Proteomic analysis of postsynaptic proteins in regions of the human neocortex

Marcia Roy[1*], Oksana Sorokina[2*], Nathan Skene[1], Clemence Simonnet[1], Francesca Mazzo[3], Ruud Zwart[3], Emanuele Sher[3], Colin Smith[1], J Douglas Armstrong[2] and Seth GN Grant[1].

* equal contribution

Author Affiliations:

1. Genes to Cognition Program, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh EH16 4SB, United Kingdom
2. School of Informatics, University of Edinburgh, Edinburgh, EH8 9AB, United Kingdom
3. Lilly Research Centre, Eli Lilly & Company, Erl Wood Manor, Windlesham, GU20 6PH, United Kingdom

Abstract:

The postsynaptic proteome of excitatory synapses comprises ~1,000 highly conserved proteins that control the behavioral repertoire and mutations disrupting their function cause >130 brain diseases. Here, we document the composition of postsynaptic proteomes in human neocortical regions and integrate it with genetic, functional and structural magnetic resonance imaging, positron emission tomography imaging, and behavioral data. Neocortical regions show signatures of expression of individual proteins, protein complexes, biochemical and metabolic pathways. The compositional signatures in brain regions involved with language, emotion and memory functions were characterized. Integrating large-scale GWAS with regional proteome data identifies the same cortical region for smoking behavior as found with fMRI data. The neocortical postsynaptic proteome data resource can be used to link genetics to brain imaging and behavior, and to study the role of postsynaptic proteins in localization of brain functions.

Introduction:

For almost two centuries, scientists have pursued the study of localization of function in the human cerebral neocortex using diverse methods including neuroanatomy, electrophysiology, imaging and gene expression studies. The frontal, parietal, temporal and occipital lobes of the neocortex have been commonly subdivided into Brodmann areas (BA)[1] based on cytoarchitectural features, and specific behavioral functions have been ascribed to these regions. One of the hallmarks of the neocortex is its high density of excitatory synaptic connections[2] and these synapses are endowed with over 1,000 postsynaptic proteins that control the behavioral repertoire and are disrupted by mutations causing over 130 brain diseases[3,4]. However, the compositional and functional differences in synapse proteomes in the neocortical regions remains unknown. The recent development of reliable methods for the direct quantification of human neocortical synapse proteomes from post-mortem tissue makes it feasible to analyze the molecular composition of synapses directly in human brain regions[5].

In the last two decades, proteomic mass spectrometry has been used to systematically characterize the protein composition of synapses and has expanded our view of the postsynaptic terminal from being a simple connector to a sophisticated signaling machine[6,7] assembled from multiprotein complexes and super-complexes[8-12]. The characterization of the postsynaptic proteome in humans[3-5], mice[3,13], rats[14-16] and zebrafish[17] reveals a highly conserved "vertebrate postsynaptic density proteome" (vPSD)[17] comprised of ~1,000 proteins. In addition to these conserved proteins, there are lineage-specific postsynaptic proteins that contribute to functional variation[17]. It is also evident that there must be compositional differences in the synapse proteome across different brain regions because immunohistochemical staining of individual synaptic proteins shows non-uniform distributions[18]. The importance of studying brain regional variation in synapse proteomes is that the large and comprehensive datasets can be used with genetic and anatomy-based resources, such as the Human Connectome Project[19], to better understand the localization of neocortical functions including behavior.

Here, we generate the first map of the human postsynaptic proteome across frontal, parietal, temporal and occipital lobes of the neocortex focusing on 12 BAs (Table 1, Figure 1A). These BAs are involved in many aspects of behavior including executive functions, language, motor functions, vision and memory. Localized pathology in these regions is associated with many disorders including schizophrenia, dementias, stroke, trauma and substance abuse. We show that diversity in the composition of PSD proteins generates a characteristic molecular signature for each BA. Integration of these data with genetic, Positron Emission Tomography (PET), functional and structural Magnetic Resonance Imaging (MRI) data show that synapse proteome data can be used to address a variety of questions about the localization of functions in the neocortex. The large-scale dataset we describe and make freely available is a valuable resource for a wide range of future studies on the human brain, behavior and brain imaging.

RESULTS

Isolation of the postsynaptic proteome from neocortex

Post-mortem human brain tissue obtained from the MRC Edinburgh Brain Bank was screened for synaptic proteome preservation using HUSPIR (HUman Synapse Proteome Integrity Ratio)[5] and four non-diseased cases suitable for synapse proteomics were identified (Supplementary Table 1). We focused on 12 BAs from frontal, parietal, temporal and occipital lobes of the neocortex (Figure 1A, Table 1). A wide range of behavioral functions has been ascribed to these areas (Table 1). The postsynaptic density (PSD) fraction was isolated from each tissue sample as previously described[5] and LC-MS/MS analysis performed on a total of 48 samples. Label-free quantitation of peptide intensity for all proteins identified 1,213 proteins (Supplementary Table 2). Proteins identified with a minimum of 2 unique peptides and having a mean peptide intensity of 1.5-fold or greater in one brain region compared to any other and determined to be significant with a p-value < 0.05 were scored as differentially expressed (Supplementary Table 2). One hundred forty-nine proteins were differentially expressed in at least one BA compared to all others (Supplementary Table 3, Supplementary Figure 1).

We first compared the complexity of this proteome dataset with two previously published human postsynaptic proteome neocortex datasets[4,5] and found an overlap of 75% between all three (Supplementary Figure 2A). Twenty five percent of proteins in this study had not been previously detected (Supplementary Table 4), which may reflect the many new BAs investigated. Consistent with this, we compared our human dataset with whole forebrain datasets obtained in mice and rats and found a 96.3% overlap (Supplementary Figure 2B). Thus, the sampling of 12 BAs across the human neocortex likely captures the diversity of postsynaptic proteins found in the mammalian brain.

Validation of pooling of individual samples

Prior to pooling datasets from the individuals, we examined reproducibility of expression between individuals using the Differential Stability (DS) approach that was previously applied in the transcriptome analysis of adult human brain regions[20]. DS analysis was performed on the MS intensity values obtained for the 1,213 proteins to identify the subset with highly reproducible expression patterns across the four independent human brains. As in Hawrylycz *et al*.[20], we quantified the average pairwise Pearson correlation ρ over a set of 12 anatomical regions and obtained the DS values ranging from 0.83 to -0.25 with 37 proteins (3% of total) having a ρ-value ≥ 0.5, highly correlated between all 4 individuals (Supplementary Table 5). The proteins with highest DS were CPNE6 (ρ-value = 0.82), LGI3 (ρ-value = 0.75), NEFH (ρ-value = 0.72) and CAMK2D (ρ-value = 0.71). The 37 PSD proteins displaying a high DS correlation in their expression patterns were enriched in membrane targeting proteins (q-value = 3.49×10^{-03}) and calcium-dependent membrane targeting proteins (q-value = 5.83×10^{-03}) (Supplementary Table 6). We next asked if these high-DS proteins corresponded to those described in the transcriptome analysis of human brain regions[20] and found a very high overlap (35/37; 95%, p-value = 2.04×10^{-16}). Although the total percentage of high DS samples is higher in the transcriptome analysis[20], the estimated correlation coefficient between the DS results for all postsynaptic proteins and their respective transcripts was small ($R^2 = 0.18$) but highly significant (p-value = 1.218×10^{-09}). As a second approach, we examined differences

between the brain regions of all 4 individuals by performing hierarchical clustering on protein abundances (Supplementary Figure 3). All 12 BAs clustered into two main branches; one large branch containing seven or eight similar regions and a second branch of three or four similar regions. In addition, Tukey's HSD test showed no significant difference between the mean values of four individuals at a confidence level of 95% (Supplementary Figure 4A, B). We also examined the effect of removing brain B (female) or brain D (longest post-mortem interval) on biochemical modules. Although this perturbs some modules (specifically modules 5 and 6), we noticed that combining these brains mutually compensated for this disturbance (see section: Neocortical architecture of the postsynaptic proteome; Supplementary Figure 5). We therefore decided to proceed with combined data from all four individuals. Individual proteins and their abundances across all 12 BAs are listed in Supplementary Table 7 and 8 (ranked regional abundance).

A core anatomically invariant component of the postsynaptic proteome

Prior to analyzing the proteins that may contribute to different functions in neocortical regions, we sought to identify the invariant proteins that may provide core functions in all 12 BAs. Using the method of Le Bihan *et al*[21] we performed a Two One-Sided T-Test (TOST-test) using mean MS intensity values (n = 4) for the 1,213 proteins and a Ɛ sensitivity factor (see Materials and Methods) of 0.045: this identified 27 proteins (Supplementary Table 9). The proteins with the lowest variability between regions were ANK3 (p-value = 0.008), RHOT1 (p-value = 0.012), MYO18A (p-value = 0.015) and DYNCLI1 (p-value = 0.018). The 27 proteins were enriched in components of the cytoskeleton (q-value = 8.91×10^{-05}), protein complex (q-value = 9.55×10^{-05}) and microtubules (q-value = 8.73×10^{-03}) (Supplementary Table 10). This indicates that this set of postsynaptic proteins with the lowest variability between brain regions is enriched in key structural elements of the cytoskeleton and are representative of a universal core of proteins that maintain the structural integrity of the postsynaptic proteome throughout the neocortex.

Neocortical architecture of the postsynaptic proteome

We proceeded to study the differences between brain regions using a series of analyses beginning with candidate proteins of known functional importance, then methods that exploit the large datasets to identify a modular molecular architecture of the postsynaptic proteome that underpins regional organization.

We first examined the expression of 41 well-known candidate proteins representing 11 protein classes including glutamate and GABA receptors, scaffold proteins and other ion channels, which are known to play a key role in synaptic physiology and cognitive functions (Supplementary Table 11). AMPA receptors are protein complexes that mediate fast synaptic transmission and we found Gria1/AMPA1 is differentially expressed throughout the neocortex, being most abundant in BA20 and least abundant in BA4 having z-scores of 0.24 and 0.34 above the mean, respectively. Conversely, other AMPA receptor subunits, Gria3/AMPA3 and Gria4/AMPA4, were detectable at very low levels and uniformly in all regions examined with average z-scores of -0.6 and -0.7 below the mean. The NMDA receptor subunits, which also assemble protein complexes[9,12] and mediate synaptic plasticity, were differentially expressed: Grin1/NR1 and Grin2B/NR2B, being least abundant in BA4 and BA19 and highly abundant in BA20. NMDA receptor complexes assemble with MAGUK (Membrane Associated Guanylate Kinase) scaffold protein complexes to form supercomplexes[9,12] and paralogs of this family (Dlg1/SAP97, Dlg2/PSD93, Dlg3/SAP102, Dlg4/PSD95) were also differentially abundant across all regions (Supplementary Table 11). These findings indicate that different regions of the neocortex express differing amounts of synaptic proteins and their complexes.

