scatter plots demonstrate this increase in genome-wide DNA methylation (Fig. 2C,F).

The Percent of In Vitro to Tissue Differences Decreases During Differentiation, but Demonstrates
Variation Dependent on Tissue-Specificity (Genes and lincRNAs) and Proximal Relation to CpG
Islands (Methylation)

In order to stratify in vitro to brain tissue comparisons into more specific and
informative datasets, statistical analysis of differential expression was performed between each
in vitro neural sample (NPC, D35, D70, D105, D140) and the endogenous brain tissue (BRAIN),
for all CCDS genes (18,062) and putative lincRNAs (8,262) using Cuffdiff from the Cufflinks
analysis package (30). Specifically, genes and lincRNAs that had both a log₂ fold-change > +2
and a p-value < 0.01 were defined as significantly upregulated in vitro, and those that had both
a log₂ fold-change < -2 and a p-value < 0.01 were defined as significantly downregulated in
vitro. Likewise, differential methylation between a subset of in vitro neural samples (NPC and
D140) and the endogenous brain tissue (BRAIN) were evaluated for all methylation sites (485,261). Specifically, methylation sites that had a beta value difference ≤ -0.2 were defined as hypomethylated in vitro, and those that had a beta value difference ≥ +0.2 were defined as hypermethylated in vitro. The results of these comparisons provide the basis for our remaining figures (Fig. 3, 4, and 5).

We first looked at both the percentage of genes and the percentage of lincRNAs with significant differential expression, assessing all CCDS genes (18,062) and putative lincRNAs (8,262) along with tissue-specific gene lists (Fig. 3A,B). For brain tissue-specific gene lists, a list of “Brain Development Genes” (443) was obtained from the GO processes network using the MetaCore search engine from GeneGo Inc. (34). As a negative control, lists of “Non-Brain Development Genes” were examined, using lists of GO processes from 8 different tissues (i.e., skin, eye, heart, bone, kidney, liver, lung, and pancreas development) (avg. 132), and the average number of differential genes was evaluated (Fig. 3A) (34). For lincRNAs, a list of “Brain Tissue-specific lincRNAs” (183) was obtained from the annotated resource provided by Cabili et al., who separated lincRNAs into tissue-specific classes (33). As a negative control, “Non-Brain Tissue-specific lincRNAs” were examined, using lists of 8 different tissues (i.e., skeletal muscle, heart, placenta, liver, kidney, colon, lung, and testes) (avg. 130) also provided by Cabili et al., and the average number of differential lincRNAs was evaluated (Fig. 3B) (33).

Trends in differential gene expression showed that the percentage of in vitro to brain tissue differences declined as our timecourse progressed; however, only the “Brain Development Genes” (443) showed a significant decreasing monotonic correlation as determined by a one-tailed non-parametric Kendall’s tau correlation test. This significant
correlation was observed for both the total percentage of differences (p < 0.01), as well as the percentage of genes downregulated in vitro (p < 0.05), but was not significant for genes that were upregulated in vitro (Fig. 3A). Likewise, these trends were not significant for all CCDS genes (18,062) or “Non-Brain Development Genes” (avg. 132). Taken together, these data show that for genes specific to brain developmental processes, there was a significant decrease in the percent of in vitro to brain tissue differences over our timecourse and that this observation was primarily due to the activation of genes that were originally downregulated in vitro (Fig. 3A).

Trends in differential lincRNA expression also showed a decrease in the percentage of in vitro to brain tissue differences. These trends showed more variation (or less linearity) than that of protein-coding genes, and the decrease only showed a significant decreasing monotonic correlation for the total number of differences of all putative lincRNAs (8,262) (p < 0.05) (Fig. 3B). Conversely, these trends were not significant for “Brain Tissue-specific” (183) or for “Non-Brain Tissue-specific” (avg. 130) lincRNAs, nor were they significant for trends specific to lincRNAs downregulated or upregulated in vitro (Fig. 3B). These data show that for all putative lincRNAs, we observed a significant decrease in the number of in vitro to brain tissue differences, but this trend was due to the combined effects of both activation of lincRNAs downregulated in vitro, and inactivation of lincRNAs upregulated in vitro (Fig. 3B). Also, there was less recovery (or progressive activation) of brain-specific lincRNAs than there was for protein-coding genes. Venn diagrams (for both CCDS genes and all putative lincRNAs) displaying the number and relative proportion of conserved in vitro to brain tissue differences.
between each adjacent step in the timecourse, as well as the differences conserved across all in vitro to tissue comparisons are shown in the online supplemental data (Supplemental Fig. S5).

