Statistical testing and power analysis for brain-wide association study

Weikang Gong[1,2], Lin Wan[3], Wenlian Lu[4,5], Liang Ma[6], Fan Cheng[4,5], Wei Cheng[4,5], Stefan Gruenewald[1,2], and Jianfeng Feng[4,5,7,†]

[1] *CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China*

[2] *University of Chinese Academy of Sciences, Beijing 100049, China*

[3] *Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China*

[4] *Centre for Computational Systems Biology, School of Mathematical Sciences, Fudan University, Shanghai 200433, China*

[5] *Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 200433, China*

[6] *Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China*

[7] *Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK*

[†] *Correspondence author: jianfeng64@gmail.com*

Abstract

Brain-wide association study (BWAS) is analogous to the successful genome-wide association study (GWAS) in the genetics field. It aims to identify the voxel-wise functional connectome variations associated with complex traits. Although it has been applied to several mental disorders, such as schizophrenia [12], autism [13] and depression [14], its statistical foundations are still lacking. Therefore, we herein report the development of a rigorous statistical framework for link-wise significance testing and theoretical power analysis based on the random field theory. Peak- and cluster-level inferences are generalized to analyze functional connectivities. A novel method to identify phenotype associated voxels based on functional connectivity pattern is also proposed. Our method reduces the computational complexity of permutation-based approach in controlling the false positive rate and provides robust and reproducible findings in several real datasets, such as the 1000 Functional Connectomes Project (1000 FCP), Autism Brain Imaging Data Exchange (ABIDE), Center for Biomedical Research Excellence (COBRE) and others.

Keywords: brain-wide association study, random field theory, functional connectivity, statistical power

1 Introduction

Characterizing the effect of brain structure and functional variations on complex phenotypes is a challenging goal for modern neuroscientists. In the past few years, with the development of neuroimaging

technology and an increasing number of publicly available datasets, such as the 1000 Functional Connectomes Project (FCP) and Human Connectome Project (HCP), large-scale, image-based association studies become possible and will improve our understanding of human brain functions.

In the genetics field, the genome-wide association study (GWAS) has successfully identified many phenotype-associated single nucleotide polymorphisms (SNPs) or genes. In the literature, most of these studies use single-locus association tests, in which each SNP is tested individually for its association with complex traits. Borrowing ideas from GWAS, the brain-wide association study (BWAS) [12, 13, 14] also uses a frequentist significance testing approach to associate each *functional connectivity* with complex traits. However, different from most of the literatures which use prior knowledge of brain region segmentation, BWAS uses the *voxels* as nodes to define the brain functional network, which is fully unbiased and data-driven. However, to conduct a systematic, well-powered BWAS, many methodological and technical issues should be addressed. First, compared with GWAS, many more statistical tests are performed in BWAS, therefore, a more stringent multiple comparison threshold should be provided. For example, a typical GWAS requires nearly 10^6 statistical tests. In comparison, a link-based BWAS requires more than 10^9 statistical tests on 3mm resolution functional MRI data consisting approximately 50000 voxels. Second, the statistical tests are dependent on the spatial structure of the fMRI data. Therefore, popular methods, such as Bonferroni correction or false discovery rate (FDR) [7], are not suited to determine the significance threshold. Third, high quality GWAS usually requires tens of thousands of samples to reach adequate statistical power, but previous BWAS on schizophrenia, autism and depression only have sample sizes of less than one thousand [12, 13, 14]. Therefore, compared with GWAS, it is natural to ask if BWAS, which is based on such limited sample size, can tolerate a larger number of hypothesis tests. If not, then the number of samples must be setted. This paper will address these questions.

Most existing connectome analysis methods are designed for region-of-interest (ROI) level studies (see [46] or [30] for comprehensive reviews). The following steps summarize these methods. First, the whole brain is segmented into several subregions, which can be either spatially continuous by anatomical segmentation (e.g. Automated Anatomical Labeling (AAL) templates [37]) or spatially discontinuous by data-driven segmentation (e.g. Independent Component Analysis (ICA) templates [6]). Second, for every subject, the correlation of blood-oxygen-level dependent (BOLD) signal between pairwise subregions is estimated. Third, the statistical model is fitted and the significance tests performed on each of the functional connectivities. Fourth, a multiple comparison correction is performed and a list of significant functional connectivities obtained. The statistical model of BWAS is quite similar to this series of steps. However, when the functional connectivities are defined on pairwise voxels, both the multiple comparison threshold and the power analysis approaches should account for the spatial structure of data.

The random field theory (RFT) is an important statistical tool in brain image analysis, and it has been widely used in the analysis of task fMRI data and structure data [4]. Statistical parametric maps (SPMs) are usually modeled as a discrete sampling of smooth Gaussian or related random fields [36]. The random field theory can control the family-wise error rate (FWER) of multiple testing by evaluating whether the observed test statistic, or the spatial extent of clusters exceeding a cluster-defining threshold (CDT), is large by chance, which is known as peak-level and cluster-level inference respectively. Since Adler's early work on the geometry of random field [1, 2], theoretical results for different types of random fields have been obtained, such as the Gaussian random field [24, 44], the

t, χ^2, F random field [41, 9], the multivariate random field [40], the correlation random field [10] and the corresponding Fisher's Z transformation [11]. However, the above results can not be directly used in BWAS, because the SPM of BWAS is not a strict Gaussian random field [11]. We propose to use a smoothness adjustment method to account for this problem. In this paper, a method based on Gaussian random field theory is developed for BWAS.