We next used the full complement of proteome data and hierarchical clustering to compare BAs. This approach allows us to group BAs into those that are similar and different in terms of PSD protein expression, as well as identify subsets (or modules) of postsynaptic proteins that are co-regulated in the BAs. Each BA had a distinct molecular signature and the similarities of these signatures was hierarchically organized into groups, including four main branches which we refer to as “Brodmann Area Groups” (BAG 1-4) (Figure 1B, C). Several organizational principles emerge: First, there is diversity of

postsynaptic proteomes in all four lobes of the cerebrum (frontal, temporal, parietal and occipital). Second, two regions from different lobes can be more similar to each other than two adjacent regions in the same lobe. For example, BA19 from the occipital lobe is similar to BA4 and BA6 in the frontal lobe (all are in BAG1), and distinct to its adjacent regions BA37 and BA39 (in BAG4). Third, adjacent regions can either be similar or different; of 10 pairs of adjacent BAs, 4 were in the same BAG. The similar pairs of adjacent regions were temporal lobe areas BA20 (inferior temporal area) and BA38 (anterior temporal lobe), both in BAG2; dorsolateral prefrontal cortex area BA9 and pars opercularis of the inferior frontal gyrus BA44; primary motor cortex BA4 and premotor cortex BA6; the angular area of the parietal cortex BA39 and the occipitotemporal area BA37 of the occipital lobe. These findings indicate that postsynaptic proteome compositional differences contribute to the diversity of BAs and that this reveals a novel hierarchical relationship between these areas.

To understand how this anatomical organization arises from the differential expression of postsynaptic proteins we identified clusters of co-regulated groups of postsynaptic proteins, which we call Postsynaptic Proteome Modules (PPM) (Figure 1B) (Supplementary Table 12). The seven PPMs were selected based on the best combination of indices provided by nbClust R package (Methods). To validate the robustness of these PPMs, we performed bootstrapping with hClust, k-means and the pam method (see Methods and Supplementary Figure 4A). To visualize the distribution of the different PPMs across the 12 BAs, Circos visualization was performed using those proteins whose abundances are above the mean and thus enriched (having a positive z-score in their protein abundance trends for a given region compared to all other regions analyzed) (Supplementary Figure 6). We found that while certain PPMs exhibited only subtle differences in distribution, others, such as PPM1 and PPM2, showed very different abundance patterns between brain regions. PPM1 contains NMDA receptor subunits Grin1/NR1 and Grin2b/NR2B and PPM2 contains MAGUK proteins Dlg4/PSD95, Dlg3/SAP102 and Dlg2/PSD93, which assemble multiprotein signaling complexes[9,12] controlling synaptic plasticity. We also found that key components of inhibitory synapses were also part of differentially distributed protein modules: PPM1 contained Gabrb1,

Gabrb3 and Gabbr2; PPM2 contained Gabra1, Gabrb2 and Gabarapl2; PPM6 contained Gabbr1, Gabra2 and Gabrg2. Taken together, these results demonstrate that differential distribution of PPMs and their key functional proteins contribute to the hierarchical organization of BAs.

Electrophysiological validation of proteome differences in Brodmann areas

To test the functional significance of the differential expression of synaptic molecules in BAs, we examined the electrophysiological properties of *Xenopus* oocytes injected with synaptosomes isolated from frozen human brain tissue[22]. We specifically examined the β1 $GABA_A$ receptor subunit between regions BA20/BA39 and BA4/BA19 (Figure 2). BA20/BA39 were chosen because they were the two regions showing the largest differential expression, in contrast to BA4/BA19, which shows small differential expression (see Gabrb1 differential distribution in Supplementary Tables 9, 10). Similar amounts of synaptosomes from BA20 and BA39 were injected in oocytes and two days later ion currents were measured using the two-electrode voltage clamp technique. Application of 1 mM of the agonist GABA resulted in much larger responses ($p<0.001$, unpaired t-test) in oocytes injected with synaptosomes from BA20 (82 ± 2 nA; n = 6) than in oocytes injected with synaptosomes from BA39 (35 ± 5 nA; n = 6) (Figure 2A). Similar results were obtained with synaptosomes from the same two brain regions obtained from two other brain donors. In these two further cases, 1 mM GABA-induced responses were 25 ± 2 nA (n = 7) for BA20 and 3 ± 1 nA (n = 7) for BA39 for one donor, and 9 ± 3 nA (n = 7) for BA20 and 1.4 ± 0.4 nA (n = 8) for BA39 for the other donor, respectively (Figure 2A). In contrast to BA20/BA39 differences, no significant differences in 1 mM GABA-evoked responses were found between BA4 and BA19, the two brain regions with similar β1 $GABA_A$ subunit protein levels (Figure 2B). Different expression levels in ion currents between donors is expected given the variation in quality of the original tissues[22]. These electrophysiological experiments indicate the differential synaptosome proteome composition reflects functional differences in ion channel activity in Brodmann areas.

Neocortical localization of biochemical pathways

We next turned our attention from the ion channels and receptor mechanisms to the wider range of postsynaptic biochemical functions found in the highly complex proteome and asked if these functions showed differential neocortical localization. We used two related strategies: KEGG pathway enrichment analysis of PPMs or in the ranked abundance of proteins in each BA. The results for the first strategy are shown in Figure 3A, revealing differential biochemical pathway enrichment in PPMs. For example, PPM1 was enriched in synaptic vesicle functions, synaptic plasticity (long-term potentiation and depression pathways), whereas PPM7 was particularly enriched in oxidative phosphorylation, Alzheimer's, Parkinson's and Huntington's disease pathways. In the second strategy, KEGG pathway enrichment was performed on the ranked protein abundances for each of the 12 BAs and hierarchical clustering performed (Figure 3B). Complementary to the organization of Brodmann Area Groups (BAGs, Figure 1B), which were based on abundance of proteins, we find that each BA has a unique biochemical "pathway signature" (Figure 3B). Interestingly, the dendrogram shows three major divisions: 1) many signal transduction mechanisms, 2) synaptic plasticity and other signaling processes, 3) neurodegenerative diseases and metabolic mechanisms.

PET imaging of postsynaptic metabolic mechanisms

Human brain imaging is the major tool for examining the localization of functions in the human neocortex. We sought to investigate the possibility that regional differences in the postsynaptic proteome was relevant to brain imaging. Positron Emission Tomography (PET) has been used to document differences between Brodmann areas for several metabolic parameters including cerebral metabolic rate for oxygen ($CMRO_2$), cerebral metabolic rate for glucose (CMRGlu), cerebral blood flow, and two derived indices (oxygen-glucose index, OGI; glycolytic index, GI), in normal resting human subjects[23]. The postsynaptic density is known to contain glycolytic and other metabolic enzymes that generate ATP used by signaling enzymes involved with synaptic plasticity[3,4,13,14,16,24]. Using published PET data[23] reporting metabolic parameters from brain regions (spatial resolution 3-5 mm), we correlated these metabolic data with protein abundances in the

PPMs and found significant correlations that were robust to perturbation analysis ($n_{regions}$ = 12, $P < 10^{-4}$)(Figure 4A)(Methods). For example, the mean regional protein abundance of PPM1 (which contains high levels of the mitochondrial Complex I & IV proteins and enzymes involved in the OXPHOS pathway) displayed a positive correlation with the OGI value and negative correlation with GI, CMRGlu and $CMRO_2$ values; the mean regional protein abundances for PPMs 2 and 3, which contain glycolytic enzymes PGAM and PKM2 proteins, were positively correlated with values of these metabolic parameters. These findings indicate that regional differences in the composition of the postsynaptic proteome correlate with functional signals measured in PET metabolic imaging.

Task-specific fMRI signals correlate with regional postsynaptic proteomes

Magnetic resonance imaging (MRI) is another imaging modality that is widely used to study structural and functional organization of the neocortex. Behavioral responses can be detected in regions of the neocortex using functional MRI (fMRI) and postsynaptic proteins are important for many human behaviors[4]. We therefore hypothesized that compositional differences between the postsynaptic proteomes in different BAs would be linked to specific behavioral responses detected with fMRI. To test this hypothesis, we obtained data from the Human Connectome Project (HCP), which used high-quality MRI data to define the structural and functional parcellation of neocortical regions[19]. The HCP uses myelin content and cortical thickness from MRI images to define the structural architecture of neocortex and has mapped correlates with the functional activation of brain regions in response to a battery of behavioral tests, called task-fMRI (tfMRI). These tasks activate well-characterized neural systems including: 1) visual and somatosensory-motor systems; 2) category-specific representations; 3) language function (semantic and phonological processing); 4) attention systems; 5) working memory/cognitive control systems; 6) emotion processing; 7) decision-making/reward processing; and 8) episodic memory systems[25] (see Methods and Supplementary Table 13 for the nomenclature of the specific tasks).

First, we identified the parcellated regions from the HCP that corresponded to the 12 BAs (Supplementary Figure 7; Methods). The fMRI structural marker *myelin* marker showed a high correlation with the expression values of four myelin proteins (Mog, Plp1, Omg and Pmp2) in our proteomic data (R^2 = 0.7-0.8, p-value < 0.0001) (Supplementary Figure 7A, Supplementary Table 14). Next, we computed the correlation of the behavioral tfMRI scores (19 scores) with proteome expression abundances (1,213 proteins) over 12 BAs (Supplementary Table 15) (Figure 4B). Perturbation analysis identified 1,154 significant (P <0.05) scores (Supplementary Table 16). We then performed hierarchical clustering of all HCP structural (myelin and cortical thickness) and behavioral scores from the fMRI datasets with postsynaptic protein abundance and found a clear division into 4 quadrants (Figure 4B).

Strikingly, the structural and functional fMRI traits were separated into two major groups: language and motor functions were clustered with the *myelin* structural marker, whereas emotion and working memory were clustered with the *cortical thickness* marker. To further explore the biochemical features of the postsynaptic proteomes associated with these behavioral traits, we correlated tfMRI scores with PPM protein abundances (similar to the PET analysis above) and found significant correlations that were robust to perturbation analysis ($n_{regions}$ = 12, $p < 0.05$) (Supplementary Figure 8). PPM5, which contains the four myelin proteins in our study (Mog, Plp1, Omg and Pmp2), correlated with the *myelin* structural marker. Other PPMs correlated with behavioral tasks, such as language/mathematical functions that correlated with PPM4 and PPM6. These data indicate that fMRI responses to behavioral tasks are associated with biochemical properties of the postsynaptic proteome in the respective brain regions.

