

OPEN
ANALYSIS

Open datasets and code for multi-scale relations on structure, function and neuro-genetics in the human brain

Antonio Jimenez-Marin^{1,2}, Ibai Diez^{3,4,5}, Asier Erramuzpe^{1,6}, Sebastiano Stramaglia⁷, Paolo Bonifazi^{1,6} & Jesus M. Cortes^{1,6,8}✉

The human brain is an extremely complex network of structural and functional connections that operate at multiple spatial and temporal scales. Investigating the relationship between these multi-scale connections is critical to advancing our comprehension of brain function and disorders. However, accurately predicting structural connectivity from its functional counterpart remains a challenging pursuit. One of the major impediments is the lack of public repositories that integrate structural and functional networks at diverse resolutions, in conjunction with modular transcriptomic profiles, which are essential for comprehensive biological interpretation. To mitigate this limitation, our contribution encompasses the provision of an open-access dataset consisting of derivative matrices of functional and structural connectivity across multiple scales, accompanied by code that facilitates the investigation of their interrelations. We also provide additional resources focused on neuro-genetic associations of module-level network metrics, which present promising opportunities to further advance research in the field of network neuroscience, particularly concerning brain disorders.

Introduction

The precise relationship between brain structural and functional connectivities has puzzled leading researchers in network neuroscience^{1–16}. Despite numerous efforts, an accurate prediction of the structural connectivity from its functional counterpart remains still a distant goal. Beyond link-level network comparisons between the two modalities, a modular-level comparison has shown bi-directional advantages in the understanding of brain structure and function^{17–29}. Although non-invasive magnetic resonance imaging (MRI) techniques have a limitation in the network resolution that they can resolve³⁰, arriving at about one millimeter resolution, multi-scale approaches between brain structural and functional modules have been studied to assess organizational aspects in healthy population^{13,26,31–33}, in relation to cognition^{34,35}, during development³⁶ and aging³⁷, or in some pathological conditions^{38,39}.

When evaluating the correspondence at the module-level (also known as aggregate level) between structure and function, there are two complementary strategies in relation to the chosen network resolution. One is the optimal strategy, in which a fixed level of network representation is established, for example, maximizing a certain network metric^{40–45}. Therefore, this optimal solution depends on the chosen metric which we have optimized to cut the hierarchical tree at a given level. The alternative approach explores the multi-scale nature of the data, using a nested hierarchy of network resolutions as potential features or input variables to be synergistically

¹Computational Neuroimaging Lab, Biobizkaia HRI, Barakaldo, Spain. ²Biomedical Research Doctorate Program, University of the Basque Country (UPV/EHU), Leioa, Spain. ³Department of Radiology, Division of Nuclear Medicine and Molecular Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, United States of America. ⁴Gordon Center for Medical Imaging, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, United States of America. ⁵Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, United States of America. ⁶IKERBASQUE Basque Foundation for Science, Bilbao, Spain. ⁷Dipartimento Interateneo di Fisica, Università Degli Studi di Bari Aldo Moro, INFN, Bari, Italy. ⁸Department of Cell Biology and Histology, University of the Basque Country (UPV/EHU), Leioa, Spain. ✉e-mail: jesus.m.cortes@gmail.com

integrated to address a specific problem^{35,37,39,46}. Therefore, for the multi-scale strategy, brain networks at different resolutions are combined, and the problem in question is solved without setting any resolution level in the hierarchical tree.

While progress has been made in assessing the module-level correspondence between structural and functional connectivities in the human brain, many questions remain unanswered. Among these inquiries, it is important to assess the extent to which the different modules within structural connectivity favor the emergence of functional modules. Moreover, it is crucial to investigate the presence of dynamic processes that influence the relationship between structural and functional modules. Identifying the scale at which the correspondence between modules in both representations reaches a maximum is also of great relevance. Understanding the insights of the pathological brain and the extent to which the interplay between structure and function is altered within this context are pressing matters. Additionally, elucidating the underlying mechanism responsible for such alterations is crucial.

In the pursuit of addressing these queries, there are brain data repositories like the Human Connectome Project⁴⁷, which offers functional connectivity matrices and diffusion data, yet lacks diverse structural connectivity matrices at varying resolutions. Another repository, MICA-MICs⁴⁸, more closely resembles our database, providing 50 structural and functional connectivity matrices across several resolutions. The latest database, utilizing data from the UK Biobank⁴⁹, contains 40,000 connectomes from healthy individuals. However, its access is somewhat restricted, as it is part of the UK Biobank database and requires payment for access. Our dataset's distinctiveness lies in its inclusion of multiscale structural-functional modules (γ -parametrized) and associated gene transcriptomic profiles.

This study aims to address this need by providing derivative functional and structural connectivity matrices at different scales and code to the scientific community, enabling progress in addressing these questions, which constitute a central focus of research in our laboratory. Specifically, in this article, we initially introduce a modeling approach to investigate the interplay between structure and function by parameterizing the degree of overlap between the two classes of connectivity. Secondly, building upon the methodology described in²⁴, we extend it to examine the module-level correspondence between the two connectivity matrices. Finally, we have extended the study on the relationship between brain structure and function by sharing also neuro-genetic code and data, delving into the underlying biological and molecular foundations of brain-related disorders, similarly to previous work⁵⁰⁻⁶⁰. We aimed to bridge the gap between transcriptomics and connectomics, particularly for those skilled in network analysis but not in data handling. This objective has been our primary drive.

Results

Structural and functional connectivity matrices at various resolutions were built making use of brain images from the open dataset "Max Planck Institut Leipzig Mind-Brain-Body Dataset" – LEMON^{61,62}, well-known for having high-quality multimodal acquisitions and preprocessed MRI sequences. Rather than using different subjects across lifespan, we specifically selected 136 young participants, aged between 20 and 30 years old, who had both the preprocessed rs-fMRI sequence and DWI sequence available. We next processed the raw images following standard neuroimaging pipelines to obtain structural connectivity (SC) and functional connectivity (FC) matrices (Fig. 1A). Subsequently, we built population connectivity matrices by choosing, for each link in the matrix, the median value across all the equivalent links in the individual connectivity matrices. Finally, we modeled the amount of interplay between structure and function connectivities using a single fusion parameter, γ . When γ equals 0 or 1, we have purely structural or functional connectivity, respectively. In intermediate situations, the amount of overlapping connectivity is modulated by γ . Furthermore, and similar to the methodology described in²⁴, for each value of the fusion parameter γ , we applied hierarchical agglomerative clustering to the resulting γ -fused structure-function matrix, γ SFC (Fig. 1B).

Optimal γ -fused structure-function modular organization. The different trees of the γ SFC matrices describe distinct scales, enabling the construction of networks at different resolutions. The determination of the optimal partition by cutting the tree at M^* modules depends on the specific metric we seek to maximize. In order to showcase the versatility of our data with various metrics, we specifically concentrate on a metric we have previously defined²⁴ and named cross-modularity χ , defined as the triple product of the modularity of the functional matrix, the modularity of the structural matrix, and the similarity between structural and functional modular configuration.

Here, we aimed to maximize χ to select the optimal value of γ for a given initial parcellation atlas (iPA). Figure S1 shows box-plots of different values of the cross-modularity χ across different dendrogram levels and for different iPAs with different size, ie., varying the number of micro-regions ranging from 183 to 2165. For further analyses, we selected the iPA with 2165 micro-regions since it exhibited a higher mean cross-modularity value together with a lower variability within a wide range of scales, particularly within the first 120 levels where the maximum χ is located. Beyond this point, there was a decreasing trend. The optimal brain partition corresponds to the iPA that maximizes χ by reducing the dendrogram to M^* modules and setting the fusion parameter to γ^* . In our case, the optimal value γ^* was determined to be 0.7, and this occurred at the level of 28 modules. However, two modules were considered invalid as they consisted of only one and two micro-regions, respectively, making it impractical to analyze the community structure within them. Henceforth, any subsequent analysis referring to the optimal brain partition will be based on the iPA that maximizes χ by reducing the dendrogram to $M^* = 26$ modules and setting the fusion parameter to $\gamma^* = 0.7$.

