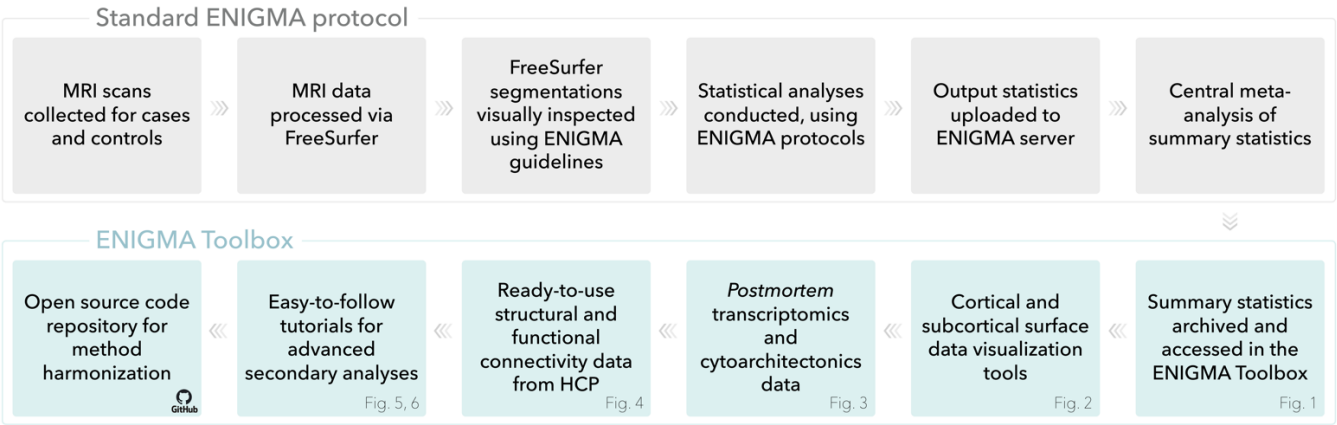
The ENIGMA Toolbox: Cross-disorder integration and multiscale neural contextualization of multisite neuroimaging datasets

Sara Larivière[1,✉], Casey Paquola[1], Bo-yong Park[1], Jessica Royer[1], Yezhou Wang[1], Oualid Benkarim[1], Reinder Vos de Wael[1], Sofie L. Valk[2,3], Sophia I. Thomopoulos[4], Matthias Kirschner[5], ENIGMA Consortium[6], Sanjay Sisodiya[7,8], Carrie R. McDonald[9], Paul M. Thompson[4], Boris C. Bernhardt[1,✉]

Among 'big data' initiatives, the ENIGMA (Enhancing NeuroImaging Genetics through Meta-Analysis) Consortium—a worldwide alliance of over 2,000 scientists diversified into over 50 Working Groups—has yielded some of the largest studies of the healthy and diseased brain. Integration of multisite datasets to assess transdiagnostic similarities and differences and to contextualize findings with respect to neural organization, however, have been limited. Here, we introduce the ENIGMA Toolbox, a Python/Matlab ecosystem for (*i*) accessing ENIGMA datasets, allowing for cross-disorder analysis, (*ii*) visualizing data on brain surfaces, and (*iii*) contextualizing findings at the microscale (*postmortem* cytoarchitecture and gene expression) and macroscale (structural and functional connectomes). Our Toolbox equips scientists with tutorials to explore molecular, histological, and network correlates of noninvasive neuroimaging markers of brain disorders. Moreover, our Toolbox bridges the gap between standardized data processing protocols and analytic workflows and facilitates cross-consortia initiatives. The Toolbox is documented and openly available at http://enigma-toolbox.readthedocs.io.

ENIGMA | multisite | multiscale | meta-analysis | mega-analysis | transcriptomics | connectivity



Introduction

The recent growth of data sharing practices and open science culture, increasingly fostered by major biomedical funding agencies encouraging grant recipients to provide full access to scientific data[1], has propelled international scientific collaborations. The fields of neuroimaging and genetics, particularly, have conquered 'big data' territory by making data and resources increasingly accessible to the community[2]. Access to large datasets enables the testing of novel hypotheses with adequate statistical power, driving new discoveries and enhancing reproducibility and generalizability of results[3, 4].

The ENIGMA (Enhancing NeuroImaging Genetics through Meta-Analysis) Consortium is one of the leading big data alliances that pools neuroimaging and genetic data from around the globe. Since its inception, ENIGMA has grown to a collaboration of thousands of

[1]Multimodal Imaging and Connectome Analysis Laboratory, McConnell Brain Imaging Centre, Montreal Neurological Institute and Hospital, McGill University, Montreal, QC, Canada; [2]Otto Hahn Research Group for Cognitive Neurogenetics, Max Planck Institute for Cognitive and Brain Sciences, Leipzig, Germany; [3]INM-7, FZ Jülich, Jülich, Germany; [4]Imaging Genetics Center, Mark & Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Marina del Rey, California, United States; [5]Department of Psychiatry, Psychotherapy and Psychosomatics, Psychiatric Hospital, University of Zurich, Switzerland; [6]http://enigma.ini.usc.edu; [7]Department of Clinical and Experimental Epilepsy, UCL Queen Square Institute of Neurology, London, UK; [8]Chalfont Centre for Epilepsy, Bucks, UK; [9]Department of Psychiatry, Center for Multimodal Imaging and Genetics, University of California San Diego, La Jolla, CA, United States. ✉e-mail: saratheriver@gmail.com; boris.bernhardt@mcgill.ca

scientists from over 45 countries[5]. Organized into several Working Groups (**FIG. 1A**), the ENIGMA Consortium has made important contributions to fundamental and clinical neuroscience, publishing some of the largest neuroimaging studies to date[6-15]. The success of the ENIGMA Consortium has heavily relied on standardized pipelines for data processing, quality control, and meta-analysis of widely used morphological measures. From these harmonized procedures, brain metrics (such as cortical thickness and subcortical volume) are extracted from raw neuroimaging data at each research center within a given Working Group and entered into site-specific linear models to test for case *vs.* control differences or to assess correlations with covariates of interest. Effect sizes and the heterogeneity of these effects across sites are then estimated via topic- and disease-specific meta-analyses, with the resulting aggregated statistical information being published and shared.

Through the sharing of these site-specific brain metrics or ensuing aggregated statistical maps[16], ENIGMA has set the stage for unprecedented analyses to compare findings across Working Groups and to contextualize findings across different scales of neural organization. Several such cross-disorder studies are underway, including a comparison of three neurodevelopmental disorders across 151 cohorts worldwide[17], and an initiative to relate structural brain abnormalities to cell-specific profiles of gene expression[18].

The neuroscience community, however, currently lacks standardized tools to analyze multicenter datasets beyond traditional structural MRI case-control meta- and mega-analyses. To fill this gap, we developed the ENIGMA Toolbox, an ecosystem for: (*i*) archiving, accessing, and integrating different ENIGMA-derived, or equivalently processed, datasets, (*ii*) visualizing data on cortical and subcortical surface models, and (*iii*) contextualizing neuroimaging findings across multiple scales of neural organization. We provide the ability to decode ENIGMA-type brain maps with respect to *postmortem* gene expression maps (from the Allen Human Brain Atlas[19]), *postmortem* cytoarchitecture (from the BigBrain project[20]), and structural as well as functional connectome properties (from the Human Connectome Project[21]). We provide several start-to-finish tutorials to show how these workflows can offer neurobiological and system-level insights into how regional effects, for example disease-related atrophy patterns, co-vary with transcriptomic, microstructural, and macrolevel properties, similar to recently published ENIGMA studies[18, 22].

Here, we present the ENIGMA Toolbox with ready-to-use and easy-to-follow code snippets. Our toolbox is available in Python and Matlab—two widely used languages in neuroimaging, neuroinformatics, and genetics communities—and is compatible with a range of subject-level, as well as meta- and mega-analytic datasets. Data and codes are openly accessible (http://github.com/MICA-MNI/ENIGMA) and complemented with expandable online documentation (http://enigma-toolbox.readthedocs.io).

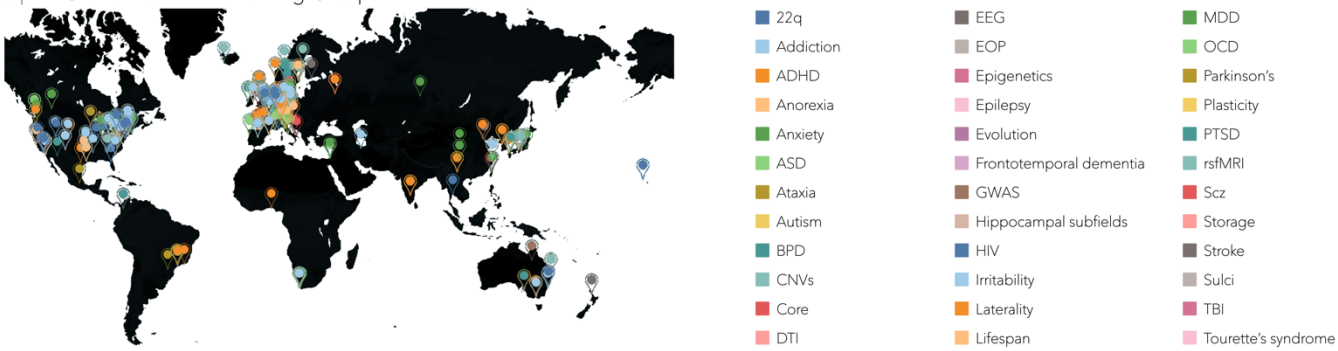
RESULTS

The ENIGMA Toolbox is an ecosystem composed of three modules. Each of these modules can be stand-alone or integrated with one another, allowing for greater flexibility, adaptiveness, and continuous development. We furthermore provide thoroughly documented workflows that users can easily adapt to their own datasets.

