Cell Systems

Article
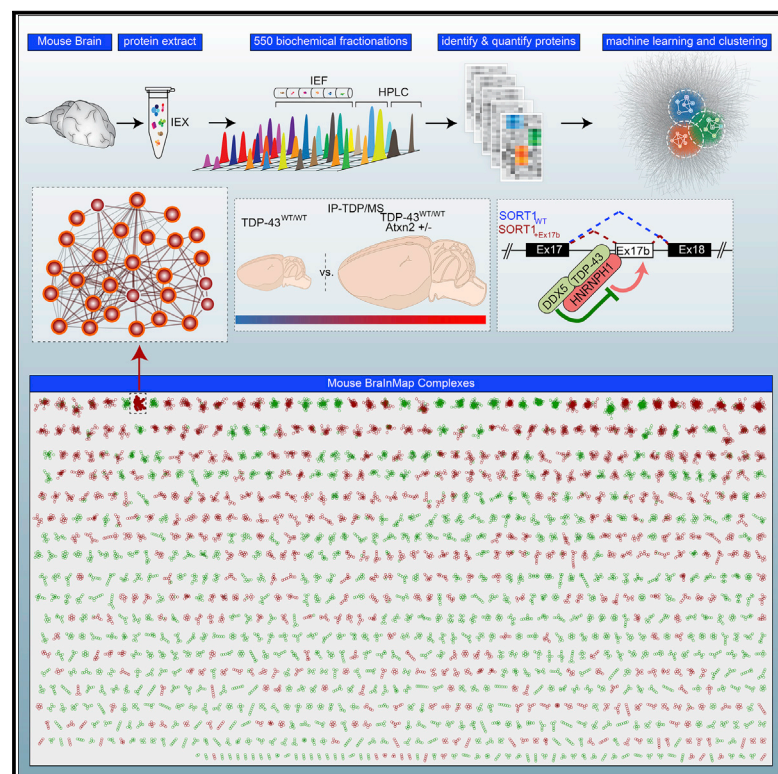
BraInMap Elucidates the Macromolecular Connectivity Landscape of Mammalian Brain

Graphical Abstract

Authors

Reza Pourhaghighi, Peter E.A. Ash, Sadhna Phanse, ..., Gary D. Bader, Benjamin Wolozin, Andrew Emili

Correspondence

bwolozin@bu.edu (B.W.), aemili@bu.edu (A.E.)

In Brief

In this ground-breaking work, Pourhaghighi et al. have carried out a survey of over one thousand multi-protein complex interactions in the mouse brain using a platform they have named BraInMap (for brain interaction map). This approach uses computer learning to reconstruct protein interactions from brain tissues that have been extensively purified. This important resource will allow neuroscientists to explore important neurobiological questions and identify pathways that are adversely affected in disease.

Highlights

- BraInMap is a global proteomic survey of over 1,000 multi-protein brain complexes

- Near-native complex identification by CF-MS and reconstruction by computer learning

- Technique interrogates complexes in normal and pathophysiological context

- Allows study of functional modules that are adversely affected in neurological diseases

CellPress

Cell Systems

Article

BraInMap Elucidates the Macromolecular Connectivity Landscape of Mammalian Brain

Reza Pourhaghighi,[1,17] Peter E.A. Ash,[2,17] Sadhna Phanse,[1,4,11,17] Florian Goebels,[1] Lucas Z.M. Hu,[1] Siwei Chen,[3] Yingying Zhang,[3] Shayne D. Wierbowski,[3] Samantha Boudeau,[2] Mohamed T. Moutaoufik,[4] Ramy H. Malty,[4] Edyta Malolepsza,[5,6] Kalliopi Tsafou,[5,6] Aparna Nathan,[5,6] Graham Cromar,[7] Hongbo Guo,[1] Ali Al Abdullatif,[2] Daniel J. Apicco,[2] Lindsay A. Becker,[8] Aaron D. Gitler,[8] Stefan M. Pulst,[9] Ahmed Youssef,[10,11,12] Ryan Hekman,[11,12] Pierre C. Havugimana,[11,12,13] Carl A. White,[11,12] Benjamin C. Blum,[11,12] Antonia Ratti,[14] Camron D. Bryant,[2] John Parkinson,[7] Kasper Lage,[5,6] Mohan Babu,[4] Haiyuan Yu,[3] Gary D. Bader,[1] Benjamin Wolozin,[2,15,16,*] and Andrew Emili[1,10,11,12,13,18,*]

[1]Donnelly Center for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada
[2]Department of Pharmacology and Experimental Therapeutics, Boston University School of Medicine, Boston, MA, USA
[3]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY, USA
[4]Department of Biochemistry, University of Regina, Regina, SK, Canada
[5]Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[6]Broad Institute of Massachusetts Institute of Technology and Harvard University, Boston, MA, USA
[7]Program in Molecular Medicine, Hospital for Sick Children and University of Toronto, Toronto, ON, Canada
[8]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA
[9]Department of Neurology, University of Utah, Salt Lake City, UT, USA
[10]Program in Bioinformatics, Boston University, Boston, MA, USA
[11]Center for Network Systems Biology, Boston University, Boston, MA, USA
[12]Department of Biochemistry, Boston University School of Medicine, Boston University, Boston, MA, USA
[13]Departments of Biochemistry and Biology, Boston University, Boston, MA, USA
[14]Department of Neurology and Laboratory of Neuroscience, IRCCS, Milan, Italy
[15]Department of Neurology, Boston University School of Medicine, Boston, MA, USA
[16]Program in Neuroscience, Boston University, Boston, MA, USA
[17]These authors contributed equally
[18]Lead Contact
*Correspondence: bwolozin@bu.edu (B.W.), aemili@bu.edu (A.E.)
https://doi.org/10.1016/j.cels.2020.03.003

SUMMARY

Connectivity webs mediate the unique biology of the mammalian brain. Yet, while cell circuit maps are increasingly available, knowledge of their underlying molecular networks remains limited. Here, we applied multidimensional biochemical fractionation with mass spectrometry and machine learning to survey endogenous macromolecules across the adult mouse brain. We defined a global "interactome" comprising over one thousand multi-protein complexes. These include hundreds of brain-selective assemblies that have distinct physical and functional attributes, show regional and cell-type specificity, and have links to core neurological processes and disorders. Using reciprocal pull-downs and a transgenic model, we validated a putative 28-member RNA-binding protein complex associated with amyotrophic lateral sclerosis, suggesting a coordinated function in alternative splicing in disease progression. This brain interaction map (BraInMap) resource facilitates mechanistic exploration of the unique molecular machinery driving core cellular processes of the central nervous system. It is publicly available and can be explored here https://www.bu.edu/dbin/cnsb/mousebrain/.

INTRODUCTION

The mammalian brain consists of intricate physical and functional protein interaction networks whose compositions are largely uncharacterized. These circuits support essential functions of a vast interconnected array of neurons, glial, oligodendrocytes, and other cell types (Elmer and McAllister, 2012; Grant and O'Dell, 2001; Migaud et al., 1998; Sherman and Brophy, 2005; Slepnev and De Camilli, 2000; Small and Petsko, 2015; Zhu et al., 2016). Proper synaptic formation and activity resulting from these net-

works is essential for core brain functions, such as neurotransmission, synaptic plasticity, and memory. These molecular circuits are perturbed in neurological syndromes by genetic variants and environmental factors, resulting in behavioral, cognitive, and neurodegenerative impairments. For example, abnormal protein-protein interactions among tau and α-synuclein lead to pathological accumulation preceding neurodegeneration (Forman et al., 2004; Ross and Poirier, 2004; Vanderweyde et al., 2016). Disease-causing disruptions in macromolecular assemblies have also been documented in

amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD) (Dormann et al., 2010), as well as Parkinson's disease (PD) (Carrion et al., 2017; Malty et al., 2017). Hence, mapping the global physical cartography of brain protein interaction networks is essential to understand normal neuronal functions, the causal mechanisms driving disease, and for discovery of new targets as a basis for more effective and selective clinical therapies. Although large-scale physical interaction maps have been reported for transformed human cell lines (Havugimana et al., 2012; Hein et al., 2015; Huttlin et al., 2017; Wan et al., 2015), to our knowledge, no direct large-scale experimental study of the mammalian brain regional protein circuitry, or "interactome," has ever been reported, thwarting clinically actionable mechanistic understanding of neuronal processes and dysfunction.

To fill this gap, we systematically isolated and characterized endogenous protein assemblies on a global scale from mammalian brain lysates. Given its experimental tractability, widespread use in the neurobiology field, and the short post-mortem intervals enabled by animal studies, we opted to study mouse as a model. Multi-protein complexes were biochemically resolved and their cognate components were identified by precision mass spectrometry, based on their reproducible co-fractionation over orthogonal separations. Using an integrative co-complex scoring pipeline, we then generated a high-resolution survey, termed the BraInMap, representing the largest experiment-based protein interaction network for the central nervous system (CNS) to date and to the best of our knowledge. BraInMap comprises hundreds of putative macromolecular assemblies, most of which are conserved in human and expressed in a regional and cell-type specific manner.

To illustrate the utility of BraInMap, we explored the functional and biophysical properties of brain-specific assemblies with significant associations to core neurological functions and disorders in humans. We provide evidence that disease-associated processes and genetic variants disrupt the physical interfaces between components of neuronal protein assemblages critical for normal brain physiological homeostasis, suggesting a common causal basis for diverse neuropathies. Particularly prevalent were assemblies enriched for RNA-binding proteins (RBPs), whose physical associations are fundamentally linked to the etiology and pathogenesis of progressive neurological disorders, such as Alzheimer's disease (AD), ALS, and FTD. While toxic gain-of-function and loss-of-function mutations in certain RBPs have previously been reported to elicit deleterious effects on splicing and RNA homeostasis (Arnold et al., 2013; Fratta et al., 2018), BraInMap describes their normal physical interactions in healthy adult brain, and therefore serves as a useful resource to bridge the gap between macro-level cell-cell connectivity studies, neuronal cell biology and epidemiological genetics, opening up new research avenues in molecular systems neuroscience.

RESULTS

Biochemical Fractionation and Precision Mass Spectrometry Reveals Brain-Selective Protein Assemblies

As illustrated schematically in Figure 1A, soluble protein extracts were prepared from homogenized whole brains from adult CD1

mice using gentle isolation procedures (STAR Methods). The lysates were subjected to extensive non-denaturing biochemical fractionation, followed by mass spectrometry to identify and quantify stably associated proteins reproducibly co-eluting together. To maximize resolution and coverage, we deployed multiple orthogonal workflows to separate native macromolecules from functionally unrelated constituents (i.e., to mitigate "chance" co-elution). This included two-dimensional (2D) separations combining isoelectric focusing (IEF) with mixed-bed ion exchange high performance liquid chromatography (IEX-HPLC). In parallel, we performed repeat IEX-HPLC-based separations using alternate chromatography procedures to selectively enrich for cytoskeletal, nuclear, membrane-bound, and synaptic protein assemblies (STAR Methods).

Altogether, 550 biochemical fractions were collected, across nine different fractionation experiments, which included replicate runs as a test for reproducibility. After enzymatic digenstion by trypsin, each fraction was analyzed by nanoflow liquid chromatography coupled to Q Exactive HF mass spectrometer. The spectra were subjected to stringent database searching and filtering (false discovery rate <1% at both the peptide- and protein-level) using multiple search algorithms, which on integration (STAR Methods) resulted in 8,389 high-confidence protein identifications (Table S1). Hierarchical clustering of the recorded protein profiles, covering two-thirds (5,505 of 9,121) of previously reported mouse brain tissue annotations (The UniProt Consortium, 2017), demonstrated the characteristic elution patterns of both neuronal and ubiquitous (housekeeping) protein assemblies (Figure 1B). In comparison to previously reported large-scale interactome studies of cultured cell lines (Havugimana et al., 2012; Huttlin et al., 2017; Wan et al., 2015), BraInMap was significantly enriched for annotated proteins known to be selectively expressed in mammalian cortex (false discovery rate [FDR] = 7.2×10^{-56}), brain (FDR = 1.34×10^{-41}) and other brain regions relative to other mouse tissues (Figure 1C; Table S6).

Scoring High-Confidence Co-complex Associations

We devised a quantitation-based computational pipeline to tally the likelihood of co-complex interactions, based on the pairwise similarity of the measured protein co-elution patterns. The premise was that stably associated components (subunits) of a multi-protein complex exhibited correlated profiles (i.e., components reproducibly co-purify together). Proteomic precursor ion (MS1) intensity profiles across all the fractions were acquired using both MaxQuant (CORE) and three additional search engines (X!Tandem, MSGF+, and Comet) (EXPANDED) to derive preliminary protein co-complex associations. We calculated five established similarity measures (APEX, Jaccard, Bayes, Euclidean Distance, and Mutual Information; see STAR Methods) that evaluated different features recorded in each experimental profile separately.

In a subsequent step (Figure 1D), protein pairs from both the CORE and EXPANDED datasets showing high similarity were input into a supervised machine-learning model (random forest classifier; STAR Methods). Two models were trained to predict high-confidence co-complex associations based on the co-fractionation patterns we observed alone, or together with other publicly available supporting functional association evidence, with reference to curated "gold standard" brain associated
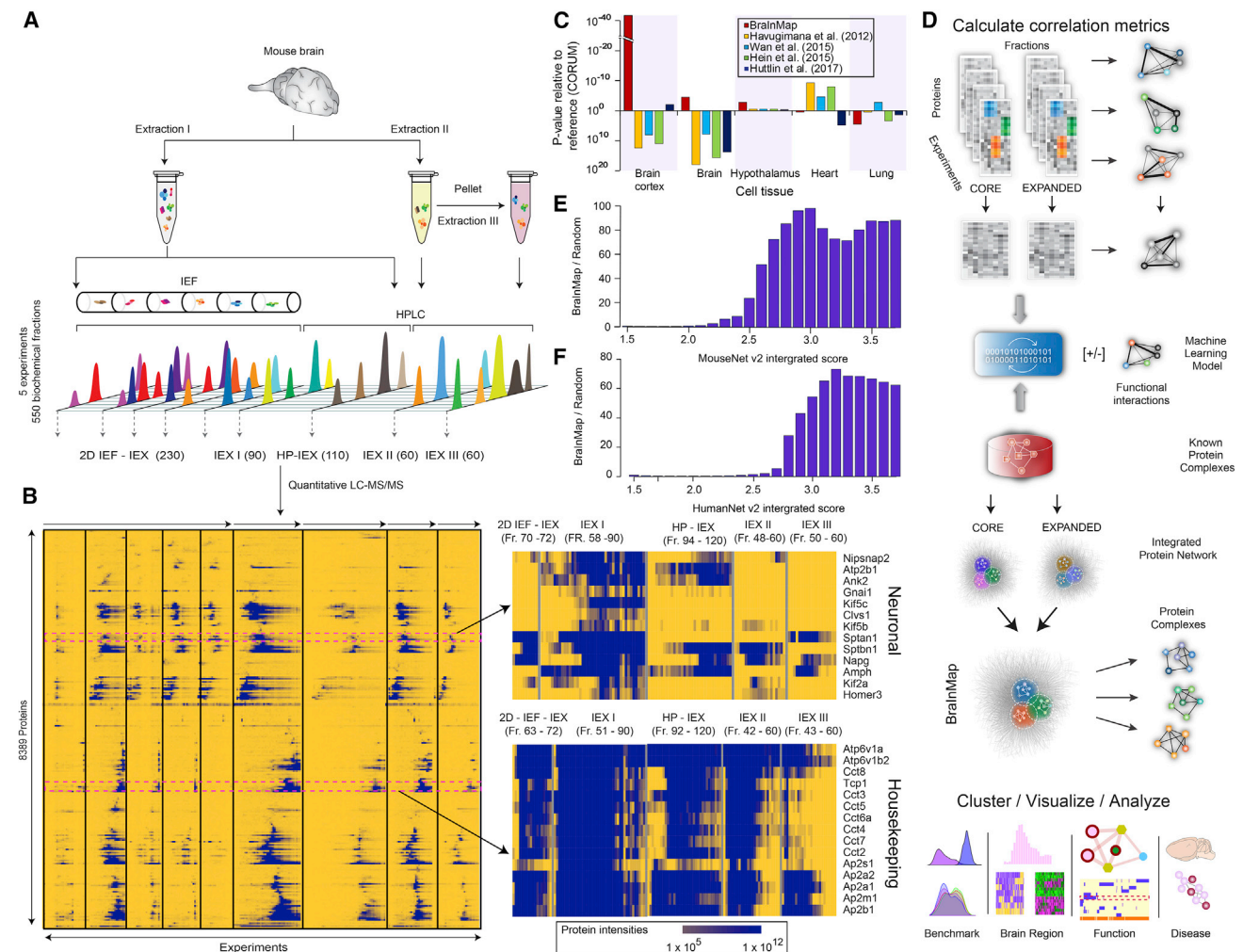
Figure 1. Integrative Workflow Used to Generate the Mammalian Brain Interactome Map (BraInMap)

(A) Multi-pronged biochemical fractionation (ion exchange chromatography, *IEX*; isoelectric focusing, *IEF*; and fraction numbers in brackets) of soluble macromolecular assemblies from mouse brain extracts.

(B) Hierarchical clustering of protein co-fractionation intensity profiles recorded by precision liquid chromatography-tandem mass spectrometry (LC-MS/MS); (*right*) neuronal (*top*) and housekeeping (*bottom*) components highlighted.

(C) Enrichment analysis (DAVID [Huang da et al., 2009]) of representative tissue annotations (UniProt) for proteins detected in this work relative to previously published interactome studies.

(D) Schematic depicting steps in the integrative BraInMap computational scoring pipeline: calculation of protein similarity (correlation) metrics, integrative classifier training (EPIC machine learning; [Hu et al., 2019]) and scoring of co-fractionation data (this study) and supporting (public) evidence to predict high-confidence co-complex interactions, followed by network partitioning, benchmarking, and meta-analysis (pathobiological relevance) of the predicted complexes.

(E) Enrichment of interacting (co-eluting) brain proteins relative to random pairs for high functional similarity based on association scores reported in MouseNet (v2) (Kim et al., 2016).

(F) Enrichment of orthologs of interacting mouse brain proteins relative to random pairs for high functional association scores in HumanNet (v2) (Hwang et al., 2019).

mammalian macromolecules (Table S2). Positive examples were obtained from public curated databases—namely CORUM (Ruepp et al., 2010), IntAct (Orchard et al., 2014), and Gene Ontology (Ashburner et al., 2000), while negatives were created from randomized combinations of components assigned to distinct clusters. To minimize classifier bias, known mouse exemplars were supplemented with annotated human protein assemblies based on strict one-to-one orthology projections (In-Paranoid) (Sonnhammer and Östlund, 2015). Moreover, complexes in the training set sharing a majority of subunits were

merged (fractional overlap >0.8), while those with more than 50 members (e.g. ribosome) were excluded.

High concordance was evident when comparing both sets of co-fractionation patterns to probabilistic functional associations previously predicted based on protein domain co-occurrence, co-expression and co-citation in both mouse (MouseNet v2 database) (Kim et al., 2016) (Figure 1E) and human (HumanNet v2) (Hwang et al., 2019) (Figure 1F) Conversion of human to mouse identifiers was done through one-to-one orthology mapping via InParanoid (Sonnhammer and Östlund, 2015), and

wherever applicable human orthologs of mouse proteins were named using uppercase letters, e.g., the human ortholog of mouse protein Tdp-43 was referred to as TDP-43. These observations establish the broad physiological relevance of our initial interactome data.

The trained classifiers were then used to generate probabilistic co-complex relationships from both the CORE and EXPANDED datasets (STAR Methods). We evaluated all possible feature combinations to optimize precision and recall. The models were merged (average) into a single final high-confidence protein-protein interaction (PPI) network, the BraInMap, consisting of 27,043 co-complex interactions (Table S3). We benchmarked the model prediction performance (precision and recall) by two-fold cross validation, using a fully independent set (i.e., exclusive of the training set) of manually curated complexes (from CORUM) for evaluation. These tests established a stringent FDR of 11% with a precision-recall area-under-the-curve AUC of 0.92 (Figure 2A).

We portioned the integrated network using the ClusterONE (Nepusz et al., 2012) clustering algorithm, which revealed 1,030 putative brain protein assemblies (Table S4). Based on their degree of connectivity and the initial source network from which a particular subunit within a given assembly was derived, each complex can be deconstructed into a "core" and "extended" set of interacting components (Figure S1A). To rigorously evaluate classifier performance at the protein complex level (rather than PPI as before), we calculated three stringent evaluation metrics (maximum matching ratio, accuracy, and overlap score) (Nepusz et al., 2012) and combined the results into a single summary "composite" quality score (F-measure) (STAR Methods).

As seen in Figure 2B, complexes based on our brain co-fractionation data alone (i.e., built without external data) produced a comparable or higher total composite score than other recently reported cell line-based interactomes, establishing the overall reliability of our scoring pipeline. We boosted classifier performance further by incorporating additional supporting functional association evidence (see STAR Methods) from MouseNet (Kim et al., 2016) and other public sources. We emphasize that external data were used primarily as a filter to reinforce the primary findings of our proteomics data and that none (zero) of the protein assemblies in BraInMap are based solely on external sources (all macromolecular complexes are derived from replicate co-fractionation data).

To establish the degree of agreement with previously known complexes, we systematically examined the overlap of BraInMap complexes with annotated assemblies using multiple similarity metrics (Figure S1B). Of the six metrics tested, we settled on average matching index (AMI) and hypergeometric score as the most inclusive and stringent criteria to define macromolecules not reported in public databases (Figure S1C). We calculated AMI as the average fraction of subunits matched between a known and predicted complex (STAR Methods).

As shown in Figures 2C and 2D, just over half (638, or 62%) of our complexes overlapped (AMI ≥ 0.25, hypergeometric p value ≤ 0.05) significantly with one or more previously reported complexes (Havugimana et al., 2012; Huttlin et al., 2017; Ruepp et al., 2010; Wan et al., 2015); of these, 146 were considered as fully annotated (AMI ≥ 0.5) while the others (492) were deemed to have additional subunits not previously reported. Using this

rigorous definition, the remaining (392) complexes appear to be reported here for the first time (Table S4). Consistent with the source tissue, over half (57%) of all the assemblies recovered by our survey consisted predominantly (≥50%) of components annotated as neuronal according to the Gene Ontology (STAR Methods), whereas only 33 assemblies (3%) lacked neuronal constituents (Figure 2E).

Brain Complexes Exhibit Recent Evolutionary Adaptations that Extend to Human

To assess the human physiological relevance of BraInMap, we compared the underlying co-complex interactions against a fully independent curated public database of high quality human PPI ("InWeb," pooled from the InWeb3 and InWeb_IM resources) (Lage et al., 2007; Li et al., 2017) and found an overall agreement of 64% (Figure 2F). Consistent with this high apparent conservation, our own independent validation experiments showed that human orthologs of putatively interacting mouse components also tended to co-elute together (i.e., have correlated co-elution profiles) in independent chromatographic fractionation experiments performed on protein extracts from cultured human neuroblastoma SH-SY5Y cells (Figure 2G; Table S5). Likewise, human orthologs of BraInMap components strikingly showed higher correlated co-fitness profiles upon mRNA knockdown in human cell culture (Pan et al., 2018) as compared with random target pairs (Figure 2H), implying functional conservation of these complexes in human brain as well. Further support for the apparent conserved roles of these putative complexes in the CNS was demonstrated by the observation that these same orthologs are highly expressed during human brain development (embryogenesis through adulthood; Figure 2I) based on messenger RNA expression data obtained from the BrainSpan Atlas (Miller et al., 2014).