Finally, differences in DNA methylation between two of the in vitro neural samples (NPC and D140) and the endogenous brain tissue (BRAIN) were also evaluated. Differential methylation was defined as a site that had a beta value difference of at least 20% (i.e., ≥ 0.2) between the brain tissue and respective in vitro sample. Trends in differential methylation were plotted for all methylation sites (485,261) on the array, as well as those specific to CpG islands (150,176), shores (North=62,827; South=49,166), and shelves (North=24,824; South=22,283). Shores are methylation regions 0-2 kilobases (kb) from CpG islands (CGI), and shelves are regions 2-4kb from CGI (37, 39, 42). Overall, there was a decrease in the total number of differentially methylated sites from the beginning (NPC) to end (D140) of our timecourse compared to the brain (BRAIN). This trend was apparent for both CGI shores and shelves; however, for CGI themselves, there was a slight increase in the percent of in vitro to brain tissue differences (Fig. 3C). The percentage of methylation sites that were hypomethylated in vitro compared to the brain tissue decreased after differentiation (D140) for all methylation regions (islands, shores and shelves), which correlates with our previous observation of a global increase in DNA methylation. Conversely, the percentage of methylation sites that were hypermethylated in vitro compared to brain increased after differentiation (D140) for most of the methylation regions examined. The only exception to this trend was the percent of north-shore (N-Shore) methylation sites hypermethylated in vitro, which decreased after differentiation (D140) (Fig. 3C).
This observation was further examined by looking at CGI associated with a subset of genes that were originally downregulated \textit{in vitro} and were transcriptionally activated during differentiation, or were originally upregulated \textit{in vitro} and were transcriptionally inactivated during differentiation. Although we observed a global increase in CpG methylation, the percent of hypermethylated sites (for both NPC and D140 compared to BRAIN) was lower for genes activated during differentiation, than it was for genes inactivated during differentiation. These trends were most pronounced in the N-shore regions of associated CGI (data not shown).

These observations are consistent with previous studies that have shown dynamic changes in CpG methylation in the developing and adult human brain (44) and during \textit{in vitro} differentiation, as well as previous reports that have identified the shore regions (0-2kb from CGI) as having the most tissue-specific methylation patterns (39, 42).

\textit{In Vitro and Brain Tissue Transcript Abundance of Selected Neurologically Relevant Genes}

We further investigated the physiological and pathological relevance of our \textit{in vitro} iPSC-derived neural cell cultures by focusing on neurologically relevant genes. This includes genes important to neurogenesis and gliogenesis (Fig. 4 and Supplemental Fig. S6), and genes associated with certain neurological disorders and related pathways (Supplement Fig. S6). Genes were selected based on their redundancy in neurological GO processes, Pubmed, OMIM, and/or commercially available arrays containing neural-specific expression markers.

