Decoding neuroproteomics: integrating the translatome with the connectome

Rob 5/5/14 7:49 PM
Deleted: Decoding neuroproteomics: integrating the proteome with the transcriptome and genome ... [1]

Robert R. Kitchen[1,2], Joel S. Rozowsky[1], Mark B. Gerstein[1,3], Angus C. Nairn[2]

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

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

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

Abstract

Whole-proteome analyses have the potential to be a powerful complement to existing or proposed genomic, epigenomic, and transcriptomic studies. Technological developments in tandem mass-spectrometry now have the potential to allow investigators to profile peptides and proteins at a sufficiently high resolution and coverage to meaningfully complement, even in complex mammalian systems, results obtained from high-throughput transcriptomic studies. In this review we discuss this state of the art in mass-spectrometry proteomics, highlight recent large-scale efforts to quantify the proteome of mammalian systems, and cover attempts to integrate proteomic data with functional genomic data for a more holistic approach to measuring gene expression. We discuss ways in which proteomic data and analysis can be made more compatible with the other high-throughput *omics data, both from more meaningful processing of the peptide data themselves to improved downstream integration in gene-expression and variation network analyses. Finally, we address issues regarding profiling the proteome of the central nervous system, paying specific attention to the immense inter- and intra-cellular heterogeneity in the mammalian brain. This heterogeneity drives the requirement for better integration of protein measurement with functional genomics and imaging and we discuss methodologies being employed to achieve finer resolution in the high-throughput *omics data obtained from neural tissues.

Rob 5/5/14 12:41 PM
Deleted:

Introduction

Since the completion of the human genome sequencing project there has been a huge amount of community effort devoted to the functional characterisation of the genome, from its structure to its molecular products. Thanks to the astounding pace of technological and methodological

innovation there are a wealth of assays available for querying the full gamut of processes accessible via the measurement of nucleic acids including 3-dimensional DNA conformation and interactions and structural variation[1], DNA-protein interactions and modifications[2,3], RNA transcription[4], post-transcriptional modifications[5], and post-transcriptional and -translational regulation[6-8]. High-profile, multi-investigator efforts have recently produced much of these genomic data either, in the case of ENCODE, Roadmap Epigenome Project, and BrainSpan, to characterise the multi-omic landscape of specific cell-types, tissues, and species or, in the case of 1000 Genomes, gEUVADIS, TCGA, and GTEx to characterise genome and transcriptome across a large set of individuals to better understand their variation in the human population and disease. There have been a number of ambitious initiatives to characterize RNA expression[9,10] and localisation[11,12] in the central nervous system and across the regions of the brain; each reinforcing observations of significant differences in gene expression between neuronal cell-types, brain regions, developmental stages, and species.

While these research efforts have lent significant insight into the incredible complexity of cellular regulatory processes and their dynamics under perturbation or disease, a notable exception has been tandem mass-spectrometry-based whole-proteome analyses (MS/MS). There are several reasons why advancements in proteome analyses have lagged compared to the other *omics. Notably it is our inability to in-vitro amplify amino acids- there is no PCR for protein -leads to more demanding requirements on the technology used for detection; although innovations in amino-acid 'sequencing' appear promising[13]. Despite this limitation, with recent technological advances it is now possible to reliably obtain quantitative observations of tens of thousands of peptides derived from up to 10,000 proteins[14]. Given the wealth of additional insight into the biosynthetic state of the cell offered by MS/MS, whole-proteome analysis is an increasingly attractive option for investigators, even those studying complex organisms.

Work in several fields has begun to deliver some results in disentangling the immense complexity of the neuronal circuitry of the mammalian brain[15]. Mapping the complete set of neuronal connections, dubbed the 'connectome', is an area of extremely active research, in which imaging tools such as fMRI, dMRI, and PET are being employed to non-invasively produce huge amounts of data[16] on the wiring of the brain at different resolutions and under different conditions[17]. Thanks to advancements in automation, electron microscopy is an increasingly attractive method for tracing neuronal processes through thousands of perfectly stacked images to trace the multitude of connections within the brain[18,19]. Further, in-situ hybridisation (ISH) and immunohistochemistry (IHC) has been used to create detailed maps of the spatial expression profiles of individual genes across the human brain[11,20]. MS/MS based

Rob 5/5/14 8:06 PM
Deleted: epigenome

Rob 5/5/14 7:55 PM
Deleted: roadmap

Rob 5/5/14 7:50 PM
Deleted: [ref, ref]

Rob 5/5/14 11:57 AM
Deleted:

Rob 6/5/14 4:50 PM
Deleted: MALDI

imaging methods allow the capture and profiling of peptides and proteins from the same brain tissue slices stained by IHC[21].

Despite advances in these imaging- and probe-based methods, MS/MS remains the only method of profiling the protein complement of the tissues of the brain with a throughput and resolving power comparable to other functional genomic methods. This is important to further unravel the normal and abnormal system-level function of CNS cells, by better understanding the relationship between RNA and protein expression, the roles of post-translational modifications, and the localisation of proteins, especially in the likely 100's of distinct neuronal cell-types each with specific transcriptome/proteome profiles. However, due to the enormous throughput of transcriptome and genome sequencing, MS/MS analyses must be used in concert with these other methodologies in order to extract the maximum information and utility from the proteome, particularly in complex tissue such as the CNS.

Biological insights from mass-spectrometry based proteomics

The standard insight offered by MS/MS is one of assaying peptide- or protein-level abundances, an overview of which is available in **Box 1** and further illustrated in **Figure 1**. These methods of spectra acquisition and quantification are complemented by the various options available for the 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 membrane[22]. This increased resolution available with MS/MS can improve sensitivity to low abundance or membrane-specific proteins that may otherwise be occluded by highly abundant nuclear and cytosolic proteins[23]. Extending beyond comparison of steady state protein expression, metabolic labelling of amenable cells (for example cell cultures or yeast) 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 RNAs[24].

Besides being the ultimate 'read-out' of the abundance of the molecular products of a genome, the proteome contains a wealth of information about the landscape of well over 90,000 sites[25] of **post-translational modifications** (PTMs) that are simply inaccessible through the analysis of nucleic acids[26]. Studies of dynamic cellular processes involving protein kinases and phosphatases, regulatory enzymes responsible for signal transduction, provide valuable insight into regulation of protein function [27]. In addition to phosphorylation, PTMs such as acetylation,

Rob 6/5/14 4:51 PM
Deleted: technology is capable of allowing

Rob 6/5/14 4:51 PM
Deleted: to MS/MS to more deeply profile the protein contents of precise brain regions and tissues

Rob 5/5/14 7:56 PM
Deleted: However, d

Rob 6/5/14 4:52 PM
Deleted: antibody

Rob 6/5/14 4:52 PM
Deleted: that

Rob 5/5/14 7:57 PM
Deleted: The purpose of this review is to summarise the potential benefits of wide adoption of MS/MS proteomics, outline several methodological improvements that might facilitate the integration of proteomic data with other functional genomic analyses, and discuss specific considerations with respect to proteome profiling of the CNS. .

Rob 6/5/14 4:54 PM
Deleted: This is similar to the nucleus- and cytoplasm-specific preps available for purifying RNA, but t

Rob 6/5/14 4:54 PM
Deleted: e

Rob 6/5/14 4:55 PM
Deleted: . For example, the protein composition of the nucleolus, nucleus, and cytoplasm has been found to be quite exclusive, with only a small number of proteins equally abundant between each compartment

Rob 5/5/14 8:00 PM
Deleted: .

Rob 5/5/14 8:00 PM
Deleted: Similar studies of dynamic cellular processes involving protein kinases and phosphatases, regulatory enzymes responsible for signal transduction, and their sites of **phosphorylation** are exclusively accessible by MS/MS proteomics[26] and provide valuable insight into regulation of protein function in health and disease[27].