Power analysis for BWAS is also challenging by the billions of statistical tests performed among spatially correlated functional connectivities. Most existing power analysis methods are designed for task fMRI data analysis, including, for example, the simulation based method [17], the non-central distribution based method [33], and the method based on non-central random field theory (ncRFT) [26]. Friston et al model the signals as a smoothed Gaussian random field with larger variance to estimate power [24, 21, 22]. Among them, the ncRFT-based method can both take into account the spatial structure of fMRI data and avoid time consuming simulation. Therefore, to analyze the power of BWAS, we adopted a methodology similar to that of the ncRFT-based method [26]. The signals at functional connectivities are modelled as a six-dimensional non-central Gaussian random field, and the power is estimated by a modified Gaussian random field theory.

The brain-wide examination of individual functional connectivities can identify coordinated phenotype -associated neural activity. Additional information may be gained from learning whether a voxel is associated with the complex traits. A voxel can be characterized by different kinds of features, including, for example, its structure (e.g. grey or white matter volume), its temporal variability [47] and its connectivity pattern (e.g. Measure of Association (MA) [12, 13, 14]). In BWAS, we focus on a voxel's connectivity pattern and its association with complex traits. This kind of analysis evaluate the simultaneous contribution of a set of functional connectivities to complex traits. Especially, this set of functional connectivities connect the same voxel. The connectivity pattern is a high dimensional feature, thus two types of methods can be used to test its association with complex traits. The first type of methods is known as the summary-statistics method. This method first tests each functional connectivity individually. It then determines whether the statistics of functional connectivities are enriched for association signals [12, 13, 14]. The second type of methods directly applies multivariate statistical tests, such as the multivariate distance matrix regression method used in connectome-wide association study (CWAS) [38]. A common problem of existing methods is that the time-consuming permutation approach is required to access the significance of their test statistics. To address this problem, a computationally efficient method is proposed based on principal component analysis (PCA).

As shown in figure 1, we first propose to develop a method for link-based BWAS to address the multiple comparison problem arised in link-wise significance testing. This method is based on Gaussian random field theory, which generalizes the peak- and cluster-level inferences to analyze functional connectivities. To test whether the proposed theory can control the family-wise error rate, both null simulations and large-scale permutations are performed using resting-state fMRI data from the 1000 Functional Connectomes Project (1000 FCP) [8], the Center for Biomedical Research Excellence (COBRE) and a Taiwan dataset [32]. Second, to characterize the voxel's functional connectivity pattern and access its association with complex traits, a voxel-wise, PCA-based model is developed for voxel-based BWAS. This model was demonstrated to provide robust and reproducible findings in the Autism Brain Imaging Data Exchange (ABIDE) dataset [18]. Third, a modified Gaussian random field theory is developed to explicitly approximate the power of peak-level inference. For brain-wide association studies on schizophrenia, autism and depression [12, 13, 14], our theoretical

analysis shows that hundreds of samples are sufficient to reach 90% power to detect at least one signal under different smoothness. The software package for BWAS can be downloaded at https://github.com/weikanggong/BWAS.

2 Material and Methods

This section consists of three parts. In Section 2.1, we describe link-based BWAS. First, the statistical model is introduced. Then, inference methods based on random field theory, including peak- and cluster-level inferences, are proposed. Finally, simulation and permutation strategies are illustrated to evaluate the proposed theory on false positive rate control. In Section 2.2, we put forward voxel-based BWAS, including the details of the model, another three statistics to be compared, and the method of evaluating its robustness and reproducibility. In Section 2.3, both the theoretical power analysis method and the simulation-based power analysis method are presented.

2.1 Link-based BWAS

2.1.1 Model

In link-based BWAS, the individual functional network is constructed by calculating the Pearson correlation coefficients between every pair of voxel time series. Let m be the number of voxels, s be the subject, and $R^{(s)} = [r_{ij}^{(s)}]_{m\times m}$ be the $m \times m$ functional network matrix for subject s. Each of the element of $R^{(s)}$ is the correlation coefficient between voxel time series i and j for subject s. An element-wise Fisher's Z transformation is then applied as $Z^{(s)} = [z_{ij}^{(s)}]_{m\times m} = [\frac{1}{2}\log(\frac{1+r_{ij}^{(s)}}{1-r_{ij}^{(s)}})]_{m\times m}$, so that $z_{ij}^{(s)}$ will approximate a normal distribution. For every functional connectivity, a general linear model (GLM) is fitted:

$$Y_{ij} = XB_{ij} + \epsilon_{ij}$$

where, $Y_{ij} = (z_{ij}^{(1)}, z_{ij}^{(2)}, \ldots, z_{ij}^{(n)})$ is an $n \times 1$ vector of functional connectivities between voxel i and j across n subjects, X is the common $n \times q$ design matrix, $B_{ij} = (\beta_{ij}^1, \beta_{ij}^2, \ldots, \beta_{ij}^q)$ is a $q \times 1$ vector of regression coefficients, and ϵ_{ij} is a $n \times 1$ vector of random error, which is assumed to be an independent and identically distributed Gaussian random variable $N(0, \sigma_{ij}^2)$. The ordinary least square estimator for B_{ij} is $\hat{B}_{ij} = (X'X)^{-1}X'Y_{ij}$, and for σ_{ij}^2 is $\hat{\sigma}_{ij}^2 = (Y_{ij} - X\hat{B}_{ij})'(Y_{ij} - X\hat{B}_{ij})/(n-q)$. A Student's t-statistics at functional connectivity between voxel i and j can be expressed as:

$$T_{ij} = \frac{\mathbf{c}\hat{B}_{ij}}{(\mathbf{c}(X'X)^{-1}\mathbf{c}'\hat{\sigma}_{ij}^2)^{\frac{1}{2}}}$$

where $\mathbf{c}$ is a $1 \times q$ contrast vector. In link-based BWAS, let β_{ij}^1 be the primary variable of interest, and $\beta_{ij}^2, \ldots, \beta_{ij}^q$ be the nuisance covariates included in the regression model. The contrast $\mathbf{c} = (1, 0, \ldots, 0)$ will be used to test the hypothesis $\beta_{ij}^1 = 0$, and the statistics T_{ij} will reflect the significance of the primary variable. Other contrasts can also be used depending on the study design.