Multi-scale γ -fused structure-function modular organization. As an illustrative example, and to gain a deeper understanding of how the multi-scale organization is influenced by γ , we investigated the network strength of the γ SFC matrix in two network representations. At the finest spatial resolution, corresponding to the

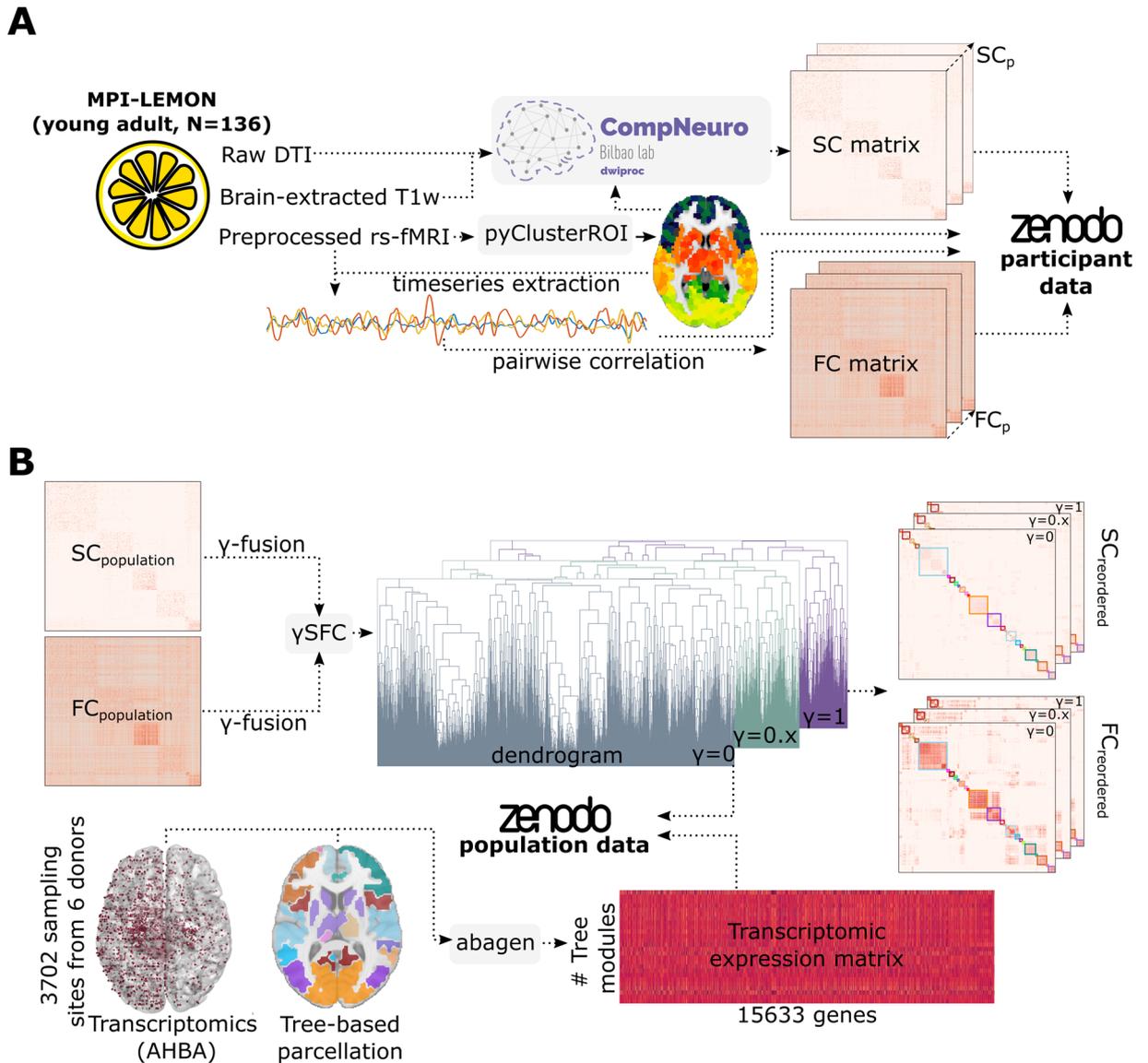


Fig. 1 Methodological sketch and pipeline. **(A)** Participants' raw MRI data from $N = 136$ healthy volunteers of the MPI-LEMON dataset were used. From preprocessed rs-fMRI images, we generated an initial brain parcellation using gray-matter masks and the open repository *pyClusterROI*. Subsequently, we employed our own publicly available code on GitHub to process the DWI images and extract the SC matrices. All post-processed data at the participant level, including the time series BOLD signal corresponding to the different iPAs, and the SC and FC matrices are available at <https://zenodo.org/record/8158914>. **(B)** We calculated population SC_p and FC_p matrices, and subsequently generated γ -fused matrices as described in the methods section. The γ SFC matrices were used to construct dendrogram trees for different γ parameters. Additionally, we employed *abagen* to generate transcriptomic expression matrices for various tree-based parcellations. The transcriptomic matrices and dendrogram trees are also accessible on Figshare. The entire pipeline and project codes can be found on GitHub at <https://github.com/compneurobilbao/bha2>.

lowest level in the hierarchical tree consisting of 2165 micro-regions (represented as brain maps in Fig. 2A), and for each of the macro-regions, including the frontal lobe, parietal lobe, occipital lobe, temporal lobe, insula, and a collection of subcortical areas (Fig. 2B). From both analyses, it is evident that increasing γ , transitioning from structure ($\gamma = 0$) to function ($\gamma = 1$), results in higher strengths shifting from subcortical areas to cortical regions. This pattern is observed consistently across all macro-regions, where the strength progressively increases with γ , peaking around $\gamma \approx 0.7$.

Beyond the two selected levels of resolution for calculating strength, a hierarchical tree organization enables obtaining different metrics at different module-levels. For instance, we first colored the $M^* = 26$ distinct modules of the optimal brain partition for ease of visualization (Figure S2, top panel). Some tree metrics we computed were the module size (MS), height (H), and multi-scale index (MSI). Very briefly, MS represents the count of micro-regions encompassed within a module. MSI quantifies the number of levels within the tree