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 stability[28]. This information is obtained directly using MS/MS due to the characteristic mass-shift that they cause in the peptide spectra[29] and, as such, provides not only the exact locations of these modifications but also quantitative measurements of their relative abundance. These 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 the simultaneous analysis of the abundance and co-occurrence of various combinations of marks, in the nucleus and their changes during ES cell differentiation[30]. As a further example, proteomic profiling of the nucleus has enabled simultaneous quantification of the complete set of expressed **transcription factors**, and, despite the obvious lack of information regarding the genomic binding locations of these nuclear proteins, proteomic profiling has been vital in identifying novel regulatory proteins[31].

Integration of proteomic and transcriptomic data

Integration of expression measurements

Several high-quality studies have employed MS/MS to profile the protein output of tissues and organisms under a variety of conditions and have proven useful as standalone resources[32-34]. However a large number of MS/MS experiments are performed as validations for findings obtained by chromatin or RNA profiling[35], similarly RNA-level data has also been used as a 'reverse-validation' of the results obtained by MS/MS[36]. Unfortunately such attempts at presenting a so-called 'integrated' analysis often appear haphazard and provide limited utility in terms of the additional information these validation data provide.

Several instances in which integrated analysis of RNA and protein were attempted have observed a **limited correlation** (between 50-70%) **in molecular abundances** between these levels[37,38], certainly not strong enough for measurements of mRNA abundance alone to be considered predictive of protein abundance[38-40] and is true regardless of the technology or experimental method used to profile the RNA or the protein. The cause for this poor correlation is more than likely a combination of biological and technical factors. General biological variables include cellular heterogeneity, alternative splicing, differential RNA stability, micro-RNA induced repression, post-translational modifications, protein-turnover, and protein localisation. The ability to resolve such biological variables is confounded by the technical noise

Rob 5/5/14 8:03 PM
Deleted: using

Rob 5/5/14 8:03 PM
Deleted: the

Rob 5/5/14 8:04 PM
Deleted: ation methods

Rob 5/5/14 8:04 PM
Deleted: mentioned earlier

Rob 5/5/14 8:04 PM
Deleted: ;

Rob 5/5/14 8:04 PM
Deleted: this approach has been exploited to reveal

Rob 6/5/14 4:56 PM
Deleted: Indeed

Rob 6/5/14 4:57 PM
Deleted: s

Rob 5/5/14 7:48 PM
Deleted: ,

Rob 5/5/14 7:48 PM
Deleted: the relative abundance of which has been shown to provide valuable transcriptional insights [ref]

Rob 6/5/14 4:57 PM
Deleted: in their own right

Rob 5/5/14 12:00 PM
Deleted: [ref]

Rob 5/5/14 8:08 PM
Deleted: However

Rob 5/5/14 8:08 PM
Deleted: add to the original

Rob 6/5/14 4:58 PM
Deleted: The situation is further complicated due to the often limited correlation of observed mRNA and protein abundances, even in well matched experiments.

Rob 6/5/14 5:00 PM
Deleted: Several studies over recent years have attempted to elucidate this

Rob 6/5/14 5:00 PM
Deleted: of the transcriptome with those of the proteome

Rob 6/5/14 5:01 PM
Deleted: ,

Rob 6/5/14 5:01 PM
Deleted: but almost all have report ... [2]

Rob 6/5/14 5:02 PM
Deleted: . This

Rob 5/5/14 8:08 PM
Deleted: including

introduced due to differences in sample preparation, measurement technology, and data handling used for the transcriptomic and proteomic analyses[41,42].

A recent development in transcriptomics has enabled investigators to directly assess translational control, known to be a significant regulatory process that determines the protein output of a transcript[43], by sequencing the very short fragments of RNA contained within the ribosome itself[7]. This so-called 'ribosome profiling' allows, for the first time, a transcriptome-wide survey of the positions of ribosomes on each transcript, including precise identification of open reading frames, and when compared to the relative abundance of those same transcripts has introduced the concept of translational efficiency to mainstream gene expression profiling. This has a large impact for the integration of proteome and transcriptome analyses as the translational efficiency may, once the methodology matures, prove a more reliable indicator of protein abundance than simple RNA expression[43]. For example, an early study of translational efficiency in yeast revealed that cells can modify their protein output whilst maintaining stable RNA abundances during different stages of meiosis, simply by increasing the density of ribosomes on selected transcripts[44]. Ribosome profiling has also shown that long non-coding RNAs in the cytosol, which are known to be spliced, capped, and polyadenylated in a similar manner to mRNAs[45], are engaged by the polyribosome[40], but do not code for protein[46,47].

Rob 5/5/14 8:09 PM
Deleted: mono-

Rob 5/5/14 8:09 PM
Deleted: complex

Rob 6/5/14 5:04 PM
Deleted: ,

Rob 6/5/14 5:04 PM
Deleted: ,

Rob 5/5/14 8:09 PM
Deleted: Finally, ribosome profiling has been used to identify sites in the 5' UTR of known transcripts that contain short open reading frames (ORFs). These upstream ORFs (uORFs) can have a variety of regulatory influences on their host transcript and can themselves produce short peptides, either of these scenarios are ripe for further exploration in searching for the relevant peptide sequences by MS/MS.

Improving the compatibility of proteomics with functional genomics

Prior to the development of massively parallel DNA sequencing, and to some extent microarrays, the volume of data output in a typical MS/MS experiment approximately matched that of a transcriptomic assay and either could be considered a valuable stand-alone resource. However it is now a simple fact that RNA-sequencing produces a substantially larger volume of usable data and, as such, there now exists significant analytical advantages to combining transcriptomic and proteomic data[14].

High-quality data integration, of course, depends upon equally high-quality data made available in well-curated and –maintained online repositories. Databases of **peptide identifications** such as PRIDE[48] and Peptide Atlas are a valuable resource for mapping spectra based on previously identified peptides, however such resources are of limited use to non-experts who may simply desire higher-level information on sites of post-translational modifications and the complement of proteins that are observed in a given disease state, tissue, or cellular compartment[49]. In the same vein, a very basic, but significant limitation of most studies that attempt to combine RNA and protein level results is that **different gene annotations** are used in the analysis of the RNA compared to the protein, which results in

Rob 5/5/14 8:25 PM
Formatted: Indent: First line: 1.27 cm

Rob 6/5/14 5:05 PM
Deleted: potentially extremely

Rob 5/5/14 8:10 PM
Formatted: Font:Bold

mundane but immediate difficulties in integrating the output of these assays. More significantly, however, is the extent to which the use of different reference annotations limit one's ability to relate observed peptides to potential transcripts, which is necessary for integrative analyses of molecular networks[50]. Adoption of a consistent, high-quality reference would benefit not only improve RNA/protein abundance comparisons, but may also facilitate the integration of peptide/protein abundance information to genome browsers such as those provided by ENSEMBL and UCSC.

The community would also benefit from a resource providing **quantitative data** on peptide or protein abundance, such as the Plant Proteome Database[51] or the Encyclopaedia of Proteome Dynamics[52]. Human gene expression atlases are well populated with in-situ and whole-transcriptome RNA abundance data[10,11], and although there are efforts to map protein expression in human tissues by immunohistochemistry[20] the real power of such resources lies in the combination of this expression localisation data and concurrent relative abundance measurements of thousands of genes. Unfortunately public quantitative proteomic datasets are currently lacking both in volume and in standardisation, particularly in terms of the processing of the raw spectra and the methods of obtaining and normalising the peptide- or protein-level abundances. A resource that combined peptide identification/quantification data in terms of a genome annotation that is consistent with genomics and transcriptomics would be of high value to the community, not least in providing a framework with which to tackle open questions in proteomics such as the non-uniform coverage of peptides belonging to the same protein.

