Comparing spatially-constrained null models for parcellated brain maps

Ross D. Markello[1,*] Bratislav Misic[1,*]
[1]*McConnell Brain Imaging Centre, Montréal Neurological Institute, McGill University, Montréal, Canada*
[*]*Correspondence to: ross.markello@mail.mcgill.ca or bratislav.misic@mcgill.ca*

Technological and data sharing advances have led to a proliferation of high-resolution structural and functional maps of the brain. Modern neuroimaging research increasingly depends on identifying correspondences between the topographies of these maps; however, most standard methods for statistical inference fail to account for their spatial properties. Recently, multiple methods have been developed to generate null distributions that preserve the spatial autocorrelation of brain maps and yield more accurate statistical estimates. Here, we comprehensively assess the performance of ten such published null frameworks in controlling the family-wise error rate in statistical analyses of parcellated neuroimaging data. We apply each framework on two prototypical analyses: (1) testing the correspondence between brain maps (e.g., correlating two activation maps) and (2) testing the spatial distribution of a feature within a partition (e.g., quantifying the specificity of an activation map within an intrinsic functional network). In agreement with previous reports, we find that naive null models that do not preserve spatial autocorrelation consistently yield unrealistically liberal statistical estimates. Performance of spatially-constrained null models depended on research context; model performance was generally consistent when testing correspondences between brain maps, but considerably more variable when testing partition specificity. Throughout these analyses, we observe minimal impact of parcellation and parcel resolution on null model performance. Altogether, our results highlight the need for continued development and standardization of statistically-rigorous methods for comparing brain maps.

Keywords: null models | spin test | brain parcellations | spatial autocorrelation | significance testing

INTRODUCTION

The brain is organized as a series of nested and increasingly multi-functional neural circuits. The connections and interactions among these circuits ultimately manifest as unique topographic distributions of multiple structural and functional properties. Recent advances in imaging, tracing, and recording technologies (Insel et al. 2013), together with global data sharing initiatives (Casey et al. 2018, Poldrack et al. 2013, Sudlow et al. 2015, Van Essen et al. 2013), have resulted in the generation of high-resolution maps of many of these properties, including gene expression (Akbarian et al. 2015, Hawrylycz et al. 2012), cytology (Scholtens et al. 2018, von Economo and Koskinas 1925), receptor densities (Beliveau et al. 2017, Norgaard et al. 2020, Zilles and Amunts 2009, Zilles et al. 2004), intracortical myelin (Burt et al. 2018, Whitaker et al. 2016), and functional organization (Bellec et al. 2010, Damoiseaux et al. 2006, Margulies et al. 2016, Murray et al. 2014, Shafiei et al. 2020, Yeo et al. 2011).

Increasingly, modern scientific discovery in neuroimaging research depends on identifying correspondences between the topographies of brain maps (Baum et al. 2020, Demirtaş et al. 2019, Gao et al. 2020, Hansen et al. 2020, Shafiei et al. 2020, Vázquez-Rodríguez et al. 2019, Wang et al. 2019); however, standard methods for statistical inference fall short when making such comparisons (Alexander-Bloch et al. 2013, 2018, Burt et al. 2020, Fulcher et al. 2020). Namely, in spatially-embedded systems—like the brain—neighboring data points are not statistically independent, violating the assumptions of many common inferential frameworks. As an example, consider computing a correlation between two brain maps. When using a standard parametric null (i.e., the Student's *t*-distribution), the spatial autocorrelation of the maps violates the inherent requirement that the model errors are independent and identically distributed (*i.i.d.*). When using a standard non-parametric null (i.e., random permutations of one of the feature maps), the spatial autocorrelation violates the requirement of exchangeability. In both instances, the calculated p-value will be inflated, yielding increased family-wise error rates (FWER, or type 1 error rates) across analyses.

In the past several years multiple frameworks have been developed to overcome the shortcomings of standard null models in neuroimaging. These frameworks generate null distributions that preserve the spatial autocorrelation of brain maps, yielding statistical inferences that more accurately reflect the underlying data. Such spatially-constrained null models fall into two broad families: non-parametric spatial permutation models (Alexander-Bloch et al. 2013, 2018) and parameterized data models (Burt et al. 2018, 2020, Vos de Wael et al. 2020). In non-parametric spatial permutation models, the cortical surface is represented as a sphere to which random rotations are applied, generating surface maps with randomized topography but identical spatial autocorrelation. In parameterized data models, spatial autocorrelation is estimated from the empirical brain maps and used to generate surrogate nulls with randomized topography and similar—though not identical—spatial autocorrelation. Since their development, these models have been adapted by several researchers (Baum et al. 2020, Cornblath et al. 2019, Váša et al. 2018, Vázquez-Rodríguez et al. 2019); to our knowledge, there have been at least ten distinct implementations of null frame-

works applied to statistical estimates of brain maps.

The earliest implementations of these null models, proposed in Alexander-Bloch et al. (2013) and later formalized in Alexander-Bloch et al. (2018), described a non-parametric method by which the cortical surface was subjected to spatial rotations. The principal challenge of implementing this method is that the medial wall—for which most brain maps contain no data—can be rotated into the cortical surface. This is an important consideration for parcellated brain maps in particular, because the loss of data caused by this medial wall rotation can bias results. To address this problem, researchers have opted to either discard the missing data (Baum et al. 2020, Cornblath et al. 2019), assign the nearest data to missing parcels (Vázquez-Rodríguez et al. 2019), or ignore the medial wall entirely (Váša et al. 2018). Other groups have devised alternative methods that do not rely on spatial rotations but use generative models instead. These parameterized frameworks vary in their conceptualization and implementation of the data-generating process, ranging from a spatial lag model (Burt et al. 2018) to spectral randomization (Vos de Wael et al. 2020, Wagner and Dray 2015) to variogram matching (Burt et al. 2020). How these different models perform when applied to the same experimental questions and datasets remains unclear.

Here, we comprehensively compare how ten published null frameworks control the FWER in statistical analyses of parcellated neuroimaging data. Using multiple open-access datasets (Neurosynth [Yarkoni et al. 2011]; Human Connectome Project [Van Essen et al. 2013]), we apply each framework on two prototypical analyses: (1) assessing the correspondence between brain maps (e.g., correlating two activation maps) and (2) assessing the spatial distribution of a feature within a partition (e.g., quantifying the enrichment of T1w/T2w ratios in resting state networks). In both analyses we systematically examine the impact of parcellation—which differentially modifies the spatial structure of the underlying data—on the performance of the null frameworks.

METHODS AND MATERIALS

Code and data availability

All code used for data processing, analysis, and figure generation is available on GitHub (https://github.com/netneurolab/markello_spatialnulls) and directly relies on the following open-source Python packages: BrainSMASH (Burt et al. 2020), BrainSpace (Vos de Wael et al. 2020), IPython (Pérez and Granger 2007), Jupyter (Kluyver et al. 2016), Matplotlib (Hunter 2007), NeuroSynth (Yarkoni et al. 2011), NiBabel (Brett et al. 2020), NumPy (Oliphant 2006, Van Der Walt et al. 2011), Pandas (McKinney 2010), PySurfer (Waskom et al. 2020b), Scikit-learn (Pedregosa et al. 2011), SciPy (Virtanen et al. 2020), and Seaborn (Waskom et al. 2020a). Additional software used in the reported analyses includes FreeSurfer (v6.0.0; http://surfer.nmr.mgh.harvard.edu/; Fischl et al. 1999) and the Connectome Workbench (v1.4.2; https://www.humanconnectome.org/software/connectome-workbench; Marcus et al. 2011).

Data

NeuroSynth association maps

To replicate the analyses described in Alexander-Bloch et al. (2018) we downloaded "association" maps from NeuroSynth (Yarkoni et al. 2011) for $t = 123$ terms overlapping with those defined in the Cognitive Atlas (https://www.cognitiveatlas.org/concepts; Poldrack 2011, Poldrack and Yarkoni 2016; refer to Table S1 for a full list of terms). This resulted in 123 volumetric statistical maps which were then inflated to a mid-gray projection of FreeSurfer's fsaverage5 surface using nearest neighbor interpolation.

Human Connectome Project

Group-averaged T1w/T2w (a proxy for intracortical myelin) data were downloaded from the S1200 release of the Human Connectome Project (HCP; Van Essen et al. 2013). To ensure consistency with the reported NeuroSynth analyses, data were resampled to FreeSurfer's fsaverage5 surface following instructions on the Human Connectome Project Wikipedia page (https://wiki.humanconnectome.org/display/PublicData/HCP+Users+FAQ).

Brain parcellations

In order to examine the impact of parcellations on the tested null models we used two multi-scale resolution atlases generated from structural (Cammoun et al. 2012) and functional (Schaefer et al. 2018) connectivity data. The former, referred to throughout the text as the "Cammoun atlas", has five resolutions ranging from 68 to 1000 parcels (68, 114, 219, 448, and 1000 parcels), generated by dividing the Desikan-Killiany atlas (Desikan et al. 2006) into equally-sized sub-units based on group-level diffusion weighted imaging data. The latter, referred to throughout the text as the "Schaefer atlas", has ten resolutions ranging from 100 to 1000 parcels in steps of 100 parcels (i.e., 100, 200, 300, and so on), generated via a gradient-weighted Markov Random Field model from resting state fMRI data (Schaefer et al. 2018).

NeuroSynth activation maps and the HCP T1w/T2w brain map were parcellated using all resolutions of the Schaefer and Cammoun atlases. Values for vertices lying on the medial wall of the surface mesh were not considered.

TABLE 1. **Null model frameworks |** Overview of the null frameworks and the implementations used in the reported analyses. The *Description* column indicates the primary point of methodological divergence for each framework. The *Duplications* column indicates whether parcel values can be duplicated in the given framework. The *Citation* column indicates the first appearance of the described implementation in the neuroimaging literature, if known. Code and/or references to code for implementing all frameworks is available at https://github.com/netneurolab/markello_spatialnulls.

Naive models			
Method	*Description*	*Duplications*	*Citation*
Parametric	Uses standard "normal" distributions (e.g., Student's *t*-distribution)	No	
Non-parametric	Reorders parcels randomly, not taking into account their inherent spatial structure	No	
Spatial permutation models			
Method	*Description*	*Duplications*	*Citation*
Vázquez-Rodríguez	Assigns parcel centroid to closest rotated centroid based on minimum Euclidean distance	Yes	Vázquez-Rodríguez et al. (2019)
Váša	Uses an iterative procedure to uniquely assign parcels based on distance to the rotated parcels	No	Váša et al. (2018)
Hungarian	Uses the Hungarian algorithm to match original and rotated parcel centroids	No	
Baum	Projects parcel data to vertices, rotates, and takes the mode of vertices in each parcel	Yes	Baum et al. (2020)
Cornblath	Projects parcel data to vertices, rotates, and takes the mean of vertices in each parcel	Yes	Cornblath et al. (2019)
Parameterized data models			
Method	*Description*	*Duplications*	*Citation*
Burt-2018	Uses a distance-dependent spatially autoregressive model to generate nulls	No	Burt et al. (2018)
Burt-2020	Uses a variogram model to estimate spatial autocorrelation to generate nulls	No	Burt et al. (2020)
Moran	Uses a distance-dependent eigendecomposition to generate nulls	No	Vos de Wael et al. (2020)

Network partitions

To investigate the partition specificity of brain maps we applied two common network definitions: the seven functional networks defined in Yeo et al. (2011) and the seven cytoarchitectonic classes proposed by von Economo and Koskinas (1925).

For the Cammoun atlas we used the fsaverage5 vertex-level Yeo network assignments provided by FreeSurfer. To derive parcel-wise network assignments from this vertex representation we applied a winner-take-all approach; that is, the mode of the vertices in each parcel was used to select the representative network. For the Schaefer atlas we used the parcel assignments for the Yeo networks provided by the original authors. To derive von Economo–Koskinas classes for parcels in both the Cammoun and Schaefer atlases we applied the classifiers provided by Scholtens et al. (2018) to the fsaverage5 surface, yielding vertex-level assignments; these assignments were then used with the winner-take-all approach to generate parcel-level network assignments.

Null model frameworks

Here, we briefly describe null model frameworks as initially proposed in the literature, and then explain the specific details of the implementations used in the current report. For a summary overview of the implementations please refer to Table 1.

Naive models

Commonly used in the neuroimaging literature prior to the development of spatially-constrained frameworks, so-called naive models do not take into account the spatial structure of the data when constructing null distributions, commonly resulting in inflated family-wise error rates. We test two such models: a *parametric* and a *non-parametric* method.

