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

Brooke E. Hjelm1,4,5, Bodour Salhia2, Ahmet Kurdoglu1, Szabolcs Szeling1,4,5, Rebecca A. Reiman3, Lucia I. Sue4,6, Thomas G. Beach4,6, Matthew J. Huentelman1,4 and David W. Craig1,4,*

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

Received November 5, 2012; Revised March 21, 2013; Accepted May 1, 2013

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 with donor-identical temporal lobe tissue. We observed in vitro progression towards the reference brain tissue, and the following three results support this conclusion: (i) 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 non-coding RNAs (lincRNAs) ($P < 0.05$), consistent with the transcriptional complexity of the brain; (ii) there was an increase in CpG methylation after neural differentiation that resembled the epigenomic signature of the endogenous tissue; and (iii) 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 neural 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 with 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 (PMI) (3.33 h), 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 (NGS) 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 non-coding RNAs (lincRNAs), both of which have been shown to exhibit tissue-specificity and are considered developmentally important (31–35). Unlike other types of non-coding 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 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 Supplementary Material, 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) (Supplementary Material, 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 Supplementary Material, 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, VLGUT1 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 with 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 post-synaptic 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% (9060), 53.5% (9662), 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. (29) in their published methods regarding RNA-Seq analysis and the CummeRbund analysis package. 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 and 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 ($R^2 \approx 0.75$ vs. $R^2 \approx 0.91$) (data not shown). Out of 8262 putative lincRNAs previously described by Cabili et al. (33), the temporal lobe sample (BRAIN) had 31.7% (2623) of lincRNA actively expressed at an FPKM > 0.01, and 31.2% (2581) had an FPKM > 1 in this sample. Comparatively, the in vitro neural samples showed 16.0% (1318), 18.7% (1544), 25.3% (2087), 22.5% (1855) and 25.4% (2097) of lincRNAs had an FPKM > 0.01, for our NPC, D35, D70, D105 and D140 samples, respectively. Likewise, 15.5% (1280), 18.1% (1496), 24.0% (1986), 21.5% (1775) and 24.4% (2020) 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 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 with protein-coding genes (33). On average, 23.2–22.5% of lincRNAs were expressed in our neural samples, for FPKM > 0.01 and > 1, respectively. In contrast, 78.3–57% 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 and E). Interestingly, for both protein-coding genes (18 062) and putative lincRNAs (8262), we saw an activation rate or slope of $\approx 2.2\%$ per timepoint, or $\approx 0.4\%$ 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 (Supplementary Material, Figs 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 with the methylated sites at the beginning of 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 and 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 data sets, statistical analysis of differential expression was performed between each in vitro neural sample (NPC, D35, D70, D105 and D140) and the endogenous brain tissue (BRAIN), for all CCDS genes (18 062) and putative lincRNAs (8262) using Cuffdiff from the Cufflinks analysis package (30). Specifically, genes and lincRNAs that had both a log_{2} fold-change $> +2$ and a P-value $< 0.01$ were defined as significantly upregulated in vitro, and those that had both a log_{2} 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 $\leq -0.2$ were defined as hypomethylated in vitro, and those that had a beta-value difference $\geq +0.2$ were defined as hypermethylated in vitro.
The results of these comparisons provide the basis for our remaining figures (Figs 3–5).

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 and 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 eight 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. (33), who separated lincRNAs into tissue-specific classes. As a negative control, ‘Non-Brain Tissue-specific lincRNAs’ were examined, using lists of eight different tissues (i.e. skeletal muscle, heart, placenta, liver, kidney, colon, lung and testes) (avg. 130) also provided by Cabili et al. (33), and the average number of differential lincRNAs was evaluated (Fig. 3B). 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 \( (P = 0.05) \), but was not significant for genes that were upregulated \( (Fig. 3A) \). 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 \( \text{in vitro} \) to brain-tissue differences over our timecourse and that this observation was primarily due to the activation of genes that were originally downregulated \( \text{in vitro} \) (Fig. 3A).

Trends in differential lincRNA expression also showed a decrease in the percentage of \( \text{in vitro} \) 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 (8262) ($P < 0.05$) (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 Supplementary Material, 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

**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 $< 0.01$. Twenty-four 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 with 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 with 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 with 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 with the endogenous brain tissue (i.e. in vitro $\neq$ 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 Supplementary Material, Data Sets S7 and S8.
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 2–4 kb 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 with 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 hypermethylated in vitro compared with 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 with brain tissue increased after differentiation (D140) for most of the methylation regions examined. The only exception to this trend was the percent of northshore (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 in vitro and were transcriptionally activated during differentiation, or were originally upregulated 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 with 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 in vitro differentiation, as well as previous reports that have identified the shore regions (0–2 kb from CGI) as having the most tissue-specific methylation patterns (39,42).

