Spatiotemporal Proteomic Profiling of Human Cerebral Development*

Ugljesa Djuric‡, Deivid C. Rodrigues§, Ihor Batruch¶, James Ellis§‖, Patrick Shannon‖**, and Phedias Diamandis‡**‡‡

Mass spectrometry (MS) analysis of human post-mortem central nervous system (CNS) tissue and induced pluripotent stem cell (iPSC)-based directed differentiations offer complementary avenues to define protein signatures of neurodevelopment. Methodological improvements of formalin-fixed, paraffin-embedded (FFPE) protein isolation now enable widespread proteomic analysis of well-annotated archival tissue samples in the context of development and disease. Here, we utilize a shotgun label-free quantification (LFQ) MS method to profile magnetically enriched human cortical neurons and neural progenitor cells (NPCs) derived from iPSCs. We use these signatures to help define spatiotemporal protein dynamics of developing human FFPE cerebral regions. We show that the use of high resolution Q Exactive mass spectrometers now allow simultaneous quantification of >2700 proteins in a single LFQ experiment and provide sufficient coverage to define novel biomarkers and signatures of NPC maintenance and differentiation. Importantly, we show that this abbreviated strategy allows efficient recovery of novel cytoplasmic, membrane-specific and synaptic proteins that are shared between both in vivo and in vitro neuronal differentiation. This study highlights the discovery potential of non-comprehensive high-throughput proteomic profiling of unfractionated clinically well-annotated FFPE human tissue from a diverse array of development and diseased states. Molecular & Cellular Proteomics 16: 10.1074/mcp.M116.066274, 1548–1562, 2017.

Interrogation of translational outputs of biological tissues unifies some of the basic tenets of molecular biology. Proteomic signatures of cellular identity shed light on cell type-specific functions and can be used to elucidate molecular underpinnings of pathological processes. MS-based proteomic workflows have made great strides in achieving near genome-wide coverage and outlining draft atlases of the human proteome (1, 2). As a result, proteomic profiling is beginning to complement genomic and transcriptomic methodologies at identifying diagnostic and predictive biomarkers of pathophysiological (3) or cellular differentiation (4, 5) states. Although these breakthroughs have yielded robust MS-based quantification of tissue-specific proteomes, assessment of human tissues across dynamic developmental milestones and organ-specific subcompartments is largely lacking. To this end, well-annotated archival formalin-fixed, paraffin-embedded (FFPE)1 tissues offer a rich source of human material across the developmental spectrum. Crosslinked protein components of FFPE tissues have provided challenges for MS analysis in the past (Reviewed by Tanca et al. 2011) but recent technological advancements have yielded reliable recovery and quantification of >8000 proteins from FFPE B-cell lymphoma and breast cancer samples (3, 7). These studies have demonstrated that measurements of abundance changes of a large number of proteins in FFPE tissues is possible with extensive peptide fractionations prior to MS analysis. However, these laborious fractionation approaches drastically increase MS run times required to analyze a single biological specimen and are thus more limited in their high throughput potential. For example, leading proteomic consortiums cite needing 10 months of machine time to run just over 100 samples (8). As the authors note, the need for such high

1 The abbreviations used are: FFPE, formalin-fixed, paraffin-embedded; LFQ, label-free quantification; iPSC, induced pluripotent stem cell; NPC, neural progenitor cell; GW, gestational week; VZ, ventricular zone; IZ, intermediate zone; SP, subplate; Cx, Cortex.
financial and infrastructure requirements severely limits the widespread application of MS-based proteomics as a research and clinical tool. Recent introduction of high-resolution mass spectrometers such as the QExactive Plus now help mitigate some of the risks of eliminating fractionation steps, such as an inability to detect changes in low abundance proteins, and provide new opportunity for rapid proteomic analysis of unfractonated samples. Furthermore, by reducing the high tissue input requirements of fractionation protocols, proteomics can be more routinely applied toward more anatomically and biological refined hypotheses (9). To assess and highlight the utility of an abbreviated FFPE tissue-based proteomics in detecting neurodevelopment-related proteome changes we set out to profile macroscopically defined cellular layers of the developing human cerebrum.

Human central nervous system (CNS) development offers an especially attractive model to investigate proteomic variations as it progresses through discrete neural milestones with well-established temporal and spatial coordinates (10). In the brain, major proliferative neural precursor cell (NPC) compartments are established in the ventricular zone (VZ) following neural tube closure in the fourth gestational week (4 GW) (11). Neurogenesis and outward migration of neuroblasts from the VZ toward the cortex, through the intermediate zone (IZ) and subplate (SP), occurs between 9 and 20 GWs (12–14). During the second half of gestation (20° GW), the VZ shifts to the predominant generation and outbound redistribution of the full complement of CNS glial cells (gliogenesis) and finally recedes just prior to birth (10). During this late developmental stage, the migrated neurons within the cortex undergo maturation through axonal and dendritic outgrowths and synaptic connections (synaptogenesis) (15).

Recently, MS-defined proteomic signatures of isolated neural cell types were described in the mouse model system (16) whereas proteomes in the human have mostly focused on static developmental timepoints and homogenized organs (1, 2, 17). Given the compartmentalization of function described above, omission of discrete anatomical structures in profiling efforts limits application to understanding neurodevelopment and complex disease states. Thus, advanced multi-group consortiums such as BrainSpan (www.brainspan.org) have compiled developmental transcriptomic landscapes of anatomically defined macro-dissected brain compartments (18, 19). Although these compartments contain heterogeneous cell populations, local and dynamic transcript changes can offer insight into development and disease. Recent proteogenomic studies however, that superimpose proteomic and genomic data, show that protein abundance is not directly related to DNA- or RNA measurements ($r = 0.23–0.45$) (8, 20, 21). As proteins represent a more functional readout of biological activity, overlaying equivalent proteomic information through CNS development represents a gap in the field. Because obtaining well-annotated anatomical structures of fresh human brains for cellular isolation from different stages of development is much more difficult, neuronal differentiations in human-derived cell culture offer a complementary strategy for defining relevant cell type-specific proteome changes. Thus, to profile cerebral neurodevelopment-related protein dynamics we compared protein abundance changes between NPC- and neuron-rich cellular layers of the FFPE cerebrum to those found when neural precursors are induced to differentiate under defined culture conditions. Our parallel screens identify compartment- and time-specific protein abundance changes. These changes include both well-established, such as NES and SATB2, and novel, previously uncharacterized markers (Filamin C, FLNC, and Cellular retinoic acid binding protein 1, CRABP1) of neurodevelopment.