We examined the domain architecture of brain-specific assemblies by assessing their corresponding Pfam A domain and family assignments (STAR Methods). By definition, domains are highly conserved sequence patterns that are presumed to represent independent folding units, whereas domain pairs in multi-domain proteins represent combinations of units operating in tandem (Cromar et al., 2016). In general, while rarely in majority, brain-specific protein domains and domain pairs were found to occur widely across the BraInMap (Figure 2J), suggesting they confer brain-specific functions. For some complexes, the occurrence of brain-specific domains was a dominant feature. The presence of unique folds, both independently and in combination, reinforces the concept that the complexes we found in the brain differ markedly from those revealed in previous interactome studies of cell lines. These observations are also consistent with specialized roles in processes linked to neurodevelopment and brain physiology. Consistent with this, many of the conserved complexes showed broad functional annotation diversity (Table S6) and enrichment for associations with specialized neuronal compartments (e.g., synapse, axon, and dendrite), processes (e.g., neurogenesis), and particular protein domains (Figure 2K).

Regionalization and Cell-Type Specificity of Brain Protein Assemblies

The mechanisms underlying the regional specification of the vertebrate CNS are of broad interest. This specification may be
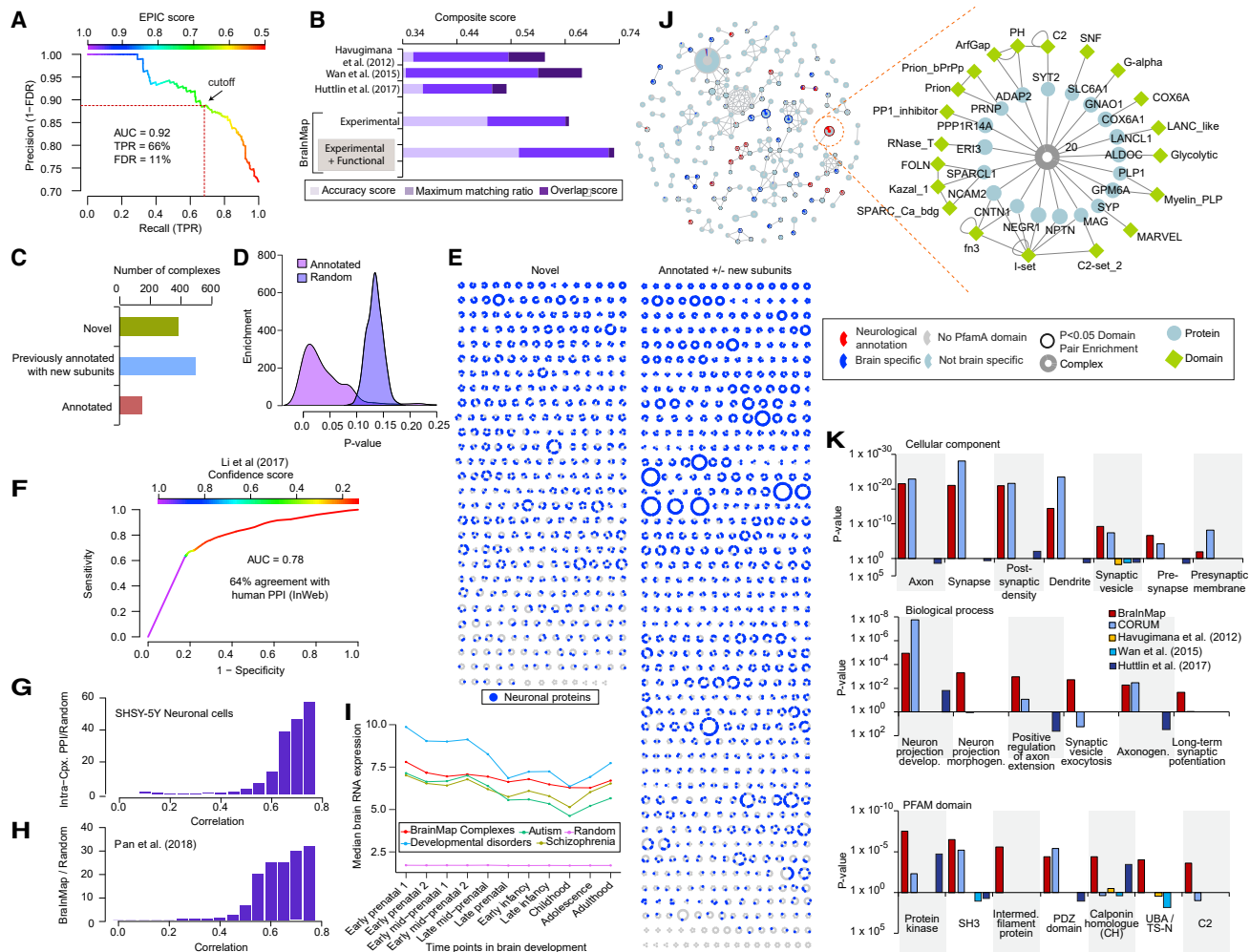
Figure 2. Benchmarking Reveals Diverse, Evolutionarily Conserved Brain Complexes

(A) Precision-recall (PR) analysis of predicted (EPIC score) co-complex interactions (CORE + EXPANDED) benchmarked against an independent (holdout) set of brain-derived reference assemblies establishes a FDR of 11%.

(B) Benchmark quality metrics of putative complexes (this work) versus other interactome maps. Bar length reflects total composite score, calculated as the sum of complex maximal matching ratio, overlap, and accuracy (see STAR Methods) relative to select reference curated brain macromolecules.

(C) Bar chart of categorized complexes (partial or complete match to annotated assemblies versus novel).

(D) Highly significant (hypergeometric p values) overlap of predicted complexes with annotated assemblies compared to randomized protein sets.

(E) Schematic of protein assemblies in BraInMap, sorted according to novelty, showing the distribution of neuron-associated components (blue).

(F) ROC analysis of predicted co-complex interactions showing high agreement with previously reported high-confidence orthologous human protein interactions in the InWeb database (Li et al., 2017).

(G) Enrichment of human orthologs of BraInMap complex subunits relative to randomized protein pairs for highly correlated co-fractionation profiles of SHSY5Y neuronal cell extracts.

(H) Enrichment of human orthologs of interacting proteins in BraInMap relative to random pairs for high functional "co-fitness" scores (Pan et al., 2018).

(I) Median expression of orthologs of BraInMap components during development of the human cortex; lines indicate levels of all interacting components (red) versus the subset associated with risk for schizophrenia (olive) (Schizophrenia Working Group of the Psychiatric Genomics, 2014), autism (green) (Sanders et al., 2015), or other neurodevelopmental disorders (cyan) (Deciphering developmental Disorders Study, 2017), as compared to random proteins (magenta).

(J) Schematic of protein domains enriched in BraInMap. Complexes (nodes) sharing two or more domains are joined according to overlap (Jaccard Index). Colors reflect the proportion of domains restricted to brain (*blue*) or linked to neuropathology (*red*). Highlighted bipartite sub-network shows relationship between subunits (*ellipses*) and domains (*diamonds*) of a representative assembly (complex 20).

(K) Annotation enrichment (DAVID; Huang da et al., 2009) in BraInMap relative to previous interactome studies: Gene Ontology (i) cellular component, (ii) biological process terms, or (iii) PFAM domains (Finn et al., 2016).

driven, in part, by differences in the abundance (expression) and composition of different macromolecular complexes. To directly examine the potential regionalization of the protein assemblies in BraInMap, we performed independent biochemical fraction-

ations (quadruplicate IEX-HPLC runs) and mass spectrometric profiling on ten distinct brain regions (Figure 3A; frontal cortex, parietal cortex, occipital cortex, hippocampus, striatum, thalamus, midbrain, hindbrain, cerebellum, and cervical spinal
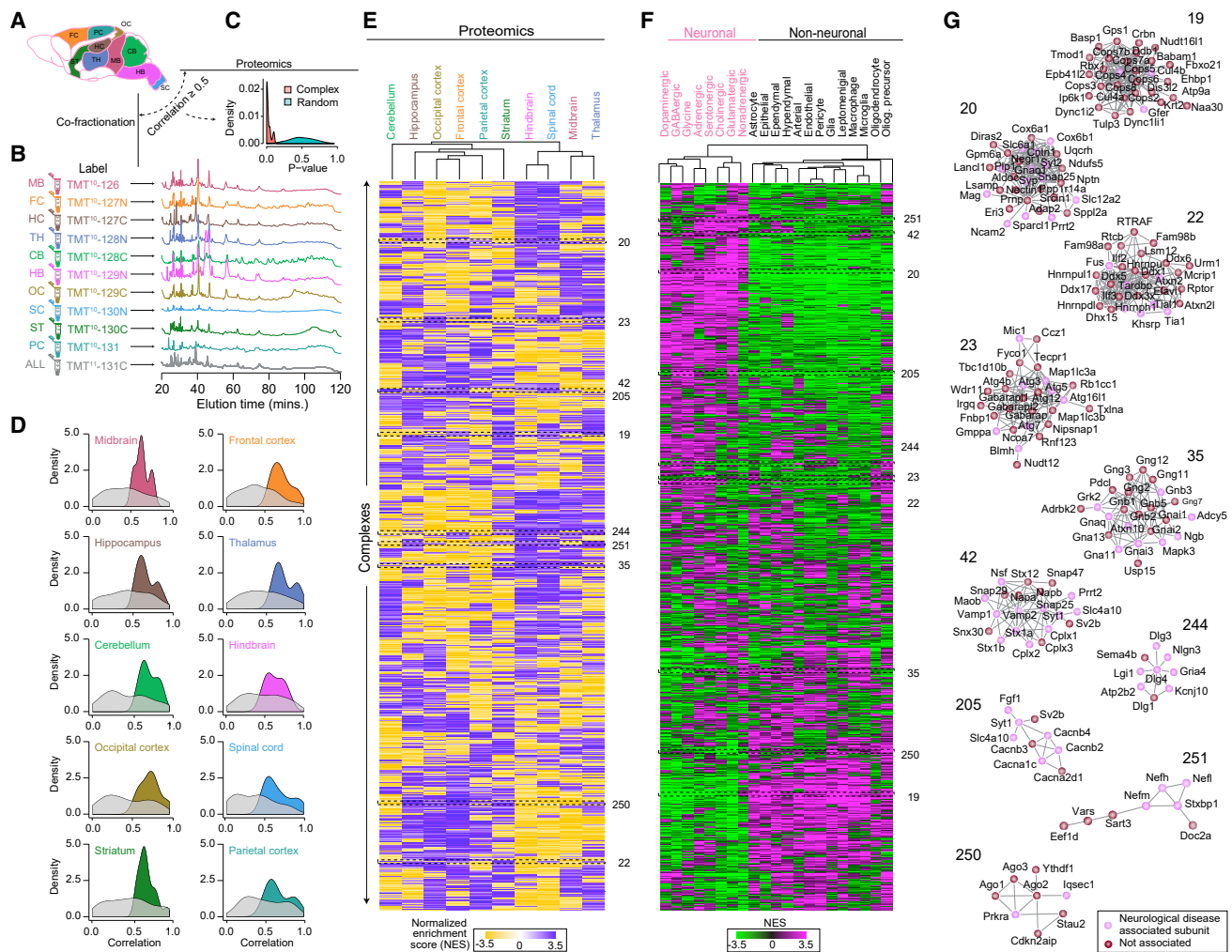
Figure 3. Regional- and Cell-Type Selective Macromolecules

(A) Schematic of ten mouse brain regions subjected to quantitative proteomic profiling and biochemical (IEX-HPLC) fractionation in parallel.

(B) Representative chromatograms and isobaric (TMT) labeling of fractionated regional assemblies.

(C) Highly significant (hypergeometric) agreement between the regional abundance patterns recorded by quantitative profiling versus co-fractionation of BraInMap components (derived by whole tissue analysis) as compared to randomized protein sets.

(D) Complex subunits with highly correlated regional co-fractionation profiles also show significantly co-enrichment (hypergeometric p value $\leq$ 0.05, relative to randomized protein pairs) in the same brain compartments as determined by quantitative proteomics (E).

(E) Heatmap clustergram showing complex regional specificity (enrichment p value $\leq$ 0.01 by Kolmogorov–Smirnov test) as measured by quantitative proteomics.

(F) Heatmap clustergram of complexes showing preferential (p $\leq$ 0.01 by KS test after normalization) component mRNA expression in neuronal versus non-neuronal cell classes based on recently published mouse brain scRNA-seq data (Zeisel et al., 2018).

(G) Representative complexes displaying regional (proteomic) and neuronal cell-type (scRNA-seq) specificity. Highlighted (red) nodes represent subunits associated with neurological disorders.

cord) isolated from age- and gender-matched CD1 mice (four males, 16 weeks) (STAR Methods). To accurately quantify differences in relative abundance, we used a two-pronged multiplexing procedure based on stable isotope labeling (tandem mass tags) to measure both the regional expression patterns and the corresponding regional co-elution profiles of most of the BraInMap assemblies in parallel (Figure 3B). We found that complexes that were significantly enriched (hypergeometric p value $\leq$ 0.05) in the brain total protein measurements (Figure 3C) were likewise enriched for subunit pairs showing significantly (hypergeometric

p value $\leq$ 0.05) and reproducibly correlated co-fractionation profiles (as compared to random pairs) across the same regions (Figure 3D), allowing us to infer the regional selectivity of most of the assemblies in BraInMap (Table S7).

To further examine the extent of specification, we overlaid BraInMap with recent mouse single-cell RNA (scRNA Seq) data (Zeisel et al., 2018). After collapsing the cell taxonomy from Zeisel et al., 2018 into 21 broad neuronal and non-neuronal cell types, we observed widespread evidence of selective cognate gene expression (Figure 3F; Table S7). For example,

complex 20, implicated in adhesion and signaling of axons with the myelin sheath, was enriched in neurons as well as abundantly expressed in the hippocampus and spinal cord (Figures 3F and 3G). Functional annotations of its membrane-associated subunits corroborate regional specificity in hippocampus (Ntpn, Prrt2, Slc6a1), and spinal cord (Lancl1, Prrt2, and Srcin1), reflecting roles supporting and maintaining axon growth signals (Gpm6a, Negr1, and Nptn), and vesicle targeting and release (Snap25, Syp, and Syt2). Another component, Slc6a1, terminates GABAergic signal through sodium-dependent reuptake to pre-synaptic terminals, leading to myoclonic-atonic seizures when mutated (Carvill et al., 2015), while Plp1, a key constituent of compact myelin, along with Mag and Cntn1, mediates adhesion of the insulating sheath to axons in the internodes and paranodes, respectively (Jahn et al., 2009). PLP1 mutations cause a spectrum of neuronal disorders from the Pelizaeus-Merzbacher disease to spastic paraplegia 2 (Hobson and Kamholz, 1993), while variants in CNTN1 cause lethal congenital myopathy (Compton et al., 2008), which may reflect an adhesion role at the neuromuscular junction. Mouse prion protein (Prnp) is also present in complex 20 and its interaction with PLP1, MAG, CNTN1, DPP6, ERI3, and SPARCL1 has previously been described (Scmitt-Ulms et al., 2004). In addition to affecting transmissible neurodegenerative disease, neuronal expression of Prnp is essential for maintaining myelination (Bremer et al., 2010). Taken as a whole, this transmembrane assembly is likely critical for formation of myelin sheaths around GABAergic axons.

Complex 251, which contains neuron-specific neurofilament light, medium, and heavy chain (Nefl, Nefm, and Nefh) axoskeletal components, showed enriched expression in neuronal cells, as well as higher abundance in cortex and hippocampus (and lower expression in midbrain, hindbrain, and spinal cord). Complex 42, comprised SNARE (soluble NSF attachment protein receptor) components necessary for neurotransmitter release, was enriched in neurons (Chen et al., 2002) as well as in cerebellum and spinal cord. Conversely, complex 35, which contains alpha, beta, and gamma subunits of guanine nucleotide-binding protein (G-protein), showed high abundance in midbrain and thalamus as well as broad expression in both neurons and non-neurons (Figures 3G and 3F). Upon extracellular ligand binding to G-protein coupled receptors (GPCRs), G-proteins are activated by GDP to GTP replacement, facilitating one of the most prevalent signaling systems in diverse cell types through downstream effectors. Notably, this assembly included β-adrenergic receptor kinases 1 and 2 (Grk2, and Adrbk2), mitogen-activated protein kinase 3 (Mapk3), as well as Ataxin 10 (Atxn10), in which a repeat expansion mutation is associated with spinocerebellar ataxia type 10 (Matsuura et al., 2000). In support of these findings, Atxn10 has previously been shown to interact with Gbeta2 (Gnb2) to potently activate the Ras/Mapk/Elk-1 signaling cascade (Waragai et al., 2006).

BraInMap assemblies showing regional enrichment in thalamus and striatum and preferential enrichment in non-neuronal cells include complex 19, which contains subunits 1 to 8 of the COP9 signalosome (responsible for deneddylation of cullin-RING ubiquitin E3 ligases), cullin4A-RING (Cul4a, Cul4b, Crbn, and Ddb1) and cullin2-RING (Rbx1) E3 ubiquitin ligases (Cavadini et al., 2016; Dubiel et al., 2015). Likewise, complex 250 is enriched in non-neurons and the hippocampus and is composed of

argonaute proteins 1–3 (Ago1-3) necessary for RNA silencing and other double-stranded RNA interacting proteins (Stau2 and Prkra). It also contains RNA-binding protein Ytfdh1 needed to facilitate learning and memory formation in the hippocampus (Shi et al., 2018).

Subcellular Compartmentalization

BraInMap identifies an array of complexes associated with neuronal subcellular compartments such as the axon, dendritic spines and synapse (Figure 4A; Table S8). The latter include assemblies that form a higher order molecular architecture on outer cell membranes as well as the synaptic vesicles involved in neurotransmission (Figure 4B; Table S8). For example, complex 42 (Figure 3F) and 51 share 14 components that encompass SNARE proteins (including Syt1, Snap25, Syntaxins 1a/1b/12, Complexins 1/2/3, and Vamp1/2) necessary for synaptic-vesicle docking (Chen et al., 2002). While complex 42 is characterized by the inclusion of additional synaptic-vesicle transmembrane factors (Sv2b, Slc4a10, and Prrt2), complex 51 is differentiated by the presence of factors mediating ER-Golgi vesicle transport and fusion (Vcp, Sec22b, Scfd1, and Arfgap2). Likewise, complex 234 and 267 share components required for Glutamatergic neurotransmission, such as Gad1/2 (Glutamate decarboxylases) and Slc17a7 (Vesicular glutamate transporter 1). Complex 234 differs by exhibiting additional interaction with components of excitatory synapses (Vdac1, Nlgn2, and Slc17a6), whereas complex 267 contains endosomal trafficking components (Rab21 and Itgb1). These observations highlight compositional variations relevant to core neuronal activities.

Manifold other complexes in BraInMap are linked to mitochondrial function (Figure 4C; Table S8), which play a crucial role in meeting the elevated energetic demands required for neuronal homeostasis. These include complex 14, which consists of mitochondrial ribosomal proteins (Mrpl/s), and the related complexes 23 and 25, which contain autophagosomal proteins involved in mitochondrial turnover. Conversely, complex 226, comprised factors involved in mitochondrial fission (Dnm1l and Mff), has links to neurological disorders through Scg3, which is involved in secretion of neuropeptides and hormones such as pre-opio-melanocortin from the CNS (Tanabe et al., 2007) as well as neurotoxin-induced apoptosis of Dopaminergic neurons in a PD model (Li et al., 2012).

BraInMap Identifies Manifold RNA-Binding Assemblies

Previously unreported complexes in BraInMap are significantly enriched for involvement in RNA metabolism (Figure 5A), including messenger RNA processing (FDR $p = 1.6 \times 10^{-2}$) and binding (FDR $p = 2.7 \times 10^{-2}$). These assemblies typically comprise RBPs (Figure 5B; Table S8), which mediate the biogenesis, distribution, and metabolism of both coding and noncoding RNAs (Hentze et al., 2018). BraInMap identifies assemblies ranging in size from eight interacting RBPs, such as complex 250, which includes Ago1/2/3 and Stau2 (Figure 3F) to larger complexes with over a dozen subunits. For instance, complex 22 (Figure 5C) contains 28 RBPs (Atxn2/2l, Ddx1/3x/5/6/17, Dhx15, Elavl1, Fam98a/b, Fus, Hnrnpdl/h1/u/ul1, Ilf2/3,
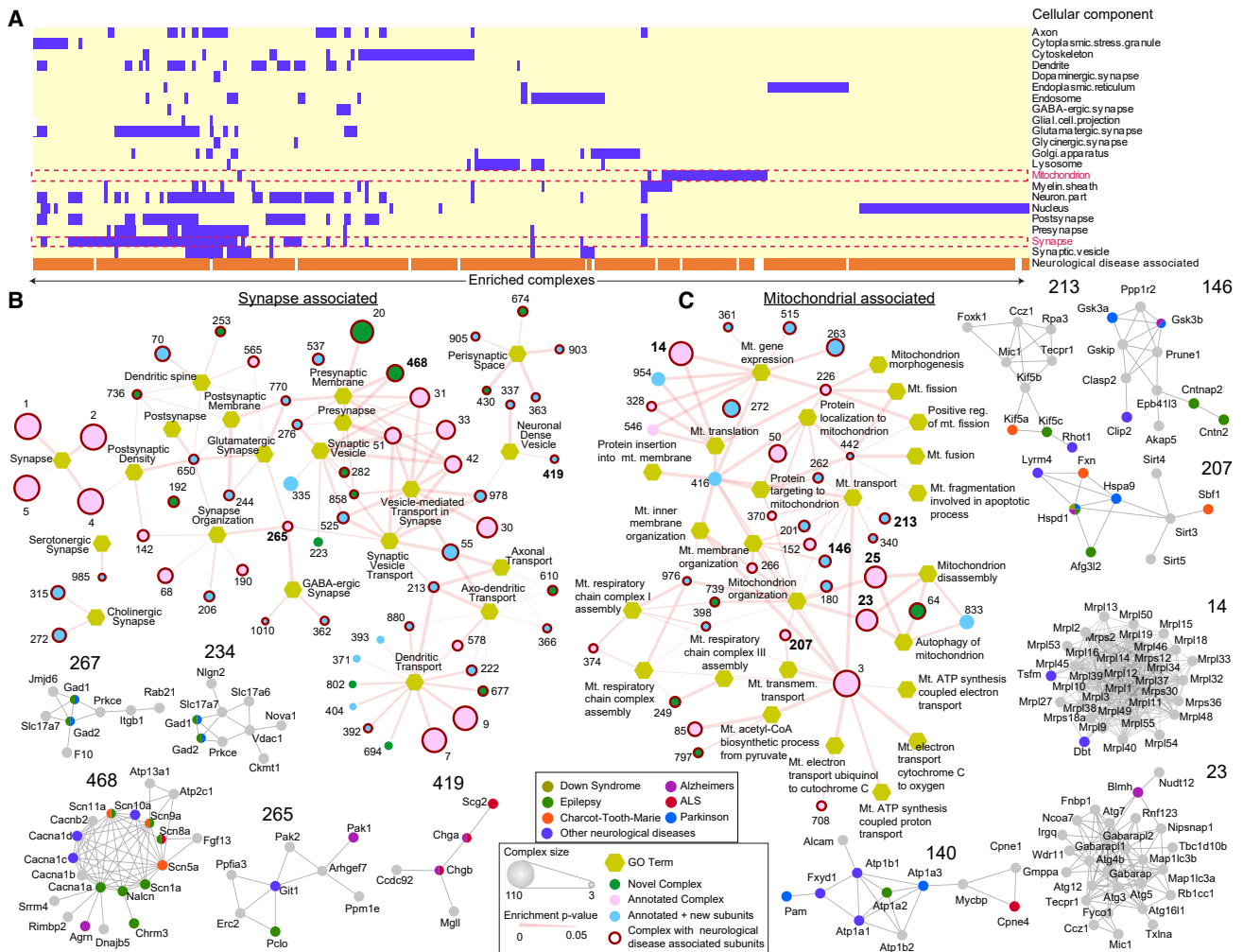
Figure 4. Compartmentalized Brain Protein Assemblies

(A) BrainMap assemblies enriched for select neuronal functions (GO annotation terms).

