Published in final edited form as:

Nat Neurosci. 2014 November ; 17(11): 1491–1499. doi:10.1038/nn.3829.

Decoding neuroproteomics: integrating the genome, translome and functional anatomy

Robert R. Kitchen^{1,2}, Joel S. Rozowsky¹, Mark B. Gerstein^{1,3}, and Angus C. Nairn²

¹Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA

²Division of Molecular Psychiatry, Abraham Ribicoff Research Facilities, Connecticut Mental Health Center, Yale University School of Medicine, New Haven, CT 06508, USA

³Department of Computer Science, Yale University, New Haven, CT 06520, USA

Abstract

The immense inter- and intra-cellular heterogeneity of the central nervous system (CNS) presents major challenges for high-throughput *omic analyses. Transcriptional, translational, and post-translational regulatory events are localised to specific neuronal cell-types, or sub-cellular compartments, resulting in discrete patterns of protein expression and activity. A spatial and quantitative knowledge of the “neuroproteome” is therefore critical to understanding normal and pathological aspects of functional genomics and anatomy of the CNS. Improvements in mass-spectrometry allow profiling of proteins at sufficient depth to complement results from high-throughput genomic and transcriptomic assays. However, there are challenges in integrating proteomic data with other data modalities and even greater challenges obtaining comprehensive neuroproteomic data with cell-type specificity. Here we discuss how proteomics should be exploited to enhance high-throughput functional genomic analysis by tighter integration of data analyses. We also discuss experimental strategies to achieve finer cellular and sub-cellular resolution in transcriptomic and proteomic studies of neural tissues.

Introduction

The complexity of the human brain is reflected in ~86 billion neurons, and at least an equal number of glial cells, which can be further subdivided into hundreds of different types based on morphology, connectivity, and molecular and electrophysiological properties. These cells are integrated into functional networks within very precise constraints and deviations from the normal course of development can lead to a variety of CNS disorders. Work in converging fields has begun to disentangle this immense complexity^{1–3}, including mapping the complete set of neuronal connections, dubbed the ‘connectome’. Imaging tools such as fMRI, dMRI, and PET (see summary of abbreviations and terms) can non-invasively produce huge amounts of data on the wiring of the brain⁴. Thanks to advancements in automation, electron microscopy is an attractive method for tracing neuronal processes through thousands of perfectly stacked images⁵. Mapping and characterizing functional neuronal circuits using more invasive methods such as viral⁶ and optogenetic approaches⁷ can provide much finer resolution. Ultimately, however, the fate and regulation of networks of neurons is dictated by exquisite choreography of the biosynthesis, trafficking, and

interactions of the molecules of the CNS that lie well beyond the resolution and scale of these methods. Fluorescent labelling of RNAs, by in-situ hybridisation (ISH), or proteins, by immunohistochemistry (IHC), can provide spatial expression profiles of individual genes across the brain^{8,9}. Such atlases, while extremely important, are limited in throughput and information regarding cell types. Importantly, none of these methods provide quantitative measurement of abundances or the state of activity of molecules within specific cell types of the CNS.

There have been ambitious efforts to characterise RNA expression across the mammalian brain, from transcriptome-wide abundance in samples of mixed cell-types^{10–12} to the finer-scale localisation of the expression of individual genes^{8,13}. Each have reinforced findings of significant differences in the transcriptome between neuronal cell-types, brain regions, developmental stages, and species. However RNA is far from the only molecule worthy of such study. Thanks to the astounding pace of technological innovation there are now a wealth of assays available for querying the full gamut of cellular processes accessible via the measurement of nucleic acids including DNA structural variation¹⁴, DNA-protein interactions/modifications¹⁵, post-transcriptional modifications¹⁶, and post-transcriptional and -translational regulation^{17,18}. High profile, multi-investigator initiatives have produced much of these functional genomic data. ENCODE, the Roadmap Epigenome Project, and BrainSpan have characterised the multi-omic landscape of specific cell-types, tissues, and species, while 1000 Genomes, gEUVADIS, TCGA, and GTEx characterise the genome and transcriptome across a large set of individuals to better understand sequence variation throughout the human population.

While these initiatives lend significant insight into the incredible complexity of cellular regulatory processes and their dynamics under perturbation, a notable exception has been a comprehensive proteomic analysis of their sample cohorts. Tandem mass-spectrometry is the most popular technology for protein-level analyses, enabling unbiased identification and quantification across the entire proteome¹⁹. Mass-spectrometry has become very flexible, thanks to options for targeted analysis of selected peptides or proteins²⁰ and methods for chemically labelling samples to obtain greater protein quantification accuracy²¹. However, methodological advances in the analysis of amino acids have lagged compared to rapid innovations in nucleic acid sequencing. Our inability to in-vitro amplify amino acids - there is no PCR for protein - leads to more demanding technical requirements for protein detection. Despite this limitation, it is now possible to reliably obtain quantitative observations for tens of thousands of peptides derived from up to 10,000 proteins per sample²², and very recently important steps have been taken toward profiling the entire human proteome^{23,24}.

Given the critical insight into the biosynthetic state of the cell offered by proteins and their modifications, whole-proteome analysis is a necessity for investigators studying all aspects of complex organisms, especially the brain. Mass-spectrometry remains the only method of profiling the neuroproteome with a throughput and resolving power comparable to other functional genomic methods²⁵. Quantitative analysis of the thousands of proteins expressed in each of the likely hundreds of distinct neuronal cell-types is vital for deeper understanding of functional anatomy and the system-level functions of CNS cells.

Integration of proteomics with other functional genomic data is also necessary to better understand the dynamics of RNA and protein expression, the roles of post-translational modifications, and the trafficking of molecules within neurons and glia.

In this review we will first discuss the advantages of increased adoption of mass-spectrometry proteomics within the context of multi-omic approaches currently used to profile the molecules of the CNS. We will then offer a brief perspective on how the field can benefit from deeper integration of proteomic and functional genomic analysis. Finally, we will discuss approaches to achieving cell-type specificity in *omic analyses and how these can be improved to better serve the needs of the neuroscience community, especially with respect to interpreting results in the framework of the functional anatomy of the brain.

Biological insights from mass-spectrometry based proteomics

The various methodologies available for proteomic profiling have been comprehensively covered elsewhere, including excellent reviews on mass-spectrometry technologies^{19,26}, analysis²⁷, interpretation^{28,29}, and specific examples of proteomics applied to the CNS³⁰. The primary insights offered by mass-spectrometry proteomics are the sequences, post-translational modifications (PTMs), and abundances of up to tens of thousands of peptides and/or proteins in a given sample. The various methods available for spectra acquisition and quantification can be complemented by purification of protein from distinct sub-cellular compartments, which allows investigators to specifically and separately assay proteins located in the nucleus, cytosol, cytoskeleton, endoplasmic reticulum, and plasma membrane³¹. Increased sub-cellular resolution combined with mass-spectrometry can improve the dynamic range and sensitivity to low abundance or membrane-specific proteins that may otherwise be occluded by highly abundant nuclear and cytosolic proteins³². Extending beyond comparison of steady state protein expression, metabolic labelling of amenable cells (for example in culture) is capable of yielding valuable insight to the rates of protein turnover and has revealed that proteins are both more abundant and have a longer half-life than do RNAs³³.