Although it may not have a profound effect on the results of a single, internally consistent, analysis, the lack of data/analysis standardisation further increases the difficulty of comparing proteomic quantifications across studies, which is a key determinant of the utility of the kinds of resources described above. Recommendations from the Human Proteome Organisation (HUPO) Brain Proteome Project (BPP) for dissemination of MS/MS data include storing the list of identified spectral peaks along with the corresponding peptide sequence and modifications as the most 'sensible' unit of measurement[53]. However this recommendation is particularly vulnerable to issues regarding the **identification and selection of the peaks** from the raw spectra. The choice of software for peak-picking, peptide identification, and quantification as well as the selection of related parameters, such as the potential PTMs to be included in the peptide identification, remains a significant source of information-loss and -variability in MS/MS[54]. Until the situation improves regarding the processing of spectral peaks, it is necessary to retain the raw MS/MS output in public resources to facilitate development of open-source analysis software and re-analysis of collections of published datasets.

Rob 6/5/14 5:05 PM
Deleted: common

Rob 5/5/14 8:18 PM
Formatted: English (US)

Rob 5/5/14 8:26 PM
Deleted: ,

Rob 6/5/14 5:06 PM
Comment [1]: Potentially cut?

Rob 5/5/14 8:26 PM
Deleted: used

Rob 6/5/14 5:11 PM
Deleted: data

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 MS/MS profiling of intact proteins as well as **computational methods** that maintain an awareness of 'upstream' information during analysis of MS/MS spectra. Utilising ribosome footprinting, for example, to identify coding sequences and translation initiation sites has been shown to produce moderate increases rates of MS/MS peptide identification in the same samples[58]. Additionally, there have been a small number of recent efforts to leverage transcript sequence information obtained from RNA-seq to improve peptide identification[55,56], which have in turn resulted in the production of somewhat basic software tools for the direct integration of such datasets[57]. However there is still a great deal of work to do to create the computational tools required to consolidate and integrate information regarding genomic variants, transcript structure and abundance, and translational efficiency in routine MS/MS analysis.

Benefits of integrating functional genomic and proteomic profiling of the CNS

Systematic integration of functional genomics assays has been crucial for deeper understanding of the complex cellular machinery. Integration, for example, of DNA variants and chromatin signals with transcriptomics has been extremely valuable for gaining a deeper intuition for the genetic, epigenetic, and post-transcriptional regulation in the CNS (for example, see brainspan.org). Integration of transcriptomic and proteomic data has the potential to be just as powerful for monitoring the sometimes subtle effects or dysfunction of protein production and localisation underlying neurodevelopment and disorder. In addition to the use of MS/MS for validating or comparing to RNA abundance or translational efficiency, a major attraction of proteomics lies in the data obtained from the peptide sequences, which enable proteome-wide validation of genomic and transcriptomic variants, allelic imbalance, and isoform identification[14] (**Figure 2**).

Confirmation of alternative splicing and isoform usage

Even small modifications to the structure, abundance, or localisation of RNAs and proteins in the brain can have profound biochemical and physiological consequences. One of the more comprehensively studied transcriptional processes is **alternative splicing** (AS), which is well known to be a highly tissue-specific process[8] that greatly increases the complexity of the

Rob 6/5/14 5:17 PM
Deleted: We have just recently started to see the

Rob 6/5/14 5:20 PM
Deleted: application of

Rob 6/5/14 5:21 PM
Deleted: and

Rob 6/5/14 5:22 PM
Deleted: analysis

Rob 6/5/14 5:24 PM
Moved (insertion) [3]

Rob 6/5/14 5:24 PM
Deleted: d

Rob 6/5/14 5:25 PM
Deleted: in the yield

Rob 6/5/14 5:25 PM
Deleted: spectra

Rob 6/5/14 5:23 PM
Deleted: aim to more tightly integrate analyses of the proteome with the transcriptome. Experimental methodologies such as ribosome profiling, single molecule RNA-sequencing, and top-down MS/MS profiling of intact proteins enable greater selectivity and sensitivity to molecules that are actively involved in the process of protein production and greater specificity to the exact structure of these transcripts a ... [3]

Rob 6/5/14 5:23 PM
Deleted: the

Rob 6/5/14 5:26 PM
Deleted: MS/MS analyses

Rob 6/5/14 5:24 PM
Moved up [3]: Utilising ribosome ... [4]

Rob 6/5/14 5:27 PM
Deleted: Utilising ribosome footpri ... [5]

Rob 6/5/14 5:31 PM
Deleted: information obtained from ... [6]

Rob 6/5/14 5:31 PM
Deleted: characterisation and

Rob 6/5/14 5:43 PM
Deleted: dynamics cellular process ... [7]

Rob 5/5/14 8:28 PM
Deleted: proteomic data in

Rob 5/5/14 12:36 PM
Formatted: Font:Bold

Rob 5/5/14 12:34 PM
Moved (insertion) [1]

Rob 5/5/14 12:36 PM
Deleted: Alternative RNA-splicing (AS)

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 disorders[59]. During CNS cell development, for example, AS in neuronal progenitor stem cells has revealed differential isoform usage at different stages of maturation[60]. Extensive profiling of AS has characterised the differential expression of DISC1 isoforms that are associated with schizophrenia[61], not to mention the translocation itself[62]. Validation of splicing events by MS/MS is quite common, either in terms of determining the contribution of known isoforms in a given experiment, or verifying the existence of novel exon junctions identified by RNA-sequencing[63].

A recent result that potentially has a significant impact on both the future analyses of the CNS and interpretation of proteomics data is the finding that most genes, in a given cell-type or tissue, tend to express only a single dominant transcript/isoform[64]. This has a direct impact on the so-called interactome, in which differential isoform usage between CNS cell types and during the progression of neurological disorders affects protein-protein interactions, as exemplified in the Autism Spliceform Interaction Network[65]. Currently full-length transcript[66] and isoform[67] profiling technologies are immature[68], however an early example of full-length transcript profiling in the brain directly observed the various isoforms of neurexin, showing their production mediates distinct protein interactions across the synapse[69]. Such investigations will only become more common, and have great potential to significantly simplify the process of integrating and interpreting genome-wide measurements of RNA and protein in the near future.

Confirmation of genomic/transcriptomic variants

Allelic diversity across the human population is well known to influence brain development[70], for example humans suffering microcephaly frequently carry a premature stop mutation in the gene ASPM, which is localised in the mitotic spindle, leading to a truncation of the protein and restricted growth of the cerebral cortex[71]. Integrated analyses of DNA- and RNA-sequence data obtained from the same individuals have resulted in a large number of discoveries relating to ADAR-mediated adenosine to inosine (A-to-I) editing of the transcriptome. **RNA-editing**, also referred to as RNA-DNA differences (RDDs), appears to have played a significant role in human brain evolution[5,72,73] and of the trio of ADAR proteins responsible for the post-transcriptional A-to-I modification, the third (ADAR3) is exclusively expressed in the brain[74]. Several mis-sense RDDs in the AMPA receptor have been shown to alter the downstream behaviour of this protein, are edited at specific stages of human brain development[75], and are required for normal brain function and phenotype in mice[76]. Another example is the serotonin receptor 5-HT_{2C}R (HTR2C),

Rob 6/5/14 5:36 PM
Deleted: alternative splicing (

Rob 6/5/14 5:36 PM
Deleted:)

Rob 6/5/14 5:38 PM
Deleted: different

Rob 6/5/14 5:37 PM
Deleted: to the final neuronal state

Rob 6/5/14 5:45 PM
Deleted: The

Rob 6/5/14 5:38 PM
Deleted: [ref]

Rob Kitchen 5/5/14 12:34 PM
Comment [2]: not brain

Rob 5/5/14 12:36 PM
Formatted: Indent: First line: 1.27 cm

Rob 6/5/14 5:46 PM
Deleted: The myelin basic protein (MBP), for example, expresses a completely different isoform in the brain compared to all other healthy human tissues[64].

