Connectomes for 40,000 UK Biobank participants: A multi-modal, multi-scale brain network resource (Dated: March 8, 2023)

Sina Mansour L.[1,*], Maria A. Di Biase[2, 3, 4], Robert E. Smith[5,6], Andrew Zalesky[1,2], and Caio Seguin[2,7]

[1]Department of Biomedical Engineering, The University of Melbourne, VIC, Australia
[2]Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne, Parkville, Victoria, Australia
[3]Department of Anatomy and Physiology, School of Biomedical Sciences, The University of Melbourne, Parkville, Victoria, Australia
[4]Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, MA, USA
[5]The Florey Institute of Neuroscience and Mental Health, Heidelberg, Victoria, Australia
[6]Florey Department of Neuroscience and Mental Health, The University of Melbourne, Parkville, Victoria, Australia
[7]Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA
[*]To whom correspondence should be addressed. Email: sina.mansour.lakouraj@gmail.com

ABSTRACT

We mapped functional and structural brain networks for more than 40,000 UK Biobank participants. Structural connectivity was estimated with tractography and diffusion MRI. Resting-state functional MRI was used to infer regional functional connectivity. We provide high-quality structural and functional connectomes for multiple parcellation granularities, several alternative measures of interregional connectivity, and a variety of common data pre-processing techniques, yielding more than one million connectomes in total and requiring more than 200,000 hours of compute time. For a single subject, we provide 28 out-of-the-box versions of structural and functional brain networks, allowing users to select, e.g., the parcellation and connectivity measure that best suit their research goals. Furthermore, we provide code and intermediate data for the time-efficient reconstruction of more than 1,000 different versions of a subject's connectome based on an array of methodological choices. All connectomes are available via the UK Biobank data sharing platform and our connectome mapping pipelines are openly available. In this report, we describe our connectome resource in detail for users, outline key considerations in developing an efficient pipeline to map an unprecedented number of connectomes, and report on the quality control procedures that were completed to ensure connectome reliability and accuracy. We demonstrate that our structural and functional connectivity matrices meet a number of quality control checks and replicate previously established findings in network neuroscience. We envisage that our resource will enable new studies of the human connectome in health, disease and aging at an unprecedented scale.

Background & Summary

Different aspects of brain connectivity can be quantified using different MRI modalities: diffusion-weighted MRI data can be utilized to map structural brain networks of white-matter connections[1,2]; alternatively, functional MRI data can be used to map functional connectivity networks describing inter-regional interactions in brain activity[3,4]. These network representations of brain connectivity are referred to as connectomes[5,6]. Establishing a large-scale community biobank of structural and functional human connectomes will enable a diverse range of research into brain networks in health and disease.

The importance of large-scale neuroimaging biobanks is increasingly recognized as key to addressing reproducibility concerns in neuroscience[7–12]. The UK biobank (UKB)—a population study containing in-depth biomedical, health, and environmental data—is the world's largest neuroimaging resource (with ∼45,000 imaging sessions acquired from ∼40,000 participants thus far)[8, 13–15]. This biobank offers tremendous potential for research on early diesase prediction and alignment of image-derived phenotypes (IDPs) with cognitive, behavioral, genetic, and medical observations. The availability of longitudinal neuroimaging data accompanying constantly updated clinical records enables prospective neuroscientific research at a population scale. To facilitate this effort, the UKB has released a range of important quantitative neuroimaging derivatives, including regional measures of brain structure, microstructure, and function. At present however, measures of brain *connectivity* are not a part of this resource. Mapping connectomes from neuroimaging data at scale is computationally burdensome and requires significant technical expertise. Establishing a brain connectivity biobank for the UKB will ensure rapid access to connectomes for researchers without expertise or computing resources for large-scale connectivity mapping, facilitate reproducible neuroscience practices, enhance UKB utilization among the research community and ultimately lead to new discoveries about brain networks.

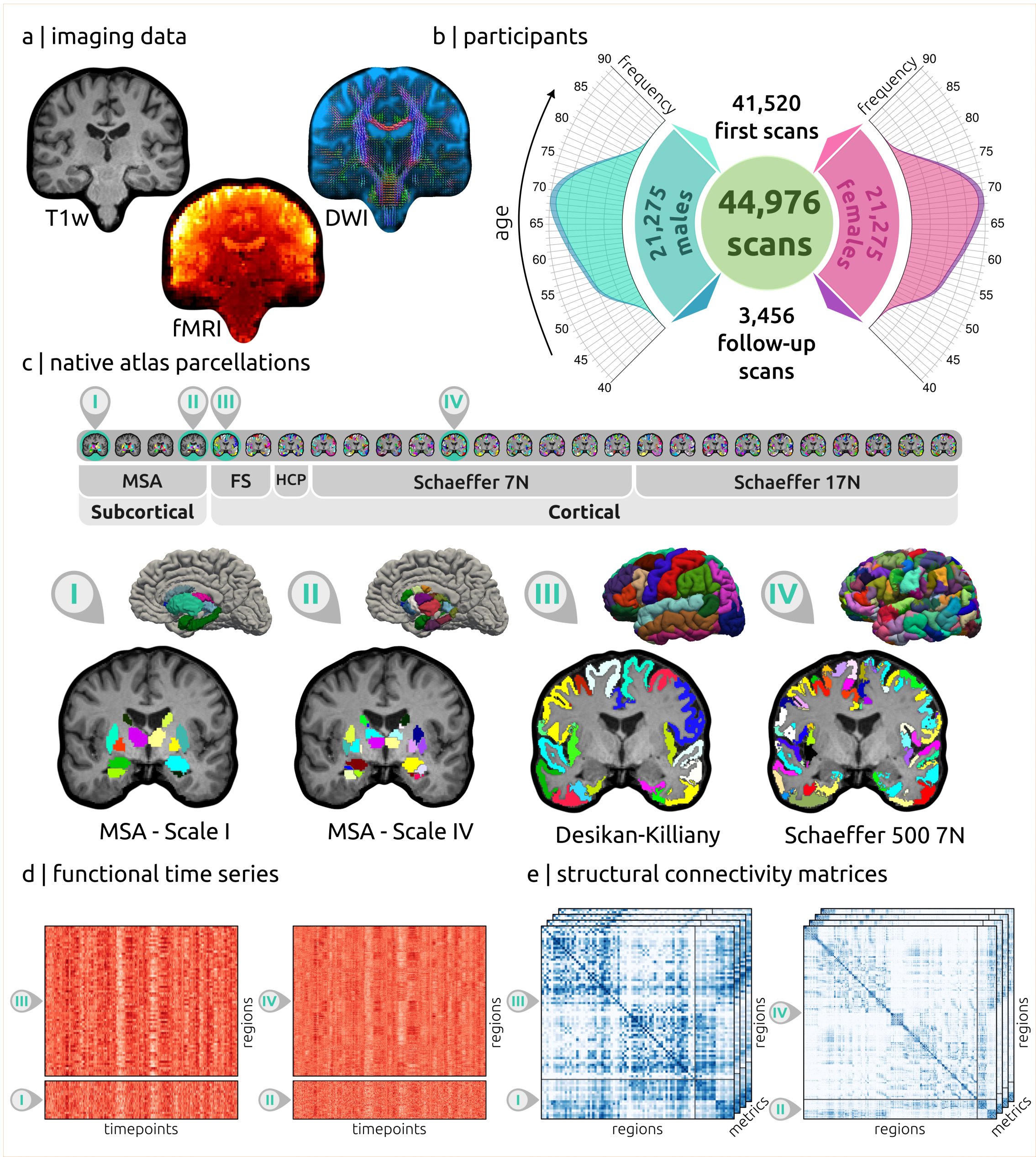


Figure 1. A comprehensive connectome biobank for the UK Biobank. (a) Structural and functional connectomes were mapped from diffusion and resting-state functional MRI data, respectively, while T1-weighted structural MRI was used for brain alignment. (b) Distribution of age and sex in the 41,520 UK Biobank participants with neuroimaging data available. Longitudinal data was available for 3,456 participants. (c) Various whole-brain atlases were used to map individual parcellations in native volumetric space. A total of 27 alternative parcellation schemes were computed. Four representative subcortical (I and II) and cortical (III and IV) parcellations of different granularities are illustrated. (d) Resting-state functional time series and (e) various structural connectivity matrices were computed for different parcellations. The time series and structural connectivity matrices are depicted for the same four sample parcellations. Abbreviations; MSA: Melbourne Subcortical Atlas, FS: FreeSurfer atlases, HCP: HCP-MMP1.0 atlas.

Here, we introduce our novel human connectome resource of brain atlases and connectivity matrices mapped for more than 40,000 adults participating in the UKB (see Figure 1). We provide functional activity time series and structural connectivity matrices for multiple parcellation schemes, several alternative measures of interregional connectivity and a variety of common data pre-processing techniques, yielding 27 brain-wide functional time series (enabling flexible and efficient access to various functional connectivity metrics), and 28 structural connectomes per imaging session, and more than one million connectomes and time-series in total. This required development of highly efficient connectome mapping pipelines and storage formats. Connectivity data is made available in compact and easy-to-use data formats and our connectome mapping pipelines are openly available. We completed extensive quality control procedures to ensure the accuracy and reliability of all connectivity matrices. This report aims to describe our connectome resource in detail for prospective users, provide insight into the key considerations that shaped the development of our connectome mapping pipeline and outline quality control procedures. We envisage that this resource will be of high utility and complement the current IDPs available in the UKB.

Methods

UK Biobank data

Background

The UKB is a large-scale dataset comprising over 500,000 participants (aged 40-69 years). The biobank is publicly available to advance health-related research[13]. Importantly, access to full health records (which are updated over time), and a wide range of longitudinal, long-term clinical, phenotypic, and genomic data is available to researchers[15]. In addition, UKB aims to acquire longitudinal neuroimaging data for 100,000 participants, with ∼45,000 separate MRI sessions for ∼40,000 participants released thus far[8, 14].

Participants

Brain MRI data sourced for the connectome biobank consists of 44,976 separate imaging sessions (23,701 females and 21,275 males) from 41,520 individuals (21,951 females and 19,569 males) aged 44–82 years ($\mu = 64.0, \sigma = 7.6$ years) at time of acquisition. For 3,456 individuals (1,750 females and 1,706 males) a single longitudinal follow-up MRI session was acquired 1–7 years ($\mu = 2.35, \sigma = 0.71$ years) after the first acquisition. Further details regarding recruitment protocols are provided elsewhere[14]. Imaging sessions that did not pass the existing UKB preprocessing and quality control pipeline[8] were excluded.

MRI data acquisition and existing preprocessing

A detailed description of the MRI data acquisition and preprocessing pipeline is provided elsewhere[8]. All modalities were acquired on 3T Siemens Skyra scanners using the standard Siemens 32-channel head coil. The T1-weighted structural brain images (Data-Field 20252) were acquired using a 3D MPRAGE acquisition at 1mm isotropic resolution with a 256mm superior-inferior field of view[8]. The preprocessing steps included gradient distortion correction (GDC)[16, 17], skull stripping[18], linear and nonlinear registration to MNI152 standard space[19–21], and defacing[8]. Additional derivative data precalculated and provided by the UKB resource include macroscopic tissue segmentation with FSL FAST[22], subcortical modeling with FSL FIRST[23], and brain segmentation including cortical surface estimation with Freesurfer[24].

Resting-state BOLD data (Data-Field 25751) were acquired with a multi-band gradient echo EPI sequence[25–27], with an acquisition time of ∼6 minutes, for a total of 490 volumes, with a spatial resolution of 2.4mm isotropic voxels (TE/TR=39/735 ms, MB=8, no in-plane acceleration, flip angle 52°, conventional fat saturation)[14]. Preprocessing steps[8] consisted of the FSL MELODIC pipeline[28] (EPI susceptibility distortion correction, GDC, motion correction with FSL MCFLIRT[20], grand-mean intensity normalization, and high-pass temporal filtering) followed by an ICA + FIX step to suppress remaining artifact components[29, 30].

Diffusion-weighted MRI (dMRI) data (Data-Field 20250) were acquired using a multi-band spin echo EPI sequence[31], with an acquisition time of ∼7 minutes, with 100 unique diffusion sensitisation directions distributed equally across two shells (b-values: 1000, 2000 s/mm^2), and 5 b=0 volumes, with a spatial resolution of 2mm isotropic voxels (MB=3, no in-plane acceleration, TE/TR=92/3600 ms, partial Fourier 6/8, conventional fat saturation). 3 additional b=0 volumes were acquired with reversed phase encoding direction to enable susceptibility field estimation[32]. The data were preprocessed with a pipeline[8] consisting of correction for eddy current and head motion[33–35] followed by GDC[36]. Additionally, FSL's `dtifit`[37] and the NODDI toolbox[38] were used to generate voxelwise microstructural parameters: fractional anisotropy (FA), tensor mode (MO), mean diffusivity (MD), b=0 signal intensity (S0), intra-cellular volume fraction (ICVF), isotropic volume fraction (ISOVF), and orientation dispersion index (ODI).

Connectomic nomenclature

Table 1 provides brief definitions for terms commonly used in this paper.

Term	Definition
Template space	A surface-based (e.g. FreeSurfer's "`fsnative`" and HCP's "`fs-LR`"), or volumetric (e.g. MNI 152) common brain space that is normally used for group-level neuroimaging studies[39].
Native space	The space in which a subject's brain imaging data is originally provided.
Registration	Computing a transformation that best aligns a subject's native space MRI data to a template space.
Brain atlas	A standard map of regional delineations segmenting the cortical and subcortical gray matter into discrete brain regions. These maps are conventionally provided in a standard template space.
Atlas parcellation	The transformation of a brain atlas from template space to subject's native space provides a parcellation map delineating different brain regions in an individual.
Brain network / Connectome	A measure of inter-regional structural or functional relationships stored in the form of a two-dimensional connectivity matrix[40]. Rows/columns of this matrics quantify inter-regional connectivity properties from a region of interest (ROI) in the gray matter. Atlas parcellarions can be used to define ROIs.
Tractography	A computational method to estimate anatomical trajectories of white matter fiber pathways from dMRI data[41].
Streamline	A set of 3D coordinates encoding trajectory estimates yielded from tractography.
Structural connectome	A connectome constructed from tractography results in which edges quantify properties describing anatomical connections between ROIs (e.g. number of streamlines, or connection length)[42].
Functional connectome	A connectome constructed from fMRI data in which edges quantify pairwise functional associations between multivariate time-series describing the regional blood-oxygen-level-dependent (BOLD) signal sampled from ROIs[43].

Table 1. A description of frequently used terms.