**EXPERIMENTAL PROCEDURES**

**iPSC Culture Cortical Neuronal Differentiation and Enrichment—** iPSC line #37 was previously described (22) and cultured in 5% CO₂ on BD hESC-qualified matrigel in mTeSR medium (Stem Cell Technologies, Vancouver, Canada). Cultures were passaged manually or with ReLeSR (Stem Cell Technologies) every 6–7 days. For neuronal induction, Embryoid Bodies (EBs) were made as described elsewhere(23). Briefly, iPSCs were lifted using 2 mg/ml collagenase type IV, and cultured as cellular aggregates in low-attachment dishes in N2 media containing laminin (1 μg/ml) with 10 μM SB431542, 2 μM DMSO and 1× penicillin-streptomycin. Media was changed daily. After 7 days, EBs were plated on poly-L-ornithine + laminin coated dishes and grown in N2 media + laminin (1 μg/ml). After 7 days, neural rosettes were manually picked and transferred to another polylaminin-coated well. Following an additional 7 days, neural rosettes were picked a second time, digested with acutase and plated on polylaminin-coated wells. At this point, the resulting neural precursor cells were grown as a monolayer and split every 5–7 days in NPC media (DMEM/F12, N2, B27, 1× nonessential amino acid (NEAA), 2 μg/ml Heparin, 1 μg/ml laminin). To generate neurons, NPCs were plated on polylaminin-coated plates at a density of $10^6$ cells per 10 cm dish and cultured for 3 weeks in neural differentiation medium (Neurobasal, N2, B27, 1 μg/ml laminin, 1× penicillin-streptomycin, 10 ng/ml BDNF, 10 ng/ml GDNF, 200 μM ascorbic acid, and 10 μM cAMP). 3-week-old heterogeneous neuronal cultures were enriched for neurons by negative selection strategy using antibodies against surface markers CD44 and CD184 (recognizing NPCs, glial progenitors and astrocytes) (24) using magnetic activated cell sorting (MACS® - Miltenyi Biotec). All steps below were performed on ice, using ice-cold PBS plus 0.5%BSA (PBS + BSA) and centrifugation steps at 4 °C, unless otherwise specified. First, −2 × $10^7$ differentiated cells from heterogeneous cultures were treated with accutase for 10 min, washed twice in PBS + BSA and centrifuged for 10 min, 300 × g. After a second wash, cells were resuspended in PBS + BSA and strained using 70 μM cell strainers (BD Biosciences). Cell number was determined after centrifugation for 5 min at 300g, and cells resuspended in 40 μl of PBS + BSA per $10^7$ cells. 20 μl of each biotin-conjugated antibody for mouse anti-human CD44 (BD Pharmigen, clone G44–26) and mouse anti-human CD184 (BD Pharmigen, clone 12G5), as well as 20 μl of FcR human blocking reagent (Milenyi) were added per 40 μl of cell suspension. Cells were incubated for 15 min at 4 °C, followed by 2 washes with PBS + BSA. After a second wash, cells were resuspended in 60 μl of PBS + BSA per $10^7$ cells. 20 μl of anti-biotin Microbeads (Milenyi) were added and cells were again incubated for 15 min at 4 °C. Cells were washed twice and resuspended in 500 μl of PBS + BSA. Cell suspensions were applied onto a large cell column (Miltenyi) (pre-washed with PBS + BSA) attached to the magnet and the flow through solution containing the neuronal
enriched fraction was collected. The columns were washed 3 times by addition of 500 µl of PBS + BSA. The final 2 ml cell suspension was centrifuged and resuspended in neuronal differentiation media (see above). Cell number was determined and cells were seeded onto matrigel-coated dishes and cultured for 7 days completing a total of 4 weeks of neuronal differentiation.

**Sample Preparation**—Fetal autopsy brain samples without neuro-pathological abnormality on microscopic exam (supplemental Table S6) were obtained from archival tissue specimens from Mt. Sinai Hospital following approval by the Research Ethics Board (REB). Brain sections were prepared from routinely prepared FFPE fetal brain samples after deparaffinization in xylene and subsequent washes in ethanol. Forty-micron-thick sections were prepared on a microtome (Leica) onto PEN membrane slides (Zeiss) and H&E stained. Following pathological inspection and delineation of developing cerebrum, regions were dissected from areas ranging in size from 8–15 mm² and placed in tissue lysis buffer (0.25% Rapigest in 200 µm Tris, 5 mm DTT). Samples were placed on ice for 30 min and sonicated for 5 pulses, 10 s each at 3.5 amplitude setting (Qsonica XL-2000 Sonicator). Lysates were placed on heat block for 45 min at 95 °C followed by 1.5 h at 80 °C with brief vortexing every 10 min. Lysates were spun down at 14,000 × g, supernatants were transferred to a clean centrifuge tube and protein was quantitated using Pierce Coomassie protein assay kit (ThermoFisher Scientific). 10 µg of protein was alkylated in 15 µl iodoacetamide, 50 µm ammonium bicarbonate (200 µl total volume) solution for 30 min at room temperature in the dark. Samples were trypsin (Sigma) digested in a 1:50 substrate/ enzyme ratio overnight in a 37 °C incubator with constant shaking. Digestion reaction was quenched with 1% trifluoroacetic acid. Samples were spun down at 14000g and supernatant transferred to a clean tube. NPC/Neuronal tissue culture samples were prepared identically but with the lysis buffer composition of 8 µm urea in 50 µm ammonium bicarbonate; boiling step was omitted in favor of a 50 °C reduction step in presence of 5 mM DTT.