Rob 5/5/14 12:31 PM
Deleted: Thus, even small modifications to the structure, abundance, or localisation of RNAs and proteins in the brain can have profound consequences.

Rob 6/5/14 5:47 PM
Formatted: Font:Not Bold

Rob 6/5/14 5:47 PM
Deleted: RNA-edit sites

which contains numerous RDDs that alter both the expressed protein sequence[77] and cause an order of magnitude reduction in efficacy in the interaction of the receptor with its G proteins[78]. Unfortunately, reliable identification RDDs is technically very challenging and has led to erroneous false-positive identifications[79], however MS/MS an extremely attractive tool for unbiased validation of mis-sense RDDs, which is capable of detecting not only the presence or absence of an RDD but provides quantitative data on the abundance of the edited and unedited copies of the proteins. The use of a proteome-wide, independent verification of mis-sense RDDs could equally be of great benefit to biochemical analyses of RNA-editing, such as ICE-seq[73], in which inosine nucleotides are directly identified through chemical modification.

Within a given individual the **allele-specific expression** (ASE) of RNA is the result of epigenetic regulatory processes that are common across species and tissues[80]. Detection of ASE relies on the discovery of heterozygous genomic variants that allow us to monitor imbalances in the abundance of RNA produced from each parental allele. A recent survey of the mouse CNS revealed 1,300 genes exhibit an allelic imbalance in expression that, interestingly, during brain development appears to favour the maternal allele, while in the adult the bias shifts toward the paternal allele[81]. Detection of allelic expression at the protein level relies on mis-sense mutations of one allele relative to the other, however such events occur with sufficient frequency to make MS/MS validations worthwhile. For example, an analysis of allele-specific protein expression in yeast reported that around 10% of heterozygous coding loci exhibit an allelic bias[82], despite somewhat weak correlation (<0.35) to that observed at the mRNA from the same genes. In addition to verification of ASE events, an interesting application of MS/MS has been to assess allele-specific transcription factor binding (ASB), in which regions of the genome that are known to be heterozygous are purified and the complex of proteins bound to these regions are subsequently profiled by MS/MS to identify differential binding[83]. The multi-omic assessment of ASE and ASB is appealing due to the cellular cascade these processes can cause; for example an allelic transcription- or splicing-factor may affect downstream interactions that are, by definition, epigenetically inherited from one parent.

Quantitative trait loci (QTL) have been extensively profiled in a variety of tissues using transcriptomics in terms of the relation of genomic variants to RNA expression changes (eQTL) and more recently, variants have also been related to protein abundance changes (pQTL)[84]. In the prefrontal cortex, for example, such variants have been found to affect the expression of more than 100 genes[85]. An analysis of individuals genotyped in the HapMap project reported that almost two-thirds of 185 detected cis-acting pQTLs were not found in a complementary analysis of the RNA[42]

Rob 6/5/14 5:47 PM
Deleted: RNA-editing sites

Rob 5/5/14 7:51 PM
Deleted: Although these RNA-DNA differences (RDDs) tend to occur in intergenic and intronic regions, several important instances RDDs have been found to alter the protein product of the transcript [ref].

Rob 6/5/14 5:47 PM
Deleted: of

Rob 5/5/14 7:52 PM
Deleted: l

Rob 6/5/14 5:47 PM
Deleted: lead

Rob 6/5/14 5:47 PM
Deleted: to an

Rob 5/5/14 8:38 PM
Deleted: and

Rob 5/5/14 8:38 PM
Deleted: this bias

Rob 5/5/14 8:38 PM
Deleted: a

Rob 5/5/14 8:38 PM
Deleted: was observed between brain regions

Rob 5/5/14 8:39 PM
Deleted: however

Rob 5/5/14 8:39 PM
Deleted: the

Rob 5/5/14 8:39 PM
Deleted: of this bias

Rob 5/5/14 8:39 PM
Deleted: level

Rob 5/5/14 8:39 PM
Deleted: in

Rob 5/5/14 8:39 PM
Deleted: was fairly poor (<0.35)

Rob 5/5/14 8:40 PM
Deleted: collection

Rob 5/5/14 8:41 PM
Deleted: in order

Rob Kitchen 1/5/14 3:48 PM
Comment [3]: not brain

Rob 5/5/14 12:34 PM
Deleted:

Rob 5/5/14 12:34 PM
Moved up [1]: Alternative RNA-splicing (AS) is well known to be a highly tissue-specific process[9] that greatly increases the complexity of the pote... [8]

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 ENCODE[86] and others[87], have reported that the union of all RNA molecules detected across a variety of tissues, cell-lines, and conditions infer 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 presence of these molecules may imply they have a functional role in cellular processes[86]. In terms of understanding the functional output of genomes of previously un-annotated organisms, transcriptomics alone is insufficient to accurately define the cohort of protein coding sequences, even when combined with ribosome profiling; high-throughput proteome profiling by MS/MS again provides the most useful avenue to this.

Challenges in assessing the proteome of the CNS and implications for future studies in quantitative neuroproteomics

There have been many proteomic studies of the CNS by MS/MS, and these have been enumerated in previous reviews[88-90]. More recent applications of MS/MS to the CNS include the profiling of neural tissue in model organisms such as the fruit fly[91] and neurodegeneration in zebrafish[92], as well as profiling cultures of specific CNS cell-types in primary culture such as neurons[93] and oligodendroglial cells[94]. Fluorescence activated cell sorting (FACS) of microglia has revealed over 100 genes that are enriched compared to neurons and other oligodendrocytes[95] and enabled MS/MS assessment of synaptic proteins[96]. A novel method, fluorescence activated nuclei sorting (FANS), in combination with an antibody against the neuronal-specific splicing protein, NeuN, has been successfully applied to purify neuronal nuclei from primary tissue in order to provide a quantitative comparison of the abundance of nuclear proteins with astrocytes and oligodendrocytes[97]. Despite these successes, there exist two particularly challenging issues regarding to proteomic profiling of the CNS, namely the immense cellular and sub-cellular heterogeneity and the dynamics of protein turnover. Resolving these challenges will be essential to understanding long-term neuronal adaptation, especially at the synapse, to external stimuli such as stress, therapeutic and other types of drugs, and adaptations related to neuropsychiatric and neurodegenerative disease.

Monitoring rates of protein production and degradation can be achieved fairly straightforwardly in cultured cells by pulse-labelling followed by SILAC proteomics[98]. Similarly,

Rob 6/5/14 5:48 PM
Deleted: .

Rob 6/5/14 5:48 PM
Deleted: lot

Rob 6/5/14 4:40 PM
Deleted: . ... [9]

Rob 6/5/14 4:40 PM
Deleted: T

Rob 6/5/14 4:41 PM
Deleted: . Both

Rob 6/5/14 4:41 PM
Deleted: are critical components of

Rob 6/5/14 4:42 PM
Deleted: the

Rob 6/5/14 4:42 PM
Deleted: of specific classes of neurons, specifically

Rob 6/5/14 4:42 PM
Deleted: of abuse

Rob 6/5/14 4:43 PM
Deleted: illness

Rob 6/5/14 11:19 PM
Deleted:

Rob 6/5/14 11:24 PM
Formatted: Indent: First line: 1.27 cm

nascent peptide chains can be captured as they are synthesised by the ribosome using biotin-puromycin labelling[99]. However these approaches do not lend themselves to assessment of rates of protein turnover in any primary mammalian or human tissue, let alone the CNS. With ribosome profiling it is possible to infer rates of protein synthesis, however direct measurement of protein turnover in primary tissues by MS/MS is currently out of reach. It may be the case that for the foreseeable future, non-invasive imaging methods that allow repeated-measurement of the same cells are potentially the most promising means by which differential rates of protein synthesis can be observed in the mammalian CNS[100].