Naive parametric method. Although the exact implementation of the parametric method varies based on the statistical test employed, all implementations share a reliance on standard null distributions. For example, when

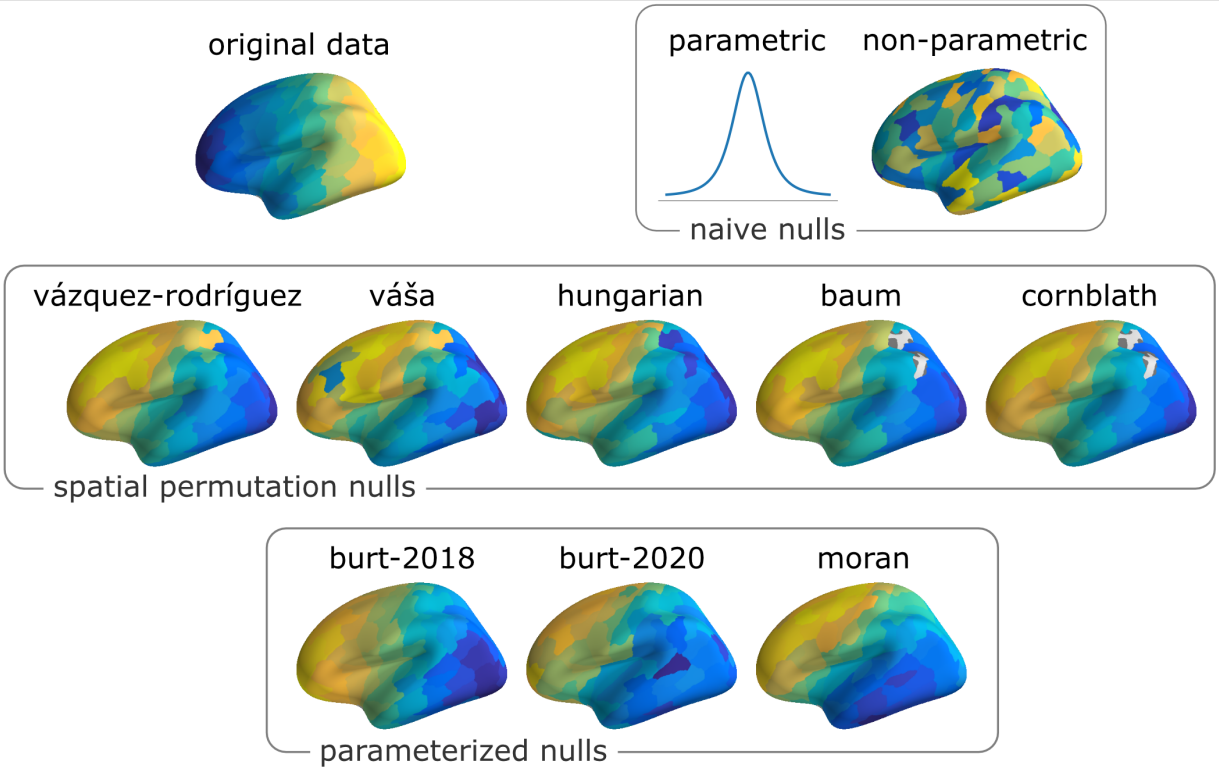

Figure 1. **Null model framework examples |** Examples of the null frameworks applied to toy data representing an anterior-posterior gradient (original data, top left) using the Cammoun 219-parcel atlas. All spatial permutation null examples were generated using the same rotation matrix to highlight variations between the different implementations. Parameterized null examples were chosen so as to maximize similarity to the spatial permutation nulls for visualization.

examining correlation values, the parametric method relies on the Student's *t*-distribution; when examining z-statistics, this method uses the standard normal distribution.

Naive non-parametric method. The naive non-parametric approach uses a random permutation (i.e., reshuffling) of the data to construct a null distribution, destroying its inherent spatial structure. Each parcel is uniquely reassigned the value of another parcel for every permutation.

Spatial permutation models

First published in Alexander-Bloch et al. (2013) and later formalized in Alexander-Bloch et al. (2018), spatial permutation models generate spatially-constrained null distributions by applying random rotations to spherical projections of the brain. A rotation matrix ($\mathbf{R}$) is applied to the three-dimensional coordinates of the brain ($\mathbf{V}$) to generate a set of rotated coordinates ($\mathbf{V_{rot}} = \mathbf{VR}$); the permutation is constructed by replacing the original values at each coordinate with those of the closest rotated coordinate. Rotations are generated independently for one hemisphere and then mirrored across the anterior-posterior axis for the other.

At the time of writing, we are aware of four published adaptations of this spatial permutation framework that have been applied to parcellated brain data: Vázquez-Rodríguez et al. (2019), Váša et al. (2018), Baum et al. (2020), and Cornblath et al. (2019). We test all four of these adaptations, in addition to one other method (*Hungarian*) that we posit is a natural extension thereof.

Vázquez-Rodríguez method. The *Vázquez-Rodríguez* method, which serves as a direct adaptation of the original framework from Alexander-Bloch et al. (2018) but applied to parcellated brain data, was first used in Vázquez-Rodríguez et al. (2019). In this adaptation, vertex coordinates are replaced with those of parcel centroids. That is, a rotation is applied to the coordinates for the center-of-mass of each parcel, and parcels are reassigned the value of the closest rotated parcel (i.e., that with the minimum Euclidean distance). As in the original formulation this method permits the duplicate reassignment of parcel values for every rotation, such that some parcel values may not be present in a given rotation and others may appear more than once.

Váša method. The first known application of spatial permutations to parcellated data, the *Váša* method (Váša et al. 2018) attempted to resolve the primary issue of the Alexander-Bloch method: duplicate reassignments.

That is, this method was created so as to yield a perfect permutation of the original data for every rotation. As with the *Vázquez-Rodríguez* method, parcel centroids are used instead of vertex coordinates. In order to avoid duplicate reassignments, parcels are iteratively assigned by (1) finding the closest rotated parcel to each original parcel, and (2) assigning the most distant pair of parcels. This two-step process is then repeated for all remaining unassigned parcels until each has been reassigned.

Hungarian method. Similar to the *Váša* method, the *Hungarian* method attempts to uniquely reassign each parcel for every rotation. Instead of using an iterative process, however, which can result in globally suboptimal assignments, this method uses the Hungarian algorithm to solve a linear sum assignment problem (Kuhn 1955). This method attempts to uniquely reassign each parcel such that the global reassignment cost is minimized, where cost is quantified as the distance between the original and rotated parcel centroid coordinates.

Baum method. Used initially in Baum et al. (2020), this method projects parcellated brain data to a high-resolution surface mesh, assigning identical values to all the vertices within a given parcel. The projected mesh is subjected to the original spatial permutation reassignment procedure and re-parcellated by taking the modal (i.e., the most common) value of the vertices in each parcel. Notably, this method can result in duplicate assignment of parcel values in each permutation.

Cornblath method. In this method implemented by Cornblath et al. (2019), parcellated data are projected to a high-resolution spherical surface mesh, rotated, and re-parcellated by taking the average (i.e., the arithmetic mean) of the vertices in each parcel. Because the data are re-parcellated the likelihood of duplicated assignments is very low (though not exactly zero); however, the distribution of re-parcellated values will be slightly different than the original data distribution.

Null model implementations The spherical projection of the fsaverage5 cortical mesh from FreeSurfer was used to define coordinates for all of the spatial permutation null models (Fischl et al. 1999). When parcel centroids were required for a model we used the procedure described in Vázquez-Rodríguez et al. (2019), *surface-projection averaging*, which includes: (1) calculating the arithmetic mean of the coordinates of all the vertices within a given parcel, and (2) using the coordinate of the surface vertex closest to this average (where closest is defined as minimizing Euclidean distance) to represent the parcel centroid. Other means of parcel centroid definition yielded similar results (see *Variability in parcel centroid definition*).

Parameterized data models

Distinct from the formulation of spatial permutation models proposed by Alexander-Bloch et al. (2018), parameterized data models do not rely on rotations to generate null distributions. Instead, these models generate surrogate null maps that retain spatial features characteristic of the data from which they are estimated. We test three such models, initially proposed for use in the neuroimaging (Burt et al. 2018, 2020) and ecology (Wagner and Dray 2015) literature.

Burt-2018 method. Described in Burt et al. (2018), this framework uses a spatial autoregressive model of the form $\mathbf{y} = \rho\mathbf{W}\mathbf{y}$ to generate surrogate data. Here, $\mathbf{y}$ refers to a Box-Cox transformed, mean-centered brain feature of interest (i.e., a brain map), $\mathbf{W}$ is a weight matrix (derived from $\mathbf{D}$, a matrix of the distance between brain regions, and d_0, a spatial autocorrelation factor), and ρ is a spatial lag parameter. The parameters ρ and d_0 are derived from the data via a least-squares optimization procedure and their estimates $\hat{\rho}$ and $\hat{d_0}$ are used to generate surrogate brain maps according to $\mathbf{y_{surr}} = (\mathbb{I} - \hat{\rho}\mathbf{W}[\hat{d_0}])^{-1}\mathbf{u}$, where $\mathbf{u} \sim \mathcal{N}(0,1)$ is a vector of random Gaussian noise. Rank-ordered values in the $\mathbf{y_{surr}}$ map are replaced with corresponding values from the original $\mathbf{y}$.

Burt-2020 method. Two years after introducing their spatial autoregressive method, Burt et al. (2020) proposed a novel model to generate surrogate data using variogram estimation. The method operates in two main steps: (1) randomly permute the values in a given brain map, and (2) smooth and re-scale the permuted values to reintroduce spatial autocorrelation characteristic of the original, non-permuted data. Reintroduction of spatial autocorrelation onto the permuted data is achieved via the transformation $|\beta|^{1/2}\mathbf{x}' + |\alpha|^{1/2}\mathbf{z}$, where $\mathbf{x}'$ is the permuted data, $\mathbf{z} \sim \mathcal{N}(0,1)$ is a vector of random Gaussian noise, and α and β are estimated via a least-squares optimization between variograms of the original and permuted data. Rank-ordered values in the surrogate map are replaced with corresponding values from the original brain map.

Moran method. Originally developed in the ecology literature (Dray 2011, Wagner and Dray 2015), Moran spectral randomization (MSR) has only been recently applied to neuroimaging data (Paquola et al. 2020, Royer et al. 2020, Vos de Wael et al. 2020). Similar to the other parameterized data methods, MSR principally relies on a spatially-informed weight matrix $\mathbf{W}$, usually taking the form of an inverse distance matrix between brain regions. However, rather than using $\mathbf{W}$ to estimate parameters via a least-squares approach, MSR uses an eigendecomposition of $\mathbf{W}$ to compute Moran's I, an estimate of spatial auto-correlation (Moran 1950). This estimate is then used to impose a similar spatial structure on random, normally distributed surrogate data.

Null model implementations All parameterized data models require an input distance or weight matrix providing information about the spatial structure of the corresponding brain maps. In the present report we calculated vertex-vertex surface distance matrices on the pial surface of the fsaverage5 cortical mesh from FreeSurfer. Shortest paths on the surface were not allowed to traverse the medial wall (see *Geodesic distances along the*

medial wall). Parcel-parcel distance matrices were calculated for each atlas and resolution by averaging the distance between every vertex in two parcels. This parcel-parcel distance matrix was used without modification for the *Burt-2018* and *Burt-2020* methods. The inverse of the parcel-parcel distance matrix—with the diagonal set to 1—was used as an input for the *Moran* method.

Analytic approaches

Although there exist numerous methods for analyzing neuroimaging data, two common approaches involve (1) examining the correspondence between two (or more) distinct brain maps, and (2) investigating whether the spatial distribution of a given brain map is circumscribed by predefined partitions (e.g., intrinsic networks, cytoarchitectonic classes, etc). We conducted analyses for both of these approaches using two datasets (see *Methods and Materials: Data*), examining the impact of each null framework on the statistical interpretation of the generated results. Null distributions for each framework were constructed from 10,000 permutations, rotations, or surrogates. When applicable, the same rotations were used across frameworks (e.g., for *Vázquez-Rodríguez*, *Váša*, and *Hungarian*).