(B) Protein complexes (*circles*; size proportional to subunit number) enriched for synaptic functions (*hexagons*). Red outlines indicate links to neurological disorders (examplars shown at bottom).

(C) Protein complexes enriched for mitochondrial (*Mt.*) functions (examplars shown at right).

Khsrp, Lsm12, Mcrip1, Rptor, Rtcb, Rtraf, Tdp-43, Tia1, Tial1, and Urm1).

We confirmed the interaction between mouse Tdp-43, Hnrnph1, Ddx5, Tia1, and Fus, key members of complex 22, by co-immunoprecipitation (coIP) of either endogenous mouse Tdp-43 or Hnrnph1 from brain cortices of wild-type C57BL/6J mice (n = 4). Whereas RBP components were absent from control IPs using either rabbit or mouse IgG, IP of Tdp-43 co-precipitated endogenous Hnrnph, Ddx5 and Tia1 (Figures 5D and S2A). Likewise, IP of endogenous Hnrnph1 reciprocally pulled-down Tdp-43, Ddx5, Tia1, and Fus (Figures 5E and S2B). Tdp-43 was also detected as a component of complex 168 that contains Dhx36, Elavl2 (HuB), Elavl3 (HuC), Elavl4 (HuD), Ewsr1, Fam98a, Hnrnpul2, Mtdh (Aeg-1), and Prpf3 (Figure 5F; Table S4). In this other RBP complex we confirmed the previously unreported association of Mtdh (Aeg-1) with Tdp-43 in mouse brain lysates by reciprocal coIP (Figure 5G).

RBP-Containing Complexes with Relevance to ALS/FTD Are Affected by Disease State

Complexes 22 and 168 are of particular interest since they contain multiple RBPs genetically linked to ALS and FTD. Mutations in *TARDBP* (Kabashi et al., 2008; Rutherford et al., 2008; Sreedharan et al., 2008), *FUS/TLS* (Kwiatkowski et al., 2009; Vance et al., 2009) and *TIA1* (Mackenzie et al., 2017) lead to the accumulation of pathological insoluble cytoplasmic inclusions in motor and cortical neurons (Mackenzie et al., 2010; Sreedharan et al., 2008). *ATXN2* is a common genetic modifier of ALS, in addition to its role in spinal cerebellar ataxia (Elden et al., 2010), and EWSR1 mutations are associated with the disease (Couthouis et al., 2012).

Given the multiple links of complex 22 to neurodegeneration, we examined a mouse model of ALS to explore the relationship of the components of this complex to disease progression. Over-expression of human TDP-43 (TDP-43$^{WT/WT}$) in mice
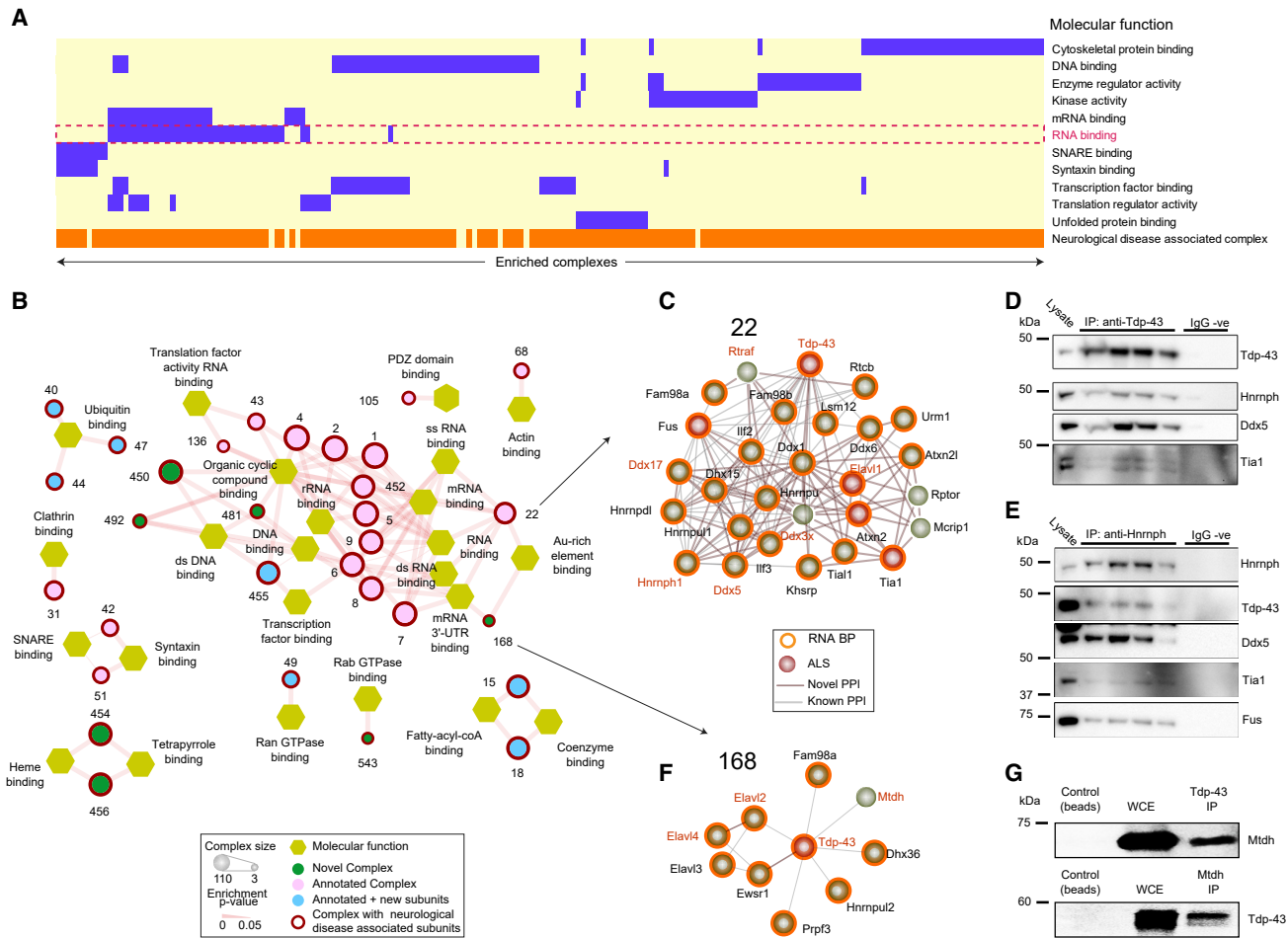
Figure 5. BraInMap Identifies Complexes with Diverse Functions

(A) BraInMap complexes enriched for RNA-binding (dashed box), other annotation terms (purple), and disease associations (orange).

(B) Sub-network of RNA-related complexes (olive); outline (red) indicates a link to neurological disorder.

(C) Putative module (complex 22), composed of 28 RBPs (orange) with links to ALS (red).

(D) Co-immunoprecipitation (western blot) analysis of endogenous Tdp-43 confirms physical associations with Hnrnph, Ddx5, and Tia1 (doublet). Lysate and replicate pull-downs provided; no non-specific signal observed using rabbit or mouse IgG (IgG −ve).

(E) CoIP analysis of endogenous Hnrnph confirms interactions with Tdp-43, Ddx5, Tia1, and FUS/TLS.

(F) Complex 168 (Tdp-43 co-complexed with Elavl2/3/4, Ewsr1, Fam98a, Dhx36, Hnrnpul2,Mdth, and Prpf3).

(G) Reciprocal coIP analysis confirms the association of Mtdh with Tdp-43 in the mouse brain.

results in rapid degeneration of motor neurons with associated pathological aggregates (Wils et al., 2010), whereas depletion of *Atxn2* (a component of complex 22) reduces aggregation of the transgenic TDP-43, increasing motor neuron survival and extending lifespan (Becker et al., 2017). We performed coIP experiments of exogenous human TDP-43 from brain cortices in both disease prone TDP-43$^{WT/WT}$*Atxn2*$^{[+/+]}$ and protected TDP-43$^{WT/WT}$*Atxn2*$^{[+/−]}$ mice (n = 4/group), and used quantitative mass spectrometry to explore changes in complex 22 components linked to neuroprotection (STAR Methods).

Immunoprecipitated TDP-43 pulled-down complex 22 RBPs Ddx1/3x/5/17, Elavl4, Fam98b, Fus, Hnrnpdl/h1/u, Khsrp, Rtcb, and Rtraf from the brains of susceptible TDP-43$^{WT/WT}$*Atxn2*$^{[+/+]}$ mice (Figure 6A), confirming the interactions detected by co-fractionation. Interestingly, in the protected TDP-43$^{WT/WT}$*Atxn2*$^{[+/−]}$ mice these interactions were all reduced,

with the exception of Ddx1 (fold change of 1.00) (Figure 6A). This finding is highlighted in the reproducible reductions observed in TDP-43 binding to Hnrnph1, Ddx3x, Ddx5, Ddx17, and Rtraf (Hnrnph1: −1.83 fold change, −Log$_{10}$ p value = 1.02; Ddx3x: −1.82 FC, −Log$_{10}$ p value = 1.28; Ddx5: −1.95 FC, −Log$_{10}$ p value = 1.09; Ddx17: −1.93 FC, −Log$_{10}$ p value = 1.06; Rtraf: −1.52 FC, −Log$_{10}$ p value = 1.10; n = 4/group) (red text labels; Figure 6A). Reduced coIP of Hnrnph1 with exogenous TDP-43 was further confirmed by immunoblot (Figures S2E and S2F).

Elavl2 and Elavl4 of complex 168 were also depleted in the coIP pull-downs from disease resistant TDP-43$^{WT/WT}$*Atxn2*$^{[+/−]}$ mice as compared to the susceptible strain (Elavl2: −2.43 FC, −Log$_{10}$ p value = 2.05; Elavl4: −1.96 FC, −Log$_{10}$, p value = 1.23; n = 4/group) (red text labels; Figure 6A). A similar trend was observed with Ewsr1, but did not reach statistical
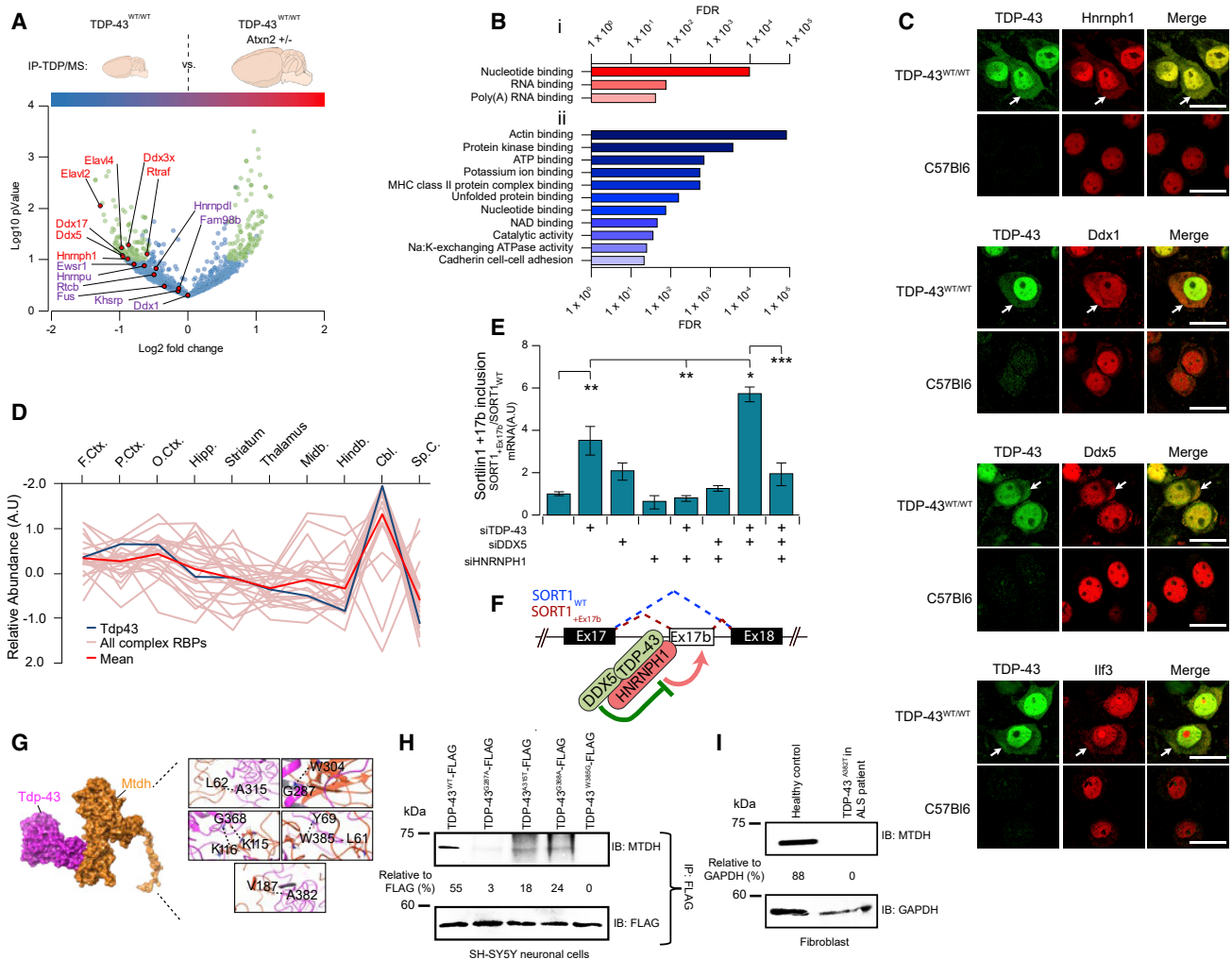
Figure 6. RBP Complexes are Affected in ALS Models

(A) Complex 22 is responsive to neuropathology. Volcano plot summarizing results from coIP pull-downs of exogenous TDP-43 from cortical lysates from diseased (TDP-43$^{WT/WT}$) versus protected (TDP-43$^{WT/WT}$Atxn2$^{[+/−]}$) transgenic mice. Precipitates were subject to quantitative mass spectrometry to define differential binding to pathogenic TDP-43 (> ±0.50 × Log$_2$-fold, −log$_{10}$P < 1, highlighted in green). Interaction of Hnrnph1, Ddx5 and Ddx17 significantly reduced in protected animals (n = 4 per group, Students t test p ≤ 0.05).

(B) Gene ontology molecular function annotations of proteins showing (i) decreased interaction and (ii) increased interaction with transgenic TDP-43 in protected TDP-43$^{WT/WT}$Atxn2$^{[+/−]}$ murine brain. Shown are terms with FDR$^{−1}$ >20.

(C) Confocal immunofluorescent microscopy showing redistribution of complex 22 RBPs (Hnrnph1, Ddx1, Ddx5, and Ilf3) into human TDP-43 positive cytoplasmic accumulations (arrows) in affected cortical neurons of transgenic TDP-43$^{WT/WT}$ mice. This is not seen in wild-type animals. Scale bar represent 20 μm.

(D) The relative brain region expression pattern of Tardbp (TDP-43; dark blue line) closely mirrors the mean complex 22 expression pattern (red line). Other RBP components are traced in pink.

(E) Knockdown (siRNA) of TDP-43 or TDP-43/DDX5 together results in the inclusion of exon 17b of sortilin1 (SORT1) in SH-SY5Y cells (quantified by qPCR), whereas knockdown of interacting partner HNRNPH1 blocks this effect. Graphs show ratio (mean ± SEM) of SORT1 transcripts with or without exon17b (SORT1$_{+Ex17b}$ versus SORT1$_{WT}$); n = 3 per group (ANOVA with Tukey's multiple comparison between all groups: * p < 0.05, ** p< 0.01, *** p < 0.001).

(F) Model of TDP-43 and DDX5 interaction illustrates coordinate inhibition of SORT1 Ex17b inclusion, dependent upon joint association with HNRNPH1.

(G) Structural model of mutations in residues of TDP-43 linked to familial ALS (A315T, G287A, G368A, W385G, and A382T) that map to the interaction interface with MTDH.

(H) CoIP analysis showing a reduced association of MTDH in SH-SY5Y cells expressing FLAG-tagged TDP-43 with ALS-relevant mutations at the predicted interaction interface.

(I) TDP-43 interaction with MTDH is abrogated in ALS-patient-derived fibroblasts carrying a pathogenic mutation (A382T), as compared to fibroblasts from a healthy control.

significance (Ewsr1: −1.74 FC, −Log$_{10}$, p value = 0.91) (purple text labels; Figure 6A). Elavl proteins are cytosolic RBPs, which suggests that Atxn2 modulates the interaction of TDP-43 with cytoplasmic RNP granules, thereby decreasing pathologic insol-

uble inclusions in TDP-43$^{WT/WT}$Atxn2$^{[+/−]}$ mice. These results point to selective dissociation of pathologic TDP-43 from different RBP assemblies upon Atxn2 reduction, consistent with reduced recruitment of TDP-43 to cytoplasmic SGs

resulting in fewer inclusions in the disease resistant strain (Becker et al., 2017; Elden et al., 2010).

In the protected TDP-43$^{WT/WT}$Atxn2$^{[+/−]}$ mice, TDP-43 showed decreased interaction with proteins associated with RNA-binding functional terms (Figure 6Bi), this is exemplified by the volcano plot distribution of heterogeneous nuclear ribonucleoproteins (HNRNPs; Figure S2C). Interestingly, TDP-43 in the protected TDP-43$^{WT/WT}$Atxn2$^{[+/−]}$ mice showed increased interaction with proteins clustering with functional categories such as protein folding, ATP binding, and sodium/potassium ion homeostasis (Figure 6Bii); this is exemplified by the volcano plot distribution of heat shock proteins and protein isomerases that form the unfolded protein response (Figure S2D). These data indicate that the interaction of pathologic TDP-43 with Hnrnph1, Ddx3x, 5 and 17, Rtraf and other RBPs involved in RNA processing is responsive to neuropathophysiological states.

Our observation that TDP-43 shows increased interaction with complex 22 components in disease-affected mice led us to predict that these RBP components would be dysregulated in neurons affected with TDP-43 pathology. Therefore, we investigated distribution of RBPs in the cortices of transgenic TDP-43$^{WT/WT}$Atxn2$^{[+/+]}$ mice by immunofluorescent confocal microscopy.

As shown previously, neurons in the TDP-43$^{WT/WT}$Atxn2$^{[+/+]}$ mice exhibited increased levels of cytoplasmic TDP-43 and pTDP-43 (Becker et al., 2017). Neurons showing cytoplasmic distribution of transgenic TDP-43 (also immuno-positive for phosphorylation at S409/410; Figure S2Hi) exhibited increased cytoplasmic distribution of complex 22 RBPs Hnrnph1, Ddx5, Ddx1, and Ilf3 (Figure 6C). The cytoplasmic distribution of these RBPs mirrored that of TDP-43; for instance, neurons showing focal accumulations of TDP-43 also showed co-localized accumulations of complex 22 RBPs (Figures 6C and S2Gi–S2Giv). These RBPs are restricted to the nuclei of neurons and non-neurons in wild-type C57Bl6 mice. The RBP U2af, which is not a component, does not redistribute out of neuronal nuclei in TDP-43$^{WT/WT}$Atxn2$^{[+/+]}$ mice, nor does the nuclear protein Histone H3 (Figures S2Hii and S2Hiii). This finding emphasizes how TDP-43 pathology may specifically disrupt complex function through subunit sequestration and aberrant cellular relocalization.

By examining our regional proteomic data (Figure 3E), we examined the distribution of complex 22 RBPs (Figure 6D and S2I). Cortical regions show a small relative increase in complex 22 expression compared to normalized whole brain expression (mean ±SEM, F.Ctx 0.27 ±0.09, P.Ctx 0.19 ±0.08, O.Ctx 0.36 ±0.10), while the thalamus, midbrain, and hindbrain exhibited a moderate decrease (mean ±SEM, thalamus −0.46 ±0.07, midbrain (Midb.) −0.25 ±0.13, and hindbrain (Hindb.) −0.46 ±0.13). The region displaying highest expression is the cerebellum (an unaffected area in ALS), while the region showing the lowest expression is the spinal cord (mean ±SEM, Cerebellum (Cbl) 1.32 ±0.23, Spinal Cord (Sp.C) −0.73 ±0.15). The regional expression of Tdp-43 (Figure 6D) and Fus clearly correlate with the pooled expression pattern of complex 22 (Pearson correlation of region mean to Tdp-43 r = 0.977, p < 0.001, to Fus r = 0.951, p < 0.001). As pathological inclusions of TDP-43 are almost universally detected in sporadic ALS-affected spinal cord and FTD-affected cortical neurons (but

rarely in cerebellar neurons) (Brettschneider et al., 2014, 2013), it appears that total regional expression of TDP-43 (or complex 22) does not directly correlate with distribution of pathology, suggesting that relative cellular expression or complexation levels may be more relevant. Notably, as indicated above, while the expression of many RBPs is ubiquitous, complex 22 RBPs are enriched in neurons (Dopaminergic, GABAergic, Glycinergic, Adrenergic, and Serotonergic) as compared with non-neuronal cells (except for oligodendrocytes; Figure 3G). Hence, factors other than absolute level of expression of TDP-43 (or complex 22) may drive motor neuron susceptibility.