Neuronal expression markers included the following: \textit{MAPT}, which was significantly downregulated \textit{in vitro}; \textit{MAP2}, which was significantly downregulated \textit{in vitro} until the last differentiation timepoint (D140) when it was expressed at similar levels to the brain; \textit{TUBB3}, which was expressed at similar levels to the brain until the last differentiation timepoint (D140).
when it was significantly upregulated \textit{in vitro}; the synaptic vesicle gene SYN1, which was also significantly downregulated \textit{in vitro} (although an increase in expression was observed over time); and \textit{DLG4}, the gene that encodes for the postsynaptic density protein, PSD-95, which was significantly downregulated \textit{in vitro} until the last differentiation timepoint (D140) when it was expressed at similar levels to the brain (Fig. 4A). There was a progressive increase in expression for the three astrocyte genes, \textit{SLC1A3(GLAST)}, \textit{GFAP} and S100B; however, all \textit{in vitro} timepoints remained significantly downregulated compared to the brain tissue, as was expected for the smaller proportion of astrocytic glia in our \textit{in vitro} neural cell cultures (Fig. 4B). Oligodendroglial markers \textit{MBP} and \textit{OLIG2} were both significantly downregulated \textit{in vitro}, and did not increase during our timecourse suggesting these neural cell cultures did not contain terminally differentiated oligodendrocytes (Fig. 4B). Additional genes evaluated for differential expression included two neurotrophic growth factors (\textit{BDNF} and \textit{NGF}), as well as genes involved in neuronal cell adhesion (\textit{CNTNAP2}), transcriptional repression (\textit{MECP2}), and calcium channel signaling (\textit{CACNA1C}) (Fig. 4C). Neuronal subtype-specific markers included genes for glutamatergic (\textit{SLC17A7}), GABAergic (\textit{GAD1}), and serotonergic (\textit{SLC6A4}) neurons, as well as the forebrain developmental markers \textit{PAX6} and \textit{FOXD1}, all of which had some degree of transcriptional activation both \textit{in vitro} and in the endogenous brain tissue (Fig. 4D). Lastly, evaluation of genes associated with excitatory (\textit{OTX1} and \textit{OTX2}) or inhibitory (\textit{NKX2-1} and \textit{GSX2}) neuronal activity demonstrated transcriptional activation occurred predominantly for the excitatory (not inhibitory) markers (Fig. 4E). Relative expression of some of these neurologically relevant genes, as well as a selected list of those associated with Alzheimer’s disease and temporal lobe epilepsy can be found in the online supplemental data (Supplemental Fig. S6).
Characterization of Genes and lincRNAs Ascribed to Early-to-Late Models of Differentiation

The differential expression results between our in vitro neural samples and the endogenous brain tissue were used to create five separate models of neural differentiation, based on the fact that we had five points in our timecourse. Genes and lincRNAs that were originally significantly different between the in vitro NPC sample and the BRAIN sample were divided into groups, based on if they changed to similar expression levels as the endogenous brain tissue (i.e., not significant vs. BRAIN) at early (D35; Model I), early-mid (D70; Model II), mid-late (D105; Model III), or late (D140; Model IV) timepoints, or if all in vitro samples remained significantly different compared to the brain (in vitro ≠ brain; Model V) (Fig. 5A). For genes that were originally upregulated in NPC compared to the brain tissue (1,123), 66% (740) fit into one of these five models, and for genes that were originally downregulated in NPC compared to the brain tissue (2,862), 77% (2,216) fit into one of these five models. Likewise, for lincRNAs upregulated in NPC compared to brain tissue (70), 86% (60) fit into one of these five models, and for lincRNAs downregulated in NPC (148), 64% (95) fit into one of these five models. The number of genes and lincRNAs in each model (for those both upregulated and downregulated in vitro) are shown, along with heatmaps of the relative expression (FPKM log2 vs. BRAIN) within each model (Fig. 5). For clarification of Models I-IV, those listed as upregulated were originally expressed in vitro at higher levels than the brain, then were inactivated to similar levels as the brain at a given step in our timecourse. Likewise, those listed as downregulated were originally expressed in vitro at lower levels than the brain, then were activated to similar levels as the brain during the differentiation timecourse. Heatmaps of differential expression for all genes in each of these models are shown, along with heatmaps of
20 selected neurologically relevant genes per model that were originally downregulated in vitro and demonstrated various rates of progressive activation (Fig. 5B). Likewise, heatmaps of differential expression for all lincRNAs in each of these models is shown (Fig. 5C).

Analysis of GO processes revealed the following sets of enriched processes for our neural differentiation models: genes inactivated during our neural differentiation timecourse (Model I-IV up) included those involved in mitosis, cell cycle control, biogenesis, chromatin assembly, and mitotic cell cycle checkpoints; genes activated during our neural differentiation timecourse (Models I-IV down) included those involved in nervous system development, neurogenesis, cell communication and signaling, and synaptic transmission. The FPKM values for each gene and lincRNA corresponding to each of these models can be found in the online supplemental data (Supplemental Data Set S7). In addition, the significant GO processes associated with each of these models can also be found in the online supplemental data (Supplemental Data S8). Future studies may use these lists as an initial association screen of which lincRNAs may be involved in mitosis, cell cycle control, neurogenesis and synaptic transmission. Likewise, genes and lincRNAs that remained significantly different between all our in vitro samples and the endogenous brain tissue (Model V) may serve as a potential list of markers associated with highly specific features or terminal stages of neural differentiation.