Data archiving and accessing

The ENIGMA Toolbox stores and accesses multiple ENIGMA datasets in a central repository. Given that data sharing practices can at times be challenging, in part due to privacy and regulatory protection, ENIGMA represents a practical alternative for standardized data processing and anonymized analysis of results (*i.e.*, meta-analysis) as well as the sharing of non-identifiable derivatives (*i.e.*, mega-analysis). Available datasets within our Toolbox consist of summary statistics from several ENIGMA Working Groups. The current release (*v*1.0.0) includes case-control summary statistics from eight Working Groups, including: 22.q11.2 deletion syndrome, attention deficit/hyperactivity disorder, autism spectrum disorder, bipolar disorder, epilepsy, major depressive disorder, obsessive-compulsive disorder, and schizophrenia (see **METHODS**). These datasets, obtained from standardized and quality-controlled protocols, represent morphological (*e.g.*, subcortical volume, cortical thickness, surface area) case-control effect sizes from previously published meta-analyses[6-8, 11-15] and can be retrieved using the `load_summary_stats()` function (**FIG. 1B, C**) and used for secondary analyses. Summary statistics from other Working Groups will be continuously added as they are published. As many ENIGMA groups have moved beyond meta-analysis to mega-analysis of subject-level data, the Toolbox is also compatible with subject-level data. As part of the ENIGMA Toolbox, we also provide example data from an individual site, accessible via `load_example_data()`. Alternatively, if users have generated their own summary statistics/subject-level

a | ENIGMA's worldwide Working Groups



b | Load summary statistics (*Python*)

```python
from enigmatoolbox.datasets import load_summary_stats

# Load summary statistics for ENIGMA-Epilepsy
sum_stats = load_summary_stats('epilepsy')

# Get case-control subcortical volume and cortical thickness tables
SV = sum_stats['SubVol_case_vs_controls_ltle']
CT = sum_stats['CortThick_case_vs_controls_ltle']

# Extract Cohen's d values
SV_d = SV['d_icv']
CT_d = CT['d_icv']
```

c | Load summary statistics (*Matlab*)

```matlab
addpath(genpath('/path/to/ENIGMA/matlab/'))

% Load summary statistics for ENIGMA-Epilepsy
sum_stats = load_summary_stats('epilepsy');

% Get case-control subcortical volume and cortical thickness tables
SV = sum_stats.SubVol_case_vs_controls_ltle;
CT = sum_stats.CortThick_case_vs_controls_ltle;

% Extract Cohen's d values
SV_d = SV.d_icv;
CT_d = CT.d_icv;
```

d | Cross-disorder effect (*Python*)

```python
from enigmatoolbox.cross_disorder import cross_disorder_effect

# Extract and visualize shared disorder effects
components, variance, names = cross_disorder_effect()
```

e | Cross-disorder effect (*Matlab*)

```matlab
% Extract and visualize shared disorder effects
[components, variance, ~, names] = cross_disorder_effect();
```

f | Shared effect across disorders

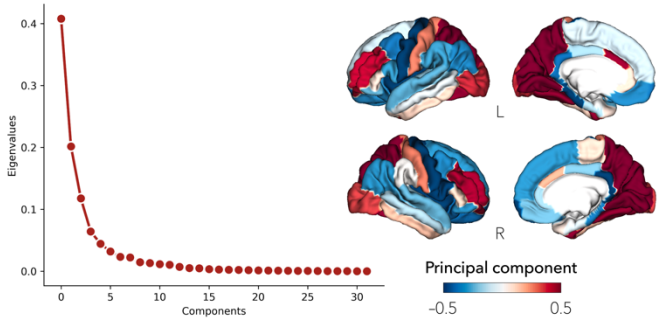


Figure 1. **Data archiving and accessing**. (**a**) World map of a subset of ENIGMA's Working Groups (color key). Each group consists of international teams of researchers and clinicians studying major brain diseases and conditions and contributing data to the ENIGMA consortium. Case-control summary statistics from published studies are archived in the ENIGMA Toolbox and easily accessible via simple Python (**b**) and Matlab (**c**) scripts. Summary statistics from the Epilepsy Working Group is shown as an example. Minimal (**d**) Python and (**e**) Matlab code snippets to identify transdiagnostic morphometric signatures. (**f**) Eigenvalues of each component are displayed in the scree plot. The principal component underlying the shared cross-disorder effect is projected to the surface template.

data that adhere to ENIGMA's harmonized processing and analysis protocols (http://enigma.ini.usc.edu/protocols/), or have in-house parcellated datasets, they can load their own data locally and still take advantage of every function our Toolbox has to offer. Online tutorials are suited for all dataset types.

To yield novel insights into brain structural abnormalities that are common or different across disorders, available summary statistics can also be harnessed to conduct cross-disorder analyses. From the `cross_disorder_effect()` function, users can explore shared and disease-specific morphometric signatures with two different approaches: (*i*) by applying a principal component analysis (PCA) to any combination of disease-specific summary statistics, resulting in shared latent components that can be used for further analysis (**Fig. 1D–F**), and (*ii*) by systematically cross-correlating patterns of brain structural abnormalities with every other set of available summary statistics, resulting in a correlation matrix (**Supplementary Fig. 1A–C**). Users can also upload local summary statistics, or equivalently processed data, for inclusion in the PCA or comparison against our database.

a | Surface data visualization (*Python*)

```python
from enigmatoolbox.utils.parcellation import parcel_to_surface
from enigmatoolbox.plotting import plot_cortical, plot_subcortical

# Map parcellated data to the surface
CT_d_fsa5 = parcel_to_surface(CT_d, 'aparc_fsa5')

# Project the results on the surface brain
plot_cortical(array_name=CT_d_fsa5, color_range=(-0.5, 0.5))
plot_subcortical(array_name=SV_d, color_range=(-0.5, 0.5))
```

b | Surface data visualization (*Matlab*)

```MATLAB
addpath(genpath('/path/to/ENIGMA/matlab/'))

% Map parcellated data to the surface
CT_d_fsa5 = parcel_to_surface(CT_d, 'aparc_fsa5');

% Project the results on the surface brain
figure, plot_cortical(CT_d_fsa5, 'color_range', [-0.5 0.5])
figure, plot_subcortical(SV_d, 'color_range', [-0.5 0.5])
```

c | Cortical thickness and subcortical volume decreases in left focal epilepsy

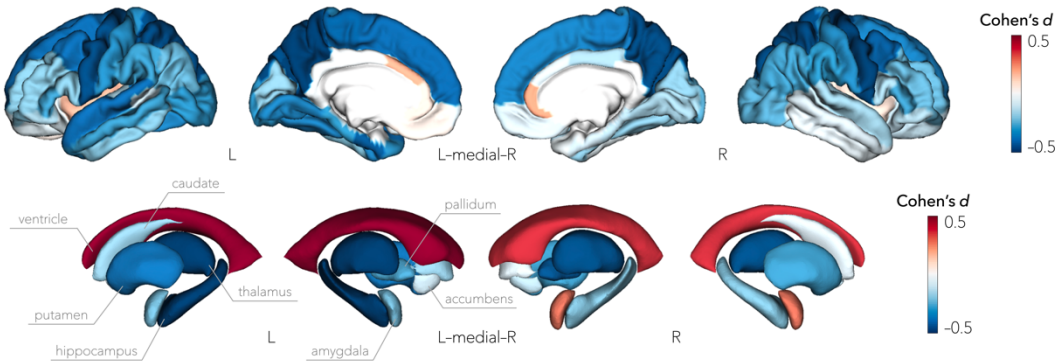


FIGURE 2. Cortical and subcortical surface visualization. Minimal (**a**) Python and (**b**) Matlab code snippets for plotting cortical thickness and subcortical volume deficits in individuals with left focal epilepsy. (**c**) Summary statistics (Cohen's *d*) comparing individuals with left focal epilepsy to healthy controls are projected to the surface templates. Profound gray matter atrophy can be visually appreciated in bilateral precentral gyrus, precuneus, thalamus, as well as left mesiotemporal regions, including the hippocampus.

Cortical and subcortical surface data visualization

The ENIGMA Toolbox provides functions to visualize cortical and subcortical data on surface models and generate publication-ready figures (**FIG. 2A, B**). To illustrate the `plot_cortical()` and `plot_subcortical()` functions, we projected cortical and subcortical gray matter atrophy in individuals with focal epilepsy relative to healthy controls to the surface templates (**FIG. 2C**). Beyond the mapping of gray matter atrophy, our surface visualization function is compatible with any neuroimaging data type, parcellation, and surface templates. Mapping to and from brain parcellations and the vertex-wise surface space can also be easily achieved using the `parcel_to_surface()` and `surface_to_parcel()` functions.

Multiscale contextualization

Transcriptomics data

The emergence of open databases for human transcriptomics yields new opportunities to associate macroscale neuroimaging findings with spatial variations at the molecular scale[19, 23-25]. The Allen Institute for Brain Science released the Allen Human Brain Atlas (AHBA)—a brain-wide gene expression atlas comprising microarray-derived measures from over 20,000 genes sampled across 3,702 spatially distinct tissue samples[19]. Using the abagen toolbox[26] and following guidelines established by Arnatkevičiūtė et al.[23], this large expression dataset was collapsed into cortical and subcortical regions of interest and combined across donors[23, 26, 27]. Genes that were consistently expressed across donors ($r \geq 0.2$, $n_{genes}=15{,}633$), can be easily obtained using the `fetch_ahba()` function (**FIG. 3A–C**). Additional details on processing and analytical choices, including parcellation compatibility and stability thresholds, are provided in the **METHODS**.

BigBrain data

The BigBrain, an ultra-high-resolution 3D reconstruction of a sliced and stained human brain[20], has supported quantitative analysis of cytoarchitecture. As part of the ENIGMA Toolbox, we characterized regional cytoarchitecture using statistical moments of staining profiles (see **METHODS**)[28]. Specifically, studying the mean intracortical staining across the mantle allows inferences on overall cellular density, whereas analysis of profile skewness indexes the distribution of cells across upper and lower layers of the cortex—a critical dimension of laminar differentiation.

a | Fetch disease-related gene expression data (*Python*)

```python
from enigmatoolbox.datasets import fetch_ahba, risk_genes

# Fetch gene expression data
genes = fetch_ahba()

# Get the names of epilepsy-related genes (Focal HS phenotype)
epilepsy_genes = risk_genes('epilepsy')['focalhs']

# Extract gene expression data for these Focal HS genes
epilepsy_gene_data = genes[genes.columns.intersection(epilepsy_genes)]
```

b | Fetch disease-related gene expression data (*Matlab*)

```matlab
addpath(genpath('/path/to/ENIGMA/matlab/'))

% Fetch gene expression data
genes = fetch_ahba();

% Get the names of epilepsy-related genes (Focal HS phenotype)
epilepsy_genes = risk_genes('epilepsy');
epilepsy_genes = epilepsy_genes.focalhs;

% Extract gene expression data for these Focal HS genes
epilepsy_gene_data = genes(:, contains(genes.Properties.VariableNames, epilepsy_genes));
```

c | Microarray expression data (Allen Human Brain Atlas)

	Label	A1BG	A1BG-AS1	A2M	A2ML1	…	ZZZ3
0	L_bankssts	0.4689	0.7002	0.367	0.3888	…	0.4824
1	L_caudalanteriorcingulate	0.6069	0.5607	0.5191	0.6108	…	0.5103
2	L_caudalmiddlefrontal	0.5121	0.6098	0.5163	0.4225	…	0.4612
…	…	…	…	…	…	…	…
79	Rpal	0.8007	0.2205	0.8173	0.5464	…	0.6012
80	Rput	0.6973	0.4478	0.5898	0.5592	…	0.6012
81	Rthal	0.3945	0.2251	0.6799	0.5665	…	0.5433

d | Focal HS-related gene expression

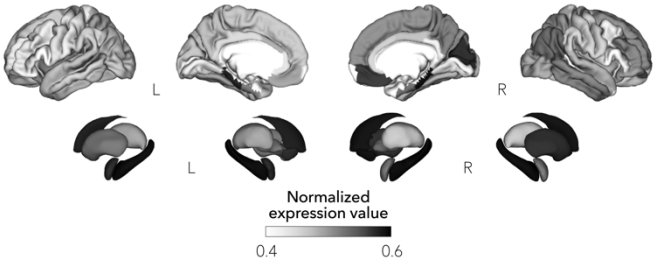


FIGURE 3. ENIGMA-friendly gene co-expression data. Minimal (**a**) Python and (**b**) Matlab code snippets to fetch disease-related gene co-expression data. Genes related to epilepsy, more specifically focal hippocampal sclerosis (Focal HS), are extracted as an example. (**c**) The complete microarray expression data can be easily accessed from the ENIGMA Toolbox and contains data from over 15,000 genes. (**d**) Disease-related gene co-expression data can be mapped to the surface templates; here, we displayed the average expression levels of Focal HS genes on the cortical and subcortical surface templates.