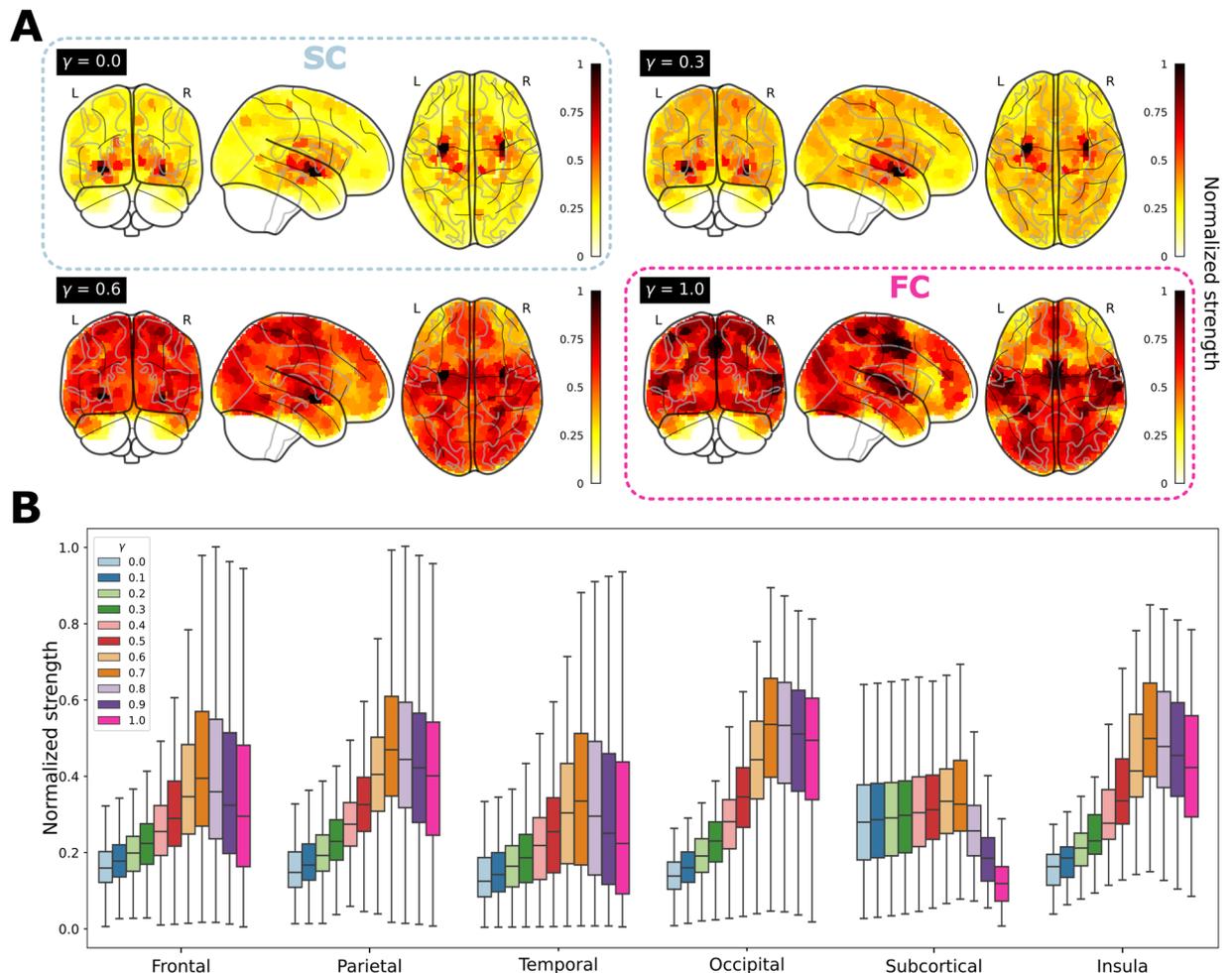


Fig. 2 Graph-node strength is modulated by γ , the amount of interplay between structural connectivity (SC) and functional connectivity (FC). **(A)** Brain maps of graph-node (normalized) strength. It is observed that for $\gamma = 0$ (pure SC), high strengths are predominantly localized in subcortical regions. However, as γ varies, there is a decrease in strength within these regions and an increase in cortical regions. **(B)** Box plots for the distribution of graph-node strength across different macro-regions: Frontal lobe, Parietal lobe, Temporal lobe, Occipital lobe, a collection of Subcortical areas, and Insula. Notably, for each macro-region, the strength shows an increasing trend from $\gamma = 0$, reaching a peak at $\gamma \approx 0.7$, followed by a subsequent decrease.

structure where a module remains intact without further division. H denotes the specific level at which a given micro-region separates from its parent module. Therefore, while H is defined for each micro-region, MS and MSI are defined at the module-level. We also defined the module height (MH), by averaging over all individual H values within the given module (Figure S2, full dendrogram and dendrogram-inset). We represented these metrics on brain plots, where we observed distinct patterns for each module (Figure S2, brain plots).

We conducted further analyses to examine the correlations among these metrics and the intra-module strength, as a proxy for module segregation (Figure S3). From the different tree measures, both MS and MH showed a high correlation ($r = 0.95$, $p < 0.001$) with module segregation, as well as between themselves ($r = -0.84$, $p < 0.001$). Interestingly, the first module $M1$ was an outlier in the metric of MSI . Located at the border between the Precuneus, Isthmus Cingulate, and Posterior Cingulate cortices (Fig. 3), $M1$ showed a remarkable resilience in remaining intact without splitting across multiple levels in the tree, possibly indicating a multi-scale functional role.

Anatomical and functional description of a given brain partition. Our methodology for defining a given partition defined by specific values of M and γ is generally data-driven. At some point, it may be of interest to characterize the anatomical and functional properties of a specific partition. As an illustrative example, we made use of the optimal partition with $M^* = 26$ modules and $\gamma^* = 0.7$ and determined the spatial locations of these 26 modules within the brain (Fig. 3, brain-glasses). We next obtained their functional and anatomical characterization by measuring respectively the amount of overlap within known Resting State Networks (RSNs)⁶³ and regions from the Desikan-Killiany atlas⁶⁴, the latter consisting of 34 cortical regions (in right and left hemispheres), 8 subcortical regions segmented from Freesurfer, and the cerebellum (Fig. 3, heatmaps). Several of our