**In vitro** and brain-tissue transcript abundance of selected neurologically relevant genes

We further investigated the physiological and pathological relevance of our in vitro iPSC-derived neural cell cultures by focusing on neurologically relevant genes. This includes genes important to neurogenesis and gliogenesis (Fig. 4 and Supplementary Material, Fig. S6), and genes associated with certain neurological disorders and related pathways (Supplementary Material, 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: MAPT, which was significantly downregulated in vitro; MAP2, which was significantly downregulated in vitro until the last differentiation timepoint (D140) when it was expressed at similar levels to the brain; TUBB3, which was expressed at similar levels to the brain until the last differentiation timepoint (D140) when it was significantly upregulated in vitro; the synaptic vesicle gene SYN1, which was also significantly downregulated in vitro (although an increase in expression was observed over time); and DLG4, the gene that encodes for the post-synaptic density protein, PSD-95, which was significantly downregulated 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, SLC1A3/GLAST, GFAP and S100B; however, all in vitro timepoints remained significantly downregulated compared with the brain tissue, as was expected for the smaller proportion of astrocytic glia in our in vitro neural cell cultures (Fig. 4B). Oligodendroglial markers MBP and OLIG2 were both significantly downregulated 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 (BDNF and NGF), as well as genes involved in neuronal cell adhesion (CNTNAP2), transcriptional repression (MECP2) and calcium channel signaling (CACNA1C) (Fig. 4C). Neuronal subtype-specific markers included genes for glutamatergic (SLC17A7), GABAergic (GAD1) and serotonergic (SLC6A4) neurons, as well as the forebrain developmental markers PAX6 and FOXG1, all of which had some degree of transcriptional activation both in vitro and in the endogenous brain tissue (Fig. 4D). Lastly, evaluation of genes associated with excitatory (OTX1 and OTX2) or inhibitory (NKX2-1 and 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 Supplementary Material, 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 with the brain (in vitro ≠ brain; Model V) (Fig. 5A). For genes that were originally upregulated in NPC compared with the brain tissue (1123), 66% (740) fit into one of these five models, and for genes that were originally downregulated in NPC compared with the brain tissue (2862), 77% (2216) fit into one of these five models. Likewise, for lincRNAs upregulated in NPC compared with 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)
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 (Models 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 Supplementary Material, Data Set S7. In addition, the significant GO processes associated with each of these models can also be found in the Supplementary Material, Data Set 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. (15) previously reported a comparison of several tissue types (i.e. hepatocytes, fibroblasts and fetal neural progenitor cells) to their iPSC-derived counterparts. Using microarrays, they identified subsets of genes that were upregulated or downregulated in all in vitro samples compared with 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 (Supplementary Material, Fig. S5). Specifically, out of 33 genes they described as consistently upregulated in iPSC-derived cell types, 7–9 of these genes were also upregulated in our samples compared with the endogenous brain tissue, 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. two-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. 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 among 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 PMI (3.33 h) 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 (2 × 10^6/cm²) so that a longer differentiation time course could be examined. NPCs were generated and maintained using the NeuroCult NS-A Proliferation Kit supplemented with bFGF (10 ng/ml), rhEGF (20 ng/ml), Heparin (2 μg/ml) (STEMCELL Technologies), 100 units/ml penicillin, 100 μg/ml streptomycin (Sigma Aldrich) and 5 μg/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, 100 μg/ml streptomycin (Sigma Aldrich) and 5 μg/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 Supplementary Material, Methods.

CNV analysis

CNV analysis between the iPSC line (2–13) and parental fibroblast (F02AA1) was performed using 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 Supplementary Material, Methods.

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 15 min at room temperature (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 10 min, then were blocked in a solution of 2.5% BSA (Sigma) in PBS-T (PBS with 0.01% Tween 20 (VWR)) for 30 min at 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. CPA-MAP2), VGLUT1 (1:1000; Synaptic Systems 135–303) and GAD67 (1:300; Abcam ab26116). Coverslips were then washed 3 × 20 min each in 2X PBS-T (PBS with 0.02% Tween 20), followed by another 30 min blocking step. Alexa Fluor-conjugated secondary antibodies [1:2000; Alexa Fluor 488 or 647 (Invitrogen)] were then incubated on the coverslips for 1 h at RT. The coverslips were washed 3 × again, and were then counterstained with DAPI (1 μg/ml) for 10 min. 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 to 15 1 μm optical sections through the Z-axis.

RNA-Seq library preparation and NGS

RNA was extracted from both the in vitro neural samples and the endogenous brain tissue using TRIzol (Ambion) and the PureLink RNA Mini Kit along with an on-column DNase treatment (Life Technologies) according to the manufacturer’s instructions. One hundred nanograms of 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 1.8 and 3.0 μg cDNA per sample was used for 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
One microgram 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. ≥ 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 MATERIAL
Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS
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 NIH (NINDS) [5U24NS051872-05, R01NS059873, U24NS072026 (National Brain and Tissue Resource for Parkinson’s Disease and Related Disorders)]; The Arizona Alzheimer’s Consortium; The NIH (NIA) [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


