Title: Insights into the molecular changes associated with postnatal human brain development: an integrated transcriptomics and proteomics study

Authors: Michael S. Breen[1,2,3†*], Sureyya Ozcan[4†], Jordan M. Ramsey[4], Nitin Rustogi[4], Michael G. Gottshalk[4], Maree J. Webster[5], Cyndi Shannon Weickert[6], Joseph D. Buxbaum[1,2,3‡], Sabine Bahn[4‡*]

Affiliations:

[1]Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA; [2]Department of Genetic and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA; [3]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York, USA; [4]Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, United Kingdom; [5]Stanley Medical Research Institute, Laboratory of Brain Research, Rockville, Maryland, USA; [6]Neuroscience Research Australia, Schizophrenia Research Institute and University of New South Wales, Sydney, NSW, Australia.

[†]Both first authors equally contributed to this work.

[‡]Both last authors equally contributed to this work.

[*]Corresponding authors.

Corresponding Authors:

Michael S. Breen[1]
Department of Psychiatry
Icahn School of Medicine at Mount Sinai
Annenberg Building
1470 Madison Ave. 22nd Floor
New York, NY, 10129, USA
Tel: +1 (212) 824-8962
michael.breen@mssm.edu

Sabine Bahn[3]
Dept. of Chemical Engineering and Biotechnology
Philippa Fawcett Drive
Tennis Court Road
Cambridge, CB3 0AS, UK
Tel:+44 1223 334160
sb209@cam.ac.uk

Keywords: protein, mRNA, unbiased, dorsolateral prefrontal cortex, neurodevelopment

ABSTRACT

Healthy cortical development depends on precise regulation of transcription and translation. However, the relationships between levels of mRNAs and their respective translated proteins have not yet been examined comprehensively in the developing human brain. We surveyed the proteomic landscape of 69 dorsolateral prefrontal cortex samples across seven stages of postnatal life and integrated these data with paired transcriptome data. We detected 911 proteins by liquid chromatography mass-spectrometry, and 83 were significantly associated with postnatal age (FDR $P<0.05$). Network analysis revealed five modules of co-regulated proteins, three of which correlated with age, including two modules with increasing expression implicated in gliogenesis and NADH-metabolism and one neurogenesis-related module with decreasing expression throughout development. These age-related protein modules overlapped with RNA modules and displayed collinear developmental trajectories between paired RNA and protein expression. We also found that the correspondence between mRNAs and proteins decrease as a function of age, especially for genes involved in neurogenesis and cytoskeleton organization. Finally, intellectual disability and developmental delay genetic risk loci converged on myelination and ATP-metabolism-related modules in the proteome and transcriptome, consistent with their neuropathological relevance. Collectively, these findings reveal dynamic aspects of protein regulation in the human prefrontal cortex, providing new insights into development, maturation and disease.

INTRODUCTION

The human prefrontal cortex plays a critical role for higher cognitive processes, including executive function, social cognition and judgment, and has been implicated in the onset and progression of many, if not most, neurodevelopmental disorders[1-6]. The development of a properly functioning prefrontal cortex depends upon the proliferation and signaling of several cell types as well as the reprograming of transcriptional and translational pathways that unfold over the first two-three decades of postnatal life[7]. During this time, the brain quadruples in size, and grows through interneuronal genesis and maturation, glial multiplication, myelination, formation of new synaptic connections and pruning of unused synaptic connections[8-10]. Such processes are orchestrated by thousands of molecules in a tightly synchronized spatiotemporal fashion, and the disruption of any one of which may result in loss of cortical integrity and homeostasis[10], leading to cognitive deficits seen in patients with neurodevelopmental abnormalities. Therefore, understanding the molecular factors governing long-term brain developmental in normal individuals is critical for the identification of neurodevelopment mechanisms and developmental vulnerability periods.

Transcriptome-wide screens of postmortem brain tissue coupled with network-based approaches have provided a systematic view of the dynamic molecular landscape across brain development and between brain regions. Initial reports of the developing human brain transcriptome revealed marked changes across development and aging, with the largest gene expression changes occurring prenatally and during infancy and early childhood[11-14]; ages when many neurodevelopmental disorders become clinically recognizable. In parallel, several studies have identified developmental transcriptional networks with regional and cell type specific expression patterns enriched within known neurodevelopmental genetic risk loci[15-18], providing mechanistic insights of how mutations in risk genes might perturb typical brain development. However, while these reports demonstrate that different genes and transcriptional networks are temporally specific, our understanding of the developing human brain proteome is limited. Indeed, proteins are the main functional components in all cells and several studies consistently report that levels of messenger RNA and their respective translated proteins often correlate poorly[19-22]. Further, most brain proteome studies are unable to infer long-term time-dependent protein expression patterns across decades of postnatal development. As such, a critical remaining question is how the brain proteome unfolds throughout development, and ultimately how this information may inform brain mechanisms governing health and disease.

The current investigation integrated liquid chromatography mass-spectrometry (LC-MS) shotgun proteomics with paired transcriptome-wide microarray screens applied to 69 dorsolateral prefrontal cortex (DLPFC) samples from healthy individuals aged 3 weeks to 49.5

years. A multistep analytic approach was used that specifically sought to: (1) identify proteins and networks of highly correlated proteins significantly associated with distinct developmental stages and that changed with human age; (2) determine the biological organization of the proteome across postnatal development and clarify the relationship between protein levels and their corresponding mRNA levels in the prefrontal cortex; (3) map known neurodevelopmental genetic risk loci onto RNA- and protein-based networks to better understand through which pathways mutations in risk genes may perturb healthy brain development; and (4) provide a critical resource for the neuroscience community aimed at accelerating our understanding of the temporally-defined molecular mechanisms governing typical and pathological prefrontal cortex development.

RESULTS

Proteome organization in the developing dorsolateral prefrontral cortex

We applied a shotgun proteomics approach to measure temporal protein abundance in 69 DLPFC samples (Brodmann area 46) from normal individuals, aged 6 weeks to 49.5 years of age. Following standardized data pre-processing (**Supplementary Figure 1**), we detected 911 proteins and 584 were present in at least 50% of our sample, with a substantial amount of protein abundance variation explained by age relative to other technical or biological factors (**Fig. 1 A-C**). Next, we sought to identify proteins that were significantly regulated as a function of age and identified 83 proteins (FDR $p<0.05$), including 66 with decreasing abundance and 17 with increasing abundance across postnatal stages (**Supplementary Table 1**). The top ten most significantly increased and decreased age-related proteins are displayed in **Table 1**. Several significant age-related proteins mapped to known neurodevelopmental genetic risk loci, including genetic loci implicated in autism spectrum disorder (ASD) (*ANK2, YWHAE, L1CAM, FABP5, NRCAM*), intellectual disability (*L1CAM, PLP1, PSAP*, *QDPR)* and schizophrenia *(NCAN, ALCAM, GNAO1, PSAP, NFASC*). In addition, we also observed that the neonatal time period (38-89 days) explained the largest fraction of protein level variability according to age, including 131 neonatal-related proteins (FDR $p<0.05$) strongly enriched for central nervous system development, neurogenesis and gliogenesis (**Supplementary Figure 2, Supplementary Table 1**).