Besides being the ultimate molecular ‘read-out’ of the genome, the proteome contains a wealth of information about the landscape of well over 90,000 sites of PTM that are simply inaccessible through the analysis of nucleic acids³⁴. Studies of dynamic cellular processes involving protein kinases and phosphatases, regulatory enzymes responsible for signal transduction, provide valuable insight into regulation of protein function³⁵. In addition to phosphorylation, PTMs such as acetylation, acylation, deamidation, glycosylation, methylation, and ubiquitination play pivotal roles in regulating almost all cellular processes including energy production and transport, DNA modification, transcription and translation of RNA, and RNA or protein stability³⁶. This information is obtained directly by mass-spectrometry due to the characteristic mass-shift that these modifications cause in the peptide spectra and, as such, provides not only the exact locations of these modifications but also measurements of their relative abundance. For example, mass spectrometry has proven to be a valuable tool for locating PTMs implicated in CNS development³⁷ and disorder³⁸, including interrogating specific hypotheses such as the p300-mediated acetylation of Tau³⁹ and the regulation of the transcriptional repressor LSD1 through phosphorylation⁴⁰. Peptide

sequence and PTM data can also be obtained from specific sub-cellular fractions to profile, for example, the fascinating landscape of histone modifications, including simultaneous analysis of the co-occurrence of various combinations of histone marks⁴¹. Further, proteomic profiling of the nucleus has enabled simultaneous quantification of the complete set of expressed transcription factors and has been vital in identifying novel regulatory proteins⁴².

Integration of proteomic, transcriptomic and genomic data

The wealth of information available from proteomics is complemented by insights offered by other functional genomic assays such as next generation RNA sequencing (RNA-seq). However, it is common for studies to focus on just one of these data modalities. There are major benefits to studying multiple layers of a system via the various functional *omic methods, but in order to obtain maximum utility the generated data need to be carefully integrated. For example, the integration of DNA variants and chromatin signals with transcriptomics has been extremely valuable for gaining a deeper intuition for genetic, epigenetic, and post-transcriptional regulation in the CNS^{43,44} (see also brainspan.org). Integration of transcriptomic and proteomic data has the potential to be just as powerful for monitoring the often-subtle effects or dysfunction of protein production and localisation underlying neurodevelopment and neuropathology. Most efforts to combine transcriptomic and proteomic profiling have compared or validated the expression of genes via the abundance of their RNAs and proteins, including characterisation of abundance across healthy tissues⁴⁵, changes due to neurological disorders⁴⁶, and localised expression in neuronal processes⁴⁷. An additional major attraction of proteomics lies in the peptide sequences, which enables proteome-wide validation of genomic and transcriptomic variants, allelic imbalance, and isoform identification²² (Figure 1). We will briefly discuss various levels of analysis where complementary proteomic data are required, suggest strategies whereby genomic and transcriptomic data can be better integrated, and highlight the potential advantages of such approaches in studies of the CNS.

Integration of expression measurements - An understanding of the “translatome” and “translational efficiency”

Several high-quality studies have employed mass-spectrometry to profile the protein output of tissues and organisms and have proven useful as standalone resources^{23,48–50}. Mass-spectrometry experiments have also been performed as validations for findings obtained by chromatin or RNA profiling⁵¹ and occasionally RNA-level data has been used as a ‘reverse-validation’ of results obtained by mass-spectrometry⁵². Although often presented as ‘integrated’, such validation-driven analyses tend to provide limited additional insight, instead serving as a secondary mechanism for controlling the rate of false discoveries in the original experiment; however there are some recent examples of deliberate multi-level *omic analysis in neuroscience⁵³. Given the tremendous volumes of data that can be produced via mass-spectrometry and the other functional genomic assays, now more than ever there exist significant analytical advantages to designing studies combining transcriptomic and proteomic data²².

Importantly, those studies that have attempted deliberate co-analysis of RNA and protein have consistently observed a limited correlation (between 0.5–0.7) in abundance. As such, measurements of mRNA abundance alone are insufficient to be considered predictive of protein abundance^{54–56}, reinforcing the requirement for cellular analysis at the proteome level. The cause for this poor correlation is likely the result of a combination of biological and technical factors⁵⁷; where biological factors include cellular heterogeneity, alternative splicing, post-transcriptional regulation, post-translational modifications, protein-turnover, and protein localisation. It is therefore essential for our deeper understanding of cellular mechanics, especially in highly heterogeneous tissues such as the CNS, to minimise technical variability such that samples subjected to multi-omic analysis are of high quality, are inherently comparable, and are processed with consistency.

A promising recent innovation allows direct profiling of the ‘translatome’, the set of actively translating RNAs engaged by the polyribosome, thus enabling analysis of RNA at a point closest to the production of protein⁵⁸ (see further discussion of so-called ribo-tag or TRAP methods and their utility in CNS below). A more detailed assessment of translational control⁵⁹ can be obtained by sequencing the very short fragments of RNA contained within the ribosome itself¹⁷. So-called ‘ribosome profiling’ allows, for the first time, a transcriptome-wide survey of the positions of ribosomes on the RNA transcript, including precise identification of open reading frames. When compared to the relative abundance of those same transcripts, the approach has formalised the concept of “translational efficiency”, a mechanism through which cells may have the ability to modify protein output whilst maintaining stable RNA abundance simply by increasing the density of ribosomes on selected transcripts⁶⁰. This has substantial consequences for the integration of proteomic and transcriptomic analyses and should, once the methodology matures, provide a more reliable predictor of protein abundance than simple RNA expression⁵⁹. Ribosome profiling of the CNS is currently in its infancy, due mainly to limitations arising from the amount and quality of input RNA. However an initial study of brain tumour material suggests for the first time that despite apparent overabundance of certain mRNAs in brain tumours, the translational efficiency of these genes is very low compared to non-tumour specific genes⁶¹. Ribosome profiling has also shown that long non-coding RNAs in the cytosol, many of which are known to be spliced, capped, and polyadenylated in a similar manner to mRNAs⁶², can be engaged by the polyribosome⁵⁶, but do not code for protein⁶³. In a somewhat related manner to ribosomal profiling, at the protein level it is possible to assay the production and degradation of polypeptides by either pulse-labelling SILAC proteomics⁶⁴ or the capture of nascent peptide chains as they are synthesised by the ribosome using biotin-puromycin labelling⁶⁵. However these approaches have rather limited application in studying intact organisms and tissues like brain due to the requirement of invasive chemical labelling.