Discussion

Data sets comparing in vitro cell culture models versus tissue-derived counterparts are beneficial for evaluating the physiological and pathological relevance of iPSC-derived cell lines and disease models. Patterson et al. previously reported a comparison of several tissue types (i.e. hepatocytes, fibroblasts, and fetal neural progenitor cells) to their iPSC-derived
counterparts (15). Using microarrays, they identified subsets of genes that were upregulated or downregulated in all in vitro samples compared to the endogenous tissues. We identified a significant number of genes in their data set that demonstrated the same pattern in our in vitro to brain tissue comparisons (Supplemental Fig. S5). Specifically, out of 33 genes they described as consistently upregulated in iPSC-derived cell types, 7 to 9 of these genes were also upregulated in our samples compared to the endogenous brain tissue, from NPC to D140, respectively. All of the timepoints were significant when tested by binomial distribution tests (data not shown). Also, out of 19 genes that were consistently downregulated in iPSC-derived cell types in their study, 10 to 2 of these genes were also downregulated in our in vitro samples, from NPC to D140, respectively. Interestingly, this was only statistically significant at the NPC, D35 and D105 timepoints (data not shown). These data support the results of Patterson et al. and, in addition, suggest that in vitro activation of downregulated genes is more easily accomplished over time in culture than is the deactivation of upregulated genes. These deductions are consistent with our previously mentioned results of progressive and global transcriptional activation. Additional studies will need to be performed in order to determine if these results are specific to neural differentiation, or are likewise supported in prolonged culture of other iPSC-derived cell types.

In this study, we focused on the in vitro to brain tissue differential expression of protein-coding genes and lincRNAs using RNA-Seq. We also evaluated some global differences in DNA CpG methylation using a genome-wide array. Future studies could also look at other functional features that may be predicted to have physiologically relevant differences. This includes, but is not limited to, microRNAs, histone methylation, protein translation and post-translational
modifications, etc. In addition, RNA-Seq data could be further investigated beyond gene and lincRNA expression, for differences in alternative splicing, alternative promoter usage, intron retention, etc.

The variable we tested in this study was time, or time in neural differentiation conditions. Temporal effects are an intriguing variable to study, because they may be relevant to many cell culture systems regardless of the specific protocol. There are many additional variables, however, that are likely to influence the physiological or pathological relevance of in vitro neurological disease models. These variables include what the neural cells were differentiated or grown on (e.g., synthetic or biological scaffolds, adherent matrices, 3-dimensional vs. 2-dimensional cell culture conditions, polymer coated surfaces, feeder layers), what cell culture media and conditions were used (e.g., differentiation protocol, growth factors and their concentrations, ambient oxygen levels, feeder layers), and what additional steps were performed that may affect the cell population complexity (e.g., purification or cell sorting steps, genetic modification, cell seeding density, passage of NPC used, etc.). Future studies may include RNA-Seq analyses on some of these in vitro variables in order to determine their effects on physiological transcription.

Here, we focused on in vitro comparisons to temporal lobe brain tissue from an elderly, male autopsy donor identified as a neurological control. In addition to this specific comparison, it will be interesting to examine the in vitro to brain tissue differences using other brain regions, as well as comparisons with donors of different ages, sexes, and those with different stages of neurological disease progression. Likewise, larger data sets that include multiple donors and/or multiple iPSC clones (potentially generated by various methods or independent research
groups) will be valuable in order to examine the variation that may exist and which aspects of the in vitro neural transcriptome are more fixed or plastic.