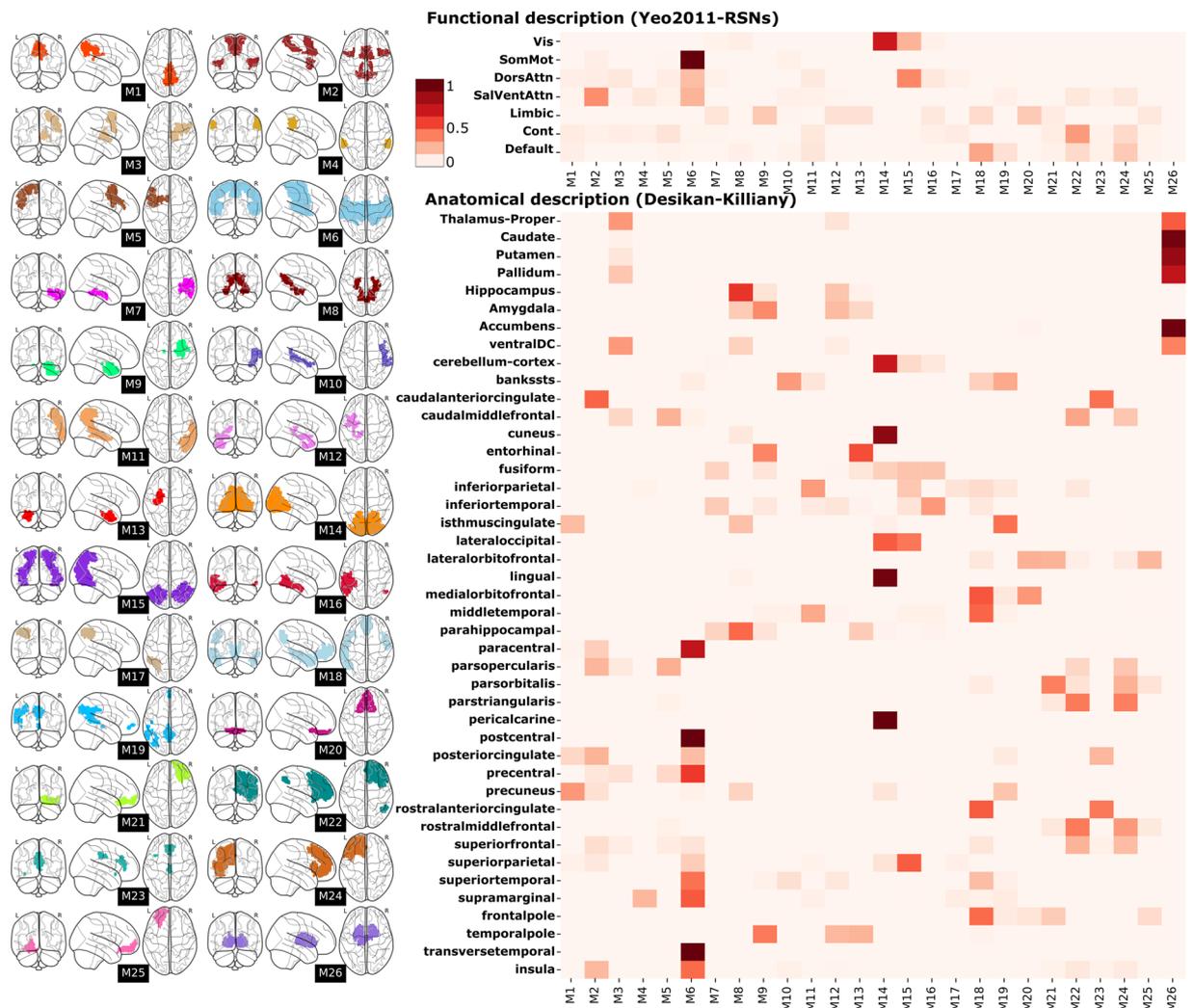


Fig. 3 Brain localization, Functional description and Anatomical description of the optimal brain partition. Left: The different M1 to M26 modules belonging to the optimal brain partition are depicted within brain representations for localization purposes. Right: Additional information is provided regarding the specific location of each module, incorporating anatomical details and functional representations derived from the overlapping analysis among different modules and established Desikan-Killiany and Resting-State Network (RSN) atlases.

modules exhibited strong overlapping within the RSNs. For instance, module M6 encompassed the Somato-motor Network (SMN) while also integrated parts of the Dorsal and Ventral Attention Networks (DAN, VAN). Module M18 showed an overlap with a portion of the Default Mode Network (DMN), and module M14 did it with the Visual Network (VIS). Moreover, certain modules displayed high overlapping with specific anatomical regions. For example, module M20 included the medial and lateral Orbitofrontal cortices, while module M26 integrated the Basal Ganglia and Thalamus.

Neurobiological relevance of a given brain partition in brain-related disorders by making use of neurogenetics data. Essential for a comprehensive biological interpretation is to characterize the modules of a given partition based on their participation in major brain-related disorders. To accomplish this, we focus again on the optimal partition and made use of publicly available transcriptomic data from the Allen Human Brain Atlas (AHBA)⁵⁰, and constructed transcriptomic expression matrices at module-level (Fig. 1B). In particular, we examined the transcriptomic expression within each module for genes associated with 40 different brain-related disorders, using a list of disorders and their corresponding genes extracted from the available data in⁶⁵. For each disorder, the average expression across samples within a module was computed for the relevant genes, and z-score values were then calculated based on the optimal partition (Fig. 4, heatmap). Modules with a z-score > 2 , indicating high expression levels, or z-score < -2 , indicating low expression levels, were marked with an asterisk. As the disorders were categorized into 7 disease groups (Psychiatric disorders, Substance abuse, Movement disorders, Neurodegenerative diseases, Tumor conditions, Developmental disorders, and Others), we also generated a z-score map representing the average expression of all genes associated with each disease group

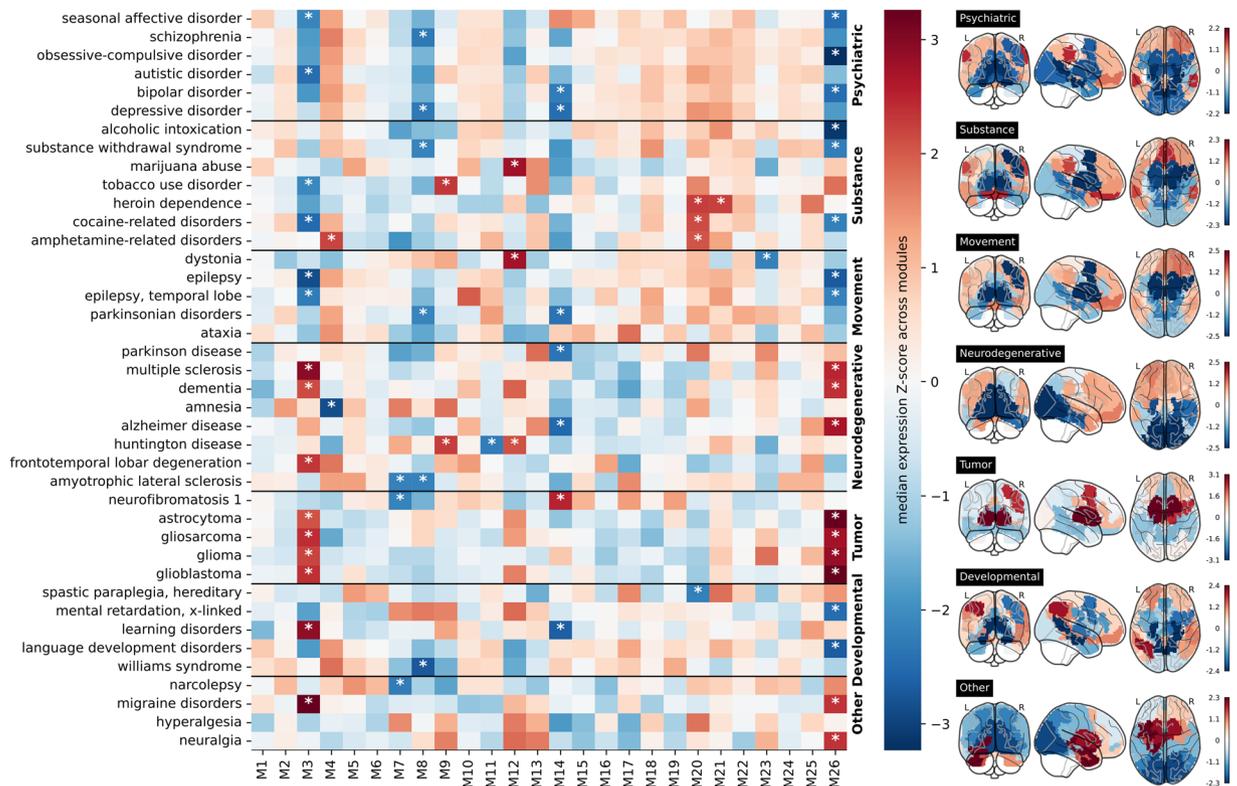


Fig. 4 Neurogenetic module-level data for disease-related transcriptomic characterization of the optimal brain partition. Left: Within each module of the optimal brain partition (M1 to M26, x-axis), we depict the median transcriptomic expression values across different sets of genes associated with various brain-related disorders. Modules with significantly high/low expression values compared to the rest are denoted by a white asterisk (*). Notice that only specific modules among the 26 exhibited significant expression values. Notably, modules encompassing the basal ganglia and thalamus (M3 and M26) demonstrated significance across all disease groups. Right: Brain maps of median transcriptomic expression, averaged over different maps belonging to the same group of brain disorders. Left, Right: Z-scores were calculated across all the modules to indicate a relative high/low gene expression levels across modules.

(Fig. 4, brain-glasses). Modules M3 and M26 exhibited high relevance across multiple disorders. M26 represented a subcortical circuit encompassing the Basal Ganglia and Thalamus, while M3 represented a Thalamo-cortical circuit comprising a portion of the right Thalamus, a small segment of the right Putamen and Pallidum, and the right Caudal Middle Frontal, Superior Frontal, and Precentral gyri. M14 was relevant for 7 disorders and included the Occipital pole and the Medial Visual Cortex. Lastly, M8 displayed relevance in 6 disorders and represented a circuit connecting the Amygdala and Hippocampus with the Precuneus via the Parahippocampal and Isthmus Cingulate cortices.