In addition to statistical moments, prior work has demonstrated robust evidence for a principal gradient of gradual cytoarchitectural variation running from primary sensory to limbic areas, mirroring spatial trends in laminar differentiation and cytoarchitectural variations[29-31]. To expand our BigBrain contextualization module, we also incorporated this microstructural similarity gradient to describe a sensory-fugal transition in intracortical microstructure (see **METHODS**). Stratifying cortical findings relative to this gradient could, for example, test whether patterns of changes are conspicuous in cortices with marked laminar differentiation (*e.g.*, sensory and motor cortices) or in those with subtle laminar differentiation (*e.g.*, limbic cortices).

Cytoarchitectural types

To further describe microscale cortical organization, the ENIGMA Toolbox includes a digitized parcellation of the von Economo and Koskinas cytoarchitectonic maps of the human cerebral cortex[32, 33]. From this mapping, five different structural types of cerebral cortex are recorded: *i*) agranular (thick with large cells but sparse layers II and IV), *ii*) frontal (thick but not rich in cellular substance, visible layers II and IV), *iii*) parietal (thick and rich in cells with dense layers II and IV but small and slender pyramidal cells), *iv*) polar (thin but rich in cells, particularly in granular layers), and *v*) granular or koniocortex (thin but rich in smalls cells, even in layer IV, and a rarified layer V)[34].

Connectivity data

Neuroimaging, particularly with functional and diffusion MRI, has become a leading tool to characterize human brain network organization *in vivo* and to identify network alterations in brain disorders. Although ongoing efforts in ENIGMA and beyond are beginning to coordinate analyses of resting-state functional[35-37] and diffusion[38-40] MRI data, connectivity measures remain sparse within the consortium. As an alternative, our Toolbox leverages high-resolution structural (derived from diffusion-weighted tractography) and functional (derived from resting-state functional MRI) connectivity data from a cohort of unrelated healthy adults from the Human Connectome Project (HCP) initiative[21]. Preprocessed cortico-cortical, subcortico-cortical, and subcortico-subcortical functional and structural connectivity matrices can be easily retrieved using the `load_fc()` and `load_sc()` functionalities (**FIG. 4A–C**). Connectivity matrices were parcellated according to several atlases, including the Desikan-Killiany atlas, and can thus be readily combined with any ENIGMA-derived, or other parcellated, datasets[27]. Details on subject inclusion, data preprocessing, and matrix generation are provided in the **METHODS**.

Analytical workflows

a | Load connectivity data (*Python*)

```python
from enigmatoolbox.datasets import load_sc, load_fc

# Load cortico-cortical functional connectivity data
fc_ctx, fc_ctx_labels, fc_sctx, fc_sctx_labels = load_fc()

# Load cortico-cortical structural connectivity data
sc_ctx, sc_ctx_labels, sc_sctx, sc_sctx_labels = load_sc()
```

b | Load connectivity data (*Matlab*)

```matlab
addpath(genpath('/path/to/ENIGMA/matlab/'))

% Load cortico-cortical functional connectivity data
[fc_ctx, fc_ctx_labels, fc_sctx, fc_sctx_labels] = load_fc();

% Load cortico-cortical structural connectivity data
[sc_ctx, sc_ctx_labels, sc_sctx, sc_sctx_labels] = load_sc();
```

c | Functional and structural connectivity matrices

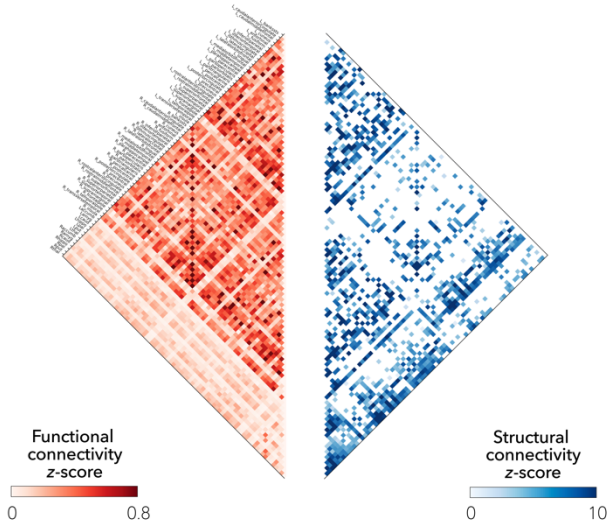


FIGURE 4. High-resolution connectivity data sharing and exploiting. Minimal (**a**) Python and (**b**) Matlab code snippets to load preprocessed functional and structural connectivity matrices. (**c**) Available connectivity data within the ENIGMA Toolbox include cortico-cortical, subcortico-cortical, and subcortico-subcortical connectivity matrices. Matrices are unthresholded and parcellated according to the Desikan-Killiany atlas[29].

As of the current release (*v*1.0.0), the ENIGMA Toolbox comprises two neural scales for the contextualization of findings: (*i*) using microscale properties, namely gene expression and cytoarchitecture, and (*ii*) using macroscale network models, such as regional hub susceptibility analysis and disease epicenter mapping. Albeit often restricted to individual site studies, similar approaches have been used to study microstructural organization in healthy[29, 41, 42] and diseased[43] brains, and to model network-level patterns of disease-related atrophy[44-46]. To ease programming and maximize transparency, analytic workflows are accompanied by comprehensive tutorials and visual assessment checkpoints. As proofs of concept, we demonstrate ready-to-use, easy-to-follow, and validated secondary analysis tutorials to relate patterns of gray matter atrophy in individuals with left focal epilepsy to transcriptomic, histological, and normative connectome properties.

Transcriptomics contextualization of findings

Motivated by the growing body of research relating gene expression to diverse properties of macroscale brain organization, toolbox users can import the AHBA microarray expression dataset and visualize brain maps of gene expression levels. This tool can also be used to identify genes that are spatially correlated with a given brain map (*e.g.*, a disease-related atrophy map). Moreover, based on reports of recently published genome-wide association studies (GWAS)[47-53], users can extract the most likely genes associated to significant genome-wide loci across a range of disorders. From these sets of genes, users can generate disease-specific gene expression maps to contextualize and decode neuroimaging findings with transcriptomics data. To illustrate the `risk_genes()` function, we selected genes related to focal epilepsy with hippocampal sclerosis[51] and displayed their average expression levels on the cortical and subcortical surface templates (**FIG. 3D**). In this example, we can observe higher gene co-expression levels of epilepsy-related genes in mesiotemporal lobe regions, overlapping with regions of profound atrophy in individuals with left focal epilepsy. Related efforts are underway to map the effects of common (single nucleotide) variants in the genome on brain structure, using visualization platforms such as ENIGMA-Vis (https://enigma-brain.org/enigmavis/)[54] and the Oxford Brain Imaging Genetics browser (http://big.stats.ox.ac.uk/).

Histological contextualization of findings

Applied conjointly with ENIGMA datasets, BigBrain-derived intracortical profile information (*i.e.*, the statistical moments and the principal gradient of microstructure differentiation) offers two complementary approaches to situate ENIGMA-type cortical findings with respect to histological findings. For the former approach, users can feed a thresholded cortical map—for instance highlighting areas of significant structural abnormalities in patients *vs.* controls—into the `bb_moments_raincloud()` function to contextualize macroscale features with respect to intensity profiles reflecting microstructural composition (*e.g.*, cellular

density, cellular distribution asymmetry) along cortical columns. This quantitative approach has notably been used to guide boundary definition and fingerprint cytoarchitecture in studies of *postmortem* data[55, 56]. Alternatively, thresholded (or unthresholded) cortical maps can be fed into the `bb_gradient_plot()` function, which discretizes the principal microstructure similarity gradient into five equally-sized bins and averages surface findings within each gradient bin. From this, neuroimaging data can be embedded into the gradient space, allowing users to make inferences about the underlying microstructural hierarchy of, for instance, atrophied regions. Contextualizing gray matter atrophy patterns in left focal epilepsy with respect to histological properties, we were able to demonstrate that atrophy predominantly affected cortical regions with greater, and more evenly distributed, cellular densities across upper and lower layers of the cortex (**SUPPLEMENTARY FIG. 2A–D**)—areas located towards the sensory apex of the cytoarchitectural gradient (**SUPPLEMENTARY FIG. 3A–D**).

Cytoarchitectornic contextualization of findings

Integration of cytoarchitecture with *in vivo* neuroimaging can consolidate microstructural differentiation with macroscale-level effects. As part of the Toolbox, users can also leverage a digitized cytoarchitectonic atlas from seminal *postmortem* work by von Economo and Koskinas[33]. From the `economo_koskinas_spider()` function, users can apply this well-established decomposition to summarize cortex-wide effects (*e.g.*, a disease-related atrophy map) and assess relationships to distinct cytoarchitectonic classes. To ease interpretability, cytoarchitectonic classification of findings are also displayed in a spider plot. To highlight how atrophy can vary with distinct architectonic cortical types, case-control Cohen's *d* measures, representing atrophy in left focal epilepsy, were averaged within each von Economo and Koskinas cytoarchitectonic class. Projecting the results in a radar (spider) plot revealed that the agranular cortex was most affected by atrophy in left focal epilepsy (**SUPPLEMENTARY FIG. 4A–C**).

Hub susceptibility model

Normative structural and functional connectomes hold valuable information for relating macroscopic brain network organization to patterns of disease-related atrophy (**FIG. 5A, B**). Prior work studying network underpinnings of morphological abnormalities in neurodegenerative and psychiatric disorders has demonstrated that hubs (*i.e.*, brain regions with many connections) typically show greater atrophy than locally-connected peripheral nodes[57, 58]. Within the ENIGMA Epilepsy Working Group, we recently tested this hypothesis using data from 1,021 individuals with epilepsy and 1,564 healthy controls, and also showed that atrophy preferentially colocalized with highly interconnected hub regions in the common epilepsies[22]. A series of follow-up ENIGMA studies are currently underway to assess these network-level effects in other disorders.

Building on the above-described functions, we can first derive weighted degree centrality maps from functional (or structural) connectivity data by computing the sum of all weighted cortico- and subcortico-cortical connections for every region, with higher degree centrality denoting hub regions (**FIG. 5C**). Spatial similarity between atrophy patterns (obtained from individuals with left focal epilepsy as an example) and hub distributions can then be compared through correlation analysis (and statistically assessed via spin permutation tests; see below), revealing that profound atrophy implicates functional cortico- and subcortico-cortical hubs more strongly than nonhub regions (**FIG. 5D**).