Proteomic assessment of alternative splicing and isoform usage

Alternative splicing (AS) is a highly tissue-specific process¹⁸ that greatly increases the complexity of the potential set of RNA molecules produced from multi-exon genes. Splicing in the brain has been extensively profiled using transcriptomics and specific instances of alternative transcript usage have been implicated in neuropsychiatric disorders¹¹. During

CNS cell development, for example, AS in neuronal progenitor stem cells has revealed differential isoform usage at different stages of maturation⁶⁶. Extensive profiling of AS has characterised the expression of DISC1 isoforms that are associated with schizophrenia⁶⁷. Validation of splicing events by mass-spectrometry is feasible (see Figure 1), either in terms of determining the abundance of known isoforms in a given cell-type, or verifying the existence of novel exon junctions identified by RNA-sequencing⁶⁹, but has not yet been systematically applied to the CNS. Proteomics may be useful, for example, in determining whether the DISC1 translocation in schizophrenia encodes a protein product⁶⁸.

A recent result that has a potentially significant impact on interpretation of proteomic data is the finding that most genes, in a given cell-type, likely express only a single ‘dominant’ transcript^{70,71}. This has direct consequences for differential isoform usage between CNS cell types and for the progression of neurological disorders affecting protein-protein interactions, as exemplified in the Autism Spliceform Interaction Network⁷². Currently full-length transcript⁷³ and isoform⁷⁴ profiling technologies are immature⁷⁵ but have fantastic potential. For example, full-length RNA profiling in the brain has been exploited to directly observe the various isoforms of neurexin, showing their production mediates distinct protein interactions at the synapse⁷⁶.

Proteomic confirmation of genomic/transcriptomic variants

RNA-editing, also referred to as RNA-DNA differences (RDDs), appears to have played a significant role in human brain evolution^{16,77} and of the trio of ADAR proteins responsible for the post-transcriptional A-to-I modification, ADAR3 is exclusively expressed in the brain¹⁶. Several mis-sense RDDs in the AMPA receptor are edited at specific stages of human brain development⁷⁸ and are required for normal brain function and phenotype in mice⁷⁹. The serotonin receptor 5-HT_{2C}R contains numerous RDDs that alter both the expressed protein sequence⁸⁰ and cause an order of magnitude reduction in efficacy in the interaction of the receptor with its G proteins⁸¹. Unfortunately reliable identification of RDDs is technically challenging, which has led to false-positive identifications⁸². Mass-spectrometry is therefore an extremely attractive tool for unbiased validation of mis-sense RDDs, and for providing quantitation of the edited and unedited copies of the proteins.

Within a given individual the allele-specific expression (ASE) of RNA or protein can be the result of epigenetic regulatory processes that are common across species and tissues⁸³. Detection of ASE relies on the discovery of heterozygous genomic variants that allow monitoring imbalances in the abundance of RNA produced from each parental allele; with mis-sense (non-synonymous) variants enabling detection of ASE by mass-spectrometry⁸⁴. A recent survey of the mouse CNS revealed 1,300 ASE transcripts that favour the maternal allele during brain development, while in the adult the bias shifts toward the paternal allele⁸⁵. Mass-spectrometry has also been used to identify and assess levels of allele-specific transcription factor binding (ASB) and define allelic regions of the genome that are disease-associated⁸⁶. Such analyses are appealing due to potential cascade effects caused by ASE/ASB; differential expression or activity of an allelic transcription- or splicing-factor, for example, may have a substantial impact on downstream cellular processes.

Quantitative trait loci (QTL) have been extensively mapped to a variety of complex traits using transcriptomics in terms of the relationship between genomic variants and RNA expression (eQTL) and, more recently, variants have also been related to protein abundance (pQTL)⁸⁷. In the prefrontal cortex, for example, such variants reportedly affect the expression of more than 100 genes⁸⁸. An analysis of individuals genotyped in the HapMap project reported that almost two-thirds of 185 observed cis-acting pQTLs were not found in a complementary analysis of the RNA⁸⁹. An example pQTL that was not detected in RNA-level analysis lies in the 3'UTR of the *IMPA1* gene, a putative target for lithium⁹⁰, and may be responsible for the post-transcriptional regulation of the expression of the gene. In this way, mass-spectrometry proteomics has the power to meaningfully complement RNA analysis in the detection and functional characterisation of phenotype-associated genomic variants.

Finally, there has been a flurry of recent activity in identifying and ascribing potential functional roles to fusion transcripts and non-protein-coding regions of the genome. Efforts including ENCODE⁷¹ and others⁹¹, have reported that more than 75% of genomic DNA is at some point transcribed to primary (un-spliced) RNA. This 'pervasive transcription' has caused some controversy and confusion, not least when considering whether the production of these RNA molecules may imply they have a functional role in cellular processes. Transcriptomic analysis alone is therefore insufficient to accurately define the cohort of protein coding sequences, even when combined with ribosome profiling, and predictions must be refined using high-throughput proteome profiling.

Improving the compatibility of proteomics with functional genomics

Improvements to encourage integration of proteomics with other data fall into three main areas - annotation, archival, and methodology.

1. Annotation: consistent sequence annotation amongst the functional genomics

A basic, but significant limitation is that different gene annotations are often used in the analysis of RNA and protein. Adoption of a high-quality proteome annotation that is consistent with genomic databases such as ENSEMBL and UCSC would not only improve RNA/protein abundance comparisons, but also facilitate the integration of peptide/protein abundance to genome browsers.

2. Archival: improved protein sequence databases

High-quality integration depends upon equally high-quality data made available in well-curated and -maintained online repositories. Databases of peptide identifications such as PRIDE⁹² and Peptide Atlas⁹³ are valuable resources for mapping spectra based on previously identified peptides; however they are of limited use to non-experts who may simply desire higher-level information about proteins observed in a given disease state, tissue, or cellular compartment⁹⁴.

3. Archival: improved protein expression databases

The neuroscience community would benefit from a large-scale resource providing quantitative data on peptide or protein abundance to complement the wealth of

RNA expression data obtained from the human brain^{8,10}. Proteomic databases exist for plant species⁹⁵ and, more recently, for human tissues^{9,23} but public data on protein localisation and abundance across the brain are currently far less mature than their RNA equivalent. A combined resource containing peptide identification/quantification data and transcriptomics would be extremely valuable and may provide a framework with which to tackle open questions in proteomics such as alternative splicing and the non-uniform peptide coverage of proteins.

4. Methodological: More consistency in spectra data analysis between mass-spectrometry platforms

The Human Proteome Organisation (HUPO) Brain Proteome Project (BPP) recommend disseminating mass-spectrometry data in the form of lists of identified spectra and corresponding peptide sequences⁹⁶. However the software and choice of parameters for peak-picking, peptide identification, PTM identification, and quantification remains a significant source of information-loss and variability⁹⁷ and the use of software is often dependent on vendor. Dissemination of the raw spectra will be required for re-analysis of published data and to facilitate development of open-source analysis software.

5. Methodological: experimental and computational assessment of the translome

The primary goal for more effective integration of transcriptomics and proteomics should be to capitalise on experimental methodologies such as ribosome profiling, single molecule RNA-sequencing, and top-down mass-spectrometry profiling of intact proteins as well as computational methods that maintain an awareness of ‘upstream’ information during analysis of spectra. Utilising ribosome footprinting, for example, to identify coding sequences and translation initiation sites has been shown to produce moderately increased rates of peptide identification in the same samples⁹⁸. Additionally, there has been some recent interest in leveraging sequence variants obtained from RNA-seq to improve peptide identification^{99,100}, which have in turn resulted in the production of somewhat basic software tools for integration of such datasets¹⁰¹. A conceptual framework in which to achieve complete integration of proteomic and functional genomic analyses is offered in Box 1 and Figure 1.