HNRNPH1 Function Is Antagonistic to That of TDP-43 and DDX5

We investigated the functional relationships of RBPs associated with TDP-43. We assayed alternative splicing upon depletion of one or more co-complex members, based on the previously reported observation that a reduction in TDP-43 leads to increase inclusion of exon 17b of *SORT1* (Sortilin 1) (Polymenidou and Cleveland, 2011; Prudencio et al., 2012). Expression of *TDP-43*, *HNRNPH1*, and *DDX5* was reduced in SH-SY5Y cells through transfection of siRNA (Figures S2J and S2K). As previously reported (Prudencio et al., 2012), knockdown of TDP-43 (siTDP-43) significantly increased the ratio of SORT1$_{+17b}$ transcript (inclusion of exon 17b) to SORT1$_{WT}$ (predominant transcript lacking exon 17b) (ANOVA with Tukey's multiple comparison between all groups: siTDP-43 = 3.54 A.U. ± 0.082, p < 0.01 versus control/siCtrl non-targeting siRNA) (Figures 6E and S2L). The knockdown of DDX5 slightly increased *SORT1* exon 17b inclusion, but the effect was not statistically significant. Strikingly, coordinated knockdown of TDP-43 and DDX5 together led to a dramatic increase in inclusion of *SORT1* exon 17b (siTDP-43/DDX5 ratio = 5.74 ± 0.36, p < 0.05 versus siTDP-43 alone). Conversely, while knockdown of HNRNPH1 alone had no significant effect on alternate splicing of *SORT1*, it exerted a profound antagonizing effect on the splicing functions of both TDP-43 and DDX5. Knockdown of HNRNPH1 blocked inclusion of *SORT1* exon 17b resulting from depletion of either TDP-43 (siTDP-43/HNRNPH1 = 0.84 ± 0.15, p < 0.01 versus siTDP-43) or TDP-43 and DDX5 together (siTDP-43/DDX5/HNRNPH1 = 1.97 ± 0.51, p < 0.001 versus siTDP-43/DDX5) (Figure 6E).

These data imply that TDP-43 and DDX5 exert cooperative (synergistic) functions in repressing HNRNPH1-mediated inclusion of alternatively spliced exons (Figure 6F). TDP-43 was known to participate with RBPs to facilitate splicing (Mohagheghi et al., 2016), and HNRNPH1 was known to bind the intron upstream of *SORT1* exon 17b (i.e. same binding region as TDP-43), but no significant effect on splicing was recorded. Our study establishes the interaction of specific RBPs as a functional module in which members exert antithetical effects on exon usage.

ALS-Mutations Disrupt TDP-43-MTDH Association within an RBP Complex

TDP-43 co-purified with other factors linked to ALS, including MTDH (metadherin, also known as Astrocyte-elevated gene-1). Molecular docking using the crystal structures of human MTDH and TDP-43 revealed that five residues of TDP-43, which when mutated cause familial ALS (A315, G287, G368, W385, and
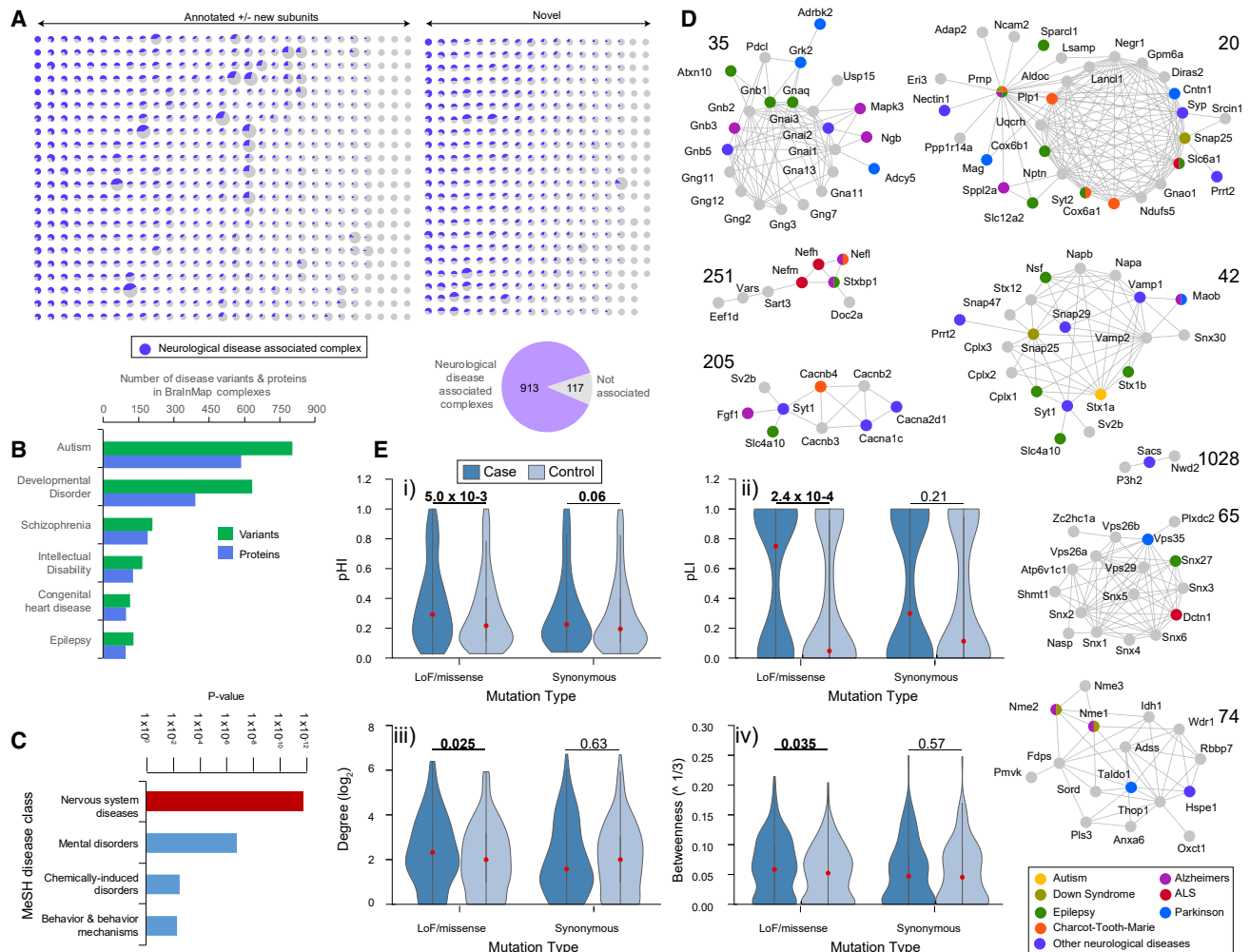
Figure 7. Macromolecular Links to Neurological Disorders

(A) Putative pathophysiological relevance of complexes in BraInMap. Proportion (purple) of subunits of each assembly linked to neurological impairment (see Table S6 for details).

(B) Number of BraInMap components (orthologs) and corresponding human genetic variants associated with specific neuropathologies (see Table S5).

(C) Enrichment (hypergeometric p value) of complex subunits with links to neuropathology as annotated in DisGeNET (Piñero et al., 2017).

(D) Representative complexes associated with Alzheimer's (magenta), autism (yellow), amyotrophic lateral sclerosis (red), epilepsy (green), Down syndrome (olive), Charcot-Toothe-Marie syndrome (orange), Parkinson's (blue), or other neurological disorders (purple).

(E) Enrichment of genes encoding BraInMap components harboring *de novo* variants for (i) haploinsufficiency (pHI) and (ii) pLI (probability a gene is intolerant to LoF mutations) versus synonymous variants in affected individuals in comparison to unaffected controls; (iii) network degree; and (iv) betweenness of genes with *de novo* LoF/missense or synonymous mutations in neurodevelopmental disorder afflicted individuals or unaffected controls. Violin plot width proportional to protein abundance (*red dot*, median); p values (one-tailed U-test; p < 0.05 in bold) shown at the top.

A382), are located in the predicted interaction interface (Figure 6G). Functional studies support the association of TDP-43 with MTDH: TDP-43 plays a role in microRNA (miRNA) biogenesis and activity via Microprocessor and RNA-induced silencing complex (RISC) (Kawahara and Mieda-Sato, 2012) and regulates miRNA loading to RISC (King et al., 2014). MTDH, an RBP, has also been shown to be physically and functionally associated with RISC activity (Yoo et al., 2011).

To independently assess the effect of these ALS-associated variants on MTDH and TDP-43 interaction, we generated four mutant TDP-43-FLAG variants (A315T, G287A, G368A, and W385G), and found that G287A and W385G variants disrupted the association in differentiated SH-SY5Y neuronal cells, while

two other variants partially impaired the interaction (Figure 6H). In addition, we demonstrated a perturbed interaction in fibroblasts isolated from an ALS patient carrying a pathogenic mutation (A382T) in TDP-43 as compared to a healthy control (Figure 6I). Taken together, our data suggest that the disruption of miRNA regulation observed in ALS pathobiology (Eitan and Hornstein, 2016) may reflect disruption of interaction between TDP-43 and factors, such as MTDH.

Modules Connected to Human Neurological Conditions

Conserved components of multiple BraInMap assemblies had strong links to diverse human neurological disorders (Figure 7A). These included macromolecules showing significant

enrichment (Figure 7B) for components associated with psychiatric conditions such as schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics, 2014), neurodevelopmental disorders such as autism (Sanders et al., 2015), and neurodegenerative diseases such as AD, PD, and ALS (Dormann et al., 2010) and for associated genetic variants (Table S9). Relative to other diseases, neurological dysfunction annotations (DisGeNet) (Piñero et al., 2017) (Figure 7C) were often associated with several subunits of certain novel brain complexes. This suggests that disruption of complex function by multiple avenues can lead to similar mechanistic and phenotypic outcomes.

An illustrative example, highlighted in Figure 7D, is complex 42, which includes SNARE protein components necessary for synaptic-vesicle fusion in neurotransmitter release (Chen et al., 2002) that are deficient in neurodegenerative impairments such as in PD (Burré et al., 2010) and Huntington's disease (HD) (Smith et al., 2007). Alterations in SNARE component SNAP25 have also been associated with psychiatric disorders, particularly attention-deficit hyperactivity disorder, in both mice and humans (Brophy et al., 2002; Bruno et al., 2007). In a similar vein, BraInMap complex 35 (discussed further above) contains the ortholog of ATXN10 in which repeat expansion mutations have been shown to cause spinocerebellar ataxia type 10 (Matsuura et al., 2000). Likewise, complex 20 (discussed above) contains a number of subunits associated with neurological disorders, including PLP1 (Hobson and Kamholz, 1993), CNTN1 (Compton et al., 2008), and PRNP. In complex 205, the human ortholog of the voltage-gated L-type calcium channel subunit Cacnb3 has been linked to bipolar disorder (Psychiatric GWAS Consortium Bipolar Disorder Working Group, 2011), while a rare mutation in the ortholog of the subunit synaptogamin1 SYT1 results in severe juvenile motor deficits and cognitive impairment (Baker et al., 2015). Complex 251 neurofilaments (Nefl, Nefm, Nefh) accumulate in certain neurodegenerative diseases and are associated genetically with ALS (Campos-Melo et al., 2018). NEFL mRNA stability is regulated by direct binding to the 3′UTR by TDP-43 and mutant SOD1, the major disease proteins of ALS, potentially dysregulating the stoichiometry of neurofilament polymerization (Ge et al., 2005; Strong et al., 2007).

Another example is complex 65 (Figure 7D), which consists of a number of retromer complex vacuolar protein sorting-associated components (Vps26a/26b/29/35) and sorting nexins (Snx1-6/27), implicating this assembly in endosomal delivery. Mutations in VPS35 impair vacuole dynamics leading to defects in macroautophagy, mitochondrial turnover and AMPA receptor trafficking (Williams et al., 2017) resulting in PD (Vilariño-Güell et al., 2011). Also present in this complex is dynactin 1 (DCNT1), important for retrograde transport of vesicles and autophagic clearance (Laird et al., 2008) and in which mutations cause ALS and Perry's syndrome (Farrer et al., 2009; Münch et al., 2004). These observations mesh with accumulating evidence that autophagic deficiencies underlie the neurological dysfunction seen in diverse clinical disorders, potentially explaining their heterogeneous etiology.

BraInMap Assemblies Are Frequent Targets of Disruptive Mutations Impairing Neurodevelopment

To evaluate whether brain complexes had an elevated rate of disease-linked mutations, we investigated the correspondence of ~21,000 de novo variants previously detected in neurodevelopmental disorder-affected individuals as compared to unaffected controls. The variant data were compiled from 40 different published studies (listed in denovo-db v.1.5), including the Deciphering Developmental Disorders project and genetic studies of autism, schizophrenia, epilepsy, and intellectual disability (Turner et al., 2017). To control for differences in abundance in the enrichment analyses, we compared subsets of proteins (2,298 per group) from BraInMap and background (detectable) proteome with matched abundance distributions for the enrichment analyses (one-tail Fisher's exact test). The same procedures were applied in Figure 7E analyses (calculating node degree, betweenness, pLI, pHI).

As summarized in Table S9, BraInMap was significantly enriched for gene products harboring de novo loss-of-function (LoF) mutations in neurologically impaired individuals (LoF, expected:observed ratio = 1.41, p value = 3.4×10^{-4} by a two-tail binomial test; see STAR Methods), but not for synonymous mutations or variants seen in unaffected controls (Figure 7E). Starting with a curated list of 1,007 known autism-associated gene products (Basu et al., 2009), we again observed significant overlap with BraInMap (expected:observed = 1.50, p value = 3.2×10^{-4}; Table S9). In contrast, gene products with rare synonymous variants from the National Heart Lung Blood Institute Exome Sequencing Project (NHLBI ESP) study exhibited the background rate (expected:observed = 0.86, p value = 1.0).

De novo variants typically occur on one copy of a gene; hence, to confer risk, should arise more frequently in genes susceptible to haploinsufficiency. In neurodevelopmental disorder-affected individuals, orthologs of BraInMap components harboring de novo LoF and missense mutations had, on average, a significantly higher probability of being haploinsufficient (pHI) (Huang et al., 2010) than unaffected controls (median 0.33 versus 0.23, binomial test p value = 5.0×10^{-3}; Figure 7Ei). Components with disruptive de novo mutations were less tolerant to genetic variation, with a higher average probability of LoF intolerance (pLI) (Lek et al., 2016) compared to controls (median 0.78 versus 0.05, p value = 2.4×10^{-4}; Figure 7Eii) or synonymous variants (median 0.33 versus 0.1, p value = 0.21). Taken together, the analyses in Table S9 show that BraInMap assemblies are frequent targets of disease-related variants that potentially impair neurodevelopment.

Disruptive missense mutations in autistic individuals reportedly impact highly connected network (hub) components (Chen et al., 2018). Consistent with this, proteins with LoF and missense mutations in affected individuals exhibited, on average, significantly higher network connectivity ("degree") relative to unaffected controls (median 0.24 versus 0.2, p value = 0.025 by one-tail U-test; Figure 7Eiii). In contrast, no significant difference was observed for synonymous variants (median 0.15 versus 0.2, p = 0.63). Similar trends were evident with other measures of network centrality such as shortest paths or "betweenness" (median 0.06 versus 0.05, p value = 0.035; Figure 7Eiv). Therefore, BraInMap offers a potential mechanistic framework for determining how genetic variants confer clinical risk through interaction perturbation.

DISCUSSION

Tissue, regional and cell-type enriched macromolecules drive brain function, physiology, and disease. However, direct

mapping of molecular connections in the CNS is challenging. While a number of experimental methods have been devised to study protein interactions that occur in neurons (see for example [Zhu et al., 2018] and [Ganapathiraju et al., 2016]), most studies have typically been executed in a piecemeal manner that does not allow for a comprehensive interrogation of the brain interactome. To address this gap, we applied a systematic, data-driven functional proteomic approach. Using deep biochemical fractionation, we identified endogenous protein complexes in murine brain in a near-native context, avoiding artifacts due to epitope tagging or ectopic over-expression. Moreover, it is complementary to existing genetic surveys (e.g. GWAS), cell connectivity maps (i.e., NIH BRAIN initiative), and single-cell transcriptome data.

In doing so, we establish an approach to identifying macromolecular protein complexes in post-mortem tissues, which could be highly useful in studying human brain samples. Rather than viewing ALS/FTD as a TDP-43 proteinopathy, a growing consensus is to consider the condition as resulting from insolubility and splicing defects of a number of RBPs. Our discovery that ALS-associated RBPs natively assemble as a functional splicing module raises the possibility that a more accurate descriptor of ALS/FTD is as an RBP "complexopathy" that results, in part, from splicing defects due to insolubility of a sub-network of RBPs. Therefore, BraInMap represents an important new tool to interrogate the composition, distribution, and function of the macromolecules of the CNS and their role in normal and diseased brain physiology.

Comprehensive characterization of the multi-protein architecture of the mammalian brain represents a milestone for neurobiology. It also provides access to unusual classes of questions. For example, of the 5,677 proteins to a complex, 2,836 (or just under 50%) were assigned to multiple complexes. These "promiscuous" interactors (or "moonlighting proteins"), which are members of multiple complexes, show an enrichment for functional roles linked to protein transport (Benjamini-Hochberg FDR $p = 9.8 \times 10^{-32}$), mRNA processing (Benjamini-Hochberg FDR $p = 1.7 \times 10^{-27}$) and translation (Benjamini-Hochberg FDR $p = 1.6 \times 10^{-17}$), as well as a higher (1.8-fold) average abundance in brain according to PAXdb (students t test $p = 4.7 \times 10^{-3}$) (Wang et al., 2015) as compared to proteins assigned only to one complex (Figure S1D; Table S8).

Given the rapid evolution of mammalian brain, we could also examine the evolutionary trajectories of these macromolecular assemblies by assigning individual proteins an "age" that represents their phylogenetic origin based on ortholog projections (see STAR Methods). In this preliminary investigation, most brain complexes exhibited a mixture of component ages (Figures S1G and S1H; Table S4), suggesting that younger, possibly less tightly bound components represent more recent evolutionary adaptations relative to more ancient assemblies. The previously unreported complexes in our network were also enriched for mammalian proteins, whereas documented assemblies exhibited a higher fraction of components of lower eukaryotic origin (Figure S1F). Taken together, these observations are consistent with the notion that mammal restricted macromolecules are more likely to mediate brain-specific functions, whereas previously described complexes are more likely associated with

house-keeping roles common to multiple lineages. Further in depth analysis of the BraInMap resource is needed to refine and rigorously test this notion.

To support such follow-up studies, BraInMap can be accessed via a dedicated web portal (https://www.bu.edu/dbin/cnsb/mousebrain/) that supports search queries, network visualization, and biological inference. The resource is currently a static representation of neuronal circuits, whereas cellular interaction networks are dynamic. The mammalian brain is impacted by changing developmental and physiological cues and contextual signaling cascades. In principle, our interactome profiling technology can be used to study these network fluxes to reveal interactions underlying particular neurological and pathological states. Thus, follow-up studies using this platform together with sophisticated functional manipulation tools, such as optogenetic control of protein interactions and complex assembly, data science methods based on newer deep learning classifiers, as well as more precise information about macromolecular localization and topology, may define the mechanistic principles driving neurological processes central to healthy and impaired brain function.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Tissue Harvest and Protein Extraction
- METHOD DETAILS
  - 2-D Biochemical Fractionation
  - Dual-Phase Heparin-Ion Exchange Fractionation of Protein Extract I
  - Sample Preparation and Trypsin Digestion
  - LC-MS/MS analysis
  - Regional Sample Preparation for Quantitative Mass Spectrometry
  - Mass Spectrometry
  - MaxQuant Search and Data Analysis
  - Regional Sample Preparation for Co-fractionation Analysis
  - Database Search and Data Analysis
  - Co-immunoprecipitation of Complexed RBPs from C57Bl/6J Mice
  - TDP-43 Immunoprecipitation from Transgenic TDP-43 Murine Brain and Proteomic Analysis
  - TDP-43 Immunofluorescence Imaging from Transgenic TDP-43 Murine Brain
  - Human Cell Culture and Differentiation
  - Human Anti-FLAG/MDTH Immunoprecipitation, Western Blotting, and Affinity-Purification/Mass Spectrometry
  - Mouse Brain Co-immunoprecipitation (Anti-Tdp-43/Anti-MTDH)
  - Structure Modeling and Docking
  - Generation of Human TDP-43 Mutants and Mutagenesis

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cels.2020.03.003.

AUTHOR CONTRIBUTIONS

A.E. conceived and supervised the project. R.P., H.G., R.H., and P.C.H. performed protein fractionation and MS analyses. F.G. and L.Z.M. performed the computational scoring with input from G.B. and A.E. P.E.A.A., R.H., P.A.H, M.T.M., and R.H.M. performed follow-up experiments with supervision from B.W., M.B., and A.E. Additional experiments were performed by S.B. and A.A.A. Fibroblasts with TDP-43 mutations were generously provided by A.R. Transgenic tissue was generated by L.A.B. and A.D.G. S.P., F.G., E.M., K.T., A.N., A.Y., S.C., Y.Z., S.D.W., and G.C. performed data analyses with guidance from A.E., K.L., H.Y., and J.P. S.P. created the web portal and figures with guidance from A.E. P.E.A.A., R.P., S.P., and A.E. wrote the manuscript with input from M.B. and B.W. All authors read and approved the manuscript.

REFERENCES

Arnold, E.S., Ling, S.C., Huelga, S.C., Lagier-Tourenne, C., Polymenidou, M., Ditsworth, D., Kordasiewicz, H.B., McAlonis-Downes, M., Platoshyn, O., Parone, P.A., et al. (2013). ALS-linked TDP-43 mutations produce aberrant RNA splicing and adult-onset motor neuron disease without aggregation or loss of nuclear TDP-43. Proc. Natl. Acad. Sci. USA 110, E736–E745.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat. Genet. 25, 25–29.

Baker, K., Gordon, S.L., Grozeva, D., van Kogelenberg, M., Roberts, N.Y., Pike, M., Blair, E., Hurles, M.E., Chong, W.K., Baldeweg, T., et al. (2015). Identification of a human synaptotagmin-1 mutation that perturbs synaptic vesicle cycling. J. Clin. Invest. 125, 1670–1678.

Basu, S.N., Kollu, R., and Banerjee-Basu, S. (2009). AutDB: a gene reference resource for autism research. Nucleic Acids Res. 37, D832–D836.

Becker, L.A., Huang, B., Bieri, G., Ma, R., Knowles, D.A., Jafar-Nejad, P., Messing, J., Kim, H.J., Soriano, A., Auburger, G., et al. (2017). Therapeutic reduction of ataxin-2 extends lifespan and reduces pathology in TDP-43 mice. Nature 544, 367–371.

Bremer, J., Baumann, F., Tiberi, C., Wessig, C., Fischer, H., Schwarz, P., Steele, A.D., Toyka, K.V., Nave, K.A., Weis, J., and Aguzzi, A. (2010). Axonal prion protein is required for peripheral myelin maintenance. Nat. Neurosci. 13, 310–318.

Brettschneider, J., Del Tredici, K., Irwin, D.J., Grossman, M., Robinson, J.L., Toledo, J.B., Fang, L., Van Deerlin, V.M., Ludolph, A.C., Lee, V.M., et al. (2014). Sequential distribution of pTDP-43 pathology in behavioral variant frontotemporal dementia (bvFTD). Acta Neuropathol. 127, 423–439.

Brettschneider, J., Del Tredici, K., Toledo, J.B., Robinson, J.L., Irwin, D.J., Grossman, M., Suh, E., Van Deerlin, V.M., Wood, E.M., Baek, Y., et al. (2013). Stages of pTDP-43 pathology in amyotrophic lateral sclerosis. Ann. Neurol. 74, 20–38.