Also, additional studies should not only look at comparisons of complex cell populations and tissues, but also investigate more simplified or specific cell types obtained after purification/isolation by flow cytometry or laser capture microdissection (LCM). One particularly intriguing question that has important technical implications is: if our neural cell cultures described herein contain both immature and mature neurons, how much of the observed transcriptional activation (and hence brain-specific progression) is due to neuronal maturation? If this accounts for even a portion of the observed transcriptional progression, one would predict that cell sorting techniques specific to mature neuronal markers could demonstrate temporal progression and an increase in physiological transcription. Future studies may include RNA-Seq evaluations of in vitro neural cell populations and in vivo brain tissue, where both samples are analyzed prior to and following single-cell isolation or cell sorting.

While homogenous neuronal cell cultures offer many technical advantages, evaluation of more complex cell culture models, that may include the additional co-culture of astrocytes, oligodendrocytes, microglia, capillary endothelial cells, etc., may also be beneficial RNA-Seq studies as they allow us to evaluate the transcriptome while different cell types are in communication or contact with one another. Some transcriptional, and possibly even pathological, effects may not be present in monotypic cell culture models.

Overall, our analysis demonstrates the power of RNA-Seq and CpG methylation array data sets in evaluating the temporal effects of in vitro development and maturation of an iPSC-
derived neural cell culture model. *In vitro* and tissue comparisons like these are not only useful resources, but also provide unique opportunities for collaboration amongst experts in genomics, stem cell biology, neuroscience, neuropathology, and bioengineering. We hope that this study provides a valuable resource for those interested in the physiological or pathological relevance of iPSC-derived neural cell culture models, and encourages others to collaborate and consider performing *in vitro* to tissue comparisons when evaluating their model system(s).

**Materials and Methods**

*Autopsy Donor and Brain Tissue Collection*

A 75-year-old male autopsy donor was identified as a neurological control by both clinical criteria and a postmortem pathology exam. This subject was enrolled in the Banner Sun Health Research Institute (BSHRI) Brain and Body Donation Program as a whole-body donor and had previously signed informed consent approved by the BSHRI Institutional Review Board (IRB) (17). The brain of this autopsy donor was removed and specific regions were frozen after a short postmortem interval (PMI) (3.33hrs) and were stored at -80°C for available use in future studies.

*In Vitro Sample Collection and Neural Differentiation*

Early-passage (p3), IPSC-derived NPCs were differentiated into neurons and glia as previously described (16), except that in these studies, NPCs were seeded at a lower cell density (2x10^3/cm²) so that a longer differentiation timecourse could be examined. NPCs were generated and maintained using the NeuroCult NS-A Proliferation Kit supplemented with bFGF (10ng/ml), rhEGF (20ng/ml), Heparin (2μg/ml) (STEMCELL Technologies), 100 units/ml penicillin, 100μg/ml streptomycin (Sigma Aldrich), and 5μg/ml Plasmocin prophylactic (Invivogen). NPCs
were expanded and passaged as monolayer cultures in wells pre-coated with BD Matrigel (BD Biosciences). NPCs were differentiated into forebrain, cortical neurons (and glia) on adherent Matrigel matrix using the Neurocult NS-A Differentiation Kit (STEMCELL Technologies) supplemented with 100 units/ml penicillin, 100μg/ml streptomycin (Sigma Aldrich), and 5μg/ml Plasmocin prophylactic (Invivogen). Neural cell cultures were allowed to differentiate for 35, 70, 105, or 140 days, at which time the RNA was extracted and frozen at -80°C for RNA-Seq studies. A brief review of the methods used for iPSC generation and neural differentiation, described in detail in our previous characterization paper (16), can also be found in the online Supplemental methods.

*Copy Number Variation (CNV) Analysis*

CNV analysis between the iPSC line (2-13) and parental fibroblast (F02AA1) was performed using whole genome sequencing (WGS) data. The methods for this type of analysis have been recently described by our group for the identification of CNVs in cancer using WGS from tumor-normal pairs (52). Additional details can be found in the online Supplemental methods.