However the difficulty in obtaining direct measurements of protein turnover really only affects biological questions regarding relatively short-term effects or changes. For processes with longer timescales such as, for example, neurodevelopment, tumour growth, or ageing the factors of interest are more the bulk changes in steady state isoform abundance between brain regions or, more importantly, neuronal cell types. Fortunately, there has been rapid and substantial progress in developing methods for interrogating the large variety of cells and intra-cellular compartments that have the potential to vastly improve upon existing whole-tissue analyses of the CNS.

Inter- and intra-cellular heterogeneity

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 their morphology, connectivity, and molecular and electrophysiological properties. All of the different CNS cell types develop and are integrated into functional networks within very precise constraints; deviations from this normal course of development can lead to a variety of disorders. Furthermore, the sub-cellular localisation of RNAs and proteins as well as the rapid and cell-type specific production of specific genes are fundamental to neuronal development, function, and disease[101]. The abundance of proteins localised, for example, at the **pre- and post-synaptic density** (PSD) is dependent on cell-type and brain region[102,103], however transcriptomic analysis is sub-optimal for assaying such differences due to the confound introduced by the RNA trafficking and/or local translation at synapses or potentially in axons[101]. Proteomic profiling can be used to directly access the protein complement of the PSD, however such purifications can introduce significant variability in the measured abundances of some of these protein. Similarly, **density/gradient centrifugation** for nuclear/organelle purification offers sub-cellular resolution[22], but this will always be confounded by inter-cellular variability

Rob 6/5/14 10:36 PM
Formatted: Font:12 pt, Bold

Rob 6/5/14 4:44 PM
Moved down [2]: There are methods to monitor protein dynamics, but these are not particularly applicable to the mammalian CNS. Monitoring rates of protein production and degradation can be achieved fairly straightforwardly in cultured cells by pulse-labelling followed by SILAC proteomics[98]. Similarly, nascent peptide chains can be captured as they are synthesised by the ribosome using a biotin-puromycin labelling approach[99]. However these approaches do not lend themselves to assessment of rates of protein turnover in mammalian and human tissues. Integration with transcriptome profiling, specifically ribosome profiling, is an increasingly attractive proxy to direct measurement of protein synthesis. Similarly, non-invasive imaging methods are potentially valuable methods by which differential rates of protein synthesis can be repeatedly observed in the mammalian CNS[100], both over long and relative short periods of time.

Rob 6/5/14 4:44 PM
Deleted: Monitoring rates of protein degradation can be achieved fairly straightforwardly in cultured cells by pulse-labelling followed by SILAC proteomics[98]. Similarly, nascent peptide chains can be captured as they are synthesised by the ribosome using a biotin-puromycin labelling approach[99]. However these approaches do not lend themselves to assessment of rates of protein turnover in mammalian and human tissues. Integration with transcriptome profiling, specifically ribosome profiling, is an increasingly attractive proxy to direct measurement of protein synthesis. ... [10]

Rob 5/5/14 12:04 PM
Deleted: Proteins

Rob 5/5/14 12:04 PM
Deleted: are of interest in the fields of addiction and substance abuse [ref]. The abundance of particular PSD proteins

Rob 6/5/14 5:51 PM
Deleted: are

Rob 5/5/14 8:49 PM
Deleted: [ref]

Rob 5/5/14 8:49 PM
Deleted:

Rob 5/5/14 8:49 PM
Deleted: [ref]

unless also applied to homogeneous collections of cells.

The molecular diversity both between and within the traditional neurotransmitter classes in the mammalian CNS has been observed not only based on the absolute presence or absence of proteins, but also in their relative abundance levels, further supporting the need for unbiased, comprehensive, and quantitative measurement of isoform expression at a higher resolution than is available at the whole-tissue level[104]. Existing whole-tissue analyses[10] are not sensitive to this small scale inter- or intra-cellular variability and suffer from confounded and diluted signals from not only different classes of neurons but also from the high proportion of glia. There are currently a variety of atlases being constructed to address this issue of heterogeneity in the mammalian brain, including spatial and temporal RNA expression, by **ISH**[11,105] and **immunohistochemistry**[20]. However this approach is severely limited by very low-throughput, is subject to variable antibody specificity and is, at best, semi-quantitative.

Analysis of single or small numbers of neural cells, obtained for example by laser-capture microscopy (LCM), followed by transcriptional profiling is an attractive approach as it provides a quantitative measure of all RNAs in a very small physical volume of tissue. However extracting neuronal cells by LCM is not guaranteed to result in cell-type specificity due to the close proximity and overlapping processes of these cells, especially between the layers of the cortex[105]. Additionally, the throughput of such an approach is still too low to meaningfully assess, for example, the response of collections of neurons in a given brain region to experimental variables such as the treatment effect a drug, especially when there is a requirement for multiple biologically independent subjects. In addition, current proteomic technologies would require larger numbers of cells to be collected by LCM in order to obtain enough material, further exacerbating the issues of contamination due to non-target cell types.

The ultimate method for monitoring the dynamics of protein production is to analyse protein abundance application in **single-cells**. For several years, researchers have been using qPCR to assess transcript abundances obtained from individual cells and very recently RNA-sequencing of single cells has expanded these analyses to the whole transcriptome. What has been found mirrors observations at a whole-tissue level[64], in that a given gene in a given cell typically transcribes a single isoform[106]. Moreover, very recent observations have even suggested that mammalian cells express these isoforms randomly from a single allele[107], revealing perhaps novel regulatory mechanisms within the cell to produce this behaviour. There is no doubt that with the continued development of MS-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]. However, as is the

Rob 5/5/14 12:06 PM
Deleted: of spatial and temporal RNA expression

Rob 5/5/14 12:10 PM
Deleted: This

Rob 5/5/14 12:10 PM
Deleted: however as it is

Rob 6/5/14 5:54 PM
Deleted: alternative

Rob 6/5/14 5:54 PM
Deleted: ,

Rob 6/5/14 5:54 PM
Deleted:

Rob 6/5/14 5:54 PM
Deleted: ing

Rob 6/5/14 5:54 PM
Deleted: individual

Rob 6/5/14 5:54 PM
Deleted: ment

Rob 6/5/14 5:55 PM
Deleted: Current

case for LCM, the utility of single-cell analyses for studying the effects of brain development, malfunction, and effect of chemical treatment, is very limited due to the issues of sample-throughput and resolution to extremely low abundance molecules. Fortunately, there are other approaches to studying not single neurons, but single populations of neurons that are more compatible with existing technologies and do not suffer the significant issues regarding throughput and contamination as do LCM / single cell techniques.

An extremely elegant means of obtaining quantitative spatial expression measurements for a specific gene is through the creation of libraries of bacterial artificial chromosomes containing fluorescent markers downstream of target regulatory elements (enhancer/promoters) that themselves lie upstream of the desired gene[12]. Such an approach, when collected for multiple regulatory elements, or when used in combination with ISH/IHC spatial expression profiles can be used to reveal cell-type specific promoter activity and RNA expression[109]. A complementary method for obtaining all cytosolic RNA expressed in a given cell-type leads directly from this identification of cell-type specific promoters, in which overexpression of a **GFP-labelled ribosomal protein** under the control of one such cell-type-specific promoter allows purification by IP of the transcripts bound by the polyribosome in the desired cells[110]. The principal advantage of this approach is that cellular material is obtained in the same way as if profiling a standard tissue extract, except the introduction of the eGFP-IP removes RNAs from non-target cell-types. Moreover, due to the labelling of the ribosomal protein, strong GFP signal is also observed in the nucleolus of target cell types enabling FACS purification of target nuclei for assessment of DNA, histone modifications, transcription factor binding, or nascent RNA transcription in the specific cell-type of interest[97,111]. The obvious disadvantage to these approaches is that they are only compatible with systems amenable to genetic modification, such as cultures or rodent models, and so of limited utility for directly studying human neurobiology. Furthermore, although the FACS approach can be used to profile proteins in the nuclei of the selected cell-type, cytosolic proteins are much more difficult to obtain by such a method thanks to the extensive processes of neuronal cells.