**LFQ Mass Spectrometry**—Each sample was concentrated using Omix C18MB (Agilent Technologies, Santa Clara, CA) tips and eluted with 3 µl of buffer A (0.1% formic acid, 65% acetonitrile). To each sample, 57 µl of buffer B (0.1% formic acid) was added, of which 10 µl (1.5 µg of peptides) was loaded from a 96-well microlate autosampler onto a C18 trap column using the EASY-nLC1000 system (Thermo Fisher Scientific, San Jose, California) and running Buffer C (0.1% formic acid). The trap column consisted of IntegraFit capillary (inner diameter 150 µm, New Objective, Woburn, MA) cut to 3.2 mm in length and packed in-house with 5 µm Pursuit C18 (Agilent Technologies). Peptides were eluted from the trap column at 300 nl/min with an increasing concentration of Buffer D (0.1% formic acid in acetonitrile) over a 60-min gradient (120 min for Neuron/NPC samples) onto a resolving 15 cm long PicoTip Emitter (75 µm inner diameter, 8 µm tip, New Objective) packed in-house with 3 µm Pursuit C18 (Agilent Technologies). The liquid chromatography setup was coupled online to Q Exactive Plus (Thermo Fisher Scientific) mass spectrometer using a nanoelectrospray ionization source (Thermo Fisher Scientific) with capillary temperature set to 275 °C and spray voltage of 2kV. A 60-min data-dependent acquisition (DDA) method was setup on the Q Exactive Plus. The full MS1 scan from 400–1500 m/z was acquired in the Orbitrap at a resolution of 70,000 in profile mode with subsequent fragmentation of top 12 parent ions using the HCD cell and detection of fragment ions in the Orbitrap using centroid mode at a resolution of 17,500. The following MS1 method parameters were used: MS1 Automatic Gain Control (AGC) target was set to 3e6 with maximum injection time (IT) of 100ms, MS2 AGC was set to 1e5 with maximum IT of 50 ms, isolation window was 1.6Da, underfill ratio 2%, intensity threshold 4e4, normalized collision energy (NCE) was set to 27, charge exclusion was set to fragment only 2+,3+ and 4+ charge state ions, apex trigger was deactivated, peptide match set to preferred and dynamic exclusion set to 45 s.

**Data Set Availability**—The mass spectrometry proteomics data have been deposited to the ProteomeXchange consortium via the PRIDE (25) partner repository with the data set identifier PXD004075 (NPC/neuron data set; name: reviewer11537@eibi.ac.uk, password: 6avg2OaZ); PXD004076 (FFPE data set; name: reviewer83184@eibi.ac.uk, password: oicZsIEn9) and PXD005065 (CMV data set; name: reviewer96639@eibi.ac.uk, password: nHX-PRezLy). Spectra files have also been deposited to MS-viewer (26) under the following search terms: vh4xmp5zq (NPC/neuron data set), owbnsn6ic (FFPE data set) and kleh8gfbx (CMV data set).

**Bioinformatics**—All mass spectrometry RAW files were uploaded into MaxQuant version 1.5.2.8 (www.coxdocs.org) and searched with built-in Andromeda search engine against a Human SwissProt protein database (January 2015 release, 89,649 entries) The following parameters were used for the search: trypsin/P enzyme with up to 2 missed cleavages allowed; carbamidomethylation of cysteine was set to fixed modification; with oxidation of methionine, phosphorylation of serine, threonine and tyrosine, N-terminal protein acetylation and ubiquitination of lysine set as variable modifications; first search peptide tolerance was set to 20 ppm against a small “human-first-search” database for the purpose of mass recalibration and main search was performed at 4.5 ppm; contaminants were included in the search; database was reversed for the purpose of calculating the peptide and protein level false-discover rate (FDR) at 1%; match between runs was set to a 0.7 min window; LFQ minimum-ratio count was set to 1.

**Experimental Design and Statistical Rationale**—For statistical comparisons we used eight fetal brains of varying gestational ages (n = 8) for regional identities. For the purposes of defining trends in temporal changes by K-means clustering we subgrouped FFPE cerebral regions into early (GW16–20, n = 3), mid (GW20, n = 2) and late (GW26–36, n = 3) gestational periods corresponding to pre-, mid- and postgoligogenesis periods, respectively. For neuronal cultured cells, three independent differentiations of iPSC#37 to NPCs (n = 3) and differentiated neurons (n = 3) were used. Data sets were imported into Perseus software platform (www.coxdocs.org, version 1.5.2.6) and contaminants, proteins only identified by site and reverse identifications were filtered out of the data set. In cases where multiple proteins from a protein family were listed in the identification the first family member was used in the analysis. For regional protein module identifications in the FFPE data set only proteins present in 50% samples in at least one group whereas in the NPC/neuron datanbset present in 1 sample in at least one group were included for performing statistical tests. LFQ values were Log2 transformed and samples with missing values for a given protein were assigned a random value using the imputation principle (downshift 1.75, width 0.4) as previously described (7). All graphs in the manuscript, PCA and hierarchical clustering were performed using this imputed data set unless otherwise stated. Differentially abundant proteins were determined using Welch’s t test with the FDR cutoffs at the indicated thresholds. Hierarchical clustering was carried out on Z-score transformed LFQ values using complete linkage of Euclidian distance. For GO term enrichment analysis, Fisher exact test was computed at the indicated FDRs (See supplemental Table S5 for statistical manipulations). For comparison of our proteomic data sets to the Brainspan_.org transcriptome, Log10 LFQ values were compared with the downloaded transformed Illumina RNA intensities (RPKM) and protein and RNA data was matched based on gene names.

**Immunocytochemistry and qRT-PCR**—Neurons were fixed with 4% paraformaldehyde in 0.4 m sucrose Krebs buffer for 10 min at room temperature. Fixed cells were permeabilized with 0.1% Triton X-100 for 10 min and blocked for 1 h at room temperature in blocking buffer (10% normal goat serum, 1% bovine serum albumin in PBS).
Primary antibodies were diluted in blocking buffer, and incubations proceeded overnight at 4 °C. Primary antibodies and dilutions were as follows: S100 (Abcam, Cambridge, UK, ab868, 1:1000), GFAP (Invitrogen, Carlsbad, CA, 1:250), SOX1 (R&D AF3369, 1:1000), CRABP1 (Sigma-Aldrich, St. Louis, MI, HPA017203, 1:500), FLNC (Sigma-Aldrich, HPA006135, 1:500) and MAP2 (Syntnic Systems, Goettingen, Germany, 188–004, 1:1000). Secondary antibodies were conjugated to Alexa Fluor dyes (Thermo-Fisher or Jackson ImmunoResearch, West Grove, PA, diluted 1:500 or 1:250, respectively). Secondary antibodies were diluted in PBS, and incubations performed for 1 h at room temperature. Nuclei were counterstained with DAPI diluted 1:5000 in PBS. After immunocytochemistry cover slips were mounted in ProLong Gold antifade mounting medium (Thermo-Fisher). To compare protein levels detected by MS in FFPE tissue to a well-accepted alternative method of protein detection, we performed heat induced epitope retrieval immunohistochemistry in EDTA buffer with antibodies against GFAP (Dakocytomation, Glostrup, Denmark, 1:500), NES (Millipore, Billerica, MA, 1:4500) and CALB2 (Dakocytomation, 1:500) on contiguous sections taken from the same blocks used for MS analysis. qRT-PCR was performed as previously described (27). Briefly, RNA was collected at the end of the neuronal differentiation period using TRIZOL (Invitrogen) and reverse transcribed using the SSII RT kit (Invitrogen). 50 ng of cDNA was used for qRT-PCR using SYBR green master mix (ABI) in triplicates and analyzed using the SSII RT kit (Invitrogen). Molecular & Cellular Proteomics 16.9