BOX 1

The ideal integrated multi-omic analysis would exploit information garnered at each stage of the gene-expression process to improve the overall utility of results obtained from RNA and protein profiling (see Figure 1). Information regarding genomic variants in a given sample/individual can be used to create a ‘personalised’ genome for that individual. These homo- and heterozygous variants are incorporated in transcriptomic analysis by RNA-seq and enable the detection of allelic imbalance (ASE), in addition to information relating to alternative splicing (AS) and RNA-editing (RDD). This cascade of information in a given sample aids in the analysis of mass-spectrometry spectra, in which peptides may be identified that support or refute the presence of non-synonymous

AS, RDD, and ASE events. From the relative abundance of these peptides it is possible to compare isoform abundance and allelic imbalance with the values estimated at the RNA-level, allowing inferences to be made based on the proteomic data, such as the presence or absence of a particular protein isoform.

Challenges in assessing the proteome(s) of the CNS

The many high-quality proteomic studies of the CNS have been enumerated in previous reviews^{30,102,103}. Recent examples include profiling neural tissue in zebrafish¹⁰⁴, cultured neurons¹⁰⁵, and oligodendroglial cells¹⁰⁶. However, most neuroproteomic studies are inherently limited by their implicit interrogation of protein expression or modification in mixed populations of neuronal cell types. This results from the distinct transcriptional profiles of intermixed neuronal and glial cell types, as well as the short- and longer-term regulatory processes initiated by differential expression of neurotransmitter and neurotrophic factor receptors. Existing whole-tissue analyses¹⁰, even combined with careful manual dissection are not sensitive to inter- or intra-cellular variability and, as a result, functional genomic information is 'diluted' amongst different classes of intermixed neurons and the high proportion of glial cells. The key, therefore, to overcoming this cellular heterogeneity lies in experimental approaches to deconvolve the multitude of cell-types, by purifying or selecting a single cell-type, prior to functional genomic or proteomic analysis.

Current approaches to overcoming inter- and intra-cellular heterogeneity

When pondering the variety of options for profiling individual cell types in a complex tissue, there is a strong temptation to wish for technological and methodological developments to converge on the profiling of single cells. This level of analytical detail will be revolutionary in terms of the additional power to resolve changes in molecular abundance during, say, neuronal differentiation from progenitor cells. However there exist drawbacks to currently available methods of single cell analysis¹⁰⁷. An obvious limitation is that multi-modal profiling of the same single cell is currently impossible and, as such, regulatory relationships between the dynamics of RNA transcription, regulation, and translation would still need to be inferred, rather than directly observed. Additionally, the throughput and resolving power of single cell approaches are still too low to meaningfully assess the response of collections of neurons in a given brain region to experimental conditions such as the chemical effect of a drug, especially when there is a requirement for multiple biologically independent subjects. There is no doubt that with the continued development of sequencing and mass-spectrometry based methods for low sample input, new and exciting biology will emerge that advances our understanding of transcriptional and translational programmes both within the cell and in cell-to-cell signalling^{108,109}.

Similarly, the analysis of single or small numbers of neural cells, obtained for example by laser-capture microscopy (LCM), followed by transcriptional and proteomic profiling is an attractive approach as it enables molecular quantitation in a well-defined physical volume of tissue. However isolating neuronal cells by LCM is a time-consuming process and is not guaranteed to result in cell-type specificity due to cellular proximity and overlapping processes. In addition, current proteomic technologies require two to three orders of

magnitude more cells, compared to an equivalent analysis of RNA, to be collected by LCM in order to obtain sufficient material to begin to interrogate their proteome. A potential solution may be application of mass-cytometry to measure 10–100s of proteins in single cells at sub-cellular resolution, however low-throughput currently limits this technique¹¹⁰.

Fluorescence activated cell sorting (FACS) is a very flexible method for separating and purifying populations of cells, typically using panels of selective antibodies. Purification of murine microglia by FACS with an antibody against *P2ry12*, a glia-specific surface protein, has revealed over 100 genes that are enriched compared to neurons and other oligodendrocytes⁵³. FACS has also enabled analysis of murine synaptic proteins by mass-spectrometry proteomics following overexpression of a GFP-tagged *VGLUT1* transgene¹¹¹. Despite the success of this sorting for purification of more granular cells such as glia, these sorting methods suffer in a similar manner to LCM when dealing with mature neurons due to loss of their extensive axonal and dendritic processes. A novel variant of FACS, fluorescence activated nuclei sorting (FANS), has recently been used to bypass the incompatibility of neurons with whole-cell sorting. For example, antibody-labelling using the neuronal-specific splicing protein, *NeuN*, was used to purify neuronal nuclei from primary tissue to provide a quantitative comparison of the abundance of nuclear proteins with astrocytes and oligodendrocytes¹¹². Additionally, purification of striatal nuclei by *NEUN* labelling has led to interesting observations of adult human neurogenesis in the hippocampus¹¹³ and striatum¹¹⁴. The full complement of nuclear proteins, specifically histones and transcription factors, are thus accessible for neuron-specific mass-spectrometry analyses.

Beyond the nucleus, the sub-cellular localisation of RNAs and proteins is fundamental to neuronal development, function, and disease¹¹⁵. The abundance of proteins localised, for example, at the pre- and post-synapse, and reflected in the content of sub-cellular fractions like the post-synaptic density (PSD) is dependent on cell-type and brain region^{116–118}. Due to the various confounds introduced by the trafficking of proteins and RNAs and/or local translation at synapses (or potentially in axons)¹¹⁵, proteomic profiling of enriched synaptic fractions can provide greater sensitivity and specificity than transcriptomics to assessing molecular regulation of synaptic proteins. The application of density/gradient centrifugation for purification of other sub-cellular fractions such as the plasma membrane or lipid rafts, or of specific types of organelles like mitochondria or the Golgi, offers sub-cellular resolution³¹, but the familiar issue of intermixing of multiple cell-types remains a significant barrier to the utility of this approach in the CNS.

A molecular approach to high-throughput analysis with cell-type specificity

An extremely elegant means of obtaining quantitative expression measurements for a specific gene from a specific cell type is through the creation of libraries of bacterial artificial chromosomes containing fluorescent marker proteins downstream of target regulatory elements (enhancer/promoters) that lie upstream of the desired gene¹³. Such an approach, when collected for multiple regulatory elements, or when used in combination with ISH/IHC spatial expression profiles has been extremely successful in revealing cell-type specific promoter activity and RNA expression¹¹⁹. A complementary method for

obtaining all ribosome-associated RNA expressed in a given cell-type leads directly from this identification of cell-type specific promoters, in which overexpression of an epitope-tagged ribosomal protein (referred to as 'TRAP' or 'Ribo-tag' methods) under the control of one such cell-type-specific promoter allows purification by immunoprecipitation of the transcripts bound by the polyribosome in the desired cells⁵⁸. An added benefit of the TRAP approach due to the labelling of the ribosomal protein is the presence of strong immunoreactive signal in the nucleolus of target cell types. This enables FANS purification of target nuclei for assessment of DNA, nascent RNA transcription in the specific cell-type of interest, as well as proteomic analysis of histone modifications or transcription factor binding^{120,121}.