For application to human neurobiology, and as a refinement of the FACS ideology, it may become possible to exploit proteins at the plasma membrane of specific neuronal subtypes for purification by antibody pulldown. Use of such cell-surface markers may be able to 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. Such an approach would thus have the power to enable truly

Rob 6/5/14 11:05 PM
Deleted: (BAC)

Rob 5/5/14 7:52 PM
Deleted: ,

Rob 6/5/14 5:56 PM
Deleted: transfection

Rob 6/5/14 11:32 PM
Deleted:

Rob 6/5/14 11:33 PM
Deleted: .

integrative, specific, and biologically meaningful analyses of the mammalian CNS.

<<BOX 1>>The various methodologies for contemporary MS/MS have been comprehensively covered elsewhere including excellent reviews on the mass-spectrometer technologies[112], computational analysis of spectra[113], and specific examples of proteomics applied to the CNS[89]. For a given sample, the tens to hundreds of thousands of peptides, typically the products obtained using trypsin digestion, are individually quantified, but the abundances from all peptides derived from a given protein can be aggregated to facilitate protein-level expression analyses in addition to the simpler peptide-level comparisons. Briefly, MS/MS experiments are distinguished by two choices, illustrated in **Figure 1**; the first being the method of quantification of the observed protein products and the second whether the experiment should be hypothesis-driven or hypothesis-free. Labelled MS/MS analysis methods such as Stable Isotope Labelling by Amino acids in cell Culture (SILAC), Stable Isotope Labelling of Mammals (SILAM), and Isobaric Tag for Relative and Absolute Quantitation (iTRAQ) are commonly employed for assessing differential abundance of proteins across phenotypes, conditions, or treatments; while label-free abundance estimates [ref] also allow relative quantitation of peptides to each other. Absolute quantitation can be achieved, in an increasingly high-throughput manner, using stable isotope dilution (SID) of a number of target proteins by spiking in synthetic labelled sequences that are exact analogues of the target sequences[114]. Improvements to both labelled and label-free quantitation can be achieved by restricting the MS/MS scans to pre-defined ranges; this approach, termed Selected Reaction Monitoring (SRM) or Multiple Reaction Monitoring (MRM)[115], essentially allows the instrument to more accurately measure the abundance of selected peptides by spending a greater fraction of total instrument time monitoring a smaller list of pre-defined peptides. Recently, however a renewed interest in data-independent spectra acquisition[116], driven by faster and more accurate mass-spectrometers, has led to the development of methods that claim to be capable of detecting 30% more proteins than conventional label-free acquisition[117].

<<END BOX 1>>

Rob 6/5/14 4:44 PM
Moved (insertion) [2]

Rob 6/5/14 10:38 PM
Deleted: There are methods to monitor protein dynamics, but these are not particularly applicable to the mammalian CNS. ... [11]

Rob 5/5/14 8:53 PM
Deleted: software such as OpenSWATH which

Rob 5/5/14 8:53 PM
Deleted: s

<<START ABBREVIATIONS>>

MS/MS	Tandem mass-spectrometry
MALDI	Matrix Assisted Laser Desorption/Ionisation
iTRAQ	Isobaric Tag for Relative and Absolute Quantitation
SILAC/M	Stable Isotope Labelling of Culture/Mammals
SID	Stable Isotope Dilution
S/MRM	Selected/Multiple Reaction Monitoring
RNA-seq	Second-generation, massively parallel RNA-sequencing
fMRI	Functional Magnetic Resonance Imaging
dMRI	Diffusion Magnetic Resonance Imaging
PET	Positron Emission Tomography
ISH	In-Situ Hybridisation
IHC	Immunohistochemistry
ES cell	Embryonic Stem cell
PTM	Post-translational modification
[u]ORF[upstream]	Open reading frame
RDD	RNA-DNA differences, arising due to RNA-editing by ADAR
ASE/B	allele specific expression/binding
e/pQTL	expression/protein quantitative trait loci
ENCODE	Encyclopaedia of DNA Elements project
gEUVADIS	Genetic European Variation in Health and Disease project
TCGA	The Cancer Genome Atlas project
GTEx	Genotype-Tissue Expression project

<<END ABBREVIATIONS>>

References

1. Abyzov, A. *et al.* Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. *Nature* **492,** 438–442 (2013).
2. Park, P. J. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* **10,** 669–680 (2009).
3. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological

basis for transcription regulation. *Cell* **148,** 84–98 (2012).
4. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10,** 57–63 (2009).
5. Li, J. B. & Church, G. M. Deciphering the functions and regulation of brain-enriched A-to-I RNA editing. *Nature Publishing Group* **16,** 1518–1522 (2013).
6. Chi, S. W., Zang, J. B., Mele, A. & Darnell, R. B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460,** 479–486 (2009).
7. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324,** 218–223 (2009).
8. Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456,** 464–469 (2008).
9. Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* 1–19 (2014). doi:10.1038/nature13185
10. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478,** 483–489 (2011).
11. Sunkin, S. M. *et al.* Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res* **41,** D996–D1008 (2013).
12. Heintz, N. Gene expression nervous system atlas (GENSAT). *Nature Neuroscience* **7,** 483 (2004).
13. Zhao, Y. *et al.* Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling. *Nat Nanotechnol* (2014). doi:10.1038/nnano.2014.54
14. Vidal, M. *et al.* The human proteome - a scientific opportunity for transforming diagnostics, therapeutics, and healthcare. *Clin Proteomics* **9,** 6 (2012).
15. Pollock, J. D., Wu, D.-Y. & Satterlee, J. S. Molecular neuroanatomy: a generation of progress. *Trends Neurosci.* **37,** 106–123 (2014).
16. Burns, R. *et al.* The Open Connectome Project Data Cluster: Scalable Analysis and Vision for High-Throughput Neuroscience. *Mol Syst Biol* (2013). doi:10.1145/2484838.2484870
17. Van Essen, D. C. & Ugurbil, K. The future of the human connectome. *Neuroimage* **62,** 1299–1310 (2012).
18. Lichtman, J. W., Livet, J. & Sanes, J. R. A technicolour approach to the connectome. *Nat Rev Neurosci* **9,** 417–422 (2008).
19. Chung, K. *et al.* Structural and molecular interrogation of intact biological systems. *Nature* **497,** 332–337 (2013).
20. Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol* **28,** 1248–1250 (2010).
21. Liu, X. *et al.* Molecular imaging of drug transit through the blood-brain barrier with MALDI mass spectrometry imaging. *Sci Rep* **3,** 2859 (2013).
22. Cox, B. & Emili, A. Tissue subcellular fractionation and protein extraction for use in mass-spectrometry-based proteomics. *Nature Protocols* **1,** 1872–1878 (2006).
23. Boisvert, F.-M., Lam, Y. W., Lamont, D. & Lamond, A. I. A quantitative proteomics analysis of subcellular proteome localization and changes induced by DNA damage. *Mol Cell Proteomics* **9,** 457–470 (2010).
24. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473,** 337–342 (2011).
25. Naegle, K. M. *et al.* PTMScout, a Web resource for analysis of high throughput post-translational proteomics studies. *Mol Cell Proteomics* **9,** 2558–2570 (2010).
26. Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat Biotechnol* **21,** 255–261 (2003).
27. Oppermann, F. S. *et al.* Large-scale proteomics analysis of the human kinome. *Mol Cell*

