

Integration across biophysical scales identifies molecular and cellular correlates of person-to-person variability in human brain connectivity

Received: 21 August 2023

Accepted: 16 September 2024

Published online: 31 October 2024

 Check for updates

Bernard Ng ^{1,2}, Shinya Tasaki ¹, Kelsey M. Greathouse³, Courtney K. Walker ³, Ada Zhang², Sydney Covitz^{4,5,6}, Matt Cieslak ^{4,5,6}, Audrey J. Weber³, Ashley B. Adamson ³, Julia P. Andrade³, Emily H. Poovey³, Kendall A. Curtis³, Hamad M. Muhammad³, Jakob Seidlitz ^{4,5,6,7}, Ted Satterthwaite ^{4,5,6}, David A. Bennett¹, Nicholas T. Seyfried ⁸, Jacob Vogel^{4,5,6,9}, Chris Gaiteri ^{1,2} ✉ & Jeremy H. Herskowitz ³ ✉

Brain connectivity arises from interactions across biophysical scales, ranging from molecular to cellular to anatomical to network level. To date, there has been little progress toward integrated analysis across these scales. To bridge this gap, from a unique cohort of 98 individuals, we collected antemortem neuroimaging and genetic data, as well as postmortem dendritic spine morphometric, proteomic and gene expression data from the superior frontal and inferior temporal gyri. Through the integration of the molecular and dendritic spine morphology data, we identified hundreds of proteins that explain interindividual differences in functional connectivity and structural covariation. These proteins are enriched for synaptic structures and functions, energy metabolism and RNA processing. By integrating data at the genetic, molecular, subcellular and tissue levels, we link specific biochemical changes at synapses to connectivity between brain regions. **These results demonstrate the feasibility of integrating data from vastly different biophysical scales to provide a more comprehensive understanding of brain connectivity.**

A long-standing goal of neuroscience is to understand how molecules and cellular structures at the microscale give rise to communication between brain regions at the macroscale. However, there is a disconnect between molecular biology and neuroimaging that hinders progress toward this goal. A paradigm shift would be to collect postmortem omics data alongside antemortem neuroimaging data from the same set of individuals. In this way, molecular levels can be tied to the functional and structural properties of the brain. Establishing such associations is imperative for translation from humans to experimental model systems that could better recapitulate neurologic health and disease.

Both molecular and neuroimaging studies have independently identified correlates of brain function and cognition, but the search for the molecular basis of functional connectivity is nascent. Recent neuroimaging genetics studies are beginning to address the question of which molecules are most related to functional connectivity^{1–5}. However, despite thousands of samples, these studies have found only a dozen independent genetic loci with modest heritability^{6,7}. Most of those genetic variants reside in noncoding regions, making it difficult to determine the implicated genes and mechanisms of action. Another class of studies tests for spatial correspondence between gene

A full list of affiliations appears at the end of the paper. ✉ e-mail: gaiteri@gmail.com; jhersko@uab.edu

expression and brain connectivity to infer the implicated genes^{8–12}. These studies typically combine the Allen Human Brain Atlas (AHBA)¹³, which contains gene expression profiles in hundreds of brain regions from six individuals, with functional magnetic resonance imaging (fMRI) data from an entirely separate cohort. Such data combinations cannot examine the covariation between gene expression and functional connectivity in a broader population. Representing such covariation is critical for identifying molecules that explain individual variation in functional connectivity and the associated cognitive phenotypes¹⁴.

In this study, we gathered data types that span multiple biophysical scales from a cohort of 98 individuals. Data types included resting-state fMRI, structural MRI, genetics, dendritic spine morphometry, proteomics and gene expression measurements from postmortem tissues of the superior frontal gyrus (SFG) and inferior temporal gyrus (ITG). Based on the stability of functional connectivity patterns within individuals¹⁵, we hypothesized that it is possible to combine postmortem molecular and subcellular data with antemortem neuroimaging data from the same individuals to prioritize molecular mechanisms underlying brain connectivity. To test this hypothesis, we built models that integrate protein measurements with dendritic spine morphometry to explain between-individual variation in functional connectivity. As a form of replication, we repeated the analysis using gene expression measurements in place of protein abundance and structural covariation as a surrogate of connectivity.

Results

Multimodal human brain data

To model how molecules at the microscale give rise to brain connectivity at the macroscale, we collected data that span multiple biophysical scales (Fig. 1) as part of the Religious Orders Study and Rush Memory and Aging Project (ROSMAP)¹⁶. Data types included genetics, gene expression, protein abundance, dendritic spine morphometry and neuroimaging measurements. Ninety-eight ROSMAP participants had all these data types measured. The average age of these participants at the time of the MRI scan and death was 88 ± 6 years and 91 ± 6 years, respectively, with an average time interval between the MRI scan and the age at death of 3 ± 2 years. The average postmortem interval (PMI) was 8.5 ± 4.6 h. Participants were 77% female and had 15 ± 3 years of education. We performed detailed characterization of each omic, cellular and neuroimaging data type and then integrated them as described next.

Neuroimaging phenotypes

Following current best practices, we organized the neuroimaging data from 1,210 ROSMAP participants into the brain imaging data structure (BIDS)¹⁷, which we validated using CuBIDS¹⁸ (Fig. 2a). For the fMRI data, we performed standard preprocessing, such as slice-time correction and spatial normalization, and regressed out motion confounds¹⁹. We made the fMRI data and their derivatives in BIDS as a free resource at radc.rush.edu to ease data sharing and re-analysis by interested investigators. To estimate functional connectivity, we first used a functional atlas comprising 100 parcels generated from resting-state fMRI data of more than 1,000 participants²⁰ to divide the brain into functionally homogeneous regions. We then averaged the time series of voxels within each parcel (Fig. 2b) and computed Pearson's correlation between all pairs of parcels for each participant (Fig. 2b). For the structural MRI data, we performed nonuniformity correction, skull-stripping, spatial normalization and tissue segmentation. We then divided the brain into 62 anatomical regions using the Desikan–Killiany–Tourville (DKT)²¹ atlas (Fig. 2d) and extracted structural attributes²² (Fig. 2e). To estimate structural covariation as a surrogate of brain connectivity, we adopted a canonical correlation analysis (CCA) approach (Supplementary Fig. 1). The motivation was based on the observation that multiple structural attributes (for example, number of vertices, surface area and curvature index) are typically cross-correlated between two regions (Fig. 2f).

Molecular measurements

We performed multiplex tandem mass tag mass spectrometry (TMT-MS) on tissue samples from SFG and ITG of 130 ROSMAP participants to generate proteomic data and applied standard preprocessing^{23–25}. To identify the major molecular systems present in each brain region, we clustered the measured proteins into covarying sets (modules) using a purely data-driven approach (Fig. 3a)²⁶. The modules are enriched for many brain-relevant structures and functions (Fig. 3b,c)²⁷ that are consistent with related datasets^{28,29}. Comparing SFG and ITG, we observed some common molecular systems with high protein overlap between their modules (Fig. 3d). In addition to protein abundance, we also collected RNA sequencing (RNA-seq) data from SFG and ITG of the same set of ROSMAP participants. We performed standard preprocessing, such as TMM normalization and confound regression with voom/limma, and clustered the measured genes into modules²⁶. These expression modules are also enriched for brain-relevant structures and functions (Supplementary Fig. 2), and we again observed common molecular systems between SFG and ITG with high gene overlap.

Dendritic spine morphometry

Dendritic spines are actin-rich protrusions along neuronal dendrites that form the majority of excitatory synapses in the brain. Spines can exhibit remarkable variability in size, shape and density, and synapse activity is inseparably linked to spine morphology³⁰. Spines can be divided into morphological subclasses based on their three-dimensional (3D) structure as thin, mushroom, stubby or filopodia^{31,32}. To measure spine density and morphology from SFG and ITG, we impregnated postmortem tissue slices with Golgi stain and imaged at $\times 60$ using a widefield microscope with a high-numerical-aperture condenser (Fig. 4a). We then reconstructed the Z stacks in 3D using Neurolucida 360 (Fig. 4b,c). We sampled between 8 and 12 pyramidal neurons from either cortical layer II or III per individual for analysis and estimated average spine density, backbone length, head diameter and volume across the reconstructed dendrites (Fig. 4d,e and Supplementary Fig. 3). We also estimated these spine attributes for each morphologic subclass—thin, mushroom, stubby and filopodia. Next, we associated each spine attribute against each measured protein (Supplementary Table 1) and applied geneset enrichment analysis (GSEA)³³. Spine density and morphological attributes were enriched for relevant neuron structures, such as synapses and the actin cytoskeleton, as well as functions, including synaptic signaling and neurotransmitter release^{30,34} (Fig. 4f). Notably, we found the level of enrichment varies across dendritic spine subclasses. Comparing the spine density and morphology between SFG and ITG (Fig. 4g) revealed differences in overall spine density ($P = 0.0310$), filopodia density ($P = 0.0038$) and mushroom spine head diameter ($P = 0.0060$).

Protein modules covary with connectivity when contextualized with dendritic spine morphology

Fundamental to brain connectivity at the macroscale are molecular sub-processes related to synaptic communication. Accordingly, we focused our analysis on protein modules most enriched for neurons and various synaptic communication gene ontology (GO) terms (Supplementary Fig. 4 and see SFG pMod6 and ITG pMod8 in Supplementary Table 2). We represented the synaptic module of each brain region by the average protein abundance of its members. To avoid association introduced by spurious correlation, we account for confounding factors including age at death, age of scan, sex, years of education, scanner, postmortem interval, side of the brain the molecular data were acquired and motion (Supplementary Fig. 5) using the extra sum of squares in testing if the synaptic modules of SFG and ITG (their main effects and interaction in aggregate) are associated with an fMRI estimate of their connectivity (Fig. 1). This initial test did not detect an association between the synaptic modules and SFG–ITG connectivity ($P = 0.6839$), and this observation holds true for other modules (Supplementary Fig. 6).

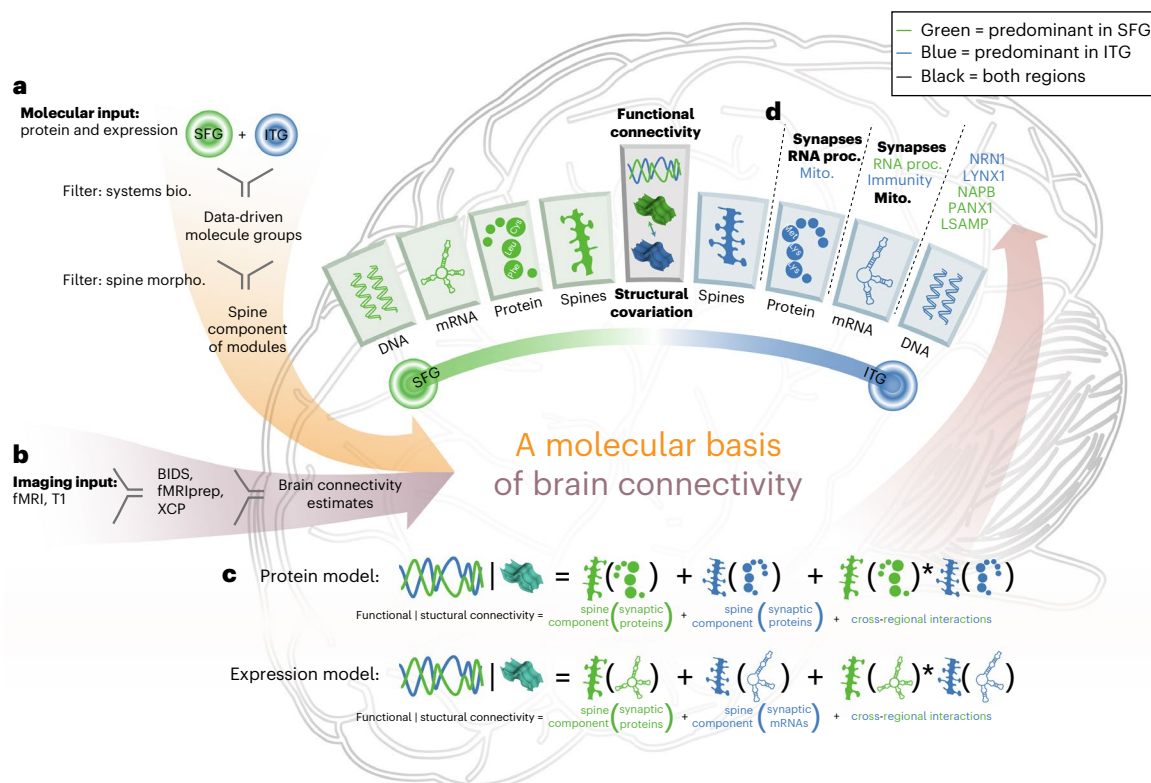


Fig. 1 | Study overview. **a**, Dendritic spine morphometric, genome-wide gene expression and protein abundance data were acquired from SFG and ITG in a cohort of 98 individuals. **b**, Structural and functional MRI data were acquired from the same 98 individuals. **c**, Molecular measurements were integrated with

dendritic spine morphologies to model functional connectivity and structural covariation between SFG and ITG. **d**, Key molecular processes and proteins associated with SFG–ITG connectivity. **Bold formatting indicates associations with both functional connectivity and structural covariation.**

that were generated at various module resolutions and by different algorithms (whether SpeakEasy (SE) or the widely used weighted gene co-expression network analysis (WGCNA)³⁵). Based on this observation, we posited that proteins reside at a biophysical scale that might be too far removed from region-level functional connectivity. Considering dendritic spines are tightly coordinated with synaptic function and even subtle alterations in spine morphology can induce marked effects on neuronal circuits^{30,36}, we hypothesized that morphologic attributes of dendritic spines could provide the cellular context to bridge the difference in biophysical scales between proteins and region-level connectivity. To test this hypothesis, we estimated the dendritic spine component of the synaptic modules by fitting the average protein abundance of their members with dendritic spine attributes (Fig. 5a–c) and repeated the analysis. Using the dendritic spine component of the synaptic modules resulted in an association with SFG–ITG connectivity ($P = 0.0174$; Fig. 5d). We next sought to test whether the association was specific to the SFG–ITG connection. Such a test would lend credence and determine the extent to which connectivity across the brain can be explained by dendritic spines and proteins gathered from SFG and ITG. Accordingly, we associated the dendritic spine component of SFG and ITG synaptic modules with the connectivity of all other region pairs. We found only 3% of other connections have higher association strength (Fig. 5e), and these connections generally emanate from either SFG and its nearby areas in the frontal cortex or from ITG. This result likely reflects how spatially proximal areas tend to have similar functional and molecular architectures^{37,38}. Moreover, the interaction between the synaptic modules of SFG and ITG explains additional variance beyond their main effects ($P = 0.0140$), with only 1.5% of other connections showing stronger interaction effects. This result suggests that not only is the synaptic module of each region individually linked to SFG–ITG connectivity but also the collective molecular state of these regions has

a role. As an exploratory analysis, we further tested the other modules and found the dendritic spine component of modules predominantly enriched for RNA processing, mitochondria and synaptic vesicles are also associated with functional connectivity (Supplementary Fig. 7).