The primary advantage of the TRAP approach is the elimination of the need to purify cells, instead all cells are lysed and processed, regardless of cell-type, and specificity is achieved through immunoprecipitation. This has several clear benefits, including increased reproducibility between experiments, high yields of RNA for sequencing, and most importantly the fact that the method performs equally well on neurons as any other cell type. In addition, TRAP benefits integrative multi-modal analysis, in that the transcripts that are captured are actively engaged by the polyribosome and therefore it is reasonable to expect their abundance to be a better proxy to protein levels than total cellular RNA.

Related immunoaffinity-based methods may be applicable to other types of organelles from specific neuronal cell types, or of sub-cellular compartments of neurons, including synapses^{111,122,123}. The utility of these methods is exemplified in a study in which transgenic mice containing both FLAG-tagged and Myc-tagged DARPP-32 were used to study phosphorylation by protein kinase A in intermixed D1 receptor-containing direct pathway and D2 receptor-containing indirect pathway neurons, respectively, of the striatum¹²⁴. The ability to isolate DARPP-32 by immunoprecipitation of either epitope-tagged protein following drug treatment provides a completely internally consistent mechanism by which to assess the effect of the drug in these two cell types. Thanks to the vastly increased resolving power of this approach, it was revealed that following cocaine administration DARPP-32 phosphorylation is increased in D1-receptor neurons but almost equally decreased in D2-receptor neurons¹²⁴.

Despite these applications, the direct utility of immunoaffinity or TRAP methods for whole-proteome analysis is currently somewhat more limited than whole-genome or whole-transcriptome profiling. Perhaps one of the more immediate ways in which proteomics will benefit is via the creation of fully quantitative, cell-type specific transcriptomes for all amenable cells of the CNS. Moreover, RNA-seq TRAP datasets could be exploited to statistically deconvolute proteomic signals obtained from comparable tissues and, in a subset of the genes, may even prove predictive of protein abundance. Finally, TRAP purified RNAs are also directly compatible with the ribosome footprinting protocol, allowing cell-type specificity in determinations of translational efficiency.

The principal disadvantage of the TRAP methodologies for studies of human tissues is the required genetic modification of the ribosomal protein via the GFP tagged bac-transgene. For application to human neurobiology it may become possible to exploit studies in

genetically tractable rodent models to identify proteins at the plasma membrane of specific neuronal subtypes. Use of such cell-surface markers could then enable cell-type specific analyses of RNA and protein without the need for genetic modification and, combined with traditional methods of tissue dissection, may provide an excellent avenue to characterise and monitor RNA and protein expression, in tandem, from a single cell-type in specific regions of the brain. For these surface proteins it may also be possible to identify extra-cellular domains, receptors, and ion channels that display specific patterns of expression that could be future targets for small molecules or drugs. Such an approach would thus have the power to enable truly integrative, specific, and biologically meaningful analyses of the mammalian CNS.

A perspective on the future of proteomics in neuroscience

The ultimate goal of neuroproteomics is to identify and quantify the complete set of proteins in the CNS, understand their interactions and relationship to RNA abundance, identify PTMs, and define specific marker proteins, especially at the plasma membrane, that can discriminate between a large number of cell-types and allow for targeting by small molecules. Analysis of single cells and single cell-types may soon define sets of RNAs and proteins that are capable, either individually or in combination, of this discrimination and produce targets for ISH imaging and other functional anatomical annotation of the brain. This will provide the mechanism for relating the molecular profiles of specific neuronal cell types to their positions and connectivity within the developing brain. Cataloguing this connectivity will in turn provide the anatomical framework with which to interpret the internal dynamics of the as yet unknown number of neuronal cell types. Promising advances have already been made in model organisms such as the mouse, but this will need to move across species in order to understand how protein regulation is negatively affected by developmental disorders and the neuropathology of degenerative diseases and aging.

For the idealisation of cell-type specific proteome profiling to be realised we will need to see improvements not only in mass-spectrometry technology and reduced sample input requirements, but also in the analytical methods by which isoforms are identified, quantified, and integrated with existing functional genomics datasets and related spatiotemporal atlases. Multimodal functional genomic analyses can be substantially enhanced with deep integration of proteomic data. With the current resources being allocated to research into the human CNS via the Brain Initiative there is a fantastic opportunity for protein-level information to be obtained that significantly improves our understanding of brain development and functional anatomy. Modeling the circuitry of the brain can only ever be a realistic goal when there exists a comprehensive understanding of its constituent parts, including a much more precise definition of the large number of neuronal cells based on both their connectivity and their internal proteomes.

Acknowledgments

Supported by NIH grants DA018343 (ACN + MBG), DA10044 (ACN) and Dept. of the Army W81XWH-09-1-0434 (ACN). Support was also obtained from the State of Connecticut, Department of Mental Health and Addiction Services.

ABBREVIATIONS

Proteomic Methods

MS/MS	Tandem mass-spectrometry
Spectra	Ionised amino acid fragments, each with a distinct mass, charge, and intensity, are recorded by the detector within the mass-spectrometer
SILAC/M	Stable Isotope Labelling of Culture/Mammals

Other Methods

WGS	Second-generation, massively parallel, whole genome DNA-sequencing
RNA-seq	Second-generation, massively parallel RNA-sequencing
PCR	Polymerase Chain Reaction
f/dMRI	Functional/Diffusion Magnetic Resonance Imaging
PET	Positron Emission Tomography
ISH	In-Situ Hybridisation
IHC	Immunohistochemistry

Project Acronyms

ENCODE	Encyclopaedia of DNA Elements project
gEUVADIS	Genetic European Variation in Health and Disease project
TCGA	The Cancer Genome Atlas project
GTE_x	Genotype-Tissue Expression project

Other Abbreviations and Definitions

PTM	Post-translational modification; chemical modifications to amino acids
RDD	RNA-DNA differences; due to RNA-editing by ADAR
ASE/B	Allele specific expression/binding; an imbalance of TF binding and/or RNA expression near a heterozygous genomic variant
e/pQTL	Expression/protein quantitative trait loci; two or more loci that underlie a quantitative phenotype of varying degree

References

1. Pollock JD, Wu DY, Satterlee JS. Molecular neuroanatomy: a generation of progress. *Trends Neurosci.* 2014; 37:106–123. [PubMed: 24388609]
2. Medland SE, Jahanshad N, Neale BM, Thompson PM. Whole-genome analyses of whole-brain data: working within an expanded search space. *Nature Neuroscience.* 2014; 17:791–800. [PubMed: 24866045]
3. Arenkiel BR, Ehlers MD. Molecular genetics and imaging technologies for circuit-based neuroanatomy. *Nature.* 2009; 461:900–907. [PubMed: 19829369]