In order to use the random field theory result, the Student's t random variable at each functional connectivity is transformed to a Gaussian random variable. This is achieved by either transforming

T-statistics to p-values and then to Z-statistics, or pooling the variance across the whole data [42]. The normal transformation makes the $\hat{\sigma}_{ij}^2$ fixed as $\hat{\sigma}_0^2$ [42, 44]. Therefore, the test statistic becomes

$$\begin{aligned} Z_{ij} &= \frac{\mathbf{c}\hat{B}_{ij}}{\sqrt{\mathbf{c}(X'X)^{-1}\mathbf{c}'\hat{\sigma}_0^2}} \\ &= \frac{\mathbf{c}(X'X)^{-1}X'Y_{ij}}{\sqrt{\mathbf{c}(X'X)^{-1}\mathbf{c}'\hat{\sigma}_0^2}} \\ &= \sum_{s=1}^{n} w^{(s)} z_{ij}^{(s)} \end{aligned}$$

where $w^{(s)}$ is the s-th element of row vector $\frac{\mathbf{c}(X'X)^{-1}X'}{\sqrt{\mathbf{c}(X'X)^{-1}\mathbf{c}'\hat{\sigma}_0^2}}$, which only depends on the subjects . It is useful to write Z_{ij} in this form in the theoretical derivation (see the next section and Appendix for details).

2.1.2 Peak-level inference

The aim of peak-level inference is to control the FWER of multiple tests, or equivalently, to control the probability of finding at least one false positive signal. This is directly related to the maximum distribution of the $\frac{m(m-1)}{2}$ test statistics. In this section, we will derive a formula to approximate it.

Let $M^{(s)}(p) = (M_1^{(s)}(p), \ldots, M_{v^{(s)}}^{(s)}(p))'$, $p \in \mathcal{P} \subset \mathbb{R}^3$ and $N^{(s)}(q) = (N_1^{(s)}(q), \ldots, N_{v^{(s)}}^{(s)}(q))'$, $q \in \mathcal{Q} \subset \mathbb{R}^3$ be two vectors of $v^{(s)}$ independent and homogeneous Gaussian random fields with mean zeros and variance one. The index s denotes subjects, and the $v^{(s)}$ can be treated as the number of time points, while p, q are the coordinates of three-dimensional Euclidean space. The six-dimensional cross-correlation random field $R^{(s)}(p,q)$ is defined as follows [10]:

$$R^{(s)}(p,q) = \frac{M^{(s)}(p)'N^{(s)}(q)}{\sqrt{M^{(s)}(p)'M^{(s)}(p)N^{(s)}(q)'N^{(s)}(q)}}$$

In link-based BWAS, the cross-correlation field is generated by calculating sample correlation coefficients between pairwise voxel time series. Next, the element-wise Fisher's Z transformation transforms this cross-correlation random field to a six-dimensional 'Gaussianized' random field as:

$$Z^{(s)}(p,q) = \frac{1}{2}\log\left[\frac{1+R^{(s)}(p,q)}{1-R^{(s)}(p,q)}\right]$$

It has mean zero and variance $\frac{1}{v^{(s)}-3}$ [28]. Our test statistic $Z_{ij}(p,q)$ forms a weighted sum of Fisher's Z transformed cross-correlation random field $Z(p,q)$ as:

$$Z(p,q) = \sum_{s=1}^{n} w^{(s)} Z^{(s)}(p,q)$$

The random field $Z(p,q)$ is a 'Gaussianized' random field with mean zero and variance one.

The expected Euler characteristic (EC) of the excursion set of random field is used to approximate the maximum distribution of the random field $Z(p,q)$ at high threshold z_0 [2, 44]:

$$\begin{aligned} P(\max\ Z(p,q) > z_0) \approx \mathbb{E}(EC) &= \sum_{d=0}^{6} \mu_d(\mathcal{P}\times\mathcal{Q})\rho_d^Z(z_0) \\ &= \sum_{i=0}^{3}\sum_{j=0}^{3} \mu_i(\mathcal{P})\mu_j(\mathcal{Q})\rho_{i+j}^Z(z_0) \end{aligned} \tag{1}$$

where $\mu_d(\cdot)$ is the d-th dimensional intrinsic volume of the random field, and $\rho_d^Z(u_0)$ is the d-th dimensional EC-density for the Gaussian random field at threshold z_0. The method for calculating $\mu_d(\cdot)$ and $\rho_d^Z(z_0)$ is illustrated in the Appendix. This formula takes into account both the number of hypothesis tests tested and the spatial structure of the tests. Intuitively, $\mu_d(\cdot)$ is a function of the Lebesgue measure of the search region spanned by $Z(p,q)$, which is directly related to the number of statistical tests. The EC-density $\rho_d^Z(z_0)$ is a function of the variance-covariance matrix of the partial derivative of $Z(p,q)$, which measures the degree of smoothness of the random field. In link-based BWAS, the following formula for the Gaussian random field is used to calculate $\rho_d^Z(z_0)$ [2]:

$$\rho_d^Z(z_0) = (2\pi)^{-\frac{d+1}{2}}|\Lambda|^{\frac{d}{2D}}e^{-\frac{z_0^2}{2}}\sum_{j=0}^{\lfloor\frac{d-1}{2}\rfloor}(-1)^j\frac{(2j)!}{j!2^j}\binom{d-1}{2j}z_0^{d-1-2j} \tag{2}$$

where D is the highest dimension of $Z(p,q)$ ($D=6$, in our case). The $|\Lambda| = |\mathrm{Var}(\dot{Z}(p,q))|$ is the determinant of the variance-covariance matrix of the partial derivative of $Z(p,q)$.