RESULTS

Proteomic Profiles of Developing Cerebral Compartments and Neuronal Cultures—To determine the applicability of MS-based FFPE tissue proteomics on neural development, we profiled four regions from eight pathologically normal human fetal brains ranging in gestational age from 16–36 weeks (GWs). For each brain, we macroscopically dissected the ventricular zone (VZ), intervening intermediate zone (IZ), subplate (SP) and frontal cerebral cortex (Cx) at the level of the anterior striatum (Fig. 1A). As a complementary analysis, we profiled proteomes of three independent iPSC-based differentiations from NPC to mature cortical neuron stage (Fig. 1A). We previously demonstrated that the employed neuronal differentiation protocol on this same iPS cell line yields electrophysiologically active cortical neurons (28). Here we further enrich for differentiated neurons by depleting astrocytes and undifferentiated NPCs using MACS (see Materials and Methods). Comparison of FFPE tissue and cell culture neurodevelopment models served two functions: (1) qualitative assessment of whether FFPE tissues yield proteins of similar cellular components and molecular functions to that of live cells of neuronal lineage and (2) quantitative comparisons of protein abundance changes between precursor and maturing compartment of both in vivo and in vitro models. For FFPE protein preparations, we made use of rapigest, an MS-compatible lysis reagent, enabling isolation and quantification of proteins from small tissue samples. Although commonly used to increase protein coverage, performing peptide fractionations was impractical because of the minute protein yields of our small macro-dissected brain regions. Following trypsin digestion and C18 stage-tip cleanup, tryptic peptides were analyzed on a Q Exactive liquid chromatography-tandem mass spectrometry with nanospray ionization. Resultant MS spectra were identified and quantitated in the MaxQuant software package (www.coxdocs.org) using the LFQ algorithm (29, 30). This abbreviated MS analysis approach, without the use of fractionation or detergent cleanup steps, significantly shortens the time from sample collection to data output. The analysis yielded 16,784 peptides mapping to a total of 3041 proteins, at a 1% peptide-level FDR, with the average number of measured proteins from ∼2200–2400 proteins per FFPE brain subcompartment (Fig. 1B, supplemental Table S1). The in vitro neuronal differentiation data set had a comparable number of 16,926 peptides mapping to 2875 proteins (supplemental Table S2), with 1967 protein identifications also found in our regional FFPE data set (Fig. 1B). The majority of proteins were quantitated in all eight samples of the developmental cohort (supplemental Fig. S1A) enabling confident temporal analysis of protein abundance within the subcompartments. 95% of the quantitated proteins spanned seven orders of magnitude and different FFPE tissue samples had comparable peptides per protein ratios and sequence coverage (supplemental Fig. S1B–S1D). Biological function and cellular component gene ontology (GO) terms (david.ncifcrf.gov) of identified proteins are highly similar in both FFPE tissue and live cell data sets (Fig. 1D). Importantly, a portion of identified proteins are localized to neuronal-specific processes, including axons and synapses, and carry out vast biological functions from cell cycle regulation to RNA splicing and neuronal differentiation. In fact, the total number of identified proteins in our FFPE tissue subcompartments are comparable to previous regional fresh frozen proteomes of human brain (31–33) and the nonformalin fixed in vitro cultures presented here. Although the recent draft of the human protein database reports much higher identification of 14,671 proteins in the human brain (2), this number reflects those identified following extensive peptide fractionations of bulk brains rather than single unfractionated brain subcompartments used in our study. Here we present the first LFQ MS data set of iPSC-derived cortical neurons and the largest FFPE-based proteomic data set of non-fractionated CNS tissue samples. We believe the benefit of anatomically and temporally defined proteomic data sets significantly outweighs the aforementioned drawbacks of using FFPE and non-fractionated tissue. This rapid and validated proteomic workflow thus provides a highly scalable, anatomically precise and dynamic tool to probe biology in practically limitless amounts of archival human tissue.

Proteomic Dynamics of iPSC-derived NPCs and Differentiated Cortical Neurons—To determine whether our abbreviated MS-based LFQ approach has satisfactory resolving power of different cellular states we compared protein abundance measurements in three independent cellular states from the iPSC to NPC to differentiated neurons following a 6-week protocol in cell culture. For this analysis, we made use
of an extensively characterized iPS cell line (hiPSC#37) which was used in neurodevelopmental disease modeling and neuronal function analysis (22, 28, 34). We previously showed by single cell Fluidigm qRT-PCR assays that the chosen differentiation strategy yields MAP2/H11001/NCAM1/H11001 neuronal cells that co-express either GAD67 and VGAT (GABAergic markers, 38% of neurons) or CAMK2/VGLUT1/2/3 (glutamatergic markers, 45% of neurons) (28). qRT-PCRs confirm that the majority of these cortical specification genes are present at the RNA level following differentiations of the samples used in our proteomic analysis (supplemental Fig. S2). Given that differentiation protocol can yield a mixture of committed neurons as well as glia and undifferentiated NPCs, we devised a MACS protocol to further enrich for neurons in culture (supplemental Fig. S2B). Pearson correlation coefficients (PCC) of global LFQ values are higher between three independent neuronal differentiation replicates compared with NPC lines (PCCNeur#1 vs. Neur#2 = 0.91 versus PCCNeur#1 vs. NPC#1 ~ 0.76) (Fig. 2A) and they are sufficient to discriminate the three stages of differentiation using the principal component analysis (Fig. 2B). Among 2875 quantitated proteins 306 with significantly different levels (FDR ~ 0.1) between precursor cells and pure neurons were identified, encompassing both high and low abundance proteins (Fig. 2C, supplemental Ta-
ble S3). Among the largest fold changing proteins are members of lamin protein family, LAMB1, LAMA1, and LAMC1 known to be up-regulated during neurite outgrowth as well as early markers of neuronal differentiation SYN1, STXBP1, TUBB3, DCX, MAP2, and NCAM1. The NPC compartment on the other hand exhibited a high expression pattern of MCF2L, an ERK signaling pathway molecule, as well as the hallmark neural progenitor markers VIM and NES. Collectively, the differentially abundant proteins largely encompass proteins with GO terms related to neuronal function or maturation, demonstrating that MS-detected changes in protein levels relate to biological processes of interest (Fig. 2D).