Postsynaptic composition localizes genetic mechanisms of behavior

Because each BA expresses a specific signature of postsynaptic proteins and human genetic studies show many diseases and behavioral traits are associated with genetic variants in postsynaptic proteins, it follows that combining our proteomic datasets with genetic data may reveal brain regions where those variants exert their influence on

phenotypes. If so, this could provide a bridge between genetics and functional brain imaging. To assess this possibility, we used Genome Wide Association Study (GWAS) data on disease and cognitive traits to ask if variation in the composition of the synaptic proteomes in different BAs revealed the neocortical localization of these genetically-influenced traits.

We utilized summary statistics from some of the largest available GWAS studies to test for association with mental disorders (schizophrenia, autism, Alzheimer's disease, migraine, eating disorders and major depression) and cognitive traits (aggression, years in education and smoking). We also included a set of other traits/disorders including height, Type 2 Diabetes, cholesterol levels, triglyceride levels, Ulcerative Colitis, Crohn's Disease and Inflammatory Bowel Disease; while these are generally thought of as non-neuronal traits and are intended to be considered as controls, some are genetically correlated with core neuronal phenotypes such as the number of years spent in education[26]. For each BA, we partitioned the genes into deciles based on their specificity for that region (Figure 5A,D), wherein specificity to brain region X is defined as the proportion of the protein's total mean expression which is found within brain region X. We assumed that SNPs would primarily have an impact upon their nearest genes. Partitioned Linkage Disequilibrium Score Regression (LDSC) to evaluate the enrichment of the common-variant single-nucleotide polymorphism (SNP) heritability for each GWAS in each decile. We then tested the hypothesis that disease association increases as brain region specificity increases by fitting a linear model to the enrichment z-scores for each of the deciles.

Prior to multiple correction testing, four GWAS studies showed significant enrichments in heritability as regional specificity increased: smoking (current vs former) (BA9, P = 0.00012)(Figure 5B), schizophrenia (BA11, P = 0.01290), smoking (never vs ever) (BA44, P = 0.01485) and years in education (BA37, P = 0.02078). Note that this does not depend on significant enrichments within any particular decile, just that the more specific deciles are significantly more associated with the trait than the least specific deciles. After correcting for multiple testing, one region remained significant: BA9, the medial

dorsolateral prefrontal cortex was associated with smoking cessation (P = 0.00012, Figure 5B). No other traits (e.g. height, Figure 5C) were significantly associated with BA9 (nor any other region), nor do other regions show enrichment for smoking (nor other) traits (e.g. BA44, Figure 5E,F). This result indicates that the postsynaptic proteins that are differentially expressed in BA9 of healthy individuals are encoded by genes that are associated with smoking cessation. Our findings are in agreement with MRI and PET imaging studies that have demonstrated the involvement of BA9 in smoking and substance abuse[27-30].

Discussion

We have analyzed the postsynaptic proteome in 12 Brodmann areas of the human neocortex and found each one has a signature of compositional differences that underlie functional differences. Key postsynaptic proteins, including ion channels and neurotransmitter receptors, and biochemical pathways controlling physiology and disease, were differentially distributed in neocortical regions. MRI and PET imaging are widely used for structural and functional assessment of brain regions and we have shown that regional diversity in postsynaptic proteomes is potentially relevant to the interpretation of these imaging modalities. We also demonstrate that synapse proteome brain maps can be used to link genetics, brain imaging and behavior.

From the compositional differences in the postsynaptic proteome, we see evidence of a molecular logic to the synaptic organization of the neocortex. The postsynaptic proteome is highly complex and comprises a diverse set of protein classes ranging from neurotransmitter receptors, adhesion and scaffold proteins to signaling, metabolic, structural and trafficking proteins[3-6,8,13,15,16,31,32]. Combinations of these proteins are physically assembled into supramolecular multiprotein complexes, which are building blocks of synapses[9,12,33]. We found differential expression of the protein components of complexes, including AMPA, NMDA and GABA receptors, indicating that postsynaptic multiprotein complexes are differentially distributed in neocortical regions. We also found that key components of postsynaptic signaling complexes were in different proteome

modules (e.g. module 1 contains Grin1/NR1, Grin2b/NR2B, and module 2 contains Dlg4/PSD95, Dlg3/SAP102 and Dlg2/PSD93 proteins), which are differentially distributed. Using electrophysiological recordings of GABA receptors, we confirmed the differential expression of functional ion channel complexes in synapses from different brain regions. Thus, the regional diversity in postsynaptic proteome composition must reflect the differential allocation of protein complexes into subtypes of synapses and populations of these subtypes define each brain region.

The regional diversity of postsynaptic proteome composition revealed organizational principles. Compositionally distinct areas were found within all four lobes of the cerebrum, with most adjacent regions showing distinct composition. Some regions in different lobes showed similar composition, suggesting anatomically distant regions are composed of synapses with similar signaling properties. The molecular differences between regions would likely influence the vulnerability of those regions to disease-causing mutations that disrupt postsynaptic proteins[4]. This is supported by the observed enrichments in some regions of disease pathways. Interestingly, we found enrichments of pathways involved with neurodegeneration and metabolism in similar regions, suggesting that the vulnerability of synapses to neurodegenerative disorders may by influenced by their metabolic properties, consistent with previous results[23,34,35].

While it has been evident from PET and MRI imaging that there is localization of behavioral traits in the neocortex, and that these traits have a genetic underpinning, the mechanistic links between genetics, traits and brain regions remain poorly understood. Our finding that GWAS data and synapse proteomics can be combined to identify brain regions associated with smoking is an important proof-of-principle result. The composition of synapse proteomes has not (to our knowledge) been previously considered as a factor in the regional signals of fMRI data. The correlations between the regional localization of postsynaptic compositional signatures and language, mathematical ability and emotional traits support the view that the genetic mechanisms for these traits are associated with the synaptic properties of the cognate brain regions. Given the rich literature describing human brain imaging in behavioral tasks and the advances in genetics of behavior, we

expect that more comprehensive datasets will lead to the identification of further links between brain regions, traits and genetics.

Our findings also provide new insight into the biological basis of smoking behavior. Our analysis indicates that the ease with which individuals can cease smoking (or their susceptibility to try to give up) may be related to the function of postsynaptic proteins in BA9 in the prefrontal cortex. We also note an interesting potential convergence between smoking, schizophrenia, the dorsolateral prefrontal cortex and the postsynaptic proteome: this brain region is involved with schizophrenia[36], eighty percent of schizophrenia patients smoke[37] and the genetic susceptibility to schizophrenia is enriched in postsynaptic proteins[38-40].

The technical obstacles for synapse proteomics from post-mortem human brain material have been largely overcome and it is now feasible to conduct much more detailed examination of human brain regions in normal and diseased cases. This comparative dataset provides a robust foundation for future studies of the human neocortex in health and disease and is freely available in the Edinburgh DataShare database (http://hdl.handle.net/10283/2362).

Methods:

PSD protein preparations for mass spectrometry:

Post-mortem human brain tissue from obtained from the MRC Brain bank – Edinburgh and processed as previously described[5]. Briefly, dissected tissue was homogenized by performing 12 strokes with a Dounce homogenizer containing 2 mL of ice cold homogenization buffer (320 mM sucrose, 1mM HEPES, pH = 7.4) containing 1X Complete EDTA-free protease inhibitor (Roche) and 1X Phosphatase inhibitor cocktail set II (Calbiochem). Insoluble material was pelleted by centrifugation at 1,000 x g for 10 minutes at 4˚C. The supernatant (S1) was removed and the pellet was resuspended in 1 mL of homogenization buffer and an additional 6 strokes were performed. Following a second centrifugation at 1,000 x g for 10 minutes at 4˚C, the supernatant (S2) was removed and pooled with S1. The combined supernatants were then applied to a sucrose gradient to isolate pure PSD fractions suitable for quantitative LC-MS/MS analysis.

Sample preparation and LC-MS/MS analysis:

All chemicals were purchased from Sigma-Aldrich, UK unless otherwise stated. Acetonitrile and water for HPLC-MS/MS and sample preparation were HPLC quality and were purchased from Thermo-Fisher, UK. Formic acid was supra-pure (90-100%) purchased from Merck (Darmstadt, Germany) while trypsin sequencing grade purchased from Promega (UK). All HPLC-MS connector fittings were purchased either from Upchurch Scientific or Valco (Hichrom and RESTEK, UK). Proteins were acetone precipitated, the pellet reconstituted in 8M urea and diluted to 2 M, samples were trypsinised in a similar manner as described[48]. Resulting peptide extracts were then dried under low pressure and acidified with 7 µl of 0.05% TFA and were filtered with Millex filter (Millipore, UK) before subjecting to HPLC-MS analysis. Nano-HPLC-MS/MS analysis was performed using an on-line system consisting of a nano-pump (Dionex Ultimate 3000, Thermo-Fisher, UK) coupled to a QExactive instrument (Thermo-Fisher, UK) with a pre-column of 300 µm x 5 mm (Acclaim Pepmap, 5 µm particle size) connected to a column

of 75 µm x 50 cm (Acclaim Pepmap, 3 µm particle size). Samples were analyzed on a 90 min gradient in data dependent analysis (1 survey scan at 70k resolution followed by the top 10 MS/MS).

Data analysis:

Data from MS/MS spectra were searched using MASCOT Version 2.4 (Matrix Science Ltd, UK) against the *Homo sapiens* (34,284 protein sequences) subset of the National Center for Biotechnology Information (NCBI) with maximum missed-cut value set to 2. Following features were used in all searches: i) variable methionine oxidation, ii) fixed cysteine carbamidomethylation, iii) precursor mass tolerance of 10 ppm, iv) MS/MS tolerance of 0.05 amu, v) significance threshold (p) below 0.05 (MudPIT scoring) and vi) final peptide score of 20.

Progenesis Version 4 (Nonlinear Dynamics, UK) was used for HPLC-MS label-free quantitation. Only MS/MS peaks with a charge of 2+, 3+ or 4+ were taken into account for the total number of 'Feature' (signal at one particular retention time and m/z) and only the five most intense spectra per 'Feature' were included. Each LC-MS run is normalized by multiplying a scalar factor. The scalar factor is a ratio in log space of the median intensity of the selected features against the median intensity of the selected feature of a reference spectra. The associated unique peptide ion intensities for a specific protein were then summed to generate an abundance value, from which was then transformed using an ArcSinH function. Based on the abundance values, within group means were calculated and from there the fold changes (in comparison to control) were evaluated. One-way analysis of variance (ANOVA) was used to calculate the p-value based on the transformed abundance values. P-values were adjusted for multiple comparisons and were calculated either from Progenesis Version 4 (Nonlinear Dynamics, UK) or using R[49] based on Benjamini and Hochberg[50]. Further analysis was performed by extracting z-score calculated on ArcsinH average group.