4. Van Essen DC, Ugurbil K. The future of the human connectome. *Neuroimage*. 2012; 62:1299–1310. [PubMed: 22245355]
5. Lichtman JW, Livet J, Sanes JR. A technicolour approach to the connectome. *Nat Rev Neurosci*. 2008; 9:417–422. [PubMed: 18446160]
6. Ekstrand MI, et al. Molecular profiling of neurons based on connectivity. *Cell*. 2014; 157:1230–1242. [PubMed: 24855954]
7. Deisseroth K, et al. Next-generation optical technologies for illuminating genetically targeted brain circuits. *Journal of Neuroscience*. 2006; 26:10380–10386. [PubMed: 17035522]
8. Sunkin SM, et al. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res*. 2013; 41:D996–D1008. [PubMed: 23193282]
9. Uhlen M, et al. Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol*. 2010; 28:1248–1250. [PubMed: 21139605]
10. Kang HJ, et al. Spatio-temporal transcriptome of the human brain. *Nature*. 2011; 478:483–489. [PubMed: 22031440]
11. Johnson MB, et al. Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron*. 2009; 62:494–509. [PubMed: 19477152]
12. Hawrylycz MJ, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. 2012; 489:391–399. [PubMed: 22996553]
13. Heintz N. Gene expression nervous system atlas (GENSAT). *Nature Neuroscience*. 2004; 7:483. [PubMed: 15114362]
14. Korbelt JO, et al. Paired-end mapping reveals extensive structural variation in the human genome. *Science*. 2007; 318:420–426. [PubMed: 17901297]
15. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*. 2009; 10:669–680. [PubMed: 19736561]
16. Li JB, Church GM. Deciphering the functions and regulation of brain-enriched A-to-I RNA editing. *Nature Publishing Group*. 2013; 16:1518–1522.
17. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*. 2009; 324:218–223. [PubMed: 19213877]
18. Licatalosi DD, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*. 2008; 456:464–469. [PubMed: 18978773]
19. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003; 422:198–207. [PubMed: 12634793]
20. Stergachis AB, MacLean B, Lee K, Stamatoyannopoulos JA, MacCoss MJ. Rapid empirical discovery of optimal peptides for targeted proteomics. *Nat Meth*. 2011; 8:1041–1043.
21. Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol*. 2005; 1:252–262. [PubMed: 16408053]
22. Vidal M, et al. The human proteome - a scientific opportunity for transforming diagnostics, therapeutics, and healthcare. *Clin Proteomics*. 2012; 9:6. [PubMed: 22583803]
23. Wilhelm M, et al. Mass-spectrometry-based draft of the human proteome. *Nature*. 2014; 509:582–587. [PubMed: 24870543]
24. Kim MS, et al. A draft map of the human proteome. *Nature*. 2014; 509:575–581. [PubMed: 24870542]
25. Cox J, Mann M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem*. 2011; 80:273–299. [PubMed: 21548781]
26. Altelaar AFM, Munoz J, Heck AJR. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet*. 2013; 14:35–48. [PubMed: 23207911]
27. Nesvizhskii AI, Vitek O, Aebersold R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Meth*. 2007; 4:787–797.
28. Ahrens CH, Brunner E, Qeli E, Basler K, Aebersold R. Generating and navigating proteome maps using mass spectrometry. *Nature Publishing Group*. 2010; 11:789–801.
29. Bensimon A, Heck AJR, Aebersold R. Mass spectrometry-based proteomics and network biology. *Annu Rev Biochem*. 2012; 81:379–405. [PubMed: 22439968]

30. Craft GE, Chen A, Nairn AC. Recent advances in quantitative neuroproteomics. *Methods*. 2013; 61:186–218. [PubMed: 23623823]
31. Cox B, Emili A. Tissue subcellular fractionation and protein extraction for use in mass-spectrometry-based proteomics. *Nature Protocols*. 2006; 1:1872–1878. [PubMed: 17487171]
32. Boisvert FM, Lam YW, Lamont D, Lamond AI. A quantitative proteomics analysis of subcellular proteome localization and changes induced by DNA damage. *Mol Cell Proteomics*. 2010; 9:457–470. [PubMed: 20026476]
33. Schwanhäusser B, et al. Global quantification of mammalian gene expression control. *Nature*. 2011; 473:337–342. [PubMed: 21593866]
34. Naegle KM, et al. PTMScout, a Web resource for analysis of high throughput post-translational proteomics studies. *Mol Cell Proteomics*. 2010; 9:2558–2570. [PubMed: 20631208]
35. Oppermann FS, et al. Large-scale proteomics analysis of the human kinome. *Mol Cell Proteomics*. 2009; 8:1751–1764. [PubMed: 19369195]
36. Deribe YL, Pawson T, Dikic I. Post-translational modifications in signal integration. *Nature Structural & Molecular Biology*. 2010; 17:666–672.
37. Edwards AVG, Edwards GJ, Schwämmle V, Saxtorph H, Larsen MR. Spatial and temporal effects in protein post-translational modification distributions in the developing mouse brain. *J Proteome Res*. 2014; 13:260–267. [PubMed: 23947802]
38. Seyfried NT, et al. Quantitative analysis of the detergent-insoluble brain proteome in frontotemporal lobar degeneration using SILAC internal standards. *J Proteome Res*. 2012; 11:2721–2738. [PubMed: 22416763]
39. Min SW, et al. Acetylation of tau inhibits its degradation and contributes to tauopathy. *Neuron*. 2010; 67:953–966. [PubMed: 20869593]
40. Toffolo E, et al. Phosphorylation of neuronal Lysine-Specific Demethylase 1LSD1/KDM1A impairs transcriptional repression by regulating interaction with CoREST and histone deacetylases HDAC1/2. *J Neurochem*. 2014; 128:603–616. [PubMed: 24111946]
41. Sridharan R, et al. Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1 γ in reprogramming to pluripotency. *Nat Cell Biol*. 2013; 15:872–882. [PubMed: 23748610]
42. Mirzaei H, et al. Systematic measurement of transcription factor-DNA interactions by targeted mass spectrometry identifies candidate gene regulatory proteins. *Proceedings of the National Academy of Sciences*. 2013; 110:3645–3650.
43. Visel A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*. 2009; 457:854–858. [PubMed: 19212405]
44. Zou F, et al. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet*. 2012; 8:e1002707. [PubMed: 22685416]
45. Kislinger T, et al. Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell*. 2006; 125:173–186. [PubMed: 16615898]
46. Shawahna R, et al. Transcriptomic and quantitative proteomic analysis of transporters and drug metabolizing enzymes in freshly isolated human brain microvessels. *Mol Pharm*. 2011; 8:1332–1341. [PubMed: 21707071]
47. Elvira G, et al. Characterization of an RNA granule from developing brain. *Mol Cell Proteomics*. 2006; 5:635–651. [PubMed: 16352523]
48. Moghaddas Gholami A, et al. Global proteome analysis of the NCI-60 cell line panel. *Cell Rep*. 2013; 4:609–620. [PubMed: 23933261]
49. Geiger T, Cox J, Ostasiewicz P, Wisniewski JR, Mann M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat Meth*. 2010; 7:383–385.
50. Yu LR, et al. Global analysis of the cortical neuron proteome. *Mol Cell Proteomics*. 2004; 3:896–907. [PubMed: 15231876]
51. Chen R, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell*. 2012; 148:1293–1307. [PubMed: 22424236]
52. Phanitiel DH, et al. Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat Meth*. 2011; 8:821–827.