![Proteomic profiles of NPCs and differentiated neurons in cell culture.](image)

**A.** Pearson correlation and protein LFQ scatterplot matrix of three independent iPS cells, NPC lines and differentiated neurons following a 6-week differentiation protocol and antibody-based neuronal enrichment. **B.** Principal component analysis of three cellular states based on the LFQ intensity values of all identified proteins within the data set. See supplemental Table S3 for the entire list of differentially abundant proteins. **D.** GO Terms with an enrichment factor above 1.3 and below 0.5 in the differentially abundant protein list using Fisher exact test (Benjamini-Hochberg FDR < 0.1). **E.** Top 24 proteins with GOBP functions of glutamate signaling or synaptic transmission. LFQ values in NPCs and neurons are shown with NaN indicating not detected by MS measurements. **F.** Euclidean hierarchical clustering of top differentially abundant proteins between NPCs and differentiated neurons (306 proteins, Welch’s t test Permutation-based FDR < 0.1).
proteins, 24 proteins with synaptic or glutamate signaling pathway associated GO functions were found (Fig. 2E). These proteins, such as RAB3A, are involved in neuronal vesicle transport and were found at higher levels in differentiated neurons while virtually undetected in NPC and iPS cell populations. Performing hierarchical clustering based on the differential proteins demonstrates that pluripotent and neuronal precursor cell populations are overall more like each other and iPS cells compared with differentiated neurons, which form their own Euclidian distance node (Fig. 2F). Two distinct clusters of LFQ intensity patterns, with GO terms of DNA organization, replication and RNA splicing and those of synaptogenesis and neurotransmitter-related functions, describe either NPC- and neuron-enriched proteins, respectively (supplemental Fig. S3). Furthermore, extracting measurements of our neuron-enriched proteins from the human proteome map resource (www.humanproteomemap.org) reveals elevated values in either fetal or adult brain compared with other human tissues (supplemental Fig. S4). This finding indicates that proteomic dynamics captured in our in vitro differentiations reflect elevated protein levels in primary human neuronal tissues. This provides confidence that the quantified proteins recovered through our workflow faithfully mirror their appropriate neural state of the profiled cell populations. Because transcription factor proteins involved in NPC maintenance or maturation, such as SOX2, are not captured in our data set, it is possible that their detection may require fractionation approaches as is known for low abundance nuclear proteins. However, it is important to note, we identified 114 differentially abundant membrane-associated proteins between neuron- or NPC-enriched cell populations. (supplemental Fig. S5). This data set is thus especially useful for future extracellular cell surface protein-based cell NPC- and neuronal cell-sorting strategies. Overall, discovery of both high (cytoskeleton-related proteins) and low abundance proteins (localized in axonal and dendritic processes) with significant abundance changes indicates that we can reliably probe differentiation-related protein dynamics in unfraccionated protein samples using our high-throughput MS workflow.

Spatial Resolution of the Developing Fetal Brain—To probe CNS proteome changes across time we compared the LFQ intensity values in 36 FFPE samples comprising four compartments from eight fetal brains, ranging in developmental age between 16–36 GWs, and gray and white cortical regions of a postnatal brain (41 GWs) and one adult brain. Cell types enriched in these regions include neural precursor cells and neuroblasts in the VZ, developing axonal processes and oligodendrocytes in the IZ/SP and maturing neuronal and glial cells in the cortex. As expected, Pearson correlation coefficient between regional neighbors showed strong concordance (e.g. VZ16GW versus IZ16GW, PCC = 0.86) and a diminishing correlation was found with increasing anatomical (e.g. VZ16GW versus Cx16GW, PCC = 0.75) and temporal (e.g. VZ16GW versus VZ36GW, PCC = 0.67) separation (Fig. 3A).

Classification based on the identified LFQ values through principal component analysis and hierarchical clustering revealed distant segregation of VZ from the neuron-rich cortex, with reduced segregation of the IZ and SP proteomic profiles (supplemental Fig. S6). Molecular function GO terms responsible for this divergence were determined by Fischer exact test where DNA replication processes were enriched in the VZ whereas synaptic assembly and transmission proteins were enriched in the cortex (Fig. 3B and supplemental Fig. S7). This analysis highlights the reliability our rapid workflow to accurately capture sub-organ level anatomic and region-specific biology within FFPE human CNS tissue.

To assess the protein to RNA relationships in the developing cerebrum, we took advantage of the publicly available developmental transcriptome RNAseq normalized RPKM data set (www.brainspan.org). Within regional brain RNAseq profiles of this resource, we mined the RNAseq information for the medial ganglionic eminence samples of developmental age 8–9 GW (corresponding to our VZ region) and the dorso-lateral prefrontal cortex (DLPCx) from 16–37 GWs (five fetal brains, corresponding to our Cx regions) to compare with our proteomic data set. By ranking protein and RNA expression levels we obtained spearman correlation values ranging from 0.48 (VZ16 GW versus MGE9 GW) to 0.39 (Cx18 GW versus DLPCx19 GW) (Fig. 3C). In general, the NPC-rich VZ had a higher RNA to protein correlation coefficient (~0.48) than the neuronal-rich cortex (~0.39–0.45) with its age matched regional counterparts. This level of similarity is in line with the current estimates of RNA to protein relationships (2). This finding is encouraging given that our analysis utilizes RNAseq and proteomic information derived from different biological samples. Although the highest levels of correlation were found when RNA and proteomes of same brain regions and developmental stage were compared, the relatively low levels of correlation reinforce the importance of complementary proteomic measurements. RNA to protein comparisons enabled us to identify GO term enrichments associated with proteins in the cerebrum with higher or lower transcript levels in the RNAseq data set (Fig. 3D). This analysis indicates that in both the cortex and the VZ compartments, genes involved in transcriptional and DNA binding activity generally have higher transcriptional than translational outputs whereas enzymatic and transporter-associated genes are relatively more abundant at the protein level (Fig. 3E).