Proteomics **8,** 1751–1764 (2009).
28. Deribe, Y. L., Pawson, T. & Dikic, I. Post-translational modifications in signal integration. *Nature Structural & Molecular Biology* **17,** 666–672 (2010).
29. Siuti, N. & Kelleher, N. L. Decoding protein modifications using top-down mass spectrometry. *Nat Meth* **4,** 817–821 (2007).
30. Sridharan, R. *et al.* Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. *Nat. Cell Biol.* **15,** 872–882 (2013).
31. 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* **110,** 3645–3650 (2013).
32. Moghaddas Gholami, A. *et al.* Global proteome analysis of the NCI-60 cell line panel. *Cell Rep* **4,** 609–620 (2013).
33. Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R. & Mann, M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat Meth* **7,** 383–385 (2010).
34. Yu, L.-R. *et al.* Global analysis of the cortical neuron proteome. *Mol Cell Proteomics* **3,** 896–907 (2004).
35. Chen, R. *et al.* Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* **148,** 1293–1307 (2012).
36. Phanstiel, D. H. *et al.* Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat Meth* **8,** 821–827 (2011).
37. Greenbaum, D., Colangelo, C., Williams, K. & Gerstein, M. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol* **4,** 117 (2003).
38. Lundberg, E. *et al.* Defining the transcriptome and proteome in three functionally different human cell lines. *Mol Syst Biol* **6,** 450 (2010).
39. Bitton, D. A., Okoniewski, M. J., Connolly, Y. & Miller, C. J. Exon level integration of proteomics and microarray data. *BMC Bioinformatics* **9,** 118 (2008).
40. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147,** 789–802 (2011).
41. Li, J. J., Bickel, P. J. & Biggin, M. D. System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ* **2,** e270 (2014).
42. Wu, L. *et al.* Variation and genetic control of protein abundance in humans. *Nature* **499,** 79–82 (2013).
43. Kong, J. & Lasko, P. Translational control in cellular and developmental processes. *Nat Rev Genet* **13,** 383–394 (2012).
44. Brar, G. A. *et al.* High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* **335,** 552–557 (2012).
45. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nat Rev Genet* **10,** 155–159 (2009).
46. Bánfai, B. *et al.* Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res* **22,** 1646–1657 (2012).
47. Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S. & Lander, E. S. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* **154,** 240–251 (2013).
48. Martens, L. *et al.* PRIDE: the proteomics identifications database. *Proteomics* **5,** 3537–3545 (2005).
49. Ahmad, Y. & Lamond, A. I. A perspective on proteomics in cell biology. *Trends Cell Biol.* **24,** 257–264 (2014).

50. Emmett, M. R. *et al.* Integrative biological analysis for neuropsychopharmacology. *Neuropsychopharmacology* **39,** 5–23 (2014).
51. Sun, Q. *et al.* PPDB, the Plant Proteomics Database at Cornell. *Nucleic Acids Res* **37,** D969–74 (2009).
52. Larance, M., Ahmad, Y., Kirkwood, K. J., Ly, T. & Lamond, A. I. Global subcellular characterization of protein degradation using quantitative proteomics. *Mol Cell Proteomics* **12,** 638–650 (2013).
53. Martens, L. *et al.* A comparison of the HUPO Brain Proteome Project pilot with other proteomics studies. *Proteomics* **6,** 5076–5086 (2006).
54. Bell, A. W. *et al.* A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. *Nat Meth* **6,** 423–430 (2009).
55. Lopez-Casado, G. *et al.* Enabling proteomic studies with RNA-Seq: The proteome of tomato pollen as a test case. *Proteomics* **12,** 761–774 (2012).
56. Wang, X. *et al.* Protein identification using customized protein sequence databases derived from RNA-Seq data. *J Proteome Res* **11,** 1009–1017 (2012).
57. Wang, X. & Zhang, B. customProDB: an R package to generate customized protein databases from RNA-Seq data for proteomics search. *Bioinformatics* **29,** 3235–3237 (2013).
58. 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* **12,** 1780–1790 (2013).
59. Johnson, M. B. *et al.* Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62,** 494–509 (2009).
60. Wu, J. Q. *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* **107,** 5254–5259 (2010).
61. Nakata, K. *et al.* DISC1 splice variants are upregulated in schizophrenia and associated with risk polymorphisms. *Proceedings of the National Academy of Sciences* **106,** 15873–15878 (2009).
62. Soares, D. C., Carlyle, B. C., Bradshaw, N. J. & Porteous, D. J. DISC1: Structure, Function, and Therapeutic Potential for Major Mental Illness. *ACS Chem Neurosci* **2,** 609–632 (2011).
63. Sheynkman, G. M., Shortreed, M. R., Frey, B. L. & Smith, L. M. Discovery and mass spectrometric analysis of novel splice-junction peptides using RNA-Seq. *Mol Cell Proteomics* **12,** 2341–2353 (2013).
64. Gonzàlez-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* **14,** R70 (2013).
65. Corominas, R. *et al.* Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. *Nat Commun* **5,** 3650 (2014).
66. Sharon, D., Tilgner, H., Grubert, F. & Snyder, M. A single-molecule long-read survey of the human transcriptome. *Nat Biotechnol* **31,** 1009–1014 (2013).
67. Tran, J. C. *et al.* Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480,** 254–258 (2011).
68. Engström, P. G. *et al.* Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat Meth* **10,** 1185–1191 (2013).
69. Treutlein, B., Gokce, O., Quake, S. R. & Südhof, T. C. Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. *Proceedings of the National Academy of Sciences* **111,** E1291–9 (2014).
70. Walsh, C. A. & Engle, E. C. Allelic diversity in human developmental neurogenetics:

insights into biology and disease. *Neuron* **68,** 245–253 (2010).
71. Bond, J. *et al.* ASPM is a major determinant of cerebral cortical size. *Nat Genet* **32,** 316–320 (2002).
72. Blow, M., Futreal, P. A., Wooster, R. & Stratton, M. R. A survey of RNA editing in human brain. *Genome Res* **14,** 2379–2387 (2004).
73. Sakurai, M. *et al.* A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. *Genome Res* **24,** 522–534 (2014).
74. Chen, C. X. *et al.* A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. *RNA* **6,** 755–767 (2000).
75. Lomeli, H. *et al.* Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science* **266,** 1709–1713 (1994).
76. Higuchi, M. *et al.* Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature* **406,** 78–81 (2000).
77. Kawahara, Y. *et al.* Dysregulated editing of serotonin 2C receptor mRNAs results in energy dissipation and loss of fat mass. *Journal of Neuroscience* **28,** 12834–12844 (2008).
78. Burns, C. M. *et al.* Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* **387,** 303–308 (1997).
79. Pickrell, J. K., Gilad, Y. & Pritchard, J. K. Comment on "Widespread RNA and DNA sequence differences in the human transcriptome". *Science* **335,** 1302–author reply 1302 (2012).
80. Buckland, P. R. Allele-specific gene expression differences in humans. *Hum Mol Genet* **13 Spec No 2,** R255–60 (2004).
81. Gregg, C. *et al.* High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* **329,** 643–648 (2010).
82. Khan, Z. *et al.* Quantitative measurement of allele-specific protein expression in a diploid yeast hybrid by LC-MS. *Mol Syst Biol* **8,** 602 (2012).
83. Butter, F. *et al.* Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. *PLoS Genet* **8,** e1002982 (2012).
84. Holdt, L. M. *et al.* Quantitative trait loci mapping of the mouse plasma proteome (pQTL). *Genetics* **193,** 601–608 (2013).
85. Liu, C. *et al.* Whole-genome association mapping of gene expression in the human prefrontal cortex. *Mol Psychiatry* **15,** 779–784 (2010).
86. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489,** 101–108 (2012).
87. Hangauer, M. J., Vaughn, I. W. & McManus, M. T. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet* **9,** e1003569 (2013).
88. Bayés, A. & Grant, S. G. N. Neuroproteomics: understanding the molecular organization and complexity of the brain. *Nat Rev Neurosci* **10,** 635–646 (2009).
89. Craft, G. E., Chen, A. & Nairn, A. C. Recent advances in quantitative neuroproteomics. *Methods* **61,** 186–218 (2013).
90. Sun, F. & Cavalli, V. Neuroproteomics approaches to decipher neuronal regeneration and degeneration. *Mol Cell Proteomics* **9,** 963–975 (2010).
91. Jüschke, C. *et al.* Transcriptome and proteome quantification of a tumor model provides novel insights into post-transcriptional gene regulation. *Genome Biol* **14,** r133 (2013).
92. Gebriel, M. *et al.* Zebrafish brain proteomics reveals central proteins involved in neurodegeneration. *J. Neurosci. Res.* **92,** 104–115 (2014).
93. Seo, J.-W., Kim, Y., Hur, J., Park, K.-S. & Cho, Y.-W. Proteomic analysis of primary cultured rat cortical neurons in chemical ischemia. *Neurochem. Res.* **38,** 1648–1660