*Immunocytochemistry (ICC) Analysis*

Differentiated neural cell cultures were grown on Matrigel-coated coverslips for 35 days (D35), at which point they were fixed in a solution of 4% paraformaldehyde (Thermo Scientific) in PBS (Invitrogen) for 15min at RT, and were rinsed (and stored at 4°C) in PBS for later use. Coverslips were permeabilized with a PBS solution of 0.1% Triton X-100 (Sigma) for 10min, then were blocked in a solution of 2.5% BSA (Sigma) in PBS-T (PBS with 0.01% Tween 20 (VWR)) for 30min at room temperature (RT). Primary antibodies were all incubated overnight on a
rotating shaker at 4°C. Primary antibodies included the following: TUBB3 (1:2000; Abcam ab18207), MAP2 (1:1000; EnCor Biotech. CPCA-MAP2), VGLUT1 (1:1000; Synaptic Systems 135-303), and GAD67 (1:300; Abcam ab26116). Coverslips were then washed 3X for 20min each in 2X PBS-T (PBS with 0.02% Tween 20), followed by another 30min blocking step. Alexa Fluor-conjugated secondary antibodies (1:2000; Alexa Fluor 488 or 647 (Invitrogen)) were then incubated on the coverslips for 1hr at RT. The coverslips were washed 3X again, and were then counterstained with DAPI (1μg/ml) for 10min. Coverslips were mounted onto slides with SlowFade Gold antifade reagent (Invitrogen) and were sealed with clear nail polish. All confocal microscopy images were obtained using an Olympus Fluoview FV1000 confocal microscope and the Olympus FV10-ASW 1.7 imaging analysis software. All confocal micrographs displayed are 3-D reconstructions from 10-15 1μm optical sections through the Z-axis.

RNA-Seq Library Preparation and Next-Generation Sequencing (NGS)

RNA was extracted from both the in vitro neural samples and the endogenous brain tissue using TRIzol (Ambien) and the PureLink RNA Mini Kit along with an on-column DNase treatment (Life Technologies) according to the manufacturer’s instructions. 100ng total RNA was linearly amplified and converted to double-stranded cDNA using the Ovation RNA-Seq System (NuGEN) as per the manufacturer’s instructions (53).

Between 1.8μg and 3.0μg cDNA per sample was used for next-generation sequencing (NGS) library preparation, similar to our previously published methods (except that samples were not bar-coded for this study) (54). Libraries were prepared using NEBNext DNA Library Prep Reagent Sets for Illumina, and were quantified and qualified using the Agilent High
Sensitivity DNA Kit. RNA-Seq libraries were sequenced with one sample per lane using the Illumina HiSeq 2000 (55).

**Sequence Alignment and RNA-Seq Analysis**

Bcl to fastq conversion was performed using Illumina software (Illumina). Fastq files were aligned to build 37 of the human genome provided by the Genome Reference Consortium (GRCh37). Transcript alignment was performed using TopHat (version 1.3.2) as previously described (28).

Analysis of differential expression and transcript abundance was performed using Cuffdiff from the Cufflinks analysis package (version 1.3.0) (29, 30). All replicates from independent experiments or tissue pieces were combined analytically during Cuffdiff analysis.

**CpG Methylation Array and Analysis**

1µg genomic DNA per sample was used for the Illumina Infinium HumanMethylation450 Bead Chip, and the chip was prepared and ran according to the manufacturer’s instructions (Illumina) (37). Differential methylation was defined as a site that had a beta value difference of at least 20% (i.e., \( \geq 0.2 \)). Analysis of differential methylation was performed for all methylation sites, as well as those specific to CpG Islands, shore or shelves.

**Data Access**

Upon acceptance, sequencing data will be deposited in the NIH database of Genotypes and Phenotypes (dbGaP) and appropriate accession numbers will be provided.

**Supplementary Materials and Methods**

Supplemental materials and methods sections are available at HMG online.
Acknowledgments

The authors would like to thank Jason Corneveaux, Alexis Christoforides, Jessica Aldrich and Dr. Christopher Kingsley for their guidance regarding statistical tools and analysis, and Dr. Kendall Van Keuren-Jensen for her advice about experiments and data presentation. In addition, we are always greatly appreciative to the whole body donors and their families, without whom this work would not be possible.

Conflict of Interest statement.

None declared.