Quantified proteins that define regional identity of subcompartments within the developing human brain include some well-established biomarkers. Proteins specific to VZ (NES and VIM), to IZ (NEFL, NEFM, and NEFH) and NPC (SATB2 and TBR1) all exhibited largely restrictive patterns of expression as measured by LFQ intensities (Fig. 4A). To determine temporal protein abundance changes within the regions we performed K-means clustering of the LFQ values and found 6 clusters of temporal protein dynamics (supplemental Fig. S8). These clusters demonstrate patterns of expression of proteins
FIG. 3. Comparison of identified proteomes between developing brain subcompartments. A, Scatterplot and Pearson correlation coefficient (PCC) matrix demonstrate time- and regionproximity related correlation values based on LFQ intensities of identified proteins in four regions of 10 fetal and postnatal FFPE brains. B, 2D GO term annotation enrichment analysis comparing average LFQ values of proteins quantitated in the VZ, the neural precursor compartment, and the mature neuron-rich cortex layer (FDR < 0.1). C, Spearman rank coefficient matrix of the LFQ intensity values compared with normalized RNAseq RPKM values from the brain regional transcriptome (www.brainspan.org), DLPCx, dorsolateral prefrontal cortex; MGE, medial ganglionic eminence. D, Comparison of protein LFQ to RPKM RNAseq values in the cortex and ventricular zone. Genes with extreme discordance between protein and RNA quantifications are highlighted (RNA > protein, red; RNA < protein, green). E, Fischer exact enrichment test of molecular function and biological processes of proteins indicated in (B), p < 0.02.
that are gradually increased or decreased in abundance over developmental time and those that rapidly change during the late developmental time period (26–36 GW). For instance, temporal protein changes are evident in several proteins, including nestin (NES, in VZ cluster 1), whose levels are highest prior to 20 GW and diminished with advancing gestational age whereas the astrocytic marker glial fibrillary acidic protein (GFAP, in VZ cluster 0) increased dramatically at late gestation coinciding with gliogenesis within VZ (Fig. 4B). Conversely in the cortex, neural-specific proteins such as calretinin (CALB2, in cortex cluster 4), a neuronal Ca\(^{2+}\) binding protein, gradually became more abundant with developmental maturation. Importantly, MS-based quantification mirrored abundance level changes observed by immunolabeling (Fig. 4C) attesting to the reliability and applicability of LFQ MS-based assays for large scale analysis of protein dynamics in development. Thus, even within heterogenous cellular regions, monitoring of spatiotemporal dynamics offer a powerful tool for identification of proteins with putative functions of interest.

Overall, we identified 609 (FDR<0.1) proteins with level differences between cortex and ventricular zone cellular layers (Fig. 5A, supplemental Table S4). Abundant proteins in the cortex include SYN3, SRGAP3, FARSB, TUBA1A, TBR1, and SATB2, all of which have drastically reduced levels in the progenitor compartment. Among the VZ-enriched proteins we identified specific regulators of cell division (e.g. complete hexomeric minichromosome maintenance (MCM) protein complex) (supplemental Fig. S9). The broad distribution of DCX within the IZ/SP and the more restricted expression of Filamin A in the VZ support the well-established regional roles of these proteins in neuronal migration (supplemental Fig. S9).

Analysis of spatiotemporal protein dynamics provides a validated discovery platform for identification of novel biomarkers of neural precursors and development. We focused on VZ-specific proteins that were also identified in cultured NPCs and had reduced levels in differentiated neuronal cultures and other areas of the cortex. One such protein was Filamin C (FLNC), an extremely poorly annotated cytoskeletal protein in
the CNS. Immunostaining of FLNC in sectioned FFPE tissues revealed exclusive localization of signal to the lateral wall of lateral ventricles (Fig. 5B), a known human neural stem cell niche that persists well into adulthood. Its highly localized and punctate staining along the lateral ventricle may reflect the recognized importance of ventricular wall contact to neural stem cell maintenance (35). Like the distribution of a more well characterized neural stem cell marker SOX2, there was no FLNC+ cells in medial aspect of the lateral ventricle (Fig. 5B). This suggests that, like nestin, FLNC may be a cytoskeletal protein that marks NPCs. This is further substantiated by loss of FLNC from cultured NPCs following neuronal differentiation, as demonstrated by both our MS and confirmatory immunostaining analysis (Fig. 5C). Consistent with a putative NPC specific marker, we find FLNC is widely distributed in well-characterized glioblastoma stem cell lines (unpublished data). These findings all converge to suggest FLNC a strong and novel candidate neural precuor marker and highlight the discovery potential of our unique data set.

Another relatively uncharacterized protein in neurodevelopment that was enriched in the VZ was CRABP1. Interestingly, unlike other precursor markers such as NES, CRABP1 levels increased in abundance with gestational age coinciding with neurogenesis and peaking around 26GW (supplemental Table...
This suggested that it may be a biomarker of a more committed neural cell within the VZ. To resolve its spatial and cellular distribution within the developing cerebrum, immunostaining of the fetal brain revealed positive cells forming large streams of cells tangentially oriented to the ventricular zone near the medial ganglionic eminence (Fig. 5D). Consistent with this origin, strong neuronal staining was found in the neurons of the corpus striatum (36) (Fig. 5D). Comparable to our MS analysis, radially migrated cortical neurons were completely negative for this marker. On the other hand, weakly staining cultured NPCs resolved into rare intensely stained cells with cytoplasmic and dendritic CRABP1 distribution following neuronal induction (Fig. 5E). These examples clearly demonstrate the power MS-derived proteomic data sets of sub-anatomical compartments across fetal development in identifying novel mediators of brain development. Further characterization of early (i.e. FLNC) and late (i.e. CRABP1) stage markers of neural precursor biology could serve as additional readouts for successful isolation of properly staged NPCs and neurons for regenerative medicine efforts.