53. Butovsky O, et al. Identification of a unique TGF- β -dependent molecular and functional signature in microglia. *Nature Neuroscience*. 2014; 17:131–143. [PubMed: 24316888]
54. Greenbaum D, Colangelo C, Williams K, Gerstein M. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol*. 2003; 4:117. [PubMed: 12952525]
55. Lundberg E, et al. Defining the transcriptome and proteome in three functionally different human cell lines. *Mol Syst Biol*. 2010; 6:450. [PubMed: 21179022]
56. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*. 2011; 147:789–802. [PubMed: 22056041]
57. Li JJ, Bickel PJ, Biggin MD. System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ*. 2014; 2:e270. [PubMed: 24688849]
58. Heiman M, et al. A translational profiling approach for the molecular characterization of CNS cell types. *Cell*. 2008; 135:738–748. [PubMed: 19013281]
59. Kong J, Lasko P. Translational control in cellular and developmental processes. *Nat Rev Genet*. 2012; 13:383–394. [PubMed: 22568971]
60. Brar GA, et al. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science*. 2012; 335:552–557. [PubMed: 22194413]
61. Gonzalez C, et al. Ribosome profiling reveals a cell-type-specific translational landscape in brain tumors. *Journal of Neuroscience*. 2014; 34:10924–10936. [PubMed: 25122893]
62. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet*. 2009; 10:155–159. [PubMed: 19188922]
63. Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell*. 2013; 154:240–251. [PubMed: 23810193]
64. Boisvert F-M, et al. A quantitative spatial proteomics analysis of proteome turnover in human cells. *Mol Cell Proteomics*. 2012; 11 M111.011429.
65. Aviner R, Geiger T, Elroy-Stein O. Genome-wide identification and quantification of protein synthesis in cultured cells and whole tissues by puromycin-associated nascent chain proteomics (PUNCH-P). *Nature Protocols*. 2014; 9:751–760. [PubMed: 24603934]
66. Wu JQ, et al. Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by short, long, and paired-end sequencing. *Proceedings of the National Academy of Sciences*. 2010; 107:5254–5259.
67. Nakata K, et al. DISC1 splice variants are upregulated in schizophrenia and associated with risk polymorphisms. *Proceedings of the National Academy of Sciences*. 2009; 106:15873–15878.
68. Soares DC, Carlyle BC, Bradshaw NJ, Porteous DJ. DISC1: Structure, Function, and Therapeutic Potential for Major Mental Illness. *ACS Chem Neurosci*. 2011; 2:609–632. [PubMed: 22116789]
69. Sheynkman GM, Shortreed MR, Frey BL, Smith LM. Discovery and mass spectrometric analysis of novel splice-junction peptides using RNA-Seq. *Mol Cell Proteomics*. 2013; 12:2341–2353. [PubMed: 23629695]
70. González-Porta M, Frankish A, Rung J, Harrow J, Brazma A. Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. *Genome Biol*. 2013; 14:R70. [PubMed: 23815980]
71. Djebali S, et al. Landscape of transcription in human cells. *Nature*. 2012; 489:101–108. [PubMed: 22955620]
72. Corominas R, et al. Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. *Nat Commun*. 2014; 5:3650. [PubMed: 24722188]
73. Sharon D, Tilgner H, Grubert F, Snyder M. A single-molecule long-read survey of the human transcriptome. *Nat Biotechnol*. 2013; 31:1009–1014. [PubMed: 24108091]
74. Tran JC, et al. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*. 2011; 480:254–258. [PubMed: 22037311]
75. Engström PG, et al. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat Meth*. 2013; 10:1185–1191.

76. Treutlein B, Gokce O, Quake SR, Südhof TC. Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. *Proceedings of the National Academy of Sciences*. 2014; 111:E1291–9.
77. Sakurai M, et al. A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. *Genome Res*. 2014; 24:522–534. [PubMed: 24407955]
78. Lomeli H, et al. Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science*. 1994; 266:1709–1713. [PubMed: 7992055]
79. Higuchi M, et al. Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature*. 2000; 406:78–81. [PubMed: 10894545]
80. Kawahara Y, et al. Dysregulated editing of serotonin 2C receptor mRNAs results in energy dissipation and loss of fat mass. *Journal of Neuroscience*. 2008; 28:12834–12844. [PubMed: 19036977]
81. Burns CM, et al. Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature*. 1997; 387:303–308. [PubMed: 9153397]
82. Pickrell JK, Gilad Y, Pritchard JK. Comment on “Widespread RNA and DNA sequence differences in the human transcriptome”. *Science*. 2012; 335:1302. author reply 1302. [PubMed: 22422963]
83. Buckland PR. Allele-specific gene expression differences in humans. *Hum Mol Genet*. 2004; 13(Spec No 2):R255–60. [PubMed: 15358732]
84. Khan Z, et al. Quantitative measurement of allele-specific protein expression in a diploid yeast hybrid by LC-MS. *Mol Syst Biol*. 2012; 8:602. [PubMed: 22893000]
85. Gregg C, et al. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science*. 2010; 329:643–648. [PubMed: 20616232]
86. Butter F, et al. Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. *PLoS Genet*. 2012; 8:e1002982. [PubMed: 23028375]
87. Holdt LM, et al. Quantitative trait loci mapping of the mouse plasma proteome (pQTL). *Genetics*. 2013; 193:601–608. [PubMed: 23172855]
88. Liu C, et al. Whole-genome association mapping of gene expression in the human prefrontal cortex. *Mol Psychiatry*. 2010; 15:779–784. [PubMed: 20351726]
89. Wu L, et al. Variation and genetic control of protein abundance in humans. *Nature*. 2013; 499:79–82. [PubMed: 23676674]
90. Agam G, et al. Knockout mice in understanding the mechanism of action of lithium. *Biochem Soc Trans*. 2009; 37:1121–1125. [PubMed: 19754464]
91. Hangauer MJ, Vaughn IW, McManus MT. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet*. 2013; 9:e1003569. [PubMed: 23818866]
92. Martens L, et al. PRIDE: the proteomics identifications database. *Proteomics*. 2005; 5:3537–3545. [PubMed: 16041671]
93. Desiere F, et al. The PeptideAtlas project. *Nucleic Acids Res*. 2006; 34:D655–8. [PubMed: 16381952]
94. Ahmad Y, Lamond AI. A perspective on proteomics in cell biology. *Trends Cell Biol*. 2014; 24:257–264. [PubMed: 24284280]
95. Sun Q, et al. PPDB, the Plant Proteomics Database at Cornell. *Nucleic Acids Res*. 2009; 37:D969–74. [PubMed: 18832363]
96. Martens L, et al. A comparison of the HUPO Brain Proteome Project pilot with other proteomics studies. *Proteomics*. 2006; 6:5076–5086. [PubMed: 16912975]
97. Bell AW, et al. A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. *Nat Meth*. 2009; 6:423–430.
98. Menschaert G, et al. Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events. *Mol Cell Proteomics*. 2013; 12:1780–1790. [PubMed: 23429522]