Funding

This work was supported by TGen’s Salt River Project (SRP) Graduate Student Support Program; The National Institute of Neurological Disorders and Stroke [SU24NS051872-05, R01NS059873, U24NS072026 (National Brain and Tissue Resource for Parkinson’s Disease and Related Disorders)]; The Arizona Alzheimer’s Consortium; The National Institute on Aging [P30AG19610 (Arizona Alzheimer’s Disease Core Center)]; The Stardust Foundation; The Marley Foundation; The Arizona Department of Health Services [Contract 211002 (Arizona Alzheimer’s Research Center)]; The Arizona Biomedical Research Commission [Contracts 4001, 0011, and 05-901 (Arizona Parkinson’s Disease Consortium)]; and The Michael J. Fox Foundation for Parkinson’s Research.
References


Legends to Figures

Figure 1. In Vitro and Tissue-Derived Neural Samples. (A) Schematic diagram of experimental design, comparing in vitro iPSC-derived neural cell populations to donor-identical brain tissue. Samples used for RNA-Seq and CpG methylation analyses are displayed, along with the corresponding sample acronyms that are used in subsequent figures (i.e., NPC, D35-D140, BRAIN). (B) Brightfield images of neural cell cultures at the beginning (NPC) and end (D140) of the experimental timecourse. NPC images show clusters of undifferentiated cell populations, while D140 images show differentiated neurons with abundant neuritic processes. Photographic image of the frozen, temporal lobe brain tissue sample is also shown (in dashed circle). All tissue pieces used for RNA-Seq and CpG methylation analyses were approximately the same size and had the same proportion of white matter (wm) and gray matter (gm). (C) Immunocytochemistry analysis of in vitro neural cell cultures at the first differentiation timepoint (D35). The neuronal marker TUBB3, glutamatergic marker VGLUT1, and GABAergic marker GAD67 (all green; Alexa Fluor 488) were counterstained against the mature neuronal marker MAP2 (red; Alexa Fluor 647). DAPI (blue) is shown in all images as a nuclear counterstain. All scale bars = 100 μm. All images (confocal micrographs) and are 3-D reconstructions from 10-15 1μm optical sections through the Z-axis. Region of interest (white box) in merged image (left most) is shown at higher magnification (2nd from left most). TUBB3/MAP2 image sets demonstrate regional heterogeneity of areas with both immature (TUBB3+/MAP2-) and mature (TUBB3+/MAP2+) neurons (1st row), and areas dominated by mature (TUBB3+/MAP2+) neurons (2nd row). VGLUT1 staining demonstrated vesicular-like localization in immature neurons and localization to both the cell body and neuritic processes in
mature neurons (3rd row). GAD67 was only detected in mature neurons (4th row). See results section for population percentages.

**Figure 2. Global Transcription and CpG Methylation.** Density plots of RNA expression (i.e., FPKM(log₁₀)) for (A) CCDS genes (N=18,062) and (B) lincRNAs (N=8,262). First plot (top left) displays the overlay of all samples. Subsequent plots display the overlay of each individual in vitro sample along with the endogenous brain tissue. (C) Density plots of CpG methylation (i.e., beta values) for all sites on the Illumina Infinium HumanMethylation450 Bead Chip (N=485,261). First plot (left) displays overlay of the three samples analyzed for CpG methylation differences (NPC, D140, BRAIN). Subsequent plots display overlay of the two in vitro samples along with the brain tissue. Rank-order scatter plots for (D) CCDS genes, and (E) lincRNAs. First plot (top left) displays the overlay of all samples’ RNA expression (y-axis: FPKM(log₁₀)), along with the Rank-order of each sample (x-axis: rank-order of FPKM, smallest to largest). Subsequent plots display the overlay of each individual in vitro sample along with the endogenous brain tissue, while using the BRAIN rank-order as the x-axis. (F) Rank-order scatter plots of 50,000 randomly selected CpG methylation sites. First plot (left) displays the overlay of the three analyzed samples’ methylation (y-axis: beta values), along with the Rank-order of each sample (x-axis: rank-order of beta values, smallest to largest). Subsequent plots display the overlay of the two in vitro samples along with the brain tissue, while using the BRAIN rank-order as the x-axis.