Discussion

In the words of Karl Friston, how rich functionality emerges from the invariant structural architecture of the brain remains a major mystery in neuroscience⁷. Previous research in network neuroscience has provided evidence for the emergence of hierarchical network modularity, facilitating the separation of neural computations at various spatial scales^{13,18,21,26,31–34,36,38}. However, the specific role played by structure in functional organization, as well as the reciprocal influence, remain elusive. A key contributing factor to this gap in knowledge is the continuous interaction that takes place across multiple temporal and spatial scales³⁰. Consequently, when employing a particular imaging modality for measurement, obtaining a comprehensive understanding of the intricate interplay between different scales becomes challenging.

For many years, our laboratory has been investigating the interplay between structure and function at the module-level. In our initial work 8 years ago²⁴, we demonstrated excellent correspondence among modules, but not at the level of links in connectivity matrices. Since then, we have expanded our research to include studies on healthy brains^{35,42,44} and pathological brains^{37,39–41,43,45,46}, exploring the multi-scale structural-functional correspondence at the module-level. Through collaborations with other groups, we have identified a lack of clear open data and repositories that facilitate this line of research. Many research groups in the field of physics and mathematics, specializing in complex network methodology, lack access to preprocessed brain networks derived from raw neuroimaging data. Motivated by this gap, our study aims to address this issue by employing state-of-the-art neuroimaging analyses, offering structural and functional networks at different scales, with both networks built

using the same set of network nodes. Not only do we provide the final connectivity matrices, but also the code to generate them. Additionally, we introduce hybrid structure-functional networks through the parameterization of their overlap using the parameter γ . Furthermore, we have also released pre-processed neurogenetic data and code, to facilitate the connection between structural-functional correspondence across different scales and underlying biological mechanisms, opening up a promising avenue for studying brain disorders.

The structure of the released data, code and analyses are as follows:

- All preprocessed structural and functional connectivity matrices, transcriptomics data, and γ -dendrograms for the different iPAs have been uploaded to (<https://zenodo.org/record/8158914>)⁶⁶.
- The code to build γ -dendrograms is available at https://github.com/compneurobilbao/bha2/blob/main/src/build_tree.py.
- All notebooks needed to reproduce all figures of this manuscript are available at <https://github.com/compneurobilbao/bha2/tree/main/notebooks>

Our choice to build brain partitions was based on the *pyClusterROI* package⁶⁷ rather than using other existing parcellations^{68,69}. This decision was driven by two primary factors. Firstly, it allows for a gradual increase in the number of ROIs across different scales. Notably, in our cross-modularity analysis (see Figure S1), we observed higher values for parcellations with more than 1000 ROIs. This suggests that, particularly for the metric of cross-modularity, there could be more optimal parcellations beyond 1000 ROIs (which is the maximum resolution provided by other parcellations eg. Schaefer's⁶⁹). This possibility might extend to other metrics, prompting us to retain this feature in our code. Secondly, and crucially for us, our choice included subcortical ROIs alongside cortical ones. This is particularly important in the study of brain pathologies, as many diseases impact subcortical regions in their early stages, and later on, in more severe stages.

In this manuscript, which aims to be a practical guide for studies using functional and structural connectivity matrices at different resolution scales, and their neurogenetic associations in relation to some brain pathologies, we have performed specific analyses that resulting in four different figures. Firstly, Fig. 1 presents a methodological overview of the various datasets and key steps involved in our analyses. Figure 2 aims to demonstrate the continuous parametric effect of γ on different network metrics, and for that, we have depicted the distributions of node-graph strength at two different spatial resolutions. One resolution encompassed 2165 distinct microregions, while the other aggregated them into six distinct macro-regions covering the entire brain: Frontal lobe, Parietal lobe, Temporal lobe, Occipital lobe, a collection of Subcortical areas, and Insula. Figure 3 depicts a widely accepted strategy employed in neuroimaging studies for the comprehensive characterization of a specific brain partition, encompassing both functional and anatomical labels for each brain region. Finally, Fig. 4 shows how gene transcriptomic imaging data can be used to characterize a given partition (in our example, the same partition chosen for Fig. 3) in relation to a set of genes with significant associations in 40 different brain disorders. These disorders are grouped into various categories such as psychiatric disorders, substance abuse, movement disorders, neurodegenerative diseases, tumor conditions, or developmental disorders.

To summarize, with this study we aim to provide valuable resources in the form of dataset-derivative functional and structural connectivity matrices at different scales and code to facilitate the exploration of the intricate interplay between brain structure and function across diverse spatial scales, including the aggregated level. It is our intention to contribute to unraveling the enigmatic nature of this structure-function relationship, which, despite significant progress, still harbors substantial gaps in our understanding. We encourage other researchers with an interest in this area to make use of the resources presented here to further contribute to the elucidation of this fascinating phenomenon.

Material and Methods

Raw neuroimaging dataset. The Max Planck Institut Leipzig Mind-Brain-Body Dataset, commonly referred to as LEMON^{61,62}, is a comprehensive multimodal dataset encompassing MRI sequences, EEG, ECG, and behavioral scores. In this study, we focused on a subgroup of $N = 136$ healthy individuals within the age range of 20 to 30 years, with a male population comprising 98 individuals (72%). The selection of young individuals stems from existing research indicating age's influence on both structural and functional network connectivity³⁷. For more considerations about sex-bias affecting the analyses, see subsection *sex bias considerations* below. For the precise selection of the IDs individuals, see <https://zenodo.org/record/8158914/files/participants.tsv>. Specifically, for this study we made use of three different MRI sequences: T1, rs-fMRI, and DWI, with the aim of extracting functional and structural connectivity matrices. For a comprehensive description of the sequences employed, please refer to Table 7 in⁶¹.

Preprocessed neuroimaging dataset. LEMON also provides access to preprocessed data. In our study, we made use of the brain-extracted T1 and rs-fMRI images as input data for further analysis. For preprocessing DWI images, we made use of a custom pipeline utilizing the the brain-extracted T1 images, MRtrix3 (v3.0_RC3), FSL (v6.0.1), and ANTs (v2.3.1). Our DWI pipeline involved cleaning the raw images using *DWIdenoise* and *DWIpreproc* tools, which included correction for eddy current distortion and susceptibility-induced distortion using the *topup* tool. Next, we performed whole-brain probabilistic tractography using the *iFOD2* tracking algorithm in MRtrix3. The tractography was initiated from seeds located at the gray matter-white matter interface, with a selection of 3 million fibers based on an angle threshold of $<45^\circ$ and a length threshold of <200 mm, following the recommendations in the MRtrix3 documentation. The tool *SIFT2*⁷⁰ was used to correct for the effect of crossing and disconnection of fibers on the quantification of the total number of streamlines.

Data-driven anatomical and functional solutions for different initial brain partitions. Following a methodology similar to²⁴, and utilizing the preprocessed rs-fMRI images, we performed an unsupervised voxel-level clustering analysis to delineate a specific number of micro-regions following the strategy proposed in⁶⁷. These micro-regions define the highest spatial resolution scale employed for subsequent analyses. In contrast to²⁴, where all brain voxels were included into the same clustering process, our methodology here involved six separate clustering procedures, each one focused solely on the voxels contained within distinct macro-regions, namely the frontal lobe, parietal lobe, occipital lobe, temporal lobe, insula, and subcortical structures (pooling together the thalamus, basal ganglia, amygdala, and hippocampus). Thus, while a functional partition was applied, the resulting regions were anatomically constrained by the defined macro-regions. By aggregating all micro-regions obtained from the six different macro-regions, we generated an initial Parcellation Atlas (iPA). We followed the same procedure to generate 9 different iPAs, each consisting of a different number of micro-regions, specifically 183, 391, 568, 729, 964, 1242, 1584, 1795, and 2165.