Disease epicenter model

To further investigate whether disease-related morphological abnormalities (*e.g.*, atrophy) follow overarching principles of connectome organization, one can identify disease epicenters. Disease epicenters are regions whose functional and/or structural connectivity profile spatially resembles a given disease-related atrophy map[22, 44-46]. Hence, disease epicenters can be identified by spatially correlating every region's healthy functional and/or structural connectivity profiles to whole-brain atrophy patterns in a given disease. This approach must be repeated systematically across the whole brain, assessing the statistical significance of the spatial similarity of every region's functional and/or structural connectivity profiles to disease-specific abnormality maps with spatial permutation tests. Cortical and subcortical epicenter regions can then be identified if their connectivity profiles are significantly correlated with the disease-specific abnormality map. Regardless of its atrophy level, a cortical or subcortical region could potentially be an epicenter if it is (*i*) strongly connected to other high-atrophy regions and (*ii*) weakly connected to low-atrophy regions. Moreover, disease epicenters do not necessarily represent hub regions, but may rather be connected to them (*i.e.*, so-called 'feeder nodes', which directly link peripheral nodes to hubs). As in prior work[22], this approach suggests that patterns of atrophy in left

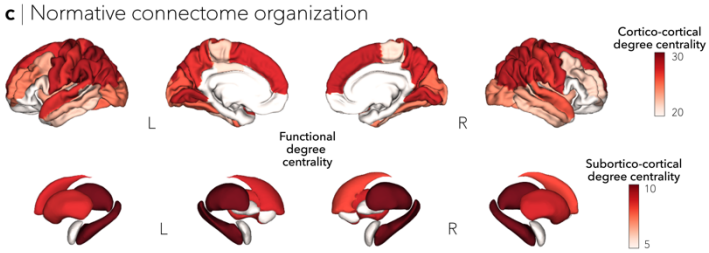
a | Hub susceptibility model (*Python*)

```python
import numpy as np

# Remove subcortical values corresponding to the ventricles
SV_d_noVent = SV_d.drop([np.where(SV['Structure'] == 'LLatVent')[0][0],
                         np.where(SV['Structure'] == 'RLatVent')[0][0]])
SV_d_noVent = SV_d_noVent.reset_index(drop=True)

# Compute weighted degree centrality measures
fc_ctx_dc = np.sum(fc_ctx, axis=0)
fc_sctx_dc = np.sum(fc_sctx, axis=1)

# Perform spatial correlations between hubs and Cohen's d
fc_ctx_r = np.corrcoef(fc_ctx_dc, CT_d)
fc_sctx_r = np.corrcoef(fc_sctx_dc, SV_d_noVent)
```

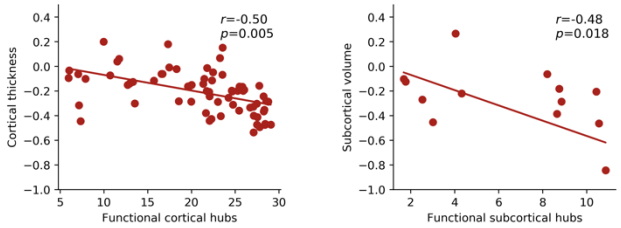
b | Hub susceptibility model (*Matlab*)

```matlab
addpath(genpath('/path/to/ENIGMA/matlab/'))

% Remove subcortical values corresponding the ventricles
SV_d_noVent = SV_d;
SV_d_noVent([find(strcmp(SV.Structure, 'LLatVent')); ...
             find(strcmp(SV.Structure, 'RLatVent'))], :) = [];

% Compute weighted degree centrality measures
fc_ctx_dc  = sum(fc_ctx);
fc_sctx_dc = sum(fc_sctx, 2);

% Perform spatial correlations between hubs and Cohen's d
fc_ctx_r  = corrcoef(fc_ctx_dc, CT_d);
fc_sctx_r = corrcoef(fc_sctx_dc, SV_d_noVent);
```

c | Normative connectome organization



d | Correlations between hubs and atrophy in left focal epilepsy



e | Permutation testing (*Python*)

```python
from enigmatoolbox.permutation_testing import spin_test, shuf_test

# Spin permutation testing for two cortical maps
fc_ctx_p, fc_ctx_d = spin_test(fc_ctx_dc, CT_d, null_dist=True)

# Shuf permutation testing for two subcortical maps
fc_sctx_p, fc_sctx_d = shuf_test(fc_sctx_dc, SV_d_noVent, null_dist=True)
```

f | Permutation testing (*Matlab*)

```matlab
addpath(genpath('/path/to/ENIGMA/matlab/'))

% Spin permutation testing for two cortical maps
[fc_ctx_p, fc_ctx_d]   = spin_test(fc_ctx_dc, CT_d);

% Shuf permutation testing for two subcortical maps
[fc_sctx_p, fc_sctx_d] = shuf_test(fc_sctx_dc, SV_d_noVent);
```

g | Permutation testing for hub-atrophy spatial correlations

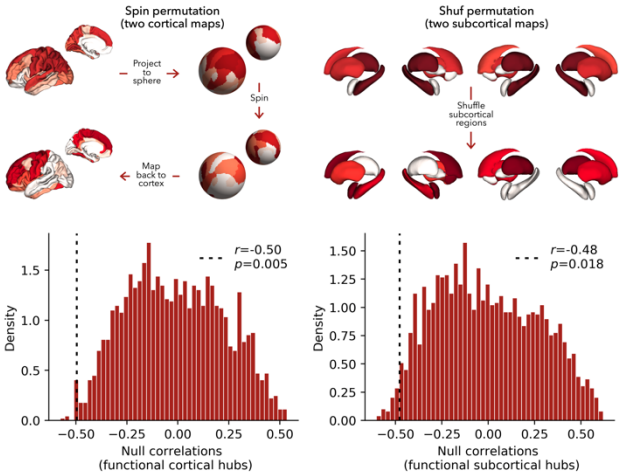


FIGURE 5. Advanced analytical workflows: hub susceptibility models and spin permutation testing. Minimal (**a**) Python and (**b**) Matlab code snippets to assess relationships between hubs and patterns of atrophy. (**c**) Functional degree centrality, derived from the HCP dataset, was used to identify the spatial distribution of hub regions. (**d**) Strong negative associations between cortical thickness/subcortical volume Cohen's d values and cortico-/subcortico-cortical degree centrality were observed, indicating that atrophy preferentially colocalized with hub regions. Minimal (**e**) Python and (**f**) Matlab code snippets to assess statistical significance of two surface maps while preserving spatial autocorrelation. (**g**) A schematic of the spin and shuf permutation frameworks for cortical and subcortical maps, respectively. The null distributions of correlations are shown; the empirical (*i.e.*, original) correlation coefficients and associated spin permuted p-values are indexed by the dashed line.

focal epilepsy are anchored to the connectivity profiles of mesiotemporal regions (**SUPPLEMENTARY FIG. 5A–C**).

Statistical comparison of spatial maps

The intrinsic spatial smoothness of brain maps—where data from neighboring regions are not statistically independent of each other—violates the underlying assumptions of several inferential statistical tests and consequently inflates the apparent significance of their spatial correlation, unless more sophisticated tests are used. To overcome these shortcomings and minimize Type I error, our toolbox includes non-parametric spatial permutation models to assess statistical significance while preserving the spatial autocorrelation of brain maps[59, 60]. With such functionality, for instance `spin_test()`, the spatial coordinates of the surface data are projected onto the surface spheres and randomly rotated to generate surface maps with randomized

topography but identical spatial autocorrelation. The empirical (*i.e.*, real) correlation coefficient can then be compared against the null distribution determined by the ensemble of correlation coefficients comparing spatially permuted surface maps. Spatial correspondence between two subcortical surface maps can be examined via a standard non-parametric null model, namely `shuf_test()`, where subcortical labels are randomly shuffled as opposed to being projected onto spheres. To illustrate these functions, we assessed the statistical significance of cortical and subcortical hub-atrophy correlations in left focal epilepsy (**FIG. 5E, F**) and displayed the empirical correlation coefficients onto the null distributions of permuted correlations (**FIG. 5G**).

DISCUSSION

The ENIGMA Toolbox is an integrated ecosystem that dovetails with ENIGMA's standardized data processing and meta-analysis strategies for integration, visualization, and contextualization of multisite results. Our Toolbox relies on a simple but efficient codebase for exploring and analyzing big data, aiming to facilitate and homogenize follow-up analyses of ENIGMA, or other, MRI datasets around the globe. At its core, the Toolbox is composed of three modules for (*i*) archiving, accessing, and integrating case-control meta-analytic datasets from specialized Working Groups, (*ii*) cortical and subcortical surface data visualization, and (*iii*) contextualizing findings based on transcriptomics, cytoarchitectonics, and connectivity data. Owing to its comprehensive tutorials, detailed functionality descriptions, and visual reports, our Toolbox is accessible to researchers and clinicians without extensive programming expertise within and beyond ENIGMA itself.

Enriching *in vivo* morphological correlates with *postmortem* microstructural information can deepen our understanding of the molecular and cellular underpinnings of healthy and diseased brain organization. To illustrate such microscale contextualization, we provide several tutorial examples that reference ENIGMA-type maps of gray matter atrophy in individuals with epilepsy against *postmortem* gene co-expression and histological measures. Based on the gene expression atlas from the Allen Institute for Brain Science[19], which compiles information on transcription of thousands of genes across the adult brain, users can compare spatial patterns of brain-wide gene expression to MRI-derived neuroanatomy measures. Importantly, microarray expression data within our toolbox were processed according to recommendations for best practice summarized in Arnatkevičiūtė and colleagues using the open-access abagen toolbox (https://abagen.readthedocs.io/en/stable/)[26], facilitating comparisons of findings across studies. Prior neuroimaging studies have already identified specific transcriptomic signatures of cortical morphometry in early brain development[61], cortical anatomy changes in youth with known genomic dosage variations[62], myeloarchitectural development in adolescence[63], and autism pathophysiology[64], bridging microstructural and macroscopic scales of brain organization (for a review, see [24]). Similarly, in a large-scale collaborative effort involving six ENIGMA Working Groups, Patel and colleagues correlated brain-wide cell-specific gene expression with group differences in cortical thickness, revealing shared neurobiological processes that underlie morphological phenotypes of multiple psychiatric disorders[18]. Moreover, as part of our toolbox, we also made the digital cell body-stained BigBrain[20] and von Economo and Koskinas cytoarchitectural atlas[32, 33] easily accessible and compatible with ENIGMA neuroimaging datasets. From the 3D histological BigBrain, several groups have made promising headway into mapping the patterns of cortical laminar architecture[65], exploring histological underpinnings of MRI-based thickness gradients in sensory and motor cortices[66], and identifying a sensory-fugal axis of microstructural differentiation[29]. On the other hand, histological atlases, such as the one by von Economo and Koskinas, are invaluable for linking brain microstructure to functional localization[34]. The digitized parcellation of von Economo and Koskinas cytoarchitectural types, thus, enables users to speculate on the underlying cytoarchitectural composition of, for instance, structurally abnormal areas in specific diseases. When combined, transcriptomic and cytoarchitectonic decoding can embed ENIGMA findings in a rich neurobiological context and yield potentially novel insights into the etiology of several brain disorders.

At the macroscopic level, network connectivity offers a vantage point to quantify brain reorganization in diseases that are increasingly being conceptualized as network disorders[67-69]. To exemplify the potential of macroscale network modeling, our toolbox provides detailed tutorials on how to relate ENIGMA-based surface maps to normative connectome properties derived from functional and diffusion MRI. Building on prior neurodegenerative[46, 57] and psychiatric[70] research, as well as recent work from our group[22], Toolbox users can build hub susceptibility models to assess the vulnerability of highly connected network hubs to disease-related effects. Given their role in integrative

processing, as well as their high topological value and biological cost, hub regions have been hypothesized to be preferentially susceptible to diverse pathological perturbations[71, 72]. Consequently, hub decoding represents an invaluable tool to interrogate whether heightened connectivity, atrophy, or metabolism properties in hubs relate to disease-specific processes. Complementing the hub susceptibility approach, another network modeling functionality available within our Toolbox is disease epicenter mapping, an analysis technique originally developed to study the spread of atrophy in neurodegenerative diseases[73]. Subsequent *in vivo* neuroimaging research provided evidence that network connectivity may herald patterns of gray matter atrophy in schizophrenia[44], Parkinson's disease[46, 74], and frontotemporal dementia syndromes[45], each linking greater atrophy to the connectivity profiles of distinct epicenters. Also applied to the common epilepsies as part of an ENIGMA-Epilepsy secondary project[22], this approach identified mesiotemporal epicenters in temporal lobe epilepsy and subcortico-cortical epicenters in generalized epilepsy, regions known to be involved in pathophysiology of each syndrome[75-79]. Combined, these two network models can significantly advance our understanding of how connectome architecture relates to morphological abnormalities across a range of disorders. Indeed, as with the microscale contextualization, our macroscale hub- and epicenter-based analytic pipelines can be applied to any ENIGMA-obtained datasets. Network models can be further enriched with microstructural properties to inject multiscale information into cortical and subcortical morphometric findings[80].

To enhance interpretability and avoid 'black-box' solutions, datasets, codes, and functionalities within our Toolbox are openly accessible and thoroughly documented. We hope that these efforts accelerate research and increase reliability and reproducibility. Notably, the modular architecture of our Toolbox allows for continuous development of analytical functionalities and tutorials. Future planned releases are poised to embrace new scientific approaches as they are published, adapting to new datasets (*e.g.*, PsychEncode consortium[81]), modalities (*e.g.*, resting-state functional MRI), and analytic pipelines (*e.g.*, structural covariance network analysis). Extension of the ENIGMA Toolbox to increase versatility of secondary analyses to additional brain parcellations, as well as vertex- and voxel-wise space, is already part of the development roadmap and will be updated to accommodate users' requests. Integration of analytic methods from users around the globe is supported and encouraged to maximize the contribution, reusability, and adaptability of any neuroimaging datasets.

In closing, by bridging the gap between pre-established data processing protocols and several analytic workflows, we hope that the ENIGMA Toolbox facilitates neuroscientific contextualization of results and cross-consortia initiatives. We are eager for researchers and clinicians to test hypotheses beyond traditional case-control comparisons. We hope that our platform will lead to novel and harmonized analyses in global neuroimaging initiatives.

Acknowledgements

Many scientists contributed to the development of ENIGMA but did not take part in the writing of this report. A full list of contributors to ENIGMA is available here: http://enigma.ini.usc.edu/about-2/consortium/members/.

The authors would like to express their gratitude to the open science initiatives that made this work possible: (*i*) The ENIGMA Consortium (core funding for ENIGMA was provided by the NIH Big Data to Knowledge (BD2K) program under consortium grant U54 EB020403 to P.M.T.), (*ii*) The Allen Human Brain Atlas, and (*iii*) The Human Connectome Project (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Funding

S.L. acknowledges funding from Fonds de la Recherche du Québec – Santé (FRQ-S) and the Canadian Institutes of Health Research (CIHR). C.P. was funded through a postdoctoral FRQ-S fellowship. O.B. was funded by a Healthy Brains for Healthy Lives (HBHL) postdoctoral fellowship. B.-y.P. was funded by the National Research Foundation of Korea (NRF-2020R1A6A3A03037088), Molson Neuro-Engineering fellowship from the Montreal Neurological Institute and Hospital, and FRQ-S. J.R. was supported by CIHR. R.V.d.W. was funded by studentships from the Savoy Foundation for Epilepsy and the Richard and Ann Sievers award. S.L.V. was supported by the Otto Hahn award of the Max Planck Society. M.K. acknowledges funding from the Swiss National Science Foundation (P2SKP3_178175). C.R.M. acknowledges funding from the National Institutes of Health (NINDS R01NS065838 and R21 NS107739). B.C.B. acknowledges research funding from

the SickKids Foundation (NI17-039), the Natural Sciences and Engineering Research Council of Canada (NSERC; Discovery-1304413), CIHR (FDN-154298), Azrieli Center for Autism Research (ACAR), an MNI-Cambridge collaboration grant, salary support from FRQ-S (Chercheur-Boursier), BrainCanada, the Helmholtz BigBrain Analytics and Learning Lab (Hiball), and the Canada Research Chairs (CRC) Program.

Author contributions

Core developers: S.L., B.C.B.
Toolbox beta testing: B.-Y.P., J.R., C.P., S.L.V., Y.W, M.K.
Writing: S.L., B.C.B.; revised and approved by other listed co-authors.

Declaration of interest

The authors declare no competing interests. P.M.T. received partial grant support from Biogen, Inc., and consulting payments from Kairos Venture Capital, Inc., for work unrelated to ENIGMA and this manuscript.

References

1. Poline, J.-B. et al. Data sharing in neuroimaging research. *Frontiers in neuroinformatics* **6**, 9 (2012).
2. Poldrack, R.A. & Gorgolewski, K.J. Making big data open: data sharing in neuroimaging. *Nat Neurosci* **17**, 1510-1517 (2014).
3. Medland, S.E., Jahanshad, N., Neale, B.M. & Thompson, P.M. Whole-genome analyses of whole-brain data: working within an expanded search space. *Nature neuroscience* **17**, 791-800 (2014).
4. Poldrack, R.A. et al. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature reviews neuroscience* **18**, 115 (2017).
5. Thompson, P.M. et al. ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl Psychiatry* **10**, 100 (2020).
6. Boedhoe, P.S.W. et al. Cortical Abnormalities Associated With Pediatric and Adult Obsessive-Compulsive Disorder: Findings From the ENIGMA Obsessive-Compulsive Disorder Working Group. *Am J Psychiatry* **175**, 453-462 (2018).
7. Hibar, D.P. et al. Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. *Mol Psychiatry* **23**, 932-942 (2018).
8. Hoogman, M. et al. Brain Imaging of the Cortex in ADHD: A Coordinated Analysis of Large-Scale Clinical and Population-Based Samples. *Am J Psychiatry* **176**, 531-542 (2019).
9. Logue, M.W. et al. Smaller Hippocampal Volume in Posttraumatic Stress Disorder: A Multisite ENIGMA-PGC Study: Subcortical Volumetry Results From Posttraumatic Stress Disorder Consortia. *Biol Psychiatry* **83**, 244-253 (2018).
10. Mackey, S. et al. Mega-Analysis of Gray Matter Volume in Substance Dependence: General and Substance-Specific Regional Effects. *Am J Psychiatry* **176**, 119-128 (2019).
11. Schmaal, L. et al. Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Mol Psychiatry* **22**, 900-909 (2017).
12. Sun, D. et al. Large-scale mapping of cortical alterations in 22q11.2 deletion syndrome: Convergence with idiopathic psychosis and effects of deletion size. *Mol Psychiatry* **25**, 1822-1834 (2020).
13. van Erp, T.G.M. et al. Cortical Brain Abnormalities in 4474 Individuals With Schizophrenia and 5098 Control Subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis (ENIGMA) Consortium. *Biol Psychiatry* **84**, 644-654 (2018).
14. van Rooij, D. et al. Cortical and Subcortical Brain Morphometry Differences Between Patients With Autism Spectrum Disorder and Healthy Individuals Across the Lifespan: Results From the ENIGMA ASD Working Group. *Am J Psychiatry* **175**, 359-369 (2018).
15. Whelan, C.D. et al. Structural brain abnormalities in the common epilepsies assessed in a worldwide ENIGMA study. *Brain* **141**, 391-408 (2018).
16. Thompson, P.M. et al. The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav* **8**, 153-182 (2014).
17. Boedhoe, P.S. et al. Subcortical brain volume, regional cortical thickness, and cortical surface area across disorders: Findings from the ENIGMA ADHD, ASD, and OCD working

groups. *American Journal of Psychiatry* **177**, 834-843 (2020).
18. Patel, Y. et al. Virtual histology of cortical thickness and shared neurobiology in 6 psychiatric disorders. *Jama Psychiatry* (2020).
19. Hawrylycz, M.J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391-399 (2012).
20. Amunts, K. et al. BigBrain: an ultrahigh-resolution 3D human brain model. *Science* **340**, 1472-1475 (2013).
21. Van Essen, D.C. et al. The Human Connectome Project: a data acquisition perspective. *Neuroimage* **62**, 2222-2231 (2012).
22. Larivière, S. et al. Network-based atrophy modeling in the common epilepsies: A worldwide ENIGMA study. *Sci Adv* **6** (2020).
23. Arnatkeviciute, A., Fulcher, B.D. & Fornito, A. A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage* **189**, 353-367 (2019).
24. Fornito, A., Arnatkevičiūtė, A. & Fulcher, B.D. Bridging the gap between connectome and transcriptome. *Trends in Cognitive Sciences* **23**, 34-50 (2019).
25. Gorgolewski, K.J. et al. NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Front Neuroinform* **9**, 8 (2015).
26. Markello, R., Shafiei, G., Y.-Q., Z. & Misic, B. abagen: A toolbox for the Allen Brain Atlas genetics data. *Zenodo* (2020).
27. 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 (2006).
28. Papoulis, A. & Pillai, S.U. Probability, random variables, and stochastic processes. (Tata McGraw-Hill Education, 2002).
29. Paquola, C. et al. Microstructural and functional gradients are increasingly dissociated in transmodal cortices. *PLoS biology* **17**, e3000284 (2019).
30. Sanides, F. The cyto-myeloarchitecture of the human frontal lobe and its relation to phylogenetic differentiation of the cerebral cortex. *Journal fur Hirnforschung* **7**, 269 (1964).
31. Mesulam, M.-M. From sensation to cognition. *Brain: a journal of neurology* **121**, 1013-1052 (1998).
32. Scholtens, L.H., de Reus, M.A., de Lange, S.C., Schmidt, R. & van den Heuvel, M.P. An mri von economo–koskinas atlas. *NeuroImage* **170**, 249-256 (2018).
33. von Economo, C.F. & Koskinas, G.N. Die cytoarchitektonik der hirnrinde des erwachsenen menschen. (J. Springer, 1925).
34. Triarhou, L.C. The Economo-Koskinas atlas revisited: cytoarchitectonics and functional context. *Stereotactic and functional neurosurgery* **85**, 195-203 (2007).
35. Adhikari, B.M. et al. Effects of ketamine and midazolam on resting state connectivity and comparison with ENIGMA connectivity deficit patterns in schizophrenia. *Human brain mapping* **41**, 767-778 (2020).
36. Adhikari, B.M. et al. Comparison of heritability estimates on resting state fMRI connectivity phenotypes using the ENIGMA analysis pipeline. *Human brain mapping* **39**, 4893-4902 (2018).
37. Adhikari, B.M. et al. Heritability estimates on resting state fMRI data using ENIGMA analysis pipeline. (2018).
38. Villalón-Reina, J.E. et al. Altered white matter microstructure in 22q11. 2 deletion syndrome: a multisite diffusion tensor imaging study. *Molecular Psychiatry*, 1-14 (2019).
39. Hatton, S.N. et al. White matter abnormalities across different epilepsy syndromes in adults: an ENIGMA-Epilepsy study. *Brain* **143**, 2454-2473 (2020).
40. Kochunov, P. et al. ENIGMA-DTI: Translating reproducible white matter deficits into personalized vulnerability metrics in cross-diagnostic psychiatric research. *Human Brain Mapping* (2020).
41. Vázquez-Rodríguez, B. et al. Gradients of structure–function tethering across neocortex. *Proceedings of the National Academy of Sciences* **116**, 21219-21227 (2019).
42. Romero-Garcia, R. et al. Structural covariance networks are coupled to expression of genes enriched in supragranular layers of the human cortex. *Neuroimage* **171**, 256-267 (2018).
43. Adler, S. et al. Topographic principles of cortical fluid-attenuated inversion recovery signal in temporal lobe epilepsy. *Epilepsia* **59**, 627-635 (2018).
44. Shafiei, G. et al. Spatial patterning of tissue volume loss in schizophrenia reflects brain network architecture. *Biological psychiatry* **87**, 727-735 (2020).

45. Brown, J.A. et al. Patient-tailored, connectivity-based forecasts of spreading brain atrophy. *Neuron* **104**, 856-868. e855 (2019).
46. Zheng, Y.-Q. et al. Local vulnerability and global connectivity jointly shape neurodegenerative disease propagation. *PLoS biology* **17**, e3000495 (2019).
47. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature genetics* **51**, 63-75 (2019).
48. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics* **51**, 431-444 (2019).
49. Stahl, E.A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature genetics* **51**, 793-803 (2019).
50. Howard, D.M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature neuroscience* **22**, 343-352 (2019).
51. Consortium, T.I.L.A.E. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. *Nature communications* **9** (2018).
52. Pardiñas, A.F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature genetics* **50**, 381-389 (2018).
53. Yu, D. et al. Interrogating the genetic determinants of Tourette's syndrome and other tic disorders through genome-wide association studies. *American Journal of Psychiatry* **176**, 217-227 (2019).
54. Stein, J.S., Natalia; Hibar, Derrek; Jahanshad, Neda; Thompson Paul ENIGMAVis: Updated interactive visualization of genetic influences on brain structure. *Organization for Human Brain Mapping (OHBM) 2016* (2016).
55. Schleicher, A., Amunts, K., Geyer, S., Morosan, P. & Zilles, K. Observer-independent method for microstructural parcellation of cerebral cortex: a quantitative approach to cytoarchitectonics. *Neuroimage* **9**, 165-177 (1999).
56. Zilles, K., Schleicher, A., Palomero-Gallagher, N. & Amunts, K. in Brain mapping: the methods 573-602 (Elsevier, 2002).
57. Zhou, J., Gennatas, E.D., Kramer, J.H., Miller, B.L. & Seeley, W.W. Predicting regional neurodegeneration from the healthy brain functional connectome. *Neuron* **73**, 1216-1227 (2012).
58. Avena-Koenigsberger, A., Misic, B. & Sporns, O. Communication dynamics in complex brain networks. *Nature Reviews Neuroscience* **19**, 17 (2018).
59. Alexander-Bloch, A.F. et al. On testing for spatial correspondence between maps of human brain structure and function. *Neuroimage* **178**, 540-551 (2018).
60. Váša, F. et al. Adolescent tuning of association cortex in human structural brain networks. *Cerebral Cortex* **28**, 281-294 (2018).
61. Ball, G. et al. Cortical morphology at birth reflects spatio-temporal patterns of gene expression in the fetal brain. *bioRxiv* (2020).
62. Seidlitz, J. et al. Transcriptomic and cellular decoding of regional brain vulnerability to neurogenetic disorders. *Nature communications* **11**, 1-14 (2020).
63. Paquola, C. et al. Shifts in myeloarchitecture characterise adolescent development of cortical gradients. *Elife* **8**, e50482 (2019).
64. Park, B.-y. et al. Connectome and microcircuit models implicate atypical subcortico-cortical interactions in autism pathophysiology. *bioRxiv* (2020).
65. Wagstyl, K. et al. Mapping cortical laminar structure in the 3d bigbrain. *Cerebral cortex* **28**, 2551-2562 (2018).
66. Wagstyl, K. et al. BigBrain 3D atlas of cortical layers: Cortical and laminar thickness gradients diverge in sensory and motor cortices. *PLoS biology* **18**, e3000678 (2020).
67. Larivière, S., Ward, N.S. & Boudrias, M.-H. Disrupted functional network integrity and flexibility after stroke: Relation to motor impairments. *NeuroImage: Clinical* **19**, 883-891 (2018).
68. Rubinov, M. & Bullmore, E. Fledgling pathoconnectomics of psychiatric disorders. *Trends in cognitive sciences* **17**, 641-647 (2013).
69. Bernhardt, B.C., Bonilha, L. & Gross, D.W. Network analysis for a network disorder: the emerging role of graph theory in the study of epilepsy. *Epilepsy & Behavior* **50**, 162-170 (2015).
70. Gollo, L.L. et al. Fragility and volatility of structural hubs in the human connectome. *Nature neuroscience* **21**, 1107-1116 (2018).
71. Crossley, N.A. et al. The hubs of the human connectome are generally implicated in the

anatomy of brain disorders. *Brain* **137**, 2382-2395 (2014).
72. Buckner, R.L. et al. Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. *Journal of neuroscience* **29**, 1860-1873 (2009).
73. Seeley, W.W. Mapping neurodegenerative disease onset and progression. *Cold Spring Harbor Perspectives in Biology* **9**, a023622 (2017).
74. Yau, Y. et al. Network connectivity determines cortical thinning in early Parkinson's disease progression. *Nature communications* **9**, 1-10 (2018).
75. Bernasconi, N. et al. Mesial temporal damage in temporal lobe epilepsy: a volumetric MRI study of the hippocampus, amygdala and parahippocampal region. *Brain* **126**, 462-469 (2003).
76. Bernhardt, B.C., Bernasconi, N., Hong, S.-J., Dery, S. & Bernasconi, A. Subregional mesiotemporal network topology is altered in temporal lobe epilepsy. *Cerebral cortex* **26**, 3237-3248 (2015).
77. Wang, Z. et al. Community-informed connectomics of the thalamocortical system in generalized epilepsy. *Neurology* **93**, e1112-e1122 (2019).
78. Bernhardt, B.C. et al. Thalamo–cortical network pathology in idiopathic generalized epilepsy: insights from MRI-based morphometric correlation analysis. *Neuroimage* **46**, 373-381 (2009).
79. Weng, Y. et al. Macroscale and microcircuit dissociation of focal and generalized human epilepsies. *Communications biology* **3**, 1-11 (2020).
80. Lariviere, S. et al. Microstructure-Informed Connectomics: Enriching Large-Scale Descriptions of Healthy and Diseased Brains. *Brain Connect* **9**, 113-127 (2019).
81. Consortium, P. Revealing the brain's molecular architecture. *Science (New York, NY)* **362**, 1262 (2018).
82. Hunter, J.D. Matplotlib: A 2D graphics environment. *Computing in science & engineering* **9**, 90-95 (2007).
83. Brett, M. et al. nipy/nibabel. *Zenodo* (2020).
84. Abraham, A. et al. Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics* **8**, 14 (2014).
85. Oliphant, T.E. A guide to NumPy, Vol. 1. (Trelgol Publishing USA, 2006).
86. Harris, C.R. et al. Array programming with NumPy. *Nature* **585**, 357-362 (2020).
87. McKinney, W. in Proceedings of the 9th Python in Science Conference, Vol. 445 51-56 (Austin, TX, 2010).
88. Waskom, M. et al. (Zenodo; 2020).
89. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **12**, 2825-2830 (2011).
90. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods* **17**, 261-272 (2020).
91. Schroeder, W.J., Lorensen, B. & Martin, K. The visualization toolkit: an object-oriented approach to 3D graphics. (Kitware, 2004).
92. Glasser, M.F. et al. A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171-178 (2016).
93. Schaefer, A. et al. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral cortex* **28**, 3095-3114 (2018).
94. Merker, B. Silver staining of cell bodies by means of physical development. *Journal of neuroscience methods* **9**, 235-241 (1983).
95. Mohlberg, H., Tweddell, B., Lippert, T. & Amunts, K. in International Workshop on Brain-Inspired Computing 15-27 (Springer, 2015).
96. Waehnert, M. et al. Anatomically motivated modeling of cortical laminae. *Neuroimage* **93**, 210-220 (2014).
97. Lewis, L. et al. in Annual Meeting of the Organization for Human Brain Mapping (2019).
98. Coifman, R.R. et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the national academy of sciences* **102**, 7426-7431 (2005).
99. de Wael, R.V. et al. BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. *Communications biology* **3**, 1-10 (2020).
100. Glasser, M.F. et al. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105-124 (2013).
101. Salimi-Khorshidi, G. et al. Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* **90**, 449-468 (2014).
102. Tournier, J.D. et al. MRtrix3: A fast, flexible and open software framework for medical image

processing and visualisation. *Neuroimage* **202**, 116137 (2019).
103. Smith, R.E., Tournier, J.-D., Calamante, F. & Connelly, A. Anatomically-constrained tractography: improved diffusion MRI streamlines tractography through effective use of anatomical information. *Neuroimage* **62**, 1924-1938 (2012).
104. Jeurissen, B., Tournier, J.-D., Dhollander, T., Connelly, A. & Sijbers, J. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage* **103**, 411-426 (2014).
105. Tournier, J.-D., Calamante, F. & Connelly, A. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage* **35**, 1459-1472 (2007).
106. 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 (2015).
107. Betzel, R.F., Griffa, A., Hagmann, P. & Misic, B. Distance-dependent consensus thresholds for generating group-representative structural brain networks. *Netw Neurosci* **3**, 475-496 (2019).

MATERIALS AND METHODS

Code data availability

All code used for data analysis and visualization is available on GitHub (http://github.com/MICA-MNI/ENIGMA). The ENIGMA Toolbox Python package relies on the following open-source Python dependencies: Matplotlib[82], NiBabel[83], nilearn[84], Numpy[85, 86], pandas[87], seaborn[88], Scikit-learn[89], SciPy[90], and VTK[91]. Users seeking help are encouraged to subscribe and post their questions to the ENIGMA Toolbox mailing list at https://groups.google.com/g/enigma-toolbox.

ENIGMA data description

Meta-analytical group comparisons (summary statistics)

ENIGMA's standardized protocols for data processing, quality assurance, and meta-analysis of individual subject data were conducted at each site (http://enigma.ini.usc.edu/protocols/imaging-protocols/). For site-level meta-analysis, all research centres within a given specialized Working Group tested for case *vs*. control differences using multiple linear regressions, where diagnosis (*e.g.*, healthy controls *vs*. individuals with epilepsy) was the predictor of interest, and subcortical volume, cortical thickness, or surface area of a given brain region was the outcome measure. Case-control differences were computed across all regions using either Cohen's *d* effect sizes or *t*-values, after adjusting for effects of age and sex (attention deficit/hyperactivity disorder [cortical thickness], bipolar disorder [cortical thickness and surface area], major depressive disorder [cortical thickness and surface area], schizophrenia [cortical thickness and surface area]), age, sex, and dataset (22.q11.2 [cortical thickness and surface area]), age, sex, and IQ (autism spectrum disorder [cortical thickness and surface area]), age, sex, and intracranial volume (ICV; attention deficit/hyperactivity disorder [surface area], epilepsy [cortical thickness, surface area, and subcortical volume], obsessive compulsive disorder [cortical thickness and surface area]). Across-site random-effects meta-analyses of Cohen's *d*/*t*-values were then performed for each of the cortical and subcortical region. These ENIGMA summary statistics can be retrieved from the ENIGMA Toolbox and contains the following data: effect sizes for case-control differences (d_icv), standard error (se_icv), lower bound of the confidence interval (low_ci_icv), upper bound of the confidence interval (up_ci_icv), number of controls (n_controls), number of patients (n_patients), observed p-values (pobs), false discovery rate (FDR)-corrected p-value (fdr_p).

Individual site or mega-analytic data

Functionalities and tutorials within the ENIGMA Toolbox are generalizable to individual site or mega-analysis datasets, which pool individual-level data. Due to restrictions of individual-level data transfer, however, we provide an ENIGMA-derived example dataset that includes subject-level data from ten healthy controls (7 females, age±SD=33.3±8.8 years) and ten individuals with drug-resistant temporal lobe epilepsy (7 females, age±SD=39.8±14.8 years). The Ethics Committee of the Montreal Neurological Institute and Hospital approved the study. Written informed consent, including a statement for open sharing of collected data, was obtained from all participants. As per ENIGMA-Epilepsy protocols, users can fetch covariate information (subject ID, diagnosis, sub-diagnosis, handedness, age at onset, duration of illness, and ICV), regional volumetric data (from 12 subcortical regions—namely bilateral accumbens, amygdala, caudate, pallidum, putamen, and thalamus, bilateral hippocampus, and bilateral ventricles), and cortical thickness and surface area from every Desikan-Killiany cortical region[27]. Given the standardized ENIGMA format, users can easily replace our example dataset with any other individual site or mega-analysis datasets of their own.

Compatibility with other datasets

To increase generalizability and usability, every function within the ENIGMA Toolbox is compatible with any neuroimaging data parcellated according to the Desikan-Killiany[27], Glasser[92], and Schaefer[93] parcellations (other parcellations will be added upon request).

Transcriptomics data and contextualization

As part of the ENIGMA Toolbox, users can fetch and manipulate preprocessed microarray expression data collected from six human donor brains and released by the Allen Institute for Brain Sciences[19]. Microarray expression data were first generated using abagen[26], a toolbox that provides reproducible workflows for processing and preparing gene co-expression data according to previously established recommendations[23]; preprocessing steps included intensity-based filtering of microarray probes, selection of a representative probe for each gene across both hemispheres, matching of microarray samples to brain parcels from the Desikan-Killiany[27], Glasser[92], and Schaefer[93] parcellations, normalization, and aggregation within parcels and across donors. Moreover, genes whose similarity across

donors fell below a threshold (r<0.2) were removed, leaving, for instance, a total of 15,633 genes for analysis for using Desikan-Killiany atlas. To accommodate users, we also provide unthresholded gene datasets with varying stability thresholds (r≥0.2, r≥0.4, r≥0.6, r≥0.8) for every parcellation (https://github.com/saratheriver/enigma-extra).

ENIGMA Toolbox users can furthermore query pre-defined lists of disease-related genes (obtained from several recently published GWAS), including gene sets for attention deficit/hyperactivity disorder (n_{genes}=26)[47], autism spectrum disorder (n_{genes}=30)[48], bipolar disorder (n_{genes}=30)[49], depression (n_{genes}=269)[50], common epilepsies (n_{genes}=21)[51], schizophrenia (n_{genes}=213)[52], and Tourette's syndrome (n_{genes}=58)[53]. These gene sets can be subsequently mapped to cortical and subcortical regions using the Allen Human Brain Atlas[19] and projected to surface templates using our surface visualization tools.

BigBrain data and contextualization

BigBrain is a ultra-high resolution, 3D volumetric reconstruction of a *postmortem* Merker-stained and sliced human brain from a 65-year-old male, with specialized pial and white matter surface reconstructions (obtained via the open-access BigBrain repository: https://bigbrain.loris.ca/main.php)[20]. The *postmortem* brain was paraffin-embedded, coronally sliced into 7400 20µm sections, silver-stained for cell bodies[94], and digitized. A 3D reconstruction was implemented with a successive coarse-to-fine hierarchical procedure[95], resulting in a full brain volume. For the ENIGMA Toolbox, we used the highest resolution full brain volume (100µm isotropic voxels), then generated 50 equivolumetric surfaces between the pial and white matter surfaces. The equivolumetric model compensates for cortical folding by varying the Euclidean distance between pairs of intracortical surfaces throughout the cortex, thus preserving the fractional volume between surfaces[96]. Next, staining intensity profiles, representing neuronal density and soma size by cortical depth, were sampled along 327,684 surface points in the direction of cortical columns.

Following seminal histological work[55, 56, 63], we characterized vertex-wise cytoarchitecture by taking two central moments of the staining intensity profiles (mean and skewness). Finally, the Desikan-Killiany atlas was nonlinearly transformed to the BigBrain histological surfaces[97] and central moments were averaged within each parcels, excluding outlier vertices with values more than three scaled median absolute deviations away from the parcel median.

The BigBrain gradient was obtained from the original publication[29] and mapped to the Desikan-Killiany[27], Glasser[92], and Schaefer[93] parcellations. In brief, the authors derived an MPC matrix by correlating BigBrain intensity profiles between every pair of regions in a 1,012 cortical node parcellation, controlling for the average whole-cortex intensity profile. The MPC matrix was thresholded row-wise to retain the top 10% of correlations and converted into a normalized angle matrix. Diffusion map embedding[98, 99], a nonlinear manifold learning technique, identified the principal axis of variation across cortical areas, *i.e.*, the BigBrain gradient. In this space, cortical nodes that are strongly similar are closer together, whereas nodes with little to no intercovariance are farther apart. To allow contextualization of ENIGMA-derived surface-based findings, we mapped the BigBrain gradient to the Desikan-Killiany atlas and partitioned it into five equally sized discrete bins.

Cytoarchitectonics data and contextualization

By adapting a previously published approach[32], we mapped the cytoarchitectonic atlas of von Economo and Koskinas[33] to cortical surface templates. Cytoarchitectonic class labels from the original five different structural types of cerebral cortex (agranular, frontal, parietal, polar, granular) were manually assigned to each parcellation region[32] and subsequently mapped to vertex-wise space.

To stratify cortex-wide effects according to the five cytoarchitectonic classes, the `economo_koskinas_atlas()` function maps parcellated data (*e.g.*, disease-related atrophy map on the Desikan-Killiany atlas) to vertex-wise space and iteratively averages values from all vertices within each class.

Connectivity data for macroscale connectome models

As in prior work[22], we selected a group of unrelated healthy adults (n=207; 83 males, mean age±SD=28.73±3.73 years, range=22-36 years) from the HCP dataset[21]. HCP data were acquired on a Siemens Skyra 3T and included: (*i*) T1-weighted images (magnetization-prepared rapid gradient echo [MPRAGE] sequence, repetition time [TR]=2,400ms, echo time [TE]=2.14ms, field of view [FOV]=224×224 mm^2, voxel size=0.7×0.7×0.7 mm, 256 slices), (*ii*) resting-state functional MRI (gradient-echo echo-planar imaging [EPI] sequence, TR=720 ms, TE=33.1 ms, FOV=208×180 mm^2, voxel size=2 mm^3, 72 slices), and (*iii*) diffusion MRI (spin-echo EPI sequence, TR=5,520

ms, TE=89.5ms, FOV=210×180, voxel size=1.25mm^3, *b*-value=1,000/2,000/3,000 s/mm^2, 270 diffusion directions, 18 b0 images). HCP data underwent the initiative's minimal preprocessing[100]. In brief, resting-state functional MRI data underwent distortion and head motion corrections, magnetic field bias correction, skull removal, intensity normalization, and were mapped to MNI152 space. Noise components attributed to head movement, white matter, cardiac pulsation, arterial, and large vein related contributions were automatically removed using FIX[101]. Preprocessed time series were mapped to standard gray ordinate space using a cortical ribbon-constrained volume-to-surface mapping algorithm and subsequently concatenated to form a single time series. Diffusion MRI data underwent b0 intensity normalization and correction for susceptibility distortion, eddy currents, and head motion. High-resolution functional and structural data were parcellated according to the Desikan-Killiany[27], Glasser[92], as well as Schaefer 100, 200, 300, and 400[93] parcellations.

Normative functional connectivity matrices were generated by computing pairwise correlations between the time series of all cortical regions and subcortical (nucleus accumbens, amygdala, caudate, hippocampus, pallidum, putamen, thalamus) regions; negative connections were set to zero. Subject-specific connectivity matrices were then *z*-transformed and aggregated across participants to construct a group-average functional connectome. Available cortico-cortical, subcortico-cortical, and subcortico-subcortical matrices are unthresholded. Normative structural connectivity matrices were generated from preprocessed diffusion MRI data using MRtrix3[102]. Anatomical constrained tractography was performed using different tissue types derived from the T1-weighted image, including cortical and subcortical gray matter, white matter, and cerebrospinal fluid[103]. Multi-shell and multi-tissue response functions were estimated[104] and constrained spherical deconvolution and intensity normalization were performed[105]. The initial tractogram was generated with 40 million streamlines, with a maximum tract length of 250 and a fractional anisotropy cutoff of 0.06. Spherical-deconvolution informed filtering of tractograms (SIFT2) was applied to reconstruct whole-brain streamlines weighted by the cross-section multipliers[106]. Reconstructed streamlines were mapped onto the 68 cortical and 14 subcortical (including hippocampus) regions to produce subject-specific structural connectivity matrices. The group-average normative structural connectome was defined using a distance-dependent thresholding procedure, which preserved the edge length distribution in individual patients[107], and was log transformed to reduce connectivity strength variance. As such, structural connectivity was defined by the number of streamlines between two regions (*i.e.*, fiber density).

The ENIGMA Toolbox: Cross-disorder integration and multiscale neural contextualization of multisite neuroimaging datasets

Supplementary Information

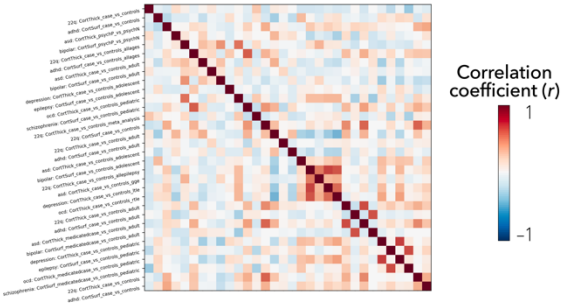
a | Cross-disorder effect (*Python*)

```python
from enigmatoolbox.cross_disorder import cross_disorder_effect

# Extract and visualize shared disorder effects
correlation_matrix, names = cross_disorder_effect(method='correlation')
```

b | Cross-disorder effect (*Matlab*)

```matlab
% Extract and visualize shared disorder effects
[~, ~, correlation_matrix, names] = cross_disorder_effect('method', 'correlation');
```

c | Shared effect across disorders



Supplementary Figure 1. **Cross-disorder effect: cross-correlation**. Minimal (**a**) Python and (**b**) Matlab code snippets to systematically correlate patterns of brain structural abnormalities across disorders. (**c**) Resulting cross-correlation matrix showing similar (red) and dissimilar (blue) transdiagnostic morphometric signatures.

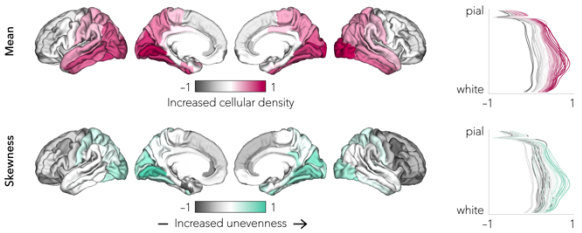
a | Stratify atrophy by BigBrain statistical moments (*Python*)

```python
import numpy as np
from enigmatoolbox.histology import bb_moments_raincloud

# Extract FDR-corrected p-values and find regions with p < 0.01
region_idx = np.where(CT['fdr_p'].to_numpy() <= 0.01)

# Stratify and plot results according to BigBrain statistical moments
bb_moments_raincloud(region_idx=region_idx)
```

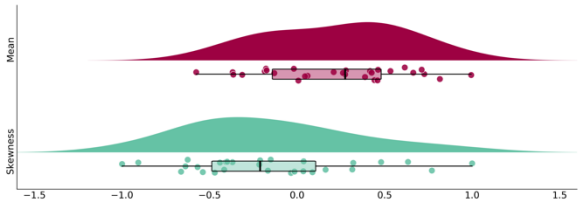
b | Stratify atrophy by BigBrain statistical moments (*Matlab*)

```matlab
% Extract FDR-corrected p-values and find regions with p < 0.01
region_idx  = find(CT.fdr_p <= 0.01);

% Stratify and plot results according to BigBrain statistical moments
f = figure,
    bb_moments_raincloud(region_idx)
```

c | BigBrain statistical moments



d | Atrophy stratified by statistical moments



Supplementary Figure 2. **Advanced analytical workflows: BigBrain statistical moments**. Minimal (**a**) Python and (**b**) Matlab code snippets to stratify significantly atrophied regions according to BigBrain statistical moments. (**c**) Regional cytoarchitecture using two statistical moments of staining profiles: (*i*) mean intracortical staining across the mantle, which allows inferences on overall cellular density, and (*ii*) profile skewness, which indexes the distribution of cells across upper and lower layers of the cortex. (**d**) Applying this approach to individuals with left focal epilepsy revealed greater cortical atrophy in regions with greater, and more evenly distributed, cellular densities.

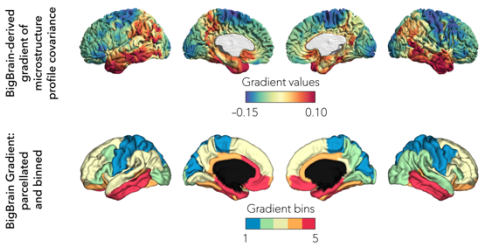
a | Stratify atrophy by the BigBrain gradient (*Python*)

```python
import numpy as np
from enigmatoolbox.histology import bb_gradient_plot

# Stratify and plot results according to the BigBrain gradient
bb_gradient_plot(data=np.where(CT_d_thr == 0, np.nan, CT_d_thr),
              axis_range=(-0.6, 0.25), yaxis_label='Cohen\'s $d$')
```

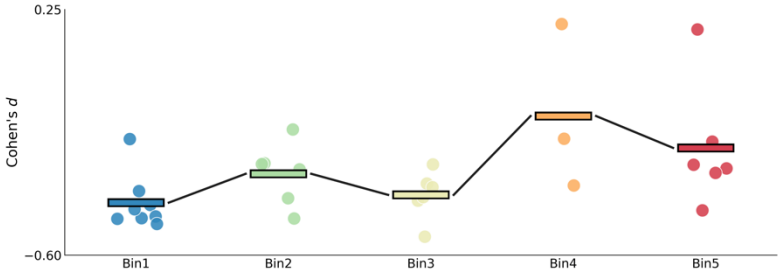
b | Stratify atrophy by the BigBrain gradient (*Matlab*)

```MATLAB
% Stratify and plot results according to the BigBrain gradient
CT_d_thr(CT_d_thr == 0) = nan;
f = figure,
    bb_gradient_plot(CT_d_thr, 'axis_range', [-0.6 0.25], ...
                 'yaxis_label', 'Cohen'' {\it d}')
```

c | BigBrain-derived gradient of microstructure profile covariance



d | Atrophy stratified by the BigBrain gradient



Supplementary Figure 3. **Advanced analytical workflows: BigBrain gradient**. Minimal (**a**) Python and (**b**) Matlab code snippets to stratify significantly atrophied regions according to the BigBrain gradient. (**c**) BigBrain-derived gradient of microstructural profile covariance gradient describes a sensory-fugal transition in intracortical microstructure (*top*). The BigBrain gradient was then mapped to several parcellations and partitioned it into five equally sized discrete bins (*bottom*). (**d**) Applying this approach to individuals with left focal epilepsy revealed greater cortical atrophy in areas located towards the sensory apex (blue/green) of the cytoarchitectural gradient.

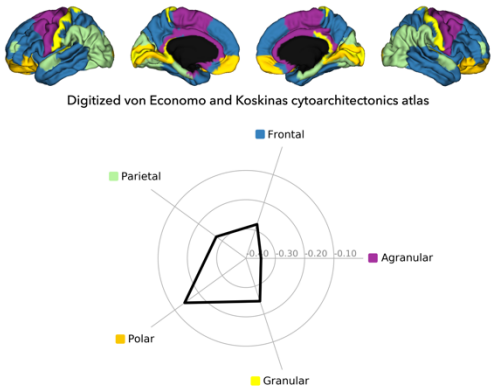
a | Stratify atrophy by Economo-Koskinas cortical types (*Python*)

```python
from enigmatoolbox.histology import economo_koskinas_spider

# Stratify cortical atrophy based on Economo-Koskinas classes
class_mean = economo_koskinas_spider(CT_d, axis_range=(-0.4, 0))
```

b | Stratify atrophy by Economo-Koskinas cortical types (*Matlab*)

```MATLAB
% Stratify cortical atrophy based on Economo-Koskinas classes
class_mean = economo_koskinas_spider(CT_d, 'axis_range', [-0.4 0])
```

c | Atrophy stratified by statistical moments



Supplementary Figure 4. **Advanced analytical workflows: von Economo and Koskinas cytoarchitectonic atlas**. Minimal (**a**) Python and (**b**) Matlab code snippets to stratify significantly atrophied regions according to cytoarchitectonic classes. (c) A digitized cytoarchitectonic atlas from seminal postmortem work by von Economo and Koskinas[30]. Contextualizing cortical atrophy patterns in individuals with left focal epilepsy revealed greater atrophy in the agranular cortex (purple).

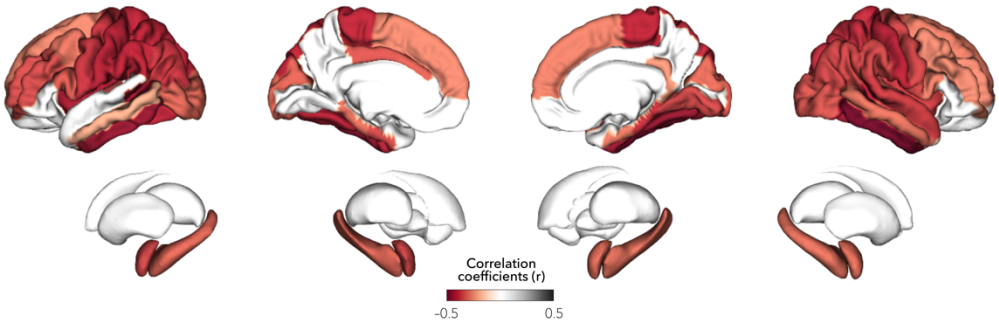
a | Disease epicenter model (*Python*)

```python
import numpy as np
from enigmatoolbox.permutation_testing import spin_test

# Identify cortical epicenters (from functional connectivity)
fc_ctx_epi = []
fc_ctx_epi_p = []
for seed in range(fc_ctx.shape[0]):
    seed_con = fc_ctx[:, seed]
    fc_ctx_epi = np.append(fc_ctx_epi, np.corrcoef(seed_con, CT_d)[0, 1])
    fc_ctx_epi_p = np.append(fc_ctx_epi_p,
                             spin_test(seed_con, CT_d, surface_name='fsa5', parcellation_name='aparc',
                                       type='pearson', n_rot=1000, null_dist=False))

# Identify subcortical epicenters (from functional connectivity)
fc_sctx_epi = []
fc_sctx_epi_p = []
for seed in range(fc_sctx.shape[0]):
    seed_con = fc_sctx[seed, :]
    fc_sctx_epi = np.append(fc_sctx_epi, np.corrcoef(seed_con, CT_d)[0, 1])
    fc_sctx_epi_p = np.append(fc_sctx_epi_p,
                              spin_test(seed_con, CT_d, surface_name='fsa5', n_rot=1000))
```

b | Disease epicenter model (*Matlab*)

```MATLAB
% Computing cortical epicenter values (from functional connectivity)
fc_ctx_epi              = zeros(size(fc_ctx, 1), 1);
fc_ctx_epi_p            = zeros(size(fc_ctx, 1), 1);
for seed = 1:size(fc_ctx, 1)
    seed_conn           = fc_ctx(:, seed);
    r_tmp               = corrcoef(seed_conn, CT_d);
    fc_ctx_epi(seed)    = r_tmp(1, 2);
    fc_ctx_epi_p(seed)  = spin_test(seed_conn, CT_d, 'surface_name', 'fsa5', 'parcellation_name', ...
                                    'aparc', 'n_rot', 1000, 'type', 'pearson');
end

% Computing subcortical epicenter values (from functional connectivity)
fc_sctx_epi             = zeros(size(fc_sctx, 1), 1);
fc_sctx_epi_p           = zeros(size(fc_sctx, 1), 1);
for seed = 1:size(fc_sctx, 1)
    seed_conn           = fc_sctx(seed, :);
    r_tmp               = corrcoef(seed_conn, CT_d);
    fc_sctx_epi(seed)   = r_tmp(1, 2);
    fc_sctx_epi_p(seed) = spin_test(seed_conn, CT_d, 'surface_name', 'fsa5', 'parcellation_name', ...
                                    'aparc', 'n_rot', 1000, 'type', 'pearson');
end
```

c | Cortical and subcortical disease epicenters



Supplementary Figure 5. **Advanced analytical workflows: disease epicenter model**. Minimal (**a**) Python and (**b**) Matlab code snippets to identify disease-specific cortical and subcortical epicenters. (**c**) Applying this approach to individuals with left focal epilepsy revealed that patterns of atrophy in left focal epilepsy were anchored to the connectivity profiles of mesiotemporal regions.