Connectome reconstruction pipelines

As detailed below, we developed novel connectome reconstruction pipelines to satisfy five competing demands:

1. *Connectome quality*, producing high-quality brain networks by using state-of-the-art pipelines to infer interregional brain connectivity from diffusion and functional neuroimaging data.
2. *Accessibility*, facilitating streamlined access to connectome data in an efficient and easy to use format.
3. *Flexibility*, enabling users to choose from connectomes mapped using a broad range of methodological preferences.
4. *Computational requirements*, as our pipeline needed to be executed on data from ∼45,000 MRI sessions, which is a considerable burden even when making use of high-performance computing (HPC) services.
5. *Storage requirements*, as the generated connectivty data needs to be stored both in the short term on local storage during calculation, and in the long term on UKB storage infrastructure.

Hence, the pipeline was developed to provide a good balance between these competing aims and prepare a comprehensive, versatile, and high-quality brain connectivity resource. This included the development of new utilities in software packages used to map conenctomes (e.g., *MRtrix3*[44]) with an explicit focus on reducing computational and storage requirements.

A crucial conflict identified within this set of demands lies at the intersection of accessibility, flexibility, and storage requirements. On the one hand, user flexibility is boosted by providing many versions of the same connectome, mapped according to different methodological choices, such as different atlases or interregional connectivity measures. On the other hand, catering for flexibility leads to a combinatorial increase in the number of connectomes, which would be prohibitively expensive to store and access for all UKB subjects. This would be particularly the case for SC matrices, for which there are many options for the metric of connectivity.

We addressed this important issue as follows. For a subset of connectome configurations, deemed to be of broad applicability, connectivity matrices were calculated and uploaded to the UKB for direct user access. Where an alternative configuration is desired, we provide both the requisite intermediate data (ie. for which the most expensive computational processes have already been performed) and a software tool that uses these data to efficiently calculate the connectome of interest, such that any combination of connectome attributes can be chosen. This tool, along with the relevant code for the pipeline in its entirety, is accessible from github.com/sina-mansour/UKB-connectomics.

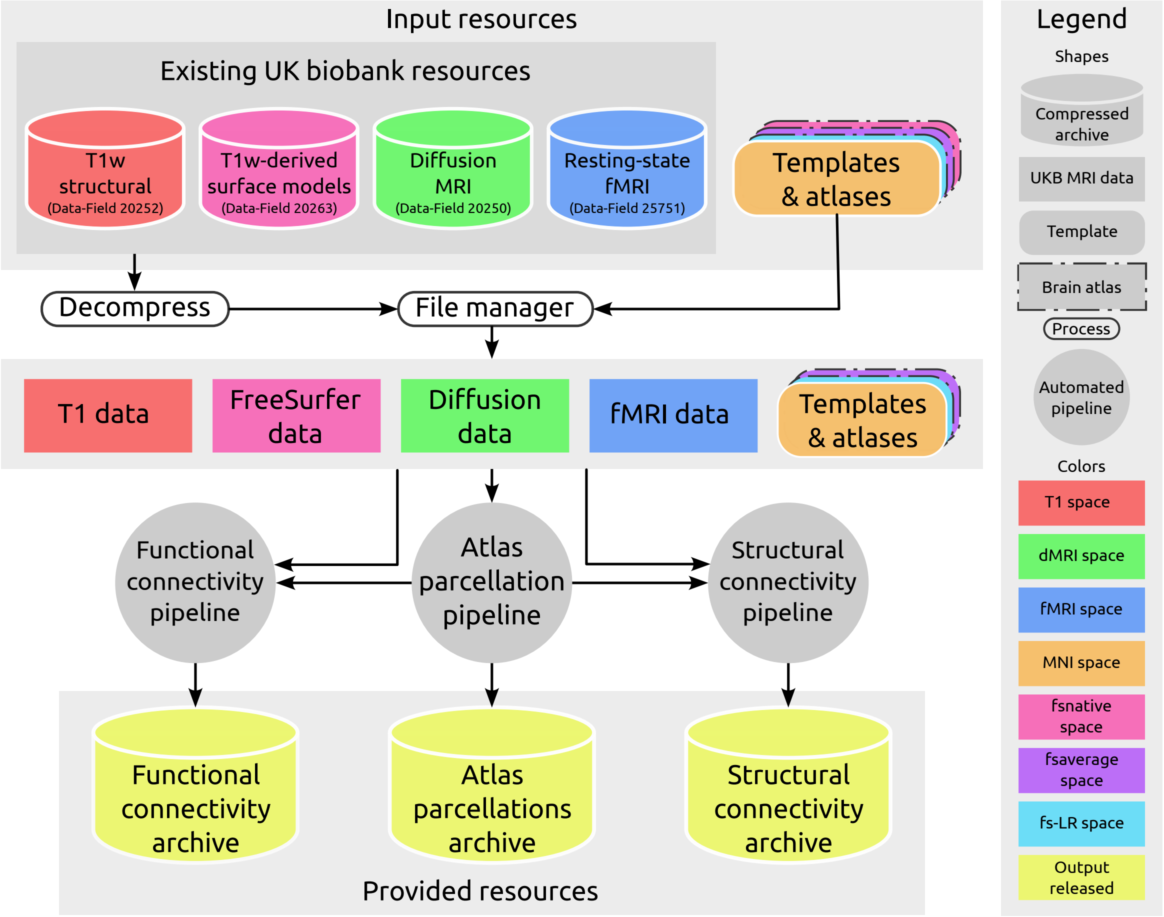


Figure 2. Schematic flowchart of the complete pipeline. Compressed bulk data archives from UKB along with publicly available brain parcellation atlases and templates were used as inputs of the connectivity mapping pipelines. These inputs were used in three separate steps to generate parcellations in subject space, that were subsequently utilized in functional and structural connectome construction pipelines. Flowcharts of these steps are detailed in the ensuing figures. For every imaging session, the outputs of these three steps are provided as separate compressed archives.

An overview flowchart of the pipeline is presented in Figure 2. In short, a set of automated pipelines were implemented to perform three main tasks:

1. *Atlas parcellations*: Generate volumetric atlas parcellations in subject space, segmenting cortical and subcortical gray matter into distinct brain regions
2. *Functional connectivity*: Derive resting-state functional time series of brain activity within these parcels to enable functional connectivity (FC) estimation
3. *Structural connectivity*: Estimate white matter fiber orientations from diffusion MRI data and perform tractography to map structural connectivity (SC) matrices

The following sections provide a detailed explanation of every step in the connectome reconstruction pipeline.

Brain atlas parcellations

For parcellation of the cortical and subcortical gray matter into distinct nodes of a connected network, here we focus strictly on the conventional approach in the domain of neuroimaging connectomics, where spatial correspondence is established between a pre-generated atlas defined in some template space and the subject-specific T1-weighted image (see Table 1)[40,42]. Importantly, it is well established that the choice of atlas, and in particular the granularity of gray matter segmentations, may impact findings in brain connectivity studies[45,46]. To address this issue, as shown in Figure 3, we consider a total of 23 cortical and 4 subcortical atlases. This provides researchers with the ability to choose the most appropriate parcellation for their study, investigate connectivity features at multiple spatial scales, and replicate analyses across different parcellation schemes.

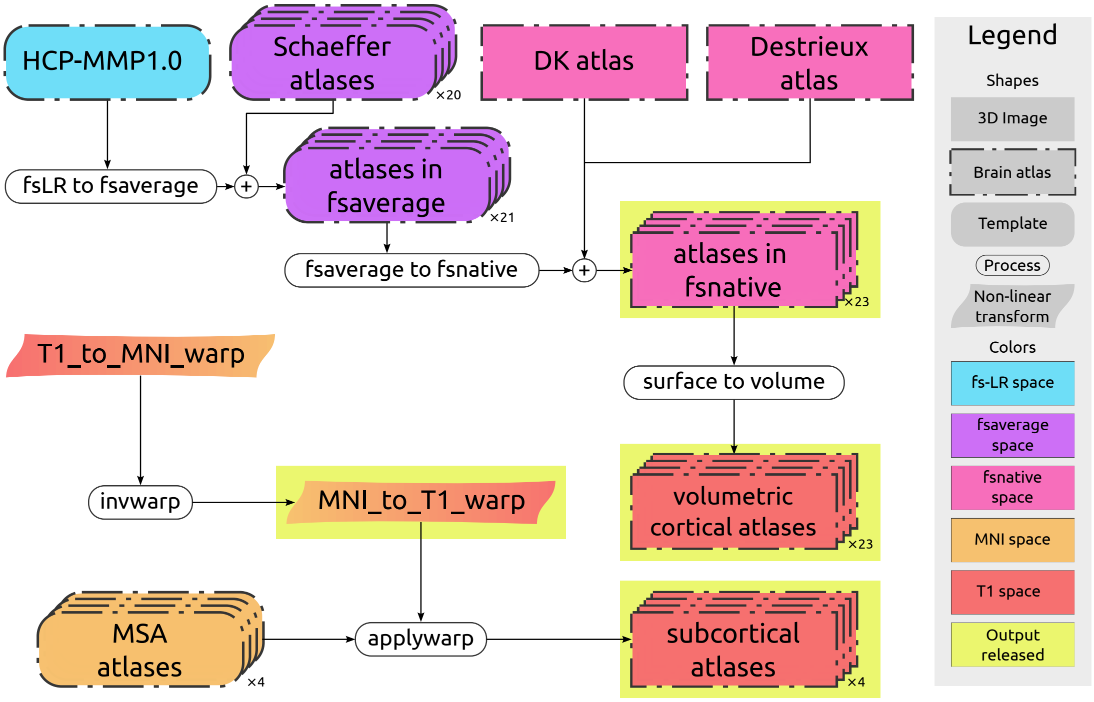


Figure 3. Flowchart for the brain atlas parcellation pipeline. A total of 23 cortical and 4 subcortical atlases, transformed from their respective template spaces, were mapped to each individual's native volumetric space. These volumetric atlas parcellations, as well as additional supporting files (volumetric warp and native surface delineations of cortical atlases), are made available in the atlas parcellations compressed archive (indicated with yellow highlight).

An important detail of our handling of parcellation data is that cortical and subcortical gray matter parcellations were processed entirely independently of one another, and are only combined immediately prior to connectome construction. This permits independent selection of both cortical and subcortical parcellations for any given connectome configuration.

We computed volumetric brain atlases in every individual's native space for a total of 23 cortical and 4 subcortical atlases of the human brain. The cortical atlases included 20 different Scheaffer parcellations[47] (derived from two different sets of functional networks[48] and sub-divided into different parcellation granularities ranging 100–1000 nodes), the Human Connectome Project's multimodal parcellation[49] (HCP-MMP1.0; also known as the Glasser atlas), the Desikan-Killiany atlas[50] (FreeSurfer's "`aparc`"), and the Destrieux Atlas[51] (FreeSurfer's "`aparc.a2009s`"). The subcortical atlases included 4 different spatial scales of the Melbourne Subcortical Atlas (MSA)[52].

All atlases were transformed to each individual's native volumetric space. For the subcortical atlases, the UKB provides non-linear warp files from the subject's T1-weighted image to the MNI152 template[21], which is the space in which the MSA parcellations are defined. We first computed the inverse of this warp (ie. from MNI152 to the subject's T1-weighted image), then applied this transformation to each of the four parcellations using nearest-neighbor interpolation (FSL's "`applywarp`"). To map subject-specific cortical atlases, we started from surface-based labels provided in different template spaces[47,49–51]. For each participant, we aligned these labels to the subject's native cortical surface representation (FreeSurfer's "`fsnative`"), as this provides superior anatomical accuracy compared to volumetric registrations. To transform surface labels into the final volumetric atlases, we first defined, for each subject, a cortical ribbon at the interface between gray and white matter. Lastly, we assigned every voxel in the ribbon the label of its nearest surface vertex. The surface transformations made use of in-house methods incorporating various functionalities from the Connectome Workbench[53], FreeSurfer[24], and NiBabel[54]. All parcellations in subject space (both surface and volumetric representations), the inverse nonlinear MNI warp, and associated conversion scripts (to expedite use of atlases not processed here) are made available (see *Data Records and Code Availability*).

Additional scripts are also provided to combine cortical and subcortical atlases, which are required for mapping SC matrices comprising combinations of cortical and subcortical parcellations beyond the ones readily provided in our resource (see *Structural connectivity: matrices* section for detail).

In some circumstances, both cortical and subcortical sources may attribute a parcel to the same voxel in native space. This occurred most predominantly in the hippocampal region when using the HCP-MMP 1.0 / Glasser atlas. The software used for connectome construction and provided to the research community prioritizes cortical labels where this occurs.

Functional connectivity

FC characterizes statistical dependencies between the BOLD time series recorded from different brain regions[40]. Several computational approaches can be used to map functional connectivity from BOLD signals[4,55]. Here, we provide regionally averaged BOLD time series for the same brain parcellations atlases used to map SC matrices. This enables researchers to easily compute FC matrices using their favorite methods, and also allows for analyses of dynamic and time-varying FC. The flowchart of FC processing is shown in Figure 4. All parcellations were first resampled from the subject's T1-weighted image voxel grid to the subject's preprocessed fMRI voxel grid (FreeSurfer's `mri_vol2vol`). For each parcel, the mean time series across the fMRI voxels ascribed to that region was calculated. By providing these parcellated BOLD time series, we enable flexibility and convenience in mapping functional connectivity across a wide range of alternative spatial resolutions and methodological approaches. For instance, FC derived from statistical correlation-based measures (e.g. Pearson's r) can be directly computed from the time series data in a few milliseconds.