Brohée, S., and van Helden, J. (2006). Evaluation of clustering algorithms for protein-protein interaction networks. BMC Bioinformatics 7, 488.

Brophy, K., Hawi, Z., Kirley, A., Fitzgerald, M., and Gill, M. (2002). Synaptosomal-associated protein 25 (SNAP-25) and attention deficit hyperactivity disorder (ADHD): evidence of linkage and association in the Irish population. Mol. Psychiatry 7, 913–917.

Bruno, K.J., Freet, C.S., Twining, R.C., Egami, K., Grigson, P.S., and Hess, E.J. (2007). Abnormal latent inhibition and impulsivity in coloboma mice, a model of ADHD. Neurobiol. Dis. 25, 206–216.

Burré, J., Sharma, M., Tsetsenis, T., Buchman, V., Etherton, M.R., and Südhof, T.C. (2010). Alpha-synuclein promotes SNARE-complex assembly in vivo and in vitro. Science 329, 1663–1667.

Campos-Melo, D., Hawley, Z.C.E., and Strong, M.J. (2018). Dysregulation of human NEFM and NEFH mRNA stability by ALS-linked miRNAs. Mol. Brain 11, 43.

Carrion, M.D.P., Marsicano, S., Daniele, F., Marte, A., Pischedda, F., Di Cairano, E., Piovesana, E., von Zweydorf, F., Kremmer, E., Gloeckner, C.J., et al. (2017). The LRRK2 G2385R variant is a partial loss-of-function mutation that affects synaptic vesicle trafficking through altered protein interactions. Sci. Rep. 7, 5377.

Carvill, G.L., McMahon, J.M., Schneider, A., Zemel, M., Myers, C.T., Saykally, J., Nguyen, J., Robbiano, A., Zara, F., Specchio, N., et al. (2015). Mutations in

the GABA transporter SLC6A1 cause epilepsy with myoclonic-atonic seizures. Am. J. Hum. Genet. *96*, 808–815.

Cavadini, S., Fischer, E.S., Bunker, R.D., Potenza, A., Lingaraju, G.M., Goldie, K.N., Mohamed, W.I., Faty, M., Petzold, G., Beckwith, R.E., et al. (2016). Cullin-RING ubiquitin E3 ligase regulation by the COP9 signalosome. Nature *531*, 598–603.

Chen, S., Fragoza, R., Klei, L., Liu, Y., Wang, J., Roeder, K., Devlin, B., and Yu, H. (2018). An interactome perturbation framework prioritizes damaging missense mutations for developmental disorders. Nat. Genet. *50*, 1032–1040.

Chen, X., Tomchick, D.R., Kovrigin, E., Araç, D., Machius, M., Südhof, T.C., and Rizo, J. (2002). Three-dimensional structure of the complexin/SNARE complex. Neuron *33*, 397–409.

Chiang, P.M., Ling, J., Jeong, Y.H., Price, D.L., Aja, S.M., and Wong, P.C. (2010). Deletion of TDP-43 down-regulates Tbc1d1, a gene linked to obesity, and alters body fat metabolism. Proc. Natl. Acad. Sci. USA *107*, 16320–16324.

Compton, A.G., Albrecht, D.E., Seto, J.T., Cooper, S.T., Ilkovski, B., Jones, K.J., Challis, D., Mowat, D., Ranscht, B., Bahlo, M., et al. (2008). Mutations in contactin-1, a neural adhesion and neuromuscular junction protein, cause a familial form of lethal congenital myopathy. Am. J. Hum. Genet. *83*, 714–724.

Couthouis, J., Hart, M.P., Erion, R., King, O.D., Diaz, Z., Nakaya, T., Ibrahim, F., Kim, H.J., Mojsilovic-Petrovic, J., Panossian, S., et al. (2012). Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. Hum. Mol. Genet. *21*, 2899–2911.

Cromar, G.L., Zhao, A., Xiong, X., Swapna, L.S., Loughran, N., Song, H., and Parkinson, J. (2016). PhyloPro2.0: a database for the dynamic exploration of phylogenetically conserved proteins and their domain architectures across the Eukarya. Database (Oxford) *2016*, baw013.

Dormann, D., Rodde, R., Edbauer, D., Bentmann, E., Fischer, I., Hruscha, A., Than, M.E., Mackenzie, I.R.A., Capell, A., Schmid, B., et al. (2010). ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. EMBO J. *29*, 2841–2857.

Dubiel, D., Rockel, B., Naumann, M., and Dubiel, W. (2015). Diversity of COP9 signalosome structures and functional consequences. FEBS Lett. *589*, 2507–2513.

Eitan, C., and Hornstein, E. (2016). Vulnerability of microRNA biogenesis in FTD-ALS. Brain Res. *1647*, 105–111.

Elden, A.C., Kim, H.J., Hart, M.P., Chen-Plotkin, A.S., Johnson, B.S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M.M., et al. (2010). Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature *466*, 1069–1075.

Elmer, B.M., and McAllister, A.K. (2012). Major histocompatibility complex class I proteins in brain development and plasticity. Trends Neurosci. *35*, 660–670.

Encinas, M., Iglesias, M., Liu, Y., Wang, H., Muhaisen, A., Ceña, V., Gallego, C., and Comella, J.X. (2000). Sequential treatment of SH-SY5Y cells with retinoic acid and brain-derived neurotrophic factor gives rise to fully differentiated, neurotrophic factor-dependent, human neuron-like cells. J. Neurochem. *75*, 991–1003.

Farrer, M.J., Hulihan, M.M., Kachergus, J.M., Dächsel, J.C., Stoessl, A.J., Grantier, L.L., Calne, S., Calne, D.B., Lechevalier, B., Chapon, F., et al. (2009). DCTN1 mutations in Perry syndrome. Nat. Genet. *41*, 163–165.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., et al. (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. *44*, D279–D285.

Forman, M.S., Trojanowski, J.Q., and Lee, V.M.Y. (2004). Neurodegenerative diseases: a decade of discoveries paves the way for therapeutic breakthroughs. Nat. Med. *10*, 1055–1063.

Fratta, P., Sivakumar, P., Humphrey, J., Lo, K., Ricketts, T., Oliveira, H., Brito-Armas, J.M., Kalmar, B., Ule, A., Yu, Y., et al. (2018). Mice with endogenous TDP-43 mutations exhibit gain of splicing function and characteristics of amyotrophic lateral sclerosis. EMBO J. *37*.

Ganapathiraju, M.K., Thahir, M., Handen, A., Sarkar, S.N., Sweet, R.A., Nimgaonkar, V.L., Loscher, C.E., Bauer, E.M., and Chaparala, S. (2016).

Schizophrenia interactome with 504 novel protein-protein interactions. NPJ Schizophr. *2*, 16012.

Ge, W.W., Wen, W., Strong, W., Leystra-Lantz, C., and Strong, M.J. (2005). Mutant copper-zinc superoxide dismutase binds to and destabilizes human low molecular weight neurofilament mRNA. J. Biol. Chem. *280*, 118–124.

Grant, S.G., and O'Dell, T.J. (2001). Multiprotein complex signaling and the plasticity problem. Curr. Opin. Neurobiol. *11*, 363–368.

Havugimana, P.C., Hart, G.T., Nepusz, T., Yang, H., Turinsky, A.L., Li, Z., Wang, P.I., Boutz, D.R., Fong, V., Phanse, S., et al. (2012). A census of human soluble protein complexes. Cell *150*, 1068–1081.

Havugimana, P.C., Wong, P., and Emili, A. (2007). Improved proteomic discovery by sample pre-fractionation using dual-column ion-exchange high performance liquid chromatography. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. *847*, 54–61.

Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F., et al. (2015). A human interactome in three quantitative dimensions organized by stoichiometries and abundances. Cell *163*, 712–723.

Hentze, M.W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. Nat. Rev. Mol. Cell Biol. *19*, 327–341.

Hobson, G.M., and Kamholz, J. (1993). PLP1-related disorders. In Gene Reviews [Internet] (University of Washington).

Hu, L.Z., Goebels, F., Tan, J.H., Wolf, E., Kuzmanov, U., Wan, C., Phanse, S., Xu, C., Schertzberg, M., Fraser, A.G., et al. (2019). EPIC: software toolkit for elution profile-based inference of protein complexes. Nat. Methods *16*, 737–742.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and Predicting Haploinsufficiency in the Human Genome. PLoS Genet *6*, e1001154.

Huttlin, E.L., Bruckner, R.J., Paulo, J.A., Cannon, J.R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M.P., Parzen, H., et al. (2017). Architecture of the human interactome defines protein communities and disease networks. Nature *545*, 505–509.

Hwang, S., Kim, C.Y., Yang, S., Kim, E., Hart, T., Marcotte, E.M., and Lee, I. (2019). HumanNet v2: human gene networks for disease research. Nucleic Acids Res. *47*, D573–D580.

Jahn, O., Tenzer, S., and Werner, H.B. (2009). Myelin proteomics: molecular anatomy of an insulating sheath. Mol. Neurobiol. *40*, 55–72.

Kabashi, E., Valdmanis, P.N., Dion, P., Spiegelman, D., McConkey, B.J., Vande Velde, C., Bouchard, J.P., Lacomblez, L., Pochigaeva, K., Salachas, F., et al. (2008). TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. Nat. Genet. *40*, 572–574.

Kawahara, Y., and Mieda-Sato, A. (2012). TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. Proc. Natl. Acad. Sci. USA *109*, 3347–3352.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Kim, E., Hwang, S., Kim, H., Shim, H., Kang, B., Yang, S., Shim, J.H., Shin, S.Y., Marcotte, E.M., and Lee, I. (2016). MouseNet v2: a database of gene networks for studying the laboratory mouse and eight other model vertebrates. Nucleic Acids Res. *44*, D848–D854.

King, I.N., Yartseva, V., Salas, D., Kumar, A., Heidersbach, A., Ando, D.M., Stallings, N.R., Elliott, J.L., Srivastava, D., and Ivey, K.N. (2014). The RNA-binding protein TDP-43 selectively disrupts microRNA-1/206 incorporation into the RNA-induced silencing complex. J. Biol. Chem. *289*, 14263–14271.

Kwiatkowski, T.J., Bosco, D.A., Leclerc, A.L., Tamrazian, E., Vanderburg, C.R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E.J., Munsat, T., et al. (2009). Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science *323*, 1205–1208.

Lage, K., Karlberg, E.O., Størling, Z.M., Ólason, P.I., Pedersen, A.G., Rigina, O., Hinsby, A.M., Tümer, Z., Pociot, F., Tommerup, N., et al. (2007). A human phenome-interactome network of protein complexes implicated in genetic disorders. Nat. Biotechnol. 25, 309–316.

Laird, F.M., Farah, M.H., Ackerley, S., Hoke, A., Maragakis, N., Rothstein, J.D., Griffin, J., Price, D.L., Martin, L.J., and Wong, P.C. (2008). Motor neuron disease occurring in a mutant dynactin mouse model is characterized by defects in vesicular trafficking. J. Neurosci. 28, 1997–2005.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Exome Aggregation Consortium Analysis of protein-coding genetic variation in 60,706 humans. Nature. 18, 285–291.

Li, F., Tian, X., Zhou, Y., Zhu, L., Wang, B., Ding, M., and Pang, H. (2012). Dysregulated expression of secretogranin III is involved in neurotoxin-induced dopaminergic neuron apoptosis. J. Neurosci. Res. 90, 2237–2246.

Li, T., Wernersson, R., Hansen, R.B., Horn, H., Mercer, J., Slodkowicz, G., Workman, C.T., Rigina, O., Rapacki, K., Stærfeldt, H.H., et al. (2017). A scored human protein-protein interaction network to catalyze genomic interpretation. Nat. Methods 14, 61–64.

Ling, S.C., Albuquerque, C.P., Han, J.S., Lagier-Tourenne, C., Tokunaga, S., Zhou, H., and Cleveland, D.W. (2010). ALS-associated mutations in TDP-43 increase its stability and promote TDP-43 complexes with FUS/TLS. Proc. Natl. Acad. Sci. USA 107, 13318–13323.

Mackenzie, I.R., Nicholson, A.M., Sarkar, M., Messing, J., Purice, M.D., Pottier, C., Annu, K., Baker, M., Perkerson, R.B., Kurti, A., et al. (2017). TIA1 mutations in amyotrophic lateral sclerosis and frontotemporal dementia promote phase separation and alter stress granule dynamics. Neuron 95, 808–816.e9.

Mackenzie, I.R., Rademakers, R., and Neumann, M. (2010). TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. Lancet Neurol. 9, 995–1007.

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21, 3448–3449.

Malty, R.H., Aoki, H., Kumar, A., Phanse, S., Amin, S., Zhang, Q., Minic, Z., Goebels, F., Musso, G., Wu, Z., et al. (2017). A map of human mitochondrial protein interactions linked to neurodegeneration reveals new mechanisms of redox homeostasis and NF-κB signaling. Cell Syst. 5, 564–577.e12.

Matsuura, T., Yamagata, T., Burgess, D.L., Rasmussen, A., Grewal, R.P., Watase, K., Khajavi, M., McCall, A.E., Davis, C.F., Zu, L., et al. (2000). Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nat. Genet. 26, 191–194.

Deciphering developmental Disorders Study (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature 542, 433–438.

Meyer, AdS., Garcia, A.A.F., Souza, A.P.d., and Souza, C.L.d., Jr. (2004). Comparison of similarity coefficients used for cluster analysis with dominant markers in maize (Zea mays L). Genet. Mol. Biol. 27, 83–91.

Migaud, M., Charlesworth, P., Dempster, M., Webster, L.C., Watabe, A.M., Makhinson, M., He, Y., Ramsay, M.F., Morris, R.G., Morrison, J.H., et al. (1998). Enhanced long-term potentiation and impaired learning in mice with mutant postsynaptic density-95 protein. Nature 396, 433–439.

Miller, J.A., Ding, S.L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., et al. (2014). Transcriptional landscape of the prenatal human brain. Nature 508, 199–206.

Mohagheghi, F., Prudencio, M., Stuani, C., Cook, C., Jansen-West, K., Dickson, D.W., Petrucelli, L., and Buratti, E. (2016). TDP-43 functions within a network of hnRNP proteins to inhibit the production of a truncated human SORT1 receptor. Hum. Mol. Genet. 25, 534–545.

Münch, C., Sedlmeier, R., Meyer, T., Homberg, V., Sperfeld, A.D., Kurt, A., Prudlo, J., Peraus, G., Hanemann, C.O., Stumm, G., and Ludolph, A.C. (2004). Point mutations of the p150 subunit of dynactin (DCTN1) gene in ALS. Neurology 63, 724–726.

Nepusz, T., Yu, H., and Paccanaro, A. (2012). Detecting overlapping protein complexes in protein-protein interaction networks. Nat. Methods 9, 471–472.

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N., et al. (2014). The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res. 42, D358–D363.

Pan, J., Meyers, R.M., Michel, B.C., Mashtalir, N., Sizemore, A.E., Wells, J.N., Cassel, S.H., Vazquez, F., Weir, B.A., Hahn, W.C., et al. (2018). Interrogation of mammalian protein complex structure, function, and membership using genome-scale fitness screens. Cell Syst. 6, 555–568.e7.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. 11, 1650–1667.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33, 290–295.

Piñero, J., Bravo, À., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., García-García, J., Sanz, F., and Furlong, L.I. (2017). DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. 45, D833–D839.

Polymenidou, M., and Cleveland, D.W. (2011). The seeds of neurodegeneration: prion-like spreading in ALS. Cell 147, 498–508.

Prudencio, M., Jansen-West, K.R., Lee, W.C., Gendron, T.F., Zhang, Y.J., Xu, Y.F., Gass, J., Stuani, C., Stetler, C., Rademakers, R., et al. (2012). Misregulation of human sortilin splicing leads to the generation of a nonfunctional progranulin receptor. Proc. Natl. Acad. Sci. USA 109, 21510–21515.

Psychiatric GWAS Consortium Bipolar Disorder Working Group (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat. Genet. 43, 977.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Ross, C.A., and Poirier, M.A. (2004). Protein aggregation and neurodegenerative disease. Nat. Med. 10, S10–S17.

Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.W. (2010). Corum: the comprehensive resource of mammalian protein complexes–2009. Nucleic Acids Res. 38, D497–D501.

Rutherford, N.J., Zhang, Y.J., Baker, M., Gass, J.M., Finch, N.A., Xu, Y.F., Stewart, H., Kelley, B.J., Kuntz, K., Crook, R.J.P., et al. (2008). Novel mutations in TARDBP (TDP-43) in patients with familial amyotrophic lateral sclerosis. PLoS Genet. 4, e1000193.

Sánchez-Taltavull, D., Ramachandran, P., Lau, N., and Perkins, T.J. (2016). Bayesian correlation analysis for sequence Count Data. PLoS One 11, e0163595.

Sanders, S.J., He, X., Willsey, A.J., Ercan-Sencicek, A.G., Samocha, K.E., Cicek, A.E., Murtha, M.T., Bal, V.H., Bishop, S.L., Dong, S., et al. (2015). Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron 87, 1215–1233.

Schizophrenia Working Group of the Psychiatric Genomics (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–427.

Scmitt-Ulms, G., Hansen, K., Liu, J., Cowdrey, C., Yang, J., DeArmond, S.J., Cohen, F.E., Prusiner, S.B., and Baldwin, M.A. (2004). Time-controlled transcardiac perfusion cross-linking for the study of protein interactions in complex tissues. Nat Biotechnol. 22, 724–731.

Sherman, D.L., and Brophy, P.J. (2005). Mechanisms of axon ensheathment and myelin growth. Nat. Rev. Neurosci. 6, 683–690.

Shi, H., Zhang, X., Weng, Y.L., Lu, Z., Liu, Y., Lu, Z., Li, J., Hao, P., Zhang, Y., Zhang, F., et al. (2018). m^{6}A facilitates hippocampus-dependent learning and memory through YTHDF1. Nature 563, 249–253.

Slepnev, V.I., and De Camilli, P. (2000). Accessory factors in clathrin-dependent synaptic vesicle endocytosis. Nat. Rev. Neurosci. 1, 161–172.

Small, S.A., and Petsko, G.A. (2015). Retromer in Alzheimer disease, Parkinson disease and other neurological disorders. Nat. Rev. Neurosci. 16, 126–132.

Smith, R., Klein, P., Koc-Schmitz, Y., Waldvogel, H.J., Faull, R.L., Brundin, P., Plomann, M., and Li, J.Y. (2007). Loss of SNAP-25 and rabphilin 3a in sensory-motor cortex in Huntington's disease. J. Neurochem. *103*, 115–123.

Sonnhammer, E.L., and Östlund, G. (2015). InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Res. *43*, D234–D239.

Sreedharan, J., Blair, I.P., Tripathi, V.B., Hu, X., Vance, C., Rogelj, B., Ackerley, S., Durnall, J.C., Williams, K.L., Buratti, E., et al. (2008). TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science *319*, 1668–1672.

Strong, M.J., Volkening, K., Hammond, R., Yang, W., Strong, W., Leystra-Lantz, C., and Shoesmith, C. (2007). TDP43 is a human low molecular weight neurofilament (hNFL) mRNA-binding protein. Mol. Cell. Neurosci. *35*, 320–327.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Tanabe, A., Yanagiya, T., Iida, A., Saito, S., Sekine, A., Takahashi, A., Nakamura, T., Tsunoda, T., Kamohara, S., Nakata, Y., et al. (2007). Functional single-nucleotide polymorphisms in the secretogranin III (SCG3) gene that form secretory granules with appetite-related neuropeptides are associated with obesity. J. Clin. Endocrinol. Metab. *92*, 1145–1154.

The UniProt Consortium (2017). UniProt: the universal protein KnowledgeBase. Nucleic Acids Res. *45*, D158–D169.

Turner, T.N., Yi, Q., Krumm, N., Huddleston, J., Hoekzema, K., F Stessman, H.A., Doebley, A.L., Bernier, R.A., Nickerson, D.A., and Eichler, E.E. (2017). de-novo-db: a compendium of human de novo variants. Nucleic Acids Res. *45*, D804–D811.

Tyanova, S., Temu, T., Carlson, A., Sinitcyn, P., Mann, M., and Cox, J. (2015). Visualization of LC-MS/MS proteomics data in MaxQuant. Proteomics *15*, 1453–1456.

Vance, C., Rogelj, B., Hortobágyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P., et al. (2009). Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science *323*, 1208–1211.

Vanderweyde, T., Apicco, D.J., Youmans-Kidder, K., Ash, P.E.A., Cook, C., Lummertz da Rocha, E., Jansen-West, K., Frame, A.A., Citro, A., Leszyk, J.D., et al. (2016). Interaction of tau with the RNA-Binding Protein TIA1 Regulates tau Pathophysiology and Toxicity. Cell Rep. *15*, 1455–1466.

Vilariño-Güell, C., Wider, C., Ross, O.A., Dachsel, J.C., Kachergus, J.M., Lincoln, S.J., Soto-Ortolaza, A.I., Cobb, S.A., Wilhoite, G.J., Bacon, J.A., et al. (2011). VPS35 mutations in Parkinson disease. Am. J. Hum. Genet. *89*, 162–167.

Wan, C., Borgeson, B., Phanse, S., Tu, F., Drew, K., Clark, G., Xiong, X., Kagan, O., Kwan, J., Bezginov, A., et al. (2015). Panorama of ancient metazoan macromolecular complexes. Nature *525*, 339–344.

Wang, M., Herrmann, C.J., Simonovic, M., Szklarczyk, D., and von Mering, C. (2015). Version 4.0 of PaxDb: protein abundance data, integrated across model organisms, tissues, and cell-lines. Proteomics *15*, 3163–3168.

Waragai, M., Nagamitsu, S., Xu, W., Li, Y.J., Lin, X., and Ashizawa, T. (2006). Ataxin 10 induces neuritogenesis via interaction with G-protein beta2 subunit. J. Neurosci. Res. *83*, 1170–1178.

Williams, E.T., Chen, X., and Moore, D.J. (2017). VPS35, the retromer complex and Parkinson's disease. J. Parkinsons Dis. *7*, 219–233.

Wils, H., Kleinberger, G., Janssens, J., Pereson, S., Joris, G., Cuijt, I., Smits, V., Ceuterick-de Groote, C., Van Broeckhoven, C., and Kumar-Singh, S. (2010). TDP-43 transgenic mice develop spastic paralysis and neuronal inclusions characteristic of ALS and frontotemporal lobar degeneration. Proc. Natl. Acad. Sci. USA *107*, 3858–3863.

Yoo, B.K., Santhekadur, P.K., Gredler, R., Chen, D., Emdad, L., Bhutia, S., Pannell, L., Fisher, P.B., and Sarkar, D. (2011). Increased RNA-induced silencing complex (RISC) activity contributes to hepatocellular carcinoma. Hepatology *53*, 1538–1548.

Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular architecture of the mouse nervous system. Cell *174*, 999–1014.e22.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., et al. (2018). Ensembl 2018. Nucleic Acids Res. *46*, D754–D761.

Zhu, F., Cizeron, M., Qiu, Z., Benavides-Piccione, R., Kopanitsa, M.V., Skene, N.G., Koniaris, B., DeFelipe, J., Fransén, E., Komiyama, N.H., and Grant, S.G.N. (2018). Architecture of the mouse brain Synaptome. Neuron *99*, 781–799.e10.