To characterize the temporal organization of the proteome we ran unsupervised weighted correlation network analysis and identified five modules of co-regulated proteins, three of which were significantly associated with development and age (**Fig. 1D, Supplementary Figures 3-4**). Two age-related modules displayed increasing expression throughout development (FDR $P<0.05$) and were enriched for processes related to myelination and gliogenesis (M1, 82 proteins, $r=0.71$, $p=1E-11$) and gluconeogenesis and NADH metabolism (M2, 30 proteins,

r=0.58, *p*=2E-07) (**Fig. 1E-G**). One age-related module displayed decreasing expression across development and was implicated in axonogenesis, neurogenesis and cytoskeleton organization (M4, 198 proteins, *r*=-0.77, *p*=2E-14) (**Fig. 1H**). The remaining two modules were not significantly associated with age nor with any other developmental stages or technical factors, and were enriched for functions related to cellular respiration and ATP metabolic processes (M3, 194 proteins) and nucleotide metabolism and oxidation-reduction processes (M5, 80 proteins) (**Supplementary Figures 4**). Collectively, all five modules were significantly enriched for direct protein-protein interactions (PPI), beyond what would be expected by chance (**Supplementary Table 2**), and PPI networks were constructed for each module (**Supplementary Figure 5-6**). Densely connected hub proteins for each age-related module included, *MBP* and *PGM1* for module M1, *ENO2* and *MDH1* for module M2 and *UBC* and *HSP90AA1* for module M4. Further, we employed cell type deconvolution on global proteomic abundance and identified substantial decreases in neuronal cell populations paralleled by increases in oligodendrocytes throughout development (**Fig. 1I**), results which correlated with transcriptome-based signatures (**Supplementary Figure 7**).

Correspondence between transcriptome and proteome organization

A total of 556 common HGNC symbols were detected between our proteomic and transcriptome-wide gene expression assays (**Fig. 2A**). Similar to the proteome, the largest amount of gene expression variability was explained by age, as compared to any other factor (**Supplementary Figure 1**). Comparably, the neonatal time period (38-89 days) also explained the largest fraction of transcriptome variation by age, albeit to a lesser extent than in the proteome, including genes primarily involved in ATP metabolic processes (**Supplementary Figure 2**). Linear regression analyses identified ~11.5% of the transcriptome was associated with postnatal development, including 1,145 genes with increasing expression and 1,181 genes with decreasing expression across all developmental stages (FDR $p<0.05$). (**Supplementary Figure 2, Supplementary Table 3**). These age-related genes also included several known neurodevelopmental genetic risk loci implicated in ASD (*LRP1, RNF135, YWHAE*), schizophrenia (*TEKT4, LRP1, DNAH1, BRSK1, INTS1, ZC3H10, METTL14*) and developmental delay (*SCYL1, PIGQ, OBSL1, SMARCB1, CEP135, SPG11, TAF1, TAT, FAM126A, RAD21*).

Weighted correlation network analysis identified eight modules of co-regulated genes (**Fig. 2B**), which displayed a high degree of reproducibility compared to existing postmortem BrainSpan data (**Supplementary Figure 8**). Four of the eight modules were significantly associated with postnatal development, including two modules implicated in gliogenesis (M4_t, 2624 genes, *r*=0.54, *p*=2E-04) and ATP metabolic processes (M6_t, 5078 genes, *r*=0.72, *p*=4E-05) with increasing expression, and two modules involved in neurogenesis (M1_t, 3046

genes, r=-0.71, p=8E-08) and synaptic signaling (M2_t, 403 genes, r=-0.53, p=2E-04) with decreasing expression throughout postnatal development (**Supplementary Figure 9**). One immune response-related module was significantly associated to the toddler postnatal age group (1.2-5.1 years) (M7_t, 384 genes, r=0.52, p=3E-04). Interestingly, a number of significant overlaps were identified between transcriptome modules and proteome modules (**Fig. 2B**). The majority of overlapping RNA- and protein-based modules also displayed a high degree of collinearity (**Fig. 2C**) and shared similar biological functions, including modules involved in gliogenesis and myelination (M1, M4_t), ATP metabolic processes (M2, M3, M5, M6_t) and neurogenesis (M4, M1_t) (**Fig. 2D-G**). Overall, all five protein modules were well represented at the RNA level (**Fig. 2G**).

Correspondence between mRNA and protein levels throughout postnatal development

We further quantified the association between the transcriptome and the proteome using paired data, first at the global level and then according to the five previously identified protein modules (**Fig. 3A**). At the global level we found weak-to-moderate correspondence within paired samples (r=0.15-0.40), with the highest correlations observed for RNA and protein products involved in gliogenesis and myelination (M1, r=0.26-0.53) and the lowest correlations for RNA and protein products involved in nucleotide and ATP metabolic processes (M5, r=-0.05-0.22). Subsequently, we explored these associations as a function of age and found that the correspondence between mRNAs and their respective translated proteins decreases throughout postnatal development (r=-0.56, p=6.4E-05), especially for RNAs and proteins involved in processes of myelination and gliogenesis (M1, r=-0.30, p=0.04) and neurogenesis and cytoskeleton organization (M4, r=-0.67, p=3.9E-07) (**Fig. 3B**). These within sample differences between expression of mRNAs and their respective translated proteins prompted us to investigate the degree of conservation in protein-based co-expression networks at the RNA level, for which we found no preservation for co-regulated RNAs implicated in nucleotide metabolic processes (M5) and gliogenesis (M1) (**Fig. 3C**). We next examined the relationships between the 566 RNA-protein pairs across human age. A high level of correspondence was observed when comparing age-related linear regression results (t-statistics) computed separately for individual RNAs and proteins (r=0.62, p=2.4E-60) (**Fig. 3D**), and a total of 47 RNA-protein pairs displayed a significant association across postnatal age (FDR p<0.05) (**Supplementary Table 4**).

Cell type and genetic risk loci enrichment analysis

We next sought to determine whether the transcriptome- and proteome-based modules were strongly linked to the underlying cellular architecture in the developing DLPFC using previously defined cell type specific markers at the RNA and protein levels (**Fig. 4A**). As expected,

several protein-based modules were significantly enriched for known cell type specific markers. Protein module M1 was significantly enriched for oligodendrocyte cell type markers (∩=12, *p*=0.001) and was nominally enriched for astrocyte cell type markers (∩=13, *p*=0.05). Protein module M4 displayed significant over-representation of neuronal cell type markers (∩=28, *p*=2.9E-06). In parallel, several transcriptome-based modules displayed significant enrichment for brain specific cell type markers, including RNA module M4_t enriched for oligodendrocyte (∩=76, *p*=5.0E-04) and astrocyte markers (∩=137, *p*=2.6E-24), module M3_t enriched for astrocyte markers (∩=57, *p*=3.4-E34) and module M7_t enriched for microglial markers (∩=148, *p*=9.2E-91).

Subsequently, we leveraged several lists of known neurodevelopmental genetic risk loci and performed developmental network enrichment analysis using our identified transcriptome- and proteome-based modules as background lists (**Supplementary Figure 9**). Our analysis, and user-friendly software tool (*see Methods for details*), identified points of convergence for intellectual disability and developmental delay genetic risk loci in gliogenesis and ATP metabolic-related modules in both the proteome (M1 and M3) and the transcriptome (M4_t, M6_t) (Bonferroni $p<0.05$) (**Fig. 4B**). We also found epilepsy genetic risk loci concentrated primarily in oxidative phosphorylation-related modules in the proteome surviving multiple adjusted significance in the proteome (M3). While we found strong enrichment for several ASD genetic risk loci in neurogenesis-related module M1_t in the transcriptome, no single proteome-based module contained enrichment for ASD risk loci, possibly a result of the lower level of protein detection relative to transcript detection (**Fig. 2A**).

DISCUSSION

Proteins are the functional components of cells in the CNS, however our understanding of the brain proteome continues to lag behind the pace of transcriptome discovery, mainly because the necessary technology has only recently matured to enable improved protein detection and coverage. To provide a foundation for an age-dependent brain proteome map, we performed in-depth proteome analysis of the temporal DLPFC (BA46) in humans. Our analysis resulted in the largest collection thus far of protein expression data in the developing human DLPFC and catalogued widespread changes in proteins and protein networks associated with distinct developmental stages and human age. By integrating these data with paired transcriptome data, we were able to investigate the higher-order organization of the developing brain transcriptome and proteome, as well as study the relationships between individual RNA levels compared to their respective protein levels across postnatal development. Finally, we examined RNA and protein networks enriched for neurodevelopmental genetic risk loci, to gain insight into how mutations in risk genes may perturb molecular pathways during healthy brain development. We discuss these points in turn below.