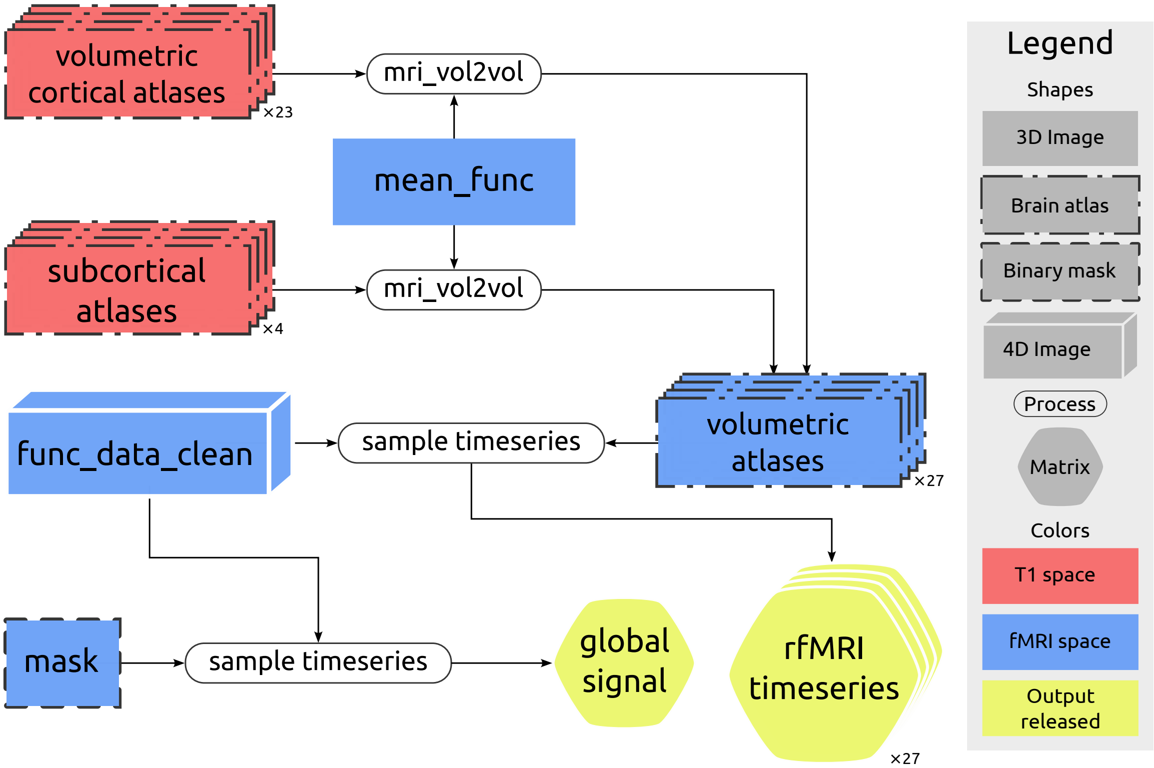


Figure 4. Flowchart for the functional connectivity mapping pipeline. BOLD regional time series are provided for 27 brain parcellation atlases, enabling rapid computation of functional connectivity matrices. The global signal time series was additionally computed. Note: yellow color indicates output resources that are made available in the functional connectivity compressed archive.

Global signal regression (GSR) is an fMRI preprocessing technique with potential merits and drawbacks that are subject to debate[56–64]. We thus calculated the global signal time series and provide those data separately to provide researchers with the flexibility to perform GSR if desired. The global signal was computed by averaging the BOLD time series over all voxels belonging to the anatomical brain mask.

Structural connectivity: tractography

Structural connectomes provide a network representation of the brain white matter axonal architecture[40]. Whole-brain tractography is used to map white matter axonal pathways from an individual's diffusion MRI data and enable estimation of the connectivity properties of these pathways using any of a number of structural connectivity measures. Figure 5 shows how whole-brain tractograms were computed using probabilistic tractography as implemented in *MRtrix3*[44].

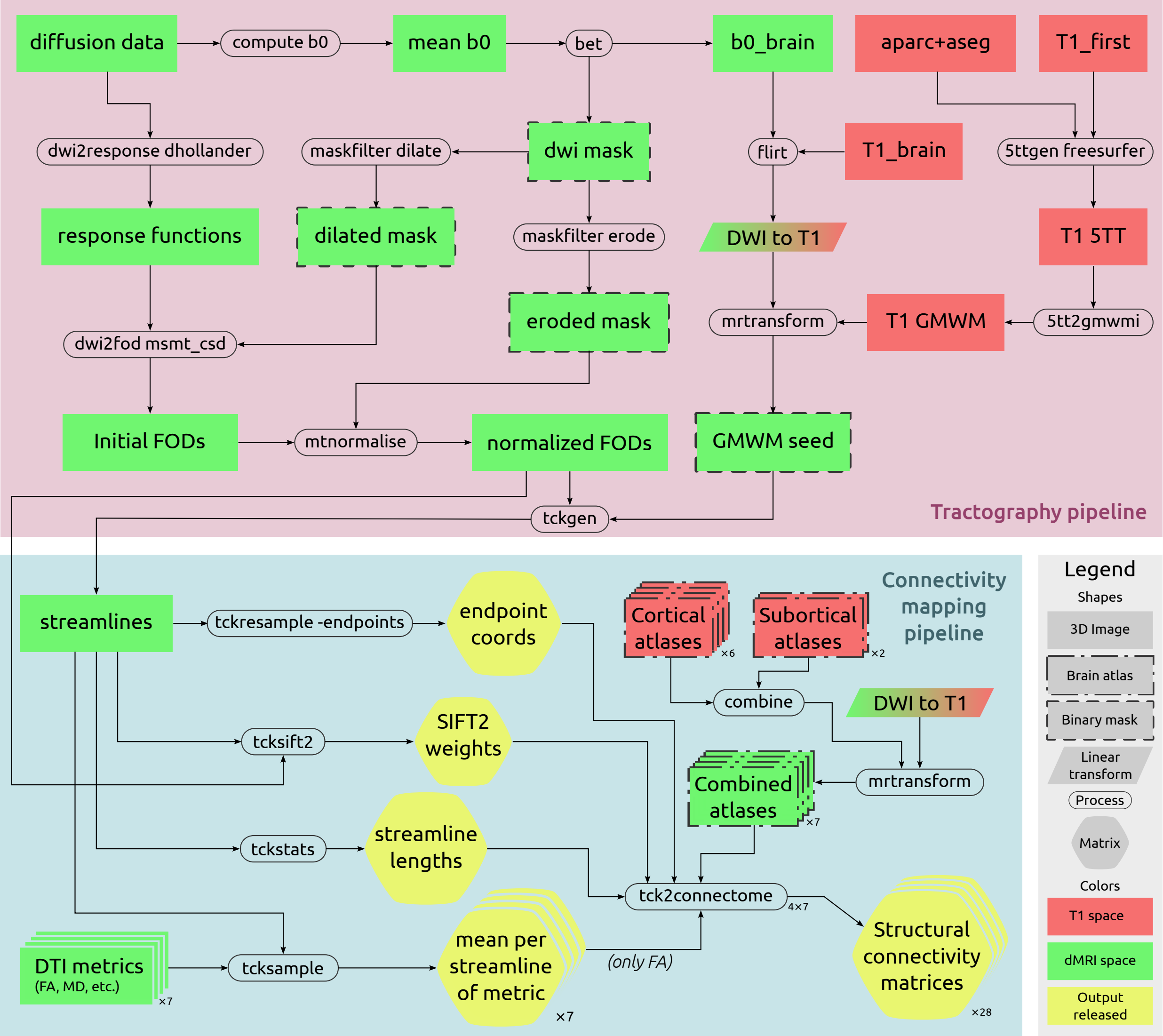


Figure 5. Flowchart for the structural connectivity mapping pipeline. The pipeline is divided into two sections. First, white matter tractography is used to generate streamlines from diffusion-weighted imaging data. Next, these streamlines are used to estimate interregional connectivity properties based on various measures of structural connectivity. The streamline endpoint coordinates, connectivity metrics per streamline, and the subset of 28 connectome matrix configurations that were explicitly generated, are all provided in the structural connectivity compressed archive (shown in yellow).

A brain mask was derived by skull-stripping the mean of DWI volumes without any diffusion weighting (b=0 volumes) and executing the FSL `bet` tool with parameters tuned for diffusion-weighted imaging (DWI) data[18]. In comparison to the DWI masks provided by the UKB, these masks were deemed more accurate and resulted in better registration between the

T1 and DWI spaces. Macroscopic tissue response functions[65] for white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) were estimated with an unsupervised heuristic[66,67]. Multi-shell, multi-tissue (MSMT) constrained spherical deconvolution (CSD)[68] was used to estimate fiber orientation distributions (FODs). This multi-tissue information was used to perform combined intensity normalization and bias field correction[69]. Liberal and conservative brain masks were used respectively for these two steps to mitigate the detrimental influences of imperfect masks in the respective processes.

Whole-brain tractography was performed as follows. A tissue-type segmentation image, intended for use of the Anatomically-Constrained Tractography (ACT) framework[70], was constructed using a combination of the FreeSurfer `aseg` image and the results of FSL FIRST[23]. From this, a mask of the interface between GM and WM was constructed for the purpose of streamline seeding. Probabilistic tractography was performed using 2nd-order integration over Fibre Orientation Distributions (iFOD2)[71]. A total of ten million streamline seeds were drawn throughout the GM-WM interface, and generated streamlines were rejected if they failed to satisfy length constraints or the ACT priors[70]. This constant number of streamline seeds is an important requirement for computational tractability that provides a more robust upper bound on execution time across sessions.

Structural connectivity: matrices

Tractography streamlines and parcellation images were next used to generate SC matrices quantifying various connectivity measures for each regional pair. To generate structural connectomes that are representative of whole-brain connectivity, we used parcellations that integrated both cortical and sub-cortical atlases. As described previously, structural connectivity matrices were pre-calculated for only a subset of all possible connectome configurations; the seven combinations of cortical and subcortical parcellations chosen—for which pre-computed out-of-the-box connectivity matrices are available—are summarized in Table 2. Nevertheless, the provided scripts and supplementary data (e.g. streamline endpoint coordinates and per-streamline metrics) enable connectivity reconstruction for any other possible combination of cortical and subcortical parcellation schemes desired.

Cortical atlas	Cortical regions	Subcortical atlas	Subcortical regions	Total regions
Desikan Killiany	68	MSA - scale I	16	84
Destrieux	148	MSA - scale I	16	164
Glasser	360	MSA - scale I	16	376
Glasser	360	MSA - scale IV	54	414
Schaeffer 200 (7 networks)	200	MSA - scale I	16	216
Schaeffer 500 (7 networks)	500	MSA - scale IV	54	554
Schaeffer 1000 (7 networks)	1000	MSA - scale IV	54	1054

Table 2. The seven combinations of cortical and subcortical parcellations for which structural connectivity matrices were precomputed, and their respective numbers of regions.

Numerous measures of structural connectivity strength are available. The most common measure is the streamline count, which quantifies the total number of streamlines connecting region pairs. Alternatively, post processing algorithms can be applied to ensure that streamline counts better reflect the underlying white matter architecture. We provided per-streamline "weights" calculated by the SIFT2 method to estimate the Fiber Bundle Capacity (FBC) between regions[72]. Additionally, the average length of all streamlines between region pairs were computed to measure connection length. Finally, microstructural parameters (i.e. FA, MD, MO, S0, ICVF, ISOVF, and ODI) were averaged for each streamline trajectory to quantify the microstructural properties of connections.

Regarding the subset of connectome configurations for which we provide pre-computed structural connectivity matrices, for each of the combined parcellation schemes as shown in Table 2, matrices were constructed utilizing the following four metrics:

- Streamline count
- Fiber Bundle Capacity (from SIFT2)
- Mean streamline length
- Mean Fractional Anisotropy (FA)

For *all* structural connectivity metrics (ie. not only those for which matrices were pre-computed), the per-streamline quantitative metrics are provided alongside the locations of the endpoints of streamlines. This facilitates the construction of structural connectomes using any preferred combination of cortical parcellation, subcortical parcellation, and connectivity metric, utilizing the provided connectome generation scripts; this requires both minimal storage (only a pair of 3-vectors and a single floating-point value per streamline) and minimal additional computation (as it is the propagation of streamlines that incurs the greatest expense; assigning streamline endpoints to a parcellation is comparably simple). Considering all different parcellations and connectivity metrics, our code and data resources allow for user-friendly and time-efficient reconstruction of ∼1000 alternate structural connectivity matrices for a single subject. We further note that the data provided in this form are entirely compatible with the adoption of recent developments in the domain such as high-resolution connectomes that consider each surface vertex as its own parcel[73], and the utilization of spatial smoothing of parcels to enhance reliability[74], for which the relevant software tools are also provided.

Computing resources

The whole pipeline was tailored for parallel execution on high-performance computing (HPC) clusters. Parallelization was implemented at the level of individual imaging sessions, with a separate computation job submitted for every session (∼45,000 parallel job submissions). While many of the underlying software tools are capable of executing multiple threads for a single processing job, this was not the case for *all* such tools, and therefore allocating a single CPU core per session was determined to yield the best CPU resource utilization. The maximal memory and wall time per job were empirically minimized to facilitate maximal parallelization on the HPC resource without compromising completion of jobs; this was chosen to be 4GB RAM and 6 hours execution time. The Spartan HPC resource provided by the University of Melbourne[75,76] was utilized for this task, which was typically capable of executing 100–200 such jobs in parallel depending on external utilization.

Time requirements

In addition to the three primary computation steps of the pipeline, each computing job also involved downloading all required UKB data and uploading the resulting derivatives. While data download would typically only require minutes, UKB servers permit only 10 parallel downloads per user, and some jobs could hence experience considerable delays in accessing their requisite data; the wall time allocated for each job was therefore set in order to tolerate such delays without resulting in job failure.

Following data download, the approximate processing time required for each stage of the pipeline was as follows. The atlas mapping pipeline required 10 minutes. Mapping resting functional data required 5 minutes. The most time-consuming step was the structural connectivity reconstruction pipeline, which required 150 minutes. Finally, subsequent conversions and upload steps required 5 minutes. Overall, the complete pipeline required 3–4 hours to finish for a single imaging session.

In total, mapping connectomes across all UKB imaging sessions required ∼200,000 CPU hours (∼20 years) of computation to complete. This substantial time requirement could only be satisfied with the extensive use of HPC resources in parallel execution which reduced the overall required time to map these connectomes from years to months[75].

Storage requirements

The storage requirements for the derivatives of the analysis pipelines are as follows:

- *Atlas parcellation pipeline*: Cortical and subcortical parcellations warped to native subject space, in addition to the non-linear warp from MNI to subject space (to facilitate reserarcher utilisation of other parcellations represented in MNI152 space), all stored in NIfTI format in a compressed archive, requires ∼100MB per session.
- *Functional connectivity pipeline*: The 28 time series (each unique cortical and subcortical parcellation as well as the global signal, as detailed in previous sections) are provided in comma-separated values (CSV) format. These files are provided as a compressed archive with a size of ∼50MB per session.
- *Structural connectivity pipeline*: The combination of parcellations, pre-calculated structural connectome matrices, streamline endpoint coordinates, and per-streamline quantitative metrics for construction of other connectome configurations, are provided as a compressed archive with a size of ∼75MB per session.

As a result, a total of ∼225MB of storage is required to store all derived data per session. The complete set of computed connectivity maps across all sessions occupies ∼10TB of storage.

