Extended Supplementary Methods For

A Multi-modal Parcellation of Human Cerebral Cortex

Matthew F. Glasser[1], Timothy S. Coalson[1]*, Emma C. Robinson[2,3]*, Carl D. Hacker[4]*, John Harwell[1], Essa Yacoub[5], Kamil Ugurbil[5], Jesper Andersson[2], Christian F. Beckmann[6], Mark Jenkinson[2], Stephen M. Smith[2], David C. Van Essen[1]

[1]Department of Neuroscience, Washington University Medical School, Saint Louis, MO, USA; [2]FMRIB centre, Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, University of Oxford, Oxford, UK; [3]Department of Computing, Imperial College London, UK; [4]Department of Biomedical Engineering, Washington University, Saint Louis, MO, USA; [5]Center for Magnetic Resonance Research (CMRR), University of Minnesota, Minneapolis, MN, USA; [6]Donders Institute for Brain, Cognition and Behavior, Radboud University, Nijmegen, The Netherlands & MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, Enschede, The Netherlands

*Co-second Authors

1. Subjects, Image Acquisition, and Initial Image Preprocessing

1.1 Subjects

We used the "HCP500" dataset (June, 2014 data release) of adults aged 22-35 drawn from a pool of twins and their non-twin siblings ((Van Essen et al., 2013); http://www.humanconnectome.org/documentation/S500/). All 449 subjects that had complete structural (at least one T1w and one T2w scan), rfMRI (4 runs X 15 minutes), and tfMRI (7 tasks, 60 minutes total) datasets were included. The subjects were divided into two main analysis groups: 210 parcellation subjects (the '210P' group) and 210 validation subjects (the '210V' group). Group independence was preserved by ensuring that no families were split between the 210P and 210V groups (210 subjects per group was the maximum number for evenly sized complete datasets from this release where all members of each family were assigned to either one group or the other). A third auxiliary group of 29 subjects (the 29T group, including some family relationships shared with the 210P group) was used only in training the cortical areal classifier as the initial 'test' dataset to prevent overfitting of random noise (see #6.6). Additionally, 27 of the 449 subjects were scanned using the entire HCP MRI acquisition protocol a second time, reanalyzed, and used to check the reproducibility of the individual subject parcellations. The HCP data were acquired using protocols approved by the Washington University institutional review board, deidentified, and publicly shared on the ConnectomeDB database (https://db.humanconnectome.org; (Hodge et al., 2015)).

1.2 Image Acquisition

The HCP MRI data acquisition has previously been described in detail ((Glasser et al., 2013; Smith et al., 2013a; Ugurbil et al., 2013);

http://www.humanconnectome.org/documentation/S500/HCP_S500+MEG2_Release_Appendix_I.pdf). Images were acquired using a customized 3T Siemens 'Connectom' Skyra having a 100mT/m SC72 gradient insert and using a standard Siemens 32-channel RF-receive head coil. At least one good quality 3D T1w MPRAGE image (as assessed by a trained rater, (Marcus et al., 2013)) was acquired at 0.7 mm isotropic resolution. At least one good quality 3D T2w SPACE image was also acquired at 0.7 mm isotropic resolution in the same session as the T1w image. Whole brain rfMRI and tfMRI data were acquired using identical multi-band EPI sequence parameters of 2 mm isotropic resolution with a TR=720 ms. Field maps were acquired during both the structural and fMRI scanning sessions to enable accurate cross-modal registrations of the T2w and fMRI images to the T1w image in each subject. Additionally, the spin echo phase reversed images acquired during the fMRI session (with matched geometry and echo spacing to the gradient echo fMRI data) were used to compute a more accurate fMRI bias field correction and to segment regions of gradient echo signal loss.

1.3 Data analysis software

Data were analyzed using the publicly released HCP pipelines (Glasser et al., 2013), plus additional pipelines for cross-subject registration and individual subject parcellation that have been or will soon be released. The software packages used for analysis included Connectome Workbench commandline tools, FSL, and FreeSurfer (Fischl, 2012; Jenkinson et al., 2012). Pipelines were written in Bash shell scripts and Matlab (https://github.com/Washington-University/Pipelines). Connectome Workbench 'wb_view' GUI (http://www.humanconnectome.org/software/connectome-workbench.html) was used for all visualization and for some interactive data analysis. All neuroanatomical (brain-related) figures for this study were generated in Connectome Workbench using the Tile Tabs (to make multiple figure panels) and Annotations (to label panels, areas, and features) capabilities and were directly exported as PNG image files, bypassing the need for image editing software (these tools are available in Connectome Workbench version 1.2.0 or higher). More importantly, these figures were also saved as Connectome Workbench "scene" files, to allow anyone who downloads the scene and associated data files to reproduce an interactive version of each figure in Connectome Workbench for further exploration of the data and comparison to their own data. The scene files and associated data files are available in the BALSA database (Brain Analysis Library of Spatial maps and Atlases) ((Van Essen et al., 2016); http://balsa.wustl.edu) to facilitate comparisons across studies and with future data, and they are directly accessible via URLs in the relevant figure legends.

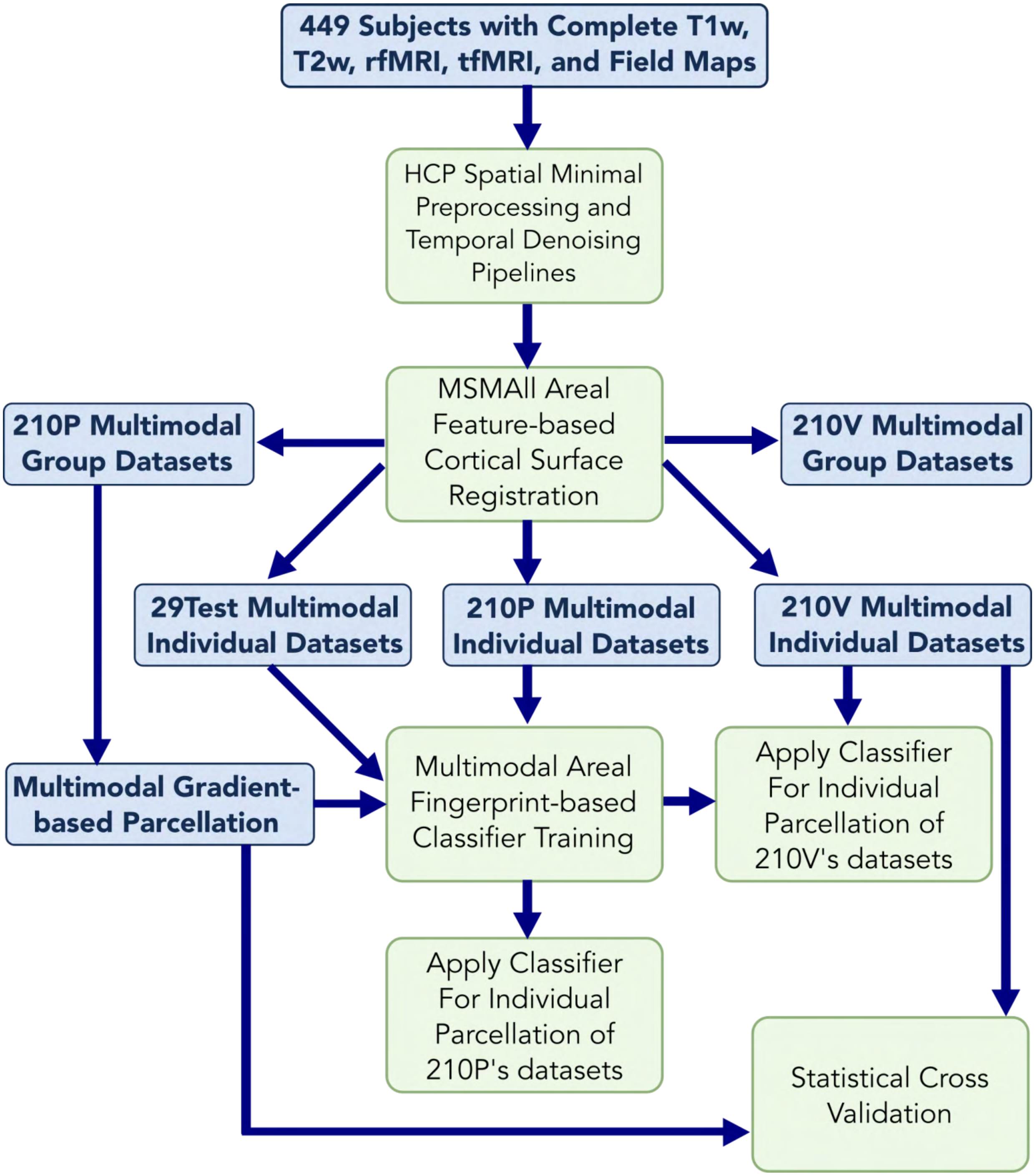


Figure 1 shows the overall workflow of the paper. After initial spatial and temporal preprocessing and areal-feature-based registration, the data were divided into groups, and group average datasets were produced. The 210P group dataset was used for semi-automated multi-modal gradient-based parcellation. This parcellation was cross-validated using the 210V individual subject datasets. The 210P individual datasets were used for training of the multi-modal areal fingerprint-based individual subject areal classifier, along with the 29T test dataset. The 210P and 210V datasets were parcellated by applying the trained classifier to their individual subject data, and group maximum probability parcellations were generated. In this and analogous figures, datasets are light blue and processes are light green.

1.4 Overview of data analysis methods

Many methods were developed *de novo* or were improved upon to produce a high quality multi-modal parcellation at the group level and to use this parcellation to identify individual subject cortical areas. Methods described in previous HCP-related publications are only summarized briefly here, whereas new or substantially improved methods are described in more detail. Figure 1 provides an overview of the major processing steps/stages that are described in subsequent sections. The minimally preprocessed (Glasser et al., 2013) data were downloaded from ConnectomeDB, and MSMAll areal-feature-based surface registration was performed on 449 subjects. The multi-modal data were then split into parcellation and validation groups of 210 subjects each, and group averages were created. The 210P subjects' group average data were used to generate a multi-modal gradient-based parcellation, and their individual data were used together with 29 additional test subjects to train a machine learning classifier to delineate and identify cortical areas in individual subjects using their multi-modal areal fingerprints. This trained classifier was applied to the 210P and 210V datasets to generate individual subject parcellations and group maximum probability maps. The 210V dataset was used for statistical cross-validation of the 210P-based original parcellation.

1.5 Initial Preprocessing: Structural

The HCP structural preprocessing pipelines (Glasser et al., 2013) include PreFreeSurfer, FreeSurfer, and PostFreeSurfer components and were run inside ConnectomeDB for the 500-subject HCP release. The PreFreeSurfer pipeline includes: (i) aligning and averaging repeated T1w and T2w scans of good or excellent quality when they exist; (ii) removing gradient nonlinearity and readout distortion (b0 distortion in 3D images) to create an unbiased "native" volume space for each subject that is rigidly aligned to the MNI template; (iii) cross-modal alignment between the T1w and T2w images with FreeSurfer's boundary based registration (BBR) method (Greve and Fischl, 2009); (iv) bias field correction using sqrt(T1w X T2w); and (v) nonlinear volume-based registration to the MNI template using FSL's FNIRT algorithm. A customized version of Freesurfer version 5.3's recon-all was used to generate white and pial cortical surfaces (using both T1w and T2w volumes at 0.7 mm resolution) plus a subcortical segmentation, all carried out in the subjects' native volume space. PostFreeSurfer converted the FreeSurfer data into standard NIFTI, GIFTI, and CIFTI file formats and also brought the data into MNI space.

The Multimodal Surface Matching (MSM) surface registration algorithm (Robinson et al., 2014) was used to perform an initial gentle non-rigid surface registration based on folding patterns (MSMSulc). This supplanted the FreeSurfer folding-based registration previously used (Glasser et al., 2013), because it achieved slightly better initial alignment of functionally corresponding regions (e.g., task fMRI) than Freesurfer's algorithm while inducing much lower local distortions (Robinson et al., 2014). This registration, together with the FNIRT nonlinear registration, was used to bring an initial version of the data into standard grayordinates space (32k standard mesh for each hemisphere's cortical surface at 2 mm average vertex spacing and 2 mm isotropic MNI-space voxels for the subcortical volume data). Myelin maps were computed using the ratio of T1w/T2w images and normalized for residual transmit field inhomogeneity (Glasser et al., 2014a; Glasser et al., 2013; Glasser and Van Essen, 2011; Robinson et al., 2014). Because gyral crowns tend to be thicker than sulcal fundi, the FreeSurfer-generated measure of cortical thickness was corrected for folding-related biases by regressing out the FreeSurfer mean curvature measure from each subjects' thickness data (Glasser and Van Essen, 2011).

1.6 Initial Preprocessing: Functional

The HCP functional preprocessing pipelines include volumetric (fMRIVolume) and surface-based (fMRISurface) components (Glasser et al., 2013; Smith et al., 2013a) applied to both rfMRI and tfMRI data. The fMRIVolume pipeline includes removing image distortions due to gradient nonlinearity and b0 inhomogeneity; motion correction; cross modal alignment to the T1w image with BBR (Greve and Fischl, 2009); concatenation of all transforms, including the nonlinear volume registration to MNI space; and resampling the original timeseries into MNI space using a single spline interpolation. Several intensity normalization steps occur, including a crude fMRI bias field correction based on the structural data from a separate imaging session (this bias field correction is replaced with a better one in later processing as described below, which has also been incorporated into the latest version of the pipelines), and grand 4D mean normalization to 10,000. The fMRISurface pipeline was then used to map grey matter timeseries data into the 91282-grayordinate standard space (2 mm average cortical vertex spacing and 2 mm subcortical voxels) using a 2 mm FWHM smoothing kernel (constrained to the cortical surface and subcortical grey matter segmentation). These steps produce a CIFTI "dense timeseries" file for each rfMRI and tfMRI run. The crude bias field correction map was also mapped into standard CIFTI space so that it could be replaced with the better one (by dividing it back out and multiplying by the new correction).

1.7 Initial Preprocessing: Resting State fMRI Temporal Denoising

For rfMRI runs, the ICA+FIX pipeline (Beckmann et al., 2005; Griffanti et al., 2014; Salimi-Khorshidi et al., 2014; Smith et al., 2013a) was used to remove spatially specific temporally structured artifacts. ICA+FIX includes several steps: 1) highpass temporal filtering with a sigma of 1,000 s (run length = 864 s) to remove linear trends in the data; 2) MELODIC independent component analysis (ICA) with auto dimensionality selection up to 250 components, producing component spatial maps and timeseries; 3) classification of these components into signal and noise categories by the FMRIB's ICA-based Xnoiseifier (FIX) trained ICA component classifier; 4) regression out of the data and all ICA components of 24 motion parameters (which were also temporal highpass filtered with sigma of 1,000 s). Regression coefficients were computed using all ICA component timeseries, and the noise component timeseries were then weighted by the regression coefficients and subtracted from the data (a "non-aggressive" regression approach). The ICA+FIX algorithms were run on the volumetric timeseries data, then the highpass filter and the nuisance regression steps were also applied to the grayordinates timeseries data. The ICA+FIX cleanup was reapplied to the rfMRI dense timeseries data after the original uncleaned native mesh data had been resampled into the standard grayordinates space according the areal-feature-based MSM surface registration (see below #2.5). Regression of the mean gray signal (aka 'global signal') was used in early analyses, but it was discontinued because it shifted some rfMRI functional connectivity (FC) gradient locations in ways that reduced cross-modal alignment. No additional spatial smoothing or temporal lowpass filtering was performed, as these types of "lossy" preprocessing steps would reduce the accuracy of the parcellations and proved unnecessary for the purposes of the current study.

1.8 Initial Preprocessing: Task fMRI Analysis

For task fMRI runs, the HCP's task analysis pipeline was run to generate grayordinate-based task analyses (Barch et al., 2013). This pipeline is an adaption of FSL's FEAT analysis pipeline (Smith et al., 2004) to CIFTI grayordinate space. A temporal highpass filter with a sigma of 100 s was run on the data to remove low frequency fluctuations presumably unrelated to the task design. These fluctuations likely include both temporal artifacts and the spontaneous intrinsic BOLD fluctuations that are the focus of rfMRI analyses. As with the rfMRI preprocessing, no additional spatial smoothing or temporal lowpass filtering was performed. FSL's FILM algorithm was used to compute the first level (single run) task fMRI statistics (including temporal autocorrelation smoothing constrained to the cortical surface and subcortical parcels), and FSL's FLAME algorithm was used in a fixed effects analysis to combine across runs within subjects. The tfMRI analysis pipeline was run after the tfMRI timeseries data had been resampled into the standard grayordinates space using areal-feature-based MSM surface registration (as explained in the MSMAll section below, see also Figure 10 and sections #3.2 and #6.4 below for a full description of the task fMRI processing used in this paper).

1.9 Initial Preprocessing: Artifact Map Generation

Three maps were generated for use as artifact indicators: a large vessel effect map, a gradient echo fMRI dropout map, and a curvature map. In addition to generating the myelin maps, the T1w/T2w ratio images were also used to estimate the effect of large blood vessels on the fMRI data. Blood vessels have a much higher T1w/T2w ratio than brain tissue and were segmented via thresholding normalized to the mean and standard deviation of the T1w/T2w image. This segmentation was regularized (dilated and eroded equal amounts, and above threshold values not near the boundary of the brain parenchyma were removed), slightly dilated, and smoothed using a 2 mm FWHM volume smoothing kernel to account for proximity effects of blood vessels. The gradient echo fMRI dropout map was created by identifying voxels having a ratio of less than 50% signal intensity in the single band reference (SBRef) gradient echo image relative to the spin echo image after using a spatial highpass filter of sigma=5 mm (smooth, then subtract from the unsmoothed data) to reduce residual low spatial frequency variation in the ratio image (related to the differences in the transmit field between the two images). This dropout segmentation was dilated one 0.7 mm voxel and smoothed at 2 mm FWHM in the volume. After removing transmit field effects and excluding dropout regions, the low spatial frequency intensity variations (sigma=5 mm) within grey matter were used to compute a more accurate, smoother receive bias field for the fMRI data. This field was also scaled to a volume mean of 10,000, just like the fMRI timeseries data were scaled by the HCP minimal preprocessing pipelines. The scaled field was used as a reference BOLD intensity image when computing bias free beta effect size maps (after having reverted the previous bias field correction). The vascular effect maps, dropout maps, and the reference images were mapped into the standard CIFTI grayordinates space. The curvature maps were taken directly from FreeSurfer and mapped into the standard CIFTI grayordinates surface space.

2. Multi-modal Areal-feature-based Surface Registration ("MSMAll")

Achieving good intersubject cortical alignment was critical to our group average gradient-based parcellation efforts. This entailed major refinements to the recently published MSM approach (Robinson et al., 2014). The following sections describe (i) the motivation for areal feature-based cortical surface registration; (ii) a novel method of weighted regression to generate more individualized and refined RSN maps in individual subjects by upweighting the most aligned regions; (iii) details on the implementation of the MSMAll method; (iv) "de-drifting" used to minimize registration-induced biases in group-average datasets; and (v) the final one-step resampling of the data from the subjects' native meshes to their standard 32k meshes. Our overall approach to surface registration is illustrated in Figure 2.

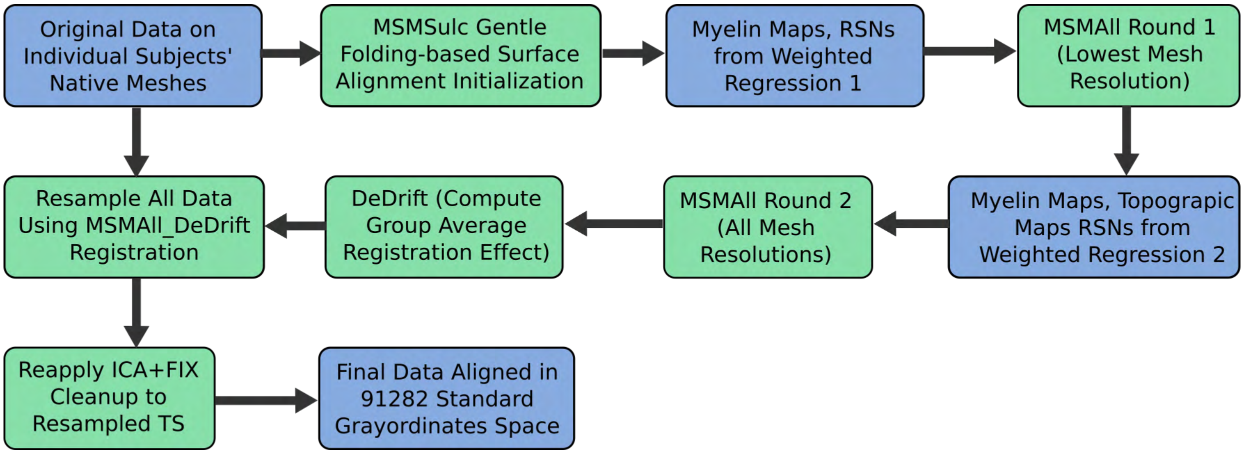
2.1 Motivation

The boundaries of cortical areas are correlated with cortical folding patterns, but the strength of this correlation is variable across the cerebral cortex. Early areas such as V1 in the calcarine sulcus and areas 4, 3a, and 3b in the central sulcus have particularly high correlations with folding patterns (Fischl et al., 2008). Association areas implicated in 'higher' levels of processing, e.g. areas 44 and 45 in the inferior frontal gyrus, often have substantially lower correlations with folding patterns (Fischl et al., 2008). Moreover, the folding patterns themselves do not necessarily have one-to-one correspondence across subjects in these higher association regions (Van Essen, 2005; Van Essen et al., 2012). To improve the alignment of cortical areas across subjects and thus the validity of our group average analyses, we used MSM (Robinson et al., 2014) to perform registration based on areal features. Registration using myelin and Resting State Network (RSN) maps can improve the cross-subject alignment of independent tfMRI datasets (Robinson et al., 2014), while at the same time decreasing the cross-subject alignment of cortical folding patterns that do not correlate with cortical areal locations.

2.2 General approach to MSMAll

The MSMSulc folding-based surface registration from the HCP minimal preprocessing pipelines (in PostFreeSurfer, see above #1.5) was used to initialize the areal feature-based registration (MSMAll). MSMAll used myelin maps, RSN maps, and rfMRI visuotopic maps (see below #4.4) in a joint multi-modal registration to drive cross-subject alignment. Because there is no temporal correspondence across subjects in the rfMRI timeseries, the only corresponding rfMRI subject-wise dataset (i.e. without reference to group data) is the rfMRI FC "dense connectome" (91,282 X 91,282 X 4 bytes = 32.5GB). This dense connectome is not suitable for driving the registration process, however, because of its sheer size and the fact that it contains an enormous amount of redundant information plus unstructured noise, neither of which are useful for registration[1]. What is needed instead is a dimensionality reduction of the dense connectome across individual subjects in which the data remains matched across subjects. We used Group ICA along with spatial multiple regression to produce RSN template maps and their corresponding individual subject maps for this purpose.

To generate the RSN template maps needed to initiate the process, melodic group ICA (Beckmann et al., 2005; Beckmann and Smith, 2004; Smith et al., 2014) was performed in CIFTI grayordinates space (Smith et al., 2013a; Smith et al., 2013b) using resting state data from the 210P group that had previously been registered with MSMRSN (Robinson et al., 2014) across a range of ICA dimensionalities (d = 22 to 100, sparsely sampled, and then searched using binary search to identify the optimal dimensionality). Thus the RSN templates that served as the input to this process were generated from the data presented in (Robinson et al., 2014) after removing the group average registration drift (see below #2.5 and (Abdollahi et al., 2014). We used two metrics to estimate the optimal group ICA dimensionality in a 28-subject "registration optimization" subset of the 210P+29T subject groups. (i) An analysis of tfMRI z-statistic cluster mass (thresholded at +/-z=2.32) using a group mixed effects analysis (Robinson et al., 2014) yielded a broad peak of similar tfMRI cluster mass values over the range of 33-42 components. (ii) An analysis of average cortical surface vertex-associated areal distortion showed a monotonic decrease in surface areal distortion (averaged across the hemisphere) when registering using higher numbers of components. Consequently, a group ICA dimensionality of 41 was used for the MSMAll registration, as this dimensionality was the local maximum in cluster mass that had the least areal distortion.[2] This included three artifactual components and four predominantly subcortical components, which were excluded from the cortical surface registration but were included in the multiple regression model that was used to generate the individual subject RSN component maps (see below #2.4).



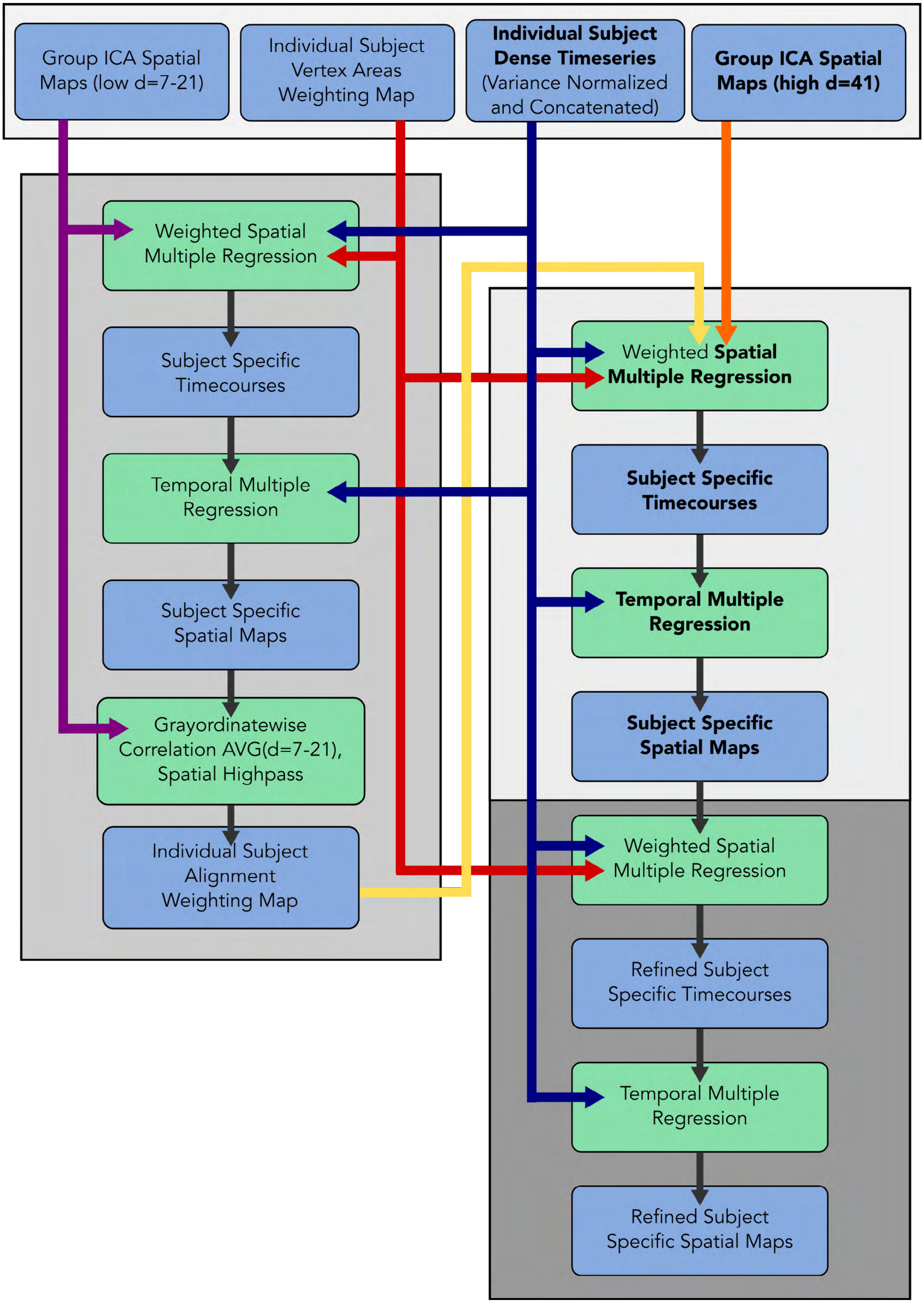
---

[1] Though see below (#3.3-3.4) for a method that downweights the effects of unstructured noise in dense connectomes without spatial or temporal smoothing

[2] In the ConnectomeDB version of the MSMAll Pipeline that was used for the 900 subject HCP data release, this process was reapplied using the registration templates generated from this study's data, and a dimensionality of 40 was found to be optimal.

Figure 2, above, shows the overall approach to surface registration used in this study. First, MSMSulc was used to perform an initial gentle non-rigid folding-based registration of the native mesh data. Then myelin maps and resting state network maps (using weighted regression, see below #2.3) were generated. The first round of MSMAll used these maps in a joint multi-modal registration for only the lowest mesh resolution in order to get both hemispheres roughly aligned based on areal features. Then myelin maps, resting state network maps (using weighted regression based on the rough alignment) and topographic maps were generated based on this alignment. The second round of MSMAll uses these maps in a joint multi-modal registration for all three mesh resolutions (see below #2.4) to get each hemisphere finely aligned by areal features. Registration drift is computed by averaging all of the individual subject registrations relative to MSMSulc—any group average registration effect is considered drift. The combined registration, MSMAll_DeDrift, is used to resample all datasets from the native mesh into the standard grayordinates space with a single interpolation step, and then the ICA+FIX cleanup is reapplied to the resampled dense timeseries (TS).

Figure 3, below, shows the steps of the Weighted Regression method used in this study and compares it to the standard dual regression steps (shown in **BOLD**). In the weighted regression method, the vertex area weighting maps are always used to weight the spatial multiple regressions. An alignment weighting map is derived from low dimensional ICA maps (d=7-21) by correlating the individual subject maps with the group maps, averaging across dimensionalities, and spatial highpass filtering (medium grey box). The first round of weighted regression uses both the vertex area map and the alignment map together to weight the spatial multiple regression of the high dimensional group ICA spatial maps to produce initial subject specific spatial maps (lightest grey box). The second round uses only the vertex area map to weight the first round initial individual subject spatial maps to produce the refined individual subject spatial maps (darkest grey box). These spatial maps can then be used for registration or areal classification. Note that to produce the visuotopic maps (not shown in diagram), a single round of weighted regression equivalent to the second round is used (where the subject specific V1-only spatial regressors take the place of the subject specific ICA spatial maps from the first round).



Group ICA Spatial Maps (low d=7-21)
Individual Subject Vertex Areas Weighting Map
Individual Subject Dense Timeseries (Variance Normalized and Concatenated)
Group ICA Spatial Maps (high d=41)
Weighted Spatial Multiple Regression
Subject Specific Timecourses
Temporal Multiple Regression
Subject Specific Spatial Maps
Grayordinatewise Correlation AVG(d=7-21), Spatial Highpass
Individual Subject Alignment Weighting Map
Weighted Spatial Multiple Regression
Subject Specific Timecourses
Temporal Multiple Regression
Subject Specific Spatial Maps
Weighted Spatial Multiple Regression
Refined Subject Specific Timecourses
Temporal Multiple Regression
Refined Subject Specific Spatial Maps

2.3 Weighted Regression for Generating Refined Individual Subject RSN Maps from Group Maps

To generate refined individual subject RSN component maps for registration and areal classification, we implemented a Weighted spatial multiple Regression (WR) technique, see Figure 3. The technique is adapted from standard dual regression (DR) (Filippini et al., 2009) which seeks to use spatial overlap between the group and individual to find individualized versions of the group maps and has two main stages (see **BOLD** portions of Figure 3): (i) subject-specific component timecourses are generated by spatial regression of the group component spatial maps onto the individual dense timeseries, equation (2) below, then (ii) subject-specific component spatial maps are generated by a temporal regression of these timecourses onto the individual dense timeseries, equation (4). Mathematically, we can consider the variables: Y = data (grayordinates X timepoints), X = group model (grayordinates X maps), T = parameter estimates (maps X timepoints), and x = individual maps (grayordinates X maps). To find the subject specific component timeseries, one can use the following equations:

$$Y = XT \tag{1}$$

$$T=\mathrm{pinv}(X)Y \tag{2}$$

Where 'pinv' is the pseudoinverse. To find the subject specific component spatial maps, one can use the following equations:

$$Y' = T'\, x' \tag{3}$$

$$x' = \mathrm{pinv}(T')Y' \tag{4}$$

Where the prime (') represents the transpose.

However, we improved this procedure substantially by incorporating a spatial weighting map in the spatial regression and by running a second round of spatial and temporal regression based on the results of the first round (i.e., 2 spatial and 2 temporal regressions total). The individual subject dense timeseries used for the weighted regression were first demeaned (removing the mean of each grayordinate's timeseries) and variance normalized to equalize unstructured noise variance (see below #3.3 for the method of computing the unstructured noise variance map), and then concatenated across all four runs (4800 timepoints/subject).

The spatial weighting map used in WR is the elementwise product of two separate maps: 1) a "Vertex Area Map" map, and 2) an "Alignment Map." The vertex area map represents the vertex areas of the individual subject's midthickness surface resampled onto the standard 32k mesh in native volume space. It compensates for three sources of distortion in surface vertex area between the native midthickness surface in the subject's physical space and the 32k standard MNI space midthickness surface. These arise from (i) the non-rigid volume registration to MNI space, (ii) projection of the anatomical surface to a sphere, and (iii) registration of the individual spherical surface to the atlas spherical surface. Correcting for these distortions is important because vertex areas (which are homogeneous on standard spherical surfaces) vary over a two-fold range (~3±1 mm^2) when measured on each subject's 32k standard mesh cortical midthickness.

The "Alignment Map" map upweights regions likely to be already well aligned to the group in the individual subject and downweights misaligned regions. This is important because misalignments contribute adversely to the individual subject component timeseries and spatial maps that are generated by DR, insofar as they cause mixing among different resting state networks (and cortical areas). This can cause the resultant DR individual subject spatial maps to be less "individualized" than the true individual subject spatial maps (i.e. more like the group maps than they should be). The problem gets worse as the dimensionality of the ICA being used in the DR increases because there is less information available to estimate each unique component timeseries and spatial map.

To generate the Alignment Map (Figure 3, left column), we used weighted dual regressions (weighted only by the Vertex Area Map) of low ICA dimensionalities. Grayordinates that are already well aligned between the individual and group will, by definition, have individual subject component spatial map values that are strongly correlated to the group map values, whereas misaligned regions will have weaker correlations. Hence, to measure alignment quality we correlated the values of the group component maps to the values of the individual subject component maps, generating a spatial map of Fisher-transformed group versus individual component correlations for each low ICA dimensionality. Dimensionalities between 7 to 21 components yielded the most consistent results when tested in a pilot analysis, so we averaged the correlation maps across these dimensionalities. Undesirable low spatial frequency SNR differences were removed by applying a spatial highpass filter with sigma=14mm.

After generation of the Alignment Map, the WR algorithm includes two rounds of weighted regression, as follows (Figure 3, right column): Round 1a) individual subject timecourses for each component were generated by *spatial* multiple regression of the group component maps into the individual subject dense timeseries, weighted according to the element-wise product of the Vertex Area Map and the Alignment Map. Round 1b) these individual subject component timecourses were then *temporally* multiple regressed into the individual subject dense timeseries to produce the initial individual subject component spatial maps. Round 2a) these inital individual subject component spatial maps were *spatially* multiple regressed into the individual subject dense timeseries, weighted this time only by the Vertex Area Map, to produce a refined set individual subject component timecourses. Round 2b) these refined individual subject component timecourses were *temporally* multiple regressed into the individual subject dense timeseries to generate a refined set of individual subject component spatial maps. The weighted form of the spatial regression looks like equation (5). (W=spatial weights, grayordinates X 1, .* is the elementwise product repeated to match the dimensions of X and Y, the square root is elementwise).

$$T=\mathrm{pinv}(X.*\sqrt{W})(Y.*\sqrt{W}) \tag{5}$$

This second round of regressions helped to refine the resultant spatial maps so that they more closely reflect the individual spatial variation of the subject's resting state data, and allowed the entire individual subject's dataset (i.e. without the Alignment Map weighting) to contribute to the spatial regression when producing the final refined individual subject component spatial maps. The resulting individual subject component spatial maps were then used in the MSMAll registration or areal classification. In practice we found that this WR approach substantially improved the ability of our classifier to delineate and identify misaligned cortical areas (see #6.3). It also had the benefit of increasing the optimal MSM ICA dimensionality (41 vs the original 26) and requiring fewer registration iterations to reach convergence (2 iterations of MSM and WR vs the original 6 iterations of MSM and DR (Robinson et al., 2014)).

2.4 Implementation of Multi-modal Areal Feature-based Surface Registration Using MSM

Of the 41 component maps, four were excluded because they were predominantly subcortical, and an additional three were excluded because they were dominated by obvious spatial artifacts, leaving 34 well-defined cortical surface RSN component maps. The following 44 maps were jointly used in the final iteration of MSMAll: 34 RSNs, the subject's myelin map, eight V1-based rfMRI visotopic regressor maps (see below #4.4), and a binary non-cortical medial wall ROI (i.e. the region of the surface mesh that does not contain neocortical grey matter). Group versions of all these maps served as the multi-modal registration target. tfMRI contrast maps were not used because (i) it was useful to preserve an independent dataset with which to evaluate whether the registration was objectively improving areal feature alignment (e.g. to check the aforementioned optimal RSN dimensionality for registration) and (ii) the individual subject task contrast maps have less useful contrast to noise (CNR) per map than the individual subject maps of the other modalities as determined by the classifier (see Supplementary Figure 12 in the SRD)[3].

For surface-based registration of spherical surfaces, we used a modified version of the MSM algorithm reported in (Robinson et al., 2014). This achieves smoother deformations by penalizing the warps of each face of the spherical mesh (rather than pairs of points). As the method uses discrete optimization, it utilizes a method for incorporating higher order terms, called Higher-order Clique reduction (HOCR) (Glocker et al., 2010; Ishikawa, 2009, 2014). An exponent of 3 was also used on the regularizer function to nonlinearly penalize larger distortions. We found that this approach achieved comparable alignment of areal features while significantly reducing distortions of the surface mesh. Registration was performed in a multi-resolution setting in which registration used a series of low-resolution control point meshes. These were placed over the original mesh and constrained the deformations so that coarse (low frequency) features of the data are aligned first, and iteratively improves alignment until the fine details are brought into correspondence. We used three resolution levels with control point spacings of 30.0 mm, 15.1 mm, and 7.6 mm. Correlation was used as a similarity metric to drive alignment of the myelin, resting state network maps, and resting state visuotopic maps in a joint multi-modal registration.

---

[3] Thus, if a substantial amount of rfMRI data were not available, tfMRI data would likely be better treated as rfMRI data in MSMAll registration or areal classification, because with a similar number of timepoints tfMRI data closely resembles rfMRI data (Cole, M.W., Bassett, D.S., Power, J.D., Braver, T.S., Petersen, S.E., 2014. Intrinsic and task-evoked network architectures of the human brain. Neuron 83, 238-251.).

Two iterations of MSM and weighted regression were used, with the first iteration performing only the lowest control point resolution level to achieve a coarse alignment. The V1-based rfMRI visuotopic regressors (see below #4.4) were included only in the second iteration to align the fine visuotopic patterns. (V1, where the visuotopic pattern is defined, is highly correlated with folding (Fischl et al., 2008) and should be well aligned across subjects at this stage of the registration.) Each map was normalized to have a spatial mean of zero and a spatial standard deviation of 1. Other registration parameters at each resolution level included regularization multiplier lambdas of 0.01, 0.05, 0.1; number of internal MSM iterations of 10, 10, 5; and internal smoothing sigmas of 2, 2, 1 mm. Sampling meshes were 7.6 mm spacing, 3.8 mm spacing, and 1.9 mm spacing. Data meshes were 3.8 mm spacing and 1.9 mm spacing. The final output of the registration was a deformed native mesh sphere—"registered sphere" for each individual hemisphere. This sphere contains the same number of vertices as the original individual subject native sphere produced by FreeSurfer, but its vertices have been moved across the surface nonlinearly to maximize alignment with the multi-modal registration target, so that like areal features are aligned across subjects.

2.5 DeDrifting and Final Resampling

The final registration stage involves removal of the group average registration "drift", which is an undesirable byproduct of repeated registration during a template generation process (Abdollahi et al., 2014). Unless corrected, this drift can cause registration template features to drift away from the typical subject in the group. As a particularly sobering example, while volumetric non-rigid registration to MNI space greatly reduces the variability in individual brain volumes (desirable), it also induces a large expansional drift in brain volume (undesirable) so that the group average brain volume is 37% larger than the typical subject's native space brain volume (Van Essen et al., 2012). For surface registration, the problem is not one of anatomical expansion (because the spherical coordinates used in registration are not tied to anatomical coordinates), but of local drifts in cortical areal size, shape, and position that differ across different registration templates and hence can lead to inaccuracies in cross-study spatial comparisons.

Surface-based atlases are fundamentally linked to standard spheres in which large and consistent brain landmarks (i.e. major cross-subject consistent sulci and gyri) have particular latitudes and longitudes in spherical mesh coordinates (Drury et al., 1998; Fischl et al., 1999a; Fischl et al., 1999b; Van Essen, 2005). We used a coordinate system based on the FreeSurfer 'fsaverage' template (Desikan et al., 2006) after registration to a standard mesh sphere that accurately aligns geographically corresponding features in the left and right hemispheres (Van Essen et al., 2012). The high-resolution version of this mesh has 163,842 vertices and is referred to as the '164k_fs_LR' mesh. For most analyses we used a lower-resolution mesh (32,492 vertices; the '32k_fs_LR' mesh, with a 2 mm average vertex spacing on the midthickness surface) matched to the 2 mm volume resolution of the HCP fMRI data (Glasser et al., 2013). Although there is inherently some arbitrariness in how this geographic coordinate system was initialized, it is important to avoid drifts away from this geographic convention to enable the most precise cross-study spatial localization, even when differing areal feature-based registration modalities and templates are used. We used the following approach to remove drifts from our MSMAll registered data. Spheres whose coordinate locations represent the MSMAll registration warp relative to the gentle MSMSulc folding-based surface alignment were resampled to the 32k_fs_LR standard mesh for each hemisphere and averaged across subjects. These average registered spheres represent the group average warp from the MSMAll registration across all 449 subjects. The inverse of this average spherical warp was applied to each individual's registered sphere, eliminating the group-average drift while maintaining the subject-specific alignment improvements. Each subjects' registration features were then resampled to the atlas sphere and averaged, creating the dedrifted registration template for use by other studies. Each subject's final registered spheres (one for each hemisphere, on the native mesh) represent the MSMAll_DeDrift surface registration (referred to as the 'MSMAll' registration in this study's other documents and in the HCP 900 subject public release).

Using the MSMAll_DeDrift registration, all data used for the parcellation analysis (myelin maps, thickness maps, artifact localization maps, resting state timeseries, and task fMRI timeseries) from all subjects were resampled from the subjects' original native meshes into the 91,282 standard grayordinates space using a single step of adaptive barycentric resampling (Glasser et al., 2013). This avoids unnecessary blurring that can occur with multiple resampling steps. The adaptive barycentric resampling algorithm insures that all native mesh data contribute when downsampling to a lower mesh resolution (~136k native meshes to 32k standard meshes). Additionally, all individual subject surfaces (e.g., midthickness, white, pial) were resampled using regular barycentric interpolation, and Connectome Workbench specification ('spec') files were generated for the MSMAll_DeDrift resampled datasets. These resampled datasets were used for all further analyses, including the following remaining individual subject preprocessing and analysis steps: (i) the task analysis pipeline described above (#1.8) was run; (ii) reapplication of the ICA+FIX resting state structured noise removal on the MSMAll_DeDrift registered and resampled CIFTI dense timeseries; and (iii) recalculation of the myelin map normalization (see above #1.5), which relies on finding very low spatial frequency differences between the individual and group myelin maps using surface registration (Glasser et al., 2013).

3. Creation of Group Average Multi-modal Datasets for Parcellation

3.1 Group Average Structural

Structural data from the 210-subject parcellation (210P) and validation (210V) groups was processed separately as follows. MSMAll_DeDrift-resampled unsmoothed myelin maps plus thickness maps corrected for curvature were averaged vertex-wise across subjects while excluding outliers greater than 3 standard deviations above or below the group mean. Though not used in the semi-automated multi-modal parcellation, average maps of folding patterns (FreeSurfer sulc and curv) were also produced from individual subject maps resampled according to MSMAll_DeDrift. Group averages of the individual MSMAll_DeDrift-resampled surfaces (i.e., averages of coordinate positions for corresponding vertices) were used to display group results.

For some analyses (e.g. gradient computation, see below #4.2-4.3), it is important to have reliable group average measures of the surface area associated with each vertex derived from the individual midthickness surfaces (via the tiles associated with each vertex, see also above #2.3). Vertex areas computed directly from group average surfaces are inherently inaccurate owing to the blurring effects caused by averaging folding patterns that differ across individuals (see Supplementary Figure 1 in the SRD). This issue arises for folding-based registration because there is not 1 to 1 correspondence of many folds across subjects, and it is exacerbated when using alignment based on areal features, because areal features are imperfectly related to folding patterns across much of the cerebral cortex. Group average vertex area maps were computed by averaging the vertex area maps of individual subjects' native volume space MSMAll_DeDrift-resampled midthickness surfaces, and were used in group computations that required accurate vertex areas such as gradients (see below #4.2-4.3).

3.2 Group Average Task fMRI

For the 86 task contrasts from the seven HCP tasks (47 unique, 39 sign reversed), cross-subject mixed effects analyses were computed using FSL's FLAME algorithm (FLAME1) applied to the individual subject cross-run results. The excellent intersubject alignment provided by MSMAll and a large number of subjects yields high z-statistic values without the need for conventional spatial smoothing. The cross-subject alignment of areal features and lack of smoothing preserves fine spatial details in the group-average maps (see Supplementary Figure 2 in the SRD) and produces high reproducibility across all non-outlier contrasts (see Supplementary Figure 3 in the SRD). For the parcellation and areal classifier we used the effect size maps (betas, after reverting the minimal preprocessing pipelines' bias field correction and applying the improved correction by dividing the beta map by the aforementioned BOLD reference image). We used effect size maps because they are a measure of the task induced effect itself, rather than the significance/precision of the measured task induced effect. Thus, the effect size maps area more analogous to the other maps used for parcellation (e.g., myelin maps).

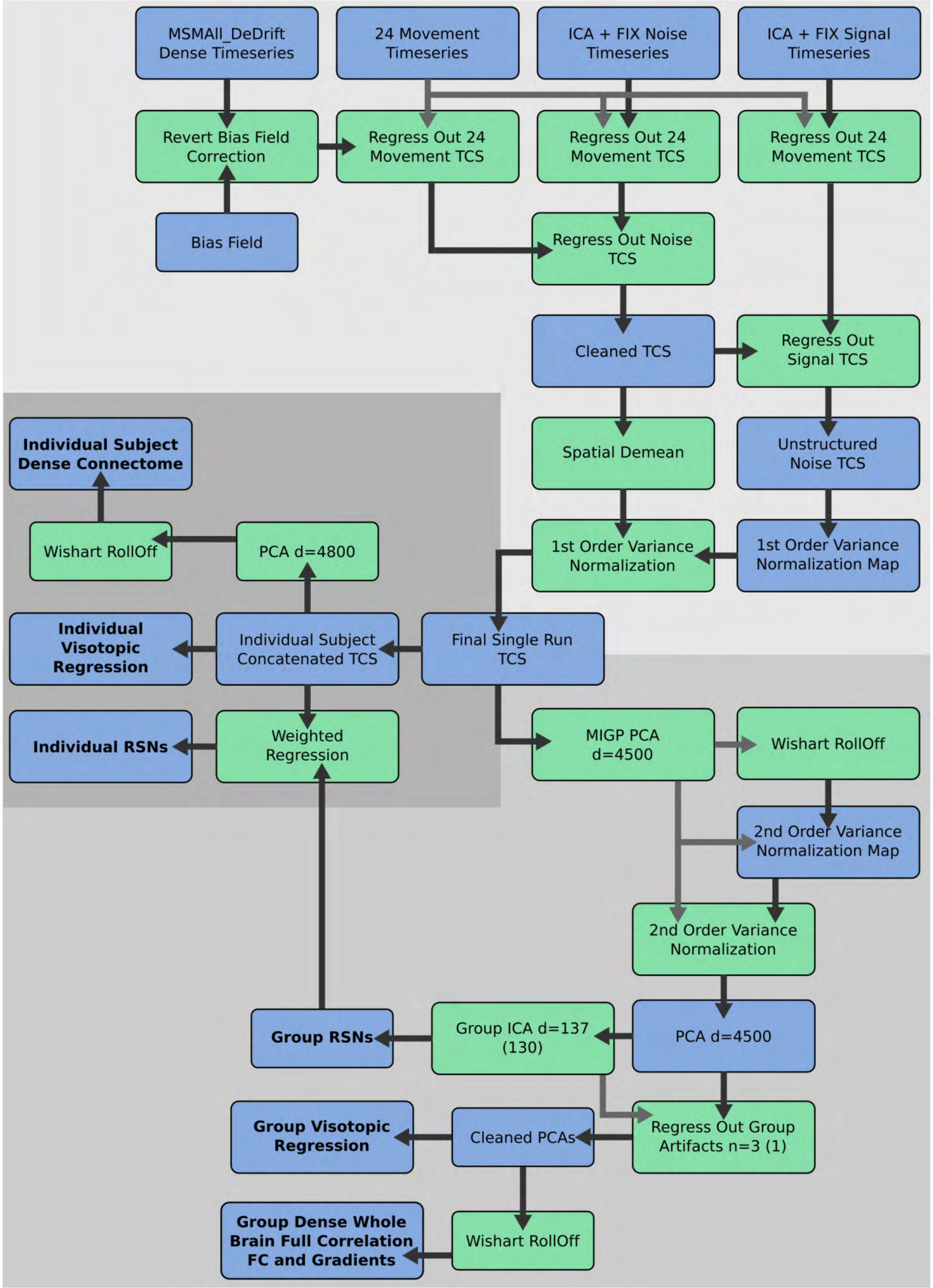

MSMAll_DeDrift Dense Timeseries
24 Movement Timeseries
ICA + FIX Noise Timeseries
ICA + FIX Signal Timeseries
Revert Bias Field Correction
Regress Out 24 Movement TCS
Regress Out 24 Movement TCS
Regress Out 24 Movement TCS
Bias Field
Regress Out Noise TCS
Cleaned TCS
Regress Out Signal TCS
Individual Subject Dense Connectome
Spatial Demean
Unstructured Noise TCS
Wishart RollOff
PCA d=4800
1st Order Variance Normalization
1st Order Variance Normalization Map
Individual Visotopic Regression
Individual Subject Concatenated TCS
Final Single Run TCS
Individual RSNs
Weighted Regression
MIGP PCA d=4500
Wishart RollOff
2nd Order Variance Normalization Map
2nd Order Variance Normalization
Group RSNs
Group ICA d=137 (130)
PCA d=4500
Group Visotopic Regression
Cleaned PCAs
Regress Out Group Artifacts n=3 (1)
Group Dense Whole Brain Full Correlation FC and Gradients
Wishart RollOff

Figure 4, above, shows the resting state processing and analysis stages used in this study separated into 3 main groups. The inputs include: the bias field, the MSMAll_DeDrift resampled dense timeseries (TCS), the 24-parameter motion timeseries, the ICA+FIX noise component timeseries, and the ICA+FIX signal component timeseries (all timeseries have been temporally filtered with a sigma of 1,000 s). The most important outputs are identified by **BOLD** font. Datasets are in light blue boxes and processes are in light green boxes. The first group (lightest gray background) includes reverting the bias field correction, repeating the ICA+FIX cleanup stages (since the MSMAll_DeDrift dense timeseries was resampled from native mesh timeseries that had not yet been cleaned), subtracting the mean of each grayordinate's timecourse, and computation of the first order variance normalization map from the square root of the variance of the unstructured noise timeseries (all cleanup done and all RSNs regressed out) and then division of the dense timeseries by this map, producing the final single run dense timeseries. At this point, the second group (darkest gray background) breaks off to continue individual subject processing. The subject's resting state runs are concatenated, and these data were used to produce the individual subject visuotopic regression maps, the individual subject RSN maps (from the group RSN maps and weighted regression), and, with PCA and Wishart RollOff, a high CNR individual subject dense functional connectome. Also breaking off at this point is the third group (medium gray background), which enables computationally efficient processing of the 210P and 210V group average rfMRI datasets. This module starts with the final single run dense timeseries and uses the MIGP PCA algorithm (d=4500) to combine and dimensionality reduce the resting state data across subjects. The output PCA Series of this algorithm has residual spatial variation unstructured noise variance and requires second order variance normalization (computed from the square root of the difference of the variance before and after a Wishart RollOff, see below). Then PCA is run again without dimensionality reduction to incorporate the spatial reweighting (d=4500). Group ICA (d=137 for 210P; d=130 for 210V) is run on the output PCA series to identify group level RSNs and artifacts (3 for 210P and 1 for 210V). The artifacts are regressed out of the data, and the result can be used for the group visuotopic regression maps. Finally a second Wishart RollOff is performed prior to group dense functional connectivity matrix generation (whole brain full correlation FC and gradients) to improve the CNR of the functional connectivity maps and to eliminate ringing artifacts.

3.3 Group Average Resting State fMRI

A variety of novel and advanced methods were used to maximize the information gained from the resting state fMRI data, while minimizing computational requirements and reducing the impact of structured and unstructured noise. Figure 4 shows an overview of the resting state processing used in this study after resampling from the native mesh dense timeseries into standard grayordinates space according to the MSMAll_DeDrift surface registration. First, the bias field correction is reverted (see below #4.4), the ICA+FIX temporal cleanup is reapplied, and the variance normalization described below is applied to make the cleaned, variance normalized individual run dense rfMRI timeseries (lightest grey box in Figure 4).

In rfMRI studies involving modest numbers of subjects, group analyses are typically carried out by temporally concatenating the individual subject dense timeseries data after subtracting the mean of each grayordinate's timecourse (and perhaps after performing variance normalization). In the present study, a concatenated dense timeseries would be prohibitively large (~342GB, based on 210 subjects X 4800 timepoints = 1,008,000 total timepoints X 91282 grayordinates X 4 floating point bytes). For a computationally tractable group rfMRI dense "timeseries" dataset that closely approximates a full concatenation, we applied the parallelizable and distributable Melodic's Incremental Group PCA (MIGP) algorithm (Smith et al., 2014) to generate a 4500 component group dense PCA series. The dimensionality reduction from 1,008,000 to 4500 removes a large amount of unstructured noise variance while retaining the strongly structured resting state BOLD signals that are consistent across subjects. The dense PCA series from MIGP, while being much more compact than the full concatenated dense timeseries, is analogous to a timeseries in many respects; however, we will invoke the term PCASeries when a PCA series is being used in an algorithm, reserving the term timeseries for genuine timeseries data.

We found that MIGP performance was improved (i) by reverting the fMRI bias field correction incorporated in the version of the HCP minimal preprocessing pipelines used in the 500 subject HCP data release and (ii) by applying variance normalization based on the amount of unstructured temporal noise in the dense timeseries of each rfMRI run. This normalization ensures that each grayordinate in each run of each subject has a comparable amount of unstructured noise variance prior to MIGP PCA (Beckmann and Smith, 2004). The result is an equal likelihood of detecting false positives across the grayordinates space and improvement in the performance of subsequent algorithms (e.g. ICA, multiple regression). The unstructured noise variance normalization map was computed as the square root of variance remaining after performing all ICA+FIX cleanup steps (high pass filter, 24 movement parameter regression, and structured noise component timeseries regression) and additionally after regressing out the structured RSN signals identified by ICA+FIX (see Figure 4, top & right group).

The large dimensionality reduction by the MIGP algorithm from 1,008,000 to 4500 PCA components results in spatial differences in the residual unstructured noise in the data and can yield undesirable ringing artifacts when visualizing seed-based functional connectivity (http://www.humanconnectome.org/documentation/mound-and-moat-effect.html). A multi-step process was used to avoid this problem (Figure 4, right & bottom module). Emulating an approach introduced by Beckman and Smith (2004), we fit a Wishart eigenspectrum distribution function (Beckmann and Smith, 2004; Wishart, 1928) to the tail of the eigenspectrum of the dense PCA series (whose eigenvectors represent mostly unstructured noise and which have the lowest eigenvalues) to estimate the profile of unstructured noise in the data in PCA space. This distribution was subtracted from the eigenvalues, thereby downweighting ('rolling off') the 'temporal' effects of unstructured noise in the data. The difference in the variances between the original PCA Series and the Wishart rolled off PCA Series represents the residual pattern of spatial variation in unstructured noise in the data. The square root of this map was used as a second order variance normalization map. The spatial variation in this map's values were inversely related (i) to the vertex areas of the surface grayordinates (smaller vertex areas had larger amounts of residual unstructured noise; larger vertex areas had lower amounts) and (ii) to the CNR of the subcortical voxels (lower CNR voxels like those in the iron rich globus pallidus had higher values relative to other voxels). After second order variance normalization, PCA was run again without dimensionality reduction (i.e. PCA d=4500) to incorporate the changes in spatial weighting across the grayordinates space. The resulting PCA series was now appropriate for group ICA.

Group ICA was performed on the 210P data to identify cross-subject consistent RSNs as well as artifactual components that slipped through the ICA+FIX cleanup. An ICA dimensionality of d=137 was chosen because when a find-the-biggest operation across components was performed, it described the data using the smallest number of spatially contiguous pieces (whose size exceeded a surface area of 25 mm^2 or a volume of 125 mm^3) relative to a wide range of dimensionalities (d=95 to 150 in unit steps and 150 to 300 in

5's). Also it was the largest local minimum before the number of subcortical pieces began to increase monotonically (around d=140), suggesting that further splitting was due to the ICA beginning to incorporate more unstructured noise into the subcortical components whose voxels have the lowest CNR and therefore are most vulnerable. Visually, the d=137 decomposition exhibited a striking degree of left/right symmetry, with RSN components either being bilaterally symmetrical or having a symmetrical pair of unilateral components (e.g. see the RSNs shown in the top four rows of Supplementary Figure 4 in the SRD). This strong bilateral symmetry may explain why d=137 was a local minimum in the spatially contiguous pieces measure. Three consistent artifactual components were identified in the d=137 decomposition of the 210P dense PCA series: the tranverse venous sinus; a fronto-polar, orbitofrontal, and anterior temporal component likely attributable to motion and/or motion's interaction with susceptibility artifacts; and an anterior and middle cerebral artery pulsation artifact. These artifactual components were regressed out of the dense PCA Series to produce a cleaned dense PCA Series appropriate for the group visuotopic regression (see below #4.4) and further processing. To produce the cleaned dense connectome for the 210V validation group in Supplementary Figures 4 and 5 in the SRD, group ICA was also performed separately on the 210V data, identifying a d=130 local minimum and only one artifactual component (the tranverse venous sinus).

Although the second order variance normalization reduces spatial inhomogeneities in unstructured noise variance in the MIGP dense PCASeries, it does not address the temporal effects of this noise on dense functional connectivity matrices (CNR reduction and ringing). To address these temporal issues, a Wishart function was fit to the second order variance normalized and cleaned PCA series eigenspectrum. For this stage, the desired output was the Wishart rolled off PCA series and dense connectome (the eigenvalues were reduced by the fitted Wishart null-eigenvalue distribution, remultiplied by the eigenvectors, and the dense connectome formed via the outer-product of the weighted eigenvectors). The Wishart roll off has several benefits: 1) It eliminates ringing induced by the hard cutoff of the PCA series at 4500 components and 2) it addresses the same goal of CNR improvement by unstructured noise reduction as do spatial and temporal smoothing. However, instead of smoothing all signals equally (whether they are of interest or not), the Wishart roll off procedure downweights less structured signals while preserving the sharpness of the structured resting state signals of interest. Highly structured RSN signals (represented by PCA components with high eigenvalues above the Wishart noise eigenvalue distribution) are not downweighted, whereas less structured signals (represented by PCA components with low eigenvalues along the Wishart noise eigenvalue distribution) are downweighted. Thus, the Wishart roll off acts as an edge-preserving spatial and temporal filter.

Empirically, the Wishart roll off improved the CNR of grayordinatewise rfMRI FC dense connectivity matrices, which otherwise have low CNR because there is no averaging across space, as in a parcellated connectivity analysis, an ICA analysis, or a multiple regression analysis, without blurring the signals of interest. Grayordinatewise rfMRI FC was performed using Pearson correlation (when computed on the PCA Series, the data are not spatially demeaned), generating a dense functional connectivity matrix. Also, the Fisher Z transform was not used because it shifted connectivity gradient peaks slightly. The decision not to use the Fisher Z transform was empirically based (not using it resulted in better alignment of resting state gradients with those of other modalities) and there is not a principled reason for doing it one way or the other. In practice, the difference in gradient peak location was relatively small, however. Despite being processed completely separately through the process shown in Figure 4, the 210P and 210V dense functional connectivity datasets show remarkable reproducibility (see Supplementary Figures 4 and 5 in the SRD). Thus, the approach used here for computationally efficient processing of group resting state data objectively shows good performance. An additional benefit of this approach is that the use of a Wishart rolloff reduces the non-specific "seed bloom" of high correlation near the seeded grayordinate, which arises from spatial autocorrelation in the noise induced by various image reconstruction (e.g. multi-band) and preprocessing steps (e.g. resampling or smoothing). The remaining local connectivity around the seed tends to follow the same areal feature boundaries as more distant functional connectivity does, along with boundaries found in other modalities, making it useful to the neuroanatomist.

3.4 Individual Resting State fMRI

The left-middle group of Figure 4 (darkest gray background) shows the processing steps after temporal concatenation of the final dense timeseries across rfMRI runs in each subject. Visuotopic regression was done using this concatenated dense timeseries (see below #4.4). Also, the individual subject RSN maps were computed using the concatenated dense timeseries and the group RSN map with the weighted regression algorithm (see above #2.3). The PCA decomposition (d=4800) and Wishart roll off technique can additionally be used for generating high CNR individual-subject rfMRI FC dense connectivity matrices without the need for spatial or temporal smoothing. The process is similar to that for the group analyses, except that no dimensionality reduction is performed and therefore there is no need for second variance normalization and a second PCA. These dense functional connectivity matrices, which do not make use of any group information, could provide a useful reference to compare with the individual subject RSN maps generated using group ICA components and weighted regression.

4. Objective, Observer Independent Assessment of Multi-modal Differences Across the Cerebral Cortex: The Gradient-based Approach

4.1 Motivation

Classical neuroanatomists used visual inspection to delineate areal boundaries based on regional differences in cytoarchitecture or myeloarchitecture, often resulting in divergent opinions and controversies as to where boundaries were located. A major advance was the advent of observer independent, objective approaches that could be applied to postmortem histological sections (Caspers et al., 2013; Schleicher et al., 1999; Schleicher et al., 2009; Zilles and Amunts, 2010). The Zilles and Amunts group used cytoarchitectonic feature vectors to characterize each location along a histological section contour and focused on peaks in a vector difference measure (the Mahalanobis distance) to identify abrupt changes that represent candidate areal boundaries. These transitions could then be tested statistically for significance and interpreted by neuroanatomists to ensure that they were not artifactual (e.g. from folding, veins, etc.). Here we apply a spatial-gradient-based approach that is similar in spirit to the histological Mahalanobis distance approach but is customized for multi-modal MRI data on 2D cortical midthickness surfaces rather than histological sections. We compute the magnitude of the first spatial derivative (gradient) of each modality, which indicates the rate of change in the modality at each spatial location. Locations of sharp change in a modality represent candidate areal borders. These spatial gradients also provide a common medium in which to compare changes across multiple modalities. As discussed below (#5), locations with transitions (gradients) in multiple modalities make particularly strong candidates for areal boundaries.

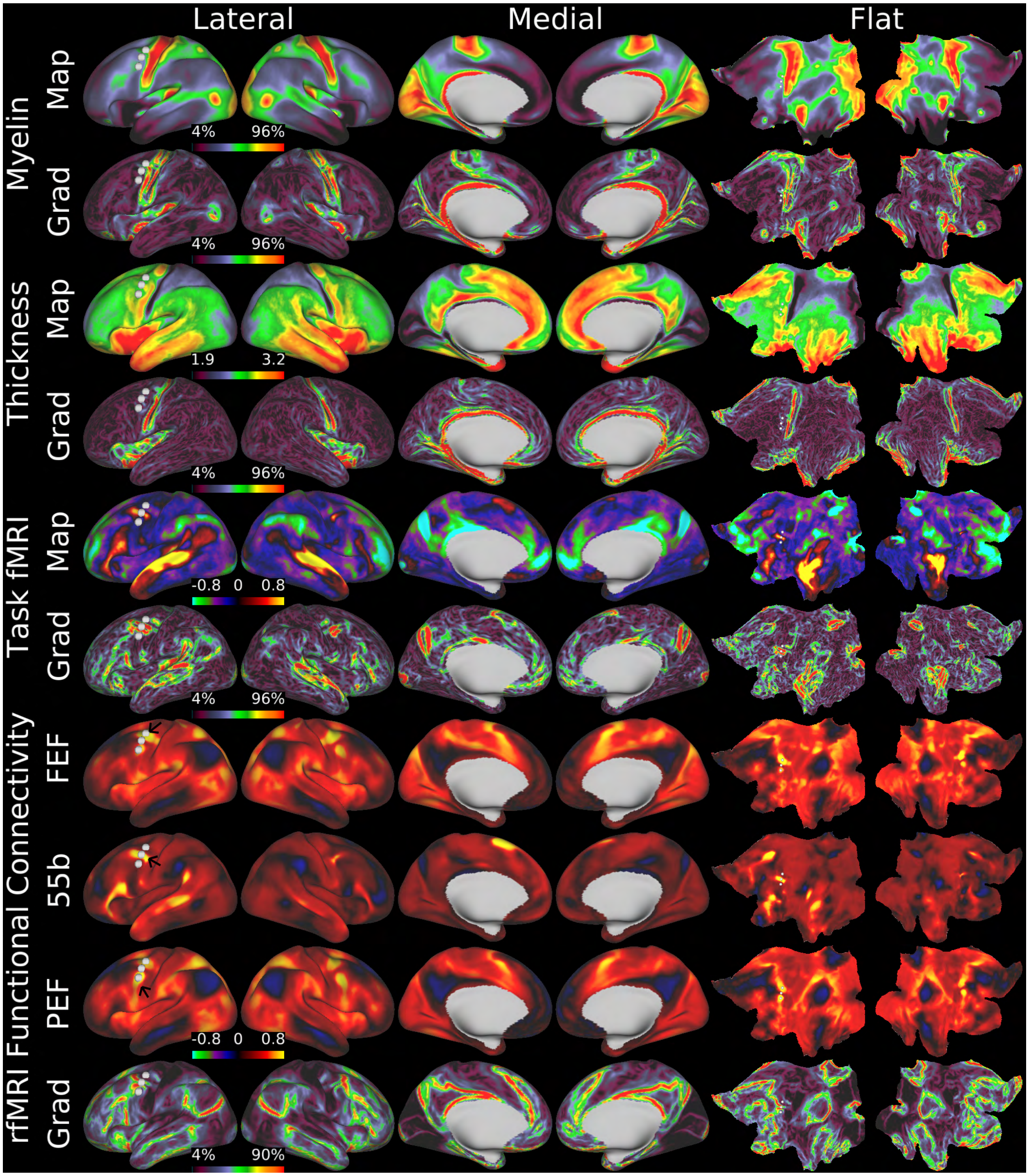


Figure 5, above, shows the four feature categories and their gradients. Rows 1 and 2 show inflated and flat map views of cortical myelin maps and gradients. Rows 3 and 4 show cortical thickness maps with curvature regressed out and their gradients. Rows 5 and 6 show an example task contrast effect size map (scaled to beta = +/-0.8) and its gradient from the LANGUAGE Story vs Baseline contrast. Task fMRI effect size maps can be treated similarly to the architectural myelin and thickness measures as far as gradient computation is concerned. Rows 7, 8, and 9 show three functional connectivity maps from the white spheres representing seed locations in FEF, 55b, and PEF (Row 7 is the most dorsal, Row 8 is the middle, and Row 9 is the most ventral). Row 10 shows the mean functional connectivity gradient map. Note that as one crosses the strong gradients surrounding area 55b from FEF to 55b to PEF, the functional connectivity pattern changes dramatically. This is correlated with smaller changes in myelin density (moderate to light to moderate, See Main Results). Also, note that the functional connectivity map from the 55b seed (Row 8) strongly resembles the LANGUAGE Story task activation map (Row 5). Data at http://balsa.wustl.edu/W1GV.

4.2 Calculation of Surface Gradients on Modality Spatial Maps

Gradients of myelin, thickness, tfMRI contrast effect size maps, and visuotopic regression maps were computed using the following approach (see Figures 5 and 8 for example modality gradients): For each vertex of the cortical midthickness surface on the 32k standard mesh, average the normals of the vertex's associated mesh triangles to obtain the vertex normal vector. Then, "unroll" the vertex and its neighbors (within the cerebral cortex ROI) onto a plane orthogonal to the vertex normal that passes through the vertex as follows (see Figure 6): (i) between the center vertex and a neighboring vertex, draw a circular arc that is tangent to the plane at the center vertex; (ii) compute the length (L_{unroll}) of this arc using the formula $L_{unroll} = L_{Euclid} * \sin^{-1}(L_{opposite} / L_{Euclid}) * L_{Euclid} / L_{opposite}$, where $L_{opposite}$ is the dot product of the vector representing the edge and the normal vector of the vertex, noting that for small angles this calculation may be unstable due to the large radius, and should be skipped for such cases as the correction is negligible; (iii) project the neighboring vertex onto the plane, then modify the projected position to keep the projected direction from the center vertex, but maintain a distance from the center vertex equal to the circular arc length. Fit a linear function within the plane ($f(t, u) = at + bu + c$, where t and u are distances along orthogonal axes within the plane) to the values and positions of the center vertex and the unrolled neighboring vertices neighbors, via regression. The gradient vector is the spatial coefficients, projected into 3D space by the unit vectors of the plane's coordinate system. If the regression is undefined, for instance if there is only one neighbor because the vertex is along the medial wall, instead average the gradient vectors obtained by taking each within-roi unrolled neighbor and using the relative position and difference in value.

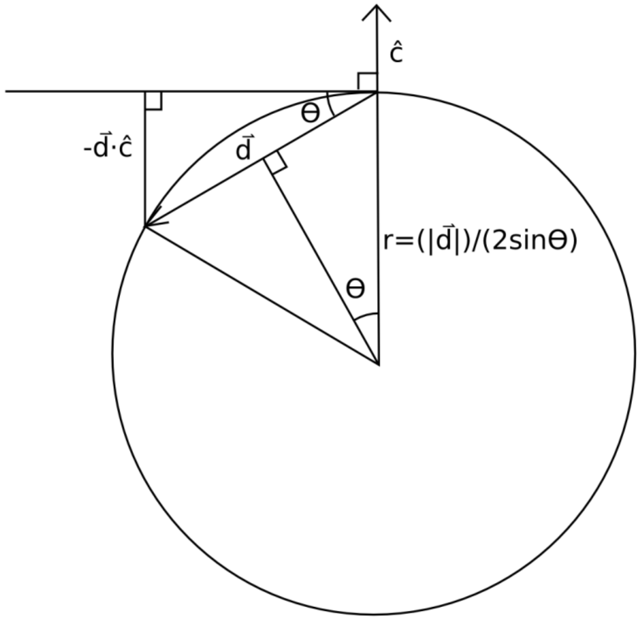


Figure 6 schematizes the geometry associated with the 'unrolling' process involved in computing gradients along the surface. $\vec{d}$ is the vector from the center vertex to the neighboring vertex being unrolled. The circle defining the arc that gets unrolled is drawn, and radii are drawn to both ends of $\vec{d}$. θ is defined as half the angle of the arc, so that the identical right triangles formed from cutting the triangle of $\vec{d}$ and the two radii in half, each have the angle θ at the center of the circle. Hence, the radius of the circle is half the length of $\vec{d}$, divided by the sine of θ. To determine the angle θ, consider the right triangle formed from a line drawn perpendicular to the tangent placed at the center vertex, connecting to the neighbor vertex being unrolled (the head of $\vec{d}$). The length of this segment is (for situations where the circle curves away from the normal vector, as in the figure) the negative of the dot product of $\vec{d}$ and the unit normal vector at the center vertex, denoted as $\hat{c}$. θ is therefore the inverse sine of this negative dot product divided by the length of $\vec{d}$. The length of an arc is the angle of the arc in radians times the radius of the circle, and we now have both, giving the final value as $|\vec{d}| * \sin^{-1}((-\vec{d} \bullet \hat{c})/|\vec{d}|) * (|\vec{d}|/(-\vec{d} \bullet \hat{c}))$, simplifying the sine θ expression as θ is defined by an inverse sine. Due to the properties of sine, the negatives can be dropped from both dot products, giving the simplified formula $|\vec{d}| * \sin^{-1}((\vec{d} \bullet \hat{c})/|\vec{d}|) * (|\vec{d}|/(\vec{d} \bullet \hat{c}))$. Whether the curvature is towards or away from the normal vector, both dot products have the same sign, and therefore the result is positive, meaning that this formula generalizes to both directions of curvature.

When taking the gradient of group-averaged data, a group-average surface is used, along with a group average of the surface areas associated with each surface vertex (referred to as the "corrected vertex areas"). These surface areas enable an approximate correction for the reduction in distance caused by the increased smoothness of the group-average surface. The length of each edge on the group average surface is assumed to be split into two parts, each part residing entirely within the area associated with the vertex at its end. Furthermore, the ratio between the lengths of these parts is assumed to be equal to the ratio of the square roots of the surface areas of the vertices, as given by just the group average surface, without using the corrected vertex areas. To find the corrected length of this edge, scale each of these parts by the ratio between the square roots of the vertex areas from the group-average surface and the corrected vertex areas. This works out to the formula: $L_{\mathrm{corrected}} = L_{\mathrm{surface}} * (\sqrt{}\mathrm{Area}_{\mathrm{corrected}}(A) + \sqrt{}\mathrm{Area}_{\mathrm{corrected}}(B)) / (\sqrt{}\mathrm{Area}_{\mathrm{surface}}(A) + \sqrt{}\mathrm{Area}_{\mathrm{surface}}(B))$, where A and B are the vertices that the edge being corrected joins, and the subscript “surface” denotes lengths and vertex areas derived only from the group surface without correction. The result of the gradient algorithm that is used for parcellation is the magnitude of the gradient vector at each vertex for each map.

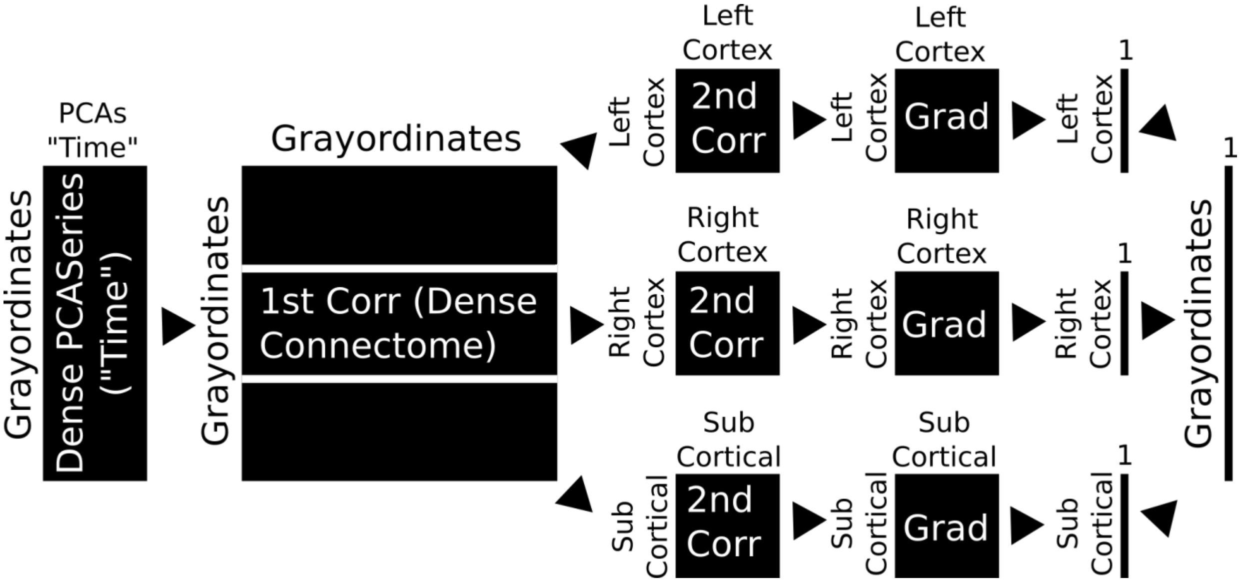


Figure 7 shows the steps used for computation of whole brain full correlation gradients. The process begins with the cleaned, Wishart rolled off dense PCA series (or individual subject timeseries), and includes generation of the dense connectome, a second correlation of functional connectivity patterns between grayordinates of each major structure (e.g. Left Cerebral Cortex or Right Thalamus), computation of the gradient map on each column of the second order correlation matrix, averaging along the rows the resultant gradient matrix, and spatially concatenating the mean gradients of all of the major structures back into an all grayordinates X 1 mean gradient map.

4.3 Calculation of Whole Brain Full Correlation rfMRI Functional Connectivity Gradients

For resting state fMRI, the 2D matrix-based nature of functional connectivity data required a more complicated approach to computing a final gradient map (see Figure 7). For all vertices in either the left or right hemisphere, the grayordinatewise FC maps were correlated to determine the similarity between each vertex’s FC map and all others from the same structure. This produced a vertices X vertices 2nd order correlation matrix (as we want gradients in the similarity of functional connectivity maps—integrating information across the whole brain—rather than gradients in the similarity of rfMRI timecourses). Then the spatial gradient was taken of each of these 2nd order correlation maps, forming a vertices X vertices gradient magnitude matrix. The mean was then taken across this matrix to produce a mean gradient map, referred to as an rfMRI FC gradient map. This gradient map represents the average rate of change in functional connectivity pattern as one moves across the surface. Subcortical gradients were computed similarly, except the voxels of each subcortical structure were used instead. This method is similar in spirit to that proposed in (Cohen et al., 2008), except that here gradients were computed on the group average midthickness surface instead of resampling to a Cartesian grid overlaid on a flattened surface, edge detection was not used, and the second similarity matrix is correlation instead of eta^2.

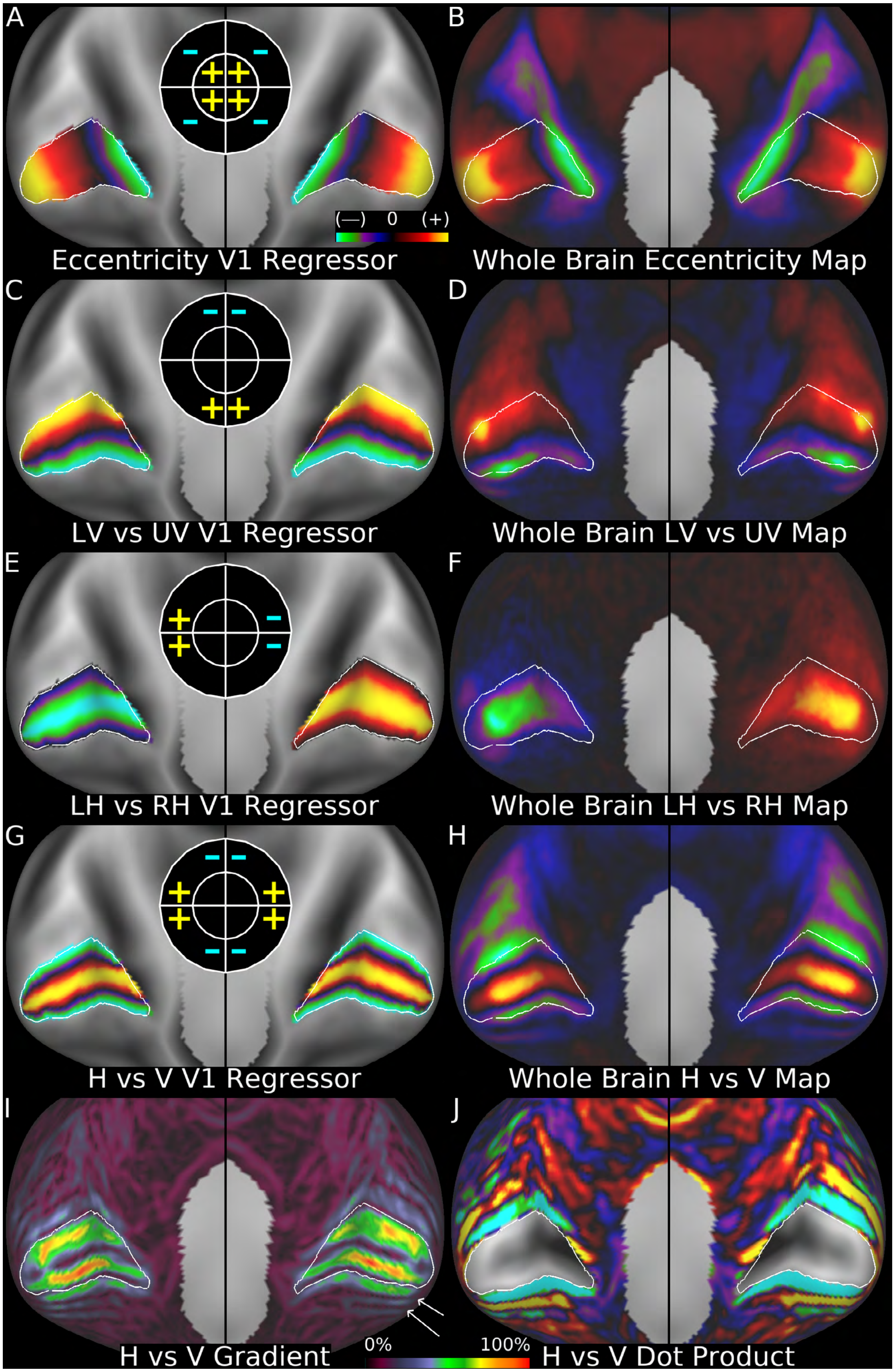

A
(—)
0
(+)
Eccentricity V1 Regressor
B
Whole Brain Eccentricity Map
C
LV vs UV V1 Regressor
D
Whole Brain LV vs UV Map
E
LH vs RH V1 Regressor
F
Whole Brain LH vs RH Map
G
H vs V V1 Regressor
H
Whole Brain H vs V Map
I
H vs V Gradient
0%
100%
J
H vs V Dot Product

Figure 8, above, shows the V1-constrained spatial regressors and their corresponding whole brain spatial maps (along with icons representing the portions of the visual field being contrasted). Panels A and B show the foveal vs peripheral contrast. Panels C and D show the upper vs lower vertical meridian contrast. Panels E and F show the left vs right horizontal meridian contrast. Panels G and H show the horizontal vs vertical meridian contrast, which is the most useful for visual area parcellation. Panel I shows the gradient magnitude of Panel H, where local gradient minima represent the locations of polar angle reversal. Panel J shows the dot product of the gradient vector of Panel H with a reference vector that points towards V1. The early visual areas alternate between positive and negative with the zero point lying on their borders (this measure is undefined inside of V1, see #6.5). Data at http://balsa.wustl.edu/J5k3.

4.4 Extraction of Topographic Information from Resting State fMRI Data

The above gradient-based methods enable objective delineation of areal boundaries based on rapid changes in spatial maps for one or more of four feature categories. Information useful for parcellation also came from continuous, topographically organized gradients in rfMRI FC within cortical areas. For example, antero-posterior gradients related to eccentricity in the visual field have previously been reported for the functional connectivity of early visual cortex (Yeo et al., 2011). Similarly, we previously described continuous topographic gradients associated with polar angle in the early visual cortex (Glasser et al., 2014b). Here, we analyzed continuous visuotopic rfMRI FC gradients using an improved, multiple regression-based method applicable to both group data (high CNR) and individual subjects (low CNR):

We first defined the full extent of area V1 using a combination of Lateral Geniculate Nucleus (LGN) rfMRI FC maps, myelin gradients, and thickness gradients (see NSR #1). Then, based on published maps (e.g. (Sereno et al., 1995)) of eccentricity and polar angle in V1, a set of nine spatial regressors based on continuous linear gradients were defined across V1 (see Figure 8), including 1) a spatial contrast between upper and lower vertical meridians (Panel C, varying from 1 to -1); 2) a spatial contrast between left and right horizontal meridians (Panel E, varying from 1 to -1); 3) a spatial contrast between the full (upper + lower) vertical meridian and the full (left + right) horizontal meridian (Panel G, varying from 1 to -1); 4) the third contrast shifted by 45 degrees (to complete the harmonic); 5) Foveal vs peripheral (Panel A, varying from 1 to -1); 6) the foveal vs peripheral gradient squared; 7) the foveal vs peripheral gradient cubed; 8) bilateral entire V1 (all ones); and 9) All grayordinates as ones. Outside of V1 the first 8 spatial regressors were all zeros. These regressors were defined using linear intensity gradients, which were computed algorithmically from a standard set of surface borders by dilation, using either the group average midthickness surface for the group maps or the individual subject midthickness surface for the individual maps (i.e., borders representing the horizontal meridian known to run down the center of V1 in the fundus of the calcarine sulcus, the vertical meridians known to run along the superior and inferior borders of V1, and the eccentricity arrangement known to run from foveal posteriorly to peripheral anteriorly). More specifically, generation of linear gradients involved a customized form of dilation on the surface, starting with a spatial map having a set of desired values along isoeccentricity or isopolar lines within V1 or along its border. For each V1 vertex that did not already have a value, the vertices within a specified distance whose straightest surface-based path to the chosen V1 vertex did not cross another vertex that already had a value were selected. From this set, the pair of vertices that had the largest value for the difference in their input values divided by the sum of their distances from the chosen vertex was selected. This was used to linearly interpolate the output value based on the pair's distances and input values.

These nine V1 regressors were multiple-regressed into the group average dense PCASeries (or individual subject dense timeseries) using a variant of the WR process described above (#2.3, see Figure 9). Specifically, the V1 regressors were first spatially regressed into the group dense PCASeries (or individual subject dense timeseries) using only the Vertex Area spatial weighting map to generate V1 regressor-specific PCACourses (or timecourses). These PCACourses (or timecourses) were then temporally regressed into the group dense PCASeries (or individual subject dense timeseries) to generate a set of whole brain spatial maps (Panels B, D, F, H), one associated with each of the originally V1 restricted spatial regressors. For the MSMAll registration, the resulting individual subject whole brain maps were registered to the group whole brain maps (an earlier version of the visuotopic maps was used for the MSMAll registration that spanned the eccentricity space as bands rather than continuous linear gradients).

The spatial maps generated using the vertical vs horizontal meridian contrast (Panel H) and the upper vs lower hemifield contrast (Panel D) were particularly useful in delineating extrastriate visual areas, as they identify the polar angle reversals that occur at areal borders. Visuotopic areal borders generally occur at representations of the vertical or horizontal meridian (Sereno et al., 1995), corresponding to local maxima or minima in these spatial maps and to local minima in the spatial gradients computed from these maps. For example, Panel I shows the spatial gradient magnitude of the map in Panel H and has multiple spatial minima (e.g. arrows) that we later show correspond to areal boundaries (see NSR #2). The topographic information revealed by rfMRI FC was particularly useful for defining early visual cortical areas, including V2, V3, V4, V3A, and parts of some higher visual areas including V6, V7, V8, VVC, LO1, ProS, V3B, V3CD, VMV1, VMV2, VMV3 (see NSR #2-4).

Additional useful characterization of visuotopic organization was obtained by computing the dot product at each vertex between the gradient vector and a vector pointing towards V1 along the surface. This map is shown in Panel J for the gradient vector associated with the vertical vs horizontal meridian contrast gradient magnitude in Panel I. This reveals where the visuotopic gradient vector changes direction at visuotopic areal borders and is analogous to the field sign measure of (Sereno et al., 1994). Because the resulting dot products are all positive or all negative in a given area, this information is directly interpretable by the areal classifier (see below #6.5), which otherwise does not have access to explicit information about intra-areal spatial patterns. This dot product can also be compared statistically across the areal boundary.

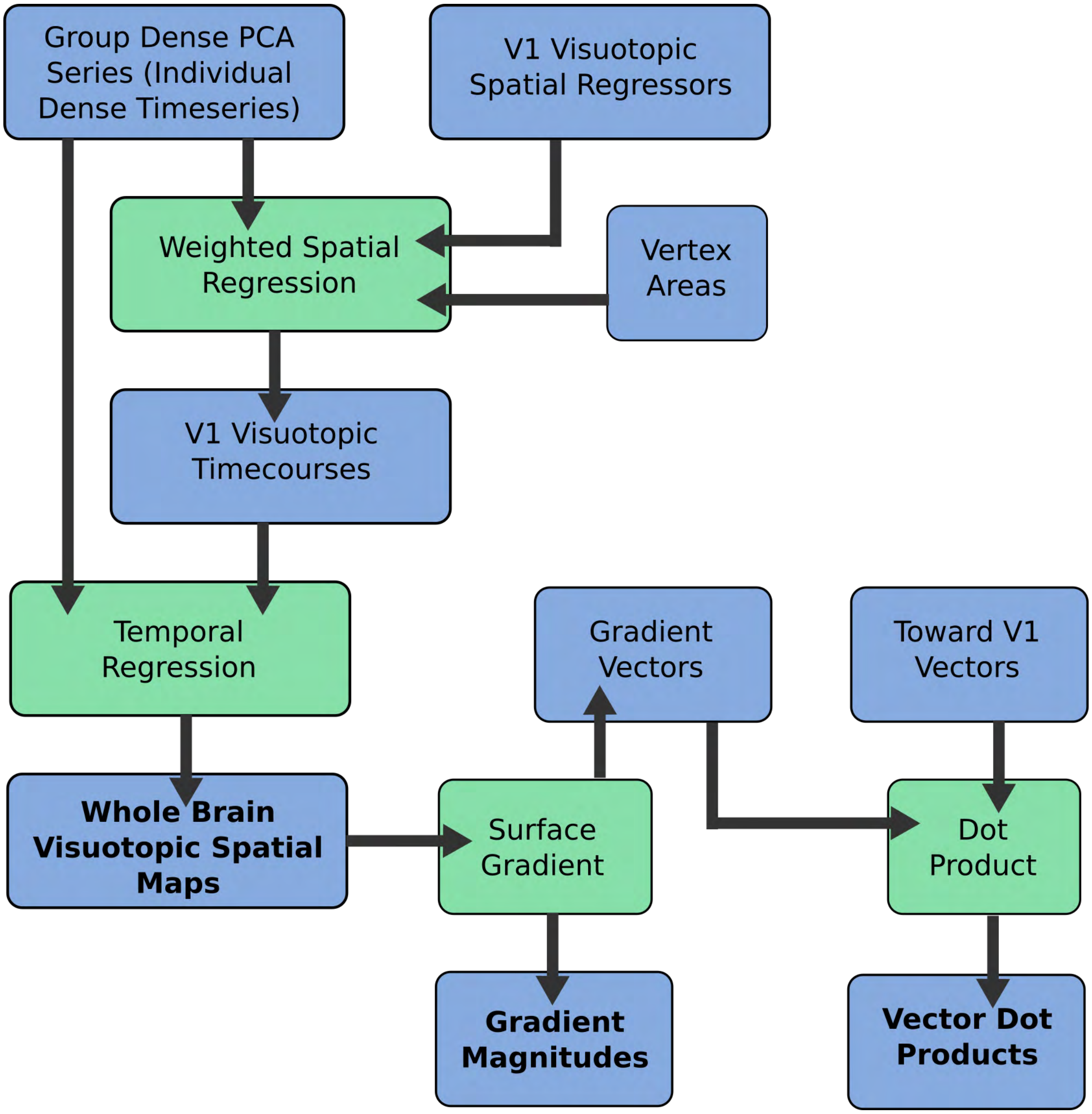


Figure 9 shows how V1-specific spatial regressors are used to produce corresponding whole brain spatial maps of various visuotopic spatial contrasts. The useful outputs are marked in **BOLD** font. The procedure can be used on either group dense PCA series or individual subject dense timeseries. It first makes use of the vertex areas in a weighted spatial regression, and the resulting V1-specific timecourses are used in a temporal regression together with the dense series to produce whole brain spatial maps. Gradient magnitudes can be computed on these maps, the local minima of which are useful for positioning areal borders (minima instead of maxima because the topographic maps themselves are gradients, and we are interested in local maxima or minima in these gradient maps where the second derivative is close to zero, as these locations delineate areal borders). Additionally, the gradient vectors can be multiplied in a dot product with the vectors that point towards V1 along the surface to determine where the visuotopic vectors switch direction (which occurs at the areal border, similar to the field sign measure of (Sereno et al., 1994)). This dot product yielded maps in which the values within a given visual area are all one sign (and different from neighboring areas), which were interpreted by the areal classifier (see below #6.5) and were also used to measure initial cross-border statistics.

5. Multi-modal Cortical Parcellation Using Gradients

5.1 Neuroanatomical Criteria for Multimodal Delineation of Cortical Areas

The preceding sections have described algorithmic methods for extracting information useful for parcellation from multiple independent modalities in a common grayordinates space (i.e. allowing visualization of modalities and their gradients on the same surfaces in Connectome Workbench). Gradient maxima from myelin maps, thickness maps, tfMRI contrast maps, rfMRI FC maps, plus gradient minima in continuous, topographically organized rfMRI FC within cortical areas together provide objective evidence for cortical areal boundaries. These non-invasive measures of architecture, function, connectivity, and topography are analogous to measures acquired invasively in non-human primates and in post-mortem human brains. To integrate this information across modalities and facilitate comparisons with published cortical parcellations, we combined a semi-automated boundary-delineation algorithm described below with neurobiologically informed interpretations based on five general criteria.

(1) To accept a putative areal boundary, we placed a strong reliance on consistent boundary location across independent modalities (for this analysis, the various ways of processing rfMRI were not considered "independent" in a full statistical sense, but agreement between them was still considered helpful).

(2) Presence of a putative boundary at corresponding locations in the left and right hemispheres increased our confidence in it. In this study we did not identify any areas that were present in one hemisphere but not the other because we did not find strong enough evidence to support such a parcellation, though some homologous areas assigned the same names show substantial lateralization of function and/or connectivity (see NSR and Main Text). A small number of areas have differing topological relationships across hemispheres or use different features for delineation in the two hemispheres. More focused future studies that find strong multi-modal evidence for hemispheric differences in cortical areal parcellation might choose to revise this criterion, however.

(3) Candidate boundaries were rejected if they were attributable to known imaging artifacts, which include: (i) fMRI signal loss from magnetic susceptibility in orbitofrontal and inferior temporal cortical regions[4]; (ii) folding-related artifacts in myelin maps that occur in thin, heavily myelinated early visual cortex (attenuated in 0.7 mm isotropic T1w and T2w images, but not eliminated); (iii) susceptibility-related artifacts in cortical thickness (in the same regions as fMRI, but less extensive, and caused by signal loss in the gradient echo T1w image); (iv) limitations related to task design in the tfMRI contrast data, including left hemisphere activation of right hand regions from button box pressing in cognitive tasks and preferential activation of the central visual fields because of the limited region of the visual field able to be stimulated by the task in the scanner; and (v) differential activation of foveal visual cortex from the visual fixation crosshair in rfMRI scans. Each of these artifacts can give rise to spatial gradients that are irrelevant to cortical parcellation. Reliable automated detection of such artifacts is currently impractical, requiring interpretation of the data by a neuroanatomist with expertise in neuroimaging.

---

[4] Susceptibility-loss regions were smaller in the high resolution HCP data than in conventional fMRI scans, but were nonetheless present.

(4) Cortex on opposite sides of a putative areal boundary should differ robustly and statistically significantly in areal features, in order to exclude gradient ridges that are simply local noise-related fluctuations.

(5) Confidence in a putative areal boundary is increased by correspondence with a published areal boundary. Many areas that we identified correspond to previously reported areas but can now be delineated in a large number of living subjects based on the high-quality HCP data and analyses.

(6) Initial runs of the areal classifier (see below #6.1-6.8) on only the 210P and 29T subject groups were invaluable for finding putative areas in the initial draft parcellation that were not consistently identifiable in individual subjects. Also, the initial areal classifier results revealed "area holes" in which no area tended to reach at least 50% probability across subjects yet were not otherwise explainable by poor multi-modal CNR (e.g. susceptibility artifacts). Some of these regions were more finely parcellated, thereby improving classifier performance (see Section #6.7). Thus, the automated individual subject areal classifier served as one objective check on the neuroanatomists' interpretation of the gradient data.

5.2 A Semi-automated Algorithm for Objective Gradient-based Delineation of Areal Borders and Initial Internal Statistical Validation

We implemented a semi-automated algorithm for objective gradient-based delineation of areal borders. When a putative areal border was identified in a given region (based on visual inspection of multiple areal features and gradients), it was manually drawn in approximately the correct location using the border-drawing tool in Connectome Workbench. An optimization region of interest that included a substantial portion of the two cortical areas on opposite sides of the border was also drawn manually. Using the border optimization function in Connectome Workbench, the modalities used by the neuroanatomists to recognize the border were selected, and the modality gradients were combined inside this ROI by a method that uses a tunable tradeoff between averaging and multiplying. Each gradient is individually rescaled, mapping its minimum and maximum value within the ROI to 0 and 1, respectively, then put through the formula "precombine = rescaled * strength + 1 – strength", where strength is a tunable parameter between 0 and 1 (default of 0.7). The resulting "precombine" values are multiplied together across modalities at each vertex, and the result is again rescaled to the range of 0 to 1 and then inverted, so that where the gradients are maximum, the value is 0. We used these values to assign a cost to each edge (line segment between adjacent surface vertices) within the entire ROI by multiplying the length of the edge by the expression "1 + followstrength * (data(i) + data(j))", where "i" and "j" are the vertices at each end of the edge being evaluated, "data" is the combined gradient data, as described above, and "followstrength" is a tunable parameter that controls how much the gradient data can influence the cost of an edge (which indirectly controls how much the optimal path can deviate from the shortest geodesic path between the endpoints, and has a default value of 5). The initial border was then revised by taking the first and last points of the border segment inside the optimization ROI and using an "A*" search (a standard graph search algorithm (Hart et al., 1968)), to construct a new path having the minimal sum of the costs of the edges in the path.

For the vast majority of the borders we kept the strength and followstrength parameters at their defaults (which were considered to be reasonable values by TSC, the algorithm designer and indeed worked well in initial testing). The approximate reasoning behind the initial choice of followstrength was that a value of 5 allows the path to be up to 5 times longer than the shortest path, as long as there is an undisputable gradient ridge to follow, which seemed sufficiently loose for the search to choose reasonable paths given realistic gradients. For the strength parameter used during the combination of gradients, a value of 0 means to ignore the map (and 0.01 for all inputs would result in something very close to pure addition of the gradient maps), 1 means pure multiplication, and 0.7 was considered a reasonable balance to achieve a more forgiving multiplication-like effect, given what the extremes represent. In some cases, the strength parameter was set to zero for the feature maps where we wished to compute cross-border statistics (so that we could report the difference between the areas in the NSR) but did not want the map to contribute to the final effective gradient. For example, if myelin, resting state, and 5 task contrasts were all significant, but we did not want the task modality to dominate the effective gradient, we would set some task contrast maps to zero. Otherwise, we kept the followstrength parameter set to its default, except in two explicity noted instances where a border passed through a susceptibility artifact and we did not want it to take a highly circuitous route (here we reduced the followstrength value, which weights the border more towards the shortest path).

Each automatically defined border segment was used to assess the effect size (Cohen's d value) and significance of the differences between feature map values on opposite sides of the border within the optimization ROI. For most modalities the p value was based on a t-test for the feature map values on either side of the border. For functional connectivity, a different approach was used. For each "seed" vertex within the optimization ROI, an average functional connectivity map was made for each side of the border within the optimization ROI; a unique average was made for each vertex to exclude vertices within 2 mm geodesic distance of the seed vertex (mainly to exclude the seed vertex itself). The seed vertex's connectivity map was then correlated with each of these average connectivity maps (average for side "A" and average for side "B"). The resulting correlation values were Fisher z-transformed and grouped according to which side the seed vertex is on, and which side the average connectivity map is from. Two t-tests were then performed on these Fisher z-transformed correlation values: one was between the group of values from seed vertices from side A to average connectivity maps on side A versus the group of values from seed vertices from side B to average connectivity maps on side A. The other was between the group of vertices from side A to side B vs side B to side B. The test having the higher p value and lower Cohen's d was used for checking if the border indeed had a robust and statistically significant effect. Two tests were used because there are inherently four groups, but two pairings are not meaningfully comparable for our purposes (A to A average with B to B average, A to B average with B to A average).

For each modality, p values were generated, and boundaries that did not have significant p values ($p<0.05$) in at least one modality were not retained. In practice, if the neuroanatomist recognized a border, the p value tended to be very low and the Cohen's d standard effect size measure was large (maximum $p<0.001$ and minimum Cohen's $d>0.5$), because the group average data for our large subject groups has relatively low noise. We used both p values and Cohen's d values in order to identify differences that were both robust and statistically significant. In initial testing, we found that the exact location of the initial border estimate and the optimization ROI had little impact on the final automatically delineated border location or the statistical test values, as long as the initial estimate was within a few mm of the combined gradient and the border optimization ROI covered a substantial portion of the cortical areas on either side of the boundary.

This initial statistical test served mainly as an additional check for the neurobiological plausibility of the boundaries being defined (i.e. criterion #4 above; see section #7.2 below for more rigorous statistical cross validation of the parcellation reproducibility and multi-modal cross-border differences using the independent 210V dataset). The end result of the semi-automated process was a set of objectively defined boundaries dividing the neocortex of both hemispheres into a mosaic of areas differing in architecture, function, connectivity, and/or topography. The primary representation of these boundaries was in the form of surface borders (strings of points projected to a surface tessellation but not constrained to discrete surface vertices). These borders were used to generate ROIs (collections of vertices) and labels for each area.

5.3 Neuroanatomical Criteria for Identifying Cortical Areas Based on Spatial Neighborhood Relationships

The final stage of the group-average parcellation process involves identification and naming of cortical areas. For purposes of continuity with prior studies, we retained cortical areal names if they were well matched to a published parcellation, whereas new names were assigned to areas that lacked such a match. When a previously identified cortical area could be subdivided in our multi-modal analysis, we used a similar area name with appropriate modifiers (See NSR). Similarly, if previous studies reported multiple subdivisions of a region identified as a single cortical area in our analysis, we incorporated both names into the area name and called it a "complex."

For some areas, we were able to use specific areal features (such as heavy myelination or visuotopic organization) as part of the evidence for matching our areas to published parcellations based on in vivo or postmortem analyses. However, in most cases spatial location and topological relationships with adjoining areas were the primary basis for matching with published parcellations. If a published parcellation and the multi-modal parcellation show a similar set of areas with similar spatial relationships, the most parsimonious explanation is that they represent corresponding cortical areas.

A major challenge in using spatial relationships to compare across parcellation studies is the diversity of methods used for intersubject alignment, of which six categories warrant brief mention. (i) We placed highest confidence in group average areal definitions that are based on cortical areal feature-based surface registration (which will produce maximally sharp group maps) and include dedrifting during template generation, so that an overlay of data from different studies has minimal bias and highest fidelity (Abdollahi et al., 2014). (ii) Next best are probabilistic maps based on folding-constrained surface registration (e.g. (Fischl et al., 2008)) that are overlaid together with the multi-modal parcellation. Comparisons with the present data should be reasonably accurate, though they may be blurrier in regions of high folding variability or high variability between areas and folds. Additionally, unless dedrifting has been done (see above #2.5), there may be spatial drift between the group average patterns of different studies (this may explain some discrepancies between our area V1 and that of (Fischl et al., 2008), see NSR #1). (iii) Surface-based figures and drawings from published studies were often useful for matching with our parcellation even when direct comparison on the same surface mesh in Connectome Workbench was not feasible. (iv) Volume-registered probabilistic maps of cortical areas (e.g. (Eickhoff et al., 2005)) were more challenging to use because of inaccuracies in mapping from group average volume to group average surface spaces, as well as inaccuracies in intersubject alignment contributing to a group-average volumetric map ((Anticevic et al., 2008; Fischl et al., 2008; Fischl et al., 1999b; Frost and Goebel, 2012; Smith et al., 2013a; Tucholka et al., 2012; Van Essen et al., 2012)). Volume-based registration can also have group average drifting issues (for instance, if using the MNI template). However we did use the JuBrain online resource ((Caspers et al., 2013); https://www.jubrain.fz-juelich.de) to help with identifying some areas. (v) Volume-based group average figures in published studies typically lack the detail required to distinguish cortical areal boundaries and their spatial relationships, so these were used sparingly. (vi) Stereotactic coordinates (e.g. MNI or Talairach), based typically on centers of volumetrically analyzed group-average task activation foci, generally lack information about cortical areal borders and are thus of limited utility for making definitive areal identifications, even though a major fraction of the neuroimaging literature uses this approach as a primary means of expressing spatial localization (see examples of this difficulty in (Glasser and Van Essen, 2011) myelin mapping Figure 11).

5.4 Coloration of the Multi-modal Parcellation

A parcellation can be colored in many ways, e.g. solving the map-coloring problem (Ringel and Youngs, 1968) with random colors, using only a single color, or coloration based on some data driven metric. For our parcellation, we wanted a color scheme that (i) utilizes a large portion of the printable and computer-displayable 3D color space (hue, saturation, and luminance) and (ii) is neurobiologicaly informative, for example by conveying how the major sensory inputs and two major anti-correlated cognitive networks are organized and how they are mixed in different cortical areas. We used rfMRI relationships between each area and five core groups of areas for this purpose. The group MIGP PCASeries was parcellated (i.e., the average PCASeries was computed inside each area) according to the multi-modal parcellation separately for each hemisphere. Core groups of areas (with each group covering similar amounts of cortical territory) were selected that are associated with auditory (A1+MBelt+LBelt+PBelt+RI), somatosensory (4+3a+3b+1), visual (V1+V2+V3), task positive (PF+PHT+23c+46), and task negative (a.k.a default mode) networks (PGi+PGs+TE1a+7m+v23ab+10r+10v). Auditory, somatosensory, and visual groups were selected because these represent the main input modalities to the brain. Task positive and task negative cognitive networks were selected because they are the strongest anticorrelated resting state networks in the brain (Fox et al., 2005). In calling these networks “task positive” and “task negative” we do not mean to suggest that tasks never activate the task negative network or always activate the task positive network (e.g. the LANGUAGE Story vs baseline task contrast in the HCP dataset activates the task negative network but not the task positive network), but rather to highlight the anti-correlated relationship between these networks and avoid making assertions about what the brain's "default mode" is. Nodes of the task positive and task negative networks were selected in parietal, temporal, cingulate, and frontal lobes.

Five area groups were chosen because that is the maximum number whose colors can be blended uniquely. The mean PCASeries of the 5 groups of areas were computed and these were temporally multiple regressed into the parcellated PCASeries, producing a 180 X 5 partial beta matrix. This matrix was then scaled to range from 0 to 1. The scaled matrix was then matrix multiplied with a primary colors matrix [255 0 0 ; 0 255 0 ; 0 0 255 ; 255 255 255 ; 0 0 0] (pure red, green, blue, white, black) to produce an 180 X 3 color components matrix. This matrix was rescaled from 0 to 255 and used as the label colors for each area (e.g. in Main Text Figure 3, see associated text for discussion of the neurobiological interpretations of the color patterns).

6. Automatically Defining and Identifying Cortical Areas in Individual Subjects Using Multi-modal Areal Fingerprints

6.1 Motivation

The group average parcellation described in the preceding methods sections and in the NSR is very useful in its own right for various purposes. However, it is also desirable to generate parcellations of individual subjects (including those who were not part of the original group-average parcellation) that are as accurate as possible at capturing individual variability. Despite the improvements in areal feature-based surface registration reported here and elsewhere (Robinson et al., 2014), there remain residual misalignments between particular individuals in particular areas and the typical subject's areas found in the population-based parcellation. We found that a major source of these residual misalignments involved topologically incompatible differences in the spatial neighborhood relationships between some areas in some individuals and the typical pattern of areal spatial neighborhood relationships in the group. For example, this includes areas that are a single contiguous region in most individuals, but with strong evidence for multiple nearby but discrete patches in some individuals (see ESDI #1.3-1.4). Topology-preserving spatial registration, such as MSM, is inherently unable to compensate for such topological differences (and if regularization was relaxed in an attempt to align such areas, the resulting estimated warps would contain excessive, neurobiologically implausible amounts of areal distortion).

One approach to individual-subject parcellation is to rely on the alignment achieved by the registration process to map the group-average parcel boundaries back to each individual. However, this does not take into account the residual misalignments after the registration process, such as the topological mismatches noted above. To enable accurate analyses of individual variability in areal size, functional activity, or connectivity, we sought a way to generate multi-modal parcellations in the individual subjects that account for the residual spatial variability in areal locations. The semi-automated methods used to generate the group parcellation above would be far too tedious, especially when dealing with a large number of subjects (e.g., the full ~1100 subjects having MRI scans for the HCP or future studies that may wish to take advantage of the multi-modal parcellation). Fortunately, we were able to capitalize on machine learning methods capable of addressing this problem once the typical subject's parcellation has been generated at the group level.

6.2 Machine Learning Classifier Design

We started with a multi-layer perceptron machine learning classifier used to classify 7 resting state networks from functional connectivity maps (Hacker et al., 2013) and adapted it to the problem of classifying cortical areas based on their multi-modal areal fingerprints. Classifying 180 areas in a single classifier run across the whole brain is an extremely hard problem, made even harder by the fact that spatially distant areas (e.g., in different lobes) may have similar areal fingerprints. Fortunately, the problem becomes much simpler by making two neurobiologically reasonable assumptions: (1). We assume that after the areal feature-based surface registration, individual cortical areas are in reasonably close proximity to the group areal definition. (2) It is sufficient to distinguish each cortical area from all of its spatially adjacent neighbors combined into a single class. Binary classification (class one=area, class two=neighbors) is the simplest and most robust kind of classification problem. Thus, the classification problem setup and classifier architecture were as follows: For each of 180 areas in each hemisphere, a multi-layer perceptron was trained to classify whether or not each vertex in an ROI containing the group areal definition plus a 30 mm surface geodesic distance surrounding it is a part of the area or not—the neuroanatomical "searchlight" for the area. The multi-layer perceptron had 3 layers (1 input, 1 hidden, 1 output) and the same nonlinear functions as in (Hacker et al., 2013) and used 9 hidden nodes (a reasonable compromise between accuracy and training speed for this problem). The training labels were the group average areal definition and the 30 mm radius spatial neighborhood surrounding it (see below #6.6 for detecting and excluding residually misaligned areas in specific subjects from classifier training).

6.3 Multi-modal Features for Areal Classification: Myelin, Thickness, and Resting State Functional Connectivity

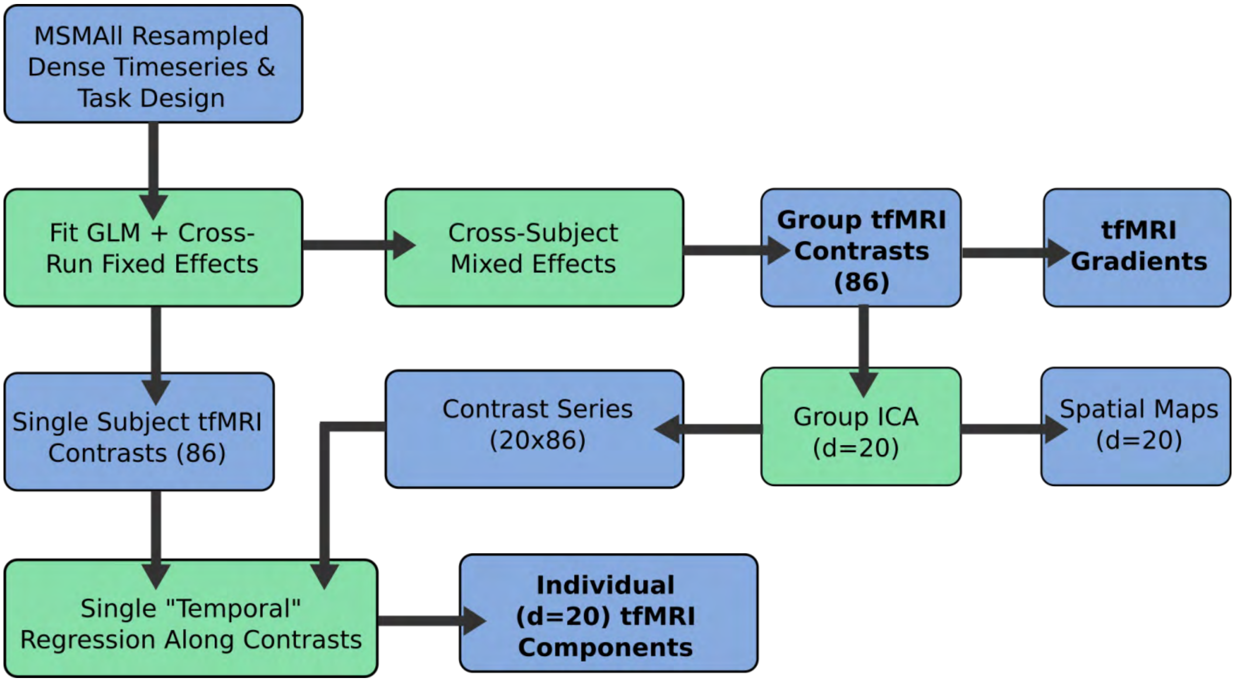
The features used by the classifier were based on the various modalities used in the semi-automated parcellation, but they were optimized for single subject analyses, classifier training speed, and accuracy. Simplest were the individual subject myelin and cortical thickness maps corrected for folding effects, for which we used the unsmoothed left and right cortical surface maps of each measure.

Because there is no temporal correspondence across subjects in the rfMRI timeseries, the only corresponding rfMRI subject-wise (i.e. without reference to group data) dataset is the rfMRI FC "dense connectome" (91,282 X 91,282 X 4 bytes = 32.5GB). As mentioned above in the section on areal feature registration, each dense connectome contains a large amount of redundant information and (unless adequately processed) a large amount of unstructured noise (see above #2.2). Such data is impractical as input to a classifier, and, as with the MSMAll registration, a dimensionality reduction with consistency across subjects is required. As with MSMAll registration, this was done using group ICA (d=137) of the MIGP PCASeries, followed by weighted regression into the individual subjects' final dense timeseries (see above #2.3) to obtain individual subject versions of the component spatial maps with the same splitting and component numbers.

The choice of RSN dimensionality for the classifier rfMRI modality entails a balance between classifier accuracy (measured by detection rate of the area—was an area found with between 0.33x and 3x the group area size) and the classifier's ability to detect misaligned areas (assessed by visual inspection in the 210P dataset for a few cortical areas determined to have substantial cross-subject variability in spatial position of areal features easily discernible for multiple modalities). Because a central objective of the classifier is to detect these misaligned areas, the classifier step is not useful if it only finds the area in the typical group location in every subject, if in fact some subjects actually have it in an atypical location.

Preliminary classifier runs using dual regression (DR), large numbers of components (e.g. 200), and with the original MSMRSN registration that tolerated substantial distortion (Robinson et al., 2014) tended to have this negative property of failing to identify misaligned areas. Better misaligned area performance was achieved with classifier runs using much lower RSN dimensionalities (e.g. 27), despite having overall lower areal detection rates. Although larger numbers of group ICA components tap deeper into the partial correlation relationships between areas, it is problematic if they begin to cause overfitting of the classifier. The effort to address this tradeoff between overall detection rate and ability to identify misaligned cortical areas for low vs high RSN dimensionalities led to a number of the methodological improvements described in this study. This includes the use of the less distorting MSMAll registration; the use of weighted regression instead of DR; the use of only the predominantly surface RSNs in the classifier (i.e. only maps with useful contrast); and the use of an RSN dimensionality that describes the data by a local minimum in the number of spatially contiguous regions, prior to the point at which the ICA starts adding substantial amounts of unstructured noise to the subcortical components (as mentioned above, at around d=140). All of these changes empirically improved classifier performance (at detecting misaligned areas at higher RSN dimensionalities) and helped avoid overfitting. Some of them also improved the speed of training the classifier by reducing the number of features used. Thus, an RSN dimensionality of d=137 was chosen because it offered high areal detection rates while still being able to detect misaligned areas well, perhaps because it contains most of the available useful rfMRI information in the HCP dataset.

Figure 10, below, shows the task processing used in this study. Starting from the task design and the MSMAll_DeDrift resampled tfMRI dense timeseries for each run, the important outputs in **BOLD** font are produced. The task GLM is fit and fixed effects cross-run within subject analyses are performed to produce 86 individual subject task contrast maps. Cross-subject mixed effects analyses are also performed to produce 86 group task contrast maps (beta maps) and their associated gradients. Because there is a lot of duplicate information in the task fMRI contrast map datasets, group ICA can then be performed on these maps at d=20 (>99% of the variance across contrasts retained), to find the unique information in the tfMRI dataset and produce spatial maps and contrast series. These contrast series can be multiple regressed into the individual subject task contrast maps using a single "temporal" regression stage to produce 20 individual subject tfMRI component maps that can be used in the areal classifier. These 20 maps contain higher CNR than the original 86 maps because of cross-contrast averaging and reduction of the redundant information across contrasts. Additionally, the mean task fMRI image is included as a feature but not shown above (because the ICA decomposition does not include the mean image).



6.4 Multi-modal Features for Areal Classification: Task fMRI

Figure 10 provides an overview of the tfMRI processing used in this paper. The HCP task analysis pipeline generates 86 task contrasts, so the tfMRI contrast information is a 91,282 X 86 contrast beta matrix for each subject (and each whole group). By both visual inspection and preliminary classifier performance, the individual subject tfMRI contrast maps had substantially lower CNR per map than the other features used in the classification (thickness may contain less overall useful information, but the information it does contain has high CNR). Additionally there is extensive redundancy of information in the group and individual tfMRI contrast maps. At the group level, this redundancy was assessed by performing group ICA on the 86 group mixed effects tfMRI contrast beta maps. Twenty ICA components sufficed to represent >99% of the variance in these 86 group contrast maps. Interestingly, by visual inspection many of these ICA components closely resemble rfMRI group ICA components, though the dimensionality at which they emerge may be higher than a d=20 rfMRI group ICA. Also some components found in the d=20 rfMRI group ICA are absent from the d=20 tfMRI contrast map group ICA. This suggests that the HCP task battery was imperfect in covering all possible cognitive brain states measurable by BOLD fMRI, even though efforts were made to provide broad spatial coverage of the brain (Barch et al., 2013). The finding of similarities between rfMRI group ICA components and tfMRI contrast map group ICA components is reminiscent of (Smith et al., 2009) and is consistent with (Cole et al., 2014)'s finding of very similar patterns when all of the HCP tfMRI data are processed like rfMRI data and compared with rfMRI data.

Because the d=20 dimensionality reduction of the group ICA components reduces redundancies, it reduced classifier computational time and also improved individual subject tfMRI contrast map CNR. Group ICA produces both component spatial maps and component "contrast series" (the equivalent of a component timeseries when using rfMRI data). Because each of the 86 tfMRI contrasts means the same thing across subjects (as opposed to resting state timeseries that do not correspond across subjects), a simple single "temporal" multiple linear regression stage was sufficient to produce individual subject dimensionality-reduced tfMRI contrast maps. The 20 group component tfMRI contrast series were "temporally" regressed into the 86 individual subject tfMRI contrast maps. This produced 20 tfMRI component contrast maps with the redundant information removed and CNR improved because of averaging across multiple contrast maps. Additionally the mean contrast map was included because the ICA decomposition does not represent the mean map, only the variance across maps. These 21 maps were used in the areal classifier to represent the tfMRI modality. Preliminary classifier runs indicated that this data was used more efficiently than using 86 contrast maps, as evidenced by higher feature weights for tfMRI. Also, in a separate analysis (see Main Text) the classifier was run without including any tfMRI data at all to see how it would perform in studies where the HCP's task fMRI battery was not available.

6.5 Multi-modal Features for Areal Classification: Within-area Smooth Gradients in Visuotopic Maps Derived from Resting State Functional Connectivity

A final source of information fed into the areal classifier was the visotopically-organized rfMRI FC used for defining early extrastriate visual areas (see above #4.4, and NSR #2-4). The multi-layer perceptron classifier cannot directly interpret smoothly varying intra-area topographic gradients within the whole-brain visuotopic spatial regression maps. This is because the classifier treats each value of each feature independently, and has no way of assessing spatial patterns in the values (it only knows if they are distinctly higher or lower in a given cortical area vs the 30 mm surrounding it). As mentioned above (see Figures 8 and 9), it was necessary to transform the topographic data into a form that the areal classifier could use. We achieved this by taking the dot product of the gradient vector of the visuotopic whole brain regression map and a reference vector whose direction is based on the start of the shortest geodesic path towards the V1 ROI.

The resulting dot product maps contain positive values when the gradient vector points towards V1 and negative values when it points away from V1. The change from positive to negative should occur at visuotopic areal boundaries, converting the within-area spatially varying pattern into a map that is consistent across the entire area and can be used by the classifier in the same way that it handles other modalities. For example, we used the gradient of the spatial contrast between horizontal and vertical meridians to define cortical areas V2 and V3 in extrastriate cortex (NSR #2). To the extent that visuotopy within areas V2 and V3 is perfectly orderly, the gradient vector of this spatial contrast should always point away from V1 within V2 (dot product of the gradient vector with the V1 ROI reference vector producing negative values), and point towards V1 within V3 (with the dot product of the vectors producing positive values). Thus, the transition between negative and positive values occurs at the V2/V3 border (in an analogous way to how the field sign changes at the areal border (Sereno et al., 1995). The classifier made use of this transformed information, just as it does for the other modalities, to define areas V2 and V3 in individual subjects. This type of transformation was computed for all four polar-angle related spatial contrasts and the first eccentricity related contrast, producing useful information for the classifier to delineate and identify visuotopically organized areas in individual subjects. The transformation was also computed on the upper vs lower vertical meridian contrast after the absolute value is taken so that areas V2, V3, and V4 have the same sign throughout their extents (otherwise for this contrast in V2 for example, the superior portion would have a positive sign and the inferior portion would have a negative sign). This is because the gradient vectors always point towards the higher number (e.g. they will point away from a local minimum).

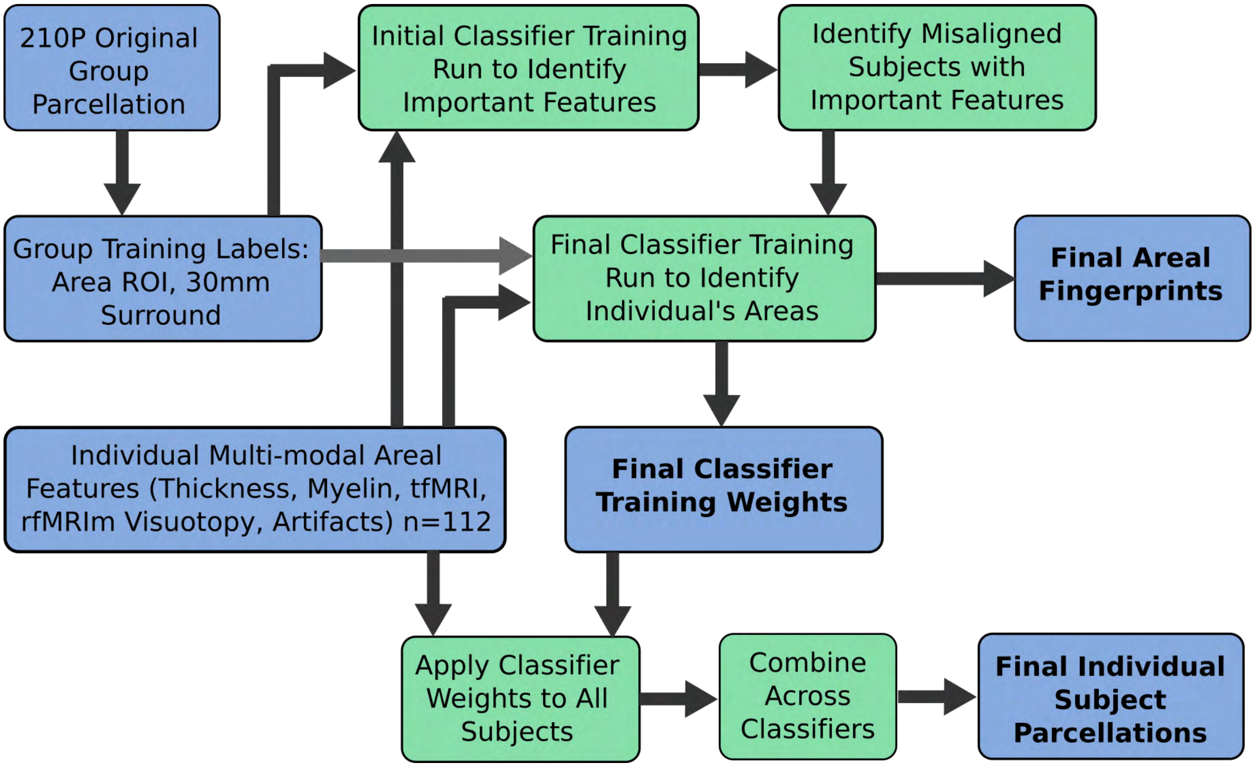


Figure 11 shows the procedure used to go from the group parcellation and individual subject multi-modal areal features to the important outputs produced by the areal classifier and listed in **BOLD** font. The individual subject multi-modal areal features are concatenated (n=112) and used together with the group area ROI and 30 mm surround as training labels for the initial classifier training. This abbreviated training is used to identify the important areal features, which are then used to detect the subjects that are likely to have a misaligned cortical area. These subjects are excluded from the final classifier training (which also uses the group training labels) so that they do not confuse the classifier (as they will be outliers for which the group training labels will not be valid). This final training produces the final classifier training weights, which can be used in the 210V group or another study. These weights are used to simply apply the classifier to the individual subject multi-modal features, producing the final areal fingerprints (what the classifier has determined are the distinctive features of each cortical area relative to it neighbors). Then the classifiers are combined across areas to produce the final individual subject parcellations.

6.6 Implementation of the Machine Learning Classifier

Figure 11 shows an overview of the implementation of the areal classifier. 112 feature maps representing all of the parcellation modalities in each individual were provided to the classifier for training (1 thickness map corrected for curvature, 1 myelin map, 1 surface curvature map, 1 mean tfMRI activation map, 20 tfMRI component contrast maps, 77 surface rfMRI RSN maps from a d=137 ICA, 1 map of vein effects, 1 fMRI dropout map, 1 map of insular surface artifacts, 1 map of the resting state timeseries standard deviation, 1 map of the mean grey timecourse beta, and 5 topographic regression dot product maps). Classifier training used these maps from the 210P (training set) and 29T (test set) subject groups together with areal distortion maps for each subject. The areal features were demeaned and normalized so that each modality had a similar spatial standard deviation (to initially give each modality equal weight in proportion to the number of features it contains). An initial "quick" 1,000-iteration classifier training was done after excluding areal distortion outliers, by using all subjects that had a mean areal distortion within the area being classified of less than 2-fold expansion or contraction (in practice few subjects in few areas were excluded by this criterion because of the improvements in the MSMAll registration used in this study).

After the initial quick classifier training, the output weight matrices were algorithmically examined to determine which features drove the areal classification of each area. Specifically, for each area class, its partial derivatives with respect to the input data values were computed and evaluated using the feature data. This measure was then multiplied by the features' gradient magnitudes to generate a classifier 'sensitivity' metric. Spatially, these sensitivities were highest in absolute value along the parcel boundaries. Features whose sensitivities were higher in absolute value contribute more to the areal classification than those whose sensitivities had lower absolute value. We used this sensitivity information to automatically select features that were particularly predictive of areal misalignment in individual subjects for a given area (because the features most strongly used in classification should also be most likely to differ if an individual cortical area is misaligned with the group average cortical areal definition—the training label). If in a particular individual for the region near a particular area deviated substantially from the group values in the region, the subject was excluded from training.

After potentially misaligned subjects were excluded from the training and test sets, the classifier was trained using a full 4,000 iterations (the learning rate nonlinearly increased with iteration number). The classifier learning rate was automatically adjusted down if the classifier was stopping too early (i.e. it detected it was overfitting itself to the 210P training dataset, insofar as performance on the 29T test dataset began to drop) and the number of iterations was extended (since the training would take longer with a lower learning rate). The overall loop (that was able to adjust the learning rate) was terminated when the most recent training run was the best and the classifier had finished the requested number of iterations, or the classifier AUC (Area Under Curve) was greater than 0.98. If the most recent run did worse than a previous run with a higher learning rate, the process was also stopped and the previous run was used. The goal was to either get an AUC of >0.98 or to complete the full number of training iterations without prematurely stopping. Classifier performance was evaluated using AUC and MSE (Mean Squared Error) as standard metrics. Once satisfactory classification had been achieved, the trained classifier was applied to every subjects' data (including the misaligned subjects). Classifier performance was also evaluated by confirming that in subjects with known misaligned areas the classifier was correctly classifying these areas (visually based on multi-modal information, see for example Supplementary Figures 9 and 10 in the SRD).

6.7 Combination Across Areal Classifiers

Classifier output values ranged from 0 to 1 for each class (class one=area, class two=neighborhood), indicating the likelihood that a given vertex in a given individual is part of either the cortical area or the neighborhood. The areal probability maps (class one) for each area in a subject were combined using a find the biggest approach. Regions where no areal probability exceeded 50% were also marked as areal "holes" (when present consistently across subjects, these represented candidates for draft parcellation revision, see above #5.1). There were several instances (e.g. among the VMV1-3 areas, particularly VMV1) where this occurred. We think the hole arises when the classifier encounters two distinct areal fingerprints within a larger draft parcel. It likely only learns the larger of these, so the smaller one is neglected. Because it is not represented by a high areal probability for any classifier, the smaller region becomes an "area hole." The issue is resolved if an appropriate subparcellation is done, as each areal classifier learns the areal fingerprint of its subparcellated area, and both will now be detected with high probability.

This initial combined parcellation was regularized (i) to remove very small patches (less than 25 mm^2); (ii) to join pieces of areas that were broken by only a few vertices of another area (as long as the piece being joined was proportionally larger than any piece being split off from another area as a result of the joining operation); and (iii) to make sure that if any multiple pieces remained, they were at least 0.33x the size of the largest piece and within 30 mm of its nearest edge. Any remaining holes in the parcellation were filled by nearest neighbor dilation. After careful consideration of the multi-modal evidence, we did not enforce a strict spatial contiguity constraint on the parcels in individual subjects. We found strong multi-modal evidence of parcels that were split into multiple pieces in some atypical individuals, although in the typical subject (i.e. group average) they were joined (see SRD #1.3-1.4). The resulting initial individual subject parcellations were checked relative to the group parcellations to see if the areas found were within 0.33x to 3x the size of the group area. If this was the case, the area was considered to have been detected in the individual subject.

6.8 Classification of Individual Subject Cortical Areas Based on Their Multi-modal Areal Fingerprints

The final classifier training weights were applied to the 210P and 210V subject groups to make final individual subject cortical parcellations. Additionally, the partial derivatives were tracked through the classifier application and multiplied by the feature map gradient magnitudes to produce the final areal fingerprints of each cortical area—that is, the information the classifier learned that uniquely distinguishes the cortical area from surrounding cortex (see above #6.1, Supplementary Figure 12 in the SRD).

7. Parcellating Data and Validating the Parcellation

7.1 Creation of Parcellated Datasets of Multiple Modalities

Parcellated datasets offer a tremendous reduction in data size and complexity, together with performing a neurobiologically valid form of spatial integration that does not blur across tissue types or areal boundaries as happens with conventional smoothing. They increase statistical sensitivity and power both by reducing unstructured noise and multiple comparisons. For most whole brain analyses, parcellated datasets should be easier to work with and more likely to show interesting effects (e.g. correlations with behavior, genetics, etc.). Also, they provide a level of coarseness appropriate for integration with non-MRI modalities such as magnetoencephalography (MEG; (Larson-Prior et al., 2013). Thus, we created parcellated datasets of each modality used in this study, except for the within area topographic gradients, which are an example of a phenomenon that is inherently more appropriate to study using dense, grayordinate-wise data.

Myelin maps and thickness maps were simply averaged across each parcel. For the fMRI modalities, cross-parcel averaging substantially reduces the unstructured noise present in the timeseries data. tfMRI timeseries data were averaged within parcels and then the tfMRI analysis pipeline (described above) was run (with autocorrelation smoothing turned off). This generated parcellated task contrast maps, which are advantageous for assessing the statistical significance of activations without resorting to conventional statistical approaches for dealing with voxelwise multiple comparisons that are likely less sensitive (e.g. Threshold-Free Cluster Enhancement, (Smith and Nichols, 2009)), and, in the case of Gaussian Random Field theory, use geometrical models that are poorly matched to areal shapes in the convoluted cortex. One still needs to do multiple comparison correction; however, because of the reduced number of tests, there is more power. rfMRI timeseries data were averaged within parcels. Parcellated functional connectomes were computed using full and partial correlation (by inverting the covariance matrix) using FSLNets ((Smith et al., 2013b) http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLNets). These connectomes were converted to Z-scores using the Fisher transform and an empirical correction for temporal autocorrelation (see FSLNets).

To produce group average parcellated datasets, myelin and thickness parcellated maps were simply averaged. tfMRI contrast maps were combined across subjects by processing the data through the group tfMRI analysis pipeline using a mixed effects analysis (FSL's FLAME). rfMRI FC parcellated connectomes were also combined across subjects using a mixed effects analysis (a one-sample t-test performed across subjects on each cell to produce group mixed effects Z statistic maps). Thus, both individual subject and group average parcellated datasets were generated for myelin, thickness, tfMRI contrasts, and rfMRI FC full and unregularized partial correlation connectomes.

7.2 Validation of the Multi-modal Parcellation with an Independent HCP Dataset

All parcellation and areal classifier training analyses described in preceding sections were performed on the 210P parcellation subjects (plus 29T classifier test subjects). The 210V validation subjects were set aside as an independent cross-validation dataset. Given the large number of subjects in each group, we expected that aggregate results would be very similar between the two groups (as was shown in the Supplementary Figures 1-5 in the SRD and #1.1 for the reproducibility of the dense features used for parcellation). To produce individual-subject parcellations in the 210V group, the multi-modal individual subject features were generated and the trained classifiers (i.e. two weight matrices and two nonlinear functions for each cortical area) were applied to each individual subject's multi-modal areal features to define and identify their cortical areas. The classifier outputs were combined as described above. Two approaches were then used to cross-validate the parcellation.

1) Parcellated datasets of 210V individuals were generated using the original semi-automated parcellation from the 210P dataset (applied as an atlas parcellation, rather than using the areal classifier) for myelin maps, cortical thickness maps, tfMRI contrast maps, and an rfMRI full correlation parcellated connectome. These data were concatenated into an areas X features X subjects matrix. t-tests across subjects were done between each pair of areas that shared a border in the original 210P parcellation for each feature to determine whether the feature was robustly and statistically significantly different between the two neighboring areas. A conservative Bonferroni threshold of $p<9*10^{-8}$ was imposed on the data (# of area pairs across both hemispheres (1050) X number of features (266) X number of tails (2) * 0.05). 63% of all possible feature/border combinations were significant, including many features that the classifier regarded as unimportant for defining the relevant area. We therefore opted to additionally threshold based on effect size (Cohen's d>1). We grouped features into 4 independent categories: myelin, thickness, tfMRI contrasts (though technically the seven tasks are independent) and the rfMRI full correlation parcellated connectome. We then binned areal pairs according to how many feature categories in which they had a robust and statistically significant difference across their mutual border. The parcellated connectome's diagonal was ignored for this test (else areas would always be different from their neighbors because of a connectivity difference with the diagonal). The vast majority of areal pairs had robust and statistically significant effect sizes in more than one independent feature category in the independent cross-validation dataset (see Main Results, SRD #1.2).

2) Probabilistic maps of each cortical area within each hemisphere were generated by averaging the areal definitions across subjects within the 210P and 210V groups separately, and a maximum probability map MPM map was calculated for each group based on the probabilistic maps. The group MPM maps were converted into binary ROIs (180 of them, one for each area), which were concatenated into a single vector. The 210P MPM vector and the 210V MPM vectors were correlated (also the 210P MPM was correlated with the original semi-automated multimodal parcellation). Because some readers found the Dice coefficient more intuitive than correlation for this application we also computed it for each parcellation reproducibility test. The Dice coefficent is $(2 * A \cap B)/(A + B)$.

3) The individual subject parcellations of the 27 individuals who were scanned and analyzed with the whole protocol on two separate occasions were also correlated and Diced as in 2) above.

References

Abdollahi, R.O., Kolster, H., Glasser, M.F., Robinson, E.C., Coalson, T.S., Dierker, D., Jenkinson, M., Van Essen, D.C., Orban, G.A., 2014. Correspondences between retinotopic areas and myelin maps in human visual cortex. Neuroimage 99, 509-524.

Anticevic, A., Dierker, D.L., Gillespie, S.K., Repovs, G., Csernansky, J.G., Van Essen, D.C., Barch, D.M., 2008. Comparing surface-based and volume-based analyses of functional neuroimaging data in patients with schizophrenia. Neuroimage 41, 835-848.

Barch, D.M., Burgess, G.C., Harms, M.P., Petersen, S.E., Schlaggar, B.L., Corbetta, M., Glasser, M.F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J.M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A.Z., Van Essen, D.C., Consortium, W.U.-M.H., 2013. Function in the human connectome: task-fMRI and individual differences in behavior. Neuroimage 80, 169-189.

Beckmann, C.F., DeLuca, M., Devlin, J.T., Smith, S.M., 2005. Investigations into resting-state connectivity using independent component analysis. Philosophical Transactions of the Royal Society of London B: Biological Sciences 360, 1001-1013.

Beckmann, C.F., Smith, S.M., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. Medical Imaging, IEEE Transactions on 23, 137-152.

Caspers, S., Eickhoff, S.B., Zilles, K., Amunts, K., 2013. Microstructural grey matter parcellation and its relevance for connectome analyses. Neuroimage 80, 18-26.

Cohen, A.L., Fair, D.A., Dosenbach, N.U., Miezin, F.M., Dierker, D., Van Essen, D.C., Schlaggar, B.L., Petersen, S.E., 2008. Defining functional areas in individual human brains using resting functional connectivity MRI. Neuroimage 41, 45-57.

Cole, M.W., Bassett, D.S., Power, J.D., Braver, T.S., Petersen, S.E., 2014. Intrinsic and task-evoked network architectures of the human brain. Neuron 83, 238-251.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31, 968-980.

Drury, H.A., Essen, D., Corbetta, M., Snyder, A., 1998. Surface based analyses of the human cerebral cortex. Brain Warping, 337.

Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25, 1325-1335.

Filippini, N., MacIntosh, B.J., Hough, M.G., Goodwin, G.M., Frisoni, G.B., Smith, S.M., Matthews, P.M., Beckmann, C.F., Mackay, C.E., 2009. Distinct patterns of brain activity in young carriers of the APOE-epsilon4 allele. Proc Natl Acad Sci U S A 106, 7209-7214.

Fischl, B., 2012. FreeSurfer. Neuroimage 62, 774-781.

Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B.T., Mohlberg, H., Amunts, K., Zilles, K., 2008. Cortical folding patterns and predicting cytoarchitecture. Cereb Cortex 18, 1973-1980.

Fischl, B., Sereno, M.I., Dale, A.M., 1999a. Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195-207.

Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M., 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. Human brain mapping 8, 272-284.

Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., Raichle, M.E., 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A 102, 9673-9678.

Frost, M.A., Goebel, R., 2012. Measuring structural-functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. Neuroimage 59, 1369-1381.

Glasser, M.F., Goyal, M.S., Preuss, T.M., Raichle, M.E., Van Essen, D.C., 2014a. Trends and properties of human cerebral cortex: correlations with cortical myelin content. Neuroimage 93 Pt 2, 165-175.

Glasser, M.F., Robinson, E.C.C., T.S. Smith, S.M. , Jenkinson, M. Van Essen, D.C. , 2014b. Retinotopic Organization of Visual Cortex Revealed by Resting State Functional Connectivity. Organization for Human Brain Mapping, Hamburg, Germany.

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C., Jenkinson, M., Consortium, W.U.-M.H., 2013. The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage 80, 105-124.

Glasser, M.F., Van Essen, D.C., 2011. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. J Neurosci 31, 11597-11616.

Glocker, B., Heibel, T.H., Navab, N., Kohli, P., Rother, C., 2010. Triangleflow: Optical flow with triangulation-based higher-order likelihoods. Computer Vision–ECCV 2010. Springer, pp. 272-285.

Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48, 63-72.

Griffanti, L., Salimi-Khorshidi, G., Beckmann, C.F., Auerbach, E.J., Douaud, G., Sexton, C.E., Zsoldos, E., Ebmeier, K.P., Filippini, N., Mackay, C.E., 2014. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. Neuroimage 95, 232-247.

Hacker, C.D., Laumann, T.O., Szrama, N.P., Baldassarre, A., Snyder, A.Z., Leuthardt, E.C., Corbetta, M., 2013. Resting state network estimation in individual subjects. Neuroimage 82, 616-633.

Hart, P.E., Nilsson, N.J., Raphael, B., 1968. A formal basis for the heuristic determination of minimum cost paths. Systems Science and Cybernetics, IEEE Transactions on 4, 100-107.

Hodge, M.R., Horton, W., Brown, T., Herrick, R., Olsen, T., Hileman, M.E., McKay, M., Archie, K.A., Cler, E., Harms, M.P., 2015. ConnectomeDB—sharing human brain connectivity data. Neuroimage.

Ishikawa, H., 2009. Higher-order clique reduction in binary graph cut. Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, pp. 2993-3000.

Ishikawa, H., 2014. Higher-order clique reduction without auxiliary variables. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1362-1369.

Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. Fsl. Neuroimage 62, 782-790.

Larson-Prior, L.J., Oostenveld, R., Della Penna, S., Michalareas, G., Prior, F., Babajani-Feremi, A., Schoffelen, J.M., Marzetti, L., de Pasquale, F., Di Pompeo, F., Stout, J., Woolrich, M., Luo, Q., Bucholz, R., Fries, P., Pizzella, V., Romani, G.L., Corbetta, M., Snyder, A.Z., Consortium, W.U.-M.H., 2013. Adding dynamics to the Human Connectome Project with MEG. Neuroimage 80, 190-201.

Marcus, D.S., Harms, M.P., Snyder, A.Z., Jenkinson, M., Wilson, J.A., Glasser, M.F., Barch, D.M., Archie, K.A., Burgess, G.C., Ramaratnam, M., Hodge, M., Horton, W., Herrick, R., Olsen, T., McKay, M., House, M., Hileman, M., Reid, E., Harwell, J., Coalson, T., Schindler, J., Elam, J.S., Curtiss, S.W., Van Essen, D.C., Consortium, W.U.-M.H., 2013. Human Connectome Project informatics: quality control, database services, and data visualization. Neuroimage 80, 202-219.

Ringel, G., Youngs, J., 1968. Solution of the Heawood map-coloring problem. Proc Natl Acad Sci U S A 60, 438.

Robinson, E.C., Jbabdi, S., Glasser, M.F., Andersson, J., Burgess, G.C., Harms, M.P., Smith, S.M., Van Essen, D.C., Jenkinson, M., 2014. MSM: a new flexible framework for Multimodal Surface Matching. Neuroimage 100, 414-426.

Salimi-Khorshidi, G., Douaud, G., Beckmann, C.F., Glasser, M.F., Griffanti, L., Smith, S.M., 2014. Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. Neuroimage 90, 449-468.

Schleicher, A., Amunts, K., Geyer, S., Morosan, P., Zilles, K., 1999. Observer-independent method for microstructural parcellation of cerebral cortex: a quantitative approach to cytoarchitectonics. Neuroimage 9, 165-177.

Schleicher, A., Morosan, P., Amunts, K., Zilles, K., 2009. Quantitative architectural analysis: a new approach to cortical mapping. J Autism Dev Disord 39, 1568-1581.

Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268, 889-893.

Sereno, M.I., McDonald, C.T., Allman, J.M., 1994. Analysis of retinotopic maps in extrastriate cortex. Cereb Cortex 4, 601-620.

Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., Kelly, M., Laumann, T., Miller, K.L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A.Z., Vu, A.T., Woolrich, M.W., Xu, J., Yacoub, E., Ugurbil, K., Van Essen, D.C., Glasser, M.F., Consortium, W.U.-M.H., 2013a. Resting-state fMRI in the Human Connectome Project. Neuroimage 80, 144-168.

Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F., 2009. Correspondence of the brain's functional architecture during activation and rest. Proc Natl Acad Sci U S A 106, 13040-13045.

Smith, S.M., Hyvarinen, A., Varoquaux, G., Miller, K.L., Beckmann, C.F., 2014. Group-PCA for very large fMRI datasets. Neuroimage 101, 738-749.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 Suppl 1, S208-219.

Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44, 83-98.

Smith, S.M., Vidaurre, D., Beckmann, C.F., Glasser, M.F., Jenkinson, M., Miller, K.L., Nichols, T.E., Robinson, E.C., Salimi-Khorshidi, G., Woolrich, M.W., Barch, D.M., Ugurbil, K., Van Essen, D.C., 2013b. Functional connectomics from resting-state fMRI. Trends Cogn Sci 17, 666-682.

Tucholka, A., Fritsch, V., Poline, J.B., Thirion, B., 2012. An empirical comparison of surface-based and volume-based group studies in neuroimaging. Neuroimage 63, 1443-1453.

Ugurbil, K., Xu, J., Auerbach, E.J., Moeller, S., Vu, A., Duarte-Carvajalino, J.M., Lenglet, C., Strupp, J., Sapiro, G., De Martino, F., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D.a., Smith, S.M., Miller, K.L., Sotiropoulos, S.N., Jbabdi, S., Andersson, J.L., Behrens, T.E.J., Glasser, M.F., Van Essen, D.C., Yacoub, E., 2013. Pushing spatial and temporal resolution for functional and diffusuion MRI in the Human Connectome Project. Neuroimage.

Van Essen, D.C., 2005. A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. Neuroimage 28, 635-662.

Van Essen, D.C., Glasser, M.F., Dierker, D.L., Harwell, J., Coalson, T., 2012. Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. Cereb Cortex 22, 2241-2262.

Van Essen, D.C., Smith, J., Glasser, M.F., Elam, J., Donahue, C.J., Dierker, D.L., Reid, E.K., Coalson, T., Harwell, J., 2016. The Brain Analysis Library of Spatial maps and Atlases (BALSA) Database. Neuroimage.

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.U.-M.H., 2013. The WU-Minn Human Connectome Project: an overview. Neuroimage 80, 62-79.

Wishart, J., 1928. The generalised product moment distribution in samples from a normal multivariate population. Biometrika, 32-52.

Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zollei, L., Polimeni, J.R., Fischl, B., Liu, H., Buckner, R.L., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol 106, 1125-1165.

Zilles, K., Amunts, K., 2010. Centenary of Brodmann's map--conception and fate. Nat Rev Neurosci 11, 139-145.