The temporal expression patterns of many proteins in this study are likely driven by a shifting cellular CNS landscape, as reflected by the observed neuron-glia oscillations. Our analysis revealed an inverse correlation between protein module M1, associated with myelination, gliogenesis and oligodendrocyte cell type specificity (increasing), relative to protein module M4, implicated in axonogenesis, neurogenesis and neuronal cell types (decreasing)[23]; a result also echoed by cell type deconvolution analysis. Though the cell division and migration of neurons are largely prenatal events, neurogenesis is known to persist throughout adult life, albeit to a limited level and produce only a small fraction of the neuronal population[24,25]. In contrast, proliferation and migration of glial progenitors, while beginning prenatally, continue for a protracted period as oligodendrocytes and astrocytes differentiate. Oligodendrocyte progenitor cells begin to differentiate by increasing myelin protein expression, as evident in the current study, and ultimately form tightly wrapped multi-layered sheaths around nearby axons. However, much uncertainty exists regarding the extent of postnatal proliferation, migration and differentiation, and about the timing of these processes relative to each other[24,25]. Our results suggest that the greatest degree of change occur between 3-15 years of age, and that these neuron-glia changes appear to play an important role in the functional organization of neural circuits during postnatal life. Further, we also observed marked increases in discretely co-regulated proteins involved in NADH metabolism and gluconeogenesis (M2) across development, consistent with the well-known energy requirements of the brain[26]. Previous work by us and others suggests that myelination is also a major energy-demanding process in the brain[27,28], especially during postnatal life. For example, myelin synthesis is an ATP-dependent process and oligodendrocytes often oxidize glucose at higher rates than neurons[29], supporting these distinct changes in protein modules across time. Moreover, glucoregulatory abnormalities, oxidative stress vulnerability and oligodendrocyte dysfunction have been prominently linked to both neuropsychiatric and neurodegenerative disorders[30,31]. Such pathways have also been implicated in the brain's high demand of glucose, which ultimately requires ATP-dependent transport across the blood-brain barrier (review[32]).

In the context of the temporally dynamic expression profiles, a fundamental question is whether RNAs and their respective translated proteins correlate throughout postnatal development. We observed Pearson correlation coefficients within paired transcriptome and proteome samples between 0.15 and 0.40. Indeed, numerous RNAs and proteins correspond significantly across postnatal development, while others do not (**Supplementary Table 4**). Several studies have also found similar low correlations in human[19,20] and murine[33] tissue. These discrepancies may be due to well-known differences in the regulation, localization and functions of mRNA and proteins, or alternatively controlled at the level of transcriptional on/off dimer switches. Moreover, we found that the correspondence of RNA and protein products within paired samples decreased as a function of age (r=-0.56), especially for those involved in

neurogenesis and cytoskeleton organization (M4, r=-0.67) and myelination and gliogenesis (M1, r=-0.30). The efficiency of myelination decreases with age, a process largely regulated by age-dependent epigenetic control of gene expression[34]. That is, during infancy, myelin synthesis is preceded by down-regulation of oligodendrocyte differentiation inhibitors, and this is associated with recruitment of histone deacetylase to promoter regions; a process that becomes less efficient in adulthood and ultimately prevents a successive surge in myelin gene expression[34,35]. In addition to the loss of epigenetic memory as a mechanism behind poor correlation of RNAs and proteins involved in myelination, post-translational events also appear to be an important factor in myelin production[36], which may explain some of these discrepancies.

One important similarity across the brain transcriptome and proteome was the consistent mapping of intellectual disability and developmental delay genetic risk loci to modules enriched for myelination, gliogenesis, and ATP metabolic processes. These modules displayed a collinear patter of expression between RNA and protein products, peaking in expression during adolescence and adulthood (**Fig. 2D-F**). As there is a close interdepency between myelin synthesis and ATP-dependent processes, a disorder affecting one of the two inevitably also leads to disturbance of the other. Indeed, defective myelination and ATP processes have been reported as key factors causing pathogenic processes involved in these disorders[15,37,38], and these data provide further substantial evidence in the broader context of long-term brain development. In parallel, ASD variants resided primarily in neuronal-based modules in the transcriptome, and not in the proteome. These results echo recent large transcriptome network studies that demonstrate the involvement of neuronal and synaptic processes involved in ASD[12,16,17]. We also mapped epilepsy genetic risk loci to oxidative phosphorylation-related modules in the proteome, consistent with growing evidence that deficits in oxidative phosphorylation complexes can result in increased oxygen and free-radical release likely implicated in the initiation and progression of epilepsy[39].

Our study also has some limitations. First, the temporal brain proteome presented here may represent an incomplete picture of the entire proteome. We applied liquid chromatography-quadrupole time-of-flight mass spectrometry and detected 911 proteins, and limited our analysis to a high-confidence set of 586 proteins, which were detected in at least half of our samples. Of this high-confidence set, 566 were represented at the mRNA level, a marginal 3.5% of the detected transcriptome. Surprisingly, however, with this level of detection we were still able to capture a considerable amount of protein variability and functionality across postnatal development compared to paired transcriptome data. That is, all of the age-related transcriptome modules (M1_t, M2_t, M4_t, M6_t: 11,150 genes total), which comprised 55.4% of the observed transcriptome, were well represented at a higher functional level in the proteome, even though fewer proteins were detected. Nonetheless, it is possible that future

reports applying deeper and more sensitive analytical techniques will enable both greater proteome coverage with fewer missing values, though we recognize that post-translational modifications will continue to present a major challenge toward resolving of a truly comprehensive and functional brain proteome. A second caveat to these data is the lack of prenatal samples, developmental stages when gene expression patterns appear to be most dynamic. For example, vast increases in expression for synapse and dendrite development genes occur prenatally and taper off in the first decade of postnatal life[10,40], and went undetected in the current investigation likely due to the lack of prenatal samples. However, in contrast to previous studies, we were able to capitalize on larger postnatal developmental group sizes (2.5× larger), thus increasing our ability to identify biologically meaningful age-related proteins and protein networks. As these concerns are addressed in the future, it will be possible to reveal further insights into the transcriptional and translational foundations of human brain development.

In sum, by taking an unbiased global approach to understanding the molecular dynamics of the developing DLPFC, we outlined the similarities and differences between mRNA and protein level expression. Although a component of the observed differences in regulation is likely to be technical, particularly with respect to the level of protein detection, it is clear that proteomic and transcriptomic data provide both overlapping and non-overlapping information. The various proteins discussed are likely to be candidates for further functional and/or synaptic developmental studies. This work also provides a framework to reach a better understanding of the control of RNA and protein expression throughout postnatal development.

METHODS

Post-mortem brain sample ascertainment

The current study analyzed fresh frozen postmortem prefrontal cortex tissue (BA46) from 69 individuals varying in age from 6 weeks to 49 years. All samples were obtained from the National Child Health and Human Development Brain and Tissue Bank for Developmental Disorders at the University of Maryland, Baltimore, USA (UMBB). All subjects were defined as healthy individuals by forensic pathologists at UMBB, having no history of psychiatric or neurological complaints, also confirmed by next of kin interviews. Samples which underwent LC-MS (*N*=69), comprised 41 males and 28 females and relevant covariates were collected: pH (6.6 ± 0.2), postmortem interval (PMI; 17.1 ± 7.0 hrs.) and ethnicity (32 Caucasian, 36 African American, 1 Hispanic). A subset of these samples also underwent microarray gene expression profiling (*N*=44), consisting of 27 males and 17 females, and covariates were recorded: pH (6.7 ± 0.15), PMI (16.9 ± 7.5 hrs.) and ethnicity (24 Caucasian, 20 African American). Full details can be found in the **Supplementary File 1**.