Zhu, J., Shang, Y., and Zhang, M. (2016). Mechanistic basis of MAGUK-organized complexes in synaptic development and signalling. Nat. Rev. Neurosci. *17*, 209–223.

 CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
(mouse) anti-Tdp-43	Ling Shuo-Chien	FL4
(rabbit) anti- hnRNP-H	Bethyl Labs	A300-511A; RRID: AB_203269
normal mouse IgG	Santa Cruz	sc-2025; RRID: AB_737182
rabbit control IgG	Proteintech	30000-0-AP; RRID: AB_2819035
(rabbit) anti-DDX5	Abcam	ab21696; RRID: AB_446484
(rabbit) anti-FUS/TLS	Proteintech	11570-1-AP; RRID: AB_2247082
(goat) anti-TIA-1	Santa Cruz	sc-1751; RRID: AB_2201433
(rabbit) anti-TDP-43	Proteintech	12892-1-AP; RRID: AB_2200505
(goat) anti-hnRNP H	Santa Cruz	sc-10042; RRID: AB_2295514
(donkey) anti-mouse-HRP	Jackson Immuno	715-035-150; RRID: AB_2340770
(donkey) anti-rabbit-HRP	Jackson Immuno	711-035-152; RRID: AB_10015282
Streptavidin-HRP	Jackson Immuno	016-030-084; RRID: AB_2337238
(mouse) anti-TARDBP	Abnova	H00023435-M01; RRID: AB_548546
(rabbit) anti-TDP-43 phosph-S409/410	Lab of Leonard Petrucelli	N/A
(rabbit) anti-DDX1	ProteinTech	11357-1-AP; RRID: AB_2092222
(rabbit) anti-ILF3	Bethyl Labs	NF90/NF110 Antibody, A303-121A; RRID: AB_10895440
(rabbit) anti-U2AF2	Novus Biologicals	NBP2-04138
(rabbit) anti-Histone H3	Abcam	ab18521; RRID: AB_732917
(chicken) anti-MAP2	Aves	MAP2
(chicken) anti-NeuN	EMD	ABN91; RRID: AB_11205760
Anti-TARDBP	Santa Cruz	sc-376532; RRID: AB_11150837
Anti-MTDH (LYRIC)	Abcam	ab227981
TMT10plex Isobaric Label Reagent Set plus TMT11-131C Label Reagent	Thermo Fisher	Cat# A34808
Pierce™ Trypsin Protease, MS Grade	Thermo Fisher	Cat# 90057
Sep-Pak C18 1 cc Vac Cartridge, 50 mg Sorbent per Cartridge, 55-105 μm Particle Size	Waters	Cat# WAT054955
Thermo Scientific™ EASY-Spray™ HPLC Column	Thermo Fisher	Cat# ES805
Bacterial and Virus Strains		
DH5α competent cells	Thermo Fisher Scientific	Cat# 18255017
E. coli NEB stable	New England Biolabs	Cat# C3040I
Biological Samples		
Control fibroblast (code: pz2)	I IRCCS, Neurology and Laboratory of Neuroscience	Female; Mutation: none; age at biopsy- 45 yrs; Healthy
Mutant TARDBP fibroblasts (code: A577 MF)	IRCCS, Neurology and Laboratory of Neuroscience	Male; Mutation TARDBP p.A382T; age at biopsy- 56 yrs.
Chemicals, Peptides, and Recombinant Proteins		
DL-Dithiothreitol	Millipore Sigma	Cat# D0632
Iodoacetamide	Millipore Sigma	Cat# I6125
Urea	Fisher	Cat# BP169
PhosSTOP™	Millipore Sigma	Cat# PHOSS-RO

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail	Millipore Sigma	Cat# 11836170001
Triethylammonium bicarbonate buffer	Millipore Sigma	Cat# T7408
Formic Acid, 99.0+%, Optima™ LC/MS Grade, Fisher Chemical	Fisher	Cat# A117
Acetonitrile, Optima™ LC/MS Grade, Fisher Chemical	Fisher	Cat# A955
Ammonium Hydroxide, ACS Reagent Grade, 28.0-30.0% as NH3	Fisher	Cat# RABA0020500
Critical Commercial Assays		
Pierce™ Quantitative Colorimetric Peptide Assay	Thermo Scientific	Cat# 23275
Pierce™ BCA Protein Assay Kit	Thermo Scientific	Cat# 23225
Deposited Data		
Protein sequences	UniProt Consortium	PMID: 29425356
Reference complexes	CORUM	PMID: 19884131
Reference complexes	IntAct	PMID:24234451
Reference complexes	GO	PMID: 10802651; PMID: 27899567
Reference complexes	Metazoan	PMID: 26344197; PMID: 26870755
Reference complexes	Human soluble	PMID: 22939629
Functional evidence	Mouse Net v2.0	PMID: 26527726
Functional evidence	Human Subcellular localization	PMID:28495876
Functional evidence	Cell type and brain region – Mouse brain Proteome	PMID:26523646
Functional evidence	Allen Brain Atlas – Mouse brain	PMID: 17151600
Functional evidence	RNA-Seq from SRA	PMID: 21062823
Functional evidence	GeneCards	PMID: 27322403
Protein abundances	PAXdb	PMID: 25656970
Disease association	DisGenNET	PMID: 27924018
Experimental PPIs	AP/MS, BF/MS, Y2H	PMID: 22939629; PMID: 26344197; PMID: 26496610; PMID: 28514442
Gene expression	Sc-RNA-Seq	PMID: 30096314; PMID: 30096299
RNA binding protein assignment	Census of RNA binding proteins	PMID: 25365966
Functional PPIs	Human Net v2.0, Mouse Net v.2	PMID: 30418591; PMID: 26527726
Domains and Orthologues	PhyloPro 2.0	PMID: 26980519
RNA-Seq database	Gene expression Omnibus	PMID: 27008011
De novo variants	denovo-db v.1.5	PMID: 27907889
Autism genes	SFARI (Mar 5, 2018)	PMID: 19015121
Population variants	NHLBI ESP exome-sequencing study	PMID: 23201682
Protein interaction interface	Interactome INSIDER	PMID: 29355848
Mouse to human gene mapping	Ensemble 92	PMID: 29155950
Human PPI databases	Inweb3, InWeb_IM	PMID: 17344885, PMID: 27892958
Human prenatal gene expression	BrainSpan	PMID: 24695229
Schizophrenia risk genes	N/A	PMID: 25056061
Autism risk genes	N/A	PMID: 26402605
Developmental disorder associated genes	N/A	PMID: 28135719
Experimental Models: Cell Lines		
SH-SY5Y	ATCC	CRL-2266; RRID: CVCL_0019
Experimental Models: Organisms/Strains		
CD1 wild type mice	N/A	N/A
C57Bl/6J	JAX labs	000664; RRID: IMSR_JAX:000664

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TDP-43WT/WT	JAX via AD. Gitler	012836; B6;SJL-Tg(Thy1-TARDBP)4Singh/J; RRID: IMSR_JAX:012836
Atxn2[+/-]	JAX via AD. Gitler	101043; B6129SF1/J; RRID: IMSR_JAX:101043
CamKCreER TDP-43 KO	Gift from P. Wong	Chiang et al., 2010. PMID: 20660762
Oligonucleotides		
siGENOME Human TARDBP	Dharmacon	siTDP, 23435
siGENOME Human HNRNPH1	Dharmacon	siH1, 3187
siGENOME Human DDX5	Dharmacon	siDDX5, 1655
siGENOME Non-Targeting siRNA Pool #1	Dharmacon	siCtrl
SORT1total_Ex15_F	Life Tech.	5'-TCCATCTGCCTCTGTTCCCTG
SORT1total_Ex16_R	Life Tech.	5'-GGTGTTCTTCTCTTCCGTACAGACAA
SORT1WT_Ex17_F	Life Tech.	5'-TGGGGTAAATCCAGTTCGAG
SORT1WT_Ex17-18_R	Life Tech.	5'-GACTTGGAATTCTGTTTTTCCGGAC
SORT1+17b_Ex17b_F	Life Tech.	5'-AATCCAGCTCTGCCTCCTCT
SORT1+17b_Ex18_R	Life Tech.	5'-TCCCACGATGGCCAGGATAA
ACTB_931F	Life Tech.	5'-GACAGGATGCAGAAGGAGAT
ACTB_1011R	Life Tech.	5'-GTACTTGCGCTCAGGAGGA
TARDBP-outer_Forward	Millipore-Sigma	CAAGATGAGCCTTTGAGAAGC
TARDBP-outer_Reverse	Millipore-Sigma	AGAGCTGCCAGGAAACAGC
TARDBP-G287A_ Forward	Millipore-Sigma	AATCAGGCTGGATTTGGTAATAGCAGAGGG
TARDBP-G287A_ Reverse	Millipore-Sigma	AAATCCAGCCTGATTCCCAAAGC
TARDBP-A315T_ Forward	Millipore-Sigma	TTGGTACGTTCAGCATTAATCCAGCC
TARDBP-A315T_Reverse	Millipore-Sigma	GAACGTACCAAAGTTCATCCCACC
TARDBP-G368A_ Forward	Millipore-Sigma	GCCTTCGCTTCTGGAAATAACTCTTATAGTGG
TARDBP-G368A_ Reverse	Millipore-Sigma	CCAGAAGCGAAGGCCTGG
TARDBP-W385G_ Forward	Millipore-Sigma	AATTGGTGGCGGATCAGCATCCAATGC
TARDBP-W385G_ Reverse	Millipore-Sigma	ATCCGCCACCAATTGCTGCACC
Recombinant DNA		
pENTR-TARDBP-G287A	Addgene	141323
pENTR-TARDBP-A315T	Addgene	141324
pENTR-TARDBP-G368A	Addgene	141325
pENTR-TARDBP-W385A	Addgene	141326
pLD-puro-Cc-TARDBP-WT-VA	Addgene	141327
pLD-puro-Cc-TARDBP-G287A-VA	Addgene	141328
pLD-puro-Cc-TARDBP-A315T-VA	Addgene	141329
pLD-puro-Cc-TARDBP-G368A-VA	Addgene	141330
pLD-puro-Cc-TARDBP-W387G-VA	Addgene	141331
Software and Algorithms		
Sequence database searching	MaxQuant 1.5.5.1 & 1.6.0.16	PMID: 19029910
Ortholog mapping	InParanoid8	PMID: 25429972
PPI Prediction	EPIC	PMID: 31308550
Complex prediction	ClusterONE	PMID: 22426491
Network visualization	Cytoscape v. 3.5.1	PMID: 14597658
Gene Set Enrichment Analysis	GSEA	PMID: 16199517; PMID: 12808457
Enrichment analysis	BinGO 3.0 Cytoscape App	PMID: 15972284
Enrichment analysis	DAVID Bioinformatics resource 6.8	PMID: 19131956
Hierarchical clustering	Cluster 3.0	PMID: 14871861
Cluster visualization	Java TreeView v 1.1.6r4	PMID: 15180930
Hypergeometric test	R function	Stats: R package

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
RNA-Seq data analysis	R function	edgeR: R Package
RNA-SEQ analysis	StringTie	PMID:25690850
RNA-SEQ analysis	HiSAT	PMID:25751142
Binomial test	Scipy function	Python package
Mann-Whitney U test	Scipy function	Python package
Network analysis	NetworkX	Python package
PIPER	Schrödinger, LLC	Protein-protein docking
ITASSER	Protein structure and function prediction	PMID: 25549265
Plots	R function	ggplot2: R package
Venn diagram	R function	VennDiagram: R package
Overlap analysis	Venn Draw Tool	http://bioinformatics.psb.ugent.be/webtools/Venn/
Surrogate Variable Analysis	R function	sva: R package
Quantile Normalization	R function	preprocessCore: R package
Heatmap	R function	ComplexHeatmap: R package
R	R version 3.5	R Foundation for Statistical Computing
IsobaricAnalyzer		
Other		
Proteomics data deposition	PRIDE	PRIDE: PXD011304
Proteomics data deposition	BioGRID	To be deposited
Nano-HPLC	Thermo Scientific	EASY-nLC™ 1200 System

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to the lead contact, Dr. Andrew Emili, by email at aemili@bu.edu. Plasmids are available upon request; this study did not generate any other new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Tissue Harvest and Protein Extraction

12 week old male CD1 mice were euthanized and the brains excised, washed several times in ice-cold phosphate-buffered saline (PBS) to remove blood, snap-frozen and stored in -80°C. For protein isolation, tissues were homogenized in ice-cold lysis buffer (10mM Tris-HCl pH7.4, 10% glycerol, 50mM NaCl, 1mM Dithiothreitol (DTT), 1mM Ethylenediaminetetraacetic acid (EDTA), 1% Triton X-114 and cOmplete Mini EDTA-free Protease Inhibitor (Roche)) using a dounce tissue grinder on ice. The lysates were kept on ice for 30 minutes and centrifuged at 20,000rcf for 10 minutes to pellet cellular debris. Prior to fractionation, lysates were treated with 100 units/ml Benzonase (Sigma) to remove nucleic acids and further clarified by centrifugation to remove debris. A Bradford assay was performed to determine protein concentrations. The lysate saved as protein extract I.

The detergent-free protein extract was prepared by homogenization of the brain tissue in lysis buffer II (containing 10mM Tris-HCl pH7.4, 20% glycerol, 50mM NaCl, 1mM DTT, 1mM EDTA and complete Mini EDTA-free Protease Inhibitor (Roche)). The suspension was incubated on ice for 30 minutes and then centrifuged at 4,000rcf for 10 minutes. The supernatant was saved as protein extract II while the pellet was resuspended in detergent containing lysis buffer (10mM Tris-HCl pH7.4, 20% glycerol, 50mM NaCl, 1mM DTT, 1mM EDTA, 1% Triton X-114 and protease inhibitors). The sample was incubated with gentle shaking for 45 min at 4°C and centrifuged at 6,000rcf for 10 min. The supernatant was saved as protein extract III. 100 units/ml Benzonase was then added to both protein extracts II and III and protein concentrations measured by Bradford assay.

METHOD DETAILS

2-D Biochemical Fractionation
Isoelectric Focusing (IEF) Fractionation
The protein extract was fractionated by isoelectric focusing using a MicroRotofor IEF cell (Bio-Rad) set up. 3mg of total protein were added to IEF running buffer (20% glycerol, 2% IPG buffer pH5-8) and an electric field at a constant power of 1W was applied to the

focusing cell while the voltage and current were limited to 350V and 2500 mA, respectively. The separation was continued 150 min and stopped after the voltage held constant for about 45 min. Five fractions per sample were collected across a pH range of 5-8.

Ion Exchange (IEX) Fractionation of IEF fractions

Each IEX fraction was subjected to IEX-HPLC separation using mixed-bed PolyCATWAX chromatography columns (200 × 2.1mm i.d., 5μm, 1000-Å) purchased from PolyLC Inc (MD, USA) without any preparation step. Depending on the pH of IEF fractions collected, an IEX buffer system of Tris pH8 or MES pH6 was used. IEX buffers were always freshly prepared with HPLC grade H_2O and comprised of a low salt buffer A (10mM Tris-HCl pH8 (or 10mM MES pH6), 0.01%-NaN_3, 5%-Glycerol) and high salt buffer B (buffer A + 1.5M NaCl). All HPLC fractionations were performed using an Agilent 1260 infinity HPLC system (Agilent Technologies, ON, Canada). The PolyCATWAX IEX column was equilibrated with running low salt buffer A for 30 minutes immediately before loading protein samples. Bound proteins were eluted from the column using a linear gradient to 30% buffer B from 5 to 95 min, followed by a gradient to 100% buffer B from 95 to 105 min and an isocratic hold at 100% B until 120 min. Protein elution was monitored by absorption at 260 and 280 nm. The gradient was run at a flow rate of 0.2ml/min and 60x 0.4ml fractions were collected (the first and last fractions with no peak at 280 nm were discarded). Fractions (i.e. 46 fractions per IEX run, 230 fractions for entire 2-D IEF-IEX fractionation experiment) were prepared and by LC-MS/MS as described.

Ion Exchange Fractionation of Protein Extract I

A total of 2mg soluble protein of protein extract I was loaded to a PolyCATWAX column (200 × 2.1mm i.d., 5μm, 1000-Å). A MES pH6 buffer system (described above) was employed and elution of bound proteins was achieved through application of a linear gradient to 15% buffer B from 2 to 80 min, followed by a gradient to 50% buffer B from until 140 min and a final 20 min long gradient of 50%–100% buffer B. An isocratic hold at 100% B applied until 180 min to elute tightly bound proteins. A total of 90x 0.4ml fractions were collected using a flow rate of 0.2ml/min.

Ion Exchange Fractionation of Protein Extracts II and III

1.2 to 1.5mg total proteins in cytoplasmic and membrane extracts were fractionated on a PolyCATWAX column (200 × 2.1mm i.d., 5μm, 1000-Å) using the MES pH6 buffer system. Protein extracts were resolved using a 120 min gradient program as follows: A linear gradient to 20% buffer B from 2 to 60 min, a gradient to 60% buffer B from 60 to 90 min followed by final 10 min gradient to 100% buffer B and 20 min run with 100% buffer B. 60 fractions by 2 min intervals and using a flow rate of 0.2 ml/min were collected.

Dual-Phase Heparin-Ion Exchange Fractionation of Protein Extract I

In order to enrich low abundance nuclear proteins, a TSKgel Heparin-5PW affinity column (75 × 7.5mm i.d., 10μm, 1000-Å) hyphenated with a PolyCATWAX mixed-bed ion exchange column (200 × 4.6mm i.d., 5μm, 1000- Å). 4mg of protein was loaded to columns and the MES pH6 buffer system was used to resolve multi-proteins complexes in protein extracts. A 240 min elution program consisting of a 10 min gradient with 100% buffer A, followed by a 120 min gradient from 0 to 15% buffer B, a 60 min gradient from 15 to 50% buffer B and a 30 min gradient to 100% buffer B followed by 30 min isocratic hold at 100% buffer B was applied to resolve and fractionate proteins. A total of 120x 0.5ml fractions were collected using a flow rate of 0.25ml/min.

Sample Preparation and Trypsin Digestion

HPLC protein fractions were precipitated overnight at 4°C by adding 10% v/v Trichloroacetic acid (TCA). The fractions then precipitated at 20,000rcf for 30 min and the pellets washed twice with 300μl ice-cold acetone. The pellets were air dried then dissolved in 90μl 50mM NH_4HCO_3. The samples were reduced by adding DTT (Thermo) to a final concentration of 5mM and incubated for 20 min at 50°C with gentle agitation. The samples were cooled to room temperature and alkylated by adding 10mM Iodoacetamide (Sigma) and incubation in the dark for 20 min. To quench excess of Iodoacetamide, 5mM DTT was added to each sample. The protein fractions were then digested by adding 1μg of mass spectrometry grade trypsin gold (Promega) and incubated overnight at 37°C with gentle agitation. The digestion was quenched by adding formic acid (FA) to 1% v/v final concentration and the peptide mixture was subjected to purifying using ziptip C18 tips (Millipore). The ziptip C18 tips were first conditioned with 10μl acetonitrile and then equilibrated with 2x 10μl of 0.1% trifluoroacetic acid (TFA). After loading the peptide mixture to ziptip C18 tips, the samples were washed three times with 0.1 v/v TFA and eluted with 2x 10μl elution buffer (80% acetonitrile, 0.1% TFA). The desalted peptides then lyophilized by using Speed-Vac (Thermo Scientific) and dissolved in 1% FA prior to LC-MS/MS analysis.

LC-MS/MS analysis

All LC-MS/MS analyses performed on an EASY nLC 1200 system (Thermo Scientific) coupled to a Q Exactive HF mass spectrometer equipped with an EASY-Spray ion source (all from Thermo Scientific). A C18 Acclaim PepMap 100 pre-column (3μm, 100 Å, 75μm × 2cm) hyphenated to a PepMap RSLC C18 analytical column (2μm, 100 Å, 75μm × 50cm) (all from Thermo Scientific) was used to separate peptide mixtures prior injection into the mass spectrometer. Depending on sample complexity in each fractionation experiment, 60 or 90-min gradients were used to elute peptides from columns. The quality of LC-MS/MS analysis was repeatedly controlled for by running Trypsin-digested BSA MS Standard (BioLabs) between sample runs.

Regional Sample Preparation for Quantitative Mass Spectrometry

Tissue from 10 brain regions were dissected from four 12 week old male CD1 mice: 1) Frontal Cortex, 2) Parietal Cortex, 3) Occipital Cortex, 4) Hippocampus, 5) Striatum, 6) Thalamus and Hypothalamus, 7) Midbrain (including Substantia Nigra), 8) Hindbrain (including Pons and Medulla), 9) Cerebellum and 10) Spinal Cord. Each was individually placed in 8M urea with phosphatase

(PhosSTOP™, Roche) and protease (cOmplete™, Roche) inhibitors, then sonicated (1 minute, in 2 second pulses) on ice. Sonicated samples were snap-frozen in liquid N$_2$ and then sonicated again as described above. Proteins were reduced for 1 hour with 5mM dithiothreitol (DTT) and alkylated for 30 minutes with 15mM iodoacetamide in the dark. Protein concentration was estimated using a BCA kit (Pierce™ BCA Protein Assay Kit, Thermo) and about 1mg of each sample was allocated for trypsin digestion. Prior to digestion, the 8M urea solution was diluted to 1M with 50mM ammonium bicarbonate. Each sample was digested overnight at 37°C with 10μg sequencing grade trypsin (Pierce™ Trypsin Protease, MS Grade, Thermo)

Prior to TMT (Tandem Mass Tag) labeling, peptides were extracted from each digested sample using c18 Sep-Pak (Waters, 50mg cartridge) and peptide concentrations were measured with a peptide quantification assay (Pierce™ Quantitative Colorimetric Peptide Assay, Thermo). Sample peptide concentrations were adjusted to 1μg/μL in 100mM triethylammonium bicarbonate (TEAB), and 100μg of the sample was aliquoted for labelling. A common pool of samples was generated by combining equal parts of each of the 40 samples. Peptides from the samples and the pool were labelled with 0.4mg and 1.6mg of TMT label, respectively (TMT11plex™ Isobaric Label Reagent Set, 1 x 0.8mg, Thermo). Ten labelled regions from each mouse along with an aliquot of the common pool were pooled and fractionated by high pH reverse-phase HPLC into 12 fractions. Mobile phase A was 0.1% ammonium hydroxide and 2% acetonitrile, mobile phase B was 0.1% ammonium hydroxide and 98% acetonitrile. Fractions were collected over a 48 minute gradient.

Mass Spectrometry

Samples were analyzed by a Q Exactive HFX mass spectrometer connected to Easy nLC 1200 reverse-phase chromatography system (Thermo Scientific). Mobile phase A was 0.1% formic acid and 2% acetonitrile, mobile phase B was 0.1% formic acid and 80% acetonitrile. Peptides were resuspended in 0.1% formic acid for loading. The samples were loaded onto a nano-trap column with mobile phase A, (75μm i.d. × 2 cm, Acclaim PepMap100 C18 3μm, 100Å, Thermo Scientific) and were separated over an EASY-Spray column, (50 cm × 75 μm ID, PepMap RSLC C18, Thermo Scientific) by an increasing mobile phase B gradient over 180 minutes at a flow rate of 250 nL/min. The mass spectrometer was operated in positive ion mode with a capillary temperature of 300°C, and with a potential of 2100V applied to the frit. All data were acquired with the mass spectrometer operating in automatic data dependent switching mode. A high resolution (60,000) MS scan (350-1500m/z) was performed using the Q Exactive to select the 12 most intense ions prior to MS/MS analysis using HCD (NCE 33, 45,000 resolution).