Distinct spatial resolution of the cortex and VZ samples was carried out through PCA and hierarchical clustering of 609 differentially abundant proteins (Fig. 6A–6B). Complete segregation of the VZ from the cortex samples was observed whereas IZ and SP regions clustered in mixed nodes suggesting that, at the analyzed developmental time points, these anatomical compartments share proteomic signatures. To compare the quantitative output of the FFPE data set to that of the in vitro neuronal differentiation model we extracted the LFQ values of the brain region proteins and compared them to the NPC/neuronal data set. Spearman correlation test reveals that the in vitro established NPCs are most similar to the 20 GW VZ (SRC = 0.6) and have a reduced similarity with both the IZ/SP (SRC, 0.21–0.47) and Cx (SRC, 0.39–0.51) compartments (Fig. 6C). On the other hand, our MAP2+ electrophysiologically active neurons share similarity with every analyzed FFPE compartment but are most closely resembling the midgestation cortex (GW 18–22, SRC, 0.58–0.64) with much reduced similarity to both postnatal and adult cortex (SRC = 0.44). Our analysis supports, at the protein level, a widely held view that further differentiation optimizations are required to achieve true mature neuronal identity in a dish. An additional explanation for our lower correlation between the two data sets as compared with regions within each data set could be the cellular heterogeneity of primary CNS tissue. The utility of
such comparisons would greatly benefit by complementary work where specific cell types of the brain can be sorted, although fresh tissue would be required, to reveal cell typespecific proteomes as has recently been carried out in the murine neuronal system (16).

Toward the goal of defining candidate proteins for such sorting efforts we merged the two proteomic data sets to identify proteins with changes in abundance in in vitro and in vivo models of neurodevelopment. By merging the two data sets, we identified 114 differentially abundant proteins between cultured NPCs compared with differentiated neurons and VZ samples compared with cortex regions (supplemental Table S7). Many of these proteins were localized to the cellular membrane, based on the GOCC analysis, as well as components of the cytosol and the nucleus. Pearson correlations of average protein intensity values in pre-gliogenesis (GW16, 18, and 20; 3 cases) and post-gliogenesis (GW26 and 36; 3 cases) developmental periods in the cerebrum demonstrate that NPCs and neurons are most similar to early stages of development, rather than postnatal or adult brain regions, of their precursor or neuron-rich brain compartment counterparts, respectively (Fig. 7A). With a few exceptions (14/114), most proteins were similarly downregulated (34/114) or up-regulated (66/114) in differentiating neurons (Fig. 7B). Proteins characterized by GO terms related to the extracellular membrane that are more abundant in neurons than NPCs include Neural cell adhesion molecule 1 (NCAM1), Synapsin 1 (SYN1), Laminin subunit gamma 1 (LAMC1) and Alpha-internexin (AINX1). Furthermore, synapse proteins Syntaxin-binding protein 1 (STXBP1) and Synaptotagmin 1 (SYT1) are undetected in NPCs and glutamate signaling-related proteins found to be elevated in the cultured neurons and the cortex include GNAQ and ATP1A3. Mining for similar proteins, that are exclusively found in differentiating neurons, serve as excellent candidates for future cell sorting experiments. On the other hand, neural precursor brain regions contain elevated perilipin 3 (PLIN3), High mobility group protein B2 (HMGB2), Alpha-actinin 1 (ACTN1), Filamin A (FLNA), Transgelin (TGLN), Vimentin (VIM) and Nestin (NES), among others. Proteins that are undetected by MS in most Cx and all cultured neuron samples but have LFQ intensity values in both VZ and NPCs include Acyl-CoA-binding domain-containing protein 7 (ACBD7), Aldehyde dehydrogenase family 16 member A1 (ALDH16A1) and Septin-10 (SEPT10). Overall, this analysis demonstrates the utility of two independent data sets for identifying neuron or neural precursor cell-enriched proteomic modules confirmed in heterogeneous in vivo tissues and purified in vitro cell populations. Furthermore, these region-specific modules can be utilized to define regional and cellular perturbations in pathological states (i.e. CMV-induced encephalitis, see supplemental Text).
DISCUSSION

With increasing technological advancements, the complex human proteome can now be resolved using mass spectrometry. Although large-scale efforts are profiling human tissues at large to define tissue-specific proteomic signatures, developmental- and region-specific proteomic discoveries are sparse. Combination of this emerging proteomic profiling technology with widely available archival human FFPE tissue promises to uncover large amounts of disease-relevant molecular information. This has recently been achieved in human B-cell lymphoma (7) and breast cancer (3). Here we use and validate a simplified and convenient adaptation of traditional proteomic profiling approaches to generate the first LFQ data set of iPS cell-mediated neuronal differentiations and a spatiotemporal snapshot of human neurodevelopment protein dynamics. Although we focused our current study to a well-defined cortical region, the practical modifications can be adapted to additional brain regions, developmental ages, and human diseases states. Higher protein yields can be achieved using carefully selected fresh or frozen tissue or by using higher protein quantities coupled with fractionation. The ability to rapidly perform profiling on microscopically defined regions however, makes this approach highly amenable for high-throughput screening of relevant complex and highly dynamic biological processes. Our modification of traditional protocols involving extensive fractionations and detergent cleanup steps, reduces the mass spectrometry time to under 2 h per sample enabling the quantification of 100 samples within 10 days of machine analysis time. This procedure is however not without its likely limitations and caveats. Primarily, a multitude of CNS proteins were not identified using our approach which could partly be a result of the limited area (anterior frontal lobe) we sampled or because of specific developmental windows we profiled. Additional regions and stages of maturation will likely increase total protein counts. With that said, we demonstrate the power and utility of spatio-temporal proteomic profiling in uncovering novel biological markers and pathological processes not possible by using highly fractionated static and bulk-tissue profiling approaches.

Tissue heterogeneity is also a factor, as with many of the large-scale transcriptomics and proteomic screening efforts (1, 2, 19). Even within our anatomically defined regions, proteins from intervening vasculature, connective tissue and blood cells can not be separated. At the protein level however, this can be easily and precisely addressed using companion cell-to-cell immunohistochemical analysis or comparison to purified collections of specific cell types. Indeed, our MACS enrichment protocol allowed us to obtain the first label-free proteome data set for human cortical neurons derived from iPSCs. Although we only focused on neuronal populations in our in vitro analysis, similar approaches with MACS-enriched astrocytes and oligodendrocytes can be used to identify additional cell type-specific developmental protein changes, as has recently been completed in the mouse brain tissue. Conversely, laser capture technologies of immunostained sections can be used to isolate specific cell populations or smaller regions prior to proteomic profiling from FFPE tissue. The added benefit of monitoring changes through time and space partly overcomes this issue as it allows selection and refinement of protein lists by linking dynamic protein changes to well understood developmental milestones. This was evident by robust detection of proteins like Filamin-C and CRABP1 which were only found in a subset of cells within the heterogeneous ventricular zone. As we show, having rich reference maps to compare with disease states also allows subtraction of specific inflammatory components of disease and analysis of specific changes in biological processes of interest (e.g. ectopic, premature or prolonged developmental milestones) in disease states like CMV encephalitis.

In summary, we demonstrate that the dynamic and evolving function of specific anatomical brain subcompartments can be molecularly defined through this scalable proteomic technique. Assembly of such molecular atlases can serve as a reference resource in the understanding molecular programs that dictate human brain development and remodeling later in life and disease. The adaptability of this cost- and time-effective technique to rich supply of well-annotated archival FFPE human material will undoubtedly prove useful in delineating disease mechanisms in primary human tissue, not only in the CNS but throughout the human body. Such proteomic profiling will not only translate into diagnostic signatures but may also be instrumental for rational individualized drug development efforts.

DATA AVAILABILITY

The mass spectrometry proteomics data have been deposited to the ProteomeXchange consortium via the PRIDE (25) partner repository (http://www.ebi.ac.uk/pride/archive/) with the data set identifier PXD004075 (NPC/neuron data set), PXD004076 (FFPE data set) and PXD005065 (CMV data set). Annotated spectra files have also been deposited to MS-Viewer (http://msviewer.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msviewer) (26) under the following search keys: vh4mxp5qze (NPC/neuron data set), owbnsrs6ic (FFPE data set) and klei8gfbx (CMV data set).

Acknowledgments—We thank Lilian Lee, Michelle Kushida and Heather Whetstone for technical expertise; Wei Wei and Alina Piekna for neuronal differentiation assistance. Authors thank all members of the Diamandis lab, Dr. Peter Dirks and Dr. Samer Hussain for helpful discussions.

* The research was supported by the Department of Pathology at the University Health Network (PD) and Ontario Brain Institute (POND network to JE). DCR was supported by a RettSyndrome.org Mentored Fellowship. Authors declare no conflict of interest.

§§ To whom correspondence should be addressed: Laboratory Medicine and Pathobiology, University of Toronto; E-mail: p.diamandis@mail.utoronto.ca.

[§] This article contains supplemental material.
Marsala, M., Gage, F. H., Goldstein, L. S. B., and Carson, C. T. (2011) Cell-surface marker signatures for the isolation of neural stem cells, glia and neurons derived from human pluripotent stem cells. *PLoS ONE* **6**.

Vitcaino, J. A., Csordas, A., Del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q. W., Wang, R., and Hermjakob, H. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, D447–D456.

Baker, P. R., and Chalkley, R. J. (2014) MS-viewer: a web-based spectral viewer for proteomics results. *Mol. Cell. Proteomics* **13**, 1392–1396.

Fussner, E., Djuric, U., Strauss, M., Hotta, A., Perez-Iratxeta, C., Lanner, F., Dilworth, F. J., Ellis, J., and Bazett-Jones, D. P. (2011) Constitutive heterochromatin reorganization during somatic cell reprogramming. *EMBO J.* **30**, 1778–1789.

Djuric, U., Cheung, A. Y., Zhang, W., Mok, R. S., Lai, W., Piekna, A., Hendry, J. A., Ross, P. J., Pasceri, P., Kim, D. S., Saltar, M. W., and Ellis, J. (2015) MECP2e1 isoform mutation affects the form and function of neurons derived from Rett syndrome patient iPSC cells. *Neurobiol. Dis.* **76**, 37–45.

Mortensen, P., Gouw, J. W., Olsen, J. V., Ong, S. E., Rigbolt, K. T., Bunkenborg, J., Cox, J., Foster, L. J., Heck, A. J., Blagoev, B., Andersen, J. S., and Mann, M. (2010) MSQuant, an open source platform for mass spectrometry-based quantitative proteomics. *J. Proteome Res.* **9**, 393–403.

Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372.

Martins-de-Souza, D., Guest, P. C., Steeb, H., Pietsch, S., Rahmoune, H., Harris, L. W., and Bahn, S. (2011) Characterizing the proteome of the human dorsolateral prefrontal cortex by shotgun mass spectrometry. *Proteomics* **11**, 2347–2353.

Martins-de-Souza, D., Guest, P. C., Guest, F. L., Bauder, C., Rahmoune, H., Pietsch, S., Roeber, S., Kretzschmar, H., Mann, D., Baborje, A., and Bahn, S. (2012) Characterization of the human primary visual cortex and cerebellum proteomes using shotgun mass spectrometry-data-independent analyses. *Proteomics* **12**, 500–504.

Fernández-Irigoyen, J., Corrales, F. J., and Santamaría, E. (2012) Proteomic atlas of the human olfactory bulb. *J. Proteomics* **75**, 4005–4016.

Hotta, A., Cheung, A. Y., Farra, N., Vijayaragavan, K., Séguin C. A., Draper, J. S., Pasceri, P., Maksakova I. A., Mager, D. L., Rossant, J., Bhatia, M., and Ellis, J. (2009) Isolation of human iPS cells using EOS lentiviral vectors to select for pluripotency. *Nat. Methods* **6**, 370–376.

Doetsch, F., García-Verdugo, J. M., and Alvarez-Buylla, A. (1999) Regeneration of a germinal layer in the adult mammalian brain. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11619–11624.

Marin, O., and Rubenstein, J. L. (2001) A long, remarkable journey: tangential migration in the telencephalon. *Nat. Rev. Neurosci.* **2**, 780–790.