Brain functional and structural connectivity matrices, and their fusion through the parameter γ . FC matrices were generated by computing the Pearson correlation between pairs of regions fMRI time-series, each one obtained by averaging voxel-wise fMRI time-series within each micro-region within a specific iPA. For SC matrices, we employed the *tck2connectome* tool to perform fiber counting of the streamlines connecting pairs of micro-regions in the iPA. Prior to constructing SC, a nonlinear registration of the iPA to the b0 image of the DWI sequence was performed.

From the individual SC and FC matrices, we computed population matrices SC_p and FC_p by calculating for each link the median value across all equivalent links in the individual matrices. Next, and to match the sparse nature of SC_p , a threshold was applied to FC_p , resulting in comparable link density. Subsequently, we binarized the two matrices and fused them into a single matrix as follows:

$$\gamma\text{SFC} \equiv \gamma \times FC_p + (1 - \gamma) \times SC_p,$$

where γ is a real number between 0 and 1. This strategy allows us to continuously parameterize the level of interplay between SC_p and FC_p , enabling the recovery of each connectivity class when $\gamma = 0$ or $\gamma = 1$, while obtaining hybrid connectivities for intermediate values.

Multi-scale hierarchical representation at module-level. For each of the 9 distinct iPAs, a hierarchical agglomerative clustering (HAC) was employed to extract different nested modules, allowing to build different network resolutions at different scales (each one determined by a cutting tree number of modules M). More specifically, HAC was performed using a weighted method on the connectivity patterns obtained from the matrix γSFC , here defined as the correlation distance between pairs of micro-regions in the γSFC matrix.

Within the context of tree or dendrogram partitions, different multi-scale metrics can be established. These metrics can be defined either at the module-level or at a more granular level (with the highest spatial resolution determined by the micro-region level). In this particular study, we have defined the module size (MS), representing the count of micro-regions encompassed within a module. Additionally, we have introduced the multi-scale index (MSI), which quantifies the number of levels within the tree structure where a module remains intact without further division. In relation to micro-regions, we have selected the height (H) metric, which denotes the specific level at which a given micro-region separates from its parent module.

The optimal brain partition based on cross-modularity maximization. After obtaining a hierarchical partitioning of the connectivity matrix, clustering micro-regions into distinct modules M , the determination of the optimal level for cutting the tree depends on the specific metric being optimized. In this study, following the methodology introduced in²⁴, the metric chosen for maximization is the cross-modularity χ , defined as:

$$\chi \equiv (Q_F \times Q_S \times T_{FS})^{\frac{1}{3}}$$

This metric simultaneously accounts for three different qualities: the modularity of the functional matrix (Q_F), the modularity of the structural matrix (Q_S), and their similarity (T_{FS}). The latter is defined as the DICE similarity between a functional module and a structural module, and averaging across all modules in the given partition. By varying M along the tree, we can calculate χ for each configuration M and select the optimal partition M^* where χ is maximized. Additionally, unlike the approach in²⁴, we here have introduced a second parameter γ^* in the maximization of χ , which controls the amount of structure-function interplay.

Transcriptomic data at module-level to assess brain-related disorders. In addition to multi-scale brain partitions and structure-function analysis, we processed the transcriptomic open data from the Allen Human Brain Atlas (AHBA)⁵⁰ by using the *abagen* tool⁵⁷, with the default settings, which follow the recommendations of⁵³. Notably, this included setting the *ibf_threshold* to 0.5, ensuring inclusion of genes where sample signal exceeds 50% of background noise, resulting in a comprehensive list of 15,633 genes. We did not apply the *abagen.keep_stable_genes* function; therefore, our gene list was not refined using differential stability criteria. *Abagen* allows us the generation of brain transcriptome module-level values for a specific parcellation, that in our case, we chose the optimal brain partition. Moreover, we examined the transcriptomic expression at module-level in relation to genes associated with 40 brain disorders introduced in⁶⁵. To do this, we calculated the mean transcriptomic expression of all genes related to each disorder within each module. Subsequently, we performed a z-score analysis for each disorder to identify modules with low transcriptomic expression ($z < -2$) and modules with high transcriptomic expression ($z > 2$). The average expression of the different genes was also grouped into

7 different disease categories: Psychiatric disorders, Substance abuse, Movement disorders, Neurodegenerative diseases, Tumor conditions, or Developmental disorders.

Sex bias considerations. Our selection of young individuals led to choosing 136 young participants from the LEMON database, with an unexpected male predominance (72%). The manuscript's repository (<https://github.com/compneurobilbao/bha2>) includes instructions for accessing the sex data of participants (<https://zenodo.org/record/8158914/files/participants.tsv>), so if a new researcher creates a partition using a sex-balanced cohort, this limitation could be effectively addressed.

Data availability

This work involves data from different sources. First, the raw MRI images can be downloaded from the LEMON database webpage https://fcon_1000.projects.nitrc.org/indi/retro/MPI_LEMON.html⁶². Second, the transcriptomic AHBA expression data can be downloaded from <https://human.brain-map.org/>. All post-processed data obtained in this manuscript, including the time-series fMRI signals corresponding to regions in the different iPAs, the SC and FC matrices, the optimal brain partition, and the γ -modulated trees are available in (<https://zenodo.org/record/8158914>)⁶⁶.

Code availability

We have used open tools available on GitHub, and we have also uploaded our own tools and code used in this manuscript. In particular,

- For defining the iPAs, we have used the open tool “pyClusterROI” (https://ccraddock.github.io/cluster_roi/).
- For obtaining the transcriptomic expression matrices, we have used *abagen* (<https://abagen.readthedocs.io/en/stable/>).
- All the analyses and code involved in the methods and results of this manuscript have been uploaded to <https://github.com/compneurobilbao/bha2>.

Received: 10 August 2023; Accepted: 12 February 2024;