MaxQuant Search and Data Analysis

Raw files were searched in MaxQuant Version 1.6.0.16 against the *Mus musculus* canonical Swiss-Prot proteome downloaded from UniProt on January 24, 2019. Two missed cleavage events were allowed and carbamidomethylation of cysteine was set as a fixed modification while variable modifications were oxidation of methionine and acetylation of protein N-termini. Reporter ion MS2 was used for quantification with 11plex TMT and a reporter mass tolerance of 0.003 Da. Peptide search tolerance was set to 4.5ppm for MS1, and MS2 fragment tolerance was set to 10ppm. Match between runs was active with an alignment window of 20 min and a match window of 0.7 min. The obtained protein intensities of each sample were first normalized to its median for each of the 10 brain regions, the replicates summed and then normalized to the intensities of the reference pool. The summed normalized intensities were used for enrichment analysis to detect regional specificity of BraInMap complexes.

Regional Sample Preparation for Co-fractionation Analysis

Snap-frozen mouse brain tissues (frontal cortex, parietal cortex, occipital cortex, hippocampus, striatum, thalamus, midbrain, hindbrain, and spinal cord) were transferred to 2ml microcentrifuge tubes and homogenized using 2x5mm stainless steel grinding beads. The tissues were homogenized for four 0.5 min cycles in Mixer Mill (MM400, Retsch Technology) at 25 Hz. Ground tissues were solubilized in a mild-detergent buffer (10mM Tris-HCl, 250mM Sucrose, 5mM MgCl$_2$, 1mM DTT, 5mM ATP, 1% DDM) containing protease and phosphatase inhibitors (Roche) and treated with Benzonase at 100 units/ml for 30 min at 4°C. The crude lysates were centrifuged at 18,000rcf for 10 mins at 4°C and the clarified supernatant collected. Protein concentration in the clarified lysates was estimated with Brad-Ford assay (Bio-Rad). The protein extracts were further clarified at 14,000rcf for 30 min at 4°C and fractionated using an optimized volatile-salt based IEX-HPLC fractionation approach (manuscript in preparation). We deployed a previously described dual IEX-HPLC elution gradient (Havugimana et al., 2007), comprising PolyWAX LP and PolyCAT A (200 x4.6 mm i.d., 5μm, 1000-A; PolyLC Inc) column in series, to generate a total of 960-IEX protein fractions (i.e., 96 fractions per mouse regional tissue). The fractions were dried in a speed vac, digested, and each set of 96 fractions was labeled with a unique Tandem Mass Tag (TMT) using the TMT-10plex kit (ThermoFisher Scientific). The TMT-labeled fractions were pooled and desalted. The desalted samples were then analyzed via LC-MS/MS using a Q Exactive Orbitrap HF mass spectrometer (ThermoFisher Scientific) (Havugimana et al., 2007).

Database Search and Data Analysis

Raw file for each fraction was searched against the *Mus musculus* canonical Swiss-Prot proteome downloaded from UniProt on January 24, 2019, using 3 search algorithms (X!Tandem, MSGF+, and Comet). MS1 intensities were extracted from the results using the utilities developed in-house as described above. The obtained protein intensities of each fraction were normalized to its median for each of the 10 brain regions and then normalized to the intensities of the reference pool. The protein-protein correlation was calculated for each of the four replicates using the co-elution profile of each protein across

all fractions. Protein pairs in BraInMap complexes that showed high correlation (≥ 0.5) in their co-elution profiles for two or more replicates and that also exhibited a high concordance in terms of their corresponding proteomic expression profiles were selected for further analysis.

Co-immunoprecipitation of Complexed RBPs from C57Bl/6J Mice

The right cortices from four 5 months old C57Bl/6J wild type mice were homogenized in lysis buffer (50mM Tris pH7.4, 150mM NaCl, 2mM EDTA, 0.2% NP-40, 0.05% SDS, 1mM PMSF, 1x HALT PIC (Pierce), PhosSTOP (Roche) and 40U/ml RNasin (Promega)) using a motorized homogenizer. Protein concentration was determined by BCA assay. Co-immunoprecipitation was performed using Direct-IP kits (Pierce) according to the manufacturer's protocol. Briefly, 5μg of either (mouse) monoclonal anti-Tdp-43 (Ling et al., 2010) (FL4; a gift from Ling Shuo-Chien) or (rabbit) polyclonal anti- hnRNP-H (Bethyl Labs; A300-511A) was conjugated to AminoLink resin, blocked (1 hr at RT) with 1% BSA in lysis buffer then washed with lysis buffer. Negative controls were performed using normal mouse IgG (Santa Cruz) and rabbit control IgG (Proteintech). 1mg of sample lysates was pre-cleared by incubation (1 hr at 4°C) with control agarose resin, before incubating overnight at 4°C in IP columns. The following day, the flow-throughs were collected then the columns washed 3 times with lysis buffer, once with lysis buffer containing 0.1% SDS then eluted by incubating at 98°C for 10 mins in TBS with 2× LDS and 1× reducing agent (Life Tech.). Samples were resolved in 4-12% BisTris Bolt gels (Thermo) with 10μg lysates, transferred to 0.45μM PVDF membrane, blocked in 5% non-fat dry milk in TBSt and probed overnight at 4°C with the antibodies as follows. IP-Tdp-43 immunoblot: 1. (rabbit) anti-hnRNP-H (Bethyl Labs; A300-511A; 1:2000), 2. (rabbit) anti-DDX5 (Abcam; ab21696; 1:2000), 3. (goat) anti-TIA-1 (Santa Cruz; sc-1751; 1:300), 4. (rabbit) anti-TDP-43 (Proteintech; 12892-1-AP; 1:2000). IP-Hnrnph immunoblot: 1. (mouse) anti-Tdp-43 (FL4; 1:4000), 2. (rabbit) anti-DDX5-Biotin (ab21696) 3. (goat) anti-TIA-1 (sc-1751), 5. (rabbit) anti-FUS/TLS-Biotin (11570-1-AP), 5. (goat) anti-hnRNP H (Santa Cruz; sc-10042; 1:2000). Where indicated, 10μg primary antibodies were conjugated to Biotin using One-Step Antibody Biotinylation kit (Miltenyi Biotec). Blots were probed with secondary (donkey) anti-mouse-HRP, (donkey) anti-rabbit-HRP (Jackson; 1:5000) or Strepavidin-HRP (Jackson; 0.2μg/ml) 1 hour at RT before activating with SuperSignal HRP substrate (Thermo) and imaging with a ChemiDoc XRS+ (Bio-Rad). Between antibodies, blots were stripped with Restore PLUS (Thermo), blocked, washed and re-probed.

TDP-43 Immunoprecipitation from Transgenic TDP-43 Murine Brain and Proteomic Analysis

Cortical sections from 21 day old wild type (gait score 0/4), TDP-43$^{WT/WT}$; $Atxn2^{[+/+]}$ (gait score 3.75/4) and TDP-43$^{WT/WT}$; $Atxn2^{[+/-]}$ (gait score 2.75/4) mice (Becker et al., 2017) (n=4 per group) were lysed in 50mM Tris-HCl pH7.4, 150mM NaCl, 0.2% NP-40 with 1mM PMSF, cOmplete PIC, PhosSTOP and 40U/ml RNasin by motorized pestle. Samples were spun at 1,000rcf for 5 mins at 4°C and the supernatants collected and assessed for concentration. Direct-IP columns were generated with 10ug each of (mouse) anti-TARDBP (Abnova; H00023435-M01) and immunoprecipitations, using 500μg lysate per column, were bound, washed and eluted as above. Negative control experiments were performed using anti-TARDBP bound columns and hippocampal lysate from a conditional *Tardbp* knockout mouse (gift from Phillip Wong (Chiang et al., 2010)) and using TDP-43$^{WT/WT}Atxn2^{[+/+]}$ lysates in columns bound with normal mouse IgG. Quantitative proteomic analysis of TDP-43 interactions was performed, as previously described (Vanderweyde et al., 2016), using LC-MS/MS, less non-specific interactions identified in negative controls and normalized to the iBAQ levels of TDP-43 in each sample. Protein interactors were excluded if not identified in duplicate or more per group. Equal amounts of TDP-43 immunoprecipitated material from TDP-43$^{WT/WT}Atxn2^{[+/+]}$ and TDP-43$^{WT/WT}Atxn2^{[+/-]}$ mice (n=3) was immunoblotted and probe for hnRNP-H (Bethyl) and DDX5 (Abcam) as detailed above.

TDP-43 Immunofluorescence Imaging from Transgenic TDP-43 Murine Brain

Hemispheres from TDP-43$^{WT/WT}Atxn2^{[+/+]}$ and wild type litter mates were drop fixed in 4% PFA for 48 hours before washing in PBS and storing in 30% sucrose in PBS. Hemispheres were sliced into 30μm sagittal sections, treated 20 mins at room temperature in 1mg/ml sodium borohydrate to block aldehydes, washed in water then mounted to slides. After drying, slides were washed in PBS, incubated 1 hour at 95°C in citrate buffer (Vector Labs; H-3300) then cooled in PBS. Tissue was permeabilized in 0.2% Tween-20 in PBS, blocked in 5% normal donkey serum, 0.05% Tween-20 in PBS then incubated overnight at 4°C with primary antibodies in PBS with 0.5% NDS and 0.05% Tween-20. Primary antibodies used were (mouse) anti-TARDBP (Abnova; H00023435-M01; 1:500), (rabbit) anti-TDP-43 phosph-S409/410 (a gift from Leonard Petrucelli; Rb3655; 1:250), (rabbit) anti-DDX1 (ProteinTech; 11357-1-AP; 1:500), (rabbit) anti-DDX5 (Abcam; ab21696; 1:1000), (rabbit) anti-hnRNP-H (Bethyl; A300-511A; 1:500), (rabbit) anti-ILF3 (Bethyl; NF90/NF110, A303-121A), (rabbit) anti-U2AF2 (Novus; NBP2-04138), (rabbit) anti-Histone H3 (Abcam; ab18521), (chicken) anti-MAP2 (Aves; MAP2) and (chicken) anti-NeuN (EMD; ABN91; 1:500). Slides were then washed with PBS and fluorescently immunolabelled as appropriate with the (donkey) anti-mouse, rabbit, or chicken with Alexafluor conjugates (Jackson Immuno; 1:750). Slides were washed again in PBS, counterstained with DAPI then autofluorescence was quenched by incubating 10 mins at RT in 0.1% Sudan Black in 50% Ethanol. The sections were coverslipped in Prolong Gold antifade reagent (ThermoFisher). Sections were then imaged at 63x on a Zeiss AxioObserver LSM700 confocal with standardized exposures given additional gain to observe cytoplasmic distribution.

Human Cell Culture and Differentiation

SH-SY5Y cells (ECACC, 94030304) were maintained in high-glucose DMEM (Millipore Sigma, D5671) medium supplemented with 10% fetal bovine serum (FBS, Lifetechnologies; 12483020), penicillin (50 u/ml), streptomycin (50μg/ml) (Lifetechnologies;

15070-063), L-glutamine (2mM) (Lifetechnologies; 25030-081). Cells were incubated at 37°C and 5% CO$_2$/95% air with saturated humidity. SH-SY5Y cells were differentiated by all-trans retinoic (at-RA, Millipore Sigma; R2625) and BDNF (eBiosciences; 14-8366-80) as described before (Encinas et al., 2000). For differentiation, cells were plated at the density of 4 X 10^4 cells/cm^2 in complete DMEM medium containing 5% FBS and at-RA acid was added to cells and the medium was changed daily for a total of 5 days. From days 6-12, cells were incubated with DMEM supplemented with penicillin (50μg/ml), streptomycin (50μg/ml), L-glutamine (2mM) and BDNF (20ng/ml) but no FBS and medium was changed every 2-3 days.

Human Anti-FLAG/MDTH Immunoprecipitation, Western Blotting, and Affinity-Purification/Mass Spectrometry

Differentiated live cells were cross linked using the cell membrane-permeable bifunctional cross linking reagent dithiobis[succinimidyl propionate] (DSP, Lomant's reagent, ThermoScientific; 22585). DSP solution was made in DMSO at the concentration of 0.25M and diluted in PBS to a final concentration of 1mM immediately prior to incubation with cells at room temperature. Excess DSP was quenched by reacting with 100mM Tris-HCl pH 7.5. Cells were then washed in 1X PBS twice, detached by incubating with versine at 37°C for 10 minutes, pelleted and lysed in RIPA buffer. Protein concentration was assayed and approximately 10mg protein was incubated in each reaction with protein-specific or control antibodies for 1 hr at 4°C with tumbling. After that, 50μl protein G magnetic microbeads were added and the mixture is incubated for an additional 4 hr at 4°C with tumbling. Subsequently, samples are purified using magnetic columns and washed using detergent-free buffers. For mass spectrometric analysis, purified proteins and their associated partners are eluted and proteolytically-digested overnight at room temperature. Samples are subsequently desalted and purified using ZipTip and analyzed using nLC-MS. For Western blotting, elution was done in Laemlli buffer and a fraction was run on SDS-PAGE, transferred to PVDF membranes and subsequently blotted with primary protein specific antibodies and subsequently with HRP-conjugated secondary antibodies. Visualization was done using luminescence.

Mouse Brain Co-immunoprecipitation (Anti-Tdp-43/Anti-MTDH)

The cell lysate was adjusted to 1ml with RIPA buffer containing PIC and 3μl of each antibody (TDP-43 or MTDH) were added. After 1 hr agitation at 4°C, 100μl of μMACS protein A or G magnetic microbeads (Miltenyi) was added with continued agitation overnight at 4°C. Microbeads suspension was passed through μMACS columns (Miltenyi) equilibrated with RIPA and PIC, and the retained microbeads were washed 2 times with 1ml of RIPA buffer containing 0.1% of detergents and PIC followed by another one wash with 1ml detergent-free RIPA buffer. Proteins bound to the microbeads were released by addition of 100μl Laemmli loading buffer 2X and heated at 95°C. Eluate was analyzed using western blot and visualized using SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher, 34095).

Structure Modeling and Docking

The I-TASSER (Iterative Threading ASSEmbly Refinement) server was used to predict non-resolved full-length structures of human MTDH and TDP-43. The structure of the lowest energy was selected, which was then refined by a fragment-guided molecular dynamic procedure, with the purpose of optimizing the hydrogen-binding network and removing steric clashes. The docking studies were carried out using PIPER. We produced 7,000 structural conformations between two structure chains. For the highest scoring docked structure, we determined residues at the complex interface using Schrodinger to measure the change in solvent accessible surface area between bound and unbound states of this complex. Residues with a minimum 15% solvent accessible surface area in the unbound state whose absolute solvent accessible surface area changes decreased by ≥1.0Å squared were considered to be at the interface of the bound structure.

Generation of Human TDP-43 Mutants and Mutagenesis

Sequence confirmed TDP-43 cDNA clone (HsCD00079870) from the human ORFeome collection (Dana Farber/Harvard Cancer Center DNA Resource Core) was used to generate the mutations, following the PCR-driven overlap extension method (PMID: 29128334). All PCR reactions were performed using the high-fidelity Phusion polymerase. A set of forward and reverse primers were designed to flank the desired mutation sites and the two unique restriction sites at NsbI and EcoRV. At each mutation site, another set of forward and reverse primers that bind at the mutation sites was used, ensuring that their T$_m$ was within ± 5°C of the outer flanking primers. For each mutation, two first-stage PCR reactions were performed with the outer flanking forward primer and the mutation site reverse primer, and vice versa. After a successful PCR, the reaction products were cleaned up using a PCR cleanup kit. For each mutation, a single second-stage overlap-extension PCR was carried out in which an equimolar mixture of two PCR products of the first-stage reactions and the outer forward and reverse flanking primers were used. Overlap extension was verified using agarose gel. Both TDP-43 plasmid (HsCD00079870) and the final PCR products were cut with restriction enzymes, NsbI and EcoRV, following manufacturer's recommendations. Gel purified products were ligated using T4-DNA ligase, transformed into competent DH5α cells, and the successful mutagenesis was verified using Sanger sequencing.

To clone TDP-43 into mammalian expression vectors, we used Gateway LR Clonase II (ThermoFisher) according to the manufacturer's instructions. Briefly, a mixture of TDP-43 wild type or mutant entry vector was mixed with the enzyme mixture and appropriate amount of the destination vector pLD-puro-Cc-VA (Addgene) containing a C-terminal versatile affinity tag containing 3× Flag, 6× histidine and 2× Streptactin epitopes (Flag and His separated by dual tobacco etch virus protease cleavage sites). After the incubation and enzyme inactivation, cloning mixture was transformed into NEB stable competent bacteria (to avoid recombination of the

repetitive lentiviral sequences in the destination vector). Successful cloning was verified using Sanger sequencing at The Centre for Applied Genomics (TCAG), Toronto Hospital for Sick Children sequencing facility.

siRNA Knockdown of Complexed Neuronal RBPs

SH-SY5Y human neuroblastoma cells were maintained in 50:50 DMEM/F12, 10% FBS, Pen/Strep, NEAA and L-glutamine using standard culturing techniques. Cells were plated (DIV0) to 6 well plates (2.0×10^5 cells/well) or 12 well plates (1.0×10^5 cells/well). The following day (DIV1), knockdown in triplicate was achieved using PepMute (SignaGen; according to manufacturer's protocol) with 50nM of the following siRNA SMARTpools (Dharmacon): siGENOME Human *TDP-43* (siTDP-43, 23435), siGENOME Human *HNRNPH1* (siH1, 3187), siGENOME Human *DDX5* (siDDX5, 1655) and siGENOME Non-Targeting siRNA Pool #1 (siCtrl). Where appropriate, siCtrl was included so as equimolar siRNA was added to each well. After 24 hours (DIV2), the media and knockdown reagents were removed and replaced for a total of 72 hours knockdown before collection (on DIV4). SH-SY5Ys from 12 well plates were collected and lysed in RIPA buffer (50mM Tris pH7.4, 150mM NaCl, 2mM EDTA, 1% NP-40, 0.1% SDS, 0.1% sodium deoxycholate, 1mM PMSF, cOmplete PIC (Roche)) and the concentrations determined by BCA reagent. Samples were immunoblotted as above.

qPCR Analysis of SORT1 Exon17b Mis-splicing

Total RNA from siRNA treated SH-SY5Ys from 6 well plates was collected using the RNeasy Minikit (Qiagen). Random hexamer primed cDNA was generated using the High Capacity cDNA Reverse Transcription Kit (Thermo). qPCR was performed using iQ SYBR Green Supermix (Bio-Rad) to detect ACTB (NM_001101.3), total Sortilin 1 (Prudencio et al., 2012) (NM_002959.6; $SORT1_{total}$; Ex15_F 5'-TCCATCTGCCTCTGTTCCCTG, Ex16_ 5'-GGTGTTCTTCTCTTCCGTACAGACAA), $SORT1_{WT}$ (omitting exon 17b; Ex17_F 5'-TGGGGTAAATCCAGTTCGAG, Ex17-18_R 5'-GACTTGGAATTCTGTTTTTCCGGAC) and $SORT1_{+17b}$ (including exon 17b; Ex17b_F 5'-AATCCAGCTCTGCCTCCTCT, Ex18_R 5'-TCCCACGATGGCCAGGATAA). Sample transcripts were normalized to ACTB levels then to the mean of the siCtrl treated group. Plotting and ANOVA with Tukey's multiple comparison posthoc statistical analysis was performed using GraphPad.

QUANTIFICATION AND STASTICAL ANALYSIS

Data Analysis

MS1 intensity elution profiles of 550 fractions from 5 experiments were determined by searching the spectra with MaxQuant version 1.6.0.16 (Tyanova et al., 2015) against the UniProt reviewed *Mus musculus* proteome protein sequence database (version: Feb 21, 2017, (The UniProt Consortium, 2017), number of sequences: 50,915). Searches were performed with fragment ion mass tolerance of 20 ppm, maximum missed cleavage of 2. Oxidation of methionine was considered as variable modification. The false discovery was controlled using a target/decoy approach with false discovery level set to 1%. Only protein groups identified with at least two or more peptides (sum of razor and unique) in more than one fraction were carried forward in the analysis. This resulted in the detection of 4,134 proteins. To increase protein coverage we ran additional database searches using 3 more search algorithms (X!Tandem, MSGF+, and Comet) subject to the same parameters as given above. The results of these 3 search engines were integrated using the MSblender integration tool which led to the identification of 8,075 proteins. Since the results were obtained in the form of MS2 spectral counts, an in-house script was developed to extract MS1 intensities. Three utilities were developed to extract MS1 intensity data. The first pair of programs scanned the X!Tandem, MSGF+ and Comet search results, producing a list of peptides identified for each fraction by a given search engine along with a range of scan numbers for each combination of identified peptide and precursor ion charge state. The range of scan numbers consisted of the lowest and highest scan numbers of MS2 spectra for which the peptide was considered identified. The third program read each list and scanned the associated spectra file, extracting and reporting the highest MS1 peak intensity within plus and minus 10 ppm of the precursor peak m/z for (1) the MS1 spectrum immediately preceding the range, (2) the first MS1 spectrum following the range, and (3) each MS1 spectrum between.

Both sets of MS1 intensities were run through the EPIC prediction tool (Hu et al., 2019) to predict PPI and complexes. Correlation scores were calculated for each experiment using 5 different methods (Euclidean, Bayes, Jaccard, Apex and Mutual information) and 15 additional functional annotation features were included to boost performance and PPI prediction. 678 complexes from GO, IntAct and CORUM (Table S2) were used as the reference set for training the data through machine learning to predict PPI. Complexes were predicted using ClusterOne (Nepusz et al., 2012)and benchmarked against a set of 78 brain specific reference complexes obtained from CORUM for mouse and other orthologs (Human, Rat, Bovine, Rabbit, and Pig; Table S2).

Given that subunits of a complex should reproducibly co-elute, chromatographic profile similarity is taken as a proxy for physical association. We applied machine-learning procedures to capture and weigh different features from the biochemical data (Hu et al., 2019). The results from each fractionation were processed by random forest classifier trained on experimentally-verified co-complexes PPIs from public curated databases to assign PPI confidence scores (CORUM (Ruepp et al., 2010), IntAct (Orchard et al., 2014), GO (Ashburner et al., 2000). To maximize coverage and accuracy, we integrate supporting functional association evidence in the random forest step. After generating a high confident co-elution network we used ClusterONE to generate a set of stable protein complexes from the co-elution network. See below for a more detailed description of each set of the data analysis pipeline and how the model evaluation was conducted (Nepusz et al., 2012).