The $|\Lambda|$ in formula (2) can be replaced by FWHM_Z, the Full Width at Half Maximum (FWHM) of the random field Z averaged across six dimensions, using the equation:

$$\mathrm{FWHM}_Z = (4\log 2)^{\frac{1}{2}}|\Lambda|^{-\frac{1}{2D}} \tag{3}$$

and FWHM_Z is a corrected smoothness parameter, which can be calculated as:

$$\mathrm{FWHM}_Z = \left(\sum_{s=1}^{n}\frac{(w^{(s)})^2}{v^{(s)}-3}\mathrm{FWHM}_{M^{(s)}}^{-2}\right)^{-\frac{1}{4}}\left(\sum_{s=1}^{n}\frac{(w^{(s)})^2}{v^{(s)}-3}\mathrm{FWHM}_{N^{(s)}}^{-2}\right)^{-\frac{1}{4}} \tag{4}$$

where $\mathrm{FWHM}_{M^{(s)}}$ and $\mathrm{FWHM}_{N^{(s)}}$ are the average FWHM of the random field vectors $M^{(s)}(p)$ and $N^{(s)}(q)$ across three dimensions.

Accordingly, formula (1) can be calculated by:

$$\begin{aligned} P(\max\ Z(p,q) > z_0) \approx &\sum_{i=0}^{3}\sum_{j=0}^{3}\mu_i(\mathcal{P})\mu_j(\mathcal{Q})\frac{(2\pi)^{-\frac{i+j+1}{2}}(4\log 2)^{\frac{i+j}{2}}}{\mathrm{FWHM}_Z^{i+j}} \\ &\times e^{-\frac{z_0^2}{2}}\sum_{k=0}^{\lfloor\frac{i+j-1}{2}\rfloor}(-1)^k\frac{(2k)!}{k!2^k}\binom{i+j-1}{2k}z_0^{d-1-2k} \end{aligned} \tag{5}$$

FWHM$_Z$, calculated by equation (4), is a function of the number of time points $v^{(s)}$ and the FWHM$_{M^{(s)}}$ and FWHM$_{N^{(s)}}$ of the individual fMRI data. Typically, the length of scanning time and image smoothness is the same for every subject in a study. Denoting them as v and FWHM, the equation (4) is then reduced to:

$$\begin{aligned}
\text{FWHM}_Z &= \left(\sum_{s=1}^{n} \frac{(w^{(s)})^2}{v-3}\text{FWHM}^{-2}\right)^{-\frac{1}{4}} \left(\sum_{s=1}^{n} \frac{(w^{(s)})^2}{v-3}\text{FWHM}^{-2}\right)^{-\frac{1}{4}} \\
&= \text{FWHM}\sqrt{v-3}\left[\left(\frac{\mathbf{c}(X'X)^{-1}X'}{\sqrt{\mathbf{c}(X'X)\mathbf{c}'\hat{\sigma}_0^2}}\right)\left(\frac{\mathbf{c}(X'X)^{-1}X'}{\sqrt{\mathbf{c}(X'X)\mathbf{c}'\hat{\sigma}_0^2}}\right)'\right]^{-\frac{1}{2}} \\
&= \text{FWHM}\sqrt{(v-3)\sigma_0^2} \\
&= \text{FWHM}
\end{aligned}$$

where we treat the sample variance σ_0^2 as the theoretical variance $\frac{1}{v-3}$. This suggests that the smoothness of random field $Z(p,q)$ equals the original image smoothness, and that the scanning time does not influence the formula (5).

2.1.3 Cluster-level inference

Cluster-level inference is also a widely used approach in brain image analysis. Here, inference is based on the observed cluster size exceeding certain cluster-defining threshold [24]. We are usually interested in whether the observed cluster size is large, i.e., where the size is on the upper tail of the distribution of maximum cluster size under the null hypothesis. We first define functional connectivity (FC) cluster, then derive a formula to approximate the maximum cluster size distribution to control the cluster-wise FWER.

Voxel-cluster is usually defined as a set of spatially connected voxels based on the d-connectivity scheme, where d is usually 6, 18 or 26 in brain image analysis. For FC-cluster, Spatial pairwise clustering (SPC) statistic is a suitable definition [46]. Suppose that there are m^* FCs with endpoints denoted as:

$$(i_1, j_1), (i_2, j_2), \ldots, (i_{m^*}, j_{m^*})$$

These m^* FCs form a FC-cluster if voxels $(i_1, i_2, \ldots, i_{m^*})$ form a voxel-cluster and $(j_1, j_2, \ldots, j_{m^*})$ form a voxel-cluster. Note that $(i_1, i_2, \ldots, i_{m^*})$ and $(j_1, j_2, \ldots, j_{m^*})$ can either be the same set of voxels, or different sets of voxels. An example is shown in Figure 2, where there are five voxel-clusters A, B, C, D, E in a two-dimensional image. The FCs between AB, BC and AD are different FC-clusters, and FCs within voxel-cluster E also form a FC-cluster.

An algorithm for finding FC-clusters can be implemented as follows. First, the voxel-clusters formed by the endpoints of a set of FCs are found. Second, each of the FCs is assigned to one of pairwise voxel-clusters iteratively. After all the iterations, FCs between the same pairs of voxel-clusters form a FC-cluster.