Testing correspondence between brain maps (NeuroSynth)

Following Alexander-Bloch et al. (2018), NeuroSynth maps were correlated to generate a term $\times$ term (123 $\times$ 123) correlation matrix, indicating the extent to which pairs of cognitive terms share a similar spatial pattern; instead of using vertex-wise values as in the original publication, correlation matrices were generated from parcellated data (see *Methods and Materials: Brain parcellations*). We applied each of the ten null frameworks to examine which of the resulting 7,503 unique correlations were significant. For each framework the NeuroSynth maps were permuted (or rotated / used to generate surrogate data, as appropriate) and re-correlated, yielding a 123 x 123 null correlation matrix for each permutation representing the correlations between the original and null maps. The largest absolute value (excluding the diagonal) of each permuted correlation matrix was retained and stored, generating a null distribution of 10,000 correlations; this procedure provides family-wise control for multiple comparisons (Alexander-Bloch et al. 2018, Westfall and Young 1993). We used this null distribution to estimate the two-tailed *p*-values for the correlations in the original term-by-term matrix, thresholding the matrix at $\alpha = 0.05$.

For the naive parametric null method, we used the Student's *t*-distribution to generate *p*-values for each correlation between cognitive maps. For the parameterized null frameworks where surrogate data depend on the input brain maps, we constructed 10,000 surrogate maps separately for each cognitive term; to maintain consistency we used the same 10,000 random data vectors (e.g., $\mathbf{u}$ in Burt-2018, $\mathbf{z}$ in Burt-2020) for every term.

Testing partition specificity (HCP)

Using parcellated T1w/T2w data (Van Essen et al. 2013), we calculated the average value of all parcels within each of seven intrinsic functional networks (Yeo et al. 2011) and seven cytoarchitectonic classes (Scholtens et al. 2018, von Economo and Koskinas 1925). We then applied each of the ten null frameworks to examine which of these averages were significantly higher or lower than would be otherwise expected. For each framework, the parcel values were permuted (or rotated / used to generate surrogate data, as appropriate) and re-averaged within the partitions, yielding a distribution of 10,000 null values for each network or class. We used these null distributions to estimate the two-tailed *p*-values for the original partition averages at $\alpha = 0.05$.

For the naive parametric null framework we used the Student's *t*-distribution to generate *p*-values, testing the distribution of parcel values for each network against zero, where the overall distribution of parcel values were z-scored prior to segregation into networks.

Null model implementation variability

While the present report is primarily interested in how the different null models compare to one another, we also wanted to investigate the extent to which methodological choices within each model impact their performance. That is, most of the null models require researchers to make certain decisions when implementing them in practice. We investigate two such choices: (1) definition of parcel centroids for the spatial permutation nulls models, and (2) definition of the geodesic distance matrix for parameterized data models.

Variability in parcel centroid definition

We examined the impact of parcel centroid definition on three spatial permutation null models: *Vázquez-Rodríguez*, *Váša*, and *Hungarian*. Parcel centroids were defined using three different techniques operating on the spherical representation of the fsaverage5 surface (Fischl et al. 1999): (1) averaging, (2) surface-projection averaging, and (3) geodesic distance minimization. In *averaging*, the coordinates for all vertices belonging to a parcel were averaged and used to represent the parcel centroid without further modification (Váša et al. 2018). Because the vertices are defined on a sphere, these averaged-coordinate centroids will always fall beneath the surface of the cortical mesh. To resolve this, *surface-projection averaging* performs the same procedure as *averaging* but then selects the coordinates of the closest vertex on the surface of the cortical mesh to represent the parcel centroid (Vázquez-Rodríguez et al. 2019). An alternative approach, *geodesic distance minimization*, calculates the vertex-by-vertex geodesic distance matrix

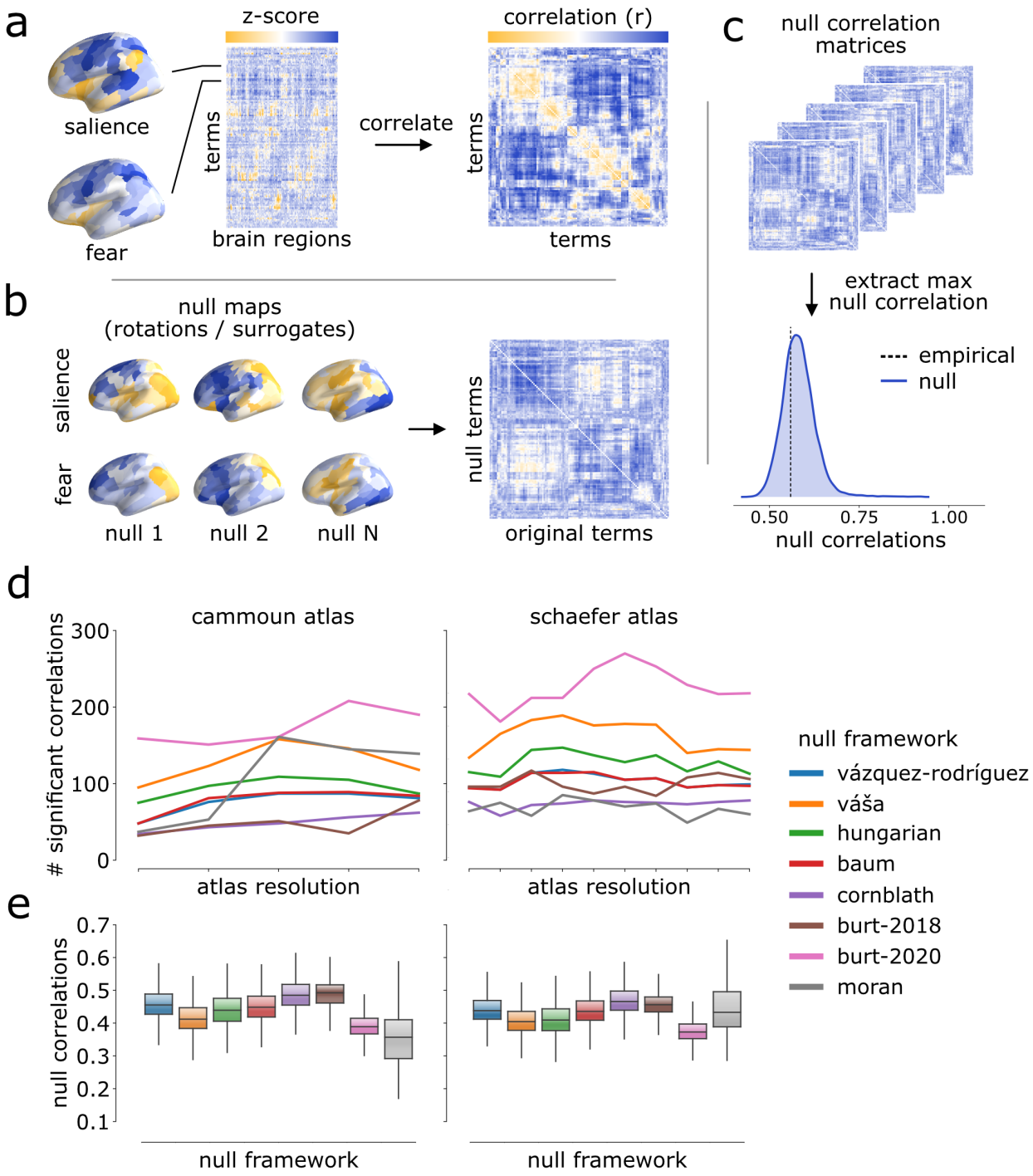

Figure 2. **Testing correspondence between brain maps |** (a) "Association test" z-score brain maps for 123 terms were downloaded from NeuroSynth. The parcellated term maps were correlated across brain regions to generate a term-by-term correlation matrix. (b) Each null framework was applied to the original term maps to generate 10,000 null maps per term. The original term maps were correlated with each of the null term maps to generate 10,000 null correlation matrices. (c) The maximum absolute correlation in each null matrix (excluding the diagonal) was extracted and retained to create a null distribution of correlations. This null distribution was used to threshold the original correlation matrix at an alpha of 0.05. (d) The number of significant correlations in the thresholded term matrix for each of the null frameworks as a function of atlas and atlas resolution. (e) The distributions of null correlations used to threshold the original term correlation matrix. Distributions are shown for the highest scale parcellation ($N = 1000$ nodes) of each atlas for all of the null frameworks. Refer to Fig. S2a for spatially-naive nulls and to Table S1 for a full list of NeuroSynth terms used.

separately for each parcel. Each distance matrix is averaged across columns and the coordinates of the vertex with the smallest average is used to represent the parcel centroid.

We generated ten sample reassignments for all nine combinations of parcel centroid definition method and the three aforementioned spatial permutation null models. The Hamming distance was used to compare the similarity between all generated reassignments (Hamming 1950).

Geodesic distances along the medial wall

We also investigated the degree to which constraining calculation of surface-based distance matrices to disallow travel along the medial wall impacts the outcomes of parameterized data models. We generated two distance matrices for each parcellation and resolution, either (1) permitting or (2) disallowing the travelled paths to intersect the medial wall. Surface distance between vertices was calculated using Dijkstra's algorithm on a graph representing the pial cortical mesh of the fsaverage5 surface (Fischl et al. 1999). Parcel-to-parcel distance was calculated as the average distance between every surface vertex in two parcels.

We used the generated distance matrices to create 1,000 surrogates for each parameterized data method, yielding 6,000 total surrogates (two distance matrices x three parameterized methods), and assessed the similarity of the resulting surrogates within each method using linear correlations. As the parameterized data models require an input brain map we used the parcellated T1w/T2w data to generate these surrogates.

RESULTS

We performed two analyses to investigate the impact of ten null models on controlling the FWER in statistical assessments of parcellated brain data. We briefly describe both analyses below and provide an overview of each of the ten null models in Table 1 (refer to *Methods and Materials* for more information; see Fig. 1 for an example null map from each model).

Testing correspondence between brain maps

We first investigate how each of the ten null models performs in a prototypical correspondence analysis, in which a user assesses the spatial correlation between two brain maps. As an example, we test correlations between meta-analytic functional activations. Following Alexander-Bloch et al. (2018), we parcellated meta-analytic "association" maps for 123 cognitive terms derived from NeuroSynth (Poldrack et al. 2011, Yarkoni et al. 2011). We used the term-specific association maps to construct a term $\times$ term linear correlation matrix, indicating the extent to which pairs of terms share a similar spatial pattern (Fig. 2a). We then applied each null model to generate a population of null term $\times$ term correlation matrices (Fig. 2b). Finally, we constructed a null distribution by retaining the maximum absolute correlation coefficient from each of the null term $\times$ term matrices, excluding the diagonal elements (Fig. 2c). This procedure yielded distributions of null correlations for each null model, which were used to threshold the empirical term $\times$ term correlation matrix (Alexander-Bloch et al. 2018).

Figs. 2d show the number of statistically significant correlations that remained in the term-by-term matrix after thresholding. All comparisons were performed at $\alpha = 0.05$. Results are organized according to two commonly-used parcellations of the cerebral cortex grey matter: one anatomical (Cammoun et al. 2012, Desikan et al. 2006) and one functional (Schaefer et al. 2018). To test the interaction between null model and parcellation resolution, analyses were repeated at multiple resolutions (5 resolutions for the Cammoun Atlas; 10 resolutions for the Schaefer atlas).

Altogether, we find consistent differences between null models (Fig. 2d,e and Fig. S2a). In particular, some of the methods are more liberal, yielding higher numbers of significant correlations (e.g. *Burt-2020*, *Váša*), while others are consistently more conservative (*Burt-2018*, *Cornblath*). Interestingly, these differences across null models do not appear to break down according to null model "family", with instances of spatial permutation and parameterized data nulls appearing as both more conservative and more liberal. Moreover, although there are differences among null models, the relative ordering of the null models is stable across multiple parcellation atlases and resolutions, suggesting that models perform consistently across multiple node definitions.

Testing partition specificity

We next compare null models on tests of partition specificity, in which a user examines whether a spatial feature (e.g. cortical thickness, intracortical myelin, functional connectivity strength) is significantly more concentrated in a partition of interest (e.g. in a specific intrinsic functional network or cytoarchitectonic class). Specifically, we tested whether the spatial distribution of the T1w/T2w ratio—thought to reflect intracortical myelin (Glasser and Van Essen 2011)—is circumscribed by intrinsic functional (Yeo et al. 2011) or cytoarchitectonic (Scholtens et al. 2018, von Economo and Koskinas 1925) network boundaries.