Tissue protein extraction

Proteins were extracted by sonicating each tissue sample (~70mg) in 350µl lysis buffer (7M urea, 2M thiourea, 4% 3-[(3-cholamidopropyl)dimethylammonio]- 1-propanesulfonate (CHAPS), 2% ASB14 and 70mM dithiotreitol (DTT), followed by sonication for 2 cycles of 15 seconds on ice using a Branson Sonifier 150 (Thistle Scientific; Glasgow, UK). A 50µl aliquot of each protein extract was precipitated for 4hr with 200µl of acetone at -20°C. The precipitates were centrifuged for 30 min at 13400 rpm, at 4°C, the acetone supernatant decanted and discarded. The resulting pellets were re-suspended in 200µl of 50mM ammonium bicarbonate (pH 8.0) and sonicated for 10 seconds. Once re-suspended, protein concentration was measured by Bradford Assay. An aliquot of each precipitated protein extract, equivalent to 100ug of protein, was reduced with 100mM DTT for 30 min at 60°C, alkylated with 200mM iodoacetamide at room temperature for 30 min in dark and digested with 4µl of 0.5µg/ul of modified sequencing grade trypsin at 37°C for 17hrs. Digestion reaction was stopped by adding 0.80µl 8.8M Hydrochloric acid (1:60). An aliquot of 5µl from each digested sample was pooled together to be used as standard[41]. A detailed schematic of tissue protein extraction can be found in the **Supplementary File 1**.

Liquid chromatography-mass spectrometry and protein quantification

Tryptic peptides were analyzed by a shotgun LC-MS approach using a 1290 Infinity LC coupled to Agilent 6550 iFunnel Q-TOF instrument (Agilent technology, USA). Peptide separation was carried out using an Agilent AdvanceBio Peptide column (2.1 μm x 250 mm, 2.7 μm) over a 90 min linear gradient of 3 to 45% ACN. The flow rate was 0.3mL/min and the column temperature was set to 50^0C. Peptides were then detected by quadrupole time-of-flight (Q-TOF) MS operated in positive mode. Acquisition was in data-dependent mode over m/z 300–1700. The top 10 precursor ions were scanned from 300-1700 and MS/MS from 50–1700. The precursor ions were then automatically isolated and fragmented using collision induced dissociation (CID) with a relative collision energy calculated using the formula, 3.6*(m/z)/100+-4.8. Data files were processed by Spectrum Mill Protein Identification software (Rev B.05.00.180, Agilent Technologies, USA). The protein identification was executed against the Swiss-Prot database (released in February 2015, Homo sapiens). Search parameters were as follows; precursor mass tolerance, 20 ppm; product ion mass tolerance, 50 ppm; maximum two missed cleavages allowed; digested by trypsin; fixed modification of carbamidomethyl cysteine; variable modifications of oxidized methionine. After MS/MS searching, auto-validation was carried out by calculating the false-discovery rate (FDR). A FDR threshold of 1.2 was applied. A detailed schematic of protein detection, identification, and quantification can be found in **Supplementary File 2**.

RNA isolation and microarray hybridization

All RNA procedures have previously been described[27]. Briefly, total RNA was extracted from dorsolateral prefrontal cortex samples using Trizol (Sigma-Aldrich, St. Louis, MO, USA) and RNA quality was assessed using a high-resolution electrophoresis system (Agilent Technologies, Santa Clara, CA, USA). Isolated total RNA was subjected to Affymetrix preparation protocol and each sample was hybridized to one HG-U133 Plus 2.0 GeneChip (Affymetrix, Santa Clara, CA, USA) to quantify transcriptome-wide gene expression.

Data pre-processing

All data pre-processing and statistical analyses were conducted in the statistical package R. First, all data were normalized to fit approximate normal distribution. Protein data were median scaled by all runs and log(e) transformed. Protein Uniprot IDs were converted to HGNC symbols using the Uniprot database (http://www.uniprot.org/uploadlists/). Microarray data were normalized using the robust multi-array average normalization with additional GC-correction (GCRMA)[42]. When multiple microarray probes mapped to the same HGNC symbol, the probe the highest average expression across all samples was used. Following normalization, all data were inspected for outlying samples using unsupervised hierarchical clustering (based on Pearson coefficient and average distance metric) and principal component analysis to identify potential outliers outside two standard deviations from these grand averages; no outliers were present in these data. Linear mixed models from the R package variancePartition[43] were used to characterize and identify biological and/or technical drivers that may affect the observed RNA and protein abundance. This approach quantifies the main sources of variation in each expression dataset attributable to differences in age, age group, gender, PMI, pH and ethnicity. Finally, to identify age-related genes and proteins, generalized linear models with Bonferroni multiple test correction were implemented. The covariates gender, pH, PMI and ethnicity were included in the models to adjust for their potential confounding influence on RNA and protein expression (lm(Age ~ Expression + PMI + gender + pH + ethnicity)). Further, a Spearman's correlation test was used to identify individual genes and proteins whose expression profile were significantly correlated with a pre-defined developmental stage or template (*e.g.* toddlers), which had been binarized (0 or 1) to quantify associations with expression.

Weighted correlation network analyses

Prior to network analysis, missing protein values were imputed using predictive mean matching in the MICE package[44] (number of multiple imputations, m=5; the number of iterations, *maxit*=50). A high confidence set of proteins detected in at least half of the sample were used

to make meaningful imputations. Weighted gene correlation network analysis[45] (WGCNA) was used to build signed co-expression networks independently for the transcriptome (*n*=20,122 genes) and proteome (*n*=584 proteins). To construct each network, the absolute values of Pearson correlation coefficients were calculated for all possible gene pairs (transcriptome data) and protein pairs (proteome data), and resulting values were transformed with an exponential weight (β) so that the final matrices followed an approximate scale-free topology (R^2). Thus, for each network we only considered powers of β that lead to a network satisfying scale-free topology (*i.e.* R^2>0.80), so the mean connectivity is high and the network contains enough information for module detection. The dynamic tree-cut algorithm was used to detect network modules with a minimum module size set to 30 and cut tree height set to 0.9999. The identified RNA and protein modules were inspected for association to age, as well as seven distinct postnatal stages and all recorded covariates. To do so, singular value decomposition of each modules expression matrix was performed and the resulting module eigengene (ME), equivalent to the first principal component, was used to represent the overall expression profiles for each module per sample. Modules were evaluated both quantitatively and qualitatively for expression patterns significantly associated with age (**Supplementary Figure 3**). Fisher's exact tests were used to assess the overlap of RNA and protein modules and correlations amongst RNA and protein ME's were explored using Pearson's correlation coefficients.

A series of module preservation analyses sought to determine whether (*i*) co-regulated modules of proteins are preserved at the RNA level and (*ii*) whether RNA modules are reproducible in independent BrainSpan data. We collected publically available BrainSpan data (http://www.brainspan.org/) and used only postnatal samples (*N*=17) to best reflect the developmental biology of our current sample (**Supplementary Figure 8**). For these analyses, module preservation was assessed using a permutation-based preservation statistic, $Z_{summary}$, implemented within WGCNA with 500 random permutations of the data[46]. $Z_{summary}$ takes into account the overlap in module membership as well as the density and connectivity patterns of genes within modules. A $Z_{summary}$ score <2 indicates no evidence of preservation, $2<Z_{summary}<10$ implies weak preservation and $Z_{summary}$ >10 suggests strong preservation.

Functional annotation and protein-protein interaction networks

All age-related RNAs and proteins identified through either linear regression or network-based analyses, were subjected to functional annotation using the ToppFun module of ToppGene Suite software[47]. We explored gene ontology terms related to biological processes and molecular factors using a one-tailed hyper-geometric tested (Benjamini-Hochberg FDR corrected) to assess the significance of the overlap. All terms must pass an FDR corrected *p*-value and a minimum of five genes/proteins per ontology were used as filters prior to pruning ontologies to less redundant terms.

The STRING database v9.1[48] was used to assess whether RNA and protein modules were significantly enriched for direct protein-protein interactions (PPIs) and to identify key genes/proteins mediating the regulation of multiple targets. For these analyses, our signature query of RNA or protein modules were used as input. STRING implements a scoring scheme to report the confidence level for each direct PPI (low confidence: <0.4; medium: 0.4-0.7; high: >0.7). We used a combined STRING score >0.04. Hub genes within the PPI network are defined as those with the highest degree of network connections. We further used STRING to test whether the number of observed PPIs were significantly more than expected by chance using a nontrivial random background model. For visualization, the STRING network was imported into CytoScape[49].