Differentially expressed proteins were only considered significant in current study if the following conditions were fulfilled: i) adjusted p-values (pair-wise) less than 0.05, ii) number of unique peptides detected and used in quantification per protein was at least 1 (for the 1,902 dataset) or 2 (for the 1,215 dataset), and iii) absolute fold change was at least 1.5 fold for up-regulated proteins and ≤ 0.667 fold for down-regulated proteins).

Preparation of Functional Synaptosomes from post-mortem Human Brain Tissue:

Synaptosomes were purified with a sucrose gradient as described[22]. In brief, human frozen tissue was homogenized ten times with a Teflon homogenizer in 9 ml/g of ice-cold 0.3 M sucrose containing 50 mM Tris-HCl (pH 7.4), 50mM EGTA, 50 mM EDTA and both protease (Roche) and phosphatase (Calbiochem) inhibitor cocktails. The homogenate was centrifuged at 1,500 g for 20 minutes at 4°C. The supernatant fraction (S1) was centrifuged at 16,000 g for 30 minutes at 4°C. The resulting pellet (P2) was re-suspended in 5 mM Tris-HCl (pH 8.0), homogenized three times with a Teflon homogenizer and left on ice for 45 minutes. The P2 fraction was homogenized again for ten times, re-suspended in sucrose to make a 34% (w/w) solution and layered onto a discontinuous sucrose density gradient consisting of, from bottom up, equal volume of sample, and buffer containing 0.85 M and 0.3 M sucrose, respectively. After centrifugation at 60,000 g for 2 hours with a Beckmann SW41Ti swing bucket rotor, the synaptosomal fraction (layer between 0.8 and 1.2 M sucrose) was collected and diluted with two volumes of ice-cold 50 mM TrisHCl (pH 7.4) and centrifuged at 48,000 g for 30 minutes. The resulting synaptosomal pellet was re-suspended in 50mM Tris.

Assay for Measurement of GABA Currents from post-mortem Human Brain Tissue:

Xenopus oocytes (stages V–VI) were removed from sacrificed frogs and de-folliculated after treatment with collagenase type I (5 mg/ml calcium-free Barth's solution) for 4 hours at room temperature. 60 nL of synaptosome (1mg/ml) suspension was injected per oocyte using a Drummond (Broomall, PA) variable volume micro-injector. After injection oocytes were incubated at 18°C in a modified Barth's solution containing 88 mM NaCl, 1 mM KCl,

2.4 mM $NaHCO_3$, 0.3 mM $Ca(NO_3)_2$, 0.41 mM $CaCl_2$, 0.82 mM $MgSO_4$, 15 mM HEPES, and 50 mg/l neomycin (pH 7.6 with NaOH; osmolarity 235 mOsm). Experiments were performed on oocytes after 2 days of incubation.

Oocytes were placed in a recording chamber (internal diameter 3 mm), which was continuously perfused with a saline solution (115 mM NaCl, 2.5 mM KCl, 1.8 mM $CaCl_2$, 1 mM $MgCl_2$, 30 µM glycine, 10 mM HEPES (pH 7.4) at a rate of approximately 10 ml/min. Dilutions of drugs in external saline were prepared immediately before the experiments and applied by switching between control and drug-containing saline using a BPS-8 solution exchange system (ALA Scientific Inc., Westbury, NY). Between responses, the oocytes were washed for 2 minutes. Oocytes were impaled by two microelectrodes filled with 3 M KCl (0.5–2.5 MΩ) and voltage-clamped using a Geneclamp 500B amplifier (Axon Instruments, Union City, CA). The external saline was clamped at ground potential by means of a virtual ground circuit using an Ag/AgCl reference electrode and a Pt/Ir current-passing electrode. The membrane potential was held at −100 mV. The current needed to keep the oocyte's membrane at the holding potential was measured. Membrane currents were low-pass filtered (four-pole low-pass Bessel filter, −3 dB at 10 Hz), digitized (50 Hz), and stored on disc for offline computer analysis. Data are expressed as nano-Ampere (nA) of current, (mean ± S.E.M). All experiments were performed at room temperature. Tests of significance were performed using the Student's t-test, and p-values less than 0.05 were considered significant.

Bioinformatic analysis and Visualization:

The majority of analysis was performed in R. Tukey tests (https://www.jstor.org/stable/3001913?seq=1#page_scan_tab_contents) were performed with *FactoMineR* (http://factominer.free.fr), which provides several useful functions for Exploratory Multivariate Analysis; correlations were visualised with *corrplot* (https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html); heat maps were generated with use of heatmap.2 function from the *gplot* R library (https://cran.r-project.org/web/packages/gplots/index.html).

Differential stability (DS) analysis was performed as described[20], namely for each protein in the list, an average Pearson correlation coefficient was calculated from six pairwise Pearson coefficients for four brain samples.

Hierarchical clustering and comparison of dendrograms was performed with *dendextend*, which allows building and comparison of multiple dendrograms Galili *et al*., 2015)[44]. The number of stable clusters was independently assessed with *nbClust* package[51], which provides several methods and indices to determine the optimal number of clusters. We selected a set of seven clusters supported by three indices, which was the second best split. Although the best split (five indices) proposed a two-cluster split, that split was not obviously biologically meaningful or useful.

Cluster robustness was estimated by a bootstrap approach using *ClusterCons*[52]. Specifically, clustering was performed 1,000 times removing 20% proteins each time and performing hierarchical clustering using *hClust* (ward.D), k-means, and pam methods. The membership of the obtained clusters was combined and for each of the individual brains as well as combinations AC, ACD and ABC and compared to assess the specific impact of potential confounders (brains B and D).

We used *ClusterProfiler* for Gene Ontology (GO) and KEGG enrichment analysis (http://www.ncbi.nlm.nih.gov/pubmed/22455463) and *ReactomePA* for pathway over-representation analysis (http://bioconductor.org/packages/release/bioc/html/ReactomePA.html)

GO enrichment for all protein modules was performed using the ranked list for all proteins identified in a given protein module based on the mean abundances (n = 4), using Gorilla (http://cbl-gorilla.cs.technion.ac.il/). GO enrichment and KEGG pathway enrichment analysis was performed using DAVID (https://david-d.ncifcrf.gov/) and were searched against a curated list of all previously published mammalian synaptic proteins as the background list. Disease enrichment was performed using Toppcluster

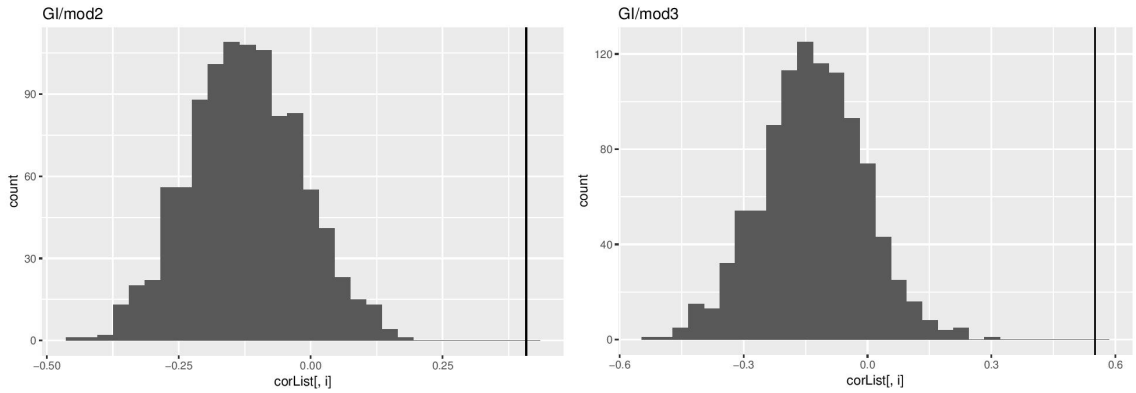
(https://toppcluster.cchmc.org/) and KEGG pathway enrichment was then performed by searching ranked protein lists obtained using GSea v 2.1.0 (http://software.broadinstitute.org/gsea/index.jsp) as previously described[53].

We assessed the non-changing proteins between the brain regions of all four individual cases by performing a Two One-Sided T-Test (TOST-test) in a similar manner as described[21], using R. An Epsilon equivalency interval of 0.045, where differences within this range are considered substantively unimportant, was used and gave the highest number of hits without including any proteins being identified as significantly differentially expressed between the different brains regions.

Circular hierarchical clustering of protein modules for the visualization of inter-region molecular interactions was performed using Circos (http://circos.ca)[54]. A Circos configuration file was created representing brain regions as 'karyotypes'. All proteins were grouped into "modules" according to their abundance similarity. Proteins that have positive abundance in more than one region were shown as links between regions. All pre-processing of the relative abundance information and generation of appropriate Circos files was performed in R.

PET analysis

To assess the significance of correlations between protein abundances with regional PET metabolic parameters[23] we performed 10,000 reshufflings of the modules; this resulted in random assignment of the module numbers (1 to 7) preserving the module size. Each randomized module list was processed to obtain the mean module abundances and correlated with PET data. The mean correlation matrix obtained by averaging each cell over 10,000 reshuffled samples is shown below for module2/GI and module3/GI pairs. Note that that the observed correlation values (vertical black line) lie far from the random distribution, indicating that the observed correlation with PET data is extremely unlikely to have occurred by chance ($p < 10^{-4}$).



MRI and fMRI analysis

MRI and fMRI data from the parcellated human cortex was taken from Glasser *et al.*,[19] via the Human Connectome DB. Corresponding Brodmann maps were generated using the Human Connectome Workbench visualization software developed by Marcus *et al*.[55] Each BA had many parcels and we therefore selected those that were most similar to one another by comparing the HCP *myelin* marker from the fMRI with the expression values of four myelin proteins (Mog, Plp1, Omg and Pmp2) in our proteomic data, which were highly correlated (R^2 = 0.7-0.8, p-value < 0.0001) (Supplementary Figure 7A). From a total of 96 parcellated regions we identified 40 (Supplementary Figure 7B, Supplementary Table 14) that most significantly correlated with average myelin expression values from our study (R^2 = 0.79, p-value = 0.001865; av1 shown in Supplementary Figure 7A). The HCP structural (myelin and cortical thickness) and behavioral scores for the selected 40 parcels were averaged with respect to the corresponding BAs. For illustration, the following four Glasser parcels: L_LO1_ROI, L_V2_ROI, L_V3_ROI, BA1 and L_V3CD_ROI were found to match to BA19, so their values for respective functions were averaged and assigned to BA19. As a result, each BA combines values from 2-4 Glasser parcels. These averaged values were tested for Pearson correlation with protein abundances across the 12 BA regions. Correlation scores are shown in Supplementary Table 15 and Figure 4B. Given the high dimensionality (1,213* 19) of the task we performed an additional perturbation analysis to test which of those correlation values were likely to be significant. For that, we shuffled the abundance values for each of the 1,213 proteins a total of 10,000 times to obtain 100,000 matrices of 1,213 proteins with

randomly assigned abundance values. Each of 10,000 matrices was correlated to the activity tasks matrix to get 10,000 correlation matrices, which were compared to the original correlation matrix. Following perturbation analysis (10,000x) and correction for multiple testing, 1,154 values remained significant ($P < 0.05$) (Supplementary table 16). We also correlated the mean abundance values for Postsynaptic Proteome Modules (PPMs) with Glasser terms (Supplementary Figure 8) and similar to the analysis for PET, described above, we assessed the significance of correlations by performing 10,000 reshufflings of the module assignment. Each randomized module list was processed to obtain the mean module abundances and correlated with fMRI data. The coefficients with absolute value above 0.53 were significant at the FDR level 5%.