We first calculated the mean T1w/T2w ratio for all constituent parcels in the seven intrinsic networks derived by Yeo et al. (2011) (Fig. 3a). We then used each of the null models to generate a distribution of network-specific T1w/T2w means. Finally, each empirical network mean was expressed as a z-score relative to its respective null distribution. A network-specific p-value was estimated by computing the proportion of absolute-valued null network means that were greater than the

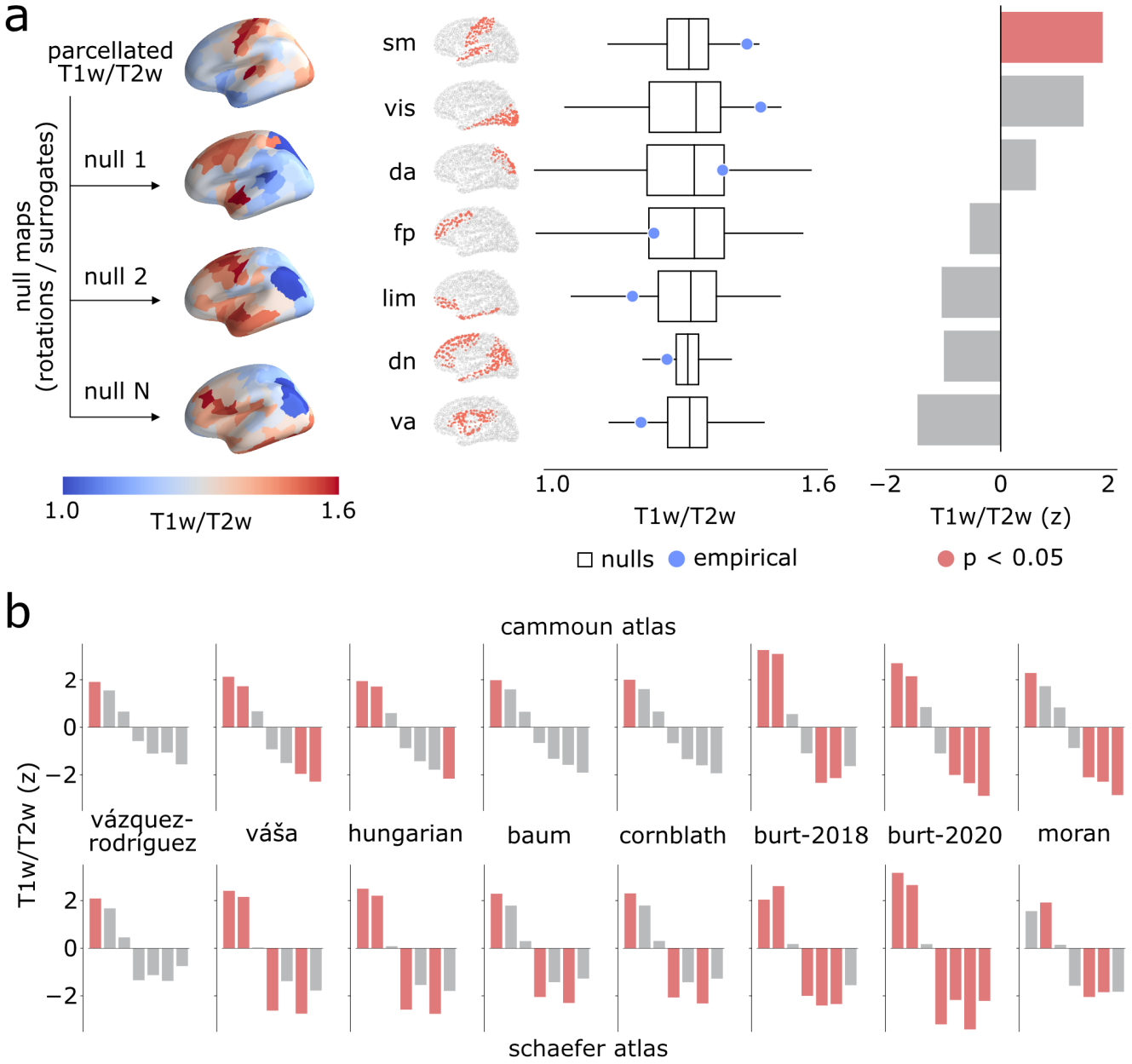

Figure 3. **Testing partition specificity |** (a) Parcellated T1w/T2w data from the Human Connectome Project were subjected to each of the described null frameworks, yielding 10,000 null T1w/T2w brain maps per framework. Parcels were averaged within each of the seven Yeo resting state networks (Yeo et al. 2011) for both the empirical (i.e., real) T1w/T2w brain map (blue dot) and the 10,000 null maps for each framework (white boxplot). Null distributions were used to normalize the empirical values, yielding a single z-score per network. A p-value was computed for each network as the proportion of null T1w/T2w network values that exceeded the real T1w/T2w value, divided by the total number of nulls. Networks with a p-value < 0.05 are shown in red. (b) The procedure described in panel (a) was performed for each of the null frameworks for both the Cammoun (top row) and Schaefer (bottom row) atlases. Results are shown for the highest scale parcellation ($N = 1000$ nodes) of each atlas. Yeo networks are shown from left-to-right in the same top-to-bottom order as in panel (a). Network assignments: sm = somatomotor, vis = visual, da = dorsal attention, fp = frontoparietal, lim = limbic, dn = default network, va = ventral attention.

absolute-valued empirical network mean (Fig. 3a), quantifying the probability that the T1w/T2w ratio is significantly greater or smaller in a particular network, above and beyond the network's size, symmetry, etc.

Fig. 3b shows network-specific z-scores for T1w/T2w ratios. Intrinsic networks with statistically significant (two-tailed, $\alpha = 0.05$) mean T1w/T2w ratios are shown in red, and non-significant networks are shown in grey. As described above, we repeated all comparisons for two commonly-used multi-scale parcellations (Cammoun et al. 2012, Desikan et al. 2006, Schaefer et al. 2018).

We observe three broad trends. First, null models are generally consistent in terms of effect direction and relative ordering of network z-scores: most methods identify over-expression of T1w/T2w ratio in the somatomotor and visual networks, and under-expression of T1w/T2w ratios in the ventral attention, default, frontoparietal and limbic networks (Fig. 3b). Second, despite these similarities, individual null models are inconsistent in the statistical inferences they provide: the number and identity of significant networks varies considerably from model to model. Third, the models are variable in terms of how conservative they are. Some models identify only 1/7 significant networks (e.g. *Vázquez-Rodríguez*), while others identify 6/7 significant networks (e.g. *Burt-2020*, Schaefer atlas). In contrast to results in the previous section, differences among models appear to break down according to model family, with spatial permutation null models being more conservative (*Vázquez-Rodríguez*, *Baum*, *Cornblath*) and parameterized nulls more liberal (e.g., *Burt-2018* and *Burt-2020*).

We also investigate differences between parcellation atlas resolutions and network partitions. Fig. 4a shows results for a structural and functional atlas (Cammoun and Schaefer) across 5 and 10 resolutions, respectively. The means are expressed as z-scores relative to a null distribution generated by a given method, and were computed for seven intrinsic functional networks (Yeo et al. 2011), as well as seven cytoarchitectonic classes (Scholtens et al. 2018, von Economo and Koskinas 1925) (Fig. 4b). The null model-, atlas-, resolution- and partition-specific z-scores are displayed in Fig. 4c and d (naive nulls are shown in Fig. S2b). Overall, the findings are consistent with the intuitions drawn above. Namely, we continue to observe consistency across nulls in terms of effect direction and ordering, and inconsistency in the number and identity of significant partitions; however, there is less differentiation between the spatial permutation and parameterized data null models. Although not specifically related to differences between models, we also note an example of how atlas and partition may potentially interact and lead to different inferences. Class #7 of the von Economo partition (limbic regions) is consistently deemed to be significantly under-enriched for T1w/T2w across Cammoun-derived parcellations, while it is generally not statistically significant across Schaefer-derived parcellations (Fig. 4d).

Implementation of null models can impact performance

While the presented results suggest that, at a broad level, selection of null model framework is an important choice for researchers, it is by no means the only choice that can impact analytical results. Here, we investigate how different implementations and seemingly minor methodological choices of individual spatial permutation and parameterized data null models can influence statistical outcomes.

Variability in parcel centroid definition

Of the five spatial permutation null models presented in the current report, three require definition of a centroid for each parcel (*Vázquez-Rodríguez*, *Váša*, and *Hungarian*). For the analyses reported, parcel centroids were generated following the procedure used by Vázquez-Rodríguez et al. (2019), which include: (1) taking the average of the coordinates of all vertices within a given parcel and (2) using the coordinates of the surface vertex closest to this average (where closest is defined as minimizing Euclidean distance). However, Váša et al. (2018) defined their parcel centroids using only the average of the vertices within each parcel (i.e., not projecting back to a surface vertex). Notably, both of these procedures fail to take into account oblong or C-shaped parcels for which the center-of-mass may fall outside the boundaries of the parcel. An alternative approach to account for this possibility is to find the coordinates of the vertex that minimizes the average surface (geodesic) distance to all other vertices within each parcel.

We assessed the extent to which these three different methods of parcel centroid definition impacted the generated rotations for the three relevant null models. We generated ten rotation matrices and applied them to the coordinates derived from each of these three parcel centroid definition methods, reassigning parcels using the approach from each of the three relevant null frameworks. We compared the similarity of the reassignments using the Hamming distance (Fig. 5a; Hamming 1950). Critically, because the same rotations were applied for every model and method, any observed differences are a result of variation in either the parcel centroid coordinates or models themselves.

Fig. 5a shows the Hamming distance between reassignments for all combinations of these parcel centroid and null framework methods across the Cammoun and Schaefer atlases (219 and 200 nodes, respectively; refer to S1 for all resolutions). Predictably, we observe the strongest differences between reassignments generated from the different null methods. In particular, the *Váša* null method seems to markedly differ from the *Vázquez-Rodríguez* and *Hungarian* methods. In addition, there is also variation between parcel centroid calculation methods; for example, the *Váša* method seems to have only moderate correspondence between different parcel centroid definitions. Altogether, these results highlight im-