Cell type and genetic risk loci enrichment analyses

Cell type enrichment was performed by cross-referencing RNA and protein modules with previously defined lists of genes and proteins known to be preferentially expressed in different cell types. Lists of cell type specific genes from RNA-sequencing[50] and mass spectrometry-based proteomics[33] were obtained and filtered by a previous study[18]. In brief, if $\log_2$ gene/protein expression ≥1.4 units, a difference of 0.8 units above the next most abundance cell type measurement was required. Mouse homologues were identified using the mygene R package[51] and the total list of RNA and proteins for each module were used as background using a Fisher's exact test. Further, neurodevelopmental genetic risk loci were curated from human genetic studies of autism spectrum disorder[52-54], epilepsy[55], developmental delay (Gene2Phenotype)[56], intellectual disability[15] and schizophrenia[57]. Similarly, over-representation of genetic risk loci was determined using a Fisher's exact test with Bonferroni multiple test correction and the identified protein and gene modules as background. Lists of neurodevelopmental genetic risk can be found in **Supplementary Table 5**.

Cell type deconvolution

The frequencies of brain cell types were estimated for proteomic using Cibersort[58] cell type de-convolution (https://cibersort.stanford.edu/). Cibersort relies on known cell subset specific marker genes and applies linear support vector regression, a machine learning approach highly robust compared to other methods with respect to noise, unknown mixture content and closely related cell types. As input, we used a curated cell type specific protein signature matrix (as above[33,49]) to distinguish between neurons, oligodendrocytes, astrocytes and microglia. Notably, we were unable to obtain sufficient enough overlap for microglial markers based on protein detection to make meaningful predictions (**Supplementary Figure 7**).

Developmental network enrichment analysis and software availability

We developed a user-friendly perl script that enables researchers to quickly test for over-representation of a user-defined list of genes or proteins within our temporal transcriptome- and proteome-based modules (**Supplementary Figure 10**). The *perl* script, developmental network enrichment analysis (DNEA.pl), contains pre-stored lists of transcriptome- and proteome-based modules, including functional and age-related properties, and performs over-representation analysis using a hyper-geometric test based on a user-defined list of HNGC symbols. It is anticipated that this tool will encourage researchers to test a specific set of genes or proteins for over-representation in the identified temporal transcriptome and proteome-based modules. This software is available in the **Supplementary File 2**. Other specific software will be made available upon request.

Data availability

Gene expression data can be downloaded from GEO using accession GSE13564. Protein expression data is available upon request.

REFERENCES

1. Mitchell, S. et al. Neuroanatomic Alterations and Social and Communication Deficits in Monozygotic Twins Discordant for Autism Disorder. ***American Journal of Psychiatry*** 166, 917-925 (2009).
2. Lesh, T., Niendam, T., Minzenberg, M. & Carter, C. Cognitive Control Deficits in Schizophrenia: Mechanisms and Meaning. ***Neuropsychopharmacology*** 36, 316-338 (2010).
3. Arnsten, A. Stress signalling pathways that impair prefrontal cortex structure and function. ***Nature Reviews Neuroscience*** 10, 410-422 (2009).
4. Gamo, N. & Arnsten, A. Molecular modulation of prefrontal cortex: Rational development of treatments for psychiatric disorders. ***Behavioral Neuroscience*** 125, 282-296 (2011).
5. Miller, E. The prefrontal cortex and cognitive control. ***Nat Rev*** 1: 59–65 (2000).
6. Sowell, E., Thompson, P., Toga, A. Mapping changes in the human cortex throughout the span of life. ***Neuroscientist*** 10: 372–392 (2004).
7. Ziats, M., Grosvenor, L., Rennert, O. Functional genomics of human brain development and implications for autism spectrum disorders. ***Translational Psychiatry*** 5, e665 (2015).
8. Nogueira, A., Nogueira, A., Veiga, J. & Teixeira, M. Extensive Migration of Young Neurons Into the Infant Human Frontal Lobe. ***Neurosurgery*** 81, E16-E18 (2017).
9. Catts, V. et al. Rethinking schizophrenia in the context of normal neurodevelopment. ***Frontiers in Cellular Neuroscience*** 7, (2013).
10. Kang H *et al.* Spatio-temporal transcriptome of the human brain. ***Nature*** 478, 483–489 (2011).
11. Colantuoni, C. *et al.* Temporal dynamics and genetic control of transcription in the human prefrontal cortex. ***Nature*** 478, 519–523 (2011).
12. Ben-David, E. & Shifman, S. Networks of neuronal genes affected by common and rare variants in autism spectrum disorders. ***PLoS Genet.*** 8, e1002556 (2012).
13. Somel, M. et al. Transcriptional neoteny in the human brain. ***Proceedings of the National Academy***

of Sciences 106, 5743-5748 (2009).

14. Weickert, C. et al. Transcriptome analysis of male–female differences in prefrontal cortical development. ***Molecular Psychiatry*** 14, 558-561 (2009).

15. Parikshak, N. N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. ***Cell*** 155, 1008–1021 (2013).

16. Willsey, A. J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. ***Cell*** 155, 997–1007 (2013).

17. Steinberg, J. & Webber, C. The roles of FMRP- regulated genes in autism spectrum disorder: single- and multiple-hit genetic etiologies. ***Am. J. Hum. Genet****.* 93, 825–839 (2013).

18. Seyfried NT, Dammer EB, Swarup V, et al. A multi-network approach identifies protein-specific co-expression in asymptomatic and symptomatic Alzheimer's disease. ***Cell Syst.*** 4(1):60–72.e4 (2017).

19. Nagaraj, N. *et al.* Deep proteome and transcriptome mapping of a human cancer cell line. ***Mol. Syst. Biol.*** 7, 548 (2011).

20. Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. ***Nature*** 509, 582–587 (2014).

21. Ernst, A. & Frisén, J. Adult Neurogenesis in Humans- Common and Unique Traits in Mammals. ***PLOS Biology*** 13, e1002045 (2015).

22. Beltaifa, S. et al. Discordant changes in cortical TrkC mRNA and protein during the human lifespan. ***European Journal of Neuroscience*** 21, 2433-2444 (2005).

23. Weissleder, C. et al. Decline in Proliferation and Immature Neuron Markers in the Human Subependymal Zone during Aging: Relationship to EGF- and FGF-Related Transcripts. ***Frontiers in Aging Neuroscience*** 8, (2016).

24. Stiles, J. & Jernigan, T. The Basics of Brain Development. ***Neuropsychology Review*** 20, 327-348 (2010).

25. Cayre, M. *et al.* Cell migration in the normal and pathological postnatal mammalian brain. ***Progress in Neurobiology*** 88 (1), 41–63 (2009).

26. Bélanger, M., Allaman, I. & Magistretti, P. Brain Energy Metabolism: Focus on Astrocyte-Neuron Metabolic Cooperation. ***Cell Metabolism*** 14, 724-738 (2011).

27. Harris, L. *et al.* Gene expression in the prefrontal cortex during adolescence: Implications for the onset of schizophrenia. ***BMC Med Genomics*** 2:28 (2009).

28. Bizzozero, O., Sanchez, P. & Tetzloff, S. Effect of ATP Depletion on the Palmitoylation of Myelin Proteolipid Protein in Young and Adult Rats. ***Journal of Neurochemistry*** 72, 2610-2616 (2002).

29. Fields, R. A new mechanism of nervous system plasticity: activity-dependent myelination. ***Nature Reviews Neuroscience*** 16, 756-767 (2015).

30. Gottschalk, M., Wesseling, H., Guest, P. & Bahn, S. Proteomic Enrichment Analysis of Psychotic and Affective Disorders Reveals Common Signatures in Presynaptic Glutamatergic Signaling and Energy Metabolism. ***International Journal of Neuropsychopharmacology*** 18, (2015).