**Figure 3. Tissue-Specific Expression, Trends in Gene and lincRNA Expression, and CpG Methylation.** Percentage of (A) genes and (B) lincRNAs that were differentially expressed between each in vitro sample and the endogenous brain tissue. Significant differential
expression was defined as a comparison that had a log₂ fold change > 2, as well as a p-value < 0.01. Trends are show for all genes/lincRNAs, as well as tissue specific lists (GeneGO and Cabili et al.). P-values shown are from a one-tailed, non-parametric Kendall’s tau correlation test. (C) Percentage of CpG methylation sites differentially methylated between the two in vitro samples analyzed (NPC and D140) and the brain tissue. Differential methylation was defined as a site that had a beta value difference of at least 20% (i.e., ≥ 0.2). Trends are show for all methylation sites, as well as CpG islands and flanking regions.

**Figure 4. Gene Expression of 24 Neurologically Relevant Genes.** Bar graphs displaying the FPKM value (log₁₀ scale) of various neurological genes for each in vitro sample along with the endogenous brain tissue. Significance (*) was defined as a comparison that had a log₂ fold change > 2, as well as a p-value < 0.01. 24 genes associated with (A) neurons and synapses, (B) astrocytic and oligodendrocytic glia, (C) neurotrophic factors, adhesion, transcriptional repression, and calcium signaling, (D) neuronal subtypes (glutamatergic, GABAergic, serotonergic) and forebrain development, and (E) excitatory or inhibitory regional specificity are displayed. Gene lists were selected based on their redundancy in GeneGO, Pubmed, OMIM, and/or commercially available arrays containing neural-specific expression markers.

**Figure 5. Models of Neural Differentiation.** Genes that were significantly different (log₂ fold change > 2 and p-value < 0.01) between the neural precursor (NPC) and brain tissue (BRAIN) samples were divided into five models of differentiation, based on if or when in the timecourse the gene became not significant (compared to the endogenous brain tissue). (A) Table of the five models (I-V), corresponding samples in each model, and the number of genes significantly upregulated or downregulated in vitro compared to the brain tissue in each model. (B)
Heatmaps of all genes in each model, and heatmaps of 20 selected neurologically relevant genes per model that were downregulated in the NPC sample compared to the BRAIN and showed progressive activation at a given timepoint. Increasing models (I-IV) demonstrate the order in which these selected genes were transcriptionally activated/inactivated in vitro (i.e., early-to-late). Model V demonstrates selected genes that remained significantly different between all the in vitro neural samples compared to the endogenous brain tissue (i.e., in vitro ≠ brain). Genes were selected based on their redundancy in GO processes related to neural development, synaptogenesis, etc. (C) Heatmaps of all lincRNAs in each model. FPKM values for all genes and lincRNAs in these models, along with the associated GO processes for each model can be found in the online supplemental data (Supplemental Data Sets S7, S8).

**Abbreviations**

*List of Gene Names and Symbols (HUGO Nomenclature)*

* TUBB3 (tubulin, beta 3)
* MAP2 (microtubule-associated protein 2)
* MAPT (microtubule-associated protein tau)
* SYN1 (synapsin 1)
* DLG4 (discs, large homolog 4)
* SLC1A3 (solute carrier family 1, member 3)
* GFAP (glial fibrillary acidic protein)
* S100B (S100 calcium binding protein B)
* MBP (myelin basic protein)
* OLIG2 (oligodendrocyte lineage transcription factor 2)
CNTNAP2 (contactin associated protein-like 2)

MECP2 (methyl CpG binding protein 2)

BDNF (brain-derived neurotrophic factor)

NGF (nerve growth factor)

CACNA1C (calcium channel, voltage-dependent, L type, alpha 1C subunit)

SLC17A7 (solute carrier family 17, member 7)

GAD1 (glutamate decarboxylase 1)

SLC6A4 (solute carrier family 6, member 4)

PAX6 (paired box 6)

FOXG1 (forkhead box G1)

OTX1 (orthodenticle homeobox 1)

OTX2 (orthodenticle homeobox 2)

NKX2-1 (NK2 homeobox 1)

GSX2 (GS homeobox 2)