Task-fMRI nomenclature

The nomenclature used for the task-fMRI (tfMRI) data is fully described in Glasser et al., 2016, Supplementary Information File 3, Table 3[19]. The tfMRI terms used for comparison with the proteomic data generated in this study are listed and fully described in Supplementary Table 13. Briefly, the tfMRI data measured the working memory (WM) of individuals tested to recognize images of body (WM_BODY), faces (WM_FACE), places (WM_PLACES) and tools (WM_TOOLS). The ability to match objects based on verbal category (MATCH) and ability to distinguish between two objects (RELATIONAL), as well as response to viewing emotional faces (EMOTION_FACES) versus neutral objects (EMOTION_SHAPES) was also measured. The various motor (MOTOR) function tasks measured by tfMRI and compared with our proteomics data, i.e., squeezing of left or right toes (MOTOR_LF_AVG or MOTOR_RF_AVG), tapping of left or right fingers (MOTOR_LH_ABG or MOTOR_RH_AVG), moving the tongue (MOTOR_T_AVG), and average of a LF, RF, LH, RH and T (MOTOR_AVG).

Calculation of regional synaptic proteome expression proportion.

A key metric used for our genetic enrichment analyzes is the proportion of expression for a given protein in a particular brain region. This is a measure of regional specificity scaled so that a value of 1 implies that the protein is exclusive to a region and a value of 0 implies the protein is not expressed in that region. We denote this specificity metric as $s_{g,c}$ for protein g and region c. Values of $s_{g,c}$ were calculated for each HGNC symbol which could be mapped from the RefSeq peptide IDs (mapping between ID's was done using *biomaRt*) (https://www.bioconductor.org/packages/devel/bioc/vignettes/biomaRt/inst/doc/biomaRt.html)

Let us denote that the proteome dataset contains measures from w cells associated with k brain regions. Each of the k regions is associated with a numerical index from the set $\{1,..,k\}$. The region annotations for sample i are stored using a numerical index in L, such that l_{1005}=5 indicates that the 1,005th sample is from the 5th region. We denote N_c as the number of samples from the region indexed by c. The expression proportion for protein g and region c (where $r_{g,i}$ is the expression of protein g in sample i) is given by:

$$s_{g,c} = \frac{\sum_{i=1}^{w} F(g,i,c)/N_c}{\sum_{r=1}^{k}(\sum_{i=1}^{w} F(g,i,r)/N_r)} \qquad F(g,i,c) = \begin{cases} r_{g,i}, l_i = c \\ 0, l_i \neq c \end{cases}$$

LD Score Regression (LDSC) and partitioning heritability

To partition heritability using LDSC (URLs)[56], it is necessary to pass LDSC annotation files (one per chromosome) with a row per SNP and a column for each sub-annotation (1=a SNP is part of that sub-annotation). To map SNPs to genes, we used dbSNP annotations (URLs, build 147 and hg19/NCBI Build 37 coordinates). All SNPs not annotated in this file were given a value of 0 in all sub-annotations. Template annotation files obtained from the LDSC Github repository were used as the basis for all region and gene set annotations ("cell_type_group.1*"). Only SNPs present in the template files were used. If an annotation had no SNPs, then 50 random SNPs from the same chromosome

were selected as part of the annotation (if no SNPs are selected then the software fails to calculate heritability).

Annotation files were created for each region for which we applied partitioned LDSC. Twelve sub-annotations were created for each region. The first represented all SNPs which map onto named regions that are not HGNC annotated genes. The second contained all SNPs which map onto genes whose protein products were not detected in the regional synaptic proteome. The other 10 sub-annotations are associated with genes with increasing levels of expression specificity for that region. To assign these, the deciles of $s_{g,c}$ were calculated over all values of g (separately for each value of c) to give ten equal length sets of genes. These are then mapped to SNPs as described above. To partition heritability amongst the gene sets (not the regions), a single set of annotation files was created with each of the gene sets used as a sub-annotation column.

LDSC was then run using associated data files from phase 3 of the 1000 Genomes Project[57]. We computed LD scores for region annotations using a 1 cM window (--ld-wind-cm 1). As recommended (LDSC Github Wiki, URLs), we restricted the analysis to using Hapmap3 SNPs, and, as in the original report[56], these analyzes excluded the major histocompatibility region due to its unusual gene density (second highest in the human genome) and exceptionally high LD (highest in the genome). The LDSC "munge_sumstats.py" script was used to prepare the summary statistics files. The heritability is then partitioned to each sub-annotation. We used LD scores calculated for HapMap3 SNPs, excluding the MHC region, for the regression weights available from the Github page (files in the 'weights_hm3_no_hla' folder).

For the LD score files used as independent variables in LD Score regression we used the full baseline model[56] and the annotations described above. We used the '--overlap-annot' argument and the minor allele frequency files ('1000G_Phase3_frq' folder via the '--frqfile-chr' argument, URLs).

Partitioned LDSC computes the proportion of heritability associated with each annotation column while taking into account all other annotations. Based on the proportion of total SNPs in an annotation, LDSC calculates an enrichment score and an associated enrichment P-value (one-tailed as we were only interested in annotations showing enrichments of heritability). A linear model was then fit to the enrichment z-scores for the 12 gene categories for each region and GWAS, and the one-tailed probability calculated that the slope is positively associated with increasing regional specificity in the synaptic proteome. The slope of this model is then used to generate the plots in Figure 5.

Comparison GWA results for other traits

We included comparisons for a selected set of brain and non-brain diseases, disorders, and traits. The GWA results were from the following sources: autism spectrum disorders and major depression[58]; schizophrenia[58] from the PGC; Migraine[59]; Anorexia (www.med.unc.edu/pgc/results-and-downloads); Alzheimer's disease[60]; aggression[61]; educational attainment[62]; smoking[63]; type 2 diabetes mellitus[64]; height[65]; Crohn's disease, inflammatory bowel disease and ulcerative colitis[66]; and low-density lipoprotein (LDL), high-density lipoprotein (HDL), total cholesterol, and triglyceride levels[67]. The summary statistics files can all be found through the PGC website (www.med.unc.edu/pgc) and LDHUB[68].

Acknowledgements

Support from the Medical Research Council (Brain Bank MR/L016400/1) and European Union Seventh Framework Programme (FP7 grant agreement no. 604102) and Horizon 2020 (agreement no. 720270). T. Le Bihan and L. Imrie at SynthSys, University of Edinburgh for mass spectrometry sample analysis. The LC-MS QExactive equipment was purchased by a Wellcome Trust Institutional Strategic Support Fund and a strategic award from the Wellcome Trust for the Centre for Immunity, Infection and Evolution (095831/Z/11/Z). Data were extracted from NIFTI (Neuroimaging Informatics Technology Initiative) files using custom automated script written by Jeremy J Roy, MEMEX, Inc.,

Burlington, Ontario, Canada. MRI data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. K. Elsegood for laboratory management. J. DeFelipe for comments on the manuscript. D. Maizels for artwork.

Author contributions

CS supplied brain tissue samples; MR performed biochemistry; FM, RZ, ES performed electrophysiology; MR, OS, NS performed bioinformatics and statistical analysis; JDA provided supervision; SG conceived and supervised the project, wrote the manuscript and secured funding.

References

1. Brodmann, K. (ed.) *On the comparative localization of the cortex. Leipzig: J.A. Barth. In G. von Bonin, Some papers on the Cerebral Cortex.* , (Charles C Thomas, Springfield, IL, 1909).
2. DeFelipe, J., Alonso-Nanclares, L. & Arellano, J.I. Microstructure of the neocortex: comparative aspects. *J Neurocytol* **31**, 299-316 (2002).
3. Bayes, A. *et al.* Comparative study of human and mouse postsynaptic proteomes finds high compositional conservation and abundance differences for key synaptic proteins. *PLoS One* **7**, e46683 (2012).
4. Bayes, A. *et al.* Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nature neuroscience* **14**, 19-21 (2011).
5. Bayes, A. *et al.* Human post-mortem synapse proteome integrity screening for proteomic studies of postsynaptic complexes. *Mol Brain* **7**, 88 (2014).
6. Coba, M.P. *et al.* Neurotransmitters drive combinatorial multistate postsynaptic density networks. *Sci Signal* **2**, ra19 (2009).
7. Li, J. *et al.* Long-term potentiation modulates synaptic phosphorylation networks and reshapes the structure of the postsynaptic interactome. *Sci Signal* **9**, rs8 (2016).
8. Fernandez, E. *et al.* Targeted tandem affinity purification of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins. *Mol Syst Biol* **5**, 269 (2009).
9. Frank, R.A. *et al.* NMDA receptors are selectively partitioned into complexes and supercomplexes during synapse maturation. *Nat Commun* **7**, 11264 (2016).
10. Husi, H., Ward, M.A., Choudhary, J.S., Blackstock, W.P. & Grant, S.G. Proteomic analysis of NMDA receptor-adhesion protein signaling complexes. *Nat Neurosci* **3**, 661-9 (2000).