31. Wesseling, H., Gottschalk, M. & Bahn, S. Targeted Multiplexed Selected Reaction Monitoring Analysis Evaluates Protein Expression Changes of Molecular Risk Factors for Major Psychiatric Disorders. ***International Journal of Neuropsychopharmacology*** 18, (2014).

32. Harris, L. et al. Schizophrenia: Metabolic aspects of aetiology, diagnosis and future treatment strategies. ***Psychoneuroendocrinology*** 38, 752-766 (2013).

33. Sharma, K. *et al.* Cell type- and brain region-resolved mouse brain proteome. ***Nature Neuroscience*** 18, 1819–1831 (2015)

34. Shen, S. et al. Age-dependent epigenetic control of differentiation inhibitors is critical for remyelination efficiency. ***Nature Neuroscience*** 11, 1024-1034 (2008).

35. Nave, K. Neuroscience: An ageing view of myelin repair. *Nature* 455, 478-479 (2008).

36. Campagnoni, A. & Macklin, W. Cellular and molecular aspects of myelin protein gene expression. ***Molecular Neurobiology*** 2, 41-89 (1988).
37. Vissers, L., Gilissen, C. & Veltman, J. Genetic studies in intellectual disability and related disorders. ***Nature Reviews Genetics*** 17, 9-18 (2015).
38. Zoghbi, H. & Bear, M. Synaptic Dysfunction in Neurodevelopmental Disorders Associated with Autism and Intellectual Disabilities. ***Cold Spring Harbor Perspectives in Biology*** 4, a009886-a009886 (2012).
39. Shin, E. et al. Role of oxidative stress in epileptic seizures. ***Neurochemistry International*** 59, 122-137 (2011).
40. Webster, M., Elashoff, M. & Weickert, C. Molecular evidence that cortical synaptic growth predominates during the first decade of life in humans. ***International Journal of Developmental Neuroscience*** 29, 225-236 (2011).
41. Broek, J., Guest, P., Rahmoune, H. & Bahn, S. Proteomic analysis of post mortem brain tissue from autism patients: evidence for opposite changes in prefrontal cortex and cerebellum in synaptic connectivity-related proteins. ***Molecular Autism*** 5, 41 (2014).
42. Gharaibeh, R., Fodor, A. & Gibas, C. Background correction using dinucleotide affinities improves the performance of ***GCRMA. BMC Bioinformatics*** 9, 452 (2008).
43. Hoffman, G. & Schadt, E. variancePartition: interpreting drivers of variation in complex gene expression studies. ***BMC Bioinformatics*** 17, (2016).
44. Buuren, S. & Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations inR. ***Journal of Statistical Software*** 45, (2011).
45. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. ***BMC Bioinformatics*** 9, 559 (2008).
46. Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module preserved and reproducible? ***PLoS Computational Biology*** 7, e1001057 (2011)
47. Chen, J., Bardes, E., Aronow, B. & Jegga, A. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. ***Nucleic Acids Research*** 37, W305-W311 (2009).
48. Franceschini, A. et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. ***Nucleic Acids Research*** 41, D808-D815 (2012).
49. Shannon, P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. ***Genome Research*** 13, 2498-2504 (2003).
50. Zhang et al., An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. ***Journal of Neuroscience*** 35, 864-866 (2015).
51. Mark A, Thompson R, Afrasiabi C and Wu C. *mygene: Access MyGene.Info_ services*. R package version 1.12.0 (2014).
52. Xu, L. et al. AutismKB: an evidence-based knowledgebase of autism genetics. ***Nucleic Acids Research*** 40, D1016-D1022 (2011).
53. De Rubeis, S. *et al*. Synaptic, transcriptional and chromatin genes disrupted in autism. ***Nature*** 515, 209–215 (2014).
54. Sanders, S. et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. ***Neuron*** 87, 1215–1233 (2015).
55. EuroEPINOMICS-RES Consortium, Epilepsy Phenome/Genome Project, Epi4K Consortium. De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. ***Am J Hum Genet*** 95, 360–370 (2014).
56. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), {June 1, 2017}. World Wide Web URL: https://omim.org/

57. Fromer, M. *et al*. Gene expression elucidates functional impact of polygenic risk for schizophrenia. ***Nature*** 506, 179–184 (2014).
58. Newman, A. et al. Robust enumeration of cell subsets from tissue expression profiles. ***Nature Methods*** 12, 453-457 (2015).

Acknowledgements. This study was supported by the Stanley Medical Research Institute. M.S.B. is funded by the Autism Science Foundation (Grant no. 17-001) and the Seaver Autism Center for Research and Treatment.

Author contributions. S.O. and S.B. conceived the study and designed the experiments. M.J.W. and C.S.W. selected and obtained the brain tissue samples. M.G.G. and N.R. prepared the samples for proteomics analysis. S.O. and J.M.R. pre-processed the data. M.S.B. processed and analyzed the data. M.S.B. and S.O. wrote the manuscript with input from S.B. and J.D.B. M.J.W. and C.S.W. also provided critical input regarding data interpretation and editing of manuscript. All other co-authors read and approved the final manuscript.

Competing financial interests. S.B. is a director of Psynova Neurotech Ltd and PsyOmics Ltd.

Table 1. Top ten up and down age-related proteins throughout postnatal development with paired RNA products.

Uniprot ID	HUGO gene symbol (name)	Proteome		Transcriptome	
		***t*-statistic**	***P*-value**	***t*-statistic**	***P*-value**
O15075	*DCLK1* (doublecortin like kinase 1)	-10.786	6.02E-16	-3.836	4.46E-04
P22676	*CALB2* (calbindin 2)	-9.154	3.51E-13	-11.731	2.31E-14
Q9BPU6	*DPYSL5* (dihydropyrimidinase like 5)	-8.951	7.86E-13	-12.522	3.05E-15
P52306	*RAP1GDS1* (Rap1 GTPase-GDP dissociation stimulator 1)	-8.950	7.90E-13	-5.331	4.39E-06
P05413	*FABP3* (fatty acid binding protein 3)	-8.511	4.57E-12	-2.295	2.72E-02
P29966	*MARCKS* (myristoylated alanine rich protein kinase C substrate)	-8.321	9.79E-12	-11.097	1.24E-13
Q12860	*CNTN1* (contactin 1)	-7.733	1.05E-10	-3.490	1.22E-03
P55072	*VCP* (valosin containing protein) *	-7.688	1.25E-10	-3.995	2.78E-04
Q6PCE3	*PGM2L1* (phosphoglucomutase 2 like 1)	-7.639	1.53E-10	-7.150	1.33E-08
O14594	*NCAN* (neurocan) ‡	-7.256	7.13E-10	-8.759	9.49E-11
P09417	*QDPR* (quinoid dihydropteridine reductase) †	7.707	1.16E-10	10.303	1.09E-12
P49189	*ALDH9A1* (aldehyde dehydrogenase 9 family member A1)	6.201	4.81E-08	2.478	1.77E-02
O94856	*NFASC* (neurofascin) ‡	5.917	1.47E-07	3.699	6.65E-04
P07602	*PSAP* (prosaposin) † ‡	5.818	2.16E-07	9.804	4.47E-12
P12277	*CKB* (creatine kinase B)	5.590	5.22E-07	N/A	N/A
O94811	*TPPP* (tubulin polymerization promoting protein)	5.257	1.86E-06	10.842	2.47E-13
P14618	*PKM* (pyruvate kinase, muscle)	5.190	2.39E-06	10.344	9.73E-13
Q16653	*MOG* (myelin oligodendrocyte glycoprotein)	5.180	2.48E-06	4.991	1.29E-05
P13611	*VCAN* (versican)	5.170	2.58E-06	-2.601	1.31E-02
Q969P0	*IGSF8* (immunoglobulin superfamily member 8)	5.038	4.21E-06	2.104	4.19E-02

Linear regression models adjusted for pH, PMI, sex and ethnicity. The direction of change (*t*-statistic) across postnatal development and *P*-value significance are displayed for each protein and paired RNA product. Candidate neurodevelopmental genetic risk loci are symbolized: *, autism; †, intellectual disability; ‡, schizophrenia.