(2013).
94. 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* **268C,** 284–296 (2014).
95. Butovsky, O. *et al.* Identification of a unique TGF-β-dependent molecular and functional signature in microglia. *Nature Neuroscience* **17,** 131–143 (2014).
96. Biesemann, C. *et al.* Proteomic screening of glutamatergic mouse brain synaptosomes isolated by fluorescence activated sorting. *EMBO J.* **33,** 157–170 (2014).
97. Dammer, E. B. *et al.* Neuron enriched nuclear proteome isolated from human brain. *J Proteome Res* **12,** 3193–3206 (2013).
98. Boisvert, F.-M. *et al.* A quantitative spatial proteomics analysis of proteome turnover in human cells. *Mol Cell Proteomics* **11,** M111.011429 (2012).
99. 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* **9,** 751–760 (2014).
100. Hughes, T. M. *et al.* Pulse wave velocity is associated with β-amyloid deposition in the brains of very elderly adults. *Neurology* **81,** 1711–1718 (2013).
101. 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).
102. Kavalali, E. T. & Jorgensen, E. M. Visualizing presynaptic function. *Nature Publishing Group* **17,** 10–16 (2014).
103. Cheng, D. *et al.* Relative and absolute quantification of postsynaptic density proteome isolated from rat forebrain and cerebellum. *Mol Cell Proteomics* **5,** 1158–1170 (2006).
104. O'Rourke, N. A., Weiler, N. C., Micheva, K. D. & Smith, S. J. Deep molecular diversity of mammalian synapses: why it matters and how to measure it. *Nat Rev Neurosci* **13,** 365–379 (2012).
105. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489,** 391–399 (2012).
106. Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498,** 236–240 (2013).
107. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343,** 193–196 (2014).
108. Romanova, E. V., Aerts, J. T., Croushore, C. A. & Sweedler, J. V. Small-volume analysis of cell-cell signaling molecules in the brain. *Neuropsychopharmacology* **39,** 50–64 (2014).
109. Portales-Casamar, E. *et al.* A regulatory toolbox of MiniPromoters to drive selective expression in the brain. *Proceedings of the National Academy of Sciences* **107,** 16589–16594 (2010).
110. Heiman, M. *et al.* A translational profiling approach for the molecular characterization of CNS cell types. *Cell* **135,** 738–748 (2008).
111. 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* **110,** 9511–9516 (2013).
112. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422,** 198–207 (2003).
113. Nesvizhskii, A. I., Vitek, O. & Aebersold, R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Meth* **4,** 787–797 (2007).
114. MacDonald, M. L. *et al.* Biochemical fractionation and stable isotope dilution liquid chromatography-mass spectrometry for targeted and microdomain-specific protein quantification in human postmortem brain tissue. *Mol Cell Proteomics* **11,** 1670–1681

(2012).
115. Stergachis, A. B., MacLean, B., Lee, K., Stamatoyannopoulos, J. A. & MacCoss, M. J. Rapid empirical discovery of optimal peptides for targeted proteomics. *Nat Meth* **8,** 1041–1043 (2011).
116. Venable, J. D., Dong, M.-Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Meth* **1,** 39–45 (2004).
117. Röst, H. L. *et al.* OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol* **32,** 219–223 (2014).

Figure Legends

Figure 1: Schematised representation of the various MS/MS analysis options

Label-free MS/MS can be performed either **a)** on single samples for the relative quantitation of proteins/peptides to all other proteins/peptides other, or **b)** on collections of samples in which protein/peptide abundance is compared between conditions to identify the subset that exhibit significantly different abundance. In either case, protein is extracted from cells/tissue, which is enzymatically digested to peptides, and separated by liquid chromatography (LC) before being injected into the first round of mass-spectrometry (MS1) where the peptide intensity is monitored over time to infer its original abundance. All of these steps occur independently for each sample in label-free quantitation, as does the second round of mass-spectrometry (MS2) in which sufficiently abundant peptides, determined in MS1, are subjected to collision induced dissociation (CID) and the resulting fragment masses are used to identify the sequence of the peptide. In label-free the integration/comparison between samples occurs at the very end in downstream data analysis, which makes it a very flexible option for proteomic analysis.

Labelled MS/MS analyses enable more accurate determination of the relative abundances of peptides between pairs or groups of samples (depending on the number of labels used). In metabolic labelling **c)** such as SILAC cells/tissues are grown in media containing heavy, normal, or light forms of amino acids that are then incorporated into all proteins produced by the cell. Given this intrinsic labelling, protein extracts from each sample can be combined in equal amounts and differences in the abundance in individual proteins/peptides can be observed directly in the MS1 quantitation and MS2 identification. In the case of chemical labelling **d)** such as iTRAQ, proteins and peptides are obtained separately from each sample, before addition of an isobaric reporter ion that incorporates into the digested peptides from each sample. Labelled peptides from different samples are then combined and the identification and quantitation of the abundance of reporter ion is performed in MS2, which reflects the relative abundance of the peptide in each sample.

Figure 2: Integrated genomic, transcriptomic, and proteomic analyses and the central dogma

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. Shown in the figure is how 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

Rob 7/5/14 12:25 PM
Formatted: Justified

Rob 7/5/14 12:01 PM
Formatted: Font:16 pt, Bold

Rob 7/5/14 12:02 PM
Formatted: Font:Bold

Rob 7/5/14 12:30 PM
Formatted: Justified, Indent: Left: 0 cm, First line: 0 cm, Tabs: 0 cm, Left + Not at 1.41 cm

Rob 7/5/14 12:23 PM
Formatted: Font:Not Bold

Rob 7/5/14 12:04 PM
Formatted: Font:Bold

Rob 7/5/14 12:04 PM
Formatted: Font:Bold

Rob 7/5/14 12:25 PM
Formatted: Justified

Rob 7/5/14 12:02 PM
Formatted: Font:Bold

Rob 7/5/14 1:19 PM
Formatted: Justified, Indent: Left: 0 cm, First line: 0 cm, Tabs: 0 cm, Left + Not at 1.41 cm

enable the detection of allelic imbalance, in addition to information relating to alternative splicing (AS) and RNA-eding (RDD). This cascade of information in a given individual finally aids in the analysis of MS/MS 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 is it possible to compare isoform abundance and allelic imbalance with the values estimated at the RNA-level. Such combined data allows inferences to be made based on the proteomic data, such as the presence or absence of a particular protein isoform.