Data Records

The generated files are organized into three separate compressed archives for every UKB imaging session. Figure 6 provides a summarized list of files provided for a single session. Supplementary Figure S1 provides a more detailed list of all files. All data are made available via UKB data returns policy and will be accessible as new bulk files.

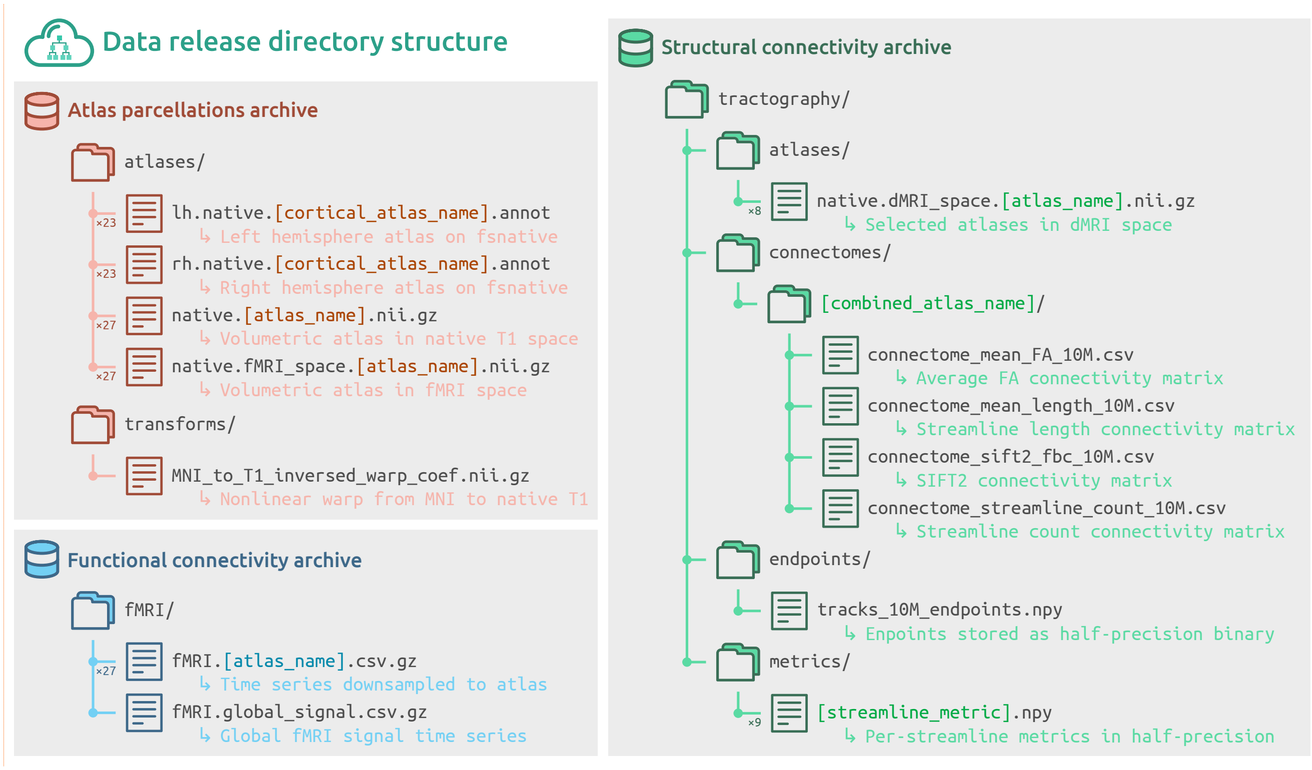


Figure 6. Summary of file archives comprising the connectome resource. Three separate compressed archives are provided for atlas parcellations (red), functional time series (blue), and structural connectivity data (green). Alternative configurations of provided data are summarized by providing a single informative placeholder (colored text in square brackets). A detailed list of all files is provided in Supplementary Information.

Atlas data

The compressed bulk file of atlas data contains derivative parcellations in native volumetric space (`atlases/native.[atlas_name].nii.gz`) for all cortical and subcortical atlases used. Cortical atlases are also provided as parcellated surface data in the FreeSurfer "`fsnative`" space (`atlases/(rh|lh).native.[atlas_name].annot`). The warp image that can be used for nonlinear transformation from MNI space to native coordinates is additionally provided (`transforms/MNI_to_T1_inversed_warp_coef.nii.gz`).

Functional data

The compressed bulk file of functional data contains time series are aggregated within the parcels of various atlases. The time series information is stored as a compressed file with comma-separated values, i.e. the `.csv.gz` format. For each atlas, all sampled time series are provided in a single bulk file (`fMRI/fMRI.[atlas_name].csv.gz`). In addition, the time series for the global signal is included (`fMRI/fMRI.global_signal.csv.gz`).

Diffusion tractography data

The compressed bulk file of diffusion data contains:

- 28 pre-calculated connectivity matrices (all combinations of the combined atlases shown in Table 2 and the four primary connectivity metrics listed in the "Structural connectivity: matrices" section), with paths:(`tractography/connectomes/[cortical_atlas_name]+[subcortical_atlas_name]/connectome_[metric_name]_10M.csv`).
- Spatial locations of all streamline endpoints: (`tractography/endpoints/tracks_10M_endpoints.npy`). This facilitates assignment of pre-generated streamlines to any parcellation of interest. This file is stored in a half-precision floating-point format (16 bits) to reduce storage requirements with minimal loss of precision.
- Nine per-streamline quantitative metrics, with paths: `tractography/metrics/[metric_name].npy`. Metrics include streamline length, SIFT2 weights, and mean values of voxel-wise quantitative metrics along the streamline

trajectories. These are also stored in half-precision floating-point format, with multiplicative factors of $1e^3$ and $1e-3$ applied to the mean MD and mean S0 metrics, respectively, to scale the magnitudes of floating-point values toward unity and therefore mitigate loss of precision.

- Statistics regarding tractogram generation (eg. why streamlines were terminated and why they were accepted or rejected).
- Statistics regarding the operation of the SIFT2 algorithm.

Data Validation

All brain imaging data sourced as inputs passed the automated quality control (QC) evaluations by UKB[8]. We additionally computed various assurance metrics to assess the quality of the provided connectomic resource. These evaluations can be divided into i) QC metrics to probe data quality and ii) analyses of topological properties of brain networks, showing that our connectomes reproduce established findings of the network neuroscience literature.

Quality control

We have computed an extensive set of connectomic QC measures. These metrics complement existing QC efforts provided by the UKB, and can be used to exclude low-quality or inaccurate connectomes. Figure 7 provides a summary of the provided QC measures.

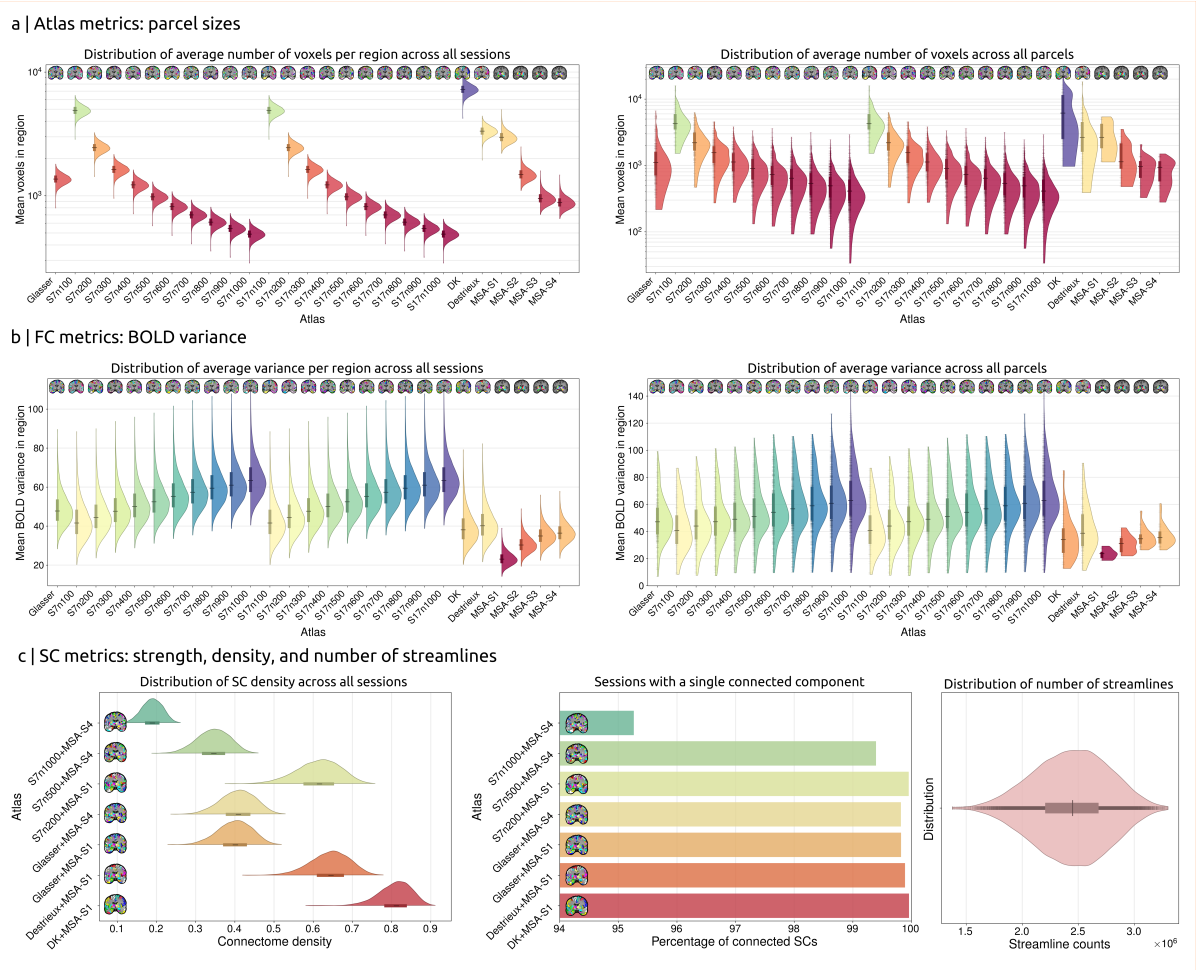


Figure 7. Quality control measures. (a) For atlas parcellations, the parcel sizes were quantified by the number of voxels within each region. Distributions of the number of voxels averaged across atlas regions (left) and across imaging sessions (right) summarize the trends in parcel size for all 27 cortical and subcortical atlases. (b) Similarly, distribution of BOLD signal variance is shown for all 27 atlases. (c) Summary of structural connectivity QC measures for connectome density, components, and reconstructed streamlines for all sessions.

Our QC measures serve as basic features that can be utilized to generate flexible data exclusion criteria tailored to specific research questions. For instance, cases where such metrics lie at the tail of the corresponding distribution could be excluded from analysis, or QC features could be matched between groups. We suggest that these QC features should be tailored to particular study aims, rather than being adopted blindly. For instance, one study may use the metric of total number of streamlines in the connectome to exclude individuals where tractography reconstruction is poor, or account for this effect as a confounding covariate; in contrast, another study aiming to assess the impact of a particular pathology on the structural integrity of the connectome may consider the total number of streamlines to be a variable of interest that should *not* be matched/regressed.

Atlas QC measures

For all volumetric parcellations (cortical and subcortical), the volume of every atlas region in an individual was computed as the total number of voxels assigned to each region. This measure can be used to investigate properties of the native atlas parcellations. Future studies may decide to exclude sessions or parcellations for which certain regions are not appropriately represented, e.g. if a region has no (or very few) voxels. In addition, these measures can optionally be used for normalization of particular structural connectivity matrices: since larger brain regions are more likely to be intersected by streamlines, one may choose to rescale connectivity based on regional volume[2, 77].

Figure 7.a shows a summary of this QC measure in the UKB sample. Violin plots on the left depict the distribution of voxel sizes averaged across all regions and plotted across sessions. Conversely, the plot on the right is averaged across sessions and depicts the distribution across regions. As anticipated, parcel volumes tend to become smaller as the granularity of the parcellation scheme increases. Furthermore, it indicates that a considerable degree of size variation exists between different regions of the same atlas; for instance, the largest regions of Schaeffer's atlas with 1000 parcels are similar in size to regions from the Destrieux atlas with only 148 parcels.

Functional connectivity QC measures

For functional time series, the variance of the aggregated signal within each region is reported across all sessions and atlas regions. This regional measure of variance can be compared to the variance of the global signal to provide an estimate of signal quality[78]. However, it is important to interpret this information with caution, as the standard deviation of an fMRI signal cannot be equated to noise strength and is known to vary with aging and cognition[79, 80]. Signal variance could still be used as a quantitative quality metric to filter out low quality scans.

Figure 7.b provides summary distribution plots of signal variance averaged across sessions and parcels. These summary plots show that relatively higher granularities (e.g. 1000 cortical regions compared to 100 cortical regions) tend to contain signals with larger variation. This is because large parcels sample fMRI over more voxels, which eliminates variance sources from localized effects and unstructured noise. These variance measures for QC could potentially be used as exclusion criteria for sessions in which a region has zero (or very low) variance. A retrospective evaluation of this QC measure indicated a limitation in UKB fMRI preprocessing pipelines impacting signal quality at the orbitofrontal cortex, which is known to be susceptible to BOLD signal loss[81, 82] (see Supplementary Information for further detail).

Structural connectivity QC measures

For structural connectivity, the number of streamlines reconstructed by tractography is reported as a measure of whole-brain tractogram reconstruction efficacy. Given that the number of seeded streamlines is constant for all individuals (10 million), a lower total streamline count indicates higher exclusion rates from streamline acceptance criteria (anatomical validity as determined by ACT and adequate length), which may indicate pathology, poor structural integrity, or low quality data. Another use of this feature is normalizing the connectivity matrices to construct connectomes with equal total strength, which can be more suitable for studying connection probability[83]. In addition, the number of connected components and total connectome density were computed for every structural connectivity matrix. This provides an additional QC feature, as connectomes are expected to form a single connected component and therefore disconnected connectomes could indicate either gross structural abnormalities or poorly reconstructed structural connectivity. Hence, future studies could exclude SC matrices with more than one connected component. As shown in Figure 7.c, SC forms a single connected component for most of the sessions ($> 95\%$ for the highest granularity, and $> 99\%$ for other granularities). Finally, the weighted nodal strength and binary nodal degree are provided, with a binarization threshold of one streamline. Figure 7.c summarizes the QC metrics extracted for structural connectomes.