Published online: 29 February 2024

References

1. Koch, M. A., Norris, D. G. & Hund-Georgiadis, M. An investigation of functional and anatomical connectivity using magnetic resonance imaging. *NeuroImage* **16**, 241–250, <https://doi.org/10.1006/nimg.2001.1052> (2002).
2. Honey, C. J., Kötter, R., Breakspear, M. & Sporns, O. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 10240–10245, <https://doi.org/10.1073/pnas.0701519104> (2007).
3. Crofts, J. J. & Higham, D. J. A weighted communicability measure applied to complex brain networks. *Journal of the Royal Society, Interface* **6**, 411–414, <https://doi.org/10.1098/rsif.2008.0484> (2009).
4. Deco, G., Jirsa, V., McIntosh, A. R., Sporns, O. & Kötter, R. Key role of coupling, delay, and noise in resting brain fluctuations. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 10302–10307, <https://doi.org/10.1073/pnas.0901831106> (2009).
5. Honey, C. J., Thivierge, J.-P. & Sporns, O. Can structure predict function in the human brain? *NeuroImage* **52**, 766–776, <https://doi.org/10.1016/j.neuroimage.2010.01.071> (2010).
6. Adachi, Y. *et al.* Functional connectivity between anatomically unconnected areas is shaped by collective network-level effects in the macaque cortex. *Cerebral Cortex (New York, N.Y.: 1991)* **22**, 1586–1592, <https://doi.org/10.1093/cercor/bhr234> (2012).
7. Park, H.-J. & Friston, K. Structural and functional brain networks: from connections to cognition. *Science (New York, N.Y.)* **342**, 1238411, <https://doi.org/10.1126/science.1238411> (2013).
8. Goñi, J. *et al.* Resting-brain functional connectivity predicted by analytic measures of network communication. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 833–838, <https://doi.org/10.1073/pnas.1315529111> (2014).
9. Messé, A., Rudrauf, D., Giron, A. & Marrelec, G. Predicting functional connectivity from structural connectivity via computational models using MRI: an extensive comparison study. *NeuroImage* **111**, 65–75, <https://doi.org/10.1016/j.neuroimage.2015.02.001> (2015).
10. Breakspear, M. Dynamic models of large-scale brain activity. *Nature Neuroscience* **20**, 340–352, <https://doi.org/10.1038/nn.4497> (2017).
11. Bansal, K., Nakuci, J. & Muldoon, S. F. Personalized brain network models for assessing structure-function relationships. *Current Opinion in Neurobiology* **52**, 42–47, <https://doi.org/10.1016/j.conb.2018.04.014> (2018).
12. Vázquez-Rodríguez, B. *et al.* Gradients of structure-function tethering across neocortex. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 21219–21227, <https://doi.org/10.1073/pnas.1903403116> (2019).
13. Suárez, L. E., Markello, R. D., Betzel, R. F. & Misisic, B. Linking structure and function in macroscale brain networks. *Trends in Cognitive Sciences* **24**, 302–315, <https://doi.org/10.1016/j.tics.2020.01.008> (2020).
14. Thiebaut de Schotten, M., Foulon, C. & Nachev, P. Brain disconnections link structural connectivity with function and behaviour. *Nature Communications* **11**, 5094, <https://doi.org/10.1038/s41467-020-18920-9> (2020).
15. Sarwar, T., Tian, Y., Yeo, B. T. T., Ramamohanarao, K. & Zalesky, A. Structure-function coupling in the human connectome: A machine learning approach. *NeuroImage* **226**, 117609, <https://doi.org/10.1016/j.neuroimage.2020.117609> (2021).
16. Zamani Esfahlani, F., Faskowitz, J., Slack, J., Mišić, B. & Betzel, R. F. Local structure-function relationships in human brain networks across the lifespan. *Nature Communications* **13**, 2053, <https://doi.org/10.1038/s41467-022-29770-y> (2022).
17. Vincent, J. L. *et al.* Intrinsic functional architecture in the anaesthetized monkey brain. *Nature* **447**, 83–86, <https://doi.org/10.1038/nature05758> (2007).
18. Meunier, D. Hierarchical modularity in human brain functional networks. *Frontiers in Neuroinformatics* **3**, <https://doi.org/10.3389/neuro.11.037.2009> (2009).
19. van den Heuvel, M. P., Mandl, R. C., Kahn, R. S. & Hulshoff Pol, H. E. Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Human Brain Mapping* **30**, 3127–3141, <https://doi.org/10.1002/hbm.20737> (2009).
20. Greicius, M. D., Supekar, K., Menon, V. & Dougherty, R. F. Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral Cortex (New York, N.Y.: 1991)* **19**, 72–78, <https://doi.org/10.1093/cercor/bhn059> (2009).

21. Meunier, D., Lambiotte, R. & Bullmore, E. T. Modular and hierarchically modular organization of brain networks. *Frontiers in Neuroscience* **4**, 200, <https://doi.org/10.3389/fnins.2010.00200> (2010).
22. Power, J. D. *et al.* Functional network organization of the human brain. *Neuron* **72**, 665–678, <https://doi.org/10.1016/j.neuron.2011.09.006> (2011).
23. Bertolero, M. A., Yeo, B. T. T. & D'Esposito, M. The modular and integrative functional architecture of the human brain. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E6798–6807, <https://doi.org/10.1073/pnas.1510619112> (2015).
24. Diez, I. *et al.* A novel brain partition highlights the modular skeleton shared by structure and function. *Scientific Reports* **5**, 10532, <https://doi.org/10.1038/srep10532> (2015).
25. Sporns, O. & Betzel, R. F. Modular Brain Networks. *Annual Review of Psychology* **67**, 613–640, <https://doi.org/10.1146/annurev-psych-122414-033634> (2016).
26. Betzel, R. F. & Bassett, D. S. Multi-scale brain networks. *NeuroImage* **160**, 73–83, <https://doi.org/10.1016/j.neuroimage.2016.11.006> (2017).
27. Rosenthal, G. *et al.* Mapping higher-order relations between brain structure and function with embedded vector representations of connectomes. *Nature Communications* **9**, 2178, <https://doi.org/10.1038/s41467-018-04614-w> (2018).
28. Fukushima, M. & Sporns, O. Comparison of fluctuations in global network topology of modeled and empirical brain functional connectivity. *PLOS Computational Biology* **14**, e1006497, <https://doi.org/10.1371/journal.pcbi.1006497> (2018).
29. Puxeddu, M. G., Faskowitz, J., Sporns, O., Astolfi, L. & Betzel, R. F. Multi-modal and multi-subject modular organization of human brain networks. *NeuroImage* **264**, 119673, <https://doi.org/10.1016/j.neuroimage.2022.119673> (2022).
30. Churchland, P. S. & Sejnowski, T. J. *The Computational Brain*. Computational Neuroscience Series (A Bradford Book, 1992).
31. Betzel, R. F. *et al.* Multi-scale community organization of the human structural connectome and its relationship with resting-state functional connectivity. *Network Science* **1**, 353–373, <https://doi.org/10.1017/nws.2013.19> (2013).
32. Kolchinsky, A. *et al.* Multi-scale integration and predictability in resting state brain activity. *Frontiers in Neuroinformatics* **8**, <https://doi.org/10.3389/fninf.2014.00066> (2014).
33. Ashourvan, A., Telesford, Q. K., Verstynen, T., Vettel, J. M. & Bassett, D. S. Multi-scale detection of hierarchical community architecture in structural and functional brain networks. *PLOS ONE* **14**, e0215520, <https://doi.org/10.1371/journal.pone.0215520> (2019).
34. Petersen, S. & Sporns, O. Brain networks and cognitive architectures. *Neuron* **88**, 207–219, <https://doi.org/10.1016/j.neuron.2015.09.027> (2015).
35. Fernandez-Iriondo, I. *et al.* Brain mapping of behavioral domains using multi-scale networks and canonical correlation analysis. *Frontiers in Neuroscience* **16**, <https://doi.org/10.3389/fnins.2022.889725> (2022).
36. Park, B.-y. *et al.* Adolescent development of multiscale structural wiring and functional interactions in the human connectome. *Proceedings of the National Academy of Sciences* **119**, e2116673119, <https://doi.org/10.1073/pnas.2116673119> (2022).
37. Bonifazi, P. *et al.* Structure–function multi-scale connectomics reveals a major role of the fronto-striato-thalamic circuit in brain aging. *Human Brain Mapping* **39**, 4663–4677, <https://doi.org/10.1002/hbm.24312> (2018).
38. Bassett, D. S. *et al.* Hierarchical organization of human cortical networks in health and schizophrenia. *Journal of Neuroscience* **28**, 9239–9248, <https://doi.org/10.1523/JNEUROSCI.1929-08.2008> (2008).
39. Diez, I. *et al.* Enhanced prefrontal functional–structural networks to support postural control deficits after traumatic brain injury in a pediatric population. *Network Neuroscience* **1**, 116–142, https://doi.org/10.1162/NETN_a_00007 (2017).
40. Rasero, J. *et al.* Group-level progressive alterations in brain connectivity patterns revealed by diffusion-tensor brain networks across severity stages in alzheimer's disease. *Frontiers in Aging Neuroscience* **9**, 215, <https://doi.org/10.3389/fnagi.2017.00215> (2017).
41. Camino-Pontes, B. *et al.* Interaction information along lifespan of the resting brain dynamics reveals a major redundant role of the default mode network. *Entropy* **20**, 742, <https://doi.org/10.3390/e20100742> (2018).
42. beim Graben, P. *et al.* Metastable resting state brain dynamics. *Frontiers in Computational Neuroscience* **13**, 62, <https://doi.org/10.3389/fncom.2019.00062> (2019).
43. Gatica, M. *et al.* High-order interdependencies in the aging brain. *Brain Connectivity* **11**, 734–744, <https://doi.org/10.1089/brain.2020.0982> (2021).
44. Fernandez-Iriondo, I. *et al.* Small variation in dynamic functional connectivity in cerebellar networks. *Neurocomputing* **461**, 751–761, <https://doi.org/10.1016/j.neucom.2020.09.092> (2021).
45. Gatica, M. *et al.* High-order functional redundancy in ageing explained via alterations in the connectome in a whole-brain model. *PLOS Computational Biology* **18**, e1010431, <https://doi.org/10.1371/journal.pcbi.1010431> (2022).
46. He, C. *et al.* Structure–function connectomics reveals aberrant developmental trajectory occurring at preadolescence in the autistic brain. *Cerebral Cortex* **30**, 5028–5037, <https://doi.org/10.1093/cercor/bhaa098> (2020).
47. Van Essen, D. C. *et al.* The Human Connectome Project: a data acquisition perspective. *NeuroImage* **62**, 2222–2231, <https://doi.org/10.1016/j.neuroimage.2012.02.018> (2012).
48. Royer, J. *et al.* An Open MRI Dataset For Multiscale Neuroscience. *Scientific Data* **9**, 569, <https://doi.org/10.1038/s41597-022-01682-y> (2022).
49. Mansour L, S., Di Biase, M. A., Smith, R. E., Zalesky, A. & Seguin, C. Connectomes for 40,000 UK Biobank participants: A multi-modal, multi-scale brain network resource. *NeuroImage* **283**, 120407, <https://doi.org/10.1016/j.neuroimage.2023.120407> (2023).
50. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399, <https://doi.org/10.1038/nature11405> (2012).
51. Diez, I. & Sepulcre, J. Neurogenetic profiles delineate large-scale connectivity dynamics of the human brain. *Nature Communications* **9**, 3876, <https://doi.org/10.1038/s41467-018-06346-3> (2018).
52. Sepulcre, J. *et al.* Neurogenetic contributions to amyloid beta and tau spreading in the human cortex. *Nature Medicine* **24**, 1910–1918, <https://doi.org/10.1038/s41591-018-0206-4> (2018).
53. Arnatkeviciute, A., Fulcher, B. D. & Fornito, A. A practical guide to linking brain-wide gene expression and neuroimaging data. *NeuroImage* **189**, 353–367, <https://doi.org/10.1016/j.neuroimage.2019.01.011> (2019).
54. Bueichekú, E. *et al.* Central neurogenetic signatures of the visuomotor integration system. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 6836–6843, <https://doi.org/10.1073/pnas.1912429117> (2020).
55. Bueichekú, E. *et al.* Divergent connectomic organization delineates genetic evolutionary traits in the human brain. *Scientific Reports* **11**, 19692, <https://doi.org/10.1038/s41598-021-99082-6> (2021).
56. Diez, I. *et al.* Early-life trauma endophenotypes and brain circuit–gene expression relationships in functional neurological (conversion) disorder. *Molecular Psychiatry* **26**, 3817–3828, <https://doi.org/10.1038/s41380-020-0665-0> (2021).
57. Markello, R. D. *et al.* Standardizing workflows in imaging transcriptomics with the abagen toolbox. *eLife* **10**, e72129, <https://doi.org/10.7554/eLife.72129> (2021).
58. Jimenez-Marin, A. *et al.* Transcriptional signatures of synaptic vesicle genes define myotonic dystrophy type I neurodegeneration. *Neuropathology and Applied Neurobiology* **47**, 1092–1108, <https://doi.org/10.1111/nan.12725> (2021).
59. Arnatkeviciute, A., Markello, R. D., Fulcher, B. D., Misis, B. & Fornito, A. Toward Best Practices for Imaging Transcriptomics of the Human Brain. *Biological Psychiatry* **93**, 391–404, <https://doi.org/10.1016/j.biopsych.2022.10.016> (2023).
60. Rasero, J. *et al.* The neurogenetics of functional connectivity alterations in Autism: Insights from subtyping in 657 patients. *Biological Psychiatry* S0006–3223(23)01230–1, <https://doi.org/10.1016/j.biopsych.2023.04.014> (2023).