Replication of module-level results with structural covariation and gene expression

Due to the interplay between functional connectivity and structural morphologies^{39–42}, we repeated the above-mentioned analysis with an MRI estimate of structural covariation between SFG and ITG as the outcome. Similar to the functional connectivity results, we did not find an association between SFG–ITG structural covariation and the synaptic protein modules ($P = 0.8815$) unless we honed into the dendritic spine component of the synaptic protein modules ($P = 0.0034$; Fig. 5f). We also observed regional specificity, with only 0.4% of other brain region pairs showing stronger association (Fig. 5g). Furthermore, the interaction effect is significant ($P = 0.0203$), with only 3% of other brain region pairs showing stronger effects. The significant associations with the same synaptic protein modules suggest that functional connectivity and structural covariation are likely to share molecular processes. As an exploratory analysis, we tested the dendritic spine component of other modules and found no association with structural covariation (Supplementary Fig. 7).

We further repeated the analysis with synaptic expression modules in place of synaptic protein modules as replication. The synaptic modules built from gene expression data (Supplementary Fig. 8 and see SFG eMod5 and ITG eMod4 in Supplementary Table 3) contain genes that highly overlap with members of the synaptic protein modules (odds ratio (OR) = 2.92, $P = 2.71 \times 10^{-32}$ for SFG and OR = 2.53, $P = 5.14 \times 10^{-23}$ for ITG) and are most enriched for neurons among expression modules (Supplementary Fig. 8). Similar to the protein module results, we did

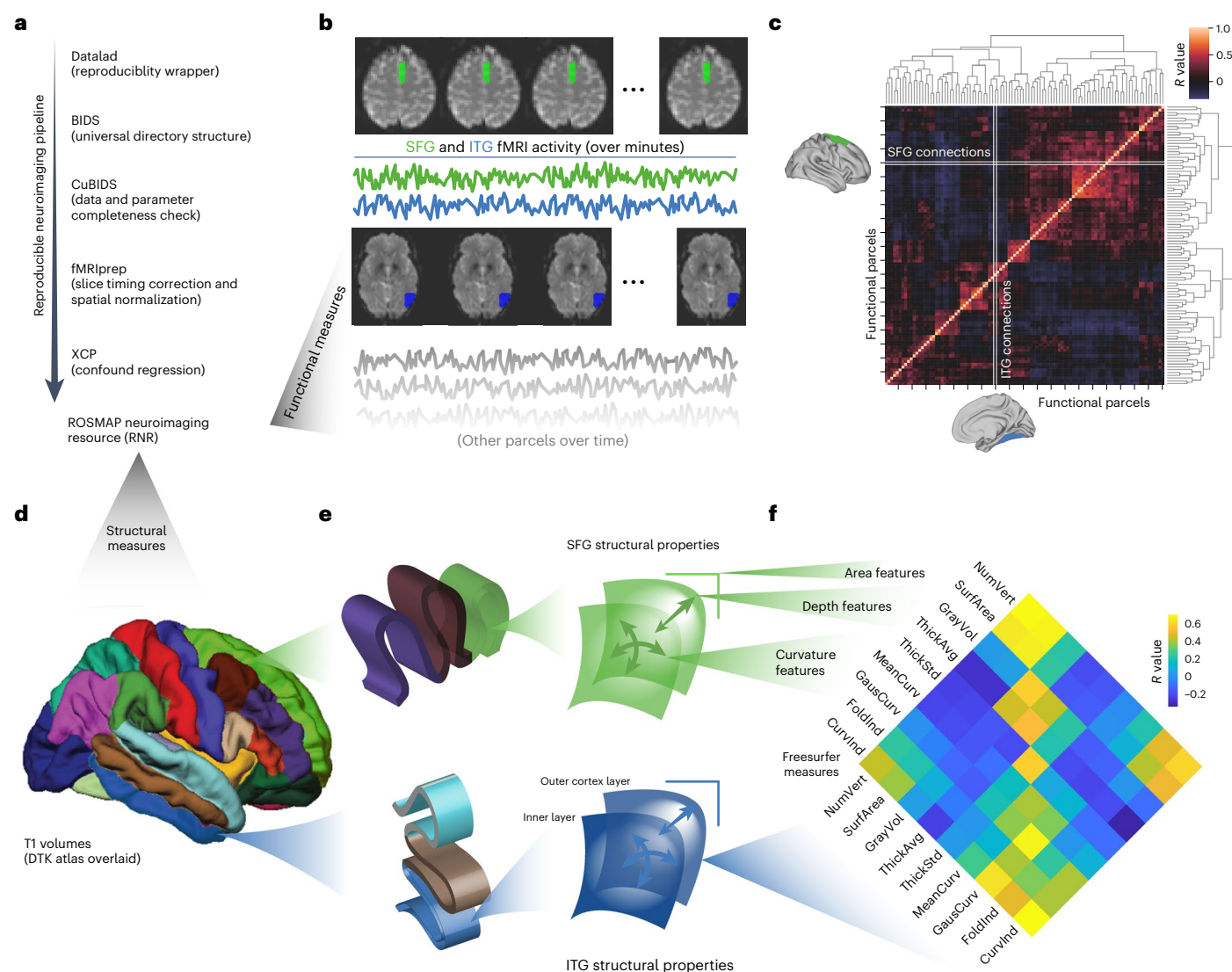


Fig. 2 | Functional and structural neuroimaging phenotypes. **a**, Preprocessed functional and structural MRI data organized into a free resource for data sharing. This process involved curation (BIDS), assessment of aberrant acquisition parameters (CuBIDS), preprocessing (fMRIPrep) and confound regression (XCP). **b**, Mean time series of hemodynamic fluctuations within brain regions from which we also measured proteomics, RNA-seq and dendritic spine

morphology. **c**, Functional connectivity estimated by correlating all pairs of parcel time series (participant average shown), with the SFG–ITG connection being of primary interest. **d**, Structural MRI data were collected during the same scan sessions in which we acquired fMRI data. **e**, Freesurfer was applied to extract structural attributes from anatomically defined brain regions. **f**, Cross-correlation between structural attributes of SFG and ITG.

not find an association between SFG–ITG connectivity and synaptic expression modules ($P = 0.0532$) unless we focus on the dendritic spine component of the synaptic expression modules ($P = 0.0396$). The observed regional specificity with protein abundance modules is also evident with gene expression modules in line with previous transcriptomic studies^{13,43,44}, with only 2% of other connections displaying stronger association (Supplementary Fig. 9a). However, we did not observe an interaction effect ($P = 0.4824$). When we associated SFG–ITG structural covariation with synaptic expression modules, we again did not find an association ($P = 0.9705$) unless we honed into their dendritic spine component ($P = 0.0234$). Regional specificity is lower compared to synaptic protein modules, with 14% of other brain region pairs showing stronger association (Supplementary Fig. 9b), and we did not observe an interaction effect ($P = 0.2538$). As an exploratory analysis, we further tested the other modules and found the dendritic spine component of modules predominantly enriched for immune systems to be also associated with functional connectivity and structural covariation (Supplementary Fig. 7). Overall, these results show that the

expression data replicate the associations between the dendritic spine component of synaptic modules and both functional connectivity and structural covariation, but with weaker interaction effects compared to protein data. More broadly, the results mentioned above indicate the benefits of cellular contextualization by using dendritic spine morphometry to bridge systems at the molecular level to connectivity at the brain region level.

Individual molecules associated with connectivity

The associations presented thus far are at the level of modules, which comprise hundreds of proteins and genes. To prioritize specific proteins among the 7,788 measured in each region, we modeled SFG–ITG connectivity by the main effect of each protein in a given region, the main effect of the synaptic protein module in the other region and the interaction between the given protein and the synaptic protein module. We used this strategy to avoid the limitation on statistical power in testing all pairwise combinations of 7,788 proteins between the two brain regions. We prioritized proteins in each brain region based on the following

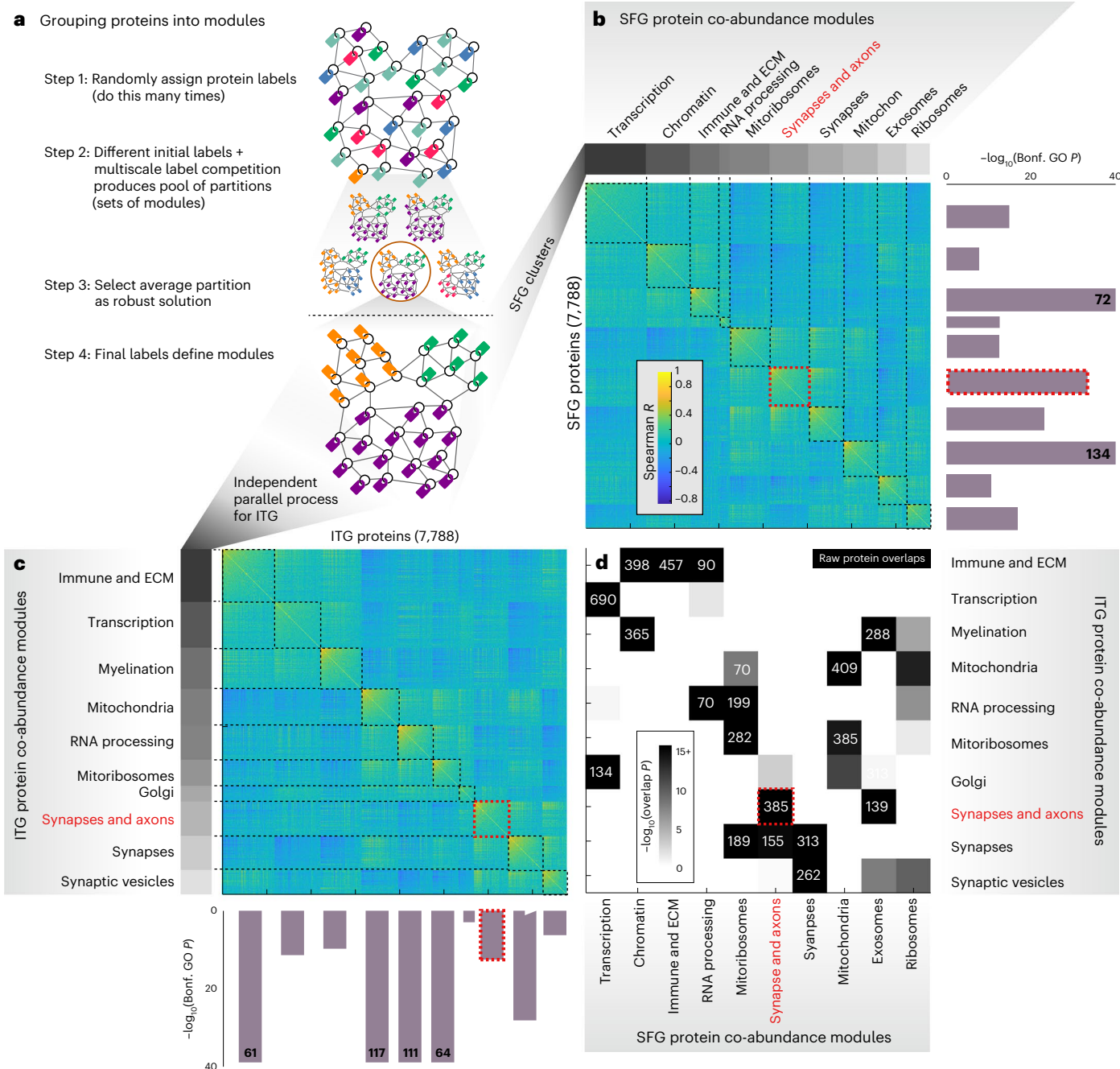


Fig. 3 | Identification of molecular systems in SFG and ITG based on protein abundance. **a**, Schematic representation of how proteins were clustered into co-abundant modules. Modules were inferred solely from proteomics data by applying a consensus clustering algorithm (SpeakEasy2 (ref. 26)) on the corresponding protein–protein correlation matrix, with no reference to ontology. **b**, SFG proteins arranged by their assigned module. Protein pairs within the same module (bounded by dashed squares) display a higher correlation than protein pairs belonging to separate modules. A short functional label was assigned to each module ('transcription', 'chromatin', etc.) based on its more

highly enriched GO terms. The P value of the most enriched GO term of each module is indicated by the corresponding mauve bar (if $-\log_{10}(P) > 40$, bars are truncated with $-\log_{10}(P)$ values stated). Modules were separately generated for SFG and ITG, with the module of interest in terms of brain connectivity highlighted in dashed red lines and red text. **c**, ITG counterpart of **b**. **d**, Most modules in one region have one or more highly overlapping modules in the other region (number of overlapping proteins shown in white text when $-\log_{10}(P) > 14$, enriched for similar GO terms. ECM, extracellular matrix.

three effects: the main effect of each protein, its interaction with the synaptic protein module and these two effects combined. We declared significance at an α of 0.05 with false discovery rate (FDR) correction across all proteins, regions and effects tested. For result interpretation, we applied GSEA³³ to the summary statistics of each effect.

Analogous to the module-level results, we did not find an association between SFG–ITG connectivity and the abundance of any proteins

without cellular contextualization (Supplementary Table 4). Therefore, we fitted each protein with all spine attributes in aggregate to extract the dendritic spine component. For 99% of the proteins, the dendritic spine fits have $R^2 > 0.1$ (Fig. 6a). Grouping the proteins based on the partial R^2 of the spine attributes resulted in clusters that well relate to processes supporting synaptic transmission (Fig. 6b,d). Notably, we observed distinct partial R^2 profiles across the clusters (Fig. 6c),

suggesting that the dominant attributes in the dendritic spine fits vary across proteins. When we repeated the protein-connectivity association analysis with the dendritic spine component, the association strength became stronger (Fig. 7a; $P < 10^{-4}$ for all three effects in both SFG and ITG), and we detected 87 SFG proteins and 344 ITG proteins (Supplementary Table 4). Overlapping these proteins with protein members of clusters derived from dendritic spine fits (Fig. 6b–d) shows that both spine density and morphologies contribute to connectivity-associated proteins. Applying GSEA to the protein-connectivity association statistics (based on the dendritic spine component) found SFG proteins to be predominantly enriched for GO terms related to synapse, energy metabolism and RNA processing (Supplementary Table 5). Detected SFG proteins include limbic system-associated membrane protein (LSAMP)⁴⁵, NAPB⁴⁶, KCNP3 (ref. 47) and PANX1 (ref. 48), which are involved in cell surface channel biology at synapses. Similarly, the protein-connectivity association statistics for ITG are enriched for GO terms related to synapse, energy metabolism and RNA processing (Supplementary Table 5). Detected ITG proteins include LYNX1, NRN1, Pleckstrin and Sec7 domain containing 2 (PSD2) and RAB11A. Like LSAMP, LYNX1 and NRN1 are glycosylphosphatidylinositol (GPI)-anchored proteins that mediate neuronal receptor activity at excitatory synapses^{49,50}. Like NAPB, PSD2 and RAB11A are critical for intracellular signal transduction and protein transport at synapses^{51,52}. These findings indicate that while the enriched GO terms of SFG and ITG overlap (OR = 48.01, $P = 1.77 \times 10^{-46}$; Fig. 7b), their detected proteins are different (Supplementary Fig. 10).