Properties of structural and functional connectomes

In this last section, we provide an initial exploration of the properties of our structural and functional brain networks. The following analyses sought to replicate previously established results in the network neuroscience literature and, as such, further demonstrate the quality and utility of our connectivity resource.

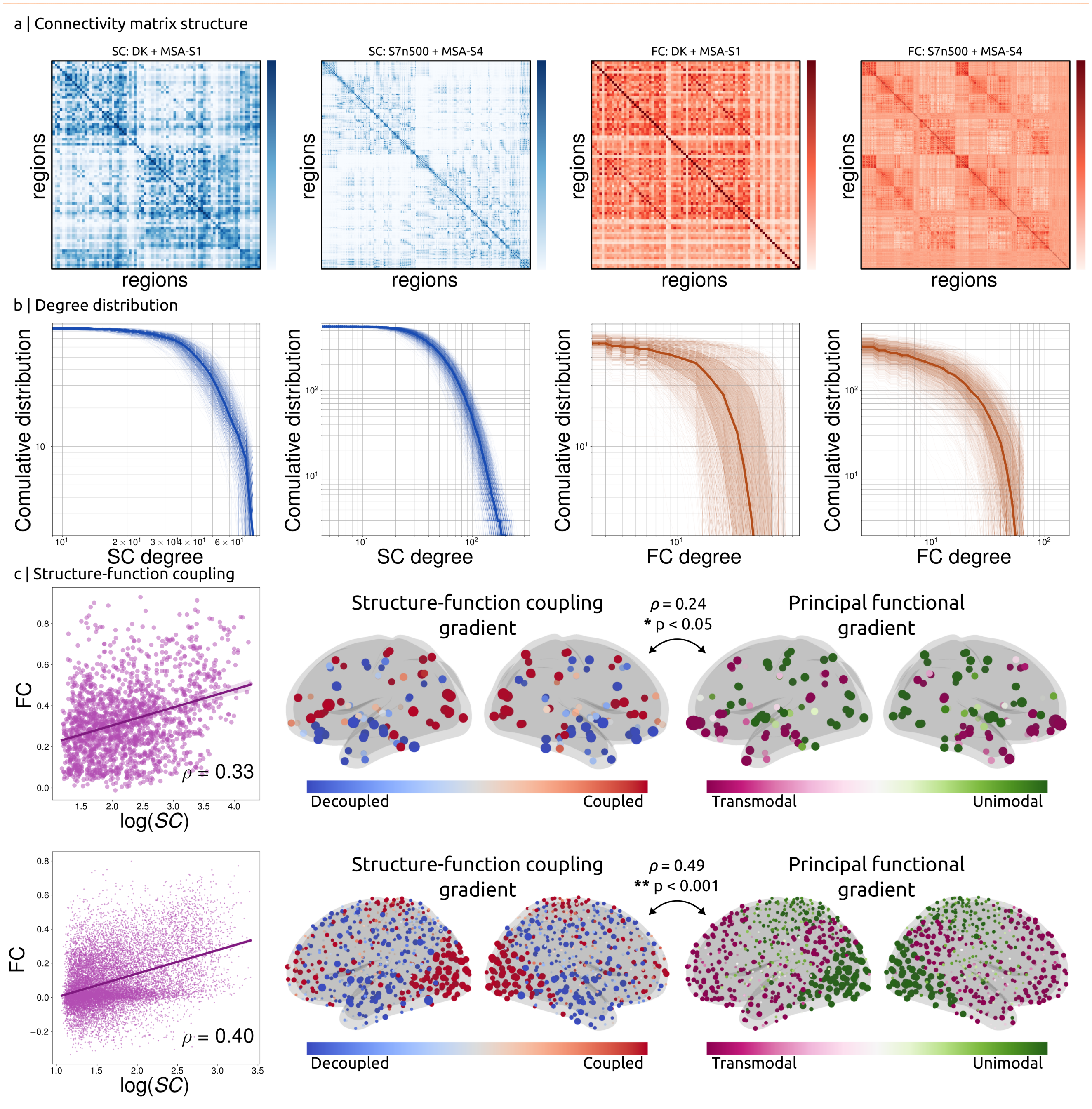


Figure 8. Properties of structural and functional brain networks. (a) Group-level connectomes were derived from 1000 random sessions. The matrices closely resemble standard brain connectivity characteristics with visual distinctions between the cortical hemispheres and the subcortex, local clusters of connectivity within each hemisphere, and strong homotopic connections forming diagonal strides. Results are presented for both SC (blue) and FC (red) matrices in two exemplar parcellations, with "DK + MSA-S1" containing the fewest and "S7n500 + MSA-S4" containing one of the greatest numbers of nodes. (b) Degree distributions of individual connectomes used to construct the group-level matrices. Thin lines indicate trajectories for individual session data; thick line presents the median; shaded regions indicate [25, 75] and [5, 95] centiles. (c) Structure-function coupling was quantified for both exemplar parcellations. Brain-wide coupling was assessed using Pearson's correlation between FC and log(SC) across all edges (scatter plots). Node-wise coupling gradients were computed to project regions to a spectrum of decoupled (blue) to coupled (red) areas. These gradients were compared with FC-derived gradients of functional hierarchy in which regions are situated on a gradient from unimodal (green) to transmodal (purple) areas.

To this end, group-level structural and functional networks were constructed from connectivity data of 1000 randomly sampled individuals. We considered two parcellation combinations comprising a total of 68 (cortex: DK, subcortex: MSA-S1) and 554 (cortex: Schaefer 500, subcortex: MSA-S4) gray matter regions. FC was mapped using GSR and the pairwise Pearson correlation of regional BOLD signals. Group-level FC was computed as the average of functional networks across the 1000 subjects and group-consensus SC was inferred using consistency-based thresholding[84].

Figure 8.a shows the resulting group average SC (blue) and FC (red) matrices for the two selected parcellation schemes. These visualizations align with previous literature[2,85,86] and exhibit features that are consistent with typical connectomes, including i) distinct inter-hemispheric and cortico-subcortical boundaries, ii) a modular structure evident within each hemisphere, and iii) evidence of homotopic connections between the two hemispheres.

Next, we assessed the scale-free network property of the connectomes. We computed the degree distribution of the FC and SC matrices for all sessions by assuming respective binarization thresholds of $\rho = 0.4$ and 1 streamline, respectively. The degree distributions of the resulting matrices are presented in Figure 8.b. The corresponding degree distribution plots replicate previous findings of characteristics observed in structural and functional connectomes with a degree distribution that follows an exponentially truncated power law[86–91].

Finally, we sought to reproduce previous findings of SC-FC coupling (Figure 8.c). The structure and function of the human brain networks are interrelated[92,93]. This relationship can be investigated by comparing the strengths of structural and functional network edges. We present data for three experiments in this regard:

1. *Structure-function correlation*: We assessed the degree of collinearity between FC strengths and the logarithm of SC strength (as quantified by streamline count). The expected positive correlation[94,95] between the strength of structural and functional connectivity was observed for atlases with high ($\rho = 0.4$) and low ($\rho = 0.33$) parcellation granularities.

2. *Structure-function coupling gradient*: The degree of structure-function coupling is reported to vary across the brain, with certain regions exhibiting stronger coupling and others displaying relatively decoupled activity[96–99]. To evaluate structure-function coupling at the level of individual regions, we used a multilinear prediction approach[96]; in short, for each node, we estimated the FC strength to all other nodes via multilinear regression based on four measures of inter-regional distance and structural communication: (i) Euclidean distance, (ii) structural connectivity, (iii) shortest path length, and (iv) communicability[100]. The accuracy of model predictions (quantified by Pearson's correlation) is indicative of local SC-FC coupling strength.

3. *Principal functional gradient*: In prior work, these local patterns of regional coupling have been reported to follow the functional organization hierarchy of unimodal to transmodal brain regions[97,98]. We thus computed the principal functional gradients by performing diffusion map embedding on group-level FC[101] and evaluated its association with the SC-FC coupling gradient. Our results (Figure 8.c) successfully replicated the expected relationship between the coupling patterns and functional organization hierarchy at low ($\rho = 0.24$) and high ($\rho = 0.49$) parcellation granularities.

The evaluations of network properties presented in this section demonstrate that the connectivity matrices provided here reproduce well-established findings in brain connectivity research. This illustrates the high quality of this connectome resource and its potential to facilitate future connectomic studies in an aging population.

Usage Notes

All data will be made available via UKB data returns policies to be accessible based on UKB material transfer agreements. Researchers can apply to access these data by filling out a UKB access application. Additional code and data (such as the label ordering of atlases) are made openly available in a publicly accessible git repository. All data are provided in formats that are readable in various programming languages. This enables use of several existing software packages for brain connectivity analysis such as the Brain Connectivity Toolbox (Matlab)[102], bctpy (Python), and Nilearn (Python), as well as general network analysis tools such as NetworkX[103].

Code availability

All scripts used to perform computations described in this manuscript (eg. generating atlas parcellations, aggregating BOLD time series, tractography, computing structural connectivity, etc.) are made publicly available in the git repository served at `github.com/sina-mansour/UKB-connectomics`. Additional scripts used to perform QC evaluations are also provided in this repository. Furthermore, sample scripts to read the generated data into various programming languages are made available to facilitate uptake of this resource by the community.

References

1. Valdés-Hernández, P. *et al.* Characterizing brain anatomical connections using diffusion weighted MRI and graph theory. *NeuroImage* **36**, 645–660, 10.1016/j.neuroimage.2007.02.012 (2007).

2. Hagmann, P. *et al.* Mapping the Structural Core of Human Cerebral Cortex. *PLoS Biol.* **6**, e159, 10.1371/journal.pbio.0060159 (2008).

3. van den Heuvel, M. P. & Hulshoff Pol, H. E. Exploring the brain network: A review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.* **20**, 519–534, 10.1016/j.euroneuro.2010.03.008 (2010).

4. Smith, S. M. *et al.* Network modelling methods for FMRI. *NeuroImage* **54**, 875–891, 10.1016/j.neuroimage.2010.08.063 (2011).

5. Sporns, O. Making sense of brain network data. *Nat. Methods* **10**, 491–493, 10.1038/nmeth.2485 (2013).

6. Bullmore, E. & Sporns, O. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10**, 186–198, 10.1038/nrn2575 (2009).

7. Van Essen, D. C. *et al.* The Human Connectome Project: A data acquisition perspective. *NeuroImage* **62**, 2222–2231, 10.1016/j.neuroimage.2012.02.018 (2012).

8. Alfaro-Almagro, F. *et al.* Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* **166**, 400–424, 10.1016/j.neuroimage.2017.10.034 (2018).

9. Avesani, P. *et al.* The open diffusion data derivatives, brain data upcycling via integrated publishing of derivatives and reproducible open cloud services. *Sci. Data* **6**, 69, 10.1038/s41597-019-0073-y (2019).

10. Mascarell Maričić, L. *et al.* The IMAGEN study: a decade of imaging genetics in adolescents. *Mol. Psychiatry* **25**, 2648–2671, 10.1038/s41380-020-0822-5 (2020).

11. Karcher, N. R. & Barch, D. M. The ABCD study: understanding the development of risk for mental and physical health outcomes. *Neuropsychopharmacology* **46**, 131–142, 10.1038/s41386-020-0736-6 (2021).

12. Markiewicz, C. J. *et al.* The openneuro resource for sharing of neuroscience data. *eLife* **10**, 10.7554/eLife.71774 (2021).

13. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine* **12**, e1001779, 10.1371/journal.pmed.1001779 (2015).

14. Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536, 10.1038/nn.4393 (2016).

15. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209, 10.1038/s41586-018-0579-z (2018).

16. Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W. & Smith, S. M. FSL. *NeuroImage* **62**, 782–790, 10.1016/j.neuroimage.2011.09.015 (2012).

17. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: An overview. *NeuroImage* **80**, 62–79, 10.1016/j.neuroimage.2013.05.041 (2013).

18. Smith, S. M. Fast robust automated brain extraction. *Hum. Brain Mapp.* **17**, 143–155, 10.1002/hbm.10062 (2002).

19. Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of brain images. *Med. Image Analysis* **5**, 143–156, 10.1016/S1361-8415(01)00036-6 (2001).

20. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage* **17**, 825–841, 10.1006/nimg.2002.1132 (2002).

21. Andersson, J. L. R., Jenkinson, M. & Smith, S. Non-linear registration, aka spatial normalization. *FMRIB technical report TR07JA2* (2010).

22. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Med. Imaging* **20**, 45–57, 10.1109/42.906424 (2001).

23. Patenaude, B., Smith, S. M., Kennedy, D. N. & Jenkinson, M. A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage* **56**, 907–922, 10.1016/j.neuroimage.2011.02.046 (2011).

24. Fischl, B. FreeSurfer. *NeuroImage* **62**, 774–781, 10.1016/j.neuroimage.2012.01.021 (2012).

25. Moeller, S. *et al.* Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain FMRI. *Magn. Reson. Medicine* **63**, 1144–1153, 10.1002/mrm.22361 (2010).

26. Setsompop, K. *et al.* Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magn. Reson. Medicine* **67**, 1210–1224, 10.1002/mrm.23097 (2012).

27. Preibisch, C., Castrillón G., J. G., Bührer, M. & Riedl, V. Evaluation of multiband EPI acquisitions for resting state fMRI. *PLoS ONE* **10**, 10.1371/journal.pone.0136961 (2015).

28. Beckmann, C. & Smith, S. Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging. *IEEE Transactions on Med. Imaging* **23**, 137–152, 10.1109/TMI.2003.822821 (2004).

29. Salimi-Khorshidi, G. *et al.* Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage* **90**, 449–468, 10.1016/j.neuroimage.2013.11.046 (2014).

30. Griffanti, L. *et al.* ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage* **95**, 232–247, 10.1016/j.neuroimage.2014.03.034 (2014).

31. Sotiropoulos, S. N. *et al.* Advances in diffusion MRI acquisition and processing in the Human Connectome Project. *NeuroImage* **80**, 125–143, 10.1016/j.neuroimage.2013.05.057 (2013).

32. Andersson, J. L., Skare, S. & Ashburner, J. How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage* **20**, 870–888, 10.1016/S1053-8119(03)00336-7 (2003).

33. Andersson, J. L. & Sotiropoulos, S. N. Non-parametric representation and prediction of single- and multi-shell diffusion-weighted MRI data using Gaussian processes. *NeuroImage* **122**, 166–176, 10.1016/j.neuroimage.2015.07.067 (2015).

34. Andersson, J. L. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage* **125**, 1063–1078, 10.1016/j.neuroimage.2015.10.019 (2016).

35. Andersson, J. L., Graham, M. S., Zsoldos, E. & Sotiropoulos, S. N. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *NeuroImage* **141**, 556–572, 10.1016/j.neuroimage.2016.06.058 (2016).