MAIN FIGURES

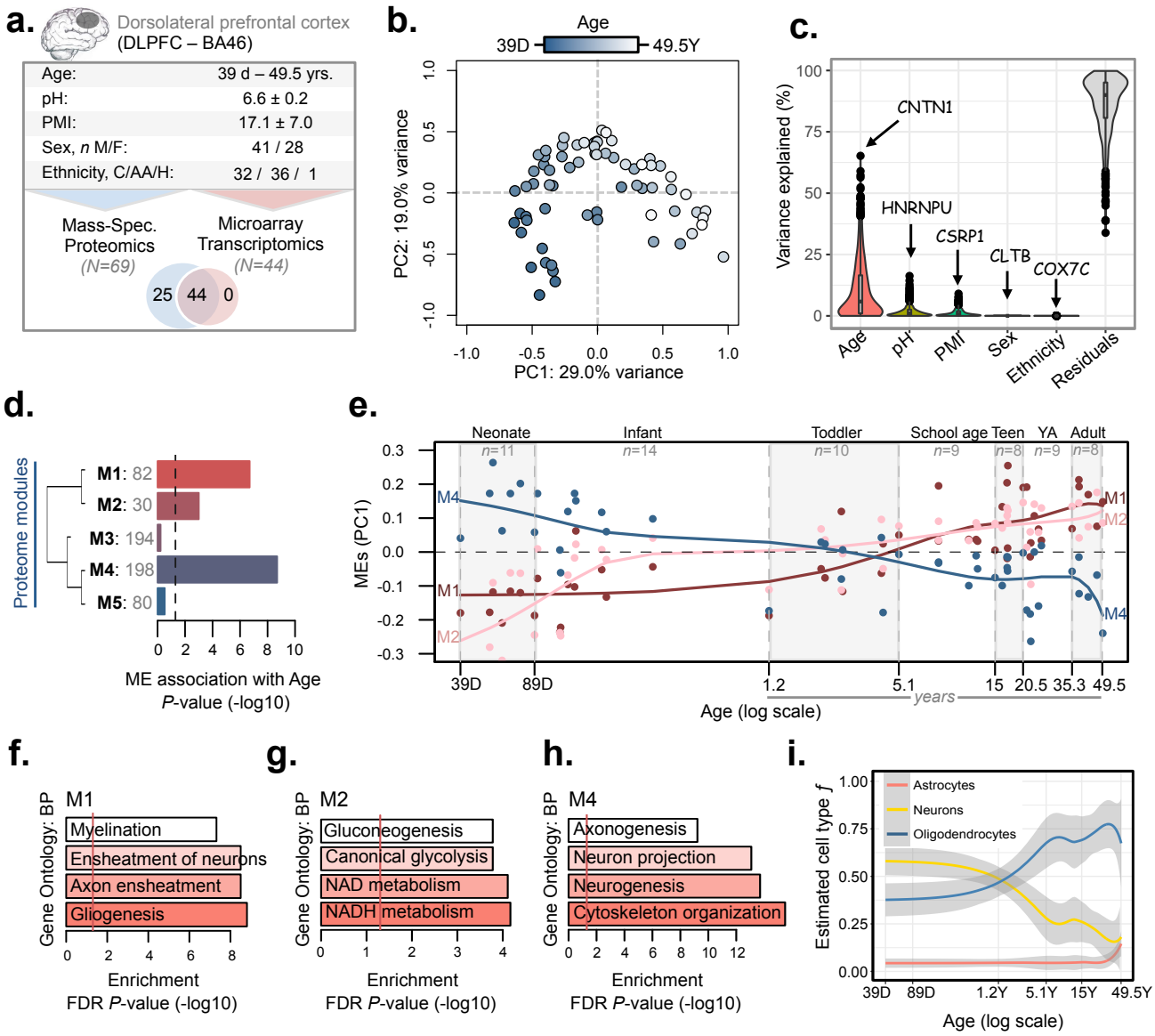

Figure 1. (**a**) Sample characteristics for LC-MS proteomic data (*N*=69). (**b**) Principal component analysis on global normalized protein abundance, samples are shaded by age. (**c**) variancePartition analysis of global protein abundance identifies age as a leading trait explaining most of observed protein variability. (**d**) Weighted correlation network analysis identified five protein modules, three significantly associated with age. Module IDs and the number of proteins within each module are displayed right of dendrogram. (**e**) Age (x-axis) relative to module eigengene (ME) expression (y-axis) for modules M1 and M2 increasing throughout development and module M4 decreasing throughout development. Smoothing linear spline models (knot=4) were fit for each ME across age. Seven post-natal developmental stages are color shaded and labeled (YA, young adult). (**f-h**) Functional annotation of age-

related modules M1, M2 and M4 according to gene ontology biological processes. (**i**) Cell-type deconvolution by Cibersort on global protein abundance. HGNC symbol abbreviations in panel **c**; *CNTN1*, contactin 1; *HNRNPU*, heterogeneous Nuclear ribonucleoprotein U; *CSRP1*, cysteine and glycine rich protein 1; *CLTB*, clathrin light chain b; *COX7C*, cytochrome c oxidase subunit 7c.