To further probe the relationship between brain function and structure, we repeated the analysis with SFG–ITG structural covariation as the outcome. In contrast to functional connectivity, even without cellular contextualization, we found 72 SFG proteins and 171 ITG proteins associated with SFG–ITG structural covariation (Supplementary Table 4). Nevertheless, focusing on the dendritic spine component increased association strength (Fig. 7c; $P < 10^{-4}$ for all effects tested except for the main effect of ITG proteins) and detected 625 SFG proteins and 1204 ITG proteins (Supplementary Table 4). These proteins overlap with proteins detected without cellular contextualization (OR = 3.61, $P = 3.94 \times 10^{-5}$ for SFG and OR = 4.60, $P = 6.75 \times 10^{-20}$ for ITG). The greater number of proteins associated with structural covariation compared to functional connectivity could be due to how anatomical structures are more stable over time. Hence, the antemortem T1 scans would better reflect the brain states at time of postmortem omic acquisition. The ranking of GO terms is again similar between SFG and ITG (OR = 215.83, $P = 9.25 \times 10^{-31}$; Fig. 7d). Notably, the enriched GO terms are similar to those found with functional connectivity as the outcome (OR = 125.73, $P = 4.68 \times 10^{-26}$ for SFG and OR = 231.90, $P = 2.86 \times 10^{-98}$ for ITG; Supplementary Table 5), with 18 SFG proteins and 32 ITG proteins in common (Fig. 7e). These findings provide validity for our detected protein-connectivity associations and indicate that certain molecular processes can explain variance in both functional connectivity and structural covariation.

Next, we repeated the analysis with gene expression data, and the overall trend was similar to the results obtained with the protein abundance data. We did not find any genes associated with SFG–ITG connectivity without cellular contextualization (Supplementary Table 6). When we focused on the dendritic spine component of gene expression, the association strength became higher ($P < 10^{-4}$ for all

tested effects except for the main effect of SFG genes). We identified 1 SFG gene and 324 ITG genes. The enriched GO terms are predominantly related to synapses, energy metabolism and RNA processing (Supplementary Table 7), similar to those found with protein abundance (OR = 29.67, $P = 1.58 \times 10^{-42}$ for SFG and OR = 6.61, $P = 2.56 \times 10^{-20}$ for ITG). However, the ranking of individual genes by association strength does not align with that of individual proteins (Supplementary Fig. 11). When we used structural covariation as the outcome, we found an association with 130 SFG genes and 18 ITG genes without cellular contextualization. Nevertheless, we detected more genes (1,023 SFG genes and 248 ITG genes) and observed higher association strength ($P < 10^{-4}$ for all tested effects except for the main effect of SFG genes with $P = 0.0178$) when we focused on the dendritic spine component of gene expression. The enriched GO terms for gene expression overlap with those found for protein abundance (OR = 66.57, $P = 7.29 \times 10^{-12}$ for SFG and OR = 164.97, $P = 2.66 \times 10^{-26}$ for ITG), but the ranking of individual genes and individual proteins do not align (Supplementary Fig. 12). Functional connectivity and structural covariation share common GO terms (OR = 68.80, $P = 1.46 \times 10^{-45}$ for SFG and OR = 13.18, $P = 6.91 \times 10^{-10}$ for ITG), but with little overlap at the gene level (Supplementary Fig. 13). Overall, our results show consistency between protein abundance and gene expression at the molecular function level, despite the lack of overlap between individual proteins and genes. This is likely due to the low correlation between mRNA level and protein abundance of the same gene⁵³.

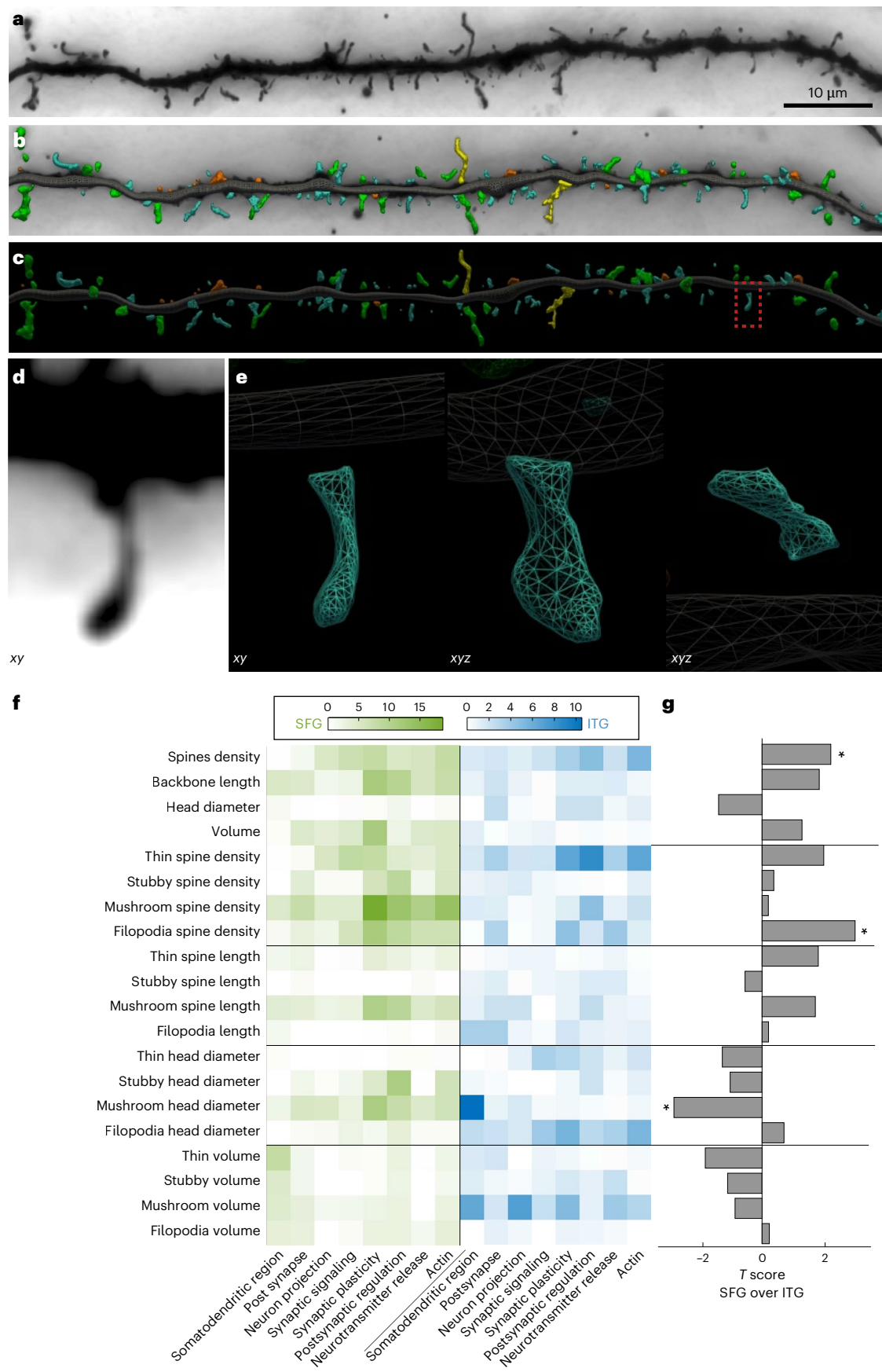
Imaging transcriptomic and genetic support for connectivity molecules

To place our connectivity-associated proteins and genes in context with other imaging transcriptomic studies, we compared with genes previously found by combining the AHBA with fMRI data of other cohorts⁵⁴. The 125 AHBA connectivity-related genes do not place highly among our proteins ranked by their effects on connectivity (Supplementary Fig. 14), possibly due to post-transcriptional regulation, which limits the correlation between mRNA level and protein abundance of the same gene⁵³. Those genes are only enriched among ITG genes ranked by their interaction effects ($P = 0.01$, based on GSEA). This result suggests that the generic set of genes found by correlating spatial patterns of gene expression in AHBA with functional connectivity estimated from other cohorts has only mild overlap with genes found by modeling covariation between gene expression and SFG–ITG connectivity within the same set of individuals.

Finally, we compared the connectivity-associated proteins and genes found herein with previous imaging genetics studies. The largest related GWAS⁶ found four independent loci (*EPHA3*, *DPP4*, *FBXO11* and *ZNF326*) that are associated with SFG–ITG connectivity (rfMRI connectivity ICA100 edge 849). These loci were derived by spatially mapping genome-wide significant genetic variants to their closest genes. Among the four loci, *EPHA3* shows the highest heritability in the edge 849 GWAS, and the dendritic spine component of *EPHA3* expression displays nominal association with SFG–ITG connectivity in our data ($P = 0.0192$ for SFG main effect and $P = 0.0140$ for ITG main effect). We could not perform the same analysis for protein abundance because tryptic peptides from *EPHA3*, *DPP4*, *FBXO11* and *ZNF326* were not measured in our dataset. Instead, we assessed our found proteins in relation to the edge 849 GWAS through a functional mapping approach. For

Fig. 4 | Measurement of dendritic spines with protein-based functional characterization. **a**, Representative $\times 60$ bright-field image of a Golgi-stained dendrite from ITG of an exemplar participant. Scale bar = 10 μ m. **b**, Digital 3D reconstruction of the dendritic segment performed on the bright-field image. **c**, Digital 3D reconstruction used for estimating spine density and morphometric attributes, including head diameter, length and volume, and assigning subclasses. Blue indicates thin spines, green indicates mushroom spines, red indicates stubby spines and yellow indicates filopodia. **d**, Representative

zoomed-in bright-field image of a single Golgi-impregnated thin spine in the xy plane (red box from **c**). **e**, Left to right, 3D digital reconstruction of the dendrite (gray) and spine (green) in the xy plane, clockwise rotation in xyz dimensions and further rotation in xyz. **f**, Each spine attribute was associated against all proteins measured from the same region with enriched GO terms indicated (SFG in green and ITG in blue). **g**, T values of contrasts between SFG and ITG for each spine morphologic attribute are shown, with * indicating nominal differences.



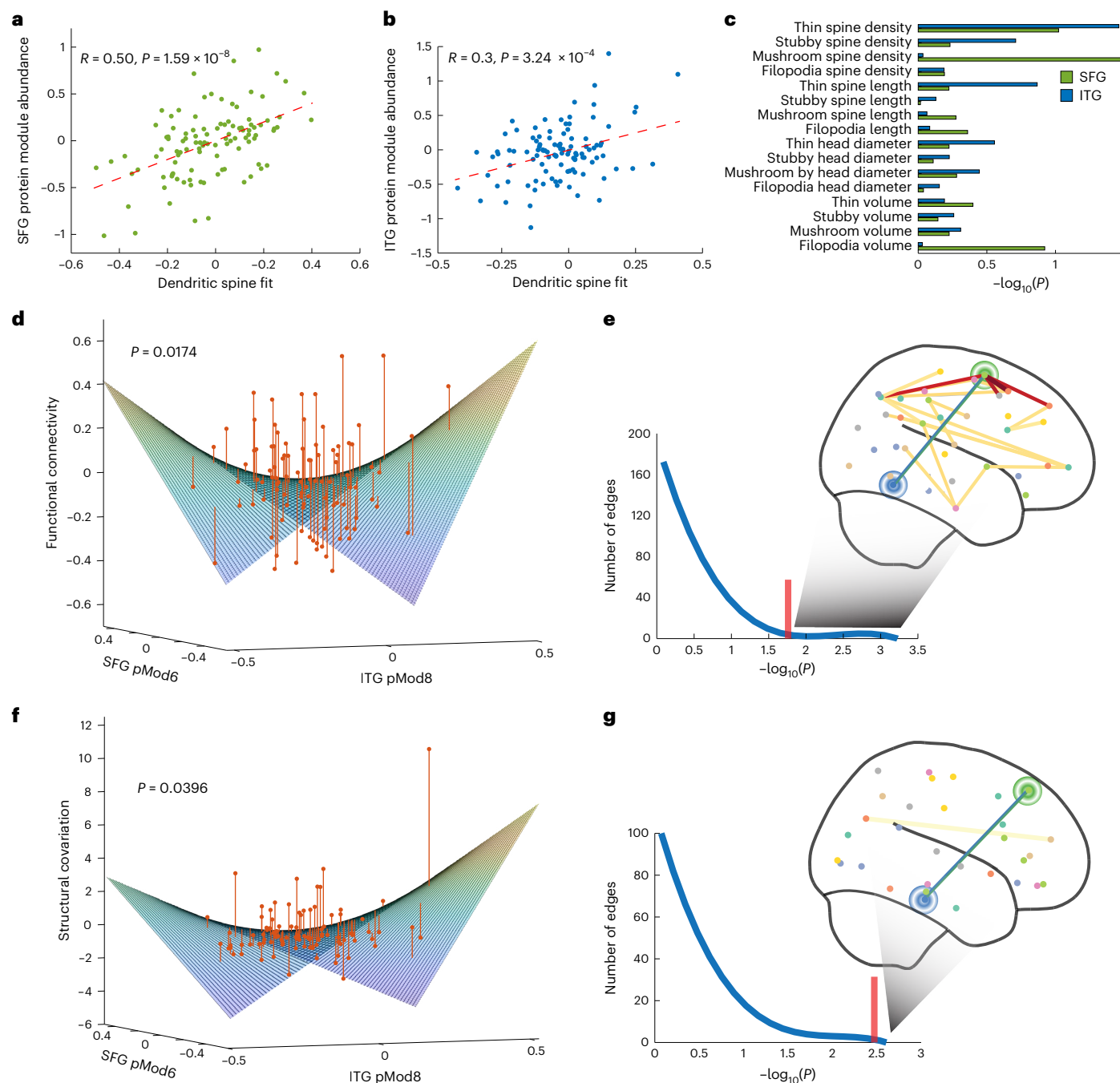


Fig. 5 | Associations between protein modules and brain connectivity.

a, Protein abundance of SFG synaptic module plotted against its dendritic spine fit. Each dot corresponds to an individual and the red dashed line corresponds to the linear fit between protein abundance and the dendritic spine fit values of the individuals. **b**, ITG counterpart of **a**. **c**, Partial contribution of each attribute toward the dendritic spine fit. **d**, Functional connectivity (with confounds regressed out) fitted by the dendritic spine component of SFG and ITG synaptic protein modules. Each red dot corresponds to the functional connectivity value of an individual with the red line visualizing how far it is from the fitted surface. The large curvature indicates a strong interaction effect of the modules

on functional connectivity. **e**, SFG–ITG connection and other connections with stronger association strength between functional connectivity and the dendritic spine component of synaptic protein modules displayed. Different color dots correspond to different brain regions. In total, 3% of other connections showed higher association strength. Connections emanating from SFG or ITG are displayed in red. Other connections are displayed in yellow. **f**, Structural covariation counterpart of **d**. **g**, Structural covariation counterpart of **e**. In total, 0.4% of other connections showed higher association strength between structural covariation and the dendritic spine component of synaptic protein modules.

each brain region, we built protein prediction models to combine the effects of genetic variants that are proximal to each protein⁵⁵. We then applied these models to the summary statistics of the edge 849 GWAS to test whether the genetic component of the proteins is associated with edge 849. Overall, the genetic component of both SFG and ITG proteins was not associated with edge 849 after FDR correction (Supplementary

Table 8), likely due to limited heritability in edge 849. Accordingly, we focused on the genetic effects of our found proteins. For SFG, 3,339 proteins have adequate heritability for model construction. Among these 3,339 proteins, the dendritic spine component of 33 proteins is associated with SFG–ITG connectivity in our data (FDR-corrected $P < 0.05$). Among those proteins, the genetic component of two proteins, NGB

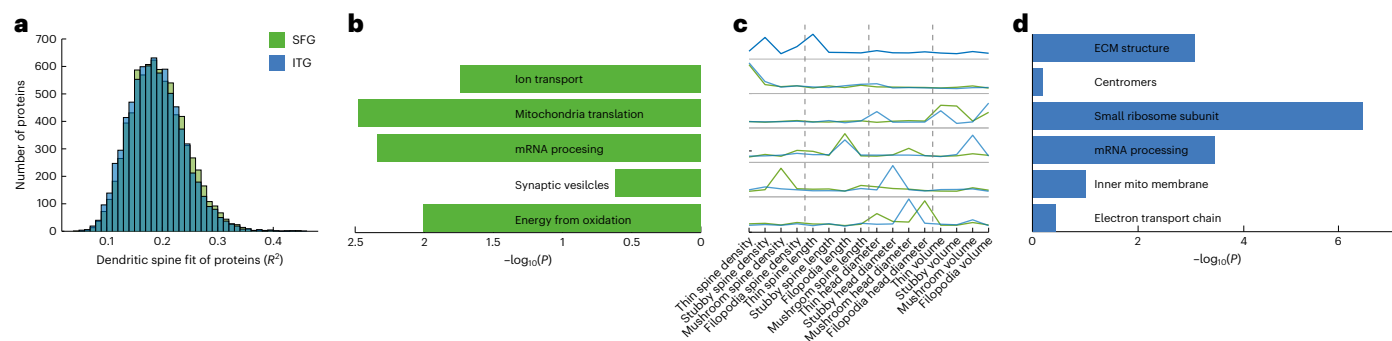


Fig. 6 | Characterization of dendritic spine fits of proteins. a, R^2 of the dendritic spine fits of all measured proteins. **b**, SFG proteins clustered by applying SpeakEasy2 to the partial R^2 of the spine attributes. Each functional label was assigned based on the most enriched GO terms of each cluster. The

P values correspond to the overlap between functional connectivity-associated proteins and protein members of each cluster. **c**, Partial R^2 of each dendritic spine attribute averaged over proteins of each cluster. Partial R^2 profiles of SFG (green) and ITG (blue) were aligned using Hungarian clustering. **d**, ITG counterpart of **b**.

($P = 0.0219$) and POLR2J ($P = 0.0423$), display nominal associations with edge 849. For the ITG region, 3,445 proteins have adequate heritability for model construction. Among these 3,445 proteins, the dendritic spine component of 147 proteins is associated with SFG–ITG connectivity in our data (FDR-corrected $P < 0.05$). Among those proteins, the genetic component of 11 proteins, including NRN1 ($P = 0.0456$), displays nominal associations with edge 849. These results provide further support for some of the protein-connectivity associations found in our data. The repeated detection of NRN1 (ref. 56), or neuritin, at the DNA and protein level, suggests that this secreted neuropeptide has a critical role in synaptic biology in supporting brain connectivity of ITG.

Discussion

A central goal of neuroscience is to develop an understanding of the brain that ultimately describes the mechanistic basis of human cognition and behavior. Independent studies are contributing to an exquisite ‘parts list’ of molecules, cell types and brain structures^{57–59}, but how these parts cohere into human cognitive function remains obscure. This research challenge is analogous to assembling a jigsaw puzzle without seeing the image on the box. Research struggles with proto-clusters of insightful pieces that are difficult to orient with respect to each other, especially when these pieces reside on different biophysical scales. To help bridge this gap, we gathered and analyzed brain omics, and cellular and neuroimaging data from the same set of 98 individuals. Focusing on the connection between SFG and ITG from which we acquired omic and cellular data, we showed that molecular measurements likely capture many properties beyond brain connectivity. Hence, only when we restricted to their components related to dendritic spine morphology were we able to establish associations with SFG–ITG connectivity. With this approach, we found molecules and molecular modules enriched for synaptic structures and functions, mitochondria-based energy metabolism and RNA processing, in line with known biology. Notably, beyond listing connectivity-related molecules, our study has broader implications in that it demonstrates the feasibility of detecting synchrony among systems of different scales in humans, which constitutes a step toward a more coherent understanding of brain function. Practically, this integration across scales could help address the long-standing challenge of brain drug discovery, in which promising molecules from cellular assays fail to influence cognitive traits.

As our study is distinctive and there are no other human datasets containing such paired brain omics and imaging data, we used consistency among data types that span different biophysical scales as a way to validate any observed synchronization. We showed that the same synaptic protein modules can explain between-participant variability in both functional connectivity and structural covariation and replicated this finding with synaptic expression modules. We also found hundreds of proteins associated with functional connectivity,

a subset of which is associated with structural covariation, and again we showed similar trends with gene expression data. Notably, all found associations are enriched for similar cellular structures and molecular processes. We further showed that the genetic component of a dozen found genes is associated with connectivity, suggesting potential causal roles of these genes. As a common consideration with fMRI is the stability of its detected effects^{60,61}, we accordingly performed all data preprocessing and normalization in a highly reproducible and stringent fashion and used the more stable structural neuroimaging data to replicate our findings.

Dendritic spines are tiny protrusions along neuronal dendrites that participate in most excitatory synapses in the brain. The head of the spine structurally supports the postsynaptic density, which collects the essential machinery for postsynaptic neurotransmission^{36,62}. These actin-rich structures are plastic and alter their shape in congruence with synaptic activity and plasticity, including long-term potentiation³⁰. These facets of spine biology were recapitulated in our GSEA of spine-protein associations (Fig. 4f), which linked actin-related and postsynaptic elements to spine morphological attributes, matching known neurobiology. Also, the enrichment for synaptic signaling and neurotransmitter release is consistent with the hypothesis that multiple aspects of spine morphology have functional consequences. The dominant spine attributes contributing to the dendritic spine fits varied across the connectivity-related proteins (Fig. 6c). Notably, spine density, which reflects the capacity for excitatory neurotransmission and hence presumably links to connectivity, is not the only dominant attribute, but spine morphological attributes also dominate for certain proteins. Although our found associations between molecules and connectivity do not describe or depend on mesoscale mechanisms between spines and fMRI signals, one major theory of fMRI generation relates to local field potentials (LFP)⁶³. LFPs are generated from synaptic transmission, which uses dendritic spines, so their involvement in large-scale signal generation is logical and plausible.

Applying GSEA to the protein-connectivity associations found enrichment for synaptic function, mitochondria-based energy metabolism and RNA biology (Figs. 6b–d and 7e). The enriched processes are consistent with normal synaptic function, which requires energy and local RNA translation. For instance, releasing and recycling synaptic vesicles at the axonal presynapse are energy-demanding mechanisms that are mediated by Ca^{2+} levels. Mitochondria localized to presynaptic terminals are positioned to buffer Ca^{2+} through the mitochondrial matrix and meet metabolic energy needs by producing adenosine triphosphate⁶⁴. In contrast, postsynaptic mitochondria are typically positioned in the dendrite rather than localized inside the dendritic spine⁶⁵. While the function of postsynaptic mitochondria is less well characterized, these organelles are commonly hypothesized to support the energy needs of multiple spines in proximity to the mitochondrial

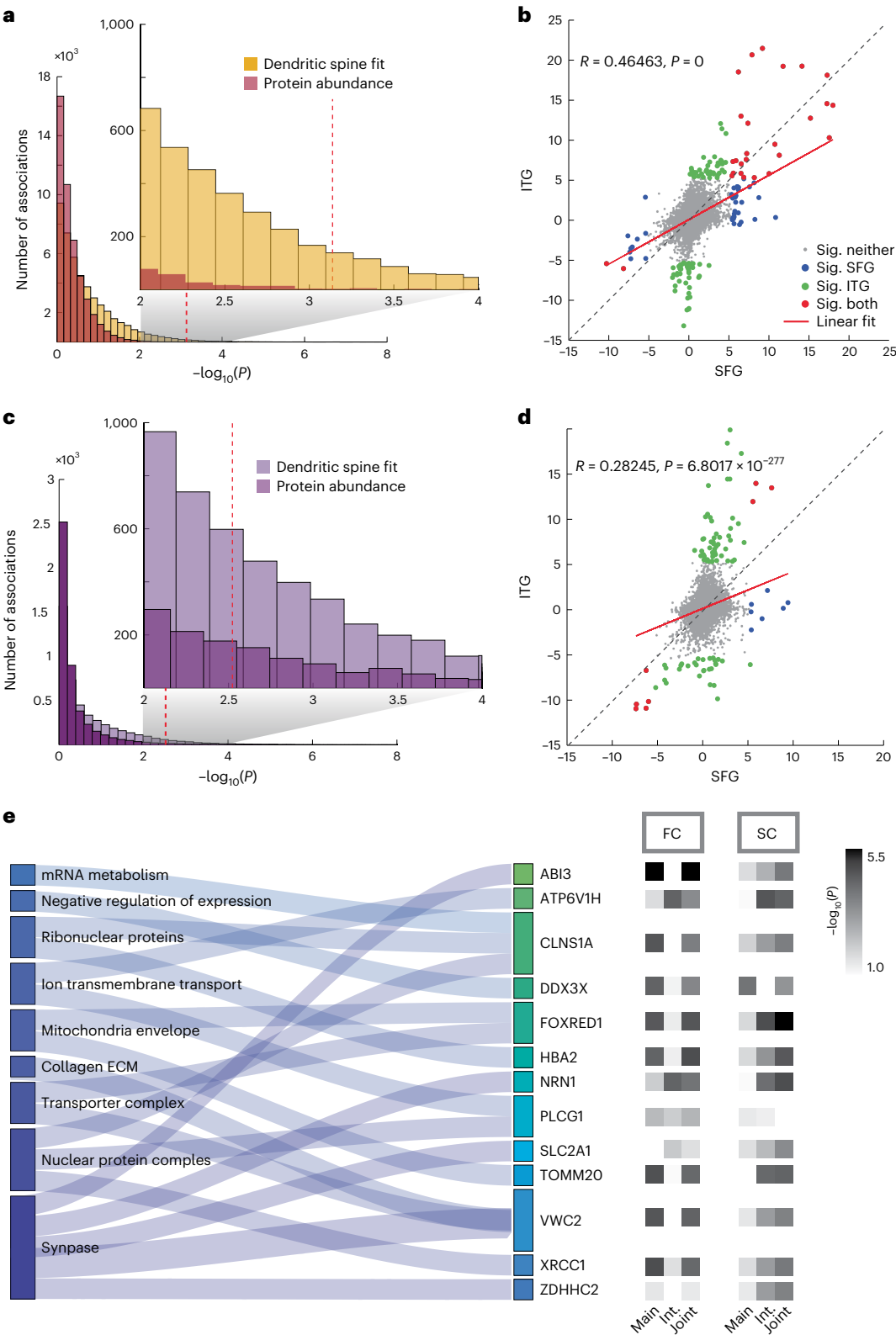


Fig. 7 | Protein-connectivity associations. **a**, Histogram on the number of associations found between proteins or their dendritic spine component and functional connectivity, aggregated over regions and effects tested. Red dotted line corresponds to the FDR threshold for the dendritic spine component. **b**, GO term enrichment based on applying GSEA to protein-connectivity association statistics. Displayed scatterplot corresponds to interaction effects. Ranking of

GO terms is similar between SFG and ITG. **c**, Structural covariation counterpart of **a**. **d**, Structural covariation counterpart of **b**. **e**, Sankey diagram displaying enriched GO terms common between functional connectivity (FC) and structural covariation (SC) as outcome, and proteins associated with both FC and SC (in terms of main effect, interaction or joint of main and interaction effect) that overlap with members of the GO terms.

location in the dendrite⁶⁶. Furthermore, synaptic plasticity partly depends on local protein synthesis, in which mRNA-binding proteins have a large role, especially localization and translation of mRNA following synaptic demand⁶⁷.

Our analysis revealed several GPI-anchored proteins that are functionally similar and localized to the synapse, including LSAMP, LYNX1 and NRN1 (Supplementary Table 4). LSAMP is expressed widely through the adult brain, and human genetics studies as well as work in rodent models suggest that LSAMP influences emotional behavior, possibly contributing to mood disorders⁶⁸. LSAMP is predicted to facilitate axon targeting to dendrites⁶⁹ and is considered a putative therapeutic target for neuropsychiatric illness⁴⁵. Similarly, LYNX1 localizes to synapses through its GPI anchor and interaction with nicotinic acetylcholine receptors. LYNX1 is hypothesized to maintain a balance between excitatory and inhibitory circuits in the mouse visual cortex of the adult brain⁷⁰. Like the others, NRN1, or candidate plasticity gene 15 (*CPG15*), is a synaptic activity-regulated gene whose expression is experience-dependent in the adult brain. NRN1 anchors to the surface of synaptic structures, regulating the generation of dendritic spines and synapse maturation⁷¹. Recent findings in older adults indicated a beneficial role for NRN1 in synaptic preservation and maintenance^{72,73}. NRN1 is a neurotrophic factor, and given the small size of its secreted form (~11 kDa), may be an attractive therapeutic target. Another set of functionally overlapping proteins between SFG and ITG converge at intracellular signaling pathways within dendritic spines. NABP, or N-ethyl-maleimide-sensitive fusion attachment protein β , interacts with α -amino-3-hydroxy-5-methyl-4-isoxazole-propionate (AMPA) receptors and functions as a molecular chaperone to facilitate processing of AMPA receptors at the postsynapse⁴⁶. PSD2, also known as the exchange factor for ARF6 (EFA6) isoform C, binds phospholipids and is localized to endo-lysosomal compartments at the postsynapse⁵². Past studies have identified PSD2 as a candidate risk gene for age-related memory disorders^{74,75}. While all these proteins are involved in connectivity-relevant processes, additional studies are required to mechanistically investigate how fluctuations in these protein abundances explain interindividual differences in functional connectivity.

Our findings suggest a number of important factors to consider when trying to draw associations between postmortem microscale molecular data and macroscale neuroimaging data. First, our results indicate that the relationships between molecular abundances and brain connectivity have substantial regional specificity. Henceforth, measuring molecular levels and spine attributes in additional brain regions from the same individuals could help define the extent of this regional specificity in determining if different proteins are indeed involved in different brain regions. Second, the number of associations found might be limited by the time interval between the antemortem scans and postmortem brain collection. From this standpoint, the molecules we detected herein could be considered a minimum set and reflective of molecular abundances whose covariation with these brain regions occurs over a particular time scale. Finally, we face the common problem of assigning molecules from bulk molecular measurements to particular types of cells or synapses. For instance, RNA-related or mitochondrial proteins we discuss in the context of spines may be from nonsynaptic areas of the neuron or nonneuronal cells. Moreover, if the found proteins are indeed from spines or synapses, they may not be globally representative, as recent efforts suggest unique molecular signatures in different synapses⁷⁶. Therefore, in addition to the need to explore regional specificity, it could be helpful to assay synapse-type specificity to investigate whether particular types of synapses are more relevant to fMRI coupling.

Overall, this study indicates that acquiring data across the major perspectives in human neuroscience from the same set of brains is foundational for understanding how human brain function is supported at multiple biophysical scales. While future research is necessary for fully determining the scope and components of multiscale brain synchrony,

we have established a robustly defined initial set of molecules whose effects likely resonate across biophysical scales.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41593-024-01788-z>.

References

1. Franzmeier, N. et al. The BDNF Val66Met SNP modulates the association between β -amyloid and hippocampal disconnection in Alzheimer's disease. *Mol. Psychiatry* **26**, 614–628 (2021).
2. Cao, H., Zhou, H. & Cannon, T. D. Functional connectome-wide associations of schizophrenia polygenic risk. *Mol. Psychiatry* **26**, 2553–2561 (2021).
3. Chen, J. et al. Genome-transcriptome-functional connectivity-cognition link differentiates schizophrenia from bipolar disorder. *Schizophr. Bull.* **48**, 1306–1317 (2022).
4. Murphy, S. E. et al. The effect of the serotonin transporter polymorphism (5-HTTLPR) on amygdala function: a meta-analysis. *Mol. Psychiatry* **18**, 512–520 (2013).
5. Cao, H. et al. The 5-HTTLPR polymorphism affects network-based functional connectivity in the visual-limbic system in healthy adults. *Neuropsychopharmacology* **43**, 406–414 (2018).
6. Elliott, L. T. et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210–216 (2018).
7. Marek, S. et al. Reproducible brain-wide association studies require thousands of individuals. *Nature* **603**, 654–660 (2022).
8. Richiardi, J. et al. BRAIN NETWORKS. Correlated gene expression supports synchronous activity in brain networks. *Science* **348**, 1241–1244 (2015).
9. Anderson, K. M. et al. Convergent molecular, cellular, and cortical neuroimaging signatures of major depressive disorder. *Proc. Natl Acad. Sci. USA* **117**, 25138–25149 (2020).
10. Rittman, T. et al. Regional expression of the *MAPT* gene is associated with loss of hubs in brain networks and cognitive impairment in Parkinson disease and progressive supranuclear palsy. *Neurobiol. Aging* **48**, 153–160 (2016).
11. Seidlitz, J. et al. Transcriptomic and cellular decoding of regional brain vulnerability to neurogenetic disorders. *Nat. Commun.* **11**, 3358 (2020).
12. Vértes, P. E. et al. Gene transcription profiles associated with inter-modular hubs and connection distance in human functional magnetic resonance imaging networks. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20150362 (2016).
13. Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
14. Gaiteri, C. et al. Gene expression and DNA methylation are extensively coordinated with MRI-based brain microstructural characteristics. *Brain Imaging Behav.* **13**, 963–972 (2019).
15. Gratton, C. et al. Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron* **98**, 439–452 (2018).
16. Bennett, D. A. et al. Religious orders study and rush memory and aging project. *J. Alzheimers Dis.* **64**, S161–S189 (2018).
17. Gorgolewski, K. J. et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* **3**, 160044 (2016).
18. Covitz, S. et al. Curation of BIDS (CuBIDS): a workflow and software package for streamlining reproducible curation of large BIDS datasets. *Neuroimage* **263**, 119609 (2022).
19. Esteban, O. et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).

20. Schaefer, A. et al. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
21. Alexander, B. et al. Desikan–Killiany–Tourville Atlas compatible version of M-CRIB neonatal parcellated whole brain atlas: the M-CRIB 2.0. *Front. Neurosci.* **13**, 34 (2019).
22. Fischl, B., Sereno, M. I. & Dale, A. M. Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. *Neuroimage* **9**, 195–207 (1999).
23. Johnson, E. C. B. et al. Large-scale deep multi-layer analysis of Alzheimer's disease brain reveals strong proteomic disease-related changes not observed at the RNA level. *Nat. Neurosci.* **25**, 213–225 (2022).
24. Ping, L. et al. Global quantitative analysis of the human brain proteome in Alzheimer's and Parkinson's disease. *Sci. Data* **5**, 180036 (2018).
25. Ping, L. et al. Global quantitative analysis of the human brain proteome and phosphoproteome in Alzheimer's disease. *Sci. Data* **7**, 315 (2020).
26. Gaiteri, C. et al. Robust, scalable, and informative clustering for diverse biological networks. *Genome Biol.* **24**, 228 (2023).
27. Gaiteri, C., Ding, Y., French, B., Tseng, G. C. & Sibille, E. Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. *Genes Brain Behav.* **13**, 13–24 (2014).
28. Mostafavi, S. et al. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat. Neurosci.* **21**, 811–819 (2018).
29. Zhang, B. et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**, 707–720 (2013).
30. Kasai, H., Ziv, N. E., Okazaki, H., Yagishita, S. & Toyozumi, T. Spine dynamics in the brain, mental disorders and artificial neural networks. *Nat. Rev. Neurosci.* **22**, 407–422 (2021).
31. Matsuzaki, M. et al. Dendritic spine geometry is critical for AMPA receptor expression in hippocampal CA1 pyramidal neurons. *Nat. Neurosci.* **4**, 1086–1092 (2001).
32. Peters, A. & Kaiserman-Abramof, I. R. The small pyramidal neuron of the rat cerebral cortex. The perikaryon, dendrites and spines. *Am. J. Anat.* **127**, 321–355 (1970).
33. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
34. Walker, C. K. et al. Cross-platform synaptic network analysis of human entorhinal cortex identifies TWF2 as a modulator of dendritic spine length. *J. Neurosci.* **43**, 3764–3785 (2023).
35. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, Article17 (2005).
36. Hayashi, Y. & Majewska, A. K. Dendritic spine geometry: functional implication and regulation. *Neuron* **46**, 529–532 (2005).
37. Distler, U. et al. Proteomic analysis of brain region and sex-specific synaptic protein expression in the adult mouse brain. *Cells* **9**, 313 (2020).
38. Lautz, J. D. et al. Synaptic protein interaction networks encode experience by assuming stimulus-specific and brain-region-specific states. *Cell Rep.* **37**, 110076 (2021).
39. Pan, Y. & Monje, M. Activity shapes neural circuit form and function: a historical perspective. *J. Neurosci.* **40**, 944–954 (2020).
40. Sallet, J. et al. Social network size affects neural circuits in macaques. *Science* **334**, 697–700 (2011).
41. He, H. et al. Reduction in gray matter of cerebellum in schizophrenia and its influence on static and dynamic connectivity. *Hum. Brain Mapp.* **40**, 517–528 (2019).
42. Koch, K. et al. Functional connectivity and grey matter volume of the striatum in schizophrenia. *Br. J. Psychiatry* **205**, 204–213 (2014).
43. Zhou, Z., Zhu, Y., Liu, Y. & Yin, Y. Comprehensive transcriptomic analysis indicates brain regional specific alterations in type 2 diabetes. *Aging* **11**, 6398–6421 (2019).
44. Hang, Y. et al. Exploration into biomarker potential of region-specific brain gene co-expression networks. *Sci. Rep.* **10**, 17089 (2020).
45. Innos, J., Koido, K., Philips, M.-A. & Vasar, E. Limbic system associated membrane protein as a potential target for neuropsychiatric disorders. *Front. Pharmacol.* **4**, 32 (2013).
46. Osten, P. et al. The AMPA receptor GluR2 C terminus can mediate a reversible, ATP-dependent interaction with NSF and α - and β -SNAPs. *Neuron* **21**, 99–110 (1998).
47. An, W. F. et al. Modulation of A-type potassium channels by a family of calcium sensors. *Nature* **403**, 553–556 (2000).
48. Whyte-Fagundes, P. & Zoidl, G. Mechanisms of pannexin1 channel gating and regulation. *Biochim. Biophys. Acta Biomembr.* **1860**, 65–71 (2018).
49. Miwa, J. M. et al. lynx1, an endogenous toxin-like modulator of nicotinic acetylcholine receptors in the mammalian CNS. *Neuron* **23**, 105–114 (1999).
50. Naeve, G. S. et al. Neuritin: a gene induced by neural activity and neurotrophins that promotes neuritogenesis. *Proc. Natl Acad. Sci. USA* **94**, 2648–2653 (1997).
51. Kelly, E. E., Horgan, C. P., McCaffrey, M. W. & Young, P. The role of endosomal-recycling in long-term potentiation. *Cell. Mol. Life Sci.* **68**, 185–194 (2011).
52. Matsuya, S. et al. Cellular and subcellular localization of EFA6C, a third member of the EFA6 family, in adult mouse Purkinje cells. *J. Neurochem.* **93**, 674–685 (2005).
53. Koussounadis, A., Langdon, S. P., Um, I. H., Harrison, D. J. & Smith, V. A. Relationship between differentially expressed mRNA and mRNA–protein correlations in a xenograft model system. *Sci. Rep.* **5**, 10775 (2015).
54. Zhang, X. et al. Dissect relationships between gene co-expression and functional connectivity in human brain. *Front. Neurosci.* **15**, 797849 (2021).
55. Barbeira, A. N. et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).
56. Yao, J.-J., Zhao, Q.-R., Lu, J.-M. & Mei, Y.-A. Functions and the related signaling pathways of the neurotrophic factor neuritin. *Acta Pharmacol. Sin.* **39**, 1414–1420 (2018).
57. Siletti, K. et al. Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046 (2023).
58. Mathys, H. et al. Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* **186**, 4365–4385 (2023).
59. Elam, J. S. et al. The human connectome project: a retrospective. *Neuroimage* **244**, 118543 (2021).
60. Botvinik-Nezer, R. et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* **582**, 84–88 (2020).
61. Ciric, R. et al. Mitigating head motion artifact in functional connectivity MRI. *Nat. Protoc.* **13**, 2801–2826 (2018).
62. Sheng, M. & Kim, E. The postsynaptic organization of synapses. *Cold Spring Harb. Perspect. Biol.* **3**, a005678 (2011).
63. Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150–157 (2001).
64. Devine, M. J. & Kittler, J. T. Mitochondria at the neuronal presynapse in health and disease. *Nat. Rev. Neurosci.* **19**, 63–80 (2018).
65. Turner, N. L. et al. Reconstruction of neocortex: organelles, compartments, cells, circuits, and activity. *Cell* **185**, 1082–1100 (2022).

66. Thomas, C. I., Ryan, M. A., Kamasawa, N. & Scholl, B. Postsynaptic mitochondria are positioned to support functional diversity of dendritic spines. *eLife* **12**, RP89682 (2023).
67. Liu-Yesucevitz, L. et al. Local RNA translation at the synapse and in disease. *J. Neurosci.* **31**, 16086–16093 (2011).
68. Koido, K. et al. Associations between *LSAMP* gene polymorphisms and major depressive disorder and panic disorder. *Transl. Psychiatry* **2**, e152 (2012).
69. Pimenta, A. F. et al. The limbic system-associated membrane protein is an Ig superfamily member that mediates selective neuronal growth and axon targeting. *Neuron* **15**, 287–297 (1995).
70. Morishita, H., Miwa, J. M., Heintz, N. & Hensch, T. K. Lynx1, a cholinergic brake, limits plasticity in adult visual cortex. *Science* **330**, 1238–1240 (2010).
71. Subramanian, J., Michel, K., Benoit, M. & Nedivi, E. CPG15/neuritin mimics experience in selecting excitatory synapses for stabilization by facilitating PSD95 recruitment. *Cell Rep.* **28**, 1584–1595 (2019).
72. Hurst, C. et al. Integrated proteomics to understand the role of neuritin (NRN1) as a mediator of cognitive resilience to Alzheimer's disease. *Mol. Cell. Proteom.* **22**, 100542 (2023).
73. Yu, L. et al. Cortical proteins associated with cognitive resilience in community-dwelling older persons. *JAMA Psychiatry* **77**, 1172–1180 (2020).
74. Kunkle, B. W. et al. Early-onset Alzheimer disease and candidate risk genes involved in endolysosomal transport. *JAMA Neurol.* **74**, 1113–1122 (2017).
75. Yu, L. et al. Association between brain gene expression, DNA methylation, and alteration of ex vivo magnetic resonance imaging transverse relaxation in late-life cognitive decline. *JAMA Neurol.* **74**, 1473–1480 (2017).
76. Niu, M. et al. Droplet-based transcriptome profiling of individual synapses. *Nat. Biotechnol.* **41**, 1332–1344 (2023).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

¹Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA. ²Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY, USA. ³Department of Neurology, Center for Neurodegeneration and Experimental Therapeutics, University of Alabama at Birmingham, Birmingham, AL, USA. ⁴Penn/CHOP Lifespan Brain Institute, University of Pennsylvania, Philadelphia, PA, USA. ⁵Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA. ⁶Penn Lifespan Informatics and Neuroimaging Center, University of Pennsylvania, Philadelphia, PA, USA. ⁷Department of Child and Adolescent Psychiatry and Behavioral Science, The Children's Hospital of Philadelphia, Philadelphia, PA, USA. ⁸Department of Biochemistry, Emory University School of Medicine, Atlanta, GA, USA. ⁹Department of Clinical Science, Malmö, SciLifeLab, Lund University, Lund, Sweden. ✉e-mail: gaiteri@gmail.com; jhersko@uab.edu

Methods

Cohort

Ninety-eight participants of the ROSMAP¹⁶ have all data types needed for this work, namely fMRI, structural MRI, genetic, dendritic spine morphometry, RNA-seq and proteomic data. All enrolled participants agreed to annual clinical evaluation and brain donation at death. Both studies were approved by an Institutional Review Board of Rush University Medical Center. Participants signed informed consent, an Anatomical Gift Act and a repository consent to allow their resources to be shared. The dendritic spine morphometry, RNA-seq and proteomic data were measured from postmortem tissue samples of the SFG and ITG. For each participant, tissue samples were drawn from either the left or the right hemisphere but not both sides. In total, 52% of participants had tissue samples drawn from the left hemisphere. The average age of participants at the time of MRI scan and death is 88 ± 6 years and 91 ± 6 years, respectively, with an average time interval between MRI scan and age at death of 3 ± 2 years. The average postmortem interval is 8.5 ± 4.6 h. In total, 77% of the participants are female, and the participants have 15 ± 3 years of education. Further demographic details on the participants were previously described¹⁶.

Neuroimaging

Structural scans were acquired using T1-weighted MRI (1.5T General Electric (GE)—1 mm³ resolution, repetition time (TR) = 6.3 ms, echo time (TE) = 2.8 ms, flip angle = 8°; 3T Siemens—1 mm³ resolution, TR = 2300 ms, TE = 2.98 ms, flip angle = 9° and 3T Philips—1 mm³ resolution, TR = 8.0 ms, TE = 3.7 ms, flip angle = 8°). Further details on scanner protocols can be found at <https://www.radc.rush.edu/docs/var/scannerProtocols.htm> (ref. 77). Nonuniformity correction, skull-stripping, spatial normalization to the MNI152 template and tissue segmentation were performed using advanced normalization tools. Freesurfer's 'recon-all' function was applied to extract morphometric attributes.

Resting-state fMRI BOLD data were acquired on multiple scanners, which are as follows: 1.5T GE Signa scanner (5 mm³ resolution, TR/TE = 2,000/33 ms, flip angle 85°), 3T Siemens Magnetom TrioTim syngo (3.3 mm³ resolution, TR/TE = 3,000/30 ms, flip angle 80°) and 3T Philips Achieva Quasar TX (3.3 mm³ resolution, TR/TE = 3,000/30 ms, flip angle 80°). Further details on scanner protocols can be found at <https://www.radc.rush.edu/docs/var/scannerProtocols.htm> (ref. 78). Raw fMRI data were preprocessed using a well-validated pipeline of robust and reproducible fMRI processing tools. First, CuBIDS was used to identify and investigate scans with deviant parameters to reduce scan heterogeneity¹⁸. fmripiprep (v.20.2.3)¹⁹ was then applied to realign and slice-time correct the fMRI volumes, with distortion correction further performed if accompanying fieldmap volumes were available. The resulting fMRI volumes were coregistered to their corresponding T1 volumes using FLIRT. Finally, the fMRI time series underwent confound regression using the eXtensible Connectivity Pipeline (XCP; https://github.com/PennLINC/xcp_d)⁶¹ (v.0.0.4 'xcp-abcd') with the 36p+Despike model⁷⁹. In brief, bandpass filtering between 0.01 and 0.08 Hz was applied, followed by despiking with AFNI⁸⁰ and regression of 36 parameters from the time series⁸¹. Regressors included six motion attributes derived during realignment, a global signal and two physiological parameters, as well as their derivatives, squares of derivatives and quadratic terms. Additionally, mean framewise displacement was calculated for each participant and used to account for any residual effect of head motions in subsequent analysis.

Proteomic

Multiplex TMT-MS was used to generate proteomic data (7,788 proteins). Brain tissue homogenization, protein digestion, TMT peptide labeling, high-pH offline fractionation and liquid chromatography–tandem mass spectrometry (LC–MS/MS), as well as protein quantification, batch correction and data preprocessing, were performed as previously described^{23–25}. Briefly, paired human tissue samples from SFG and ITG

were obtained from 98 ROSMAP participants. Approximately 100 mg of tissue was homogenized in 8 M urea, 10 mM Tris and 100 mM NaH₂PO₄ (pH 8.5) buffer with Halt protease and phosphatase inhibitor cocktail (Thermo Fisher Scientific) using a Bullet Blender (Next Advance). Protein concentration was determined by bicinchoninic acid assay (Pierce), and 1D SDS–PAGE gels were run with Coomassie blue staining for quality control before protein digestion. Lysyl endopeptidase (Wako) at 1:100 (wt/wt) was added to each 100 µg sample for digestion overnight. Trypsin (Promega) was added at 1:50 (wt/wt), and digestion occurred for another 16 h. The peptide solutions were acidified and desalted. The samples were then loaded onto the column, washed and eluted. An equal amount of peptide from each sample was aliquoted and pooled as the global internal standard (GIS), which was split and labeled in each TMT batch. The eluates were then dried to completeness using a SpeedVac. Before TMT labeling, cases were randomized by covariates (age, sex, PMI, diagnosis, etc.) into 26 total batches. Peptides from each individual case and the GIS pooled standard or bridging sample (at least one per batch) were labeled using the TMT 11-plex kit (Thermo Fisher Scientific, 90406). For each batch, up to two TMT channels were used to label GIS standards, while the remaining TMT channels were used for samples after randomization. Next, high-pH offline fractionation was performed, and 96 individual equal-volume fractions were collected across the gradient and then pooled by concatenation into 24 fractions and finally dried with a SpeedVac. Fractions were resuspended in an equal volume of loading buffer and analyzed by LC–MS/MS. Peptide eluents were separated on a self-packed C18 (1.9 µm; Dr. Maisch) fused silica column (25 cm × 75 µm internal diameter; New Objective) by a Dionex UltiMate 3000 RSLCnano liquid chromatography system (Thermo Fisher Scientific). Peptides were monitored on an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific). Full MS scans were collected at a resolution of 120,000 (400–1,400 *m/z* range, 4×10^5 AGC, 50-ms maximum ion injection time). All HCD MS/MS spectra were acquired at a resolution of 60,000 (1.6 *m/z* isolation width, 35% collision energy, 5×10^4 AGC target, 50-ms maximum ion time). Dynamic exclusion was set to exclude previously sequenced peaks for 20 s within a 10-ppm isolation window. Peptide spectral matches (PSMs) were filtered to an FDR of less than 1% using the Percolator node. For database searching and protein quantification, raw MS data files were analyzed in Proteome Discover software suite (v2.3; Thermo Fisher Scientific), and MS/MS spectra were searched against the UniProtKB human proteome database. Following spectral alignment, peptides were assembled into proteins and filtered based on the combined probabilities of their constituent peptides to a final FDR of 1%. Reporter ions were quantified from MS2 scans using an integration tolerance of 20 ppm with the most confident centroid setting. Only PSMs with less than 50% isolation interference were used for quantification, and only unique and razor (that is, parsimonious) peptides were considered for quantification. Multiconsensus was performed to achieve parsimony across individual batches. In cases of redundancy, shared peptides were assigned to the protein sequence in adherence with the principles of parsimony. Finally, batch correction and data preprocessing were performed. A total of 10,426 high confidence, master proteins were identified across the 26 TMT batches, but only proteins quantified in >50% of samples were included in subsequent analyses. log₂ abundances were normalized as a ratio divided by the central tendency of pooled standards. Batch correction was performed using a Tunable Approach for Median Polish of Ratio (<https://github.com/edammer/TAMPOR>). Multidimensional scaling plots were used to visualize batch contributions to variation before and after batch correction. Network connectivity was used to remove outliers, defined as samples that were greater than 3 s.d. away from the mean.

Gene expression

RNA was extracted using the Chemagic RNA tissue kit (PerkinElmer, CMG-1212). RNA was concentrated (Zymo, R1080), and RQN (RIN score)

was calculated using Fragment Analyzer (Agilent, DNF-471). RNA concentration was determined using the Qubit broad-range RNA assay (Invitrogen, Q10211) according to the manufacturer's instructions. In total, 500 ng total RNA was used for RNA-seq library generation, and rRNA was depleted with RiboGold (Illumina, 20020599). A Zephyr G3 NGS workstation (PerkinElmer) was used to generate TruSeq stranded sequencing libraries (Illumina, 20020599) with custom unique dual indexes (IDT) according to the manufacturer's instructions with the following modifications. RNA was fragmented for 4 min at 85 °C. The first strand synthesis was extended to 50 min. Size selection post adapter ligation was modified to select larger fragments. According to the manufacturer's instructions, library size and concentrations were determined using an NGS fragment assay (Agilent, DNF-473) and Qubit dsDNA assay (Invitrogen). The modified protocol yielded libraries with an average insert size of around 330–370 bp. Libraries were normalized for molarity and sequenced on a NovaSeq 6000 (Illumina) at 80–100 million reads, 2 × 150 bp paired-end. RNA-seq data processing was implemented using the following three parallel pipelines: an RNA-seq quality control (QC) pipeline, a gene/transcripts quantification pipeline and a 3'-UTR quantification pipeline. In the QC pipeline, paired-end RNA-seq data were first aligned by STAR (v2.6) to a human reference genome. The primary assembly of the reference genome fasta file and transcriptome annotation came from Gencode (Release 27 GRCh38). Picard tools were applied to the aligned BAM files to assess the quality of RNA-seq data. In the quantification pipeline, transcript raw counts were calculated by Kallisto (v0.46).

We preprocessed the RNA-seq data following the pipeline previously described⁸². In brief, we applied TMM normalization (using edgeR calcNormFactors) to the raw counts to estimate the effective library size of each participant. We then applied voom/limma to regress out confounds and convert the counts into log counts per million ($\log_2(\text{CPM})$). Technical confounds included batch, study (ROS or MAP), RNA integrity number, postmortem interval, library size, percentages of aligned reads, coding and intergenic bases, ribosomal and untranslated region bases, duplicated reads, 3' bias, 5' over 3' bias and coefficient of variation in coverage. Only genes with mean $\log_2(\text{CPM}) > 2$ were kept.

Dendritic spine imaging and processing

Golgi-Cox staining, using the FD Rapid Golgi Stain Kit (FD Neurotechnologies, PK401), was used to visualize dendrites and dendritic spines in postmortem samples from SFG and ITG. The following adjustments were made from the standard operating procedure in the manual from the kit. All steps were performed at room temperature. Solutions A (potassium dichromate and mercuric chloride) and B (potassium chromate) were combined 48 h before tissue submersion. In total, 2 ml of solution A + B were placed in wells of a 12-well plate (Thermo Fisher Scientific, 08-772-29). Frozen tissue blocks, approximately 10 × 10 × 10 mm, were dropped into the A + B solution in each well. The chromate solution was replaced after 24 h, and the tissue blocks remained in the solution for 21 days. Next, the tissue blocks were transferred to a new 12-well dish containing solution C. Solution C was replaced after 24 h. After a total of 72 h in solution C, each tissue block was sliced into 125 µm sections in solution C using a Leica Vibratome (VT1000 S). Free-floating sections were placed in a six-well dish (Thermo Fisher Scientific, 353046) that contained solution C in each well. Tissue sections were sequentially moved from solution D to solution E and to distilled water, similar to the manufacturer's instructions. Slices were dehydrated with alcohols (70%, 90% and 100% ethanol) and cleared with xylenes (Thermo Fisher Scientific, X3P). Slices were then placed on glass slides (Thermo Fisher Scientific, 12-550-15) with a single slide spacer (Electron Microscopy Sciences, 70327-20S), sealed with Permount (Thermo Fisher Scientific, SP15-100) and coverslipped with 24 × 50 mm, thickness 0.13–0.17 mm, glass (Carolina Biological, 633153). Slides were dried in the dark for 7 days before microscopy imaging.

Dendrites were imaged by blinded experimenters using bright-field microscopy. Each case had multiple slides with multiple slices available for imaging. From each tissue slice, one to two pyramidal neurons from layers two or three were randomly selected to image for analysis. Between 8 and 12 neurons were sampled per individual. A single dendritic segment was imaged per neuron. While we used random site sampling, dendrite segments were imaged if the following criteria were met: (1) located centrally within the total tissue depth, (2) not obscured by staining debris and/or intersecting neighboring neurons and (3) fully impregnated. Additionally, only dendrites that were over 30 µm in length and had an approximate diameter of 1 µm were imaged. Tissue slices were visualized under 10× magnification to determine tissue quality. Once suitable dendrites were identified, tissue slices were visualized at 60× magnification using Type F immersion oil (Nikon, MXA22168). Z stacks were captured at a step size of 0.1 µm at ×60 using a Nikon Plan Apo ×60/1.40 numerical aperture (NA) oil-immersion objective in combination with a high-NA oil condenser (Nikon, MEL59500) on a Nikon Eclipse Ti2 inverted microscope with a Lumencor SOLA light engine and Hamamatsu ORCA-flash 4.0 digital camera. All images were 1,024 × 1,024 pixels.

Reconstructions of dendrites and dendritic spines were conducted by blinded experimenters using Neurolucida 360 (MBF Biosciences, v2.70.1). Nikon ND2 files were converted to 16-bit TIFF files using ImageJ and then imported to Neurolucida 360. Dendritic segments were traced using a user-guided semi-automated directional kernel algorithm. Initiation and termination points for dendrite reconstruction were determined with the following criteria: (1) ≥5 µm away from the distal tip of the dendrite, (2) consistent diameter, (3) level axis with limited smear in the z plane and (4) ≥20 µm in length. After the dendritic segment was traced, the experimenter verified that the points located on the dendrite were accurate in x, y and z planes and, if not, made manual adjustments to the trace. The diameter of the dendrite at each point was also verified to ensure the accuracy of the trace. Dendritic spines were traced using voxel clustering. The following parameters were used for spine identification: outer range, 7 µm; minimum height, 0.3 µm; detector sensitivity, 90–125% and minimum count, 8 voxels. The morphology of each dendritic spine was examined to verify that the axial smear did not cause misrepresentation. Merge and slice tools in Neurolucida 360 were used to correct inaccuracies. The position of every spine's backbone point was verified by the blinded experimenter. If the backbone caused a misrepresentation, the dendritic spine was visualized from the z plane, and the experimenter's re-oriented backbone points in the x–y plane. Notably, repositioning in the x–z or y–z plane was performed while the spine was being viewed from the lateral angle. Morphometric analysis was conducted for each spine, and measurements categorized spines into thin, stubby, mushroom and filopodia classes. The following parameters were used for spine classifications: head-to-neck ratio, 1.1; length-to-head ratio, 2.5; mushroom head size, 0.35 µm and filopodium length, 3.0 µm. Spines with a head-to-neck ratio >1.1 and head diameter >0.35 µm were classified as mushroom. Spines were classified as filopodia or thin if the head-to-neck ratio was <1.1 and either (1) length-to-head ratio was >2.5 or (2) head size was <0.35 µm. Moreover, if the total length of the spine was >3.0 µm, the spine was classified as filopodia, and if the total length was <3.0 µm, then the spine was classified as thin. Dendrite and spine reconstructions were exported to Neurolucida Explorer (MBF Biosciences, v2.70.1), where data were collected for quantitative analysis. The dendritic spine measurements were exported and collected in Microsoft Excel. Spine density was calculated by determining the quantity of spines per 10 µm of dendrite length. Spine length was defined as the curvilinear backbone length from the insertion point to the most distal point of the spine head. Head diameter was defined as the breadth of the spine head at its widest cross-sectional point.

Dendritic spines display an enormous array of morphologies, including varying head diameters and neck lengths, which are

inextricably linked with spine function. For example, spine head diameter can reflect the density of protein receptors at the postsynapse, which correlates with synapse strength³⁶. The width and length of the spine neck influence the diffusion of signaling molecules and the degree to which individual synapses affect neuronal activity⁸³. Measurements of spine head diameter and spine length as a ratio form the basis of classification into the common morphologic subclasses. Thin spines have short necks and small heads, whereas mushroom spines have longer necks and larger heads. Stubby spines are hypothesized to be transitional structures with a short neck and wide head. Subtle alterations in spine morphology and related biology can induce marked effects on connectivity patterns of neuronal circuits at microscale and downstream cognitive behavior³⁰.