36. Glasser, M. F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80**, 105–124, 10.1016/j.neuroimage.2013.04.127 (2013).

37. Basser, P., Mattiello, J. & Lebihan, D. Estimation of the Effective Self-Diffusion Tensor from the NMR Spin Echo. *J. Magn. Reson. Ser. B* **103**, 247–254, 10.1006/jmrb.1994.1037 (1994).

38. Zhang, H., Schneider, T., Wheeler-Kingshott, C. A. & Alexander, D. C. NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. *NeuroImage* **61**, 1000–1016, 10.1016/j.neuroimage.2012.03.072 (2012).

39. Brett, M., Johnsrude, I. S. & Owen, A. M. The problem of functional localization in the human brain. *Nat. Rev. Neurosci.* **3**, 243–249, 10.1038/nrn756 (2002).

40. Fornito, A., Zalesky, A. & Breakspear, M. Graph analysis of the human connectome: Promise, progress, and pitfalls. *NeuroImage* **80**, 426–444, 10.1016/j.neuroimage.2013.04.087 (2013).

41. Zhang, F. *et al.* Quantitative mapping of the brain's structural connectivity using diffusion MRI tractography: A review. *NeuroImage* **249**, 118870, 10.1016/j.neuroimage.2021.118870 (2022).

42. Sporns, O. The human connectome: Origins and challenges. *NeuroImage* **80**, 53–61, 10.1016/j.neuroimage.2013.03.023 (2013).

43. Bullmore, E. T. & Bassett, D. S. Brain Graphs: Graphical Models of the Human Brain Connectome. *Annu. Rev. Clin. Psychol.* **7**, 113–140, 10.1146/annurev-clinpsy-040510-143934 (2011).

44. Tournier, J.-D. *et al.* MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage* **202**, 116137, 10.1016/j.neuroimage.2019.116137 (2019).

45. Zalesky, A. *et al.* Whole-brain anatomical networks: Does the choice of nodes matter? *NeuroImage* **50**, 970–983, 10.1016/j.neuroimage.2009.12.027 (2010).

46. Fornito, A., Zalesky, A. & Bullmore, E. T. Network scaling effects in graph analytic studies of human resting-state fMRI data. *Front. Syst. Neurosci.* **4**, 1–16, 10.3389/fnsys.2010.00022 (2010).

47. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb. Cortex* **28**, 3095–3114, 10.1093/cercor/bhx179 (2018).

48. Thomas Yeo, B. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165, 10.1152/jn.00338.2011 (2011).

49. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178, 10.1038/nature18933 (2016).

50. 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, 10.1016/j.neuroimage.2006.01.021 (2006).

51. Destrieux, C., Fischl, B., Dale, A. & Halgren, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* **53**, 1–15, 10.1016/j.neuroimage.2010.06.010 (2010).

52. Tian, Y., Margulies, D. S., Breakspear, M. & Zalesky, A. Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nat. Neurosci.* **23**, 1421–1432, 10.1038/s41593-020-00711-6 (2020).

53. Marcus, D. S. *et al.* Human Connectome Project informatics: Quality control, database services, and data visualization. *NeuroImage* **80**, 202–219, 10.1016/j.neuroimage.2013.05.077 (2013).

54. Brett, M. *et al.* nipy/nibabel. *Zenodo* 10.5281/zenodo.591597 (2020).

55. Li, K., Guo, L., Nie, J., Li, G. & Liu, T. Review of methods for functional brain connectivity detection using fMRI. *Comput. Med. Imaging Graph.* **33**, 131–139, 10.1016/j.compmedimag.2008.10.011 (2009).

56. Saad, Z. S. *et al.* Trouble at Rest: How Correlation Patterns and Group Differences Become Distorted After Global Signal Regression. *Brain Connect.* **2**, 25–32, 10.1089/brain.2012.0080 (2012).

57. Yan, C. G., Craddock, R. C., Zuo, X. N., Zang, Y. F. & Milham, M. P. Standardizing the intrinsic brain: Towards robust measurement of inter-individual variation in 1000 functional connectomes. *NeuroImage* **80**, 246–262, 10.1016/j.neuroimage.2013.04.081 (2013).

58. Gotts, S. J. *et al.* The perils of global signal regression for group comparisons: A case study of Autism Spectrum Disorders. *Front. Hum. Neurosci.* **7**, 1–20, 10.3389/fnhum.2013.00356 (2013).

59. Murphy, K. & Fox, M. D. Towards a consensus regarding global signal regression for resting state functional connectivity MRI. *NeuroImage* **154**, 169–173, 10.1016/j.neuroimage.2016.11.052 (2017).

60. Liu, T. T., Nalci, A. & Falahpour, M. The global signal in fMRI: Nuisance or Information? *NeuroImage* **150**, 213–229, 10.1016/j.neuroimage.2017.02.036 (2017).

61. Power, J. D., Plitt, M., Laumann, T. O. & Martin, A. Sources and implications of whole-brain fMRI signals in humans. *NeuroImage* **146**, 609–625, 10.1016/j.neuroimage.2016.09.038 (2017).

62. Billings, J. & Keilholz, S. The Not-So-Global Blood Oxygen Level-Dependent Signal. *Brain Connect.* **8**, 121–128, 10.1089/brain.2017.0517 (2018).

63. Li, J. *et al.* Global signal regression strengthens association between resting-state functional connectivity and behavior. *NeuroImage* **196**, 126–141, 10.1016/j.neuroimage.2019.04.016 (2019).

64. Aquino, K. M., Fulcher, B. D., Parkes, L., Sabaroedin, K. & Fornito, A. Identifying and removing widespread signal deflections from fMRI data: Rethinking the global signal regression problem. *NeuroImage* **212**, 116614, 10.1016/j.neuroimage.2020.116614 (2020).

65. Tournier, J. D., Calamante, F., Gadian, D. G. & Connelly, A. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage* **23**, 1176–1185, 10.1016/j.neuroimage.2004.07.037 (2004).

66. Dhollander, T., Raffelt, D. & Connelly, A. Unsupervised 3-tissue response function estimation from single-shell or multi-shell diffusion MR data without a co-registered T1 image. In *ISMRM Workshop on Breaking the Barriers of Diffusion MRI*, 5 (2016).

67. Dhollander, T., Mito, R., Raffelt, D. & Connelly, A. Improved white matter response function estimation for 3-tissue constrained spherical deconvolution. *Proc. Intl. Soc. Mag. Reson. Med.* **555** (2019).

68. 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, 10.1016/j.neuroimage.2014.07.061 (2014).

69. Dhollander, T. *et al.* Multi-tissue log-domain intensity and inhomogeneity normalisation for quantitative apparent fibre density. *Proc. ISMRM* **29**, 2472 (2021).

70. 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, 10.1016/j.neuroimage.2012.06.005 (2012).

71. Tournier, J. D., Calamante, F. & Connelly, A. Improved probabilistic streamlines tractography by 2 nd order integration over fibre orientation distributions. *Ismrm* **88**, 2010 (2010).

72. Smith, R., Raffelt, D., Tournier, J.-D. & Connelly, A. Quantitative streamlines tractography: methods and inter-subject normalisation. *Aperture Neuro* **2**, 1–23, 10.52294/ApertureNeuro.2022.2.NEOD9565 (2022).

73. Mansour L, S., Tian, Y., Yeo, B. T., Cropley, V. & Zalesky, A. High-resolution connectomic fingerprints: Mapping neural identity and behavior. *NeuroImage* **229**, 117695, 10.1016/j.neuroimage.2020.117695 (2021).

74. Mansour L, S., Seguin, C., Smith, R. E. & Zalesky, A. Connectome spatial smoothing (CSS): Concepts, methods, and evaluation. *NeuroImage* **250**, 118930, 10.1016/j.neuroimage.2022.118930 (2022).

75. Meade, B., Lafayette, L., Sauter, G. & Tosello, D. Spartan HPC-Cloud Hybrid: Delivering Performance and Flexibility, 10.4225/49/58ead90dceaaa (2017).

76. Lafayette, L. & Wiebelt, B. Spartan and NEMO: Two HPC-cloud hybrid implementations. *Proc. - 13th IEEE Int. Conf. on eScience, eScience 2017* 458–459, 10.1109/eScience.2017.70 (2017).

77. Sotiropoulos, S. N. & Zalesky, A. Building connectomes using diffusion MRI: why, how and but. *NMR Biomed.* **32**, 1–23, 10.1002/nbm.3752 (2019).

78. Welvaert, M. & Rosseel, Y. On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data. *PLoS ONE* **8**, 10.1371/journal.pone.0077089 (2013).

79. Garrett, D. D., Kovacevic, N., McIntosh, A. R. & Grady, C. L. Blood oxygen level-dependent signal variability is more than just noise. *J. Neurosci.* **30**, 4914–4921, 10.1523/JNEUROSCI.5166-09.2010 (2010).

80. Garrett, D. D., Kovacevic, N., McIntosh, A. R. & Grady, C. L. The importance of being variable. *J. Neurosci.* **31**, 4496–4503, 10.1523/JNEUROSCI.5641-10.2011 (2011).

81. Jezzard, P. & Clare, S. Sources of distortion in functional MRI data. *Hum. Brain Mapp.* **8**, 80–85, 10.1002/(SICI)1097-0193(1999)8:2/3<80::AID-HBM2>3.0.CO;2-C (1999).

82. Stenger, V. A. Technical considerations for BOLD fMRI of the orbitofrontal cortex. In *The Orbitofrontal Cortex*, 423–446, 10.1093/acprof:oso/9780198565741.003.0017 (Oxford University PressOxford, 2006).

83. Jbabdi, S. & Johansen-Berg, H. Tractography: Where Do We Go from Here? *Brain Connect.* **1**, 169–183, 10.1089/brain.2011.0033 (2011).

84. Roberts, J. A., Perry, A., Roberts, G., Mitchell, P. B. & Breakspear, M. Consistency-based thresholding of the human connectome. *NeuroImage* **145**, 118–129, 10.1016/j.neuroimage.2016.09.053 (2017).

85. Fornito, A., Zalesky, A. & Bullmore, E. T. *Fundamentals of Brain Network Analysis* (Elsevier, 2016).

86. Gong, G. *et al.* Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cereb. Cortex* **19**, 524–536, 10.1093/cercor/bhn102 (2009).

87. Bassett, D. S. & Bullmore, E. Small-world brain networks. *Neuroscientist* **12**, 512–523, 10.1177/1073858406293182 (2006).

88. Achard, S., Salvador, R., Whitcher, B., Suckling, J. & Bullmore, E. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* **26**, 63–72, 10.1523/JNEUROSCI.3874-05.2006 (2006).

89. Iturria-Medina, Y., Sotero, R. C., Canales-Rodríguez, E. J., Alemán-Gómez, Y. & Melie-García, L. Studying the human brain anatomical network via diffusion-weighted MRI and Graph Theory. *NeuroImage* **40**, 1064–1076, 10.1016/j.neuroimage.2007.10.060 (2008).

90. Hayasaka, S. & Laurienti, P. J. Comparison of characteristics between region-and voxel-based network analyses in resting-state fMRI data. *NeuroImage* **50**, 499–508, 10.1016/j.neuroimage.2009.12.051 (2010).

91. Joyce, K. E., Laurienti, P. J., Burdette, J. H. & Hayasaka, S. A new measure of centrality for brain networks. *PLoS ONE* **5**, 10.1371/journal.pone.0012200 (2010).

92. Honey, C. J., Kötter, R., Breakspear, M. & Sporns, O. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc. Natl. Acad. Sci. United States Am.* **104**, 10240–10245, 10.1073/pnas.0701519104 (2007).

93. 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. *Hum. Brain Mapp.* **30**, 3127–3141, 10.1002/hbm.20737 (2009).

94. Honey, C. J. *et al.* Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci. United States Am.* **106**, 2035–2040, 10.1073/pnas.0811168106 (2009).

95. Honey, C. J., Thivierge, J. P. & Sporns, O. Can structure predict function in the human brain? *NeuroImage* **52**, 766–776, 10.1016/j.neuroimage.2010.01.071 (2010).

96. Vázquez-Rodríguez, B. *et al.* Gradients of structure–function tethering across neocortex. *Proc. Natl. Acad. Sci. United States Am.* **116**, 21219–21227, 10.1073/pnas.1903403116 (2019).

97. Preti, M. G. & Van De Ville, D. Decoupling of brain function from structure reveals regional behavioral specialization in humans. *Nat. Commun.* **10**, 1–7, 10.1038/s41467-019-12765-7 (2019).

98. Suárez, L. E., Markello, R. D., Betzel, R. F. & Misic, B. Linking Structure and Function in Macroscale Brain Networks. *Trends Cogn. Sci.* **24**, 302–315, 10.1016/j.tics.2020.01.008 (2020).

99. Baum, G. L. *et al.* Development of structure–function coupling in human brain networks during youth. *Proc. Natl. Acad. Sci.* **117**, 771–778, 10.1073/pnas.1912034117 (2020).

100. Estrada, E. & Hatano, N. Communicability in complex networks. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **77**, 1–12, 10.1103/PhysRevE.77.036111 (2008).

101. Margulies, D. S. *et al.* Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci.* **113**, 12574–12579, 10.1073/pnas.1608282113 (2016).

102. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage* **52**, 1059–1069, 10.1016/j.neuroimage.2009.10.003 (2010).

103. Hagberg, A., Swart, P. & Chult, D. S. Exploring network structure, dynamics, and function using NetworkX (2008).

Acknowledgements

This project was funded by CS's Early Career Researcher Grant from the University of Melbourne. The data analysis was supported by Spartan High-Performance Computing infrastructure[75, 76], and dedicated data storage solutions provided by the Research Computing Services at the University of Melbourne. RS is a fellow of the National Imaging Facility, a National Collaborative Research Infrastructure Strategy (NCRIS) capability, at the Florey Institute of Neuroscience and Mental Health.

Author contributions statement

S.M.L.: Pipeline development, Software, Data curation, Formal analysis, Writing - original draft, Writing - review & editing. **M.D.B**: Pipeline development, Writing - original draft, Writing - review & editing. **R.S.**: Pipeline development, Software, Writing - original draft, Writing - review & editing. **A.Z.**: Supervision, Pipeline development, Software, Writing - original draft, Writing - review & editing. **C.S.**: Supervision, Funding acquisition, Pipeline development, Writing - original draft, Writing - review & editing.