We utilize Gaussian random field theory to approximate the null distribution of maximum FC-cluster size. In brief, let M be the number of FCs exceeding CDT z_0, N be the number of FC-clusters,

and S be the FC-cluster size. Suppose that separate FC-clusters are independent, then the distribution of maximum cluster size S_{max} for Gaussian random field is [1, 24]:

$$P(S_{max} > s) = 1 - \exp\left[-\mathbb{E}(N)P(S > s)\right]$$

The expected number of FC-clusters $\mathbb{E}(N)$ at high CDT z_0 can be approximated by the expected EC of Gaussian random field using equation (5):

$$\mathbb{E}(N) \approx \mathbb{E}(EC) = \sum_{i=0}^{3}\sum_{j=0}^{3} \mu_i(\mathcal{P})\mu_j(\mathcal{Q})\rho^Z_{i+j}(z_0)$$

The distribution of S can be approximated by [1, 35]:

$$P(S > s) = \exp\left[-\left(\frac{\Gamma(D/2+1)\mathbb{E}(N)s}{\mathbb{E}(M)}\right)^{2/D}\right]$$

where D=6 in our case, and

$$\mathbb{E}(M) = \frac{m(m-1)}{2}[1 - \Phi(z_0)]$$

where m is the number of voxels, and $\Phi(\bullet)$ is the cumulative distribution function of standard normal distribution.

The above theory is a generalization of three-dimensional result [24, 25], except that we use the six-dimensional 'Gaussianized' random field result derived in the previous section to approximate the expected number of clusters $\mathbb{E}(N)$. For a complete overview of the methodology, see [24] and [25] as examples. Thus, small-sized FC-clusters are more likely to be identified as false positives and filtered out by our method, e.g. the red links in Figure 2. Cluster-level inference provides evidence of an experimental effect by evaluating whether there is a large number of FCs between two voxel-clusters survives the cluster-defining threshold. Peak-level inference can reflect the effect of single FC, which is sensitive to local intense signals, cluster-level inference can reflect the simultaneous effect of a set of FCs, which is sensitive to spatial signals.

2.1.4 Simulation study

We generate null data with relatively small sizes (30 voxels per dimension) to evaluate whether the proposed random field theory can actually control the FWER in both peak- and cluster-level inferences. Previous study have shown that the random field theory tends to perform better when the search region becomes larger [44].

The synthetic null fMRI data are generated in four steps. First, we generate two sets of 10000 three-dimensional independent Gaussian white noise images, with 30 voxels per dimension. Second, the images are smoothed with different Gaussian kernels (FWHM = 1, 2, ..., 6 voxels). Third, the balls with radius r $(5, 5.5, \ldots, 9.5$ voxels) centered at the cube are extracted. This guarantees the uniform smoothness of the images. Fourth, every 20 images are combined to form 500 simulated four-dimensional fMRI data. We denote the images in the first set as $(A_1, A_2, \ldots, A_{500})$ and the images in the second set as $(B_1, B_2, \ldots, B_{500})$.

The empirical distribution of maximum statistics Z_{max} and maximum cluster size N_{max} is generated by the following steps. First, the Pearson correlation coefficients are calculated between every pair of voxel time series of images A_i and B_i, and a Fisher's Z transformation is performed. Second, two groups of images from two sets are randomly selected, each group consists of 450 samples. A Z-map is generated by fitting each functional connectivity to a general linear model to compare two groups. Third, The second step is repeated 5000 times. Each time, for peak-level inference, the maximum Z statistic is recorded, and for cluster-level inference, the maximum cluster size exceeding CDT (4, 4.5, or 5 in our analysis) is recorded. This steps form the empirical distributions of the maximum statistics Z_{max} and maximum cluster size N_{max} separately. Fourth, the FWER threshold at α is estimated as the upper α quantiles of the empirical distribution.

2.1.5 Real data permutation study

To evaluate whether the random field theory can actually control the FWER in real data analysis, we perform permutation tests using resting-state fMRI data in three datasets: (1) 197 samples (normal people) from the Cambridge dataset in 1000 FCP; (2) 120 samples (67 normal people and 53 chronic schizophrenia patients) from the COBRE dataset; (3) 387 samples (241 normal people and 146 chronic schizophrenia patients) from the Taiwan dataset. All data are masked by a 3mm resolution AAL template, and 47636 voxels within 90 cerebrum regions are extracted. The subjects are then randomly divided into two groups with equal sample sizes, and link-based BWAS is performed 500 times on each random dataset. The maximum statistic Z_{max} and maximum cluster size N_{max} are recorded at each random permutation. Since the group labels are randomly permuted, the expected number of the false positive peaks or clusters n_{fp} should be:

$$n_{fp} = 500\alpha$$

where α is the FWER we want to control, say 0.05. We compare n_{fp} with the result obtained by the random field theory to evaluate performance.

2.2 Voxel-based BWAS

2.2.1 Model

To characterize the voxel's functional connectivity pattern and access its association with phenotype, a voxel-based BWAS approach is proposed and analyzed. The model is based on voxel-wise principal component analysis and general linear model. Let X_i be an $n \times m$ functional connectivity matrix, with n as the number of subjects and m as the number of functional connectivities connecting voxel i across the whole brain. We assume that X_i has had nuisance covariates removed and is mean-centered by column. Let Y be the $n \times 1$ primary phenotype of interest. First, voxel-based BWAS performs the principal component analysis on each X_i. Next, we select the top k_i principal component based on a typical criteria, e.g. the percent of the explained total variance, or, alternatively, we use the distribution-based method [15]. Thus, for each voxel i, we have,

$$S_i = X_i L_i$$

where S_i is an $n \times k_i$ projected data matrix in the principal component space, with each of its columns representing a principal component, and L_i is an $m \times k_i$ loading matrix. S_i is a low-dimensional

representation of functional connectivity pattern of voxel i. Then, a general linear model is fitted to each S_i to test its association with phenotype as

$$f(Y) = \beta_{i0} + \beta_{i1}\text{PC}_{i1} + \cdots + \beta_{ik_i}\text{PC}_{ik_i} + \epsilon_i$$

where PC_{ij} is the j-th principal component of voxel i. In voxel-based BWAS, a multivariate linear model is used if Y is a continuous variable, and a logistic regression model is used if Y is a binary variable. Finally, a likelihood ratio test is performed to compare the full model with the constant model ($f(Y) = \beta_{i0} + \epsilon_i$), and a chi-square statistics with $n - k_i - 1$ degree of freedom is obtained.

2.2.2 Comparing voxel-based BWAS with summary-statistics method

We compare voxel-based BWAS with the following three summary statistics methods.

The first method, measure of association (MA), has been used in our previous brain-wide association studies [12, 13, 14]. For a voxel i, MA-statistic is defined as the number of significant functional connectivities connecting this voxel across the whole brain:

$$\text{MA}_i = \#\{|Z_{ij}| > z_0\}$$

where Z_{ij} is the Z statistic of a functional connectivity between voxel i and j, and z_0 is the multiple comparison threshold to declare significance. This MA-value has two drawbacks. Its null distribution can not be obtained easily, and it is sensitive to the choice of threshold.

The second method, sum of chi-square (SOCS), is a threshold-free statistics which has been widely used in gene/pathway-based GWAS. For a voxel i, SOCS-statistic is defined as [31]:

$$\text{SOCS}_i = \sum_{j=1}^{n} Z_{ij}^2$$

where n is the total number of functional connectivities connecting voxel i. Under the null hypothesis, it is subject to a weighted sum chi-square distribution with 1 degree of freedom ($\text{SOCS}_i \sim \sum_{j=1}^{n} \lambda_j \chi_1^2$), which can only be computed numerically.

The third method, global threshold-free cluster enhancement (gTFCE), borrows ideas from [39], which considers both the 'height' and 'size' of a voxel's functional connectivity map. For a voxel i, gTFCE-statistic is defined as:

$$\text{gTFCE}_i = \int_{-\infty}^{+\infty} s(z)^{0.5} z^2 dz$$

where z is the threshold and $s(z)$ is the number of the supra-threshold voxels. The powers 0.5 on $s(z)$ and 2 on z are derived in the original paper [39]. A large number of permutations are needed to obtain the null distribution of gTFCE.

2.2.3 Using meta-analysis to evaluate reproducibility

Meta-analysis is a widely used method to integrate results from multiple studies. Comparing results of different meta-analysies can evaluate whether the findings are reproducible. ABIDE is a large resting-state fMRI dataset consisting of autism patients with matched controls from multiple imaging sites.

After unified data preprocessing and quality control, a total of 1653 samples from 25 sites were used to evaluate the reproducibility of voxel-based BWAS.

We introduce the meta-analysis method for voxel-based BWAS. Within each site, voxel-based BWAS is performed. Fisher's method is used to combine the p-value SPMs of n_0 sites by,

$$\chi^2_{2n_0}(i) = -2\sum_{j=1}^{n_0}\log(p_{ij})$$

where p_{ij} is the p-value of voxel i at site j, and $\chi^2_{2n_0}(i)$ is a central chi-square variable with $2n_0$ degree of freedom under the null hypothesis.

The following steps are performed for evaluation. First, the voxel-based BWAS is performed separately within each of the 25 sites. The criterion for selecting a principal component is whether its explained variance is larger than the average. Second, the SPMs of randomly selected 12 randomly selected sites and the remaining 13 sites are combined by Fisher's method separately. The resulting chi-square maps are first converted to p-value maps, and then to a positive Z maps. Third, the proportion of overlap ρ is calculated as:

$$\rho = \frac{2\#\{\text{Map}_1 > Z_0 \ \&\ \text{Map}_2 > Z_0\}}{\#\{\text{Map}_1 > Z_0\} + \#\{\text{Map}_2 > Z_0\}}$$

where Map_1 and Map_2 are the SPMs of two meta-analysies, and Z_0 is the CDT (3, 3.5, 4 in our analysis). Step two and three are repeated 10000 times, and the empirical distributions of ρ under three different CDTs are generated.

2.3 Statistical power analysis

2.3.1 Theoretical analysis

In this section, we derive a formula to estimate the statistical power of peak-level inference. The power is defined as the probability of finding at least one true positive signal in a search region, with the false positive rate α controlled at a certain level [24]. To estimate the power, four parameters should be specified: (1) the threshold of controlling the FWER α; (2) the effect size of true signal γ; (3) the sample size n; (4) the smoothness of statistical map FWHM.

First, under the null hypothesis $H_0 : \beta^1_{ij} = 0$, suppose that the whole search region A is a Gaussian random field with mean zero and variance one. The threshold z_0 to control the FWER at α is obtained by the random field theory (5):

$$\alpha = P(\max_{(p,q)\in A} Z(p,q) > z_0 | H_0)$$

where (p, q) are the coordinates of the functional connectivities.

If we assume that the primary variable of interest, β^1_{ij}, is subject to a normal distribution $N(\mu_{ij}, \sigma^2_{ij})$ under the alternative hypothesis H_1, and that the sample size is n, then the test statistics Z_{ij} will subject to $N(\sqrt{n}\mu_{ij}/\sigma_{ij}, 1)$. The $\gamma_{ij} = \mu_{ij}/\sigma_{ij}$ is called effect size at FC_{ij}. We further assume that the distribution of signals will be the same in region B; that is, all β^1_{ij} is subject to the same normal

distribution $N(\mu, \sigma^2)$. Region B is a non-central Gaussian random field $Z^*(p,q)$ with mean $\sqrt{n}\gamma$ and variance one. The power in the search region $B \subset A$ can be expressed as:

$$\text{Power} = P(\max_{(p,q)\in B} Z^*(p,q) > z_0 | H_1)$$

The non-central Gaussian random field $Z^*(p,q)$ can be transformed to a central Gaussian random field by the following element-wise transformation:

$$Z(p,q) = Z^*(p,q) - \sqrt{n}\gamma$$

which allows the power in region B to be calculated using formula (5) by

$$\text{Power} = P(\max_{(p,q)\in B} Z(p,q) > z_0 - \sqrt{n}\gamma | H_0)$$

Three issues remain. The first involves selecting region B. When estimating power, we select region B as consisting of functional connectivities between two three-dimensional balls, each of ball has radius r equalling $\frac{\sqrt{2}}{2}\times$FWHM of the image. Thus, the signal is located in a six-dimensional ball whose radius equals the smoothness (FWHM) of the image (Figure 1 power analysis). The idea is similar to the non-central random field theory framework [26]. The matched filter theorem [43] suggests that the signal is best detected when the width of the smooth kernel matches the width of the signal.

The second issue involves the random field theory which can only approximate the right tail of the maximum distribution. Thus, the theory may provide an inaccurate estimation when the statistic $z_0 - \sqrt{n}\gamma$ is small. To address this problem, we propose to use the following modification:

$$\text{Power} = 1 - \exp\left[-P(\max_{(p,q)\in B} Z(p,q) > z_0 - \sqrt{n}\gamma | H_0)\right]$$

This formula ensures that the power is between zero and one, which shows excellent performance in the simulation.

The last issue concerns estimating the effect size, which is typically estimated from the statistical map of a pilot study using the same study design. Suppose that the pilot link-based BWAS study used $n^\star$ samples. Then, the estimated effect size at FC_{ij} is [27]:

$$\hat{\gamma}_{ij} = Z_{ij}/\sqrt{n^\star}$$

Using the above formula, the power can be estimated for each functional connectivity to form a power map on six-dimensional space. This is similar to the power map in the three-dimensional case [27], but it is quite difficult to visualize such maps. To report the power of a study, we estimate the effect size of every FC to form an empirical distribution. The power curves of different sample sizes and effect sizes under certain power (e.g. 90% power) are analyzed and reported.

2.3.2 Simulation study

The power simulation study is quite similar to the simulation procedures in peak-level inference. It aims to generate the empirical distribution of maximum statistics under the alternative hypothesis.

The following procedures are carried out. First, a ball whose radius r equals to $\frac{\sqrt{2}}{2}\times$FWHM is extracted from the two sets of simulated images $(A_1, A_2, \ldots, A_{500})$ and $(B_1, B_2, \ldots, B_{500})$. In our simulation, we choose to use FWHM $= 1, 2, \ldots, 6$ voxels. Second, the Pearson correlation coefficients between every pair of voxel time series of images A_i and B_i are calculated. A Fisher's Z transformation is then applied. Third, two groups of images are randomly selected, each consisting of n^* samples. An effect size is added to the second group of functional connectivities. Fourth, a general linear model is then fitted to each functional connectivity to compare two groups, and a non-central statistical map is generated. Fifth, step four is repeated 10000 times. The maximum statistics $Z_{max} = (Z_1, Z_2, \ldots, Z_{10000})$ are recorded and the effect size γ is estimated. The power is estimated by the empirical distribution Z_{max} as:

$$\text{Power}(n^*, \gamma) = \frac{\#\{\ Z_{max} > z_0\}}{10000}$$

where z_0 is the FWER threshold estimated by the central random field theory.

3 Results

Figure 1 shows a graphical overview of our proposed methods, including link-based BWAS, voxel-based BWAS and theoretical power analysis. The Manhattan plot (Figure 5) shows an example of link-based BWAS result, as visualized using 1653 samples of the ABIDE dataset. The $-\log_{10}$(p-values) of functional connectivities, which are grouped by the 94 cerebrum regions of AAL2 template, are plotted in this figure. Each functional connectivity can appear once or twice in this figure, depending on whether it connects different parts of the same region, or it connects two different brain regions, which is the most frequent case.

3.1 Performance of link-based BWAS

3.1.1 Simulation

The estimated FWERs of the peak-and cluster-level inference are the proportion of simulations with any significant signals declared by the random field theory. The details of data generation and evaluation methods are illustrated in the Material and Methods, Section 2.1.4. Figure 3 shows the results of controlling the FWER at 0.05. For peak-level inference, we find that it is conserved when the smoothness is low. Typically, it is more conserved than Bonferroni correction when the FWHM is smaller than 2 voxels. For cluster-level inference, we find that it can provide reliable results if FWHM is larger than 3 voxels and the CDT is high. When FWHM is low, the results are invalid.

3.1.2 Permutation test

We use three resting-state fMRI datasets to evaluate the performance of the random field theory. The detailed data information, preprocessing and permutation methods are illustrated in the Material and Methods, Section 2.1.5, and the Appendix. In brief, link-based BWAS is performed on the label-permuted data for 500 times. The estimated FWERs of the peak- and cluster-level inference are the proportion of permutations with any significant signals declared by the random field theory. Figure 4 shows the results of comparing the random field theory with the permutation approach at FWER 0.05.

For peak-level inference, the estimated FWERs are between 0.04 and 0.11. For cluster-level inference, if the CDTs are between 5 and 6, the estimated FWERs are between 0.03 and 0.12. Based on our analysis, we conclude that our method shows excellent performance if the sample size is large (over 120), the smooth kernel is moderate or large (larger than 3 voxles), and the CDT is high (larger than 5, p $\approx 3 \times 10^{-7}$) empirically. To get a more precise understanding of the performance, more real data permutations should be conducted in the future.

3.2 Performance of voxel-based BWAS

Based on 1653 autism patients with matched controls from ABIDE dataset, we evaluate the performance of voxel-based BWAS.

We first compare the result of voxel-based BWAS with three summary-statistics methods (MA, SOCS and gTFCE). See section 2.2.2 for their definitions. Link-based BWAS and voxel-based BWAS are performed on each site of the ABIDE dataset separately. The three statistics and the p-value of the first principal component of each voxel are recorded. In Figure 6, the scatter plot shows their relationship in the ABIDE-KKI dataset. A positive correlation can be observed between voxel-based BWAS and the MA, SOCS, and gTFCE-statistics, indicating that our method can both capture the characteristics of functional connectivities pattern of a voxel and provide a statistics with theoretical null distribution.

As detailed in section 2.2.3, we then use meta-analysis to show that voxel-based BWAS can provide robust and reproducible results. Figure 7 shows the result for one of the 10000 random meta-analysies using the threshold of CDT=3.5 ($p \approx 2.3 \times 10^{-4}$). Several overlapped voxel-clusters found by the two meta-analysies are shown to be associated with autism. They are partly located in some key regions, such as Thalamus, Precentral and Postcentral gyrus, Supplementary motor area, Fusiform gyrus, Frontal gyrus and Temporal gyrus. We also generate the distribution of the proportion of overlaps of 10000 random meta-analysies using different CDTs. As shown in Figure 8, the average proportion of overlap is larger than 45% at the threshold of CDT=3, but it becomes smaller when we raise the threshold. Nearly half of the findings can be replicated by our method in this multi-site analysis. At the same time, results in a multi-site dataset can be biased in several ways, including, for example, small sample size in each site, and acquisition of samples from different scanners using different proposals.

3.3 Theoretical power analysis

Figure 9 compares the power estimated by the modified random field theory (Section 2.3.1) and the simulation (Section 2.3.2). Each subfigure shows the relationship between sample size and power obtained by the two methods under different image smoothness (FWHM from 1 to 6 voxels). Although the original random field theory can only approximate the upper tail of maximum distribution, results show that performance is excellent in the context of determine the sample size after a suitable modification.

We then analyze the power of link-based BWAS studies on schizophrenia, autism and depression [12, 13, 14]. The effect size γ distribution is estimated on the basis of previous Z maps. With the FWER controlled at 0.05, the estimated effect sizes are around 0.2 to 0.23 (figure 10A). The power curves of effect size and sample size to achieve 90% power at different smoothness are shown in Figure

10B. It can be seen that about 300 samples are needed to reach 90% power at moderate smoothness (FWHM= 3 to 4 voxels). Even at low smoothness (FWHM= 1 voxel), 700 samples are enough to conduct such an analysis if we assume that the effect sizes are similar.

4 Discussion

In this article, we develop a rigorous statistical framework for BWAS. First, both peak- and cluster-level inferences are introduced for the analysis of voxel-wise functional connectomes, and the random field theory is developed to control the family-wise error rate and estimate statistical power. Second, a novel method to identify phenotype associated voxels based on the functional connectivity pattern is also proposed. All of them are carefully validated using several real fMRI datasets. Link-based BWAS shows excellent performance for controlling false positive rate, and voxel-based BWAS can provide reliable and reproducible findings, while the theoretical power analysis can estimate power accurately and rapidly.

Various advantages shown in this paper and previous studies [12, 13, 14] indicate that BWAS is well-suited to the identification of functional variations of human connectome. A large number of robust and reproducible functional connectivities and brain regions have been identified, shedding light on the fundamental mechanisms of the human brain and their relationship to psychiatric disorders. However, we are still far from voxel-level comprehensive mappings of the human connectome. Various studies have supported the superiority of BWAS to ROI-level analysis. For example, as shown in a recent study [29], neurons in different parts of the basolateral amygdala are both structurally and genetically distinct, and they are activated by stimuli that elicit different behaviors. In our paper reporting on depression [14], we found that the functional connectivities connecting different parts of the Orbitofrontal cortex (OFC) show significant difference between patients with depression and normal people. These findings indicate that the neurons in the same brain regions may carry out different functions. The averaging approach used in ROI-level analysis may provide wrong results. With more and more data available and higher resolution images, BWAS can make full use of the information provided by the data. It is completely unbiased and can provide evidence about the exact location of the functional connectivity variations.

To the best our knowledge, link-based BWAS is the first method to use the random field theory to analyze the voxel-wise functional connectome. However, the random field theory makes various strong assumptions. As shown in a recent article [19], it may lead to inflated false positive rate in task fMRI analysis. The non-parametric permutation method has been introduced to account for this problem and it has found successful applications [34]. Although its computational complexity is very high, it is particularly useful when the parametric assumptions are not met, or the null distribution of the statistic of interest is hard to derive analytically. In this paper, we used several real datasets to compute the empirical family-wise error rates for both peak- and cluster-level inference using permutation method. We found that the random field theory and permutation method can provide similar thresholds. Our analysis shows that the random field theory can provide a reliable result with large sample size (over 120), moderate or large smoothness (FWHM$>$ 3 voxels), and high CDT (p-value$<$ 10^{-7} on 3mm resolution data).

Large-scale computational resources are needed to perform BWAