A multiregional proteomic survey of the postnatal human brain

Becky C. Carlyle1, Robert R. Kitchen2, Jean E. Kanyo3, Edward Z. Voss3, Mihovil Pletikos4, André M. M. Sousa4, TuKiet T. Lam2,3, Mark B. Gerstein2, Nenad Sestan4,5,6* and Angus C. Nairn1,6*

Detailed observations of transcriptional, translational and post-translational events in the human brain are essential to improving our understanding of its development, function and vulnerability to disease. Here, we exploited label-free quantitative tandem mass-spectrometry to create an in-depth proteomic survey of regions of the postnatal human brain, ranging in age from early infancy to adulthood. Integration of protein data with existing whole-transcriptome sequencing (RNA-seq) from the BrainSpan project revealed varied patterns of protein–RNA relationships, with generally increased magnitudes of protein abundance differences between brain regions compared to RNA. Many of the differences amplified in protein data were reflective of cytoarchitectural and functional variation between brain regions. Comparing structurally similar cortical regions revealed significant differences in the abundances of receptor-associated and resident plasma membrane proteins that were not readily observed in the RNA expression data.

Work from converging fields is beginning to disentangle the immense complexity of the mammalian brain, from the mapping of the connectome1,2 to ever deeper molecular characterization of brain regions3–6 and cell types7–15. Extensive maps of RNA abundance across the mammalian brain have been available since 201116. The focus of projects such as BrainSpan16, PsychENCODE17 and organizations such as the Allen Brain Institute (http://www.brain-map.org/) is to increase the depth of these datasets, cataloging cell types, coding and noncoding RNA expression and epigenetic modifications17. Technological advances in microfluidics and nucleic acid sequencing have led to single-cell and single-cell-type transcriptome analyses of increasingly high quality18–20. Automation of electron microscopy is providing unprecedented resolution with which to survey the connections of neurons in fixed tissue21, while optogenetics allows us to define functionally important circuits in rodent models22. Despite this huge increase in our collective ability to interrogate the mammalian nervous system, comprehensive protein-level data, particularly in human brains, is a notable exception. Given the likely poor correlation between mRNA and protein abundance23,24, and given that protein levels are closer to the biosynthetic output of the cell21,22, a systematic survey of the human brain proteome is vital. Here we present a detailed proteomic analysis of seven major regions of the human brain across postnatal development from early infancy through adulthood.

Tandem liquid chromatography–mass spectrometry (LC-MS/MS) is the gold-standard method for obtaining unbiased, high-resolution proteomic data22,23. Current technology makes it possible to obtain quantitative observations of peptides derived from several thousand proteins per sample24–26. However, the depth of mass spectrometry is limited by the complexity of the sample being investigated. In heterogeneous brain tissue, sensitivity to low-abundance proteins is likely decreased due to the presence of a large number of spectra from high-abundance proteins, such as actin or tubulin21. This effect is magnified when highly diverse samples are combined in an isobaric labeling experiment27,28. For this reason, we adopted a dual approach to analyzing human brain samples similar to that of a recent high-quality study of the mouse brain proteome29. First was a discovery phase, designed to create a highly sensitive, heavily fractionated spectral library for each adult brain region. Then, to produce a quantitative analysis for each adult and postnatal development sample, ‘single shot’ label-free LC-MS/MS runs were used to accurately quantify proteins on the basis of detected precursor peptide ion masses.

To maximize compatibility with existing -omic profiles of the human brain, the post-mortem tissue samples used for quantitative proteomics were exact matches to those already profiled by RNA-seq in the BrainSpan16 project (http://www.brainspan.org/); thus these matched RNA and protein samples constitute a detailed and deeply integrated resource. Finally, we compared the human discovery and quantitative proteomic data to the mouse brain proteome29 to create a point of reference for the neuroscience community.

Results

Powdered frozen tissue was obtained from postnatal samples previously profiled by RNA-seq as part of the BrainSpan Project (Fig. 1a). Of the 16 regions profiled in BrainSpan, we selected 7 regions that showed large inter-regional differences in gene expression by RNA-seq: cerebellar cortex (CBC), striatum (STR), mediodorsal thalamic nucleus (MD), amygdala (AMY), hippocampus (HIP), primary visual cortex (V1C) and dorsolateral prefrontal cortex (DFC). Developmental samples spanned the period from early infancy (1 year after conception) to adulthood (42 years)16 and were derived from a near-equal mix of male and female donors (Fig. 1a and Supplementary Table 1). Consistent and stringent quality control
Peptide library of seven adult human brain regions. To create a peptide library for each region, we generated seven region-specific collections of tissue homogenate pooled from adult subjects. Trypsin-digested peptides from these pools were separated offline by high pH reverse phase chromatography into 15 fractions before online LC-MS/MS. Mass spectra were searched in MaxQuant against the Ensembl/Gencode proteome. A total of 111,456 peptides, corresponding to 8,980 proteins, were detected in at least one region, with an average of 7,945 proteins per region (Fig. 1c, Supplementary Figs. 1 and 2, and Supplementary Table 2). On average, we detected over 9 peptides per protein (Supplementary Fig. 2) resulting in a mean-average coverage of 28% of the observed proteins (Supplementary Table 3a,b). Data from several recent, large scale proteomic projects suggest that the number of proteins detectable in a single biological sample tends to saturate at approximately 11,000. These data are in rough agreement with RNA analyses, which frequently report that ~50% of protein-coding genes are expressed in a given tissue or cell type. In our data, the observed proteins constituted the majority of coding genes detected by RNA-seq, with a strong bias toward proteins corresponding to more highly expressed RNAs (Fig. 1d). Over 60% of genes expressed at >1 read per kilobase per million mapped (RPKM) and over 80% of genes expressed at >10 RPKM were detected as proteins (Fig. 1e). Peptides were detected for proteins across a wide range of molecular weights, from two 44-amino-acid variants of β-thymosin (TMSB10 and TMSB4X) to fragments of TTN, the largest known protein. Analysis of the gene ontology category ‘integral membrane component’ showed no significant enrichment or depletion of these potentially difficult-to-extract proteins. Taken together, these data indicate a broad and detailed protein library.

Quantitative single shot proteomics of postnatal human brain regions. Single-shot, label-free protein quantifications (LFQs) were obtained from the seven brain regions of 16 individuals spanning 1 year after conception to 42 years. LFQ was used to maximize the number of protein identifications per sample and also for comparability with the recent high-quality study of the mouse brain. As with the fractionated peptide atlas, mass spectral data from the single-shot runs were searched in MaxQuant against the Ensembl/Gencode proteome. To increase the number of protein identifications, single-shot data were searched alongside the fractionated samples, using the ‘match between runs’ feature in MaxQuant. A total of...
63,478 peptides derived from 7,244 proteins were detected in one or more of the 77 single-shot samples analyzed, detecting on average 18,835 peptides from 3,612 proteins per sample. The use of match between runs produced ~50% more protein identifications per sample than would have otherwise been detected (Supplementary Fig. 3). A large fraction (5,151; 71%) of the 7,244 total proteins were reliably quantified by LFQ (Supplementary Table 4).

Single-shot LFQs were filtered, normalized and log transformed (see Methods and Supplementary Fig. 4) before differential expression (DEX) analysis, jointly modeling protein changes over time and between brain regions by regression and ANOVA. A total of 1,804 proteins were found to have significant DEX between one or more brain region. Given the limited number of developmental samples available (six time points spanning from infancy to adulthood), we had comparatively low power to detect significant time-dependent changes, likely resulting in a small number of proteins with significant DEX across development. We detected 123 proteins with significant DEX across developmental periods (Bonferroni adjusted (adj.) \( P < 0.05 \), Fig. 2a and Supplementary Table 5). Using the proteins found to be differentially expressed between brain regions, samples clustered in a predictable manner, with CBC a clear outlier (Fig. 2b and Supplementary Fig. 5).

We performed cluster analysis in which proteins showing DEX were grouped according to their relative abundance across all single-shot samples. Proteins formed distinct groups that reflected the substantial differences in their abundance between brain regions. Clusters comprising proteins enriched or depleted in the major non-cortical structures, such as the CBC (for example, Fig. 2c, clusters 20 and 21) and STR (clusters 26 and 31), were the most striking (see Methods, Supplementary Table 5a–c and Supplementary Figs. 6 and 7). The largest CBC-enriched cluster (cluster 7) contained a significant number of proteins annotated as involved in mRNA processing, such as RNPS1, SF3A2, SNRPA1, SRRM1, SRSF10, SRSF11, SRSF6 and SRSF9 (Fisher’s exact test, 'RNA processing' adj. \( P = 9.6 \times 10^{-2} \), cellular component 'nucleus' adj. \( P = 1.85 \times 10^{-4} \)). This functional enrichment is reflective of many of the CBC-enriched clusters and presumably reflects the much higher density of nuclei in the cerebellar granular layer compared to the other brain regions. Several clusters

Fig. 2 | Differentially expressed proteins across human brain regions and postnatal development. a. The majority of significantly differentially expressed genes were between brain regions, rather than over developmental periods (two-way ANOVA, significant DEX defined as Bonferroni-corrected \( P < 0.05 \); adj. \( P \) values of \( n = 5,141 \) genes across 7 regions (6 d.f., Fig. 1a) and 6 time points (1 d.f., Fig. 1a) in Supplementary Table 5). b. Clustering all samples subjected to MS/MS using proteins significantly differentially expressed between brain regions revealed expected bulk differences between brain regions. Samples are defined by the same color scheme used to depict regions and developmental period in Fig. 2c,d (see also fully labeled, zoomable version in Supplementary Fig. 5). CBC and STR are clear outliers (lower left), as they are by RNA-seq, and the remaining samples cluster well by region with the exception of occasional outlying samples derived from the youngest subject, HSB139 (dark blue). c. Clustering all proteins significantly differentially expressed across human brain regions and postnatal development. A large fraction (5,151; 71%) of the 7,244 total proteins were reliably quantified by LFQ (Supplementary Table 4).
highlighted the diversity of the other brain regions. For example, clusters 26 and 31 (Fig. 2c) contain PDE10A (cluster 26), TH (cluster 26) and CHAT (cluster 31), proteins known to be functionally important in the striatum. This was confirmed by network analysis (STRING) of clusters 26 and 31 combined, which showed a strong enrichment for established protein–protein interactions (adj. \( P = 1.35 \times 10^{-15} \)) and for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways involved in stimulant addiction (cocaine addiction adj. \( P = 2.55 \times 10^{-6} \), ‘amphetamine addiction’ adj. \( P = 0.9606 \), Supplementary Fig. 8) and ‘dopaminergic synapse’ (adj. \( P = 0.001 \), well established processes relevant to the human striatum).

Quantitative single-shot proteomics of brain development. Several clusters of temporally coexpressed proteins highlighted increasing or decreasing abundance over postnatal brain development (Fig. 2d, Supplementary Fig. 9 and Supplementary Table 5). A striking number of the time-DEX proteins appeared to be exclusively expressed during early infancy, a time characterized by corticocortical and long afferent axon reorganization, synaptogenesis and spinogenesis. While there were no statistically enriched gene ontologies in this cluster (cluster 1, Fig. 2d), it included many candidates, such as EFNB3, BZW2, CNTN3, CELSR2, NYAP2, TENM4, CSPG4, SEZ6, RAPHI and LRRMT4, known to function in cell–cell adhesion and downstream signaling, a critical process for appropriate neuronal maturation.

Cluster 6 was a small cluster of six proteins that increased in abundance across postnatal development and that may be linked to several neurological disorders. NFASC, a protein present in cluster 6 that is a product of alternative splicing and that anchors myelin to axons, is known to be disrupted in multiple sclerosis and decreased methylation of the body of the gene encoding TPPP, another member of cluster 6, has been linked to depression in children subject to early life stress. Decreased methylation of the body of the gene encoding TPPP, another member of cluster 6, has been linked to depression in children subject to early life stress. TPPP may also be found in alpha-synuclein-positive protein deposits in Parkinson’s and Lewy body dementia.

Quantitative spatiotemporal comparison of mRNA and protein abundance. All of the brains used in this study have been previously characterized by RNA-seq as part of the BrainSpan project (Supplementary Table 6a,b), allowing the ready integration of gene-level RNA expression data with quantitative proteomic data (5,039 genes in total). This allowed us to disambiguate the protein-level expression of 264 genes (5%) for which the peptide mapping was ambiguous (Supplementary Tables 6a,b and Methods). Principal component analysis of protein- and RNA-level data showed a similar proportion of variance attributed to the first and second components (90.5% protein and 91.2% RNA; Fig. 3a). In these components, however, the CBC showed a greater separation from the other regions in the protein data, while also showing clearer grouping of the other regions compared to RNA.

The correlation between absolute mRNA and protein abundance within one sample is typically lower than the correlation between mRNA and protein fold changes. We therefore directly compared inter-regional fold changes between RNA and protein across all possible region pairs. This approach produced a median Pearson correlation coefficient of 0.32 (Fig. 3b). Genes found to have significant DEX between brain regions at the protein level were significantly skewed toward a more positive correlation (median correlation coefficient of 0.48; Kolmogorov–Smirnov test (KS) \( P < 1 \times 10^{-15} \); Fig. 3b). A shift to the right was observed for proteins with DEX over the developmental time course (median correlation coefficient of 0.41), but this shift was not statistically significant (\( P = 0.8911, n = 123 \)). Region-level summary expression and differential expression results for both the RNA and protein data are provided in Supplementary Table 7.

Using the region-averaged protein and RNA expression data, we identified the abundance and relative expression of the 20 most enriched proteins and RNAs in each brain region (Fig. 4). In general, the top 20 genes were enriched to a greater degree in CBC or STR than in neocortical regions by both RNA and protein. Furthermore, the fraction of genes that appeared in the top 20 for both protein and RNA was greater in these noncortical regions (−50% in the CBC and STR) than in neocortical regions (−20% in the V1C and DFC).

For all possible region pairs, we categorized gene products based on their fold-change similarity between RNA and protein (Fig. 5a and Supplementary Fig. 10a–g). Over all region pairs, the majority of genes (62.5%) showed a significant but not substantial (≤2-fold) inter-regional abundance difference by both RNA and protein (Fig. 5a,b; all two-region comparisons in Supplementary Fig. 11a).
These were considered the ‘no change’ group. A further 17.4% showed larger (>2-fold) region-specific expression differences in the same direction in both RNA and protein. These were considered the ‘agree’ and ‘partially agree’ (same direction of change, different magnitude of fold change) groups. However, we observed over twice as many genes with ≥2-fold DEX by protein (‘protein only’) but not
RNA (11.5%) compared to ≥2-fold DEX by RNA (‘RNA only’) but not protein (5.2%). A small fraction (3.3%) of genes disagreed in the direction of change between RNA and protein. These were defined as the ‘not agree’ category. Finally, the differences in a gene’s protein abundance between regions were generally greater than the differences in RNA abundance (Supplementary Fig. 11b).

Across all unique pairs of regions, genes found to be consistently differentially expressed (Supplementary Table 8) by protein-only—that is, DEX in protein but not RNA—were largely nuclear proteins (cellular component ‘nucleus’, adj. P = 8.6 × 10^{-28}, Supplementary Table 9a,b), and within the enriched nuclear GO terms (Supplementary Table 9a) there was a bias toward ‘RNA processing’ (adj. P = 3.3 × 10^{-28}). The enrichment of these GO terms in the ‘RNA-only’ category showed enrichment for ontological terms that describe proteins that may be transported away from the nucleus to the synapse and distal regions of neurons. Thus, while most mRNA in a cell is found in the cell body within the region measured, synaptic proteins are often transported along projections, typically into a different region. Enriched GO terms (Supplementary Table 9a) in the ‘RNA-only’ category include ‘ion transmembrane transport’ (adj. P = 3.2 × 10^{-28}), ‘signaling’ (adj. P = 5.9 × 10^{-28}) and ‘synaptic transmission’ (adj. P = 3 × 10^{-4}). Synaptic transmission was also significantly enriched in the ‘partially agree’ category, wherein RNA and protein changes occurred in the same direction within a region but to a different magnitude.

Looking more closely at regional comparisons, DFC and STR are well established as functionally linked regions^{41}, but differ with respect to their cytoarchitecture and developmental origin. In gene products differentially expressed between these two regions, there was still a protein-only enrichment of nuclear proteins in DFC (Fig. 5c and Supplementary Fig. 10h (annotated and zoomable), GO term ‘RNA processing’ adj. P = 0.03). ‘Synaptic transmission’ was found in the ‘RNA-only’ category, as well as in the ‘partially agree’ group, reflecting the distal transport of some of these proteins out of the region containing the cell bodies (Fig. 5c and Supplementary Fig. 10h, adj. P = 2.5 × 10^{-4}, 6.8 × 10^{-28}). In the ‘agree’ category, many GO terms reflected axonal proteins and development (Supplementary Table 9c,d), although signaling terms such as ‘regulation of GTPase activity’ (Fig. 5c and Supplementary Fig. 10h, adj. P = 3.2 × 10^{-28}) were also found in this category. As in all two-region comparisons (Fig. 5b), the largest number of gene products in the DFC–STR comparison were found in the ‘no change’ category (Supplementary Fig. 11a). These included proteins with stable mRNA and proteins with conserved functions^{41} across all cells, such as ‘translation’ (Fig. 5c and Supplementary Fig. 10h, adj. P = 1.8 × 10^{-3}) and ‘cellular respiration’ (Fig. 5c and Supplementary Fig. 10h, adj. P = 3.7 × 10^{-4}).

Comparing the two regions in our dataset with the highest developmental and cytoarchitectural similarity, the neocortical areas DFC and V1C, we observed much more pronounced protein-level
Fig. 6 | Comparison of the human and mouse brain proteomes. a, The cumulative frequency of Pearson correlations for each 1:1 ortholog protein between human and mouse shows a median correlation of 0.3 (n = 4,052). When considering only those proteins significantly differentially expressed at the protein level between human brain regions (n = 1,517), these correlations are significantly increased (median 0.65; KS P < 10−16). b, Human and mouse protein abundance differences between two example brain regions, PFC and STR, shows a lower overall degree of consistency between organisms compared to that between human RNA and protein (see Fig. 5a). As before, genes are colored on the basis of their agreement or disagreement between human and mouse; genes for which the human variability between regions was < 2-fold of that reported for mouse were considered consistent (green, mouse and human protein consistently (within 2-fold; gray lines) enriched in one of the regions, gray, mouse and human proteins not sufficiently different between regions to pass the > 2-fold change criterion). Purple genes are those with consistent direction but variable magnitude of change between the regions of human and mouse, while red genes disagree even in the direction of change. Blue and orange genes vary between regions according to human but not mouse and vice versa, respectively. c, Distribution of the number of genes annotated in each of the color categories defined in b across all unique pairs of brain regions. The center line indicates the median, limits indicate the IQR, and whiskers either 1.5 times the IQR or the minimum/maximum value if it falls within a factor of 1.5 of the IQR. d, Genes with poor correlation between human and mouse regions tend to have more sequence differences in their coding sequence compared to those genes with greater correlation (red vs. green (P = 0.04) or gray (P = 8.97 × 10−4)). Box plots are defined as in c, with the individual points representing outliers that fall more than a factor of 1.5 below the IQR and the notch representing the 95% confidence interval of the median.

Integration with mouse protein abundance from the Mouse Brain Proteome project. Given wide-ranging interest in the use of mouse models for human disorders, we compared the human protein data with a recent study on mouse brain regions27. No temporal data from whole tissue were available from the mouse; however, five brain regions matched those in our study: CBC, STR, thalamus, HIP and prefrontal cortex. Of the 16,217 1:1 human–mouse orthologs annotated by Ensembl, 4,052 were detected by LC-MS/MS in both the mouse and human datasets. Mouse and human protein expression showed a similar correlation to that of human RNA–protein (median Pearson correlation = 0.3); however, when only proteins differentially expressed between human regions were considered, this correlation was significantly higher (median Pearson correlation = 0.65, KS P < 10−16; Fig. 6a). This generally positive correlation with mouse tissue, in which the post-mortem interval was presumably controlled, was encouraging given the potential for nontryptic protease action during the post-mortem interval in human tissue28.

We computed fold changes for the human proteins and their mouse orthologs between the five shared regions (Fig. 6b, Supplementary Figs. 13 and 14a–e, and Supplementary Table 10).

No substantial region specific differences (≤ 2-fold) were detected for 47.5% of proteins in both human and mouse (Fig. 6c). A further 27.2% of proteins showed large (> 2-fold) region-specific expression differences, in the same direction, in both human and mouse. Around half as many proteins appeared to vary substantially in the mouse but not human (5.6%) as to vary in human but not mouse (11.6%). Compared to the human RNA vs. protein comparison, there were more genes and proteins that completely disagreed on their direction of change in human (8.1%). There was a weak but significant relationship between the conservation of protein sequence between human and mouse and the correlation of the human and mouse inter-regional fold changes (Fig. 6d). Those proteins with a fold change ≥ 2 in human only had a significantly lower percentage sequence identity than those genes exhibiting no substantial regional difference (P = 8.97 × 10−4) or those genes in high agreement (P = 0.04) between the two organisms (Fig. 6d).

Discussion

We anticipate this large human brain proteomic dataset will be of great use to the neuroscience community to investigate proteins of interest, their expression patterns, and conservation with mouse models. Overall, we directly observed 111,456 peptides, derived from 8,980 proteins, and obtained reliable quantifications for 5,151 of these proteins. A large fraction (~35%) of the quantified proteins (1,804) were found to be differentially expressed between brain regions, with much fewer (123) being identified as changing over the postnatal developmental time course. The small number of proteins with significant DEX between infancy and adult is likely a result of low power. A large proportion of the differential expression was driven by differences between other regions and the CBC. There were also very clear STR-enriched clusters of genes, which were reflected in the robust sample clustering of CBC and STR. We observed fewer developmental changes, which is likely a reflection...
of the fact that the most dramatic changes observed in gene expression occur at earlier time points, during prenatal development\textsuperscript{6,16}. Integration of protein data with existing whole-transcriptome sequencing revealed generally greater magnitudes of protein abundance differences between brain regions compared to RNA abundance difference. While many of these differences reflect cytoarchitectural, developmental and functional differences between brain regions, comparison of the more similar neocortical regions in our study revealed the presence of potential ‘protein only’ region markers. Comparison of human and mouse data showed a generally positive correlation between protein orthologs, which was significantly increased in those proteins found to be DEX between human regions.

In complex tissues such as the brain, protein abundance is highly dependent on the sites of synthesis and localization. It is also the case that different proteins and, to a lesser extent, mRNAs possess vastly different half-lives\textsuperscript{19,21}. To be able to accurately assess these relationships between RNA and protein, therefore, it is a great advantage to measure gene product abundance in the same exact biological samples, as we have done here. We have shown that although a majority of genes had similar magnitudes of change in their mRNA and protein abundance between brain regions, there were many that showed consistent differences in relative extent and direction of change. As such, mRNA expression is an incomplete measure of protein abundance, requiring direct proteomic assessment to generate an overview of gene expression and regulation in different regions of the brain.

Genes identified as differentially expressed between regions by ‘protein only’ tended to be enriched for nuclear functions, with a strong bias toward ‘RNA processing’ over other ontological terms, such as ‘chromatin modification’. RNA-processing genes have been reported to have stable protein but unstable RNA and would therefore be expected to be more enriched compared to genes related to chromatin modification, which have both unstable RNA and unstable protein\textsuperscript{21}. The stability of proteins involved in aspects of nuclear function likely contributed strongly to the significant differences between regions that arise from the number, relative density and size of nuclei. The most extreme example is the cerebellum, which contains approximately 68 billion of the 86 billion neurons estimated to make up the human brain\textsuperscript{45}, and this high density of nuclei was reflected in the highly significant enrichment for nuclear GO terms seen in many CBC-enriched clusters.

Previous analyses of the mRNA from the brain samples used revealed subtle expression differences between the regions of the neocortex\textsuperscript{11-13}. Thus, the 11 neocortical regions were so similar by mRNA-seq that they were collapsed when compared to the subcortical regions and over the developmental time course\textsuperscript{16}, a phenomenon that is even more apparent in preadolescence\textsuperscript{8}. Meanwhile, total-RNA analysis of the neocortex, which is not restricted to measuring only coding RNAs, suggests that it is lower-abundance noncoding RNAs that are largely responsible for the differences between neocortical regions. However, our proteomic analysis of the DFC and V1C found many more differentially expressed proteins between these neocortical regions than we found differentially expressed mRNAs. Furthermore, we found little evidence of nuclear enrichment in these protein-only DEX genes, likely due to the similar cytoarchitecture of these two regions. Instead we observed significant enrichments for genes implicated in ‘receptor activity’ or ‘localization to the plasma membrane’. It is therefore plausible to hypothesize that coding genes substantially contribute to the diversity and molecular differentiation of the neocortical regions, but that these differences are being magnified in their protein abundance compared to their mRNA expression.

In neurons, the extent of localized (non-cell-body) protein translation is unclear. As high-throughput techniques for measuring the impact of local translation improve\textsuperscript{46}, it seems likely that production of protein from mRNA transcripts occurs primarily in the cell body, with a subset of functionally important proteins being synthesized in the dendritic and potentially axonal compartments\textsuperscript{47}. It is therefore possible that differences in mRNA and protein abundance are driven by highly mobile proteins and mRNAs located far from their brain region of origin. For example, we observed an enrichment of synaptic proteins in the ‘RNA only’ category, suggesting that while the mRNA would largely be present in the source region containing the cell soma but not in the axonal target region, protein that is subsequently transported between the brain regions would be detected in both. Such genes would only be differentially expressed between the regions in the ‘RNA only’ data. Therefore, while it is important to understand where and when a gene is expressed, this may not accurately reflect the abundance of the final protein product.

Despite inherent problems working with the post-mortem human brain, notably the highly variable post-mortem intervals that limit observability of post-translational modifications and affect nonproteic protease digestion\textsuperscript{48,49,50}, the data presented here are a rich source of information. For example, the database of observable peptides in each brain region (Supplementary Table 2) may be used to design targeted assays as a replacement for low-throughput methods such as immunoblotting. Peptides observed in the single-shot data can be used as a high-confidence subset for quantitation, and the protein quantifications themselves (Supplementary Table 4) can be used to estimate stoichiometry for molecular modeling of these brain regions. In cases where the mRNA and protein are highly correlated (Supplementary Tables 7 and 8), RNA measurement is likely to be preferable in terms of ease, cost and replicability. It is also straightforward to compare regional protein abundance in the mouse vs. human (Supplementary Table 10), an important consideration for animal models of disease and neuropharmacology. These data have the potential to become a valuable resource for neuroscientists, and the approaches used are a step toward broader adoption of proteomic techniques by the neuroscience community.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41593-017-0011-2.

Received: 3 January 2017; Accepted: 27 September 2017;
Published online: 13 November 2017

References

1. Silbereis, J. C., Pochareddy, S., Zhu, Y., Li, M. & Sestan, N. The cellular and molecular landscapes of the developing human central nervous system. Neuron 89, 248–268 (2016).
2. Kashuth, N. et al. Saturated reconstruction of a volume of neocortex. Cell 162, 648–661 (2015).
3. Swanson, L. W. & Lichtman, J. W. From Cajal to connectome and beyond. Annu. Rev. Neurosci. 39, 197–216 (2016).
4. Oh, S. W. et al. A mesoscale connectome of the mouse brain. Nature 508, 207–214 (2014).
5. Cui, J. et al. Spatial mapping of structural and connectional imaging data for the developing human brain with diffusion tensor imaging. Methods 73, 37–37 (2015).
6. Lerner, T. N., Ye, L. & Deisseroth, K. Communication in neural circuits: tools, opportunities, and challenges. Cell Rep. 164, 1136–1150 (2016).
7. Furtuzinhos, S. et al. Laminar and temporal expression dynamics of coding and noncoding RNAs in the mouse neocortex. Cell Rep. 6, 938–950 (2015).
8. Pletikos, M. et al. Temporal specification and bilaterality of human neocortical topographic gene expression. Neuron 81, 321–332 (2014).
9. Bakken, T. E. et al. A comprehensive transcriptional map of primate brain development. Nature 535, 367–375 (2016).
10. BrainSpan Consortium. Technical white paper: transcriptome profiling by RNA sequencing and exon microarray. http://help.brain-map.org/download/attachments/3506181/Transcriptome_Profiling.pdf?api=v2 (2013).
11. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161, 1202–1212 (2015).
12. Marques, S. et al. Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. Science 352, 1326–1329 (2016).
13. Zeisel, A. et al. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
14. Krishnaswami, S. R. et al. Using single nuclei for RNA-seq to capture the transcriptional states of postmortem neurons. *Nat. Protoc.* **11**, 499–524 (2016).
15. Heiman, M. et al. A translational profiling approach for the molecular characterization of CNS cell types. *Cell* **135**, 738–748 (2008).
16. Kang, H. J. et al. Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
17. Akbarian, S. et al. The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015).
18. Burgess, D. J. Technology: a drop in single-cell challenges. *Nat. Rev. Genet.* **16**, 376–377 (2015).
19. Schwabhauser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
20. Li, J. J., Bickel, P. J. & Biggin, M. D. System wide analyses have underestimated epigenetic abstractions and the importance of transcription in mammals. *PeerJ* **2**, e270 (2014).
21. Schwabhauser, B., Wolf, J., Selbach, M. & Busse, D. Synthesis and degradation jointly determine the responsiveness of the cellular proteome. *Bioessays* **35**, 597–601 (2013).
22. Kitchen, R. R., Rozovsky, J. S., Gerstein, M. B. & Nairn, A. C. Decoding neuroproteomics: integrating the genome, transcriptome and functional anatomy. *Nat. Neurosci.* **17**, 1491–1499 (2014).
23. Mann, M., Kulak, N. A., Nagaraj, N. & Cox, J. The coming age of complete, accurate, and ubiquitous proteomes. *Mol. Cell* **49**, 583–590 (2013).
24. Nagaraj, N. et al. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7**, 548 (2011).
25. Beck, M. et al. The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7**, 549 (2011).
26. Aebersold, R. & Mann, M. Mass-spectrometric determination of proteome structure and function. *Nat. Rev. Genet.* **7**, 347–357 (2006).
27. Wang, H., Alvarez, S. & Hicks, L. M. Comprehensive comparison of iTRAQ and label-free LC-based quantitative proteomics approaches using two *Chlamydomonas reinhardtii* strains of interest for biofuels engineering. *J. Proteome Res.* **11**, 487–501 (2012).
28. Latoisinka, A. et al. Comparative analysis of label-free and 8-plex iTRAQ approach for quantitative tissue proteomic analysis. *PLoS One* **10**, e0137048 (2015).
29. Sharma, K. et al. Cell type- and brain region-resolved mouse brain proteome. *Nat. Neurosci.* **18**, 1819–1831 (2015).
30. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
31. Djebsi, S. et al. Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
32. Li, Z. et al. Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. *J. Proteome Res.* **11**, 1582–1590 (2012).
33. Steiner, H. & Tseng, K. Y., eds. *Neuroproteomics: integrating the genome, translatome and functional standards.* Springfield, Virginia, USA, (2017).
34. Szklarczyk, D. et al. STRINGv10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
35. Volkov, N. D., Mori, T. & Morales, M. The brain on drugs: from reward to addiction. *Cell* **162**, 712–725 (2015).
36. Tessier-Lavigne, M. & Goodman, C. S. The molecular biology of axon guidance. *Science* **274**, 1123–1133 (1996).
37. Scheiffele, P. Cell-cell signaling during synapse formation in the CNS. *Annu. Rev. Neurosci.* **26**, 485–508 (2003).
38. Lindner, M., Ng, I. K. M., Hochmeister, S., Meinl, E. & Linington, C. Neurofascin 186 specific autoantibodies induce axonal injury and exacerbate disease severity in experimental autoimmune encephalomyelitis. *Exp. Neurol.* **247**, 259–266 (2013).
39. Mathey, E. K. et al. Neurofascin as a novel target for autobody-mediated axonal injury. *J. Exp. Med.* **204**, 2363–2372 (2007).
40. Weder, N. et al. Child abuse, depression, and methylation in genes involved with stress, neural plasticity, and brain circuitry. *J. Am. Acad. Child Adolesc. Psychiatry* **53**, 417–424.e5 (2014).
41. Montalvo-Oritz, J. L. et al. The role of genes involved in stress, neural plasticity, and brain circuitry in depressive phenotypes: Convergent findings in a mouse model of neglect. *Behav. Brain Res.* **315**, 71–74 (2016).

42. Kovács, G. G. et al. Natively unfolded tubulin polymerization promoting protein TIPP/p25 is a common marker of alpha-synucleinopathies. *Neurobiol. Dis.* **17**, 155–162 (2004).
43. Hintiryan, H. et al. The mouse cortico-striatal projectome. *Nat. Neurosci.* **19**, 1100–1114 (2016).
44. Seyfried, N. T. et al. Quantitative analysis of the detergent-insoluble brain proteome in frontotemporal lobar degeneration using SILAC internal standards. *J. Proteome Res.* **11**, 2721–2738 (2012).
45. Lillas, R. R., Walton, K. D. & Lang, E. J. Cerebellum in The Synaptic Organization of the Brain (ed. Shepherd, G. M.) 271–310, https://doi.org/10.1039/acropf0870151 (2003).
46. Namjoshi, S. V. & Raab-Graham, K. F. Screening the molecular framework underlying local dendritic mRNA translation. *Front. Mol. Neurosci.* **10**, 45 (2017).
47. Holt, C. E. & Schuman, E. M. The central dogma decentralized: new perspectives on RNA function and local translation in neurons. *Neuron* **80**, 648–657 (2013).
48. Dammer, E. B. et al. Neuron enriched nuclear proteome isolated from human brain. *J. Proteome Res.* **12**, 3193–3206 (2013).
49. Tagawa, K. et al. Comprehensive phosphoproteome analysis unravels the core signaling network that initiates the earliest synapse pathology in preclinical Alzheimer’s disease brain. *Hum. Mol. Genet.* **24**, 540–558 (2015).

Acknowledgements

We thank S. Leslie and D. Li for discussions. Data were generated as part of the PsychENCODE Consortium, supported by U01MH103339, U01MH103365, U01MH103392, U01MH103346, U01MH103346, R01MH105472, R01MH094714, R01MH105898, R21MH102791, R21MH105881, R21MH103877 and P50MH106934 awarded to S. Akbarian (Icahn School of Medicine at Mount Sinai), G. Crawford (Duke University), S. Dracheva (Icahn School of Medicine at Mount Sinai), P. Farnham (University of Southern California), M.B.G. (Yale University), D. Geschwind (University of California, Los Angeles), T.M. Hyde (Lieber Institute for Brain Development), A. Jaffe (Lieber Institute for Brain Development), J.A. Knowles (University of Southern California), C. Liu (University of Illinois at Chicago), D. Pinto (Icahn School of Medicine at Mount Sinai), N.S. (Yale University), P. Sklar (Icahn School of Medicine at Mount Sinai), M. State (University of California, San Francisco), P. Sullivan (University of North Carolina), E. Vaccarino (Yale University), S. Weissman (Yale University), K. White (University of Chicago) and P. Zandi (Johns Hopkins University). This work was supported by the Yale/NIDA Neuroproteomics Centre (DA018343-12), by NIA grant AG047270-02, by NIMH grant MH119026, by NIH SIG grants 1S10DD009670-0 and 1S10DD008304-01, and by the State of Connecticut, Department of Mental Health & Addiction Services. B.C.C. was supported by a 2014 NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation.

Author contributions

B.C.C. designed the experiments, performed the experiments, analyzed the data and wrote the manuscript. R.R.K. designed the experiments, analyzed the data and wrote the manuscript. J.E.K. performed the experiments. E.Z.V. performed the experiments. M.P. contributed to tissue and sample processing, A.M.S. contributed to tissue and sample processing. T.T.L. performed the experiments and wrote the manuscript. M.B.G. contributed to RNA-seq data generation and provided computational resources. N.S. designed the experiments, contributed to tissue and sample processing, contributed to RNA-seq data generation and wrote the manuscript. A.C.N. designed the experiments and wrote the manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-017-0011-2.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to N.S. or A.C.N.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Methods

**Tissue procurement.** This study was conducted using frozen post-mortem human brain specimens from a number of tissue collections facilitated by the BrainSpan and PsychENCODE consortia (see Kang et al.1 for specific sample handling and preservation details). Briefly, tissue was only obtained after obtaining parental or next-of-kin consent and with approval by the institutional review boards at the Yale School of Medicine, the National Institutes of Health (NIH) and each institution from which tissue specimens were obtained. Tissue was handled in accordance with ethical guidelines and regulations for the research use of human brain tissue set forth by the NIH (https://policymanual.nih.gov/3014) and the WMA Declaration of Helsinki (https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/).

Appropriate informed consent was obtained and all available nonidentifying information was recorded for each specimen. Specimens ranged in age from 1 year after conception to 40 years. The post-mortem interval (PMI) was defined as hours between time of death and time when tissue samples were frozen. To ensure appropriate representation of the dissected sample region, frozen samples were pulverized in liquid nitrogen using a ceramic mortar and pestle.

Of the 16 brain regions profiled in BrainSpan, the 7 regions that showed the largest inter-regional differences in gene expression by RNA-seq were selected: cerebellar cortex (CBC), striatum (STR), medio-dorsal thalamic nucleus (MD), amygdala (AMY), hippocampus (HIP), primary visual cortex (V1C) and dorsolateral prefrontal cortex (DFC). Developmental samples spanned early infancy (1 year after conception; developmental period 8) to adulthood (42 years; developmental period 13)2; the subjects were a near equal mix of males and females (Fig. 1a and Supplementary Table 1). No statistical methods were used to determine sample sizes, but our sample sizes are similar to those reported by Sharma et al.3.

**Sample dissections.** Please see the supplement of Kang et al.4 for more detailed descriptions of the dissections of the regional samples. Briefly, the dissections were as follows.

DFC was sampled from approximate border between the anterior and middle third of the medial frontal gyrus. DFC corresponds approximately to Brodmann areas (BA) 9 and 46.

V1C was sampled from the area surrounding the calcarine fissure. Only samples in which the stria of Gennari could be recognized were included. V1C corresponds to BA 17 or the primary visual cortex.

HIP was sampled from the middle third of the retrocommisural hippocampal formation, located on the medial side of the temporal lobe. Sampled areas always contained the dentate gyrus and the cornu ammonis.

AMY included the whole amygdala.

STR included the head of the caudate nucleus and the putamen, separated by the internal capsule and ventrally connected to the nucleus accumbens.

MD included the whole mediodorsal nucleus of the thalamus (MD) sampled from the dorsal and mediodorsal thalamus. Small quantities of surrounding thalamic nuclei may be present in the sample.

CBC was sampled from the lateral part of the posterior lobe. The sampled area contained all three layers of cerebellar cortex and underlying white matter but not the deep cerebellar nuclei.

**Sample preparation.** Frozen powdered brain samples were weighed and added to lysis buffer (50 μL M in 4 M urea, 0.2% TFA), and protease inhibitor (Roche) at 1:10 weight:volume. Samples were homogenized by sonication and cleared by centrifugation at 16,000 × g, 4 °C, 10 min on a desktop centrifuge. Lysates were quantified by BCA assay and adjusted to 100 μg/mL in 50 μL 8 M urea, 0.4 M ammonium bicarbonate. pH was confirmed to be ~8. Dithiothreitol (45 mM at one-tenth volume sample) was added to lysates for 30 min at 37 °C, followed by addition of iodoacetamide (100 mM at one-tenth sample volume) for 30 min in the dark at room temperature. Samples were diluted to 2 μL of urea with deionized water before addition of trypsin in a 1:20 trypsin:protein ratio. Proteins were digested for 16 h at 37 °C. The digestion was quenched by adjusting the pH to below 3 by the addition of 20% trifluoroacetic acid and the sample was desalted using C18 Macro Spin Columns (Millipore) according to the manufacturer’s instructions. Peptides eluted from the column were dried in a SpeedVac (Savant, ThermoFisher Scientific) and stored at −20 °C. Dried peptides were dissolved in 3.5% formic acid, 0.1% trifluoroacetic acid (TFA), and peptide concentrations were estimated from A280 absorbance using a Thermo Scientific Nanodrop 2000. Aliquots were diluted additionally with 0.1% TFA to a final concentration of 0.04 μg/μL, with 0.2 μg loaded on column for mass spectrometric analysis. This procedure was used for both fractionated and single-shot runs.

**Fractionation of samples.** Master region pools were produced by adding 40 μg of homogenate from each of the adult (period 13) subjects, to make 200 μg total per region. Pooled lysates were processed for proteomic analysis as described above. For the fractionated samples, peptides were first dissolved in 53 μL 0.1% TFA and injected onto a Waters ACQUITY UPLC (BEH) C18 column (130 Å, 1.7 μm × 75 mm i.d. × 100 mm) at a flow rate of 0.4 mL/min. An orthogonal high pH reverse phase separation was carried out (using buffers A, 100% water with 10 mM ammonium acetate, pH 10; and B, 90% water, 10% acetonitrile with 10 mM ammonium acetate, pH 10) with a gradient of 0.0 min 2% B, 1.29 min 2% B, 19.83 min 37% B, 26.85 min 75% B, 33.06 min 98% B, 34.53 min 98% B, 37.47 min 2% B and 44.75 min 2% to separate the peptides. Forty-eight fractions were collected from each brain region and were then pooled on the basis of their estimated concentration from analysis of the chromatogram into 15 pools. Each pool was subsequently analyzed individually by LC-MS/MS.

Mass spectrometry analysis (LC-MS/MS). For both the pooled fractionated and single-shot quantitative runs, MS analyses were performed on a Q Exactive Plus mass spectrometer (ThermoFisher Scientific) coupled online to a Waters nanoAcquity UPLC in the low-pH condition. Peptides were separated over a 180-min gradient run using a Waters Symmetry C18 trap (1.7 μm, 180 μm × 20 mm) and a ACQUITY UPLC (BEH) C18 column (130 Å, 1.7 μm × 75 μm i.d. × 250 μm) at a flow rate of 0.3 μL/min with a gradient of 5% water, 0.1% formic acid and 1% buffer B (0.1% formic acid in acetonitrile), before eluting with linear gradients that reached 0.0 min 1% B, 140 min 30% B, 155 min 40% B and 160 min 85% B with a flow rate of 300 nL/min. MS1 (300 to 1,500 m/z, target value 3 × 105, maximum ion injection times 45 ms) were acquired and followed by higher energy collisional dissociation (normalized-collision energy 28). A resolution of 70,000 at m/z 200 was used for MS1 scans, and up to 20 dynamically chosen, most abundant precursor ions were fragmented (isolation window 1.7 m/z). The MS2 scans were acquired at a resolution of 17,500 at m/z 200 (target value 1 × 104, maximum ion injection times 100 ms). Samples were run in regional blocks, with control samples interspersed throughout to allow correction of batch effects.

**Data analysis.** Data collection and analysis were not performed blind to the conditions of the experiments. No samples were excluded from the analysis. Mass spectra were processed using MaxQuant29 (v1.5.2.1). Spectra were searched against the Homo sapiens protein sequences annotated in Gencode5 (version 21; hg38) using the Andromeda search engine51. This search included a fixed modification, cysteic acid, and methionine oxidation. Peptides shorter than seven amino acids were not considered for further analysis because of lack of uniqueness, and a 1% false-discovery rate (FDR) was used to filter poor identifications at both the peptide and protein level. Where possible, peptide identification information was matched between runs of the fractionated and single-shot samples within MaxQuant29. This exploited the accurate mass and retention times across LC-MS runs to infer the identity of a peptide in a particular run in which the precursor ion was detected but was not selected for identification by MaxS2.

For the single-shot spatiotemporal analysis of human brain samples, mass spectra were matched between the fractionated brain-region-specific samples (Supplementary Table 2) and the adult single-shot runs. Due to the number of extra protein identifications using this feature, we endeavored to estimate the likelihood of protein misidentification by the match between runs (MBR) feature. We took the 148 proteins observed only in the CBC fractionated data (CBC specific) and counted how many times, and by what means, these same proteins were identified in the adult single-shot samples (Supplementary Table 11). Only 49 of these CBC-specific proteins were identified by any means in the single-shot samples, a reflection of the substantially increased depth of fractionated vs. single-shot proteomics. As would likely be expected in this small subset of proteins, there were no extra identifications by methods in the CBC single-shot samples. 3.4% of possible observations from non-CBC single-shot samples occurred only at the MS1 match level. These observations may reflect a false identification of an isobaric ‘imposter’ peptide by the MBR feature, or they may simply reflect a peptide missed by the fractionated proteomic profiling of the other regions. Although it is not possible to confidently differentiate between these two possibilities, we estimate an upper limit of 3.4% for the protein misidentification rate and 1.8% for the rate at which proteins were undetected in the non-CBC fractionated runs.

To be included in the region-specific dataset, peptides were required to have at least two MS1/MS2 scans. Protein identification required at least one unique or razor peptide per protein group. Quantification in MaxQuant was performed using the label-free quantification (maxLFQ) algorithm52. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org/) via the PRIDE partner repository with the dataset identifier PXD005445.

**Sample preparation:** RNA extraction, library preparation and sequencing. Details of tissue handling, RNA sample preparation, library construction and sequencing can be found in the white paper at http://help.brain-map.org/download/attachments/3306181/Transcriptome_Profile.pdf.

**RNA sequencing alignment and expression quantification.** RNA-seq reads were realigned and quantified using a custom analysis pipeline. This involved filtering reads against known laboratory contaminant sequences in UniVec before further filtering by explicit alignment to the human 5.8S and 45S rRNA sequences. Reads aligned to these sequences were removed. Reads that did not map to UniVec or rRNA were aligned against the human genome (hg38) and known splice junctions.
The minimum cluster size was varied between the different gene sets to produce a threshold of 1:517 genes (region-averaged mouse mass-spectrometry data were obtained from the supplement of Sharma et al.29. Human:mouse 1:1 orthologs were obtained from BioMart (Ensembl Genes 85). Ortholog gene pairs were defined by matching gene symbols in the human with those in the mouse.

Statistics. Normalized LFQ human protein and RNA data (5,141 proteins, 77 samples from 17 individuals) were subjected to a gene-wise linear model to predict the coefficients of development period (continuous variable, 6 periods) and brain region (discrete variable, 7 regions). Data distribution was assumed to be normal, but this was not formally tested. Statistical significance was computed by two-way ANOVA and P-values were corrected for multiple hypothesis testing using Bonferroni’s method. A threshold of P < 0.05 was used to define significantly differentially expressed (DEX) genes across either developmental period or brain region, with an additional criterion of a minimum fold change of 2 for genes included in GO analysis.

DEX genes over developmental period or brain region were clustered using the R package ‘dynamicTreeCut’32. Euclidean distances between genes were clustered using the ‘hclust’ function (method = average) before cluster discretization using a dynamic tree cut (method = hybrid, deepSplit = T, pamStage = T, maxPamDist = 0). The minimum cluster size was varied between the different gene sets to produce more easily interpretable results; for period-DEX genes the minimum cluster membership was set to 2 and for region-DEX proteins the minimum cluster membership was set to 10.

For the protein/RNA comparison, KS tests were performed to assess whether RNA/protein correlation was altered in regionally DEX genes (n = 1,776) versus all detected genes (n = 5,039). For the human/mouse comparison, only those data from the five regions (CBC, STR, MD, HIP and frontal cortex: dorsolateral PFC in primate, PFC in mouse) and 4,052 genes represented in both datasets were used (1,517 genes were region significant). Human differential expression analysis was performed using the subset of genes and regions in the same manner as before. For the mouse, equivalent replicate-level data were unavailable from the supplement of Sharma et al.29, so instead human-equivalent coefficients were computed from the region-averaged data available in their Supplementary Information. This simply involved a rescaling of the reported abundances from log₂, to log₁₀.

To assess gene-set enrichments in the regionally differentially expressed protein and RNA data, we downloaded gene sets from known gene ontology terms and acccessions for each of the three GO domains using the biomaRt package in R. All ontological analysis was performed using the R package topGO (https://biocoductor.org/packages/release/bioc/html/topGO.html) and all P-values corrected for multiple tests using the Benjamini–Hochberg method33. Briefly, genes were classified on the basis of their RNA vs. protein fold-change agreement in each region pair, visually summarized in Fig. 5. We sought to identify enriched ontological terms in genes that either agree, partially agree or disagree in the fold-change reported by RNA-seq and proteomics, including genes that appear highly differentially expressed in one of these assays but not the other. Only significantly differentially expressed genes were input (background) and the classification based on ±2-fold changes between region pairs.

We also classified each gene on the basis of its consistency of RNA/protein fold-change agreement across all pairs of regions. For each gene, regions with fold changes <2 in both RNA and protein were discarded and genes appearing in the same category (e.g., protein-only DEX) in more than 50% of the remaining region pairs were annotated as consistent—in this example, as ‘proteinOnly’.

Ontological enrichment analysis was performed as before with the individual region pairs.

Finally, we sought to increase the scope and relevance of this enrichment analysis by including all hallmark, positional, BioCarta, KEGG, reactome, miRNome, and TTP targets available in the Molecular Signatures Database (v5.2)34. Enrichments here were calculated using Fisher’s exact tests and P-values again corrected for multiple comparisons using the Benjamini–Hochberg method.

Exact N values for each functional enrichment test can be found in the “Annotated” and “Significant” columns of Supplementary Fig. 9.

To generate the network figures in Supplementary Figs. 7 and 12, STRING criteria were set at “medium” stringency (0.4) and included all sources of interactions (text mining, experiments, databases, coexpression, neighborhood, gene fusion and co-occurrence). For the STR figure (Supplementary Fig. 8), the P value survives when stringency is increased to “highest” (0.9). All enrichment analyses were performed using a background list of genes that were significantly differentially expressed as RNA or protein. (This same list was used to generate the scatter plots.)

Immunoblotting. Lysates prepared in 8 M urea were separated by SDS-PAGE (4–20% Tris-glycine gels, Life Technologies). Proteins were transferred onto 0.2 µm nitrocellulose (Bio-Rad). Primary antibodies used were anti-miGluR2/3 (GRM2/3, EMD Millipore, 06–676, 1:1,000), anti-CB1 (CNR1, Cell Signaling, D5NSC, 1:1,000), anti-PDE4D (Millipore ABSS2, 1:1,000), anti-TrkC (NTRK3, Cell Signaling, C4AHS, 1:1,000 and anti-GAPDH (Calbiochem clone CO1, 1:5,000). The antibodies used have been validated for this assay in other species, but not in human brain. Comprehensive validation of antibodies for use in human brain tissue is extremely difficult, as discussed by the Antibody Validation Working group35, but given that for all antibodies except TrkC the antibodies showed a comparable pattern of expression to the LC-MS/MS data and bands of the appropriate size, we can presume these antibodies are working appropriately. In TrkC’s case, the disagreement may depend on the epitope used by the antibody, dropout of peptides due to post-translational modification or a lack of sensitivity by the antibodies for a relatively modest change. The primary antibodies were visualized using anti-mouse or anti-rabbit HRP (Vector Laboratories, PI-2000 Ms, 1:1,000 Rh 13,1000) and a secondary bio-Rad Odyssey anti-IRDye 800 (Bio-Rad) or in the case of GAPDH, the Licor IRDye 800 anti-mouse secondary antibody (Rockland 610-102-041, 1:15,000) and a Licor Odyssey Infra Red Scanner. Bands were quantified using ImageJ and normalized within lane to GAPDH, and pairwise Student’s t-tests were performed using Prism 7 (n = 5 biological replicates per group, all data shown on blots/graphs).

A summary of important reproducibility-related information from these methods can be found in the accompanying Life Sciences Reporting Summary.

Life Sciences Reporting Summary. Further information on experimental design and reagents is available in the Life Sciences Reporting Summary.

Data availability. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE9 partner repository with the data identifier PXD000443.

Code availability. Analysis code and the required input tables have been provided as a zipped Supplementary Software file. This large supplementary file (>600 MB unzipped) contains a readme file with system-specific instructions for running the code.

References
50. Harrow, J. et al. ENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 22, 1760–1774 (2012).
51. Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. 10, 1794–1805 (2011).
52. Cox, J. et al. Accurate mass spectrometry to gene ontology terms: a large-scale format for delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol. Cell. Proteomics 13, 2513–2526 (2014).
53. Habegger, L. et al. RSEQtools: a modular framework to analyze RNA-Seq data using compact, anonymized data summaries. Bioinformatics 27, 281–283 (2011).

54. Huber, W. et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods 12, 115–121 (2015).

55. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8, 118–127 (2007).

56. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 24, 719–720 (2008).

57. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing on JSTOR. J. R. Stat. Soc. B 57, 289–300 (1995).

58. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550 (2005).

59. Uhlen, M. et al. A proposal for validation of antibodies. Nat. Methods 13, 823–827 (2016).

60. Vizcaíno, J. A. et al. 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. 44 (D1), D447–D456 (2016).
Experimental design

1. Sample size
   Describe how sample size was determined.
   No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in Sharma et al’s study of the mouse brain (Nature Neuroscience 2015).

2. Data exclusions
   Describe any data exclusions.
   No data were excluded from the analysis.

3. Replication
   Describe whether the experimental findings were reliably reproduced.
   Of the proteins we tested by immunoblotting in the DFC vs V1C, 3 out of 4 proteins validated.

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups.
   Samples were run on the LC-MS/MS in regional blocks, with control samples interspersed throughout to allow for correction of batch effects.

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   Data collection and analysis were not performed blind to the conditions of the experiments.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters
   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

   n/a  | Confirmed
   ☒   | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
   ☒   | A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
   ☒   | A statement indicating how many times each experiment was replicated
   ☒   | The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
   ☒   | A description of any assumptions or corrections, such as an adjustment for multiple comparisons
   ☒   | The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
   ☒   | A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
   ☒   | Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.
Software

Describe the software used to analyze the data in this study.

- Maxquant v1.5.2.1 to map/quantify peptides & proteins
- ComBat to correct for batch effects in the proteomic data
- STAR aligner v2.4.2a to map RNAseq reads
- RSEQ tools to compute RNA abundance
- Custom R/Bioconductor code provided in accompanying zip file to analyse differential expression/enrichment/generate figures

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

All materials used (not including the samples themselves) are available through standard commercial vendors.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Primary antibodies used were anti mGluR2/3 (GRM2/3, EMD Millipore, 06-676, 1:1000), anti CB1 (CNR1, Cell Signalling, D5N5C, 1:1000), anti PDE4D (Millipore ABS22, 1:1000), anti TrkC (NTRK3, Cell Signalling, C44H5, 1:1000) and anti GAPDH (Calbiochem CB1001, 1:5000). They have been validated for this assay in other species, but not in human brain. The bands we observe are of equivalent size to those shown in validation data. Comprehensive validation of antibodies for use in human brain tissue is extremely difficult, as discussed by the Antibody Validation Working group (Uhlen et al, Nature Methods, Oct 2016), but given that for all antibodies except TrkC the antibodies show a comparable pattern of expression to the LC-MS/MS data, and bands of the appropriate size, we can presume these antibodies are working appropriately. In TrkC's case, the disagreement may depend on the epitope used by the antibody, drop out of peptides due to post translational modification, or a lack of sensitivity by the antibodies for a relatively modest change.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

N/A

b. Describe the method of cell line authentication used.

N/A
c. Report whether the cell lines were tested for mycoplasma contamination.

N/A
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

N/A

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

N/A
12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

All demographic information regarding the human samples used can be found in Table S1.