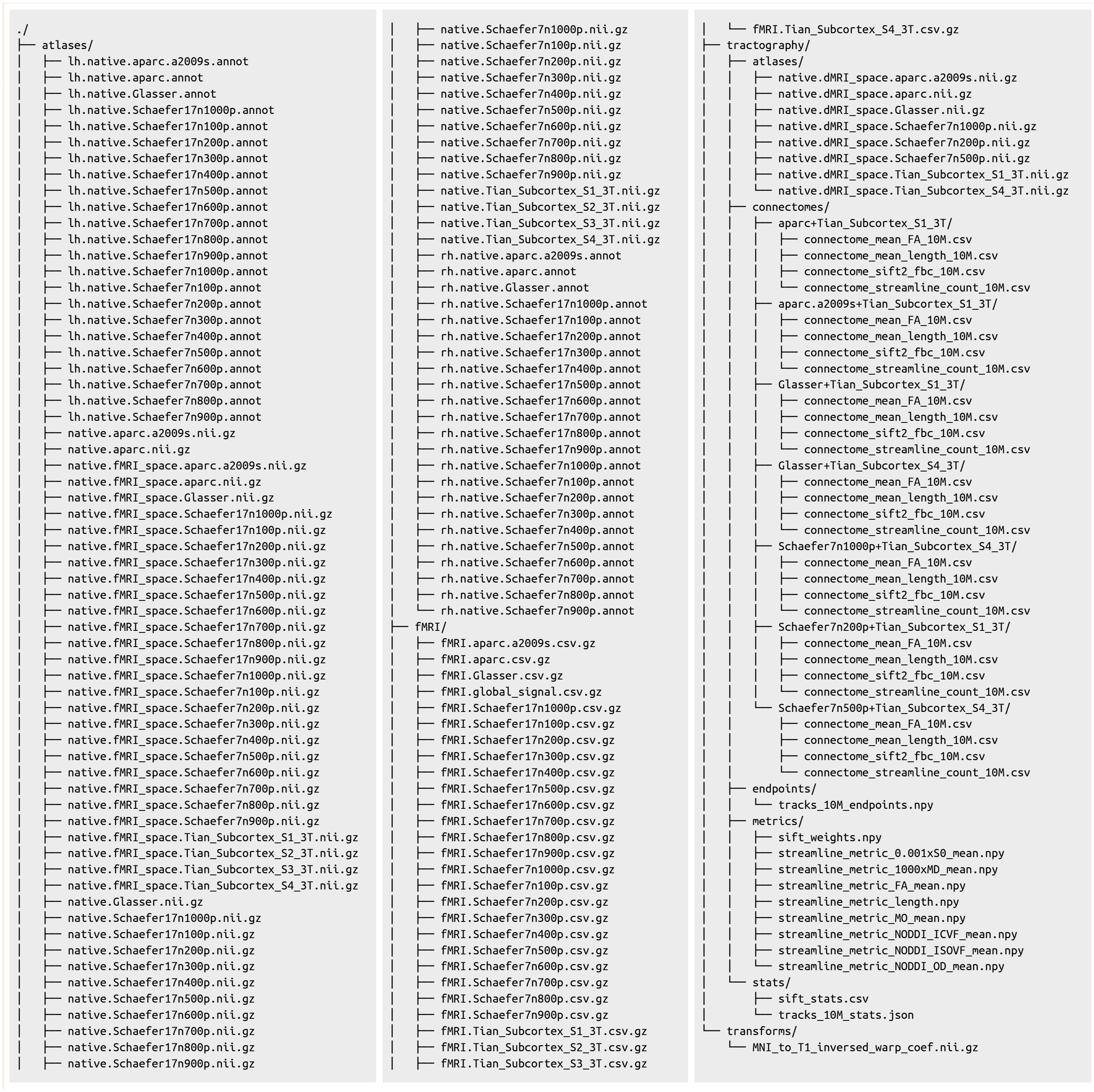
Competing interests

The authors declare no competing interests.

Supplementary Information

Comprehensive directory tree

A summarized directory tree was provided in the main text to list all released data. Figure S1 provides an extensive list of all files provided in the compressed archives of a session.



```
./
├── atlases/
│   ├── lh.native.aparc.a2009s.annot
│   ├── lh.native.aparc.annot
│   ├── lh.native.Glasser.annot
│   ├── lh.native.Schaefer17n1000p.annot
│   ├── lh.native.Schaefer17n100p.annot
│   ├── lh.native.Schaefer17n200p.annot
│   ├── lh.native.Schaefer17n300p.annot
│   ├── lh.native.Schaefer17n400p.annot
│   ├── lh.native.Schaefer17n500p.annot
│   ├── lh.native.Schaefer17n600p.annot
│   ├── lh.native.Schaefer17n700p.annot
│   ├── lh.native.Schaefer17n800p.annot
│   ├── lh.native.Schaefer17n900p.annot
│   ├── lh.native.Schaefer7n1000p.annot
│   ├── lh.native.Schaefer7n100p.annot
│   ├── lh.native.Schaefer7n200p.annot
│   ├── lh.native.Schaefer7n300p.annot
│   ├── lh.native.Schaefer7n400p.annot
│   ├── lh.native.Schaefer7n500p.annot
│   ├── lh.native.Schaefer7n600p.annot
│   ├── lh.native.Schaefer7n700p.annot
│   ├── lh.native.Schaefer7n800p.annot
│   ├── lh.native.Schaefer7n900p.annot
│   ├── native.aparc.a2009s.nii.gz
│   ├── native.aparc.nii.gz
│   ├── native.fMRI_space.aparc.a2009s.nii.gz
│   ├── native.fMRI_space.aparc.nii.gz
│   ├── native.fMRI_space.Glasser.nii.gz
│   ├── native.fMRI_space.Schaefer17n1000p.nii.gz
│   ├── native.fMRI_space.Schaefer17n100p.nii.gz
│   ├── native.fMRI_space.Schaefer17n200p.nii.gz
│   ├── native.fMRI_space.Schaefer17n300p.nii.gz
│   ├── native.fMRI_space.Schaefer17n400p.nii.gz
│   ├── native.fMRI_space.Schaefer17n500p.nii.gz
│   ├── native.fMRI_space.Schaefer17n600p.nii.gz
│   ├── native.fMRI_space.Schaefer17n700p.nii.gz
│   ├── native.fMRI_space.Schaefer17n800p.nii.gz
│   ├── native.fMRI_space.Schaefer17n900p.nii.gz
│   ├── native.fMRI_space.Schaefer7n1000p.nii.gz
│   ├── native.fMRI_space.Schaefer7n100p.nii.gz
│   ├── native.fMRI_space.Schaefer7n200p.nii.gz
│   ├── native.fMRI_space.Schaefer7n300p.nii.gz
│   ├── native.fMRI_space.Schaefer7n400p.nii.gz
│   ├── native.fMRI_space.Schaefer7n500p.nii.gz
│   ├── native.fMRI_space.Schaefer7n600p.nii.gz
│   ├── native.fMRI_space.Schaefer7n700p.nii.gz
│   ├── native.fMRI_space.Schaefer7n800p.nii.gz
│   ├── native.fMRI_space.Schaefer7n900p.nii.gz
│   ├── native.fMRI_space.Tian_Subcortex_S1_3T.nii.gz
│   ├── native.fMRI_space.Tian_Subcortex_S2_3T.nii.gz
│   ├── native.fMRI_space.Tian_Subcortex_S3_3T.nii.gz
│   ├── native.fMRI_space.Tian_Subcortex_S4_3T.nii.gz
│   ├── native.Glasser.nii.gz
│   ├── native.Schaefer17n1000p.nii.gz
│   ├── native.Schaefer17n100p.nii.gz
│   ├── native.Schaefer17n200p.nii.gz
│   ├── native.Schaefer17n300p.nii.gz
│   ├── native.Schaefer17n400p.nii.gz
│   ├── native.Schaefer17n500p.nii.gz
│   ├── native.Schaefer17n600p.nii.gz
│   ├── native.Schaefer17n700p.nii.gz
│   ├── native.Schaefer17n800p.nii.gz
│   ├── native.Schaefer17n900p.nii.gz
│   ├── native.Schaefer7n1000p.nii.gz
│   ├── native.Schaefer7n100p.nii.gz
│   ├── native.Schaefer7n200p.nii.gz
│   ├── native.Schaefer7n300p.nii.gz
│   ├── native.Schaefer7n400p.nii.gz
│   ├── native.Schaefer7n500p.nii.gz
│   ├── native.Schaefer7n600p.nii.gz
│   ├── native.Schaefer7n700p.nii.gz
│   ├── native.Schaefer7n800p.nii.gz
│   ├── native.Schaefer7n900p.nii.gz
│   ├── native.Tian_Subcortex_S1_3T.nii.gz
│   ├── native.Tian_Subcortex_S2_3T.nii.gz
│   ├── native.Tian_Subcortex_S3_3T.nii.gz
│   ├── native.Tian_Subcortex_S4_3T.nii.gz
│   ├── rh.native.aparc.a2009s.annot
│   ├── rh.native.aparc.annot
│   ├── rh.native.Glasser.annot
│   ├── rh.native.Schaefer17n1000p.annot
│   ├── rh.native.Schaefer17n100p.annot
│   ├── rh.native.Schaefer17n200p.annot
│   ├── rh.native.Schaefer17n300p.annot
│   ├── rh.native.Schaefer17n400p.annot
│   ├── rh.native.Schaefer17n500p.annot
│   ├── rh.native.Schaefer17n600p.annot
│   ├── rh.native.Schaefer17n700p.annot
│   ├── rh.native.Schaefer17n800p.annot
│   ├── rh.native.Schaefer17n900p.annot
│   ├── rh.native.Schaefer7n1000p.annot
│   ├── rh.native.Schaefer7n100p.annot
│   ├── rh.native.Schaefer7n200p.annot
│   ├── rh.native.Schaefer7n300p.annot
│   ├── rh.native.Schaefer7n400p.annot
│   ├── rh.native.Schaefer7n500p.annot
│   ├── rh.native.Schaefer7n600p.annot
│   ├── rh.native.Schaefer7n700p.annot
│   ├── rh.native.Schaefer7n800p.annot
│   └── rh.native.Schaefer7n900p.annot
├── fMRI/
│   ├── fMRI.aparc.a2009s.csv.gz
│   ├── fMRI.aparc.csv.gz
│   ├── fMRI.Glasser.csv.gz
│   ├── fMRI.global_signal.csv.gz
│   ├── fMRI.Schaefer17n1000p.csv.gz
│   ├── fMRI.Schaefer17n100p.csv.gz
│   ├── fMRI.Schaefer17n200p.csv.gz
│   ├── fMRI.Schaefer17n300p.csv.gz
│   ├── fMRI.Schaefer17n400p.csv.gz
│   ├── fMRI.Schaefer17n500p.csv.gz
│   ├── fMRI.Schaefer17n600p.csv.gz
│   ├── fMRI.Schaefer17n700p.csv.gz
│   ├── fMRI.Schaefer17n800p.csv.gz
│   ├── fMRI.Schaefer17n900p.csv.gz
│   ├── fMRI.Schaefer7n1000p.csv.gz
│   ├── fMRI.Schaefer7n100p.csv.gz
│   ├── fMRI.Schaefer7n200p.csv.gz
│   ├── fMRI.Schaefer7n300p.csv.gz
│   ├── fMRI.Schaefer7n400p.csv.gz
│   ├── fMRI.Schaefer7n500p.csv.gz
│   ├── fMRI.Schaefer7n600p.csv.gz
│   ├── fMRI.Schaefer7n700p.csv.gz
│   ├── fMRI.Schaefer7n800p.csv.gz
│   ├── fMRI.Schaefer7n900p.csv.gz
│   ├── fMRI.Tian_Subcortex_S1_3T.csv.gz
│   ├── fMRI.Tian_Subcortex_S2_3T.csv.gz
│   ├── fMRI.Tian_Subcortex_S3_3T.csv.gz
│   └── fMRI.Tian_Subcortex_S4_3T.csv.gz
├── tractography/
│   ├── atlases/
│   │   ├── native.dMRI_space.aparc.a2009s.nii.gz
│   │   ├── native.dMRI_space.aparc.nii.gz
│   │   ├── native.dMRI_space.Glasser.nii.gz
│   │   ├── native.dMRI_space.Schaefer7n1000p.nii.gz
│   │   ├── native.dMRI_space.Schaefer7n200p.nii.gz
│   │   ├── native.dMRI_space.Schaefer7n500p.nii.gz
│   │   ├── native.dMRI_space.Tian_Subcortex_S1_3T.nii.gz
│   │   └── native.dMRI_space.Tian_Subcortex_S4_3T.nii.gz
│   ├── connectomes/
│   │   ├── aparc+Tian_Subcortex_S1_3T/
│   │   │   ├── connectome_mean_FA_10M.csv
│   │   │   ├── connectome_mean_length_10M.csv
│   │   │   ├── connectome_sift2_fbc_10M.csv
│   │   │   └── connectome_streamline_count_10M.csv
│   │   ├── aparc.a2009s+Tian_Subcortex_S1_3T/
│   │   │   ├── connectome_mean_FA_10M.csv
│   │   │   ├── connectome_mean_length_10M.csv
│   │   │   ├── connectome_sift2_fbc_10M.csv
│   │   │   └── connectome_streamline_count_10M.csv
│   │   ├── Glasser+Tian_Subcortex_S1_3T/
│   │   │   ├── connectome_mean_FA_10M.csv
│   │   │   ├── connectome_mean_length_10M.csv
│   │   │   ├── connectome_sift2_fbc_10M.csv
│   │   │   └── connectome_streamline_count_10M.csv
│   │   ├── Glasser+Tian_Subcortex_S4_3T/
│   │   │   ├── connectome_mean_FA_10M.csv
│   │   │   ├── connectome_mean_length_10M.csv
│   │   │   ├── connectome_sift2_fbc_10M.csv
│   │   │   └── connectome_streamline_count_10M.csv
│   │   ├── Schaefer7n1000p+Tian_Subcortex_S4_3T/
│   │   │   ├── connectome_mean_FA_10M.csv
│   │   │   ├── connectome_mean_length_10M.csv
│   │   │   ├── connectome_sift2_fbc_10M.csv
│   │   │   └── connectome_streamline_count_10M.csv
│   │   ├── Schaefer7n200p+Tian_Subcortex_S1_3T/
│   │   │   ├── connectome_mean_FA_10M.csv
│   │   │   ├── connectome_mean_length_10M.csv
│   │   │   ├── connectome_sift2_fbc_10M.csv
│   │   │   └── connectome_streamline_count_10M.csv
│   │   └── Schaefer7n500p+Tian_Subcortex_S4_3T/
│   │       ├── connectome_mean_FA_10M.csv
│   │       ├── connectome_mean_length_10M.csv
│   │       ├── connectome_sift2_fbc_10M.csv
│   │       └── connectome_streamline_count_10M.csv
│   ├── endpoints/
│   │   └── tracks_10M_endpoints.npy
│   ├── metrics/
│   │   ├── sift_weights.npy
│   │   ├── streamline_metric_0.001xS0_mean.npy
│   │   ├── streamline_metric_1000xMD_mean.npy
│   │   ├── streamline_metric_FA_mean.npy
│   │   ├── streamline_metric_length.npy
│   │   ├── streamline_metric_MO_mean.npy
│   │   ├── streamline_metric_NODDI_ICVF_mean.npy
│   │   ├── streamline_metric_NODDI_ISOVF_mean.npy
│   │   └── streamline_metric_NODDI_OD_mean.npy
│   └── stats/
│       ├── sift_stats.csv
│       └── tracks_10M_stats.json
└── transforms/
    └── MNI_to_T1_inversed_warp_coef.nii.gz
```

Figure S1. Complete list of all files included in provided archives. Three separate compressed archives are provided, for each of atlas parcellation (↪`atlases/*`), functional time series (↪`fMRI/*`), and structural connectivity (↪`tractography/*`) pipelines.