Our spine morphology and density measurements are consistent with similar studies assessing spine structure characteristics in human and nonhuman primates using confocal and light microscopy^{84–87}. In a particularly elegant study, ref. 87 assayed fresh human brain tissue obtained from the neurosurgical operation and used iontophoretic microinjection of Lucifer yellow and high-resolution confocal microscopy. Our dendritic spine density and length measurements are highly consistent with details in ref. 87, which reported an average spine density of 1 per μm on layer III pyramidal neuron dendrites in the cingulate cortex of an 80-year-old male. The mean spine density in SFG is 1.3 per μm and 1.2 per μm in ITG (Supplementary Fig. 3c). The details in ref. 87 reported an average spine length of 1.4 μm on layer III dendrites in the cingulate cortex. The mean spine lengths are 1.54 μm in SFG and 1.46 μm in ITG (Supplementary Fig. 3d). However, spine volume measurements in ref. 87 averaged 0.36 μm^3 , whereas the mean spine volume is 0.26 μm^3 in SFG and 0.27 μm^3 in ITG (Supplementary Fig. 3e). This disconnect could be attributed to the lack of resolution using Golgi stain to measure the volume of the spine neck. The fluorescence microscopy approach by ref. 87 allows greater signal-to-noise ratios for more precise 3D measurement of the spine neck, which is approximately 100 nm³. However, bright-field microscopy of Golgi stain relies on light absorption, which may not provide sufficient signal-to-noise to accurately measure the volume of the spine neck.

Genetics

Genotyping was performed using the Infinium Global Screening Array, Affymetrix Genome-Wide Human SNP Array 6.0 and Illumina OmniQuad Express platform^{88,89}. The genotyping data underwent sample-level exclusions, including removing duplicated samples, those with a genotyping success rate below 95%, and samples with discordant gender information. At the probe level, additional filtering was applied based on specific quality-control criteria—a Hardy–Weinberg equilibrium P value threshold ($<1 \times 10^{-50}$), genotype call rate (<0.9) and misshape test ($<1 \times 10^{-9}$). For imputation on the Haplotype Reference Consortium r1.1, we used the Michigan Imputation Server, Minimac3 (v1.0.4) and Eagle (v2.3). Before imputation, the input data were prepared using the HRC-1000G-check-bim_ts.pl script (available at <https://www.well.ox.ac.uk/~wrayner/tools/>). Imputed genotypes with an information score greater than 0.3 were converted to a Plink binary file format using Plink 1.9 for downstream analysis.

Functional connectivity estimation

To estimate functional connectivity, we first used a functional atlas comprising 100 parcels (Schaefer2018_100Parcels_17Networks_order_FSLMNI152) generated from resting-state fMRI data of >1,000 participants²⁰ to divide the brain into functionally homogenous regions. We then averaged the time series of voxels within each parcel and computed Pearson's correlation between all pairs of parcels for each participant. To select the brain connection (that is, the parcel pair) most relevant to our molecular data, we examined the spatial overlap between each functional parcel and the brain areas at which we drew tissue samples. Parcels 25 (left hemisphere) and 74 (right hemisphere)

have the greatest overlap with our SFG tissue samples, and parcels 14 (left hemisphere) and 65 (right hemisphere) have the greatest overlap with our ITG tissue samples. For associating with the molecular data, we only used the SFG–ITG connectivity estimate that matches the side of the brain from which we drew tissue samples for each participant, that is, the connectivity estimate between parcels 25 and 14 if we drew brain tissue samples from the left hemisphere for a given participant, and vice versa. We used the connectivity estimates of other parcel pairs for assessing regional specificity ('Module-level association analysis').

Structural covariation estimation

To estimate structural covariation, we used the nine regional morphological attributes (number of vertices, surface area, gray matter volume, average cortical thickness, standard deviation of cortical thickness, mean curvature, Gaussian curvature, fold index and curvature index) from Freesurfer cortical surface reconstruction with regions defined anatomically based on the DKT atlas²². Let Z_i be the $n \times 9$ morphological attribute matrix of region i with n being the number of participants. The standard approach⁹⁰ estimates the structural covariation between regions i and j by correlating row k of Z_i with row k of Z_j for $k = 1$ to n . The assumption is that morphological attributes have a one-to-one correspondence across regions, that is, the fold index in region i covaries with the fold index in region j , the curvature index in region i covaries with the curvature index in region j , etc. However, the fold index of region i might also covary with, for example, the curvature index of region j . Therefore, to mitigate the one-to-one attribute correspondence assumption, we applied CCA to find combinations of morphological attributes that maximally covary between region pairs (Supplementary Fig. 1). Specifically, we applied CCA to Z_i and Z_j to find 9×9 projection matrices A_i and A_j that maximize their correlation, that is, $\max \text{tr}(A_i^T Z_i^T Z_j A_j)$ subject to $A_i^T Z_i^T Z_i A_i = I$ and $A_j^T Z_j^T Z_j A_j = I$. Adopting a recent approach where co fluctuation in fMRI intensity at single time point resolution is used to estimate connectivity between two regions⁹¹, we used $S_{ij} = (Z_i A_i) \circ (Z_j A_j)$ as estimates of structural covariation between regions i and j , where the symbol \circ denotes element-wise product and each CCA dimension (that is, each column of S_{ij}) captures a different mode of shape covariation. For associating with the molecular data, we split the participants based on the side of the brain from which we drew tissue samples and separately applied CCA to those two sets of participants. We focused our analysis on the structural covariation between 'superior frontal' and 'inferior temporal' in the DKT atlas, which best overlaps with where we sampled brain tissues and used other anatomical region pairs for assessing regional specificity ('Module-level association analysis').

Molecular module estimation

To determine the molecular systems of interacting proteins in each brain region, we clustered the proteins into co-abundance modules using a fully data-driven approach called SpeakEasy⁹², which takes in a similarity matrix and joins nodes into modules based on both local connectivity and global network structure. To estimate the degree of interaction between each pair of proteins, we correlated their abundance levels across participants. We repeated this estimation for all protein pairs to generate a protein correlation matrix for each region and applied SpeakEasy, which extracted ten modules for SFG and ten modules for ITG (Supplementary Table 2). Considering our goal is to model brain connectivity, which presumably arises from subprocesses related to synaptic communication, we applied Fisher's exact test to annotate each module by GO terms and focused our analysis on modules (pMod) whose protein members are most enriched for synaptic communication terms. Using this criterion, we selected pMod6 for SFG and pMod8 for ITG, which happen to have the greatest protein overlap among all protein module pairs ($OR = 13.16$, $P = 3.99 \times 10^{-186}$). We also annotated the modules by brain cell types by applying GSEA³³ with cell type-specific mean expression⁹³ as scores and proteins within our modules as genesets. The selected protein modules also happen

to be the most enriched for neurons among modules (Supplementary Table 2). To represent a protein module, we used the average abundance levels of its protein members. We applied the same procedures to the gene expression data, which extracted nine modules for SFG and nine modules for ITG (Supplementary Table 3). Based on synaptic GO term enrichment, we selected and focused our analysis on eMod5 for SFG and eMod4 for ITG, which also happen to have the greatest gene overlap among all expression module pairs ($OR = 57.25$, $P = 4.94 \times 10^{-323}$) and most enriched for neurons (Supplementary Table 3).

Proteomic characterization of dendritic spine attributes

To provide a simple characterization/quality check of the dendritic spine data, we associated each dendritic spine attribute with the protein data and examined its GO term enrichment to check if the dendritic spine attributes are enriched for the expected neuronal structures and functions. Specifically, we applied regression with each measured protein from a given brain region as the response and each spine attribute from the same brain region as a covariate of interest, while accounting for age of death, sex, PMI and year of education as confounding factors. We then applied GSEA to the set of t values of each dendritic spine attribute and declared significance at an α of 0.05 with Bonferroni correction for the number of GO terms ($P < 5 \times 10^{-6}$).

Module-level association analysis

Let \mathbf{Y}_{ij} be a $n \times 1$ vector corresponding to neuroimaging estimates of a brain attribute (functional connectivity or structural covariation between regions i and j) of n participants, \mathbf{X}_i and \mathbf{X}_j be $n \times 1$ vectors corresponding to the average molecular values (protein abundance or gene expression level) of members within the synaptic modules of regions i and j and \mathbf{C} be a $n \times d$ confound matrix. We modeled \mathbf{Y}_{ij} as follows: $\mathbf{Y}_{ij} = \alpha_0 + \mathbf{C}\alpha + \mathbf{X}_i\beta_i + \mathbf{X}_j\beta_j + \mathbf{X}_i \circ \mathbf{X}_j\beta_{ij} + \boldsymbol{\epsilon}$, with age at death, age of scan, sex, years of education, scanner, postmortem interval and side of brain molecular data as acquired and mean framewise displacement as confounds. Our primary question is whether $\mathbf{X}_i\beta_i + \mathbf{X}_j\beta_j + \mathbf{X}_i \circ \mathbf{X}_j\beta_{ij}$ can significantly explain variance in \mathbf{Y}_{ij} across participants for $i = \text{SFG}$ and $j = \text{ITG}$. Considering that the biophysical scale at which proteins reside might be far from region-level attributes, we tested whether cellular contextualization of these modules can better reveal their association with our brain attributes of interest. Accordingly, we extracted the dendritic spine component of the synaptic modules by using regression to fit the average molecular values of their members as responses with morphometric attributes of dendritic spines as covariates and used these components (that is, the fitted values) as \mathbf{X}_i and \mathbf{X}_j . To assess regional specificity, we fitted the same model with \mathbf{Y}_{ij} set to brain attribute values of other region pairs (that is, non-SFG–ITG connections) and computed the percentage of other region pairs with lower P values. For associating with structural covariation, which has nine CCA dimensions, we separately fitted the model to each CCA dimension and applied the aggregated Cauchy association test (ACAT) to combine P values across the nine dimensions⁹⁴. We opted to use ACAT over other P value combination tests because it provably provides optimal power in sparse settings^{94,95}, which accentuates the CCA dimensions that have a stronger association with the synaptic modules, in contrast to, for example, Fisher's method that weights all dimensions equally. In addition to the synaptic modules, we also tested the association between other modules and connectivity as an exploratory analysis. To aid interpretation as well as reduce the number of statistical tests, we used Hungarian clustering to first pair up SFG and ITG modules that represent similar systems based on ORs of their gene overlap. We then associated each module pair against functional connectivity and structural covariation.

Molecule-level association analysis

To find the specific molecules (proteins and genes) associated with our brain attributes of interest, we modeled \mathbf{Y}_{ij} as follows: $\mathbf{Y}_{ij} = \alpha_0 + \mathbf{C}$

$\alpha + \mathbf{x}_{mi}\beta_{mi} + \mathbf{X}_j\beta_j + \mathbf{x}_{mi} \circ \mathbf{X}_j\beta_{mij} + \boldsymbol{\epsilon}$, where \mathbf{x}_{mi} is a $n \times 1$ vector of molecular values (abundance level of protein m or expression level of gene m in region i) of n participants. The other variables are as defined in the module-level association model. We used this model to avoid testing all pairwise combinations of molecules between SFG and ITG (7,788² protein pairs or ~12,000² gene pairs), which limits statistical power. We tested for molecules involved in SFG by setting i to SFG and j to ITG, and vice versa. We focused on testing the following three terms: $\mathbf{x}_{mi}\beta_{mi}$, $\mathbf{x}_{mi} \circ \mathbf{X}_j\beta_{mij}$, and $\mathbf{x}_{mi}\beta_{mi} + \mathbf{x}_{mi} \circ \mathbf{X}_j\beta_{mij}$, that is, effects that do not exclusively involve the synaptic modules. We declared significance at 0.05 with FDR correction over the combinations of molecules, regions and effects of interests. We considered a molecule as significant if any of the three effects is significant with FDR correction. Considering a molecule could be involved in multiple molecular processes, we again applied cellular contextualization by substituting the molecular values with their dendritic spine components and repeated the analysis. To test if cellular contextualization improves association strength, for each combination of effect and region, for example, the main effect of SFG, we compared the $-\log_{10}(P)$ of an effect with versus without cellular contextualization using a permutation test. Let \mathbf{d} be a $q \times 1$ vector with elements corresponding to the differences in $-\log_{10}(P)$ between with and without cellular contextualization for q molecules. For each permutation, we randomly selected $q/2$ molecules, flipped the sign of their $-\log_{10}(P)$ differences and computed the average difference over molecules. We repeated this procedure 10^4 times and computed the proportion of times the nonpermuted average $-\log_{10}(P)$ difference is smaller than the permuted average $-\log_{10}(P)$ differences, that is, a permutation-based P value. We further performed GSEA³³ on GO terms and a connectivity-related geneset⁵⁴ for result interpretation. We applied GSEA to the t values of all measured molecules (separately for proteins and genes) for each of the three effects on functional connectivity (separately for SFG and ITG) and declared significance at 0.05 with Bonferroni correction ($P < 5 \times 10^{-6}$). Using t values of all measured molecules control for background genes. We considered a GO term as significant if any of the three effects is significant. For $\mathbf{x}_{mi}\beta_{mi} + \mathbf{x}_{mi} \circ \mathbf{X}_j\beta_{mij}$, we only have an F value for each molecule. Therefore, we approximated a t value for each molecule by taking the square root of its F value and multiplying it by the sign of the mean t value of $\mathbf{x}_{mi}\beta_{mi}$ and $\mathbf{x}_{mi} \circ \mathbf{X}_j\beta_{mij}$, that is, taking the sign of the more dominant effect. For structural covariation, we only have P values derived by aggregating the CCA dimensions. Therefore, for each molecule, we approximated its effect direction by averaging its t values on a given effect across the CCA dimensions and taking the sign of this average. We used signed $-\log_{10}(P)$ as input to GSEA.

MetaXcan analysis

To compare our connectivity-associated proteins against previous imaging genetics studies, we used a functional mapping approach, called MetaXcan⁵⁵, to convert SNP-level GWAS summary statistics to protein level. For each protein m , We first built a model to extract the genetic component of its protein abundance: $\mathbf{P}_m = \sum_{s \in S_m} \mathbf{w}_{ms} \mathbf{S}_s + \boldsymbol{\epsilon}_m$, where \mathbf{P}_m is a $n \times 1$ vector containing the abundance levels of protein m ; \mathbf{S}_s is a $n \times 1$ vector containing the dosage of SNP s ; \mathbf{w}_{ms} is the s th element of a $l_m \times 1$ model weight vector, \mathbf{w}_m , to be estimated; and S_m is the set of l_m SNPs within ± 1 Mb from the transcription start site of gene m that produces protein m . Following MetaXcan, we estimated \mathbf{w}_{ms} using elastic net regression by applying GLMNET with its default settings, that is, tenfold cross-validation to set the sparsity parameter. We regressed out confounds including age of death, sex, postmortem interval, top three ancestry principal components and top ten protein principal components from the protein abundance measurements before fitting the models. Given \mathbf{w}_{ms} , protein level GWAS z score, z_m , of protein m can be estimated by the following: $z_m = \sum_{s \in S_m} \mathbf{w}_{ms} \sigma_s / \sigma_m \times z_s$, where z_s is the GWAS z score of SNP s , and σ_m and σ_s are the variance of protein m , and SNP s , respectively. We built models for SFG and ITG

proteins separately and packaged the models into SQLite databases (<https://doi.org/10.7303/syn52087434>) so users can directly use them with the public MetaXcan software (<https://github.com/hakyimlab/MetaXcan>). We applied MetaXcan with our protein models to a GWAS on functional connectivity between SFG and ITG (rfMRI connectivity ICA100 edge 849 (ref. 6)) and examined which proteins found associated with SFG–ITG connectivity in our data show significant genetic effects based on z_m .

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Protein and dendritic spine data are available through <https://adknowl-edgeportal.synapse.org/>. ROSMAP resources, including neuroimaging, gene expression and genetics, can be requested at <https://www.radc.rush.edu>. MS/MS spectra were searched against the UniProtKB human proteome database.

Code availability

All analyses described in the Methods used standard MATLAB and R functions as well as code provided in referenced publications without modifications. Custom code was not generated in this study.

References

77. Fleischman, D. A. et al. Gray-matter macrostructure in cognitively healthy older persons: associations with age and cognition. *Brain Struct. Funct.* **219**, 2029–2049 (2014).
78. Han, S. D. et al. Neural intrinsic connectivity networks associated with risk aversion in old age. *Behav. Brain Res.* **227**, 233–240 (2012).
79. Ciric, R. et al. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage* **154**, 174–187 (2017).
80. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).
81. Satterthwaite, T. D. et al. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage* **64**, 240–256 (2013).
82. Felsky, D. et al. The Caribbean–Hispanic Alzheimer’s disease brain transcriptome reveals ancestry-specific disease mechanisms. *Neurobiol. Dis.* **176**, 105938 (2023).
83. Tønnesen, J., Katona, G., Rózsa, B. & Nägerl, U. V. Spine neck plasticity regulates compartmentalization of synapses. *Nat. Neurosci.* **17**, 678–685 (2014).
84. Benavides-Piccione, R., Ballesteros-Yáñez, I., DeFelipe, J. & Yuste, R. Cortical area and species differences in dendritic spine morphology. *J. Neurocytol.* **31**, 337–346 (2002).
85. Tang, G. et al. Loss of mTOR-dependent macroautophagy causes autistic-like synaptic pruning deficits. *Neuron* **83**, 1131–1143 (2014).
86. Young, M. E., Ohm, D. T., Dumitriu, D., Rapp, P. R. & Morrison, J. H. Differential effects of aging on dendritic spines in visual cortex and prefrontal cortex of the rhesus monkey. *Neuroscience* **274**, 33–43 (2014).
87. Benavides-Piccione, R., Fernaud-Espinosa, I., Robles, V., Yuste, R. & DeFelipe, J. Age-based comparison of human dendritic spine structure using complete three-dimensional reconstructions. *Cereb. Cortex* **23**, 1798–1810 (2013).
88. Shulman, J. M. et al. Genetic susceptibility for Alzheimer disease neuritic plaque pathology. *JAMA Neurol.* **70**, 1150–1157 (2013).
89. De Jager, P. L. et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer’s disease research. *Sci. Data* **5**, 180142 (2018).
90. Carmon, J. et al. Reliability and comparability of human brain structural covariance networks. *Neuroimage* **220**, 117104 (2020).
91. Sporns, O., Faskowitz, J., Teixeira, A. S., Cutts, S. A. & Betzel, R. F. Dynamic expression of brain functional systems disclosed by fine-scale analysis of edge time series. *Netw. Neurosci.* **5**, 405–433 (2021).
92. Gaiteri, C. et al. Identifying robust communities and multi-community nodes by combining top-down and bottom-up approaches to clustering. *Sci. Rep.* **5**, 16361 (2015).
93. Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer’s disease. *Nature* **570**, 332–337 (2019).
94. Liu, Y. et al. ACAT: a fast and powerful p value combination method for rare-variant analysis in sequencing studies. *Am. J. Hum. Genet.* **104**, 410–421 (2019).
95. Liu, Y. & Xie, J. Cauchy combination test: a powerful test with analytic p -value calculation under arbitrary dependency structures. *J. Am. Stat. Assoc.* **115**, 393–402 (2020).

Acknowledgements

We are grateful to the participants in the Religious Order Study, the Rush Memory and Aging Project (ROSMAP) and also the staff and investigators at the Rush Alzheimer’s Disease Center for providing and processing high-quality data. This study was funded by the National Institutes of Health (R01AG061800 to B.N., S.T., C.G. and J.H.; R01AG061798 to B.N., A.Z., S.T. and C.G.; R01AG057911 and U01AG079847 to C.G., B.N. and S.T.; F31AG067635 to C.W.; R01AG054719, AG063755, R21AG085379 and AG068024 to J.H.; T32NS061788 to C.W. and P30AG10161, P30AG72975, R01AG15819, R01AG17917, U01AG46152 and U01AG61356 to D.B.).

Author contributions

B.N., S.T., C.G. and J.H. conceptualized the project. B.N., C.G., S.T. and J.H. developed the methodology. B.N., S.T., C.G., J.H., T.S., J.V., K.G., C.W., S.C., M.C., A.W., A.A., J.A., E.P., K.C. and H.M. conducted the investigation. B.N., S.T. and C.G. handled visualization. D.B., C.G., T.S. and J.H. secured funding. B.N., C.G., S.T. and J.H. wrote the original draft. B.N., S.T., A.Z., J.S., T.S., D.B., N.S., J.V., C.G. and J.H. did the writing, reviewing and editing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41593-024-01788-z>.

Correspondence and requests for materials should be addressed to Chris Gaiteri or Jeremy H. Herskowitz.

Peer review information *Nature Neuroscience* thanks Stefano Berto, Michael Hawrylycz and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection The following software was used for data collection: Proteome Discover software suite (version 2.3), fmrip (v. 20.2.3), eXtensible Connectivity Pipeline (v. 0.0.4 "xcp-abcd"),

Data analysis Data analysis and normalization was performed in: STAR v2.6, Kallisto (v0.46), edgeR 3.42.4, voom/limma 3.56.2, Neurolucida 360 (version 2.70.1), Minimac3 version 1.0.4, Eagle version 2.3, PLINK 1.9, SpeakEasy 2.0, MATLAB 2023b, Python 3.8.5, R 4.3.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Protein and dendritic spine data available through <https://adknowledgeportal.synapse.org/>. ROSMAP resources, including neuroimaging, gene expression, and genetics can be requested at <https://www.radc.rush.edu>. MS/MS spectra were searched against the UniProtKB human proteome database.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

The individuals who provided data for this study were drawn from the Religious Orders and Memory and Aging Project (ROSMAP) studies, which are community-based longitudinal studies of health and aging. All individuals were 55 year of age and extensive testing on entry assures they were cognitively healthy. See Bennett et al 2018 for more details. The cohort is majority female and in the subset we analyzed 75/98 individuals were female. Information of subject sex and all other variables recorded for these individuals is available upon request through radc.rush.edu. This sample size of individuals with neuroimaging and brain omics analysed here was insufficient for division by sex.

Reporting on race, ethnicity, or other socially relevant groupings

The ROS study enrolls nuns, priests and brothers from across the United States. MAP enrolls lay persons from northeastern Illinois. The individuals in this substudy were 100% Caucasian. The subset of individuals from both studies in the current analysis all resided within driving distance of Chicago, as they were providing brain scans via Rush University.

Population characteristics

Average age of death was 89.1 years old, and total years of education are 16.9. Population characteristics we controlled include study (ROS or MAP), sex and age of death and age at scan.

Recruitment

The ROSMAP studies primarily recruited from retirement communities and also from Section 8 and Section 202 housing subsidized by the Department of Housing and Urban Development, retirement homes, and through local churches and other social service agencies serving minorities and low-income elderly. There is really nothing more to say on recruitment, and the level of detail provided here is equal to how the studies are represented in hundreds of other publications. It's not possible to fully define self-selection bias, but participants are more highly educated and Caucasian compared to national averages. Relative to the ROS study, MAP participants may have different life experiences, as they are not members of Religious Orders, and overall lower levels of education. However the cohort was designed to be otherwise compatible in terms of test and geography, and years of education is regressed from the data. These aspects of the study are covered in detail in the cited references.

Ethics oversight

The parent studies and sub-studies were all approved by an Institutional Review Board of Rush University Medical Center and all participants signed an informed consent, Anatomical Gift Act, and a repository consent to share data and biospecimens.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences

☐ Behavioural & social sciences

☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

The sample was all individuals who provided both ante-mortem fMRI and post-mortem brain proteomics. This cohort size is consistent with that necessary for robust neuroimaging and human brain molecular biology effects from a broad population.

Data exclusions

We only included subjects with RNAseq, proteomics, dendritic spine morphology, and neuroimaging data as well as their corresponding confounding variables.

Replication

As there is no other cohort that has paired fMRI and brain proteomics, we replicated findings by incorporating gene expression, and also structural neuroimaging in this same cohort which showed related and sometimes stronger effects as the main fMRI-protein tests. We also utilize genetic effects to provide replication, comparing an external GWAS for functional connectivity between our regions of interest with our data, which show significant overlap (see text for detailed analysis). The structural MRI scans and gene expression data were collected from the same subjects from whom we collected fMRI and protein data. The GWAS data was generated from an independent cohort.

Randomization

We compare continuously distributed phenotypes (neuroimaging and omics), with no binarization or binning.

Blinding

Neuroimaging was performed prior to any brain omics, and hence intrinsically blinded, while proteomics were assays by investigators with no access to neuroimaging data, and even prior to the neuroimaging data being processed, so they were also blinded.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).
Research sample	State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.
Sampling strategy	Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.
Data collection	Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.
Timing	Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Non-participation	State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.
Randomization	If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.
Research sample	Describe the research sample (e.g. a group of tagged <i>Passer domesticus</i> , all <i>Stenocereus thurberi</i> within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.
Sampling strategy	Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.
Data collection	Describe the data collection procedure, including who recorded the data and how.
Timing and spatial scale	Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Reproducibility	Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.
Randomization	Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.
Blinding	Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	<i>Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).</i>
Location	<i>State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).</i>
Access & import/export	<i>Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).</i>
Disturbance	<i>Describe any disturbance caused by the study and how it was minimized.</i>

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input type="checkbox"/>	<input checked="" type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	<i>Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.</i>
Validation	<i>Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.</i>

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	<i>State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.</i>
Authentication	<i>Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.</i>
Mycoplasma contamination	<i>Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.</i>
Commonly misidentified lines (See ICLAC register)	<i>Name any commonly misidentified cell lines used in the study and provide a rationale for their use.</i>

Palaeontology and Archaeology

Specimen provenance	<i>Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.</i>
Specimen deposition	<i>Indicate where the specimens have been deposited to permit free access by other researchers.</i>

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals

For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.

Wild animals

Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.

Reporting on sex

Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.

Field-collected samples

For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.

Study protocol

Note where the full trial protocol can be accessed OR if not available, explain why.

Data collection

Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.

Outcomes

Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.

Dual use research of concern

Policy information about [dual use research of concern](#)

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

- | No | Yes |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Public health |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> National security |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Crops and/or livestock |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Ecosystems |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Any other significant area |

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes
<input checked="" type="checkbox"/>	<input type="checkbox"/> Demonstrate how to render a vaccine ineffective
<input checked="" type="checkbox"/>	<input type="checkbox"/> Confer resistance to therapeutically useful antibiotics or antiviral agents
<input checked="" type="checkbox"/>	<input type="checkbox"/> Enhance the virulence of a pathogen or render a nonpathogen virulent
<input checked="" type="checkbox"/>	<input type="checkbox"/> Increase transmissibility of a pathogen
<input checked="" type="checkbox"/>	<input type="checkbox"/> Alter the host range of a pathogen
<input checked="" type="checkbox"/>	<input type="checkbox"/> Enable evasion of diagnostic/detection modalities
<input checked="" type="checkbox"/>	<input type="checkbox"/> Enable the weaponization of a biological agent or toxin
<input checked="" type="checkbox"/>	<input type="checkbox"/> Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.
Novel plant genotypes	Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.
Authentication	Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.

ChIP-seq

Data deposition

- ☐ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- ☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links May remain private before publication.	For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.
Files in database submission	Provide a list of all files available in the database submission.
Genome browser session (e.g. UCSC)	Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates	Describe the experimental replicates, specifying number, type and replicate agreement.
Sequencing depth	Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.
Antibodies	Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.
Peak calling parameters	Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.
Data quality	Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.
Software	Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

Flow Cytometry

Plots

Confirm that:

- ☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☐ All plots are contour plots with outliers or pseudocolor plots.
- ☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

- Sample preparation *Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
- Instrument *Identify the instrument used for data collection, specifying make and model number.*
- Software *Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
- Cell population abundance *Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
- Gating strategy *Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*
- ☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

- Design type *resting state*
- Design specifications *Subjects were instructed to lie at rest as we performed fMRI scans.*
- Behavioral performance measures *The subjects were not performing any task, hence we did not measure behavioural performance.*

Acquisition

- Imaging type(s) *functional; structural*
- Field strength *1.5 and 3.0T*
- Sequence & imaging parameters *Resting state fMRI BOLD data were acquired on multiple scanners: 1.5T GE Signa scanner (5 mm3 resolution, TR/TE=2000/33 ms, flip angle 85°), 3T Siemens Magnetom TrioTim syngo (3.3 mm3 resolution, TR/TE=3000/30 ms, flip angle 80°) and 3T Philips Achieva Quasar TX (3.3 mm3 resolution, TR/TE=3000/30 ms, flip angle 80°). Further details on scanner protocols can be found at <https://www.radc.rush.edu/docs/var/scannerProtocols.htm>.*
- Area of acquisition *whole brain*
- Diffusion MRI ☐ Used ☒ Not used

Preprocessing

- Preprocessing software *cuBIDS was used to identify and investigate scans with deviant parameters. fmriprep (v. 20.2.3)5 was then applied to realign and slice-time correct the fMRI volumes, with distortion correction further performed if accompanying fieldmap volumes were available.*
- Normalization *The fMRI volumes were co-registered to their corresponding T1 volumes using FLIRT.*
- Normalization template *All fMRI and T1 volumes were spatially normalized to MNI152 template.*
- Noise and artifact removal *The fMRI timeseries underwent confound regression using the eXtensible Connectivity Pipeline (XCP; https://github.com/PennLINC/xc_p_d) (v. 0.0.4 "xcp-abcd") with the 36p+Despike model. In brief, bandpass filtering between 0.01 and 0.08 Hz was applied, followed by despiking with AFNI8 and regression of 36 parameters from the timeseries. Regressors included six motion attributes derived during realignment, global signal, and two physiological parameters, as well as their derivatives,*

squares of derivatives, and quadratic terms. Additionally, mean framewise displacement was calculated for each subject and used to account for any residual effect of head motions in subsequent analysis.

Volume censoring

Scans with excessive motion or distortion were excluded and remaining motion effects regressed out.

Statistical modeling & inference

Model type and settings

We used mainly univariate linear regression models. Canonical correlation analysis and elastic net regression were also used.

Effect(s) tested

Association between omic data and resting state fMRI-based connectivity estimates (as well as MRI-based structural covariation as replication). No task or stimulus was involved.

Specify type of analysis: ☐ Whole brain ☒ ROI-based ☐ Both

Statistic type for inference

t-values and p-values from linear regression were used for inferring associations between omic and neuroimaging data.

(See [Eklund et al. 2016](#))

Correction

FDR correction for all effects in the regression models.

Models & analysis

n/a | Involved in the study

☐ ☒ Functional and/or effective connectivity

☒ ☐ Graph analysis

☒ ☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Pearson's correlation

Graph analysis

We did not perform graph analysis in terms of extracting graph properties, such as node degree, betweenness centrality, clustering coefficients etc. for this study.