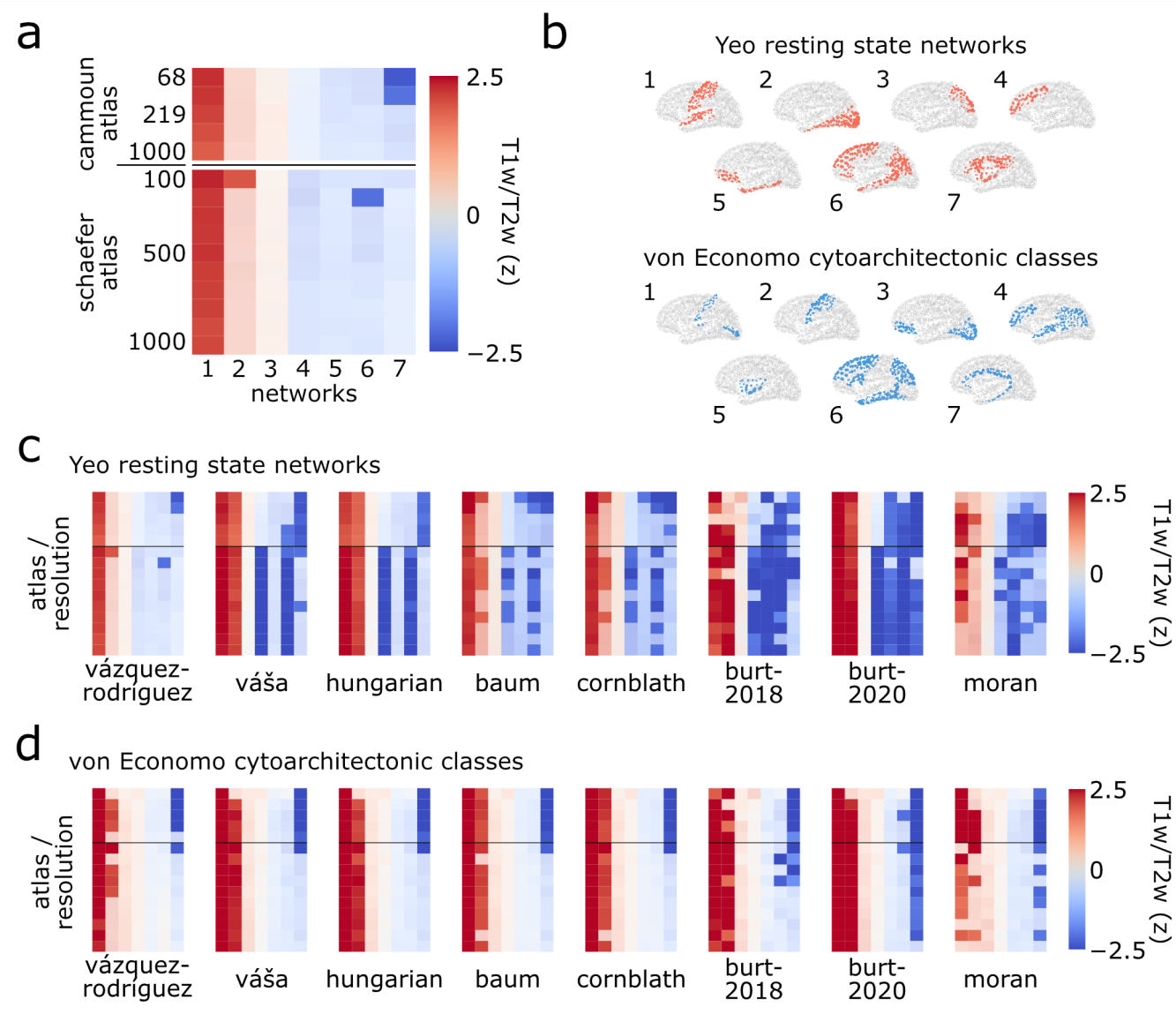

Figure 4. **Partition-specific definition of T1w/T2s maps |** (a) Overview of the results assessing partition specificity of T1w/T2w measurements. Each cell in the heatmap represents the T1w/T2w z-score for a given network (x-axis) and atlas / resolution (y-axis). Bold colors represent network z-scores that are significant at the $\alpha = 0.05$ level. Note that the fifth and fifteenth row of each heatmap represent the same values shown in Fig. 3b. (b) The network labels for both the seven Yeo resting state networks (Yeo et al. 2011) and the seven von Economo-Koskinas cytoarchitectonic classes (Scholtens et al. 2018, von Economo and Koskinas 1925). (c, d) The results for all eight spatially-constrained null models for all atlases and resolutions for the (c) Yeo resting state networks and (d) von Economo cytoarchitectonic classes. While there is some notable variation amongst atlases and resolutions, primary differences are observable between null frameworks and partitions. Refer to Fig. S2b for spatially-naive nulls. Yeo resting state networks: 1 = somatomotor, 2 = visual, 3 = dorsal attention, 4 = frontoparietal, 5 = limbic, 6 = default, 7 = ventral attention. von Economo cytoarchitectonic classes: 1 = primary sensory cortex, 2 = primary motor cortex, 3 = primary/secondary sensory cortex, 4 = association cortex, 5 = insular cortex, 6 = association cortex, 7 = limbic regions.

portant differences not only between spatial permutation null models but within the implementation of each model itself.

Geodesic distances along the medial wall

One notable disadvantage of spatial permutation null models is that the rotations on which they rely will cause the medial wall to be displaced into the cortical sheet

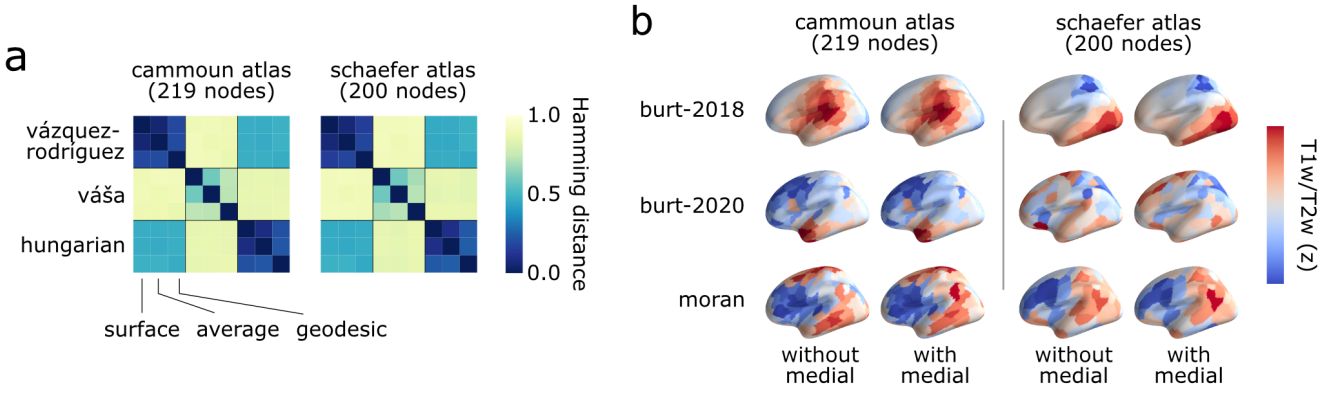

Figure 5. **Implementation of null models can impact performance |** (a) The impact of parcel centroid definition on three spatial permutation null models. Heatmaps show the correspondence (Hamming distance) between null reassignments generated using different null frameworks and parcel centroid methods, where lower values (blue colors) indicate greater correspondence. The heatmaps are broken into nine sections (black outlines) which delineate different null frameworks; cells within each section represent the different parcel centroid methods. For results from all resolutions of the Cammoun and Schaefer atlases refer to S1. (b) Example surrogates generated from distances matrices allowing ("with medial") and disallowing ("without medial") surface distance travel along the medial wall for three parameterized data null models (Burt-2018, Burt-2020, and Moran).

(Fig. 1). The resulting null distribution will have missing data for this portion of the cortex and discarded data for the cortex rotated into the medial wall, yielding uneven sample sizes between null distributions and potentially biasing results (refer to Burt et al. 2020 for an illustration of this concept).

While the family of parameterized data models do not rely on rotations and thus are not subject to this constraint, they do rely on a user-defined weight matrix that provides an estimation of the spatial structure of the related brain maps. In most cases—as in the current report—this weight matrix takes the form of a geodesic- or surface-based distance matrix calculated from shortest paths taken along the surface mesh projection of a brain. However, the calculation of these distance matrices often fails to take into account that paths along the brain's medial wall, where the cortical sheet is discontinuous due to underlying white matter and subcortical nuclei, should not be possible. Failing to account for this discontinuity yields distance matrices that categorically underestimate the actual distance between many brain regions.

We tested the extent to which generating distance matrices (dis-)allowing travel along the medial wall biases the surrogate data maps generated from the three parameterized null models. Fig. 5b shows an example surrogate map for these null models for the Cammoun and Schaefer atlases. Overall, surrogates generated from the different distance matrices show strong similarities. To examine this in more detail we correlated 1,000 surrogates generated from the two distance matrices for each method, and found relatively moderate to high correspondence: 0.85 ± 0.27 (*Burt-2018*), 0.85 ± 0.15 (*Burt-2020*), 0.94 ± 0.04 (*Moran*) (mean $\pm$ standard deviation across all atlases and resolutions). Altogether, although parameterized data nulls seem relatively robust to this issue, how to handle the medial wall when constructing the required distance matrices is still an important choice for researchers to make when implementing these models.

DISCUSSION

In the present report we systematically assess the effect of ten null frameworks on statistical inference in two prototypical analyses. Within the spatially-constrained null frameworks we observe limited differences in performance. Although in some cases there appears to be a primary distinction between the spatial permutation and parameterized families—with the former tending to be more conservative and the latter more liberal—this strongly depends on research context and question. Encouragingly, we find minimal impact of parcellation and parcel resolution on these null frameworks. Across all of our analyses, however, we find that spatially-naive nulls models consistently yield both unrealistically liberal statistical estimates and a strong dependence on parcel resolution, confirming previous reports that these methods are unsuitable for most neuroimaging analyses (Alexander-Bloch et al. 2018, Burt et al. 2020).

From the perspective of a researcher, the first choice when applying one of these frameworks is which family to choose from (Tables 1 and S2 provide summaries of their relative differences and limitations). On

the one hand, the spatial permutation null models are computationally-efficient. Since they are not conditioned on the data itself—simply the geometry of the cortical surface—the same null can be applied to any dataset defined on the same cortical mesh. On the other hand, nulls generated by the parameterized models (with the exception of the Moran method), are dependent on the data itself, dramatically increasing the computational cost for multivariate analyses. However, the benefit of parameterized nulls is that they can be used for all data geometries. That is, they can be applied to both volumetric and surface data, and the models are consistent regardless of whether the data are defined at the vertex-, voxel-, or parcel-level (Burt et al. 2020).

Implementation differences within null families entail additional methodological choices that can ultimately lead to variability in statistical inference (Table 1). Within the spatial permutation nulls, differences are primarily driven by handling of the medial wall (i.e., *Vázquez-Rodríguez*, *Váša*, *Hungarian*) and whether to project parcellated data to the vertex-level (i.e., *Baum*, *Cornblath*). The parameterized nulls mostly vary in how they implement the data-generating process. More specifically, these frameworks differ in their underlying assumptions of the nature of the spatial autocorrelation itself and how it is estimated in the first place. Our results clearly show that these implementation differences have an observable effect on generated statistical estimates, but more research is needed to better understand how differences in performance arise from specific methodological choices.

More broadly, we find that null framework performance depends on the research question itself, such that differences between frameworks are more pronounced for some types of analyses than others. In particular, we see greater variability between nulls when examining partition specificity (Fig. 3) as compared to brain map correspondence (Fig 2). Within partition specificity analyses we observe increased variability between frameworks when comparing intrinsic resting state networks than when examining cytoarchitectonic classes (Fig 4). Indeed, the main limitation of the present report is that it is focused on single instances of two prototypical analyses. That is, we sample only a small space of the application of these null frameworks and do not exhaustively assess their performance in all conceivable research contexts.

Interestingly, we find that parcellation and parcel resolution seem to have minimal effect on null framework performance. How brain regions are defined and ultimately represented in analyses is becoming an increasingly important choice in neuroimaging (Eickhoff et al. 2018). Recent work has highlighted the role of brain region definition in analyses including test-retest reliability (Arslan et al. 2018), structure-function coupling (Messé 2020, Vázquez-Rodríguez et al. 2019), individual fingerprinting (Finn et al. 2015), modelling of brain dynamics (Proix et al. 2016), and prediction of behavior and disease (Dadi et al. 2020, Kong et al. 2019). Although the choice of parcellation is clearly an important methodological consideration, our results suggest that it does not strongly influence spatially-constrained null models. This ostensible "parcellation invariance" is encouraging, and will hopefully support broader adoption of these models in future studies.

More generally, this investigation builds on increasing efforts in the neuroimaging literature to benchmark the effects of methodological choices. Recently, the broad impacts of these choices has been demonstrated in structural MRI (Bhagwat et al. 2020, Kharabian Masouleh et al. 2020), task fMRI (Botvinik-Nezer et al. 2020, Carp 2012), resting state fMRI (Ciric et al. 2017, Parkes et al. 2018), and diffusion MRI (Maier-Hein et al. 2017, Oldham et al. 2020, Schilling et al. 2019) research. Concomitant with an increasing awareness of the importance of these choices is a developing trend to share and analyze "raw" or un-thresholded brain maps (Gorgolewski et al. 2015, Witt et al. 2020). Convergence of these trends has naturally opened new research questions that revolve around comparing such brain maps. Methods that appropriately consider the inherent spatial structure of the brain are thus going to be critical to ensuring valid inferences can be drawn from these lines of inquiry. The present study is a first step towards better understanding the variable implementation of these statistical methods and ultimately standardizing or synthesizing their application.

Collectively, the present report comprehensively examines ten recently-developed null models for comparing brain maps. We find relatively consistent performance among the null models that account for spatial autocorrelation; however, we note some differences between nulls that depend on analytic framework. Our results highlight the need to closely consider the variability across implementation of these null methods and to more explicitly compare their performance across a wider range of research contexts.

ACKNOWLEDGEMENTS

We thank Laura Suárez, Golia Shafiei, Justine Hansen, Vincent Bazinet, Bertha Vázquez-Rodríguez, and Elizabeth DuPre for their comments and suggestions. This research was undertaken thanks in part to funding from the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives initiative. RDM acknowledges support from the Fonds du Recherche Québec - Nature et Technologies. BM acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery Grant RGPIN #017-04265) and from the Canada Research Chairs Program.

Akbarian, S., Liu, C., Knowles, J. A., Vaccarino, F. M., Farnham, P. J., Crawford, G. E., Jaffe, A. E., Pinto, D., Dracheva, S., Geschwind, D. H., Mill, J., Nairn, A. C., Abyzov, A., Pochareddy, S., Prabhakar, S., Weissman, S., Sullivan, P. F., State, M. W., Weng, Z., Peters, M. A., White, K. P., Gerstein, M. B., Amiri, A., Armoskus, C., Ashley-Koch, A. E., Bae, T., Beckel-Mitchener, A., Berman, B. P., Coetzee, G. A., Coppola, G., Francoeur, N., Fromer, M., Gao, R., Grennan, K., Herstein, J., Kavanagh, D. H., Ivanov, N. A., Jiang, Y., Kitchen, R. R., Kozlenkov, A., Kundakovic, M., Li, M., Li, Z., Liu, S., Mangravite, L. M., Mattei, E., Markenscoff-Papadimitriou, E., Navarro, F. C. P., North, N., Omberg, L., Panchision, D., Parikshak, N., Poschmann, J., Price, A. J., Purcaro, M., Reddy, T. E., Roussos, P., Schreiner, S., Scuderi, S., Sebra, R., Shibata, M., Shieh, A. W., Skarica, M., Sun, W., Swarup, V., Thomas, A., Tsuji, J., van Bakel, H., Wang, D., Wang, Y., Wang, K., Werling, D. M., Willsey, A. J., Witt, H., Won, H., Wong, C. C. Y., Wray, G. A., Wu, E. Y., Xu, X., Yao, L., Senthil, G., Lehner, T., Sklar, P., and Sestan, N. (2015). The PsychENCODE project. *Nature Neuroscience*, 18(12):1707–1712.

Alexander-Bloch, A., Raznahan, A., Bullmore, E., and Giedd, J. (2013). The convergence of maturational change and structural covariance in human cortical networks. *Journal of Neuroscience*, 33(7):2889–2899.

Alexander-Bloch, A. F., Shou, H., Liu, S., Satterthwaite, T. D., Glahn, D. C., Shinohara, R. T., Vandekar, S. N., and Raznahan, A. (2018). On testing for spatial correspondence between maps of human brain structure and function. *NeuroImage*, 178:540–551.

Arslan, S., Ktena, S. I., Makropoulos, A., Robinson, E. C., Rueckert, D., and Parisot, S. (2018). Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. *Neuroimage*, 170:5–30.

Baum, G. L., Cui, Z., Roalf, D. R., Ciric, R., Betzel, R. F., Larsen, B., Cieslak, M., Cook, P. A., Xia, C. H., Moore, T. M., Ruparel, K., Oathes, D. J., Alexander-Bloch, A. F., Shinohara, R. T., Raznahan, A., Gur, R. E., Gur, R. C., Bassett, D. S., and Satterthwaite, T. D. (2020). Development of structure–function coupling in human brain networks during youth. *Proceedings of the National Academy of Sciences*, 117(1):771–778.

Beliveau, V., Ganz, M., Feng, L., Ozenne, B., Højgaard, L., Fisher, P. M., Svarer, C., Greve, D. N., and Knudsen, G. M. (2017). A high-resolution in vivo atlas of the human brain's serotonin system. *Journal of Neuroscience*, 37(1):120–128.

Bellec, P., Rosa-Neto, P., Lyttelton, O. C., Benali, H., and Evans, A. C. (2010). Multi-level bootstrap analysis of stable clusters in resting-state fMRI. *Neuroimage*, 51(3):1126–1139.

Bhagwat, N., Barry, A., Dickie, E. W., Brown, S. T., Devenyi, G. A., Hatano, K., DuPre, E., Dagher, A., Chakravarty, M. M., Greenwood, C. M., Mišić, B., Kennedy, D. N., and Poline, J.-B. (2020). Understanding the impact of preprocessing pipelines on neuroimaging cortical surface analyses. *bioRxiv*.

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, A., Avesani, P., Baczkowski, B., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., Benoit, R., Berkers, R., Bhanji, J., Biswal, B., Bobadilla-Suarez, S., Bortolini, T., Bottenhorn, K., Bowring, A., Braem, S., Brooks, H., Brudner, E., Calderon, C., Camilleri, J., Castrellon, J., Cecchetti, L., Cieslik, E., Cole, Z., Collignon, O., Cox, R., Cunningham, W., Czoschke, S., Dadi, K., Davis, C., De Luca, A., Delgado, M., Demetriou, L., Dennison, J., Di, X., Dickie, E., Dobryakova, E., Donnat, C., Dukart, J., Duncan, N. W., Durnez, J., Eed, A., Eickhoff, S., Erhart, A., Fontanesi, L., Fricke, G. M., Galvan, A., Gau, R., Genon, S., Glatard, T., Glerean, E., Goeman, J., Golowin, S., González-García, C., Gorgolewski, K., Grady, C., Green, M., Guassi Moreira, J., Guest, O., Hakimi, S., Hamilton, J. P., Hancock, R., Handjaras, G., Harry, B., Hawco, C., Herholz, P., Herman, G., Heunis, S., Hoffstaedter, F., Hogeveen, J., Holmes, S., Hu, C.-P., Huettel, S., Hughes, M., Iacovella, V., Iordan, A., Isager, P., Isik, A. I., Jahn, A., Johnson, M., Johnstone, T., Joseph, M., Juliano, A., Kable, J., Kassinopoulos, M., Koba, C., Kong, X.-Z., Koscik, T., Kucukboyaci, N. E., Kuhl, B., Kupek, S., Laird, A., Lamm, C., Langner, R., Lauharatanahirun, N., Lee, H., Lee, S., Leemans, A., Leo, A., Lesage, E., Li, F., Li, M., Lim, P. C., Lintz, E., Liphardt, S., Losecaat Vermeer, A., Love, B., Mack, M., Malpica, N., Marins, T., Maumet, C., McDonald, K., McGuire, J., Melero, H., Méndez Leal, A., Meyer, B., Meyer, K., Mihai, P., Mitsis, G., Moll, J., Nielson, D., Nilsonne, G., Notter, M., Olivetti, E., Onicas, A., Papale, P., Patil, K., Peelle, J. E., Pérez, A., Pischedda, D., Poline, J.-B., Prystauka, Y., Ray, S., Reuter-Lorenz, P., Reynolds, R., Ricciardi, E., Rieck, J., Rodriguez-Thompson, A., Romyn, A., Salo, T., Samanez-Larkin, G., Sanz-Morales, E., Schlichting, M., Schultz, D., Shen, Q., Sheridan, M., Shiguang, F., Silvers, J., Skagerlund, K., Smith, A., Smith, D., Sokol-Hessner, P., Steinkamp, S., Tashjian, S., Thirion, B., Thorp, J., Tinghög, G., Tisdall, L., Tompson, S., Toro-Serey, C., Torre, J., Tozzi, L., Truong, V., Turella, L., van't Veer, A. E., Verguts, T., Vettel, J., Vijayarajah, S., Vo, K., Wall, M., Weeda, W. D., Weis, S., White, D., Wisniewski, D., Xifra-Porxas, A., Yearling, E., Yoon, S., Yuan, R., Yuen, K., Zhang, L., Zhang, X., Zosky, J., Nichols, T. E., Poldrack, R. A., and Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*.

Brett, M., Markiewicz, C. J., Hanke, M., Côté, M.-A., Cipollini, B., McCarthy, P., Jarecka, D., Cheng, C. P., Halchenko, Y. O., Cottaar, M., Ghosh, S., Larson, E., Wassermann, D., Gerhard, S., Lee, G. R., Wang, H.-T., Kastman, E., Kaczmarzyk, J., Guidotti, R., Duek, O., Rokem, A., Madison, C., Morency, F. C., Moloney, B., Goncalves, M., Markello, R., Riddell, C., Burns, C., Millman, J., Gramfort, A., Leppäkangas, J., Sólon, A., van den Bosch, J. J., Vincent, R. D., Braun, H., Subramaniam, K., Gorgolewski, K. J., Raamana, P. R., Nichols, B. N., Baker, E. M., Hayashi, S., Pinsard, B., Haselgrove, C., Hymers, M., Esteban, O., Koudoro, S., Oosterhof, N. N., Amirbekian, B., Nimmo-Smith, I., Nguyen, L., Reddigari, S., St-Jean, S., Panfilov, E., Garyfallidis, E., Varoquaux, G., Legarreta, J. H., Hahn, K. S., Hinds, O. P., Fauber, B., Poline, J.-B., Stutters, J., Jordan, K., Cieslak, M., Moreno, M. E., Haenel, V., Schwartz, Y., Baratz, Z., Darwin, B. C., Thirion, B., Papadopoulos Orfanos, D., Pérez-García, F., Solovey, I., Gonzalez, I., Palasubramaniam, J., Lecher, J., Leinweber, K., Raktivan, K., Fischer, P., Gervais, P., Gadde, S., Ballinger, T., Roos, T., and Reddam, V. R. (2020). nipy/nibabel. *Zenodo*, doi:10.5281/zenodo.591597.

Burt, J. B., Demirtaş, M., Eckner, W. J., Navejar, N. M., Ji, J. L., Martin, W. J., Bernacchia, A., Anticevic, A., and Murray, J. D. (2018). Hierarchy of transcriptomic specialization across hu-

man cortex captured by structural neuroimaging topography. *Nature Neuroscience*, 21(9):1251–1259.

Burt, J. B., Helmer, M., Shinn, M., Anticevic, A., and Murray, J. D. (2020). Generative modeling of brain maps with spatial autocorrelation. *NeuroImage*.

Cammoun, L., Gigandet, X., Meskaldji, D., Thiran, J. P., Sporns, O., Do, K. Q., Maeder, P., Meuli, R., and Hagmann, P. (2012). Mapping the human connectome at multiple scales with diffusion spectrum MRI. *Journal of Neuroscience Methods*, 203(2):386–397.

Carp, J. (2012). On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments. *Frontiers in Neuroscience*, 6:149.

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E., Teslovich, T., Dellarco, D. V., Garavan, H., Orr, C. A., Wager, T. D., Banich, M. T., Speer, N. K., Sutherland, M. T., Riedel, M. C., Dick, A. S., Bjork, J. M., Thomas, K. M., Chaarani, B., Mejia, M. H., Jr., D. J. H., Cornejo, M. D., Sicat, C. S., Harms, M. P., Dosenbach, N. U. F., Rosenberg, M., Earl, E., Bartsch, H., Watts, R., Polimeni, J. R., Kuperman, J. M., Fair, D. A., Dale, A. M., and the ABCD Imaging Acquisition Workgroup (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, 32:43–54.

Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., Shinohara, R. T., Elliott, M. A., Eickhoff, S. B., Davatzikos, C., Gur, R. C., Gur, R. E., Bassett, D. S., and Satterthwaite, T. D. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage*, 154:174–187.

Cornblath, E. J., Ashourvan, A., Kim, J. Z., Betzel, R. F., Ciric, R., Adebimpe, A., Baum, G. L., He, X., Ruparel, K., Moore, T. M., Gur, R. C., Gur, R. E., Shinohara, R. T., Roalf, D. R., Satterthwaite, T. D., and Bassett, D. S. (2019). Temporal sequences of brain activity at rest are constrained by white matter structure and modulated by cognitive demands.

Dadi, K., Varoquaux, G., Machlouzarides-Shalit, A., Gorgolewski, K. J., Wassermann, D., Thirion, B., and Mensch, A. (2020). Fine-grain atlases of functional modes for fMRI analysis. *NeuroImage*.

Damoiseaux, J. S., Rombouts, S., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., and Beckmann, C. F. (2006). Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences*, 103(37):13848–13853.

Demirtaş, M., Burt, J. B., Helmer, M., Ji, J. L., Adkinson, B. D., Glasser, M. F., Van Essen, D. C., Sotiropoulos, S. N., Anticevic, A., and Murray, J. D. (2019). Hierarchical heterogeneity across human cortex shapes large-scale neural dynamics. *Neuron*, 101(6):1181–1194.

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., Albert, M. S., and Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980.

Dray, S. (2011). A new perspective about Moran's coefficient: Spatial autocorrelation as a linear regression problem. *Geographical Analysis*, 43(2):127–141.

Eickhoff, S. B., Yeo, B. T., and Genon, S. (2018). Imaging-based parcellations of the human brain. *Nature Reviews Neuroscience*, 19(11):672–686.

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., and Constable, R. T. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18(11):1664–1671.

Fischl, B., Sereno, M. I., Tootell, R. B., and Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, 8(4):272–284.

Fulcher, B. D., Arnatkeviciute, A., and Fornito, A. (2020). Overcoming bias in gene-set enrichment analyses of brain-wide transcriptomic data. *bioRxiv*.

Gao, R., van den Brink, R. L., Pfeffer, T., and Voytek, B. (2020). Neuronal timescales are functionally dynamic and shaped by cortical microarchitecture. *bioRxiv*.

Glasser, M. F. and Van Essen, D. C. (2011). Mapping human cortical areas in vivo based on myelin content as revealed by T1-and T2-weighted MRI. *Journal of Neuroscience*, 31(32):11597–11616.

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., Sochat, V. V., Nichols, T. E., Poldrack, R. A., Poline, J.-B., Yarkoni, T., and Margulies, D. S. (2015). Neurovault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics*, 9:8.

Hamming, R. W. (1950). Error detecting and error correcting codes. *The Bell System Technical Journal*, 29(2):147–160.

Hansen, J. Y., Markello, R. D., Vogel, J. W., Seidlitz, J., Bzdok, D., and Mišić, B. (2020). Molecular signatures of cognition and affect. *bioRxiv*.

Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., van de Lagemaat, L. N., Smith, K. A., Ebbert, A., Riley, Z. L., Abajian, C., Beckmann, C. F., Bernard, A., Bertagnolli, D., Boe, A. F., Cartagena, P. M., Chakravarty, M. M., Chapin, M., Chong, J., Dalley, R. A., Daly, B. D., Dang, C., Datta, S., Dee, N., Dolbeare, T. A., Faber, V., Feng, D., Fowler, D. R., Goldy, J., Gregor, B. W., Haradon, Z., Haynor, D. R., Hohmann, J. G., Horvath, S., Howard, R. E., Jeromin, A., Jochim, J. M., Kinnunen, M., Lau, C., Lazarz, E. T., Lee, C., Lemon, T. A., Li, L., Li, Y., Morris, J. A., Overly, C. C., Parker, P. D., Parry, S. E., Reding, M., Royall, J. J., Schulkin, J., Sequeira, P. A., Slaughterbeck, C. R., Smith, S. C., Sodt, A. J., Sunkin, S. M., Swanson, B. E., Vawter, M. P., Williams, D., Wohnoutka, P., Zielke, H. R., Geschwind, D. H., Hof, P. R., Smith, S. M., Koch, C., Grant, S. G. N., and Jones, A. R. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, 489(7416):391.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95.

Insel, T. R., Landis, S. C., and Collins, F. S. (2013). The NIH BRAIN initiative. *Science*, 340(6133):687–688.

Kharabian Masouleh, S., Eickhoff, S., Zeighami, Y., Lewis, L., Dahnke, R., Gaser, R., Gaser, C., Chouinard-Decorte, F., Lepage, C., Scholtens, L., Hoffstaedter, F., Glahn, D., Blangero, J., Evans, A., Genon, S., and Valk, S. L. (2020). Influence of processing pipeline on cortical thickness measurement. *Cerebral Cortex*.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., and the Jupyter development team (2016). Jupyter Notebooks–A publishing format for reproducible computational workflows. In Loizides, F. and Scmidt, B., editors, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87–90. IOS Press.

Kong, R., Li, J., Orban, C., Sabuncu, M. R., Liu, H., Schaefer, A., Sun, N., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., and Yeo, B. T. T. (2019). Spatial topography of individual-specific cortical networks predicts human cognition, personality, and emotion. *Cerebral Cortex*, 29(6):2533–2551.

Kuhn, H. W. (1955). The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97.

Maier-Hein, K. H., Neher, P. F., Houde, J.-C., Côté, M.-A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F.-C., Lin, Y.-C., Ji, Q., Reddick, W. E., Glass, J. O., Chen, D. Q., Feng, Y., Gao, C., Wu, Y., Ma, J., He, R., Li, Q., Westin, C.-F., Deslauriers-Gauthier, S., González, J. O. O., Paquette, M., St-Jean, S., Girard, G., Rheault, F., Sidhu, J., Tax, C. M. W., Guo, F., Mesri, H. Y., Dávid, S., Froeling, M., Heemskerk, A. M., Leemans, A., Boré, A., Pinsard, B., Bedetti, C., Desrosiers, M., Brambati, S., Doyon, J., Sarica, A., Vasta, R., Cerasa, A., Quattrone, A., Yeatman, J., Khan, A. R., Hodges, W., Alexander, S., Romascano, D., Barakovic, M., Auría, A., Esteban, O., Lemkaddem, A., Thiran, J.-P., Cetingul, H. E., Odry, B. L., Mailhe, B., Nadar, M. S., Pizzagalli, F., Prasad, G., Villalon-Reina, J. E., Galvis, J., Thompson, P. M., Requejo, F. D. S., Laguna, P. L., Lacerda, L. M., Barrett, R., Dell'Acqua, F., Catani, M., Petit, L., Caruyer, E., Daducci, A., Dyrby, T. B., Holland-Letz, T., Hilgetag, C. C., Stieltjes, B., and Descoteaux, M. (2017). The challenge of mapping the human connectome based on diffusion tractography. *Nature Communications*, 8(1):1–13.

Marcus, D., Harwell, J., Olsen, T., Hodge, M., Glasser, M., Prior, F., Jenkinson, M., Laumann, T., Curtiss, S., and Van Essen, D. (2011). Informatics and data mining tools and strategies for the human connectome project. *Frontiers in Neuroinformatics*, 5:4.

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., Bezgin, G., Eickhoff, S. B., Castellanos, F. X., Petrides, M., Jeffries, E., and Smallwood, J. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences*, 113(44):12574–12579.

McKinney, W. (2010). Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX.

Messé, A. (2020). Parcellation influence on the connectivity-based structure–function relationship in the human brain. *Human Brain Mapping*, 41(5):1167–1180.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23.

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D., and Wang, X.-J. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience*, 17(12):1661–1663.

Norgaard, M., Beliveau, V., Ganz, M., Svarer, C., Pinborg, L. H., Keller, S. H., Jensen, P. S., Greve, D. N., and Knudsen, G. M. (2020). A high-resolution in vivo atlas of the human brain's benzodiazepine binding site of $GABA_A$ receptors. *bioRxiv*.

Oldham, S., Arnatkeviciute, A., Smith, R. E., Tiego, J., Bellgrove, M. A., and Fornito, A. (2020). The efficacy of different preprocessing steps in reducing motion-related confounds in diffusion MRI connectomics. *bioRxiv*.

Oliphant, T. E. (2006). *A guide to NumPy*, volume 1. Trelgol Publishing USA.

Paquola, C., Seidlitz, J., Benkarim, O., Royer, J., Klimes, P., Bethlehem, R. A. I., Lariviere, S., Vos de Wael, R., Hall, J. A., Frauscher, B., Smallwood, J., and Bernhardt, B. C. (2020). The cortical wiring scheme of hierarchical information processing. *bioRxiv*.

Parkes, L., Fulcher, B., Yücel, M., and Fornito, A. (2018). An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional mri. *NeuroImage*, 171:415–436.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct):2825–2830.

Pérez, F. and Granger, B. E. (2007). IPython: A system for interactive scientific computing. *Computing in Science & Engineering*, 9(3):21–29.

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, 72(5):692–697.

Poldrack, R. A., Barch, D. M., Mitchell, J., Wager, T., Wagner, A. D., Devlin, J. T., Cumba, C., Koyejo, O., and Milham, M. (2013). Toward open sharing of task-based fMRI data: the OpenfMRI project. *Frontiers in Neuroinformatics*, 7:12.

Poldrack, R. A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., Parker, D. S., Sabb, F. W., and Bilder, R. M. (2011). The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. *Frontiers in Neuroinformatics*, 5:17.

Poldrack, R. A. and Yarkoni, T. (2016). From brain maps to cognitive ontologies: Informatics and the search for mental structure. *Annual Review of Psychology*, 67:587–612.

Proix, T., Spiegler, A., Schirner, M., Rothmeier, S., Ritter, P., and Jirsa, V. K. (2016). How do parcellation size and short-range connectivity affect dynamics in large-scale brain network models? *NeuroImage*, 142:135–149.

Royer, J., Paquola, C., Larivière, S., de Wael, R. V., Tavakol, S., Lowe, A. J., Benkarim, O., Evans, A. C., Bzdok, D., Smallwood, J., Frauscher, B., and Bernhardt, B. C. (2020). Myeloarchitecture gradients in the human insula: Histological underpinnings

and association to intrinsic functional connectivity. *NeuroImage*, page 116859.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., and Yeo, B. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, 28(9):3095–3114.

Schilling, K. G., Nath, V., Hansen, C., Parvathaneni, P., Blaber, J., Gao, Y., Neher, P., Aydogan, D. B., Shi, Y., Ocampo-Pineda, M., Schiavi, S., Daducci, A., Girard, G., Barakovic, M., Rafael-Patino, J., Romascano, D., Rensonnet, G., Pizzolato, M., Bates, A., Fischi, E., Thiran, J.-P., Canales-Rodríguez, E. J., Huang, C., Zhu, H., Zhong, L., Cabeen, R., Toga, A. W., Rheault, F., Theaud, G., Houde, J.-C., Sidhu, J., Chamberland, M., Westin, C.-F., Dyrby, T. B., Verma, R., Rathi, Y., Irfanoglu, M. O., Thomas, C., Pierpaoli, C., Descoteaux, M., Anderson, A. W., and Landman, B. A. (2019). Limits to anatomical accuracy of diffusion tractography using modern approaches. *NeuroImage*, 185:1–11.

Scholtens, L. H., de Reus, M. A., de Lange, S. C., Schmidt, R., and van den Heuvel, M. P. (2018). An MRI Von Economo–Koskinas atlas. *NeuroImage*, 170:249–256.

Shafiei, G., Markello, R. D., de Wael, R. V., Bernhardt, B. C., Fulcher, B. D., and Misic, B. (2020). Topographic gradients of intrinsic dynamics across neocortex. *bioRxiv*.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., and Collins, R. (2015). UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, 12(3).

Van Der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The NumPy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., and the Wu-Minn HCP Consortium (2013). The WU-Minn human connectome project: an overview. *NeuroImage*, 80:62–79.

Váša, F., Seidlitz, J., Romero-Garcia, R., Whitaker, K. J., Rosenthal, G., Vértes, P. E., Shinn, M., Alexander-Bloch, A., Fonagy, P., Dolan, R. J., Jones, P. B., Goodyer, I. M., the NSPN consortium, Sporns, O., and Bullmore, E. T. (2018). Adolescent tuning of association cortex in human structural brain networks. *Cerebral Cortex*, 28(1):281–294.

Vázquez-Rodríguez, B., Suárez, L. E., Markello, R. D., Shafiei, G., Paquola, C., Hagmann, P., Van Den Heuvel, M. P., Bernhardt, B. C., Spreng, R. N., and Mišić, B. (2019). Gradients of structure–function tethering across neocortex. *Proceedings of the National Academy of Sciences*, 116(42):21219–21227.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., İlhan Polat, Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and the SciPy 1.0 Contributors (2020). Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, pages 1–12.

von Economo, C. F. and Koskinas, G. N. (1925). *Die cytoarchitektonik der hirnrinde des erwachsenen menschen*. J. Springer.

Vos de Wael, R., Benkarim, O., Paquola, C., Lariviere, S., Royer, J., Tavakol, S., Xu, T., Hong, S.-J., Langs, G., Valk, S., Mišić, B., Milham, M., Margulies, D., Smallwood, J., and Bernhardt, B. C. (2020). BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. *Communications Biology*, 3(1):1–10.

Wagner, H. H. and Dray, S. (2015). Generating spatially constrained null models for irregularly spaced data using Moran spectral randomization methods. *Methods in Ecology and Evolution*, 6(10):1169–1178.

Wang, P., Kong, R., Kong, X., Liégeois, R., Orban, C., Deco, G., Van Den Heuvel, M. P., and Yeo, B. T. (2019). Inversion of a large-scale circuit model reveals a cortical hierarchy in the dynamic resting human brain. *Science Advances*, 5(1):eaat7854.

Waskom, M., Botvinnik, O., Ostblom, J., Gelbart, M., Lukauskas, S., Hobson, P., Gemperline, D. C., Augspurger, T., Halchenko, Y., Cole, J. B., Warmenhoven, J., de Ruiter, J., Pye, C., Hoyer, S., Vanderplas, J., Villalba, S., Kunter, G., Quintero, E., Bachant, P., Martin, M., Meyer, K., Swain, C., Miles, A., Brunner, T., O'Kane, D., Yarkoni, T., Williams, M. L., Evans, C., and Fitzgerald, C. (2020a). mwaskom/seaborn. *Zenodo*, doi:10.5281/zenodo.592845.

Waskom, M., Larson, E., Brodbeck, C., Gramfort, A., Burns, S., Luessi, M., Weidemann, C. T., Bitzer, S., Markiewicz, C., LaPlante, R., Halchenko, Y., Engemann, D. A., van Vliet, M., Ghosh, S., Klein, N., Piantoni, G., Brett, M., Gwilliams, L., King, J.-R., and Liu, D. (2020b). nipy/pysurfer. *Zenodo*, doi:10.5281/zenodo.592515.

Westfall, P. H. and Young, S. S. (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*, volume 279. John Wiley & Sons.

Whitaker, K. J., Vértes, P. E., Romero-Garcia, R., Váša, F., Moutoussis, M., Prabhu, G., Weiskopf, N., Callaghan, M. F., Wagstyl, K., Rittman, T., Tait, R., Ooi, C., Suckling, J., Inkster, B., Fonagy, P., Dolan, R. J., Jones, P. B., Goodyer, I. M., the NSPN Consortium, and Bullmore, E. T. (2016). Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. *Proceedings of the National Academy of Sciences*, 113(32):9105–9110.

Witt, S. T., van Ettinger-Veenstra, H., Salo, T., Riedel, M. C., and Laird, A. R. (2020). What executive function network is that? An image-based meta-analysis of network labels. *bioRxiv*.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8):665.

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., and Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3):1125.

Zilles, K. and Amunts, K. (2009). Receptor mapping: architecture of the human cerebral cortex. *Current Opinion in Neurology*, 22(4):331–339.

Zilles, K., Palomero-Gallagher, N., and Schleicher, A. (2004). Transmitter receptors and functional anatomy of the cerebral cortex. *Journal of Anatomy*, 205(6):417–432.

TABLE S1. **List of terms used in NeuroSynth analyses |** Terms that overlapped between NeuroSynth (Yarkoni et al. 2011) and Cognitive Atlas (Poldrack et al. 2011) corpuses used in the reported analyses. For a machine-readable format please refer to https://github.com/netneurolab/markello_spatialnulls.

action	eating	insight	naming	semantic memory
adaptation	efficiency	integration	navigation	sentence comprehension
addiction	effort	intelligence	object recognition	skill
anticipation	emotion	intention	pain	sleep
anxiety	emotion regulation	interference	perception	social cognition
arousal	empathy	judgment	planning	spatial attention
association	encoding	knowledge	priming	speech perception
attention	episodic memory	language	psychosis	speech production
autobiographical memory	expectancy	language comprehension	reading	strategy
balance	expertise	learning	reasoning	strength
belief	extinction	listening	recall	stress
categorization	face recognition	localization	recognition	sustained attention
cognitive control	facial expression	loss	rehearsal	task difficulty
communication	familiarity	maintenance	reinforcement learning	thought
competition	fear	manipulation	response inhibition	uncertainty
concept	fixation	meaning	response selection	updating
consciousness	focus	memory	retention	utility
consolidation	gaze	memory retrieval	retrieval	valence
context	goal	mental imagery	reward anticipation	verbal fluency
coordination	hyperactivity	monitoring	rhythm	visual attention
decision	imagery	mood	risk	visual perception
decision making	impulsivity	morphology	rule	word recognition
detection	induction	motor control	salience	working memory
discrimination	inference	movement	search	
distraction	inhibition	multisensory	selective attention	

TABLE S2. **Null framework limitations |** Overview of some of the drawbacks of each of the null frameworks described in the current report. Each of the families also suffer from drawbacks impacting all constituent null frameworks (i.e., spatial permutation models can only handle surface data; parameterized models tend to be more computationally complex); refer to *Discussion* for more information.

Naive models	
Method	*Drawbacks*
Parametric	Assumes errors are *i.i.d.*; does not account for spatial structure
Non-parametric	Does not account for spatial structure
Spatial permutation models	
Method	*Drawbacks*
Vázquez-Rodríguez	Allows duplicate parcel assignments
Váša	Generates sub-optimal assignments with respect to global cost
Hungarian	Undesirable, high-cost (i.e., distant) assignments can be made to optimize global cost; time consuming for high-resolution atlases
Baum	Projection to vertices involves upsampling; allows duplicate parcel assignments
Cornblath	Projection to vertices involves upsampling; null data will not match original due to re-averaging
Parameterized data models	
Method	*Drawbacks*
Burt-2018	Assumes exponential distance model provides accurate estimation of $\hat{\rho}$ and $\hat{d}_0$ from data
Burt-2020	Surrogate data can be unbalanced between hemispheres; variogram models may not properly model some spatial distributions
Moran	Dependent on provided weight matrix, $\mathbf{W}$, which varies based on neighborhood size; tends to yield highly auto-correlated surrogates

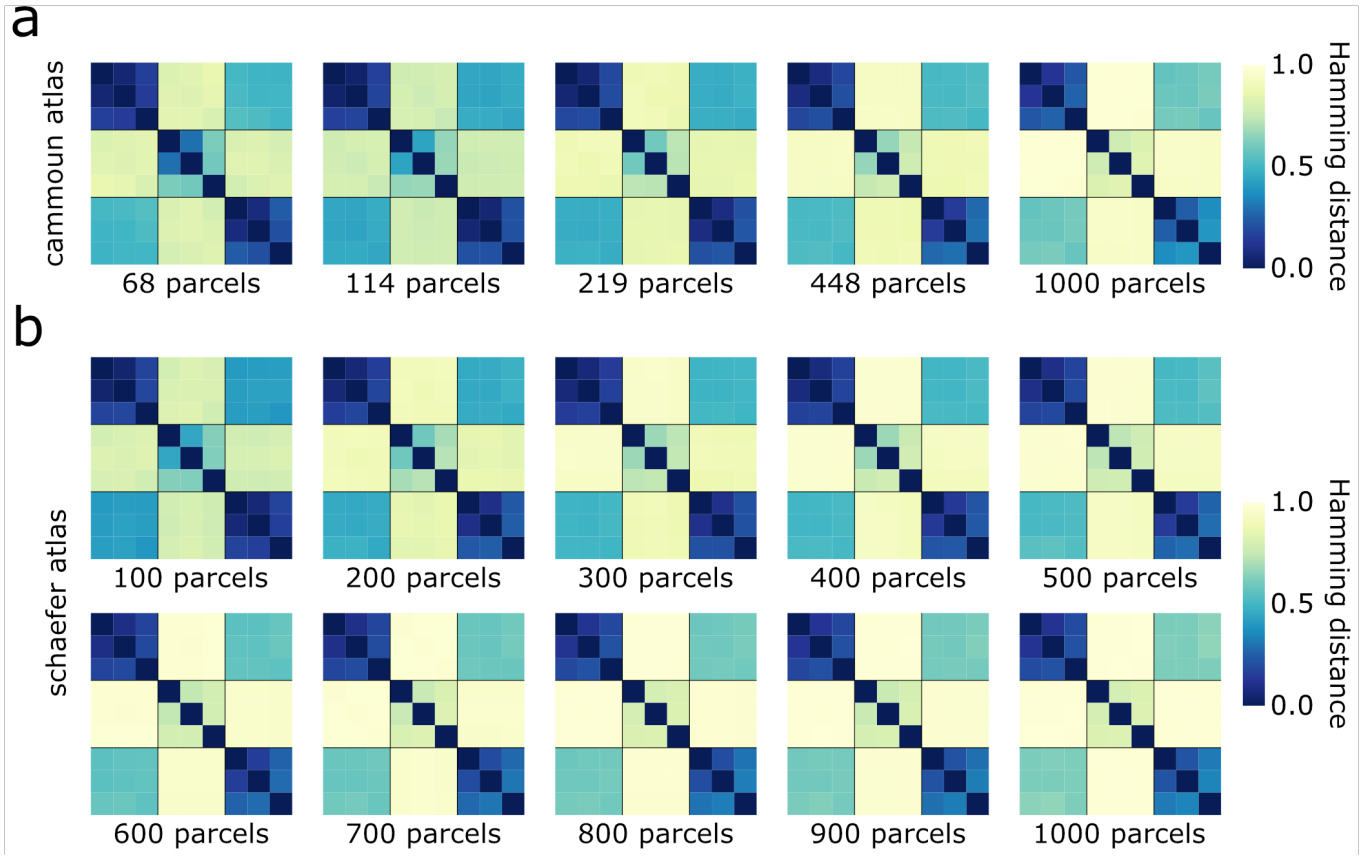

Figure S1. **Variability in parcel centroid definition |** Reproduction of results shown in Fig. 5a for all resolutions of the (a) Cammoun and (b) Schaefer atlases. Black lines delineate different null frameworks (*Vázquez-Rodríguez*, *Váša*, *Hungarian* from top-to-bottom) and cells within each section of the heatmap represent different parcel centroid definition methods (*surface*, *average*, *geodesic* from top-to-bottom). Refer to *Methods and Materials: Null model implementation variability* for more information.

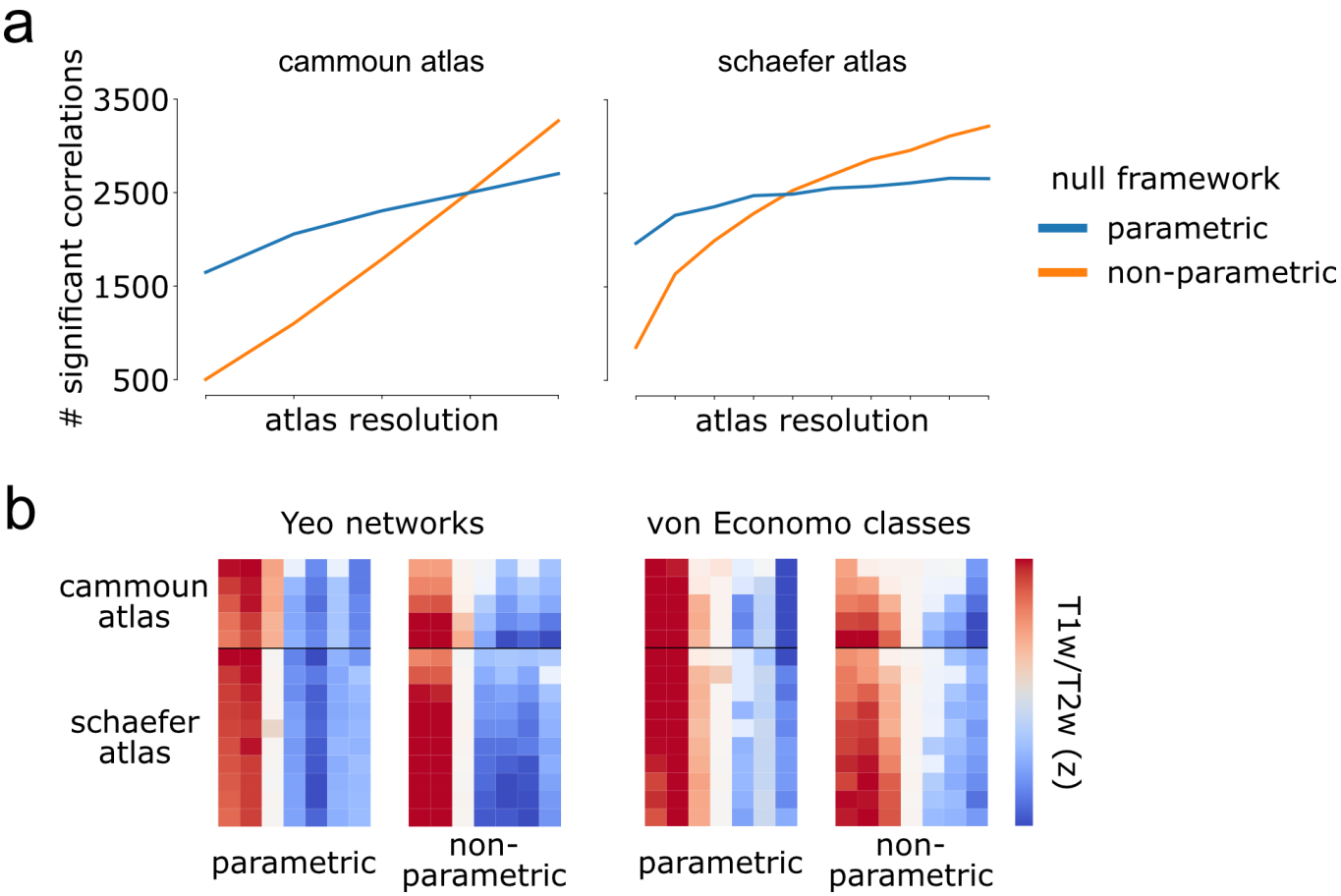

Figure S2. **Naive null models |** (a) Results of the spatially-naive null models applied to the NeuroSynth analyses described in *Results: Testing correspondence between brain maps*; refer to Fig. 2 for description of plots. Note the different scale of the y-axis as compared to Fig. 2d. (b) Results of the spatially-naive null models applied to the T1w/T2w analyses described in *Results: Testing partition specificity*; refer to Fig. 4 for description of heatmaps. Color scale for *parametric* nulls ranges from -1 to 1 and represents the average T1w/T2w ratio in each network. Color scale for *non-parametric* nulls ranges from -7.5 to 7.5 and represents the z-scored T1w/T2w ratio in each network.