Resting-state signal loss

The BOLD signal was collected using a gradient echo EPI acquisition which can experience geometric distortion as well as signal loss in regions of strong susceptibility[81]. Computing a brain mask from empirical fMRI data, using any method based on image intensity, may lead to exclusion of such regions. All voxels outside of this brain mask were explicitly zero-filled in UKB fMRI preprocessing pipeline[8]. For a region that can be clearly delineated on an effectively undistorted anatomical image, but its entire volume resides outside of the fMRI brain mask, the aggregate fMRI signal within that region would consequently be zero-filled. These cases can be identified by computing the QC measure of regional signal variance. Hence, signal variance can be used as a proxy to assess signal loss due to exclusion of susceptible regions.

To evaluate the cases in which UKB preprocessing led to such problems, we counted the number of sessions for which a variance of zero (constant signal) was reported for a brain region. This information was then sorted to pinpoint the regions with highest detrimental impact of signal loss. Table S1 provides a list of percentages of signal loss for 100 mostly affected brain regions (across different atlases) sorted based on signal loss severity. This indicated a pattern of severe signal loss within the orbitofrontal cortex (OFC). The signal loss severity was higher for higher granularity levels such that the cortical atlases with 1000 regions had signal loss in more than a third of the cohort. However, the severity was considerably lower for lower granularities such that atlases with fewer than 700 regions had signal loss in less than 5% of the cohort.

We hereby report this issue as a limitation of the fMRI acquisition and preprocessing and suggest either i) conducting the analysis at lower granularities (especially if OFC is of interest), or ii) excluding the impacted regions of OFC in cases where the analyses are to be conducted on higher granularities. Finally, to verify that signal loss was due to impacts of preprocessing on OFC and not a result of the connectivity mapping pipeline manual visual inspections were conducted for a handful of sessions. Figure S2 provides an example that shows a complete lack of signal in the OFC parcel (`Schaefer7n1000p:7Networks_LH_Limbic_OFC_6`) due to zero-filling leading to a zero variance QC metric.

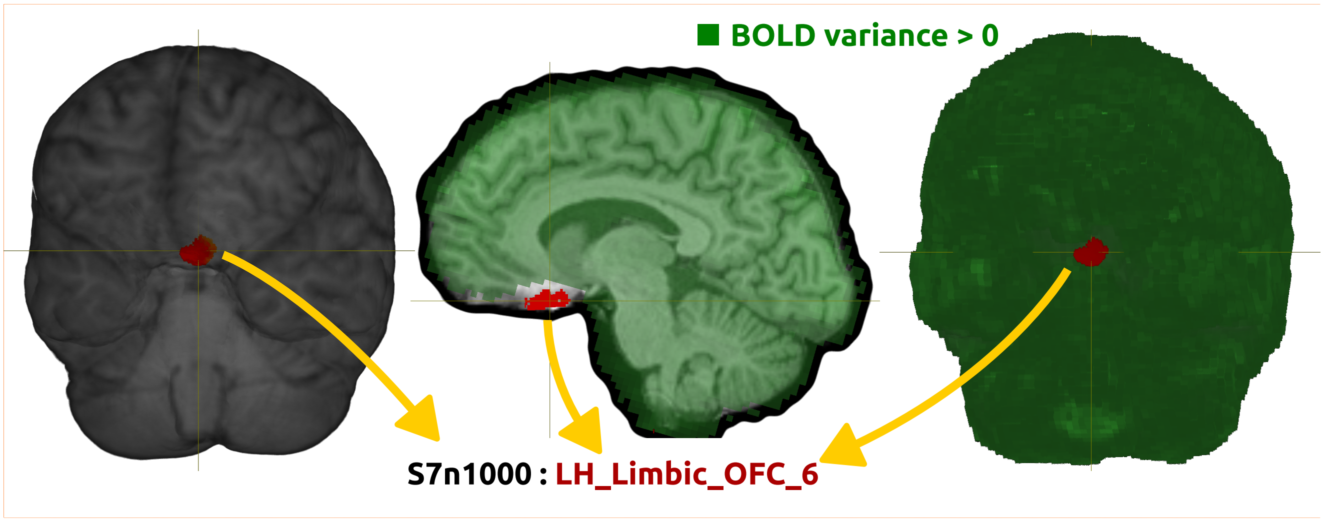


Figure S2. Illustration of signal loss near certain OFC regions for a single session. The `LH_Limbic_OFC_6` regions from `S7n1000` atlas is highlighted (red) on a 3D rendered T1 brain [left]. A sagittal slice of the brain showing the same parcel (red) along with a highlighted map of regions within the rfMRI brain mask (green) for which signal variance is positive [middle]. A similar 3D render showing the particular OFC parcel falling outside of the fMRI mask [right]. The images indicate that the preprocessed fMRI signal fails to cover certain cortical gray matter regions belonging to the OFC.

Table S1. One hundred regions that were most severely affected by BOLD signal loss. The percentage of sessions for which zero variance was recorded in a certain region was quantified. This quantifies the frequency by which a region falls outside of the fMRI brain mask due to effects of signal loss.

Signal loss Table	
Atlas:Parcel	Percentage of signal loss (over all sessions)
Schaefer7n1000p:7Networks_LH_Limbic_OFC_6	36.5
Schaefer17n1000p:17Networks_LH_LimbicB_OFC_6	36.5
Schaefer7n1000p:7Networks_LH_Limbic_OFC_3	15
Schaefer17n1000p:17Networks_LH_LimbicB_OFC_3	15
Schaefer7n1000p:7Networks_RH_Limbic_OFC_4	13.6
Schaefer7n900p:7Networks_RH_Limbic_OFC_4	13.6
Schaefer17n1000p:17Networks_RH_LimbicB_OFC_4	13.6
Schaefer17n900p:17Networks_RH_LimbicB_OFC_4	13.6
Schaefer17n900p:17Networks_LH_LimbicB_OFC_7	11.4
Schaefer7n900p:7Networks_LH_Limbic_OFC_7	11.4
Schaefer7n1000p:7Networks_LH_Limbic_OFC_11	10.2
Schaefer17n1000p:17Networks_LH_LimbicB_OFC_11	10.2
Schaefer17n800p:17Networks_LH_LimbicB_OFC_7	8.4
Schaefer7n800p:7Networks_LH_Limbic_OFC_7	8.4
Schaefer7n1000p:7Networks_LH_Limbic_OFC_2	7.9
Schaefer17n1000p:17Networks_LH_LimbicB_OFC_2	7.9
Schaefer7n600p:7Networks_RH_Limbic_OFC_2	7.5
Schaefer17n600p:17Networks_RH_LimbicB_OFC_2	7.5
Schaefer17n700p:17Networks_RH_LimbicB_OFC_2	7.5
Schaefer7n700p:7Networks_RH_Limbic_OFC_2	7.5
Schaefer17n800p:17Networks_RH_LimbicB_OFC_2	6.6
Schaefer7n800p:7Networks_RH_Limbic_OFC_2	6.6
Schaefer7n800p:7Networks_RH_Limbic_OFC_7	6.3
Schaefer17n800p:17Networks_RH_LimbicB_OFC_9	6.3
Schaefer7n900p:7Networks_RH_Limbic_OFC_3	6.3
Schaefer17n900p:17Networks_RH_LimbicB_OFC_3	6.3
Schaefer7n1000p:7Networks_LH_Limbic_OFC_4	6.3
Schaefer17n1000p:17Networks_LH_LimbicB_OFC_4	6.3
Schaefer7n1000p:7Networks_RH_Limbic_OFC_8	6.2
Schaefer17n1000p:17Networks_RH_LimbicB_OFC_10	6.2
Schaefer17n700p:17Networks_LH_LimbicB_OFC_1	5.8
Schaefer7n700p:7Networks_LH_Limbic_OFC_1	5.8
Schaefer7n900p:7Networks_LH_Limbic_OFC_1	5.4
Schaefer17n900p:17Networks_LH_LimbicB_OFC_1	5.4
Schaefer7n800p:7Networks_RH_Limbic_OFC_4	5.3
Schaefer17n800p:17Networks_RH_LimbicB_OFC_4	5.3
Schaefer7n900p:7Networks_RH_Limbic_OFC_8	5.3
Schaefer17n900p:17Networks_RH_LimbicB_OFC_8	5.3
Schaefer7n700p:7Networks_LH_Limbic_OFC_3	4.5
Schaefer17n700p:17Networks_LH_LimbicB_OFC_3	4.5
Schaefer7n1000p:7Networks_RH_Limbic_OFC_3	3.9
Schaefer17n1000p:17Networks_RH_LimbicB_OFC_3	3.9
Schaefer7n600p:7Networks_RH_Limbic_OFC_6	3.9
Schaefer17n600p:17Networks_RH_LimbicB_OFC_7	3.9
Schaefer7n800p:7Networks_RH_Limbic_OFC_3	3.7
Schaefer17n800p:17Networks_RH_LimbicB_OFC_3	3.7
Schaefer7n900p:7Networks_LH_Limbic_OFC_3	3.2
Schaefer17n900p:17Networks_LH_LimbicB_OFC_3	3.2
Schaefer17n800p:17Networks_LH_LimbicB_OFC_2	2.9

Continuation of signal loss Table S1

Atlas:Parcel	Percentage of signal loss (over all sessions)
Schaefer7n800p:7Networks_LH_Limbic_OFC_2	2.9
Schaefer7n700p:7Networks_RH_Limbic_OFC_3	2.6
Schaefer17n700p:17Networks_RH_LimbicB_OFC_3	2.6
Schaefer7n600p:7Networks_RH_Limbic_OFC_3	2.6
Schaefer17n600p:17Networks_RH_LimbicB_OFC_3	2.6
Schaefer7n700p:7Networks_LH_Limbic_OFC_5	2.6
Schaefer17n700p:17Networks_LH_LimbicB_OFC_5	2.6
Schaefer17n600p:17Networks_LH_LimbicB_OFC_4	2.5
Schaefer7n600p:7Networks_LH_Limbic_OFC_4	2.5
Schaefer17n800p:17Networks_LH_LimbicB_OFC_11	2.4
Schaefer7n800p:7Networks_LH_Limbic_OFC_10	2.4
Glasser:L_pOFC_ROI	2.4
Schaefer17n800p:17Networks_LH_LimbicB_OFC_4	2
Schaefer7n800p:7Networks_LH_Limbic_OFC_4	2
Schaefer7n500p:7Networks_RH_Limbic_OFC_2	1.9
Schaefer17n500p:17Networks_RH_LimbicB_OFC_2	1.9
Schaefer7n800p:7Networks_RH_Limbic_OFC_9	1.6
Schaefer7n400p:7Networks_RH_Limbic_OFC_5	1.6
Schaefer17n800p:17Networks_RH_LimbicB_OFC_11	1.6
Schaefer17n400p:17Networks_RH_LimbicB_OFC_5	1.6
Schaefer7n400p:7Networks_LH_Limbic_OFC_1	1.4
Schaefer17n400p:17Networks_LH_LimbicB_OFC_1	1.4
Schaefer17n1000p:17Networks_LH_LimbicB_OFC_12	1.3
Schaefer7n1000p:7Networks_LH_Limbic_OFC_12	1.3
Schaefer17n500p:17Networks_LH_LimbicB_OFC_1	1.3
Schaefer7n500p:7Networks_LH_Limbic_OFC_1	1.3
Schaefer17n600p:17Networks_LH_LimbicB_OFC_2	1.2
Schaefer7n600p:7Networks_LH_Limbic_OFC_2	1.2
Schaefer7n1000p:7Networks_LH_Limbic_OFC_1	1.2
Schaefer17n1000p:17Networks_LH_LimbicB_OFC_1	1.2
Schaefer7n900p:7Networks_RH_Limbic_OFC_2	1.2
Schaefer17n1000p:17Networks_RH_LimbicB_OFC_2	1.2
Schaefer7n1000p:7Networks_RH_Limbic_OFC_2	1.2
Schaefer17n900p:17Networks_RH_LimbicB_OFC_2	1.2
Schaefer7n800p:7Networks_LH_Limbic_OFC_1	0.9
Schaefer17n800p:17Networks_LH_LimbicB_OFC_1	0.9
Schaefer17n1000p:17Networks_RH_LimbicB_OFC_5	0.8
Schaefer7n1000p:7Networks_RH_Limbic_OFC_5	0.8
Glasser:R_25_ROI	0.8
Schaefer7n900p:7Networks_RH_Limbic_OFC_5	0.7
Schaefer17n900p:17Networks_RH_LimbicB_OFC_5	0.7
Schaefer17n900p:17Networks_RH_LimbicA_TempPole_2	0.7
Schaefer7n900p:7Networks_RH_Limbic_TempPole_2	0.7
Glasser:L_25_ROI	0.6
Schaefer17n700p:17Networks_RH_LimbicB_OFC_4	0.6
Schaefer7n700p:7Networks_RH_Limbic_OFC_4	0.6
Schaefer7n500p:7Networks_LH_Limbic_OFC_5	0.5
Schaefer17n500p:17Networks_LH_LimbicB_OFC_5	0.5
Schaefer17n900p:17Networks_LH_LimbicA_TempPole_7	0.5
Schaefer7n900p:7Networks_LH_Limbic_TempPole_7	0.5
Schaefer7n700p:7Networks_LH_Limbic_TempPole_12	0.5