Removing Low Scoring Proteins

To ensure proper quality across different co-elution experiments we integrated two preprocessing steps: a) removing low scoring proteins, and b) normalizing peptide counts. Both of the filtering steps treated each co-elution experiment as an individual entity and we merged all experiments at a later stage. First, we removed all proteins for which peptides were observed in one fraction only. For example, if protein A was only observed in fraction 21 in a co-elution experiment that protein was discarded. We justify this filtering based on the fact that calculating any kind of co-elution is impossible for a protein with exactly one observation. We observed that some fractions contain more peptide than others, to minimize this fraction bias we performed a column-wise normalization followed by a row-wise normalization. In the column-wise normalization, we divided the number of identified peptides for each protein for each fraction by the total number of peptides in that fraction. For row wise normalization, we divided the number of peptides of a protein in a specific fraction by the total number of identified peptides of that protein.

Co-elution Scores

We expect proteins that are physically interacting will co-elute in our fractionation experiments and thus the elution profile of interaction proteins should be similar. To measure this relationship we deploy several methods that capture elution profile similarity. At its core, these methods are different correlation metrics that are tuned to measure different aspects of correlation. In the formulas for each correlation metric: p_a, p_b denote protein a and protein b in the same co-fractionation experiment, N denotes the total number of proteins and M is the total number of fractions.

Euclidean Distance

The Euclidean distance denotes the distance of two vectors (or two points) in a high-dimensional space (also known as 2-norm). The two points, for which the distance is calculated, represent the protein pair and the number of fractions is the dimension of the space to which the Euclidean distance applies. The Euclidean distance feature uses normalized counts and lies between 0 and the square root of 2, where identical elution profiles have a distance of 0 and elution profiles that differ greatly have a distance close to square root of 2.

Bayes Correlation

In this work, we integrated a novel method (Sánchez-Taltavull et al., 2016) that utilizes a Bayesian framework for calculating correlation scores between two MS2 spectral counts based vectors. Originally, this method was proposed to process RNA-Seq gene expression data that is based on sequence counts for various genes under various conditions. Here we propose to use the same method for protein peptide counts for various proteins across various biochemical fractions. The main advantage of the Bayesian correlation over Pearson correlation is that it considers both measured signal magnitudes and associated uncertainties in those magnitudes. Thus, Bayesian correlation will retain high correlation values if the measurement confidence is high and will prevent high correlation values when the measurement confidence is low. Moreover, it was shown that the Bayesian correlation could be used as a kernel in any kernel based machine-learning method, such as support vector machines, which makes Bayesian correlation a useful feature for our co-fractionation pipeline. To integrate Bayesian correlation we downloaded the R script (http://www.perkinslab.ca/sites/perkinslab.ca/files/Bayes_Corr.R) and integrated into the python pipeline using the rpy python package that allows the import of R code into python. Bayesian correlation calculation scores support three different assumptions of how the priors distributed: uniform, Dirichlet-marginalized and zero count-motivated. Zero count was used here, as it performed better than the others (unpublished data).

Mutual Information (MI)

Mutual information (MI), unlike linear correlation metrics such as PCC, considers information about both linear and nonlinear dependencies. The initial step in calculating MI is to binarize the spectral count vector elements into 'with protein' and 'without protein', since mutual information measures statistical dependence between the two given proteins based on their relative co-elution frequency (% co-eluted fractions) and each protein's individual relative frequency (% fractions containing the respective protein). We binarize the elution matrix by temporarily changing each protein peptide count to 1 if there were spectral counts observed in the fraction and to 0 if not. Thus, P(p_a =1) denotes the individual relative frequency of p_a, which is calculated by dividing the total number of fractions with value 1 for protein p_a by the total number of fractions in the corresponding co-fractionation experiment. Whereas, the joint relative co-elution frequency of protein p_a and p_b named P(p_a =1, p_b =1) is calculated by counting the total number of fractions that contain both p_a and p_b and dividing this number by the total number of fractions. MI is calculated as follows:

$$MI(p_a, p_b) = H(p_a, p_b) - H(p_a) - H(p_b)$$

In the formula above, H(p_a) denotes the entropy of protein a and H(p_a, p_b) the joint entropy with the following formulas:

$$H(p_a) = -\sum_{i}^{\{0,1\}} P(p_a = i) * \log_2(P(p_a = i)) \quad H(p_a, p_b) = -\sum_{j}^{\{0,1\}} \sum_{i}^{\{0,1\}} P(p_a = i, p_b = j) * \log_2(P(p_a = i, p_b = j))$$

Jaccard Score

Jaccard score computes the ratio of how often proteins are eluted in the same fractions and how often proteins are eluted in different fractions. Thus the Jaccard score between two proteins is calculated by counting the number of fractions that contain both proteins and dividing by the number of fractions that have at least one of the two proteins.

Apex

Most proteins tend to elute only at a specific time, and thus the fraction that contains the largest amount of a particular protein is also the most critical fraction for the given protein. Thus, two proteins are considered to be more likely to interact with each other if the fractions that have the largest amount of proteins across all fractions are the same. Based on this assumption, the previous co-fractionation experiments utilized the co-apex score, which scores protein co-elution profiles highly if their respective peak fraction is the same (apex score = 1) or not (apex score = 0).

Functional Evidence

We enriched our experimental data with high quality functional evidence and other brain-related experiments taken from various sources. In order to prevent circular reasoning, we removed all evidences that used information derived from protein complexes. Adding functional evidence only slightly increased the composite score.

MouseNetV2

MouseNetV2 is a functional gene network for the laboratory mouse that combines various functional evidence from both mouse and other model organisms mapped to mouse (Kim et al., 2016).

Allen Brain Atlas

The Allen brain atlas is a gene expression database for mouse brain that contains exhaustive *in situ* experimental data for various mouse brain regions. We extracted expression values for each gene for all available brain regions that are: Isocortex, Olfactory areas, Hippocampal formation, cortical subplate, striatum, pallidum, thalamus, hypothalamus, midbrain, pons, medulla, cerebellum. The expression is measured in expression energy, which is calculated as follows: Within a given area A (voxel or structure), expression energy = (sum of intensity of expressing pixels in A) / (sum of all pixels in A). The final interaction score is derived by calculating Pearson correlation for all protein pairs based on their expression energy.

Published Brain Networks

We also integrated brain data from various other sources. A recently published work on mouse brain proteasome that contains MS expression analysis for 12934 proteins across major brain regions and cell types was integrated. Additionally, we mapped a study of the human subcellular location to their respective mouse orthologs via InParanoid. We calculated Pearson correlation for each protein pair in each experiment respectively.

RNA-Seq Data

RNA-Seq data was extracted from the Gene expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/gds) using their R library, and selected using several criteria to extract high quality brain data. Only transcriptomic RNA-Seq data for adult mouse generated using Illumina HiSeq 2000 and 2500 and having as source tissue one of the following descriptors: brain, cortex, thalamus, striatum, cerebellum, cerebellum dentate, or olfactory bulb was selected. We then use the Sequence read archive (SRA) tool to map each GEO to their respective raw data set on the SRA. In accordance with a recently published Nature protocol (Pertea et al., 2016), we processed the raw RNA-Seq read data by using StringTie (Pertea et al., 2015) and HISAT (Kim et al., 2015) with the Ensembl (Zerbino et al., 2018) mouse reference genome (Mus_musculus.GRCm38.84.gtf) as an annotation source. Once we quantified RNA expression in each experiment, we calculated co-expression based on those experiments using Pearson correlation.

Reference Complexes

We created a comprehensive set of 678 mouse protein complexes by extracting known protein complexes from CORUM, IntAct, and GO. We downloaded a recent set of complexes from CORUM and only kept those complexes annotated with biochemical evidence. We further expanded this set by adding experimentally verified mouse protein complexes from IntAct and protein complexes that we constructed by using GO annotation. GO complexes were derived by taking all genes that are annotated with a complex specific GO annotation. We identified them by taking all experimentally validated GO cellular component annotations that are a leaf annotation (lowest level, i.e. most specific) and are a descendent of the protein complex GO term. Genes with the same GO annotation are grouped together in the same complex. In the next step, we repeated the same procedure for Human protein complexes, followed by a strict one-to-one mapping of human proteins to mouse proteins using only human-mouse pairs that have a 100% InParanoid score (highest confidence score). After obtaining this set of complexes, we performed several preprocessing steps. In the first step, we removed all proteins, for which we have no elution profile, followed by removing all assemblies that have more than 50 members. In an effort to eliminate redundancy we merged all protein complexes that have an overlap coefficient of 0.8 or more. In addition, for complex benchmarking we used a distinct set (i.e. not a part of the reference complexes) of 78 brain associated complexes from CORUM.

Positive and Negative Protein Complexes

As in previous work (Havugimana et al., 2012; Wan et al., 2015), we generated a set of positive co-occurring protein pairs by taking all possible protein pairs that were observed in the same protein complex. Negative protein pairs are all possible protein pairs that are never observed in the same protein complex. For example, we observed protein A and B are members of the same complex we consider them to be part of the positive training set, and if we never observe protein A and C together in any of our reference complexes we would consider them to be part of the negative training set. Furthermore, previous studies showed that co-elution prediction works best when having a ratio of one to five between positive and negative protein pairs. We created that ratio of positive to negative training data points by under sampling negative protein pairs since there are considerably more negative PPIs than positives.

Model Evaluation

Protein complexes consist of multiple proteins and determining if two complexes are matching is a non-trivial problem. The most common way of measuring it is using the overlap coefficient. The overlap between two protein complexes A and B is calculated as follows (note that |A| denotes the number of proteins in complex A):

$$O(A, B) = \frac{|A \cap B|^2}{|A| * |B|}$$

We defined two protein complexes as matching when the overlap score between them is larger than 0.25 since two clusters of the same size would have this score if the intersection set is half of the complex size.

Additionally, we calculated prediction sensitivity, accuracy, positive predictive value, and cluster separation (Brohée and van Helden, 2006). For the following scores, we considered $a_1, \ldots, a_i, \ldots, a_m$ predicted complexes, which we compared to a set of $b_1, \ldots, b_j, \ldots, b_n$ reference complexes, and $T_{i,j}$ denotes the number of proteins that were found in both complexes i and j.

Sensitivity (Sn): the fraction of proteins in predicted complexes that were found in reference complexes.

$$S_n = \frac{\sum_{i=1}^{n} max_{j=1}^{m} t_{i,j}}{\sum_{i=1}^{n} |b_i|}$$

Positive predictive value (PPV): indicates how specific and complete the predicted complexes match the reference complexes. A score of 1 indicates that each predicted complex only overlaps with exactly one reference complex, and a low score indicates low or redundant overlap with the reference.

$$PPV = \frac{\sum_{j=1}^{m} max_{i=1}^{n} T_{i,j}}{\sum_{j=1}^{m} \sum_{i=1}^{n} T_{i,j}}$$

Accuracy (Acc): shows the trade-off between PPV and Sn.

$$Acc = \sqrt{Sn * PPV}$$

Maximum matching ratio (MMR): The MMR was developed to cope with some of the limitations of the PPV. PPV tends to be lower if there is substantial overlap in the reference data (Nepusz et al., 2012), but those overlaps are common in biological data sets such as CORUM. Our merging step only removes highly overlapping clusters, but smaller overlaps are still present. Thus, even if EPIC perfectly predicts the reference complexes it will not achieve a score of 1 for PPV and Sep (clustering-wise separation score suggested previously by (Brohée and van Helden, 2006)). MMR addresses this problem:

$$MMR = \frac{\sum_{i=1}^{n} max_{i=1}^{m} O(n_i, m_j)}{|max_{i=1}^{n} O(n_i, m) > 0|}$$

As established by others (Nepusz et al., 2012), we summarized MMR, overlap score, and accuracy to create the composite score, and we considered the parameter combination with the highest composite score to be the best combination.

Cross Fold Evaluation

Our primary goal was to accurately infer stable protein complexes from the experimental data in order to properly evaluate our performance. Therefore, we measured how well we could reconstruct known reference complexes from our experimental data. We performed a two-fold cross validation to ensure that we have the same amount of complexes for training and validation. To train the model we first split our set of reference complexes 50:50 and then generated positive and negative PPIs for one set and then trained a random forest model to distinguish them. Next, we predicted all PPIs for which we have elution data and retained all PPIs pairs with a random forest score greater than 50%. We then generated protein clusters from these interactions using ClusterOne with default parameters. The performance was evaluated using overlap score, MMR, and accuracy of those predicted clusters against the separate set of brain specific reference complexes. To perform an extensive benchmark, we tried out all possible combinations of co-elution scores and found the best result using Apex, Jaccard, Bayes, Euclidean distance, and mutual information. We performed a global optimization to select the elution profile correlation metrics that generated the highest composite score. We also noted that adding functional evidence considerably increased the composite score.

Scored protein co-fractionation networks were calculated by correlation analysis (Apex, Jaccard, Bayes, Euclidean distance, and mutual information) based on the protein intensities recorded across each set of fractions (STAR Methods). Weighted networks were constructed based on functional evidence reported in MouseNet v2 (Kim et al., 2016) omitting mammalian protein interaction data to minimize circularity that might bias our association predictions. For the machine-learning classifier, we used the Fast Random Forest implementation (STAR Methods) to integrate all generated networks. Cross-validated decision trees were learned and benchmarked using independent training and test sets of reference complexes (Ruepp et al., 2010) (STAR Methods). Clusters were defined using ClusterONE parameter settings maximizing the bipartite matching ratio between the predicted complexes and set of cluster-training complexes (STAR Methods).

Random Forest Cut-off

The final output of the random forest returns a confidence score on how likely two proteins are interacting based on their functional evidence and their co-elution. This score ranges between 0 and 1, and we would only consider two proteins to be interacting if they have a score of at least 0.5. Higher cut-off results in better composite scores and better-predicted complexes, but at a cost of reducing the number of complexes predicted. To explore the effects of this parameter, we evaluated prediction performance for each random forest score cutoff between 0.5 and 1 for two fold evaluation. We observed a steady increase in scores for cut-off scores 0.5 to 0.683, with a drop in MMR and accuracy for higher cut-offs. At the same time, we see an increase in overlap score, which in turn causes a significant increase in composite score. We see that the number of PPIs and predicted clusters declines for higher cut-off and the jump in overlap score is most likely caused by over fitting. Thus, we select a random forest confidence score cut-off of 0.683.

Classification of BraInMap Complexes

There are many metrics for measuring cluster agreement (overlap), but none is universally accepted in the field. To define novelty in a stringent and transparent manner, we applied 6 independent similarity measures reported in previous interactome publications. These include the Jaccard, Sorensen-Dice, Anderberg, Ochai (Meyer et al., 2004), and Overlap scores (Nepusz et al., 2012), and the hypergeometric distribution, to define the overlap between our predicted protein complexes and known assemblies in CORUM. Though these established similarity metrics gave generally similar results (Figure S1B), they did not account for instances wherein the subunits of a small (known) complex were found as part of a larger predicted assembly. Hence, to address this shortcoming, we then calculated an average matching index (AMI) that looked at overlaps with respect to both the vantage of the annotated and the predicted complex as follows:

$$Average\ matching\ index(AMI) = \left(\frac{\overline{(p \cap k)}}{p} + \frac{\overline{(p \cap k)}}{k} \right)$$

where p and k represent the number of subunits in predicted and known complexes respectively and ($p \cap k$) the number of subunits present in both.

As both a pragmatic and stringent solution, we classified putative complexes with an average matching index $\geq$ 0.5 as "annotated," those between $\geq$ 0.25 & < 0.5 as "previously reported assemblies with new subunits," and finally only those complexes with < 0.25 average matching index that are also not statistically significant (p-value > 0.05) by hypergeometric test as "novel" (Figure S1C).

Selection of Neurological and Other Disease Annotation

Neurological and other disease associations for BraInMap complexes were compiled from disease annotations in DisGeNET 5.0 (Piñero et al., 2017). We used high quality curated associations obtained by applying stringent filtering to exclude associations with EI (Evidence Index) < 0.9 and DisGeNET score < 0.005. (Figure S1E), to map 1710 members of BraInMap complexes to various neurological diseases (Table S8).

Enrichment Analysis

Enrichment analysis was carried out with Gene Ontology (GO) version 1.2 (downloaded on 2019-03-07), and mouse gene associations downloaded from Gene Ontology (Ashburner et al., 2000). A subset of the gene ontology comprising ~3,221 GO terms were defined using goslim synapse and selected neuronal terms for enrichment. Interaction space was constrained to only those interactions between pairs of proteins that were observed both in our high-confidence PPIs and in the target annotated dataset.

Over-representation analysis of gene ontology terms was performed using the Cytoscape app BiNGO Version 3.0.3 (Maere et al., 2005). Enrichment for each annotated term among genes in each of the 1030 complexes was calculated using the hypergeometric test (p < 0.05) with Benjamini-Hochberg FDR correction, using genes in our high confidence network as the reference set.

Gene set enrichment analysis (GSEA) (Subramanian et al., 2005) of brain regions and cell types was performed to determine brain-specific and cell type specific complexes in BraInMap. In each case, our 1030 complexes were used as gene sets. To determine brain-region specificity we used normalized protein intensity data from regional mouse brain co-fractionations performed in our lab (see STAR Methods below) restricted to proteins in our high confidence network, while for cell type specificity we used the sc-RNA-Seq gene expression data from mouse brain (Zeisel et al., 2018). Average normalized CPM values were computed using the edgeR package for R (Robinson et al., 2010) and grouped into representative neuronal and non-neuronal cell types. The gene

expression data was again constrained to genes present in our high confidence network. Results were visualized using the Cytoscape Enrichment map app and hierarchical clustering.

Enrichment of Neurodevelopmental Disorder-Related Genes in BrainMap PPI Network
The overlap between a given gene set and our network genes was evaluated using a binomial model

$$X \sim Binomial(n,p)$$

Where:*n* is the number of genes in the gene set being examined

p is the probability of observing a random protein-coding gene in our brain PPI network, which is calculated as the fraction of 2,304 genes in the network over all 20,210 mouse protein-coding genes (The UniProt Consortium, 2017).

Domain Enrichment
Domain architectures for all mouse proteins were obtained from PhyloPro 2.0 (Cromar et al., 2016) for the longest peptide associated with each gene. Domain predictions are based on Hidden Markov Models of curated seed alignments comprising Pfam A Domains and Families. To avoid frequency biases, all architectures were stripped of domain repeats using a custom Perl script (e.g. AABBAAA becomes ABA). The resulting architectures were then used to determine domain pairs as follows. Domain architectures within proteins were determined by ordering domains by sequence start site and creating adjacent pairs. These were used to define brain specific pairs as seen in the overlap analysis (see below). Domain architectures within complexes were compared to produce all possible combinations of cross-protein domain pairs, ignoring adjacent domains. For example, comparing ABC to DEF would yield AD, AE, AF, BD, BE, BF, CD, CE, CF but not AB, BC, DE or DF). We did this because, at the complex level, we were interested to discover domain associations' particular to the complex rather than the proteins themselves. Neurologically associated domains are defined as those appearing in proteins that are annotated to one or more neurological diseases. To determine unique brain and neurologically associated domains, an overlap analysis was performed as follows. A list of domains in each complex was obtained by pooling the domain architectures of proteins in each complex. This was also done for complexes in the assembly CORUM, Havugimana et al., 2012, and Wan et al., 2015 data sets. These lists were compared using (http://bioinformatics.psb.ugent.be/webtools/Venn/) to determine domain overlaps between the four data sets and identify domains unique to the brain. Brain specific and neurologically associated proteins were determined similarly. To determine statistical significance of features, we constructed 10,000 random data sets consisting of complexes of the same size as the real data set by selecting random genes (and their associated domain architectures) from a list comprising all mouse proteins with domain predictions. Custom Perl scripts were used to calculate the frequency of specific proteins, domains and domain pairs for the real data set and compare them with the sum of frequency of occurrences in the random networks, counting a score of 1 for each random network in which the protein, domain or domain pair was present as frequently or more frequently than in the real network. For the domain similarity network (Main Figure), domains and domain pairs were classified as either brain specific (b), neurologically associated (n) or non-brain specific (nb) and enrichments were determined by category. The p-value is the ratio of the real frequency to the score of the random frequencies. Network construction and visualization was done in Cytoscape.

Phylogenetic Conservation of Complexes
Ortholog predictions for all proteins were obtained from PhyloPro 2.0 (Cromar et al., 2016) and clustered using Cluster 3.0 (City Block, Complete Linkage) to group proteins with similar phylogenetic conservation patterns across 164 taxa. Taxa were phylogenetically arranged and grouped into: Eukaryotes, Opisthokonts, Metazoans, Vertebrates, and Mammals. Within each group, we scored the presence or absence of an ortholog prediction for each gene and used an unbiased, consensus approach to predict gene origin. To account for gene losses in some clades within a group we defined an arbitrary cutoff of 30% representation as a requirement to score a gene as being present within a group. To determine whether the group comprising novel complexes was enriched for proteins of a particular age category versus the group comprising non-novel complexes, the assignment of complexes to novel or non-novel groups was randomly shuffled and the frequency of proteins of different ages was compared between real groups versus 10,000 random assignments. A tally was kept in which the frequency of proteins in an age category equaled or exceeded the frequency in the real group.

DATA AND CODE AVAILABILITY

All raw proteomic (co-fractionation) data from this work is submitted to the PRIDE repository (PRIDE: PXD011304) at the European Bioinformatics Institute, in accordance with the data sharing policy. Codes used in generating the results are described above in detail.

Update

Cell Systems

Correction

BraInMap Elucidates the Macromolecular Connectivity Landscape of Mammalian Brain

Reza Pourhaghighi, Peter E.A. Ash, Sadhna Phanse, Florian Goebels, Lucas Z.M. Hu, Siwei Chen, Yingying Zhang, Shayne D. Wierbowski, Samantha Boudeau, Mohamed T. Moutaoufik, Ramy H. Malty, Edyta Malolepsza, Kalliopi Tsafou, Aparna Nathan, Graham Cromar, Hongbo Guo, Ali Al Abdullatif, Daniel J. Apicco, Lindsay A. Becker, Aaron D. Gitler, Stefan M. Pulst, Ahmed Youssef, Ryan Hekman, Pierre C. Havugimana, Carl A. White, Benjamin C. Blum, Antonia Ratti, Camron D. Bryant, John Parkinson, Kasper Lage, Mohan Babu, Haiyuan Yu, Gary D. Bader, Benjamin Wolozin,* and Andrew Emili*

*Correspondence: bwolozin@bu.edu (B.W.), aemili@bu.edu (A.E.)
https://doi.org/10.1016/j.cels.2020.08.006

(Cell Systems 10, 333–350.e1–e14; April 22, 2020)

In the original published version of this paper, the Atxn2[+/−] mouse was described incorrectly in the Key Resources Table. The mouse was not sourced from the Gitler laboratory. The correct source is the Pulst laboratory, University of Utah, as originally described in Kiehl et al. (2006). The genetic background of the Atxn2[+/−] mouse was also reported incorrectly as B6. The correct background is C57B6/Fvb129 hybrid. This information has been corrected and the authors apologize for any confusion these errors may have caused.

REFERENCES

Kiehl, T.R., Nechiporuk, A., Figueroa, K.P., Keating, M.T., Huynh, D.P., and Pulst, S.M. (2006). Generation and characterization of Sca2 (ataxin-2) knockout mice. Biochem. Biophys. Res. Commun. 339, 17–24.