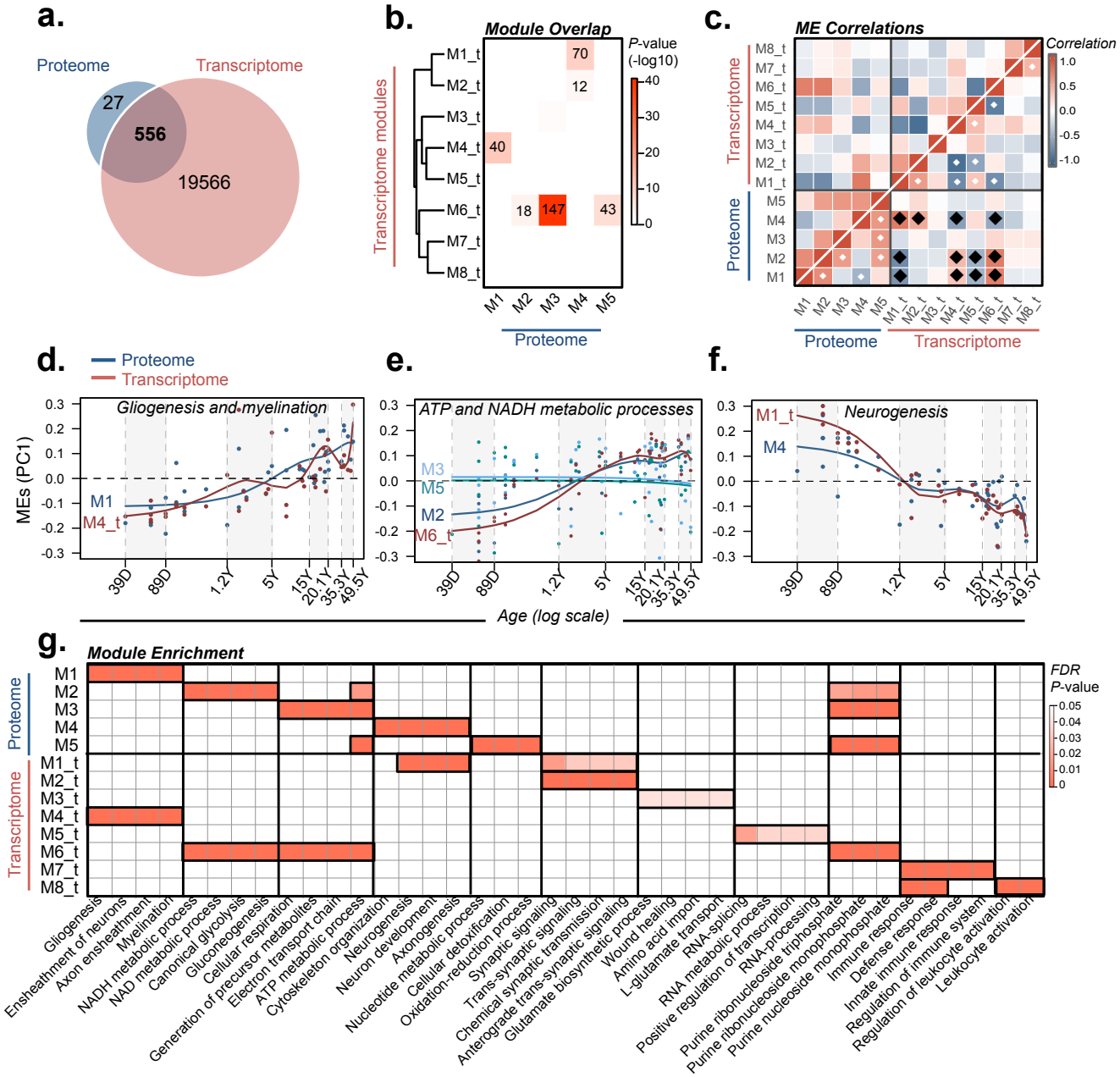

Figure 2. (**a**) Venn diagram of detected the proteome and transcriptome. (**b**) Overlap analysis of transcriptome- and proteome-based modules. The number of overlapping HGNC symbols are displayed for significant intersects. (**c**) Pearson's correlation coefficient analysis of module eigengenes (MEs) within and across transcriptome- and proteome-based modules. Black diamonds indicate $p<0.05$. (**d-g**) Convergent ME expression patterns for modules with significant overlap and shared function between transcriptome and proteome. Smoothing linear spline models (knot=4) were fit for each ME across seven postnatal stages (shaded grey/white). (**h**) Overlap and correspondence of the top four functional categories for all identified RNA- and protein-based modules.

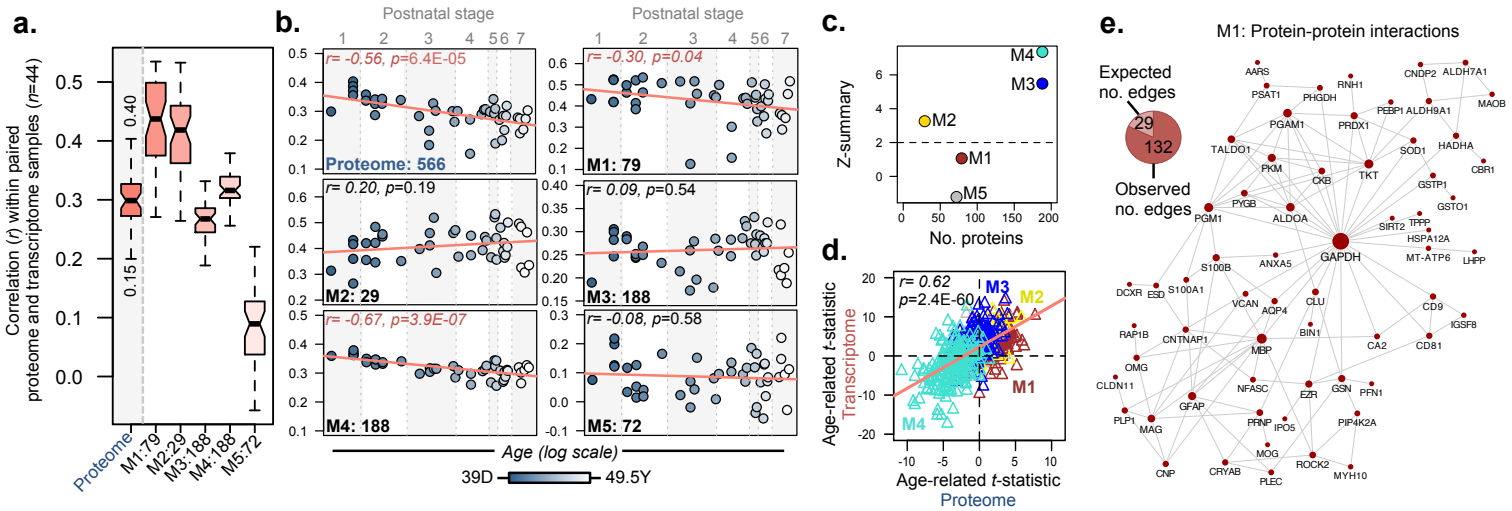

Figure 3. (**a**) Average Pearson correlation coefficient within paired DLPFC transcriptome and proteome samples (*N*=44). Associations were tested for all paired RNA and protein molecules (proteome) and then by focusing on RNA and protein content accordingly to protein module status. The number of molecules compared within each module are listed on the x-axis. (**b**) Average Pearson correlation within paired transcriptome and proteome samples (y-axis) measured as a function of age (x-axis). Samples are ranked according to age, and shaded accordingly to postnatal developmental period (1, neonate; 2, infant; 3, toddler; 4, school age; 5, teenager; 6, young adult; 7, adult). (**c**) Preservation of protein co-regulation patterns at the RNA level; $Z_{summary}$ <2 indicates no preservation, 2< $Z_{summary}$ <10 indicates weak preservation. (**d**) Scatterplot of linear regression age-related *t*-statistics computed for the overlapping 566 mRNAs and proteins. RNAs and proteins are colored according to protein module membership. (**e**) Direct protein-protein interaction network of protein module M1, displaying significant decrease in correlation between RNA and proteins across postnatal development.

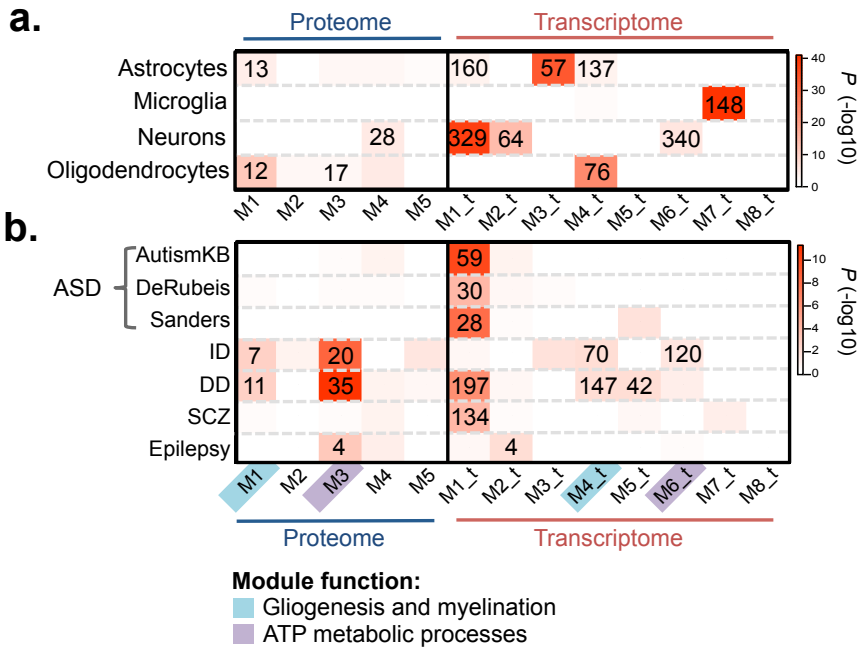

Figure 4. (**a**) Cell type enrichment analysis for the identified transcriptome- and proteome-based modules. The number of significantly overlapping HGNC symbols are displayed. (**b**) Genetic risk loci enrichment analysis according to transcriptome- and proteome-based modules. Seven different lists of neurodevelopmental risk loci were used, including intellectual disability (ID), epilepsy, schizophrenia (SCZ), developmental delay (DD) and three specific to autism spectrum disorder (ASD).