61. Babayan, A. *et al.* A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults. *Scientific Data* **6**, 180308, <https://doi.org/10.1038/sdata.2018.308> (2012).
62. Babayan, A. *et al.* Max Planck Institut Leipzig Mind-Brain-Body Dataset - LEMON. *Functional Connectomes Project International Neuroimaging Data-Sharing Initiative*, https://doi.org/10.15387/fcp_indi.mpi_lemon (2018).
63. Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology* **106**, 1125–1165, <https://doi.org/10.1152/jn.00338.2011> (2011).
64. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *NeuroImage* **31**, 968–980, <https://doi.org/10.1016/j.neuroimage.2006.01.021> (2006).
65. Zeighami, Y. *et al.* A comparison of anatomic and cellular transcriptome structures across 40 human brain diseases. *PLOS Biology* <https://doi.org/10.1371/journal.pbio.3002058> (2023).
66. Jimenez-Marin, A. *et al.* Brain Hierarchical Atlas 2 (BHA2). *Zenodo* <https://doi.org/10.5281/zenodo.8158914> (2023).
67. Craddock, R. C., James, G., Holtzheimer, P. E., Hu, X. P. & Mayberg, H. S. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping* **33**, 1914–1928, <https://doi.org/10.1002/hbm.21333> (2012).
68. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178, <https://doi.org/10.1038/nature18933> (2016).
69. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex (New York, N.Y.: 1991)* **28**, 3095–3114, <https://doi.org/10.1093/cercor/bhx179> (2018).
70. Smith, R. E., Tournier, J.-D., Calamante, F. & Connelly, A. SIFT2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography. *NeuroImage* **119**, 338–351, <https://doi.org/10.1016/j.neuroimage.2015.06.092> (2015).

Acknowledgements

This work would not have been possible without the tremendous effort of numerous colleagues with whom we have actively contributed over the past 8 years to various studies exploring the multi-scale nature vs. optimal partitions of structure-function multi-scale networks. The list of researchers includes: N. Aginako, C. Alonso-Montes, J.C. Arango-Lasprilla, P. beim Graben, M.P. Boisgontier, B. Camino-Pontes, Hu. Chen, He. Chen, R. Cofré, M. Desroches, D. Drijkoningen, X. Duan, F.E. Rosas, I. Escudero, M. Fernandez, I. Fernandez-Iriondo, I. Gabilondo, M. Gatica, J. Gooijers, X. Guo, C. He, X. Huang, L. Li, W. Liao, D. Marinazzo, B. Mateos, P.A.M. Mediano, M.A. Muñoz, L. Olabarrieta-Landa, P. Orío, S.P. Swinnen, L. Pauwels, J. Raseró, L. Remaki, S. Rodrigues, F.E. Rosas, W. Sheng, B. Sierra, L.Q. Uddin and J. Xiao. AJM is funded by a predoctoral contract from the Department of Education of the Basque Country (PRE-2019-1-0070). JMC, PB and AE are funded by Ikerbasque: The Basque Foundation for Science. JMC is funded by the Health Department of the Basque Country (grants 2022111031 and 2023111002). PB is funded by the Spanish Ministry for Science and Innovation (MICINN; grant reference PID2021-127163NB-I00, funded by the MCIN/AEI/10.13039/501100011033/FEDER, UE), and by Maratoia EITB (grant reference: BIO22/ALZ/010/BCB). AE is funded by the Spanish Ministry of Science and Innovation, grant RYC2021-032390-I.

Author contributions

J.M.C., P.B., I.D., A.E., S.S. and A.J.M. designed the project. A.J.M. performed the data analyses. S.S., P.B., and J.M.C. supervised the methodology. A.J.M. and J.M.C. made the figures. A.J.M. and J.M.C. drafted the manuscript. All authors wrote and reviewed 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/s41597-024-03060-2>.

Correspondence and requests for materials should be addressed to J.M.C.

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

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 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

© The Author(s) 2024