99. Lopez-Casado G, et al. Enabling proteomic studies with RNA-Seq: The proteome of tomato pollen as a test case. *Proteomics*. 2012; 12:761–774. [PubMed: 22539427]
100. Wang X, et al. Protein identification using customized protein sequence databases derived from RNA-Seq data. *J Proteome Res*. 2012; 11:1009–1017. [PubMed: 22103967]
101. Wang X, Zhang B. customProDB: an R package to generate customized protein databases from RNA-Seq data for proteomics search. *Bioinformatics*. 2013; 29:3235–3237. [PubMed: 24058055]
102. Bayés A, Grant SGN. Neuroproteomics: understanding the molecular organization and complexity of the brain. *Nat Rev Neurosci*. 2009; 10:635–646. [PubMed: 19693028]
103. Sun F, Cavalli V. Neuroproteomics approaches to decipher neuronal regeneration and degeneration. *Mol Cell Proteomics*. 2010; 9:963–975. [PubMed: 20019051]
104. Gebriel M, et al. Zebrafish brain proteomics reveals central proteins involved in neurodegeneration. *J Neurosci Res*. 2014; 92:104–115. [PubMed: 24123299]
105. Seo JW, Kim Y, Hur J, Park KS, Cho YW. Proteomic analysis of primary cultured rat cortical neurons in chemical ischemia. *Neurochem Res*. 2013; 38:1648–1660. [PubMed: 23670089]
106. Liu X, et al. Proteomics reveal energy metabolism and mitogen-activated protein kinase signal transduction perturbation in human Borna disease virus Hu-H1-infected oligodendroglial cells. *Neuroscience*. 2014; 268C:284–296. [PubMed: 24637096]
107. Macaulay IC, Voet T. Single cell genomics: advances and future perspectives. *PLoS Genet*. 2014; 10:e1004126. [PubMed: 24497842]
108. Romanova EV, Aerts JT, Croushore CA, Sweedler JV. Small-volume analysis of cell-cell signaling molecules in the brain. *Neuropsychopharmacology*. 2014; 39:50–64. [PubMed: 23748227]
109. Liu X, et al. Molecular imaging of drug transit through the blood-brain barrier with MALDI mass spectrometry imaging. *Sci Rep*. 2013; 3:2859. [PubMed: 24091529]
110. Giesen C, et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat Meth*. 2014; 11:417–422.
111. Biesemann C, et al. Proteomic screening of glutamatergic mouse brain synaptosomes isolated by fluorescence activated sorting. *EMBO J*. 2014; 33:157–170. [PubMed: 24413018]
112. Dammer EB, et al. Neuron enriched nuclear proteome isolated from human brain. *J Proteome Res*. 2013; 12:3193–3206. [PubMed: 23768213]
113. Spalding KL, et al. Dynamics of Hippocampal Neurogenesis in Adult Humans. *Cell*. 2013; 153:1219–1227. [PubMed: 23746839]
114. Ernst A, et al. Neurogenesis in the Striatum of the Adult Human Brain. *Cell*. 2014; 156:1072–1083. [PubMed: 24561062]
115. Holt CE, Schuman EM. The central dogma decentralized: new perspectives on RNA function and local translation in neurons. *Neuron*. 2013; 80:648–657. [PubMed: 24183017]
116. Kavalali ET, Jorgensen EM. Visualizing presynaptic function. *Nature Publishing Group*. 2014; 17:10–16.
117. Cheng D, et al. Relative and absolute quantification of postsynaptic density proteome isolated from rat forebrain and cerebellum. *Mol Cell Proteomics*. 2006; 5:1158–1170. [PubMed: 16507876]
118. O'Rourke NA, Weiler NC, Micheva KD, Smith SJ. Deep molecular diversity of mammalian synapses: why it matters and how to measure it. *Nat Rev Neurosci*. 2012; 13:365–379. [PubMed: 22573027]
119. Portales-Casamar E, et al. A regulatory toolbox of MiniPromoters to drive selective expression in the brain. *Proceedings of the National Academy of Sciences*. 2010; 107:16589–16594.
120. Jordi E, et al. Differential effects of cocaine on histone posttranslational modifications in identified populations of striatal neurons. *Proceedings of the National Academy of Sciences*. 2013; 110:9511–9516.
121. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*. 2009; 324:929–930. [PubMed: 19372393]

122. Selimi F, Cristea IM, Heller E, Chait BT, Heintz N. Proteomic studies of a single CNS synapse type: the parallel fiber/purkinje cell synapse. *PLoS Biol.* 2009; 7:e83. [PubMed: 19402746]
123. Fernández E, et al. Targeted tandem affinity purification of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins. *Mol Syst Biol.* 2009; 5:269. [PubMed: 19455133]
124. Bateup HS, et al. Cell type-specific regulation of DARPP-32 phosphorylation by psychostimulant and antipsychotic drugs. *Nature Publishing Group.* 2008; 11:932–939.

Figure 1. Integrated genomic, transcriptomic and proteomic analyses and the central dogma Illustrated are the three canonical processes of the central dogma (**a–c**), the common methods by which genome-wide data are obtained, and the information provided by these data. (**a**) Whole-genome or exome sequencing provides genomic variant information. An example multi-exon gene is annotated with two isoforms, differentiated by a skipped penultimate exon, and contains several non-synonymous mutations identified by whole genome sequencing (WGS), exome sequencing, or genotyping microarray (SNPchip). Homo- and heterozygous missense mutations each cause a change in the codon at their respective positions, and a nonsense mutation that leads to a premature stop codon in the penultimate exon of the first isoform. These mutations are reflected in the maternal and paternal transcripts produced for each isoform (**b**) and are detected in the reads obtained from RNA-sequencing, which also detects an RNA-edit site in all transcripts. RNA-seq quantitation reveals that the second isoform is predominantly expressed, based on the subset of reads (red) that are able to distinguish the first from the second isoform. Also evident from the quantitation is that the maternal transcript of isoform 2 is primarily expressed (allele-specific expression; ASE), based on a different subset of reads that span the heterozygous variant. Whole-proteome analysis by mass-spectrometry (**c**) produces spectra that can be matched against a database derived from the maternal and paternal transcripts and produce peptides that span the positions of the homo- and heterozygous variants. The abundances of these peptides support the RDD and the allelic bias observed in RNA-seq. The spectra also suggest the presence of a post-translational modification and the absence of peptides toward the C-terminus of the first isoform lead for example to the inference that the premature stop potentially leads to the ubiquitination and destruction of this protein.