11. Komiyama, N.H. *et al.* SynGAP regulates ERK/MAPK signaling, synaptic plasticity, and learning in the complex with postsynaptic density 95 and NMDA receptor. *J Neurosci* **22**, 9721-32 (2002).
12. Frank, R.A.W., Zhu, F., Komiyama, N.H. & Grant, S.G.N. Hierarchical organisation and genetically separable subfamilies of PSD95 postsynaptic supercomplexes. *J Neurochem* (2017).
13. Collins, M.O. *et al.* Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome. *J Neurochem* **97 Suppl 1**, 16-23 (2006).
14. Cheng, D. *et al.* Relative and absolute quantification of postsynaptic density proteome isolated from rat forebrain and cerebellum. *Mol Cell Proteomics* **5**, 1158-70 (2006).
15. Li, K.W. *et al.* Proteomics analysis of rat brain postsynaptic density. Implications of the diverse protein functional groups for the integration of synaptic physiology. *J Biol Chem* **279**, 987-1002 (2004).
16. Peng, J. *et al.* Semiquantitative proteomic analysis of rat forebrain postsynaptic density fractions by mass spectrometry. *J Biol Chem* **279**, 21003-11 (2004).
17. Bayes, A. *et al.* Evolution of complexity in the zebrafish synapse proteome. *Nat Commun* **8**, 14613 (2017).
18. Emes, R.D. *et al.* Evolutionary expansion and anatomical specialization of synapse proteome complexity. *Nature neuroscience* **11**, 799-806 (2008).
19. Glasser, M.F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171-8 (2016).
20. Hawrylycz, M. *et al.* Canonical genetic signatures of the adult human brain. *Nat Neurosci* **18**, 1832-44 (2015).
21. Le Bihan, T. *et al.* Label-free quantitative analysis of the casein kinase 2-responsive phosphoproteome of the marine minimal model species Ostreococcus tauri. *Proteomics* **15**, 4135-44 (2015).
22. Mazzo, F. *et al.* Reconstitution of synaptic Ion channels from rodent and human brain in Xenopus oocytes: a biochemical and electrophysiological characterization. *J Neurochem* **138**, 384-96 (2016).
23. Vaishnavi, S.N. *et al.* Regional aerobic glycolysis in the human brain. *Proc Natl Acad Sci U S A* **107**, 17757-62 (2010).
24. Wu, K., Aoki, C., Elste, A., Rogalski-Wilk, A.A. & Siekevitz, P. The synthesis of ATP by glycolytic enzymes in the postsynaptic density and the effect of endogenously generated nitric oxide. *Proc Natl Acad Sci U S A* **94**, 13273-8 (1997).
25. Barch, D.M. *et al.* Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* **80**, 169-89 (2013).
26. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
27. Brody, A.L. *et al.* Differences between smokers and nonsmokers in regional gray matter volumes and densities. *Biol Psychiatry* **55**, 77-84 (2004).
28. Gowin, J.L., Mackey, S. & Paulus, M.P. Altered risk-related processing in substance users: imbalance of pain and gain. *Drug Alcohol Depend* **132**, 13-21 (2013).
29. Janes, A.C. *et al.* Brain reactivity to smoking cues prior to smoking cessation predicts ability to maintain tobacco abstinence. *Biol Psychiatry* **67**, 722-9 (2010).
30. Brody, A.L. *et al.* Brain metabolic changes during cigarette craving. *Arch Gen Psychiatry* **59**, 1162-72 (2002).
31. Bayes, A. *et al.* Zebrafish synapse proteome complexity, evolution and ultrastructure. *Nat Commun* (2017).

32. Pocklington, A.J., Cumiskey, M., Armstrong, J.D. & Grant, S.G. The proteomes of neurotransmitter receptor complexes form modular networks with distributed functionality underlying plasticity and behaviour. *Molecular systems biology* **2**, 2006 0023 (2006).
33. Broadhead, M.J. *et al.* PSD95 nanoclusters are postsynaptic building blocks in hippocampus circuits. *Sci Rep* **6**, 24626 (2016).
34. Buckner, R.L. *et al.* Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. *J Neurosci* **25**, 7709-17 (2005).
35. Vlassenko, A.G. *et al.* Spatial correlation between brain aerobic glycolysis and amyloid-beta (Abeta) deposition. *Proc Natl Acad Sci U S A* **107**, 17763-7 (2010).
36. Weinberger, D.R., Berman, K.F. & Zec, R.F. Physiologic dysfunction of dorsolateral prefrontal cortex in schizophrenia. I. Regional cerebral blood flow evidence. *Arch Gen Psychiatry* **43**, 114-24 (1986).
37. Dalack, G.W., Healy, D.J. & Meador-Woodruff, J.H. Nicotine dependence in schizophrenia: clinical phenomena and laboratory findings. *Am J Psychiatry* **155**, 1490-501 (1998).
38. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179-84 (2014).
39. Kirov, G. *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry* **17**, 142-53 (2012).
40. Purcell, S.M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185-90 (2014).
41. Hallahan, B.P. *et al.* In vivo brain anatomy of adult males with Fragile X syndrome: an MRI study. *Neuroimage* **54**, 16-24 (2011).
42. Mahajan, R. & Mostofsky, S.H. Neuroimaging endophenotypes in autism spectrum disorder. *CNS Spectr* **20**, 412-26 (2015).
43. Philip, R.C. *et al.* A systematic review and meta-analysis of the fMRI investigation of autism spectrum disorders. *Neurosci Biobehav Rev* **36**, 901-42 (2012).
44. Stanfield, A.C. *et al.* Towards a neuroanatomy of autism: a systematic review and meta-analysis of structural magnetic resonance imaging studies. *Eur Psychiatry* **23**, 289-99 (2008).
45. Stigler, K.A., McDonald, B.C., Anand, A., Saykin, A.J. & McDougle, C.J. Structural and functional magnetic resonance imaging of autism spectrum disorders. *Brain Res* **1380**, 146-61 (2011).
46. Hawrylycz, M.J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391-9 (2012).
47. Blazquez-Llorca, L., Garcia-Marin, V. & DeFelipe, J. GABAergic complex basket formations in the human neocortex. *J Comp Neurol* **518**, 4917-37 (2010).
48. Le Bihan, T., Grima, R., Martin, S., Forster, T. & Le Bihan, Y. Quantitative analysis of low-abundance peptides in HeLa cell cytoplasm by targeted liquid chromatography/mass spectrometry and stable isotope dilution: emphasising the distinction between peptide detection and peptide identification. *Rapid Commun Mass Spectrom* **24**, 1093-104 (2010).
49. Team, R.C. *R: a language and environment for statistical computing.* , (R Foundation for Statistical Computing, Vienna, Austria, 2013).
50. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* **57**, 289-300 (1995).
51. Charrad, M., Ghazzali, N., Boiteau, V. & Niknafs, A. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *J. Statistical Software* **61**(2014).
52. Simpson, T.I., Armstrong, J.D. & Jarman, A.P. Merged consensus clustering to assess and improve class discovery with microarray data. *BMC Bioinformatics* **11**, 590 (2010).

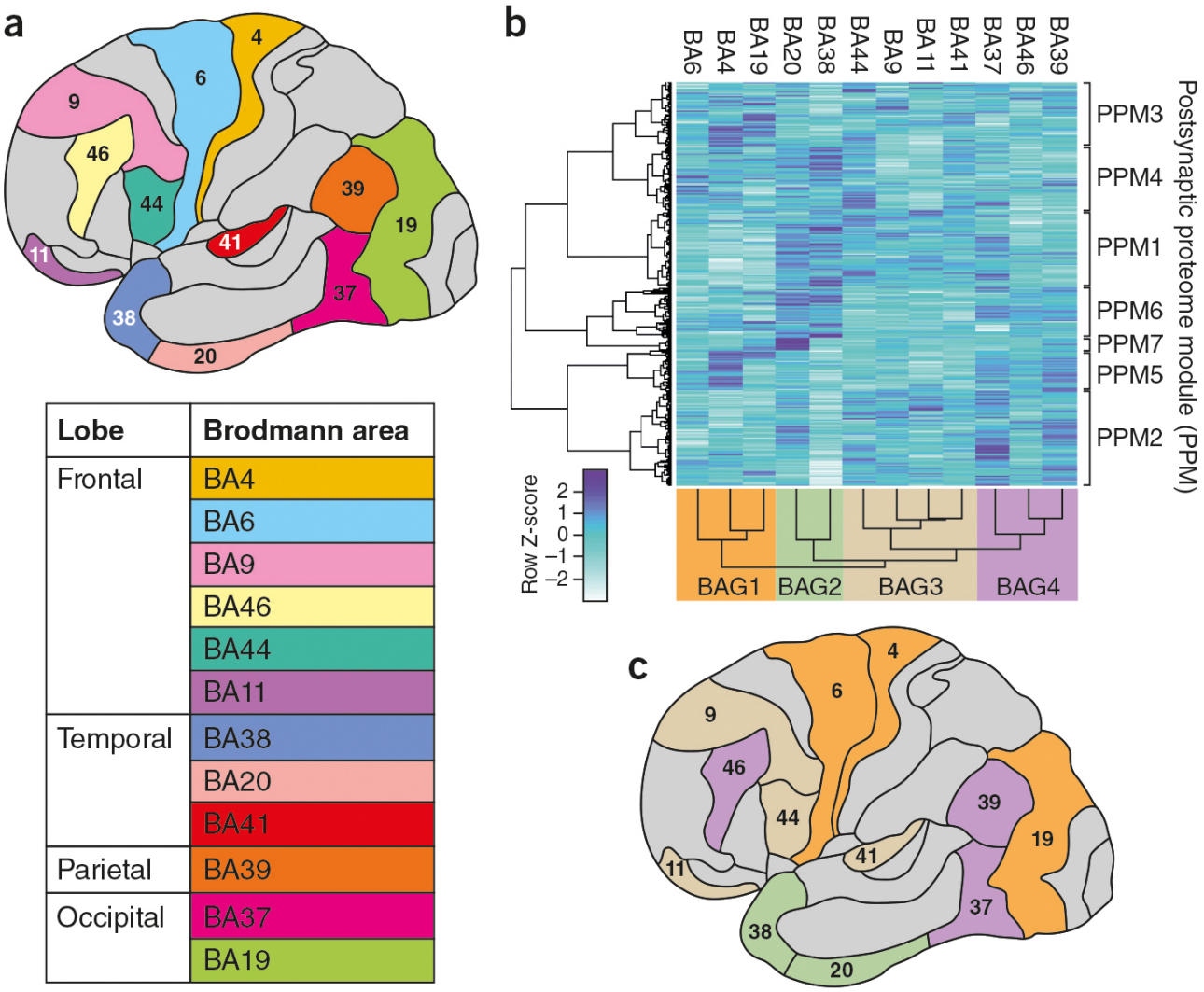
53. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).
54. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-45 (2009).
55. Marcus, D.S. *et al.* Informatics and data mining tools and strategies for the human connectome project. *Front Neuroinform* **5**, 4 (2011).
56. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
57. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
58. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-7 (2014).
59. Gormley, P. *et al.* Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat Genet* **48**, 856-66 (2016).
60. Lambert, J.C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452-8 (2013).
61. Pappa, I. *et al.* A genome-wide approach to children's aggressive behavior: The EAGLE consortium. *Am J Med Genet B Neuropsychiatr Genet* **171**, 562-72 (2016).
62. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539-42 (2016).
63. Tobacco & Genetics, C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42**, 441-7 (2010).
64. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* **44**, 981-90 (2012).
65. Wood, A.R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* **46**, 1173-86 (2014).
66. Liu, J.Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* **47**, 979-86 (2015).
67. Global Lipids Genetics, C. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-83 (2013).
68. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279 (2017).

Table 1

Lobe	Brodmann area	Name	Functions	Pathology
Frontal	BA4	Primary motor cortex	Motor function, homunculus	Upper motor neuron motor disorders; hemiparesis
	BA6	Premotor cortex	Movement control, motor sequencing	Cerebrovascular disease
	BA9	Medial part of dorsolateral prefrontal cortex	Executive functions, inhibition, abstract thinking	Schizophrenia, substance abuse
	BA46	Dorsolateral prefrontal cortex	Attention, working memory	Alzheimer's disease; neoplasia; dysexecutive syndrome
	BA44	Pars opercularis of the inferior frontal gyrus	Broca's area, language production	Cerebrovascular disease, expressive dysphas
	BA11	Orbitofrontal cortex	Decision making, emotion, reward	Traumatic brain injury
Temporal	BA38	Anterior temporal lobe	Linguistic, semantic, visuospatial, audiovisual	Traumatic brain injury
	BA20	Inferior temporal area	Language, vision, working memory	Cerebrovascular disease
	BA41	Anterior transverse temporal gyrus	Auditory processing, memory, speech perception	Cerebrovascular disease
Parietal	BA39	Angular area	Language, mathematics, visuospatial	Cerebrovascular disease; Gerstmann syndrome
Occipital	BA37	Occipitotemporal area	Language processing, memory, visual processing	Cerebrovascular disease
	BA19	Peristriate area	Language processing, memory, visual processing, retinotopic maps	Cerebrovascular disease

Table 1. Brodmann areas in frontal, temporal, parietal and occipital lobes of the neocortex and summary of their functions and pathology. Color code as in Figure 1A.

Figure 1



Lobe	Brodmann area
Frontal	BA4
	BA6
	BA9
	BA46
	BA44
	BA11
Temporal	BA38
	BA20
	BA41
Parietal	BA39
Occipital	BA37
	BA19

Figure 1. Postsynaptic proteome composition in 12 Brodmann Areas (BA). (A) Twelve BAs distributed into frontal, temporal, parietal and occipital lobes (color coded as in Table 1). (B) Hierarchical clustering by BA (x-axis) and protein abundance (y-axis) shows each BA has a unique signature of postsynaptic proteome composition. The 12 BAs were clustered into four Brodmann Area Groups (BAG 1-4) and the 1,213 proteins into seven Postsynaptic Proteome Modules (PPM 1-7). (C) Neuroanatomical map of BAGs color coded as in (B).

Figure 2

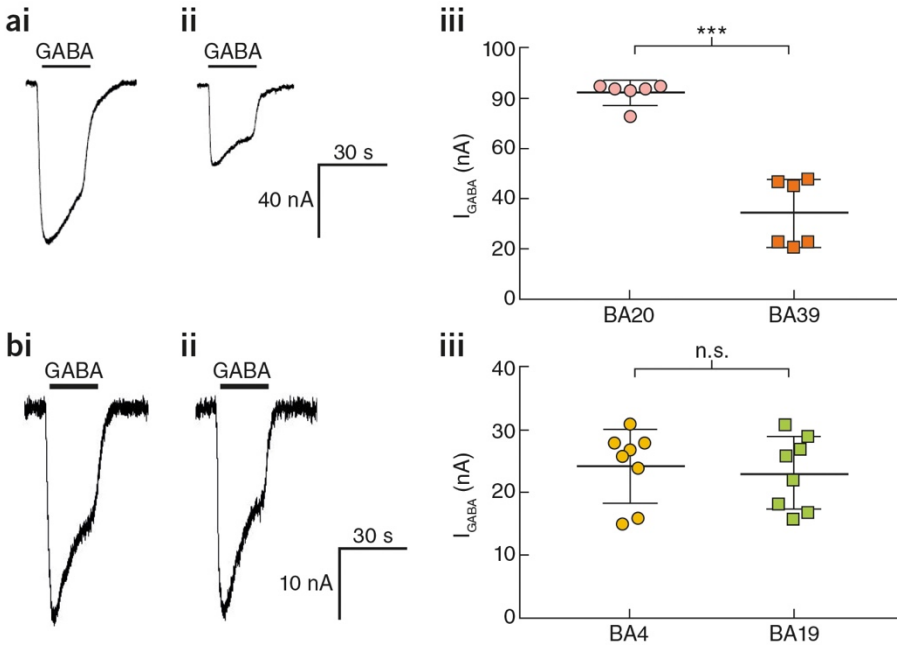


Figure 2. Functional characterization of GABAA receptors from different cortical brain areas reconstituted in *Xenopus* oocytes. ai) Oocytes injected with synaptosomes prepared from brain area BA20 responded with large 1 mM GABA-evoked chloride currents and aii) oocytes injected with synaptosomes prepared from brain area BA39 responded with smaller 1 mM GABA-evoked chloride currents. aiii) Summary of the average size of the GABA-evoked currents obtained with the two brain areas. The 1 mM GABA evoked ion currents had amplitudes of 82 ± 2 nA (n = 6) and 35 ± 5 nA (n = 6) for BA20 and BA39, respectively. These values were statistically different ($p<0.001$) b) Oocytes injected with synaptosomes prepared from brain areas BA4 and BA19 responded to application of 1 mM GABA with ion currents that did not statistically differ in their amplitudes. bi) Example of 1 mM GABA-induced response of an oocyte expressing GABAA receptors from BA4. bii) Example of 1 mM GABA-induced response of an oocyte expressing GABAA receptors from BA19. biii) Summary of the average size of the GABA-evoked currents obtained with these two brain areas. 1 mM GABA evoked ion currents with amplitudes of 24 ± 2 nA (n = 8) and 23 ± 2 nA) (n = 8), for BA4 and BA19, respectively.

Figure 3

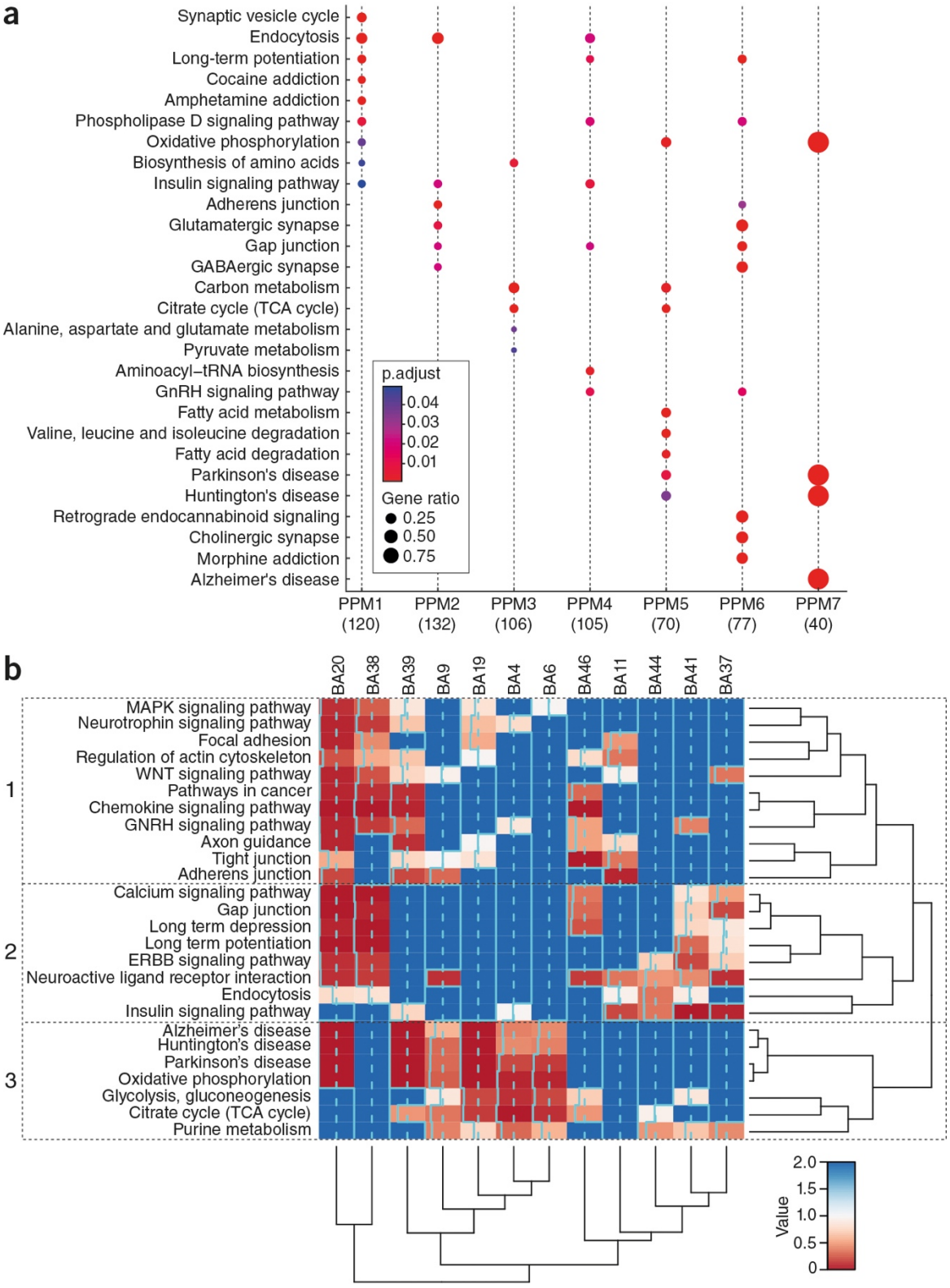

a
Synaptic vesicle cycle
Endocytosis
Long-term potentiation
Cocaine addiction
Amphetamine addiction
Phospholipase D signaling pathway
Oxidative phosphorylation
Biosynthesis of amino acids
Insulin signaling pathway
Adherens junction
Glutamatergic synapse
Gap junction
GABAergic synapse
Carbon metabolism
Citrate cycle (TCA cycle)
Alanine, aspartate and glutamate metabolism
Pyruvate metabolism
Aminoacyl−tRNA biosynthesis
GnRH signaling pathway
Fatty acid metabolism
Valine, leucine and isoleucine degradation
Fatty acid degradation
Parkinson's disease
Huntington's disease
Retrograde endocannabinoid signaling
Cholinergic synapse
Morphine addiction
Alzheimer's disease
p.adjust
0.04
0.03
0.02
0.01
Gene ratio
0.25
0.50
0.75
PPM1 (120)
PPM2 (132)
PPM3 (106)
PPM4 (105)
PPM5 (70)
PPM6 (77)
PPM7 (40)
b
BA20
BA38
BA39
BA9
BA19
BA4
BA6
BA46
BA11
BA44
BA41
BA37
1
MAPK signaling pathway
Neurotrophin signaling pathway
Focal adhesion
Regulation of actin cytoskeleton
WNT signaling pathway
Pathways in cancer
Chemokine signaling pathway
GNRH signaling pathway
Axon guidance
Tight junction
Adherens junction
2
Calcium signaling pathway
Gap junction
Long term depression
Long term potentiation
ERBB signaling pathway
Neuroactive ligand receptor interaction
Endocytosis
Insulin signaling pathway
3
Alzheimer's disease
Huntington's disease
Parkinson's disease
Oxidative phosphorylation
Glycolysis, gluconeogenesis
Citrate cycle (TCA cycle)
Purine metabolism
Value
2.0
1.5
1.0
0.5
0

Figure 3. Biochemical pathways and functions in Brodmann areas (BA) and Postsynaptic Proteome Modules (PPM). (A) KEGG pathway terms (y-axis) enrichment in PPMs (x-axis) with the number of proteins contributing to KEGG enrichment indicated in brackets. Size of the dots represents the number of genes associated with that pathway (GeneRatio) and the significance indicated by the p-adjust color bar. (B) Heatmap of the KEGG biochemical pathway and disease enrichment terms (y-axis) based on the ranked abundance of post-synaptic proteins in each BA (x-axis). Shown is the significance (p-value) of enrichment, where red (near 0 values) corresponds to highly enriched terms (< 0.01) and blue (default value set to 2) - to non-enriched. Three clusters of terms are bracketed: 1) many signal transduction mechanisms, 2) synaptic plasticity and other signaling processes, 3) neurodegenerative diseases and metabolic mechanisms. The dashed blue line represents mean value.

Figure 4

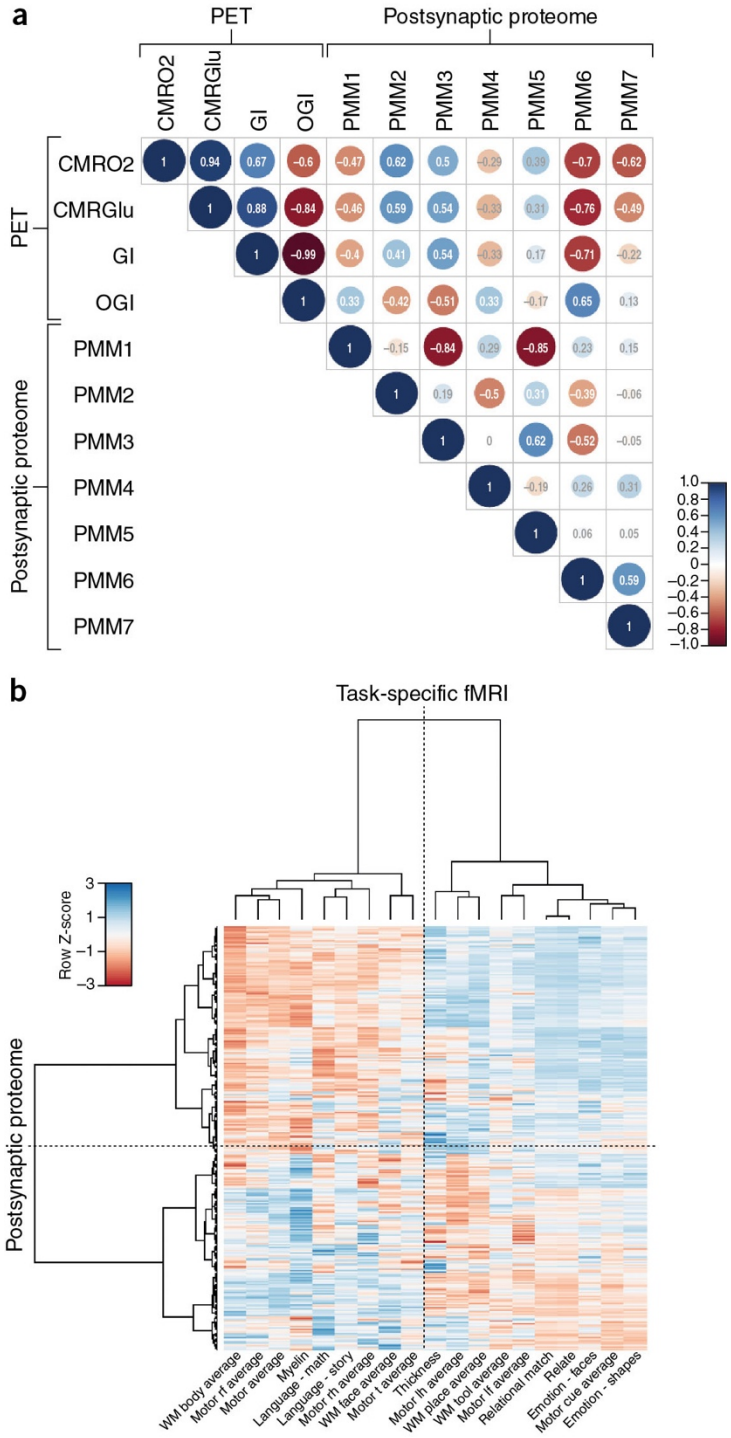


Figure 4. PET imaging of metabolism and fMRI imaging of behavior. (A) Correlation between brain region-specific metabolism in human brain regions and protein modules. PET data of metabolic measures ($CMRO_2$, cerebral metabolic rate for oxygen; CMRGlu,

cerebral metabolic rate for glucose; GI, glycolytic index; OGI, oxygen-glucose index)[23] correlated with PPM abundances (PPM1-7). Color scale bar and Pearson's R^2 shown on right-hand side proportional to dot size. (B) Hierarchical clustering between neocortical localization of behavioral responses (task-specific fMRI) and brain region-specific postsynaptic proteome abundance. Two major clusters of behaviors and postsynaptic proteomes indicated by dotted lines. Each row corresponds to one protein; color bar represents the transformed correlation values from Supplementary Table 13 on the spatial correlation between that protein abundance levels and activation.

Figure 5

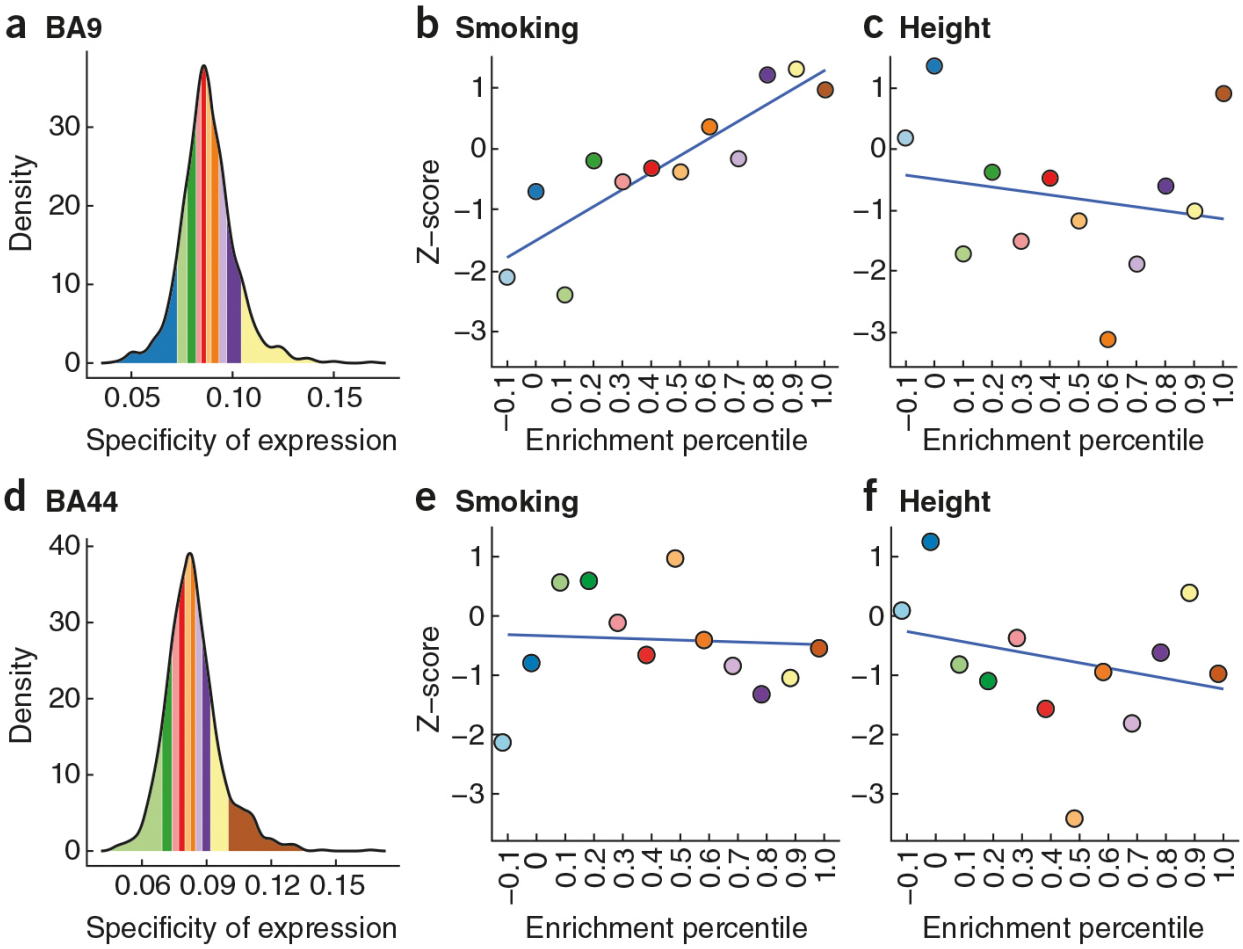


Figure 5. Genetic association of smoking to BA9

The genetic association with smoking increases, the more specifically the proteins are localized to BA9 (A-C). (A) Distribution of BA9 specificity values for all synaptic proteins divided into deciles (represented by colored blocks). B) Genetic association with smoking (current vs former smokers) increases in decile groups that are more specific to BA9 (color coded as in A). The two leftmost dots do not represent specificity decile groups, and instead represent respectively: all SNPs which map onto named regions which are not HGNC annotated genes; and all SNPs which map onto genes whose protein products were not detected in the regional synaptic proteome. The values shown are enrichment values and not enrichment z-scores. Error bars indicate the 95% confidence intervals. C)

Absence of a genetic association with height as the decile groups become more specific to BA9. D) Distribution of BA44 specificity values for all synaptic proteins divided into deciles (represented by colored blocks). E) Absence of a genetic association with smoking as the decile groups become more specific to BA44. F) Absence of a genetic association with height as the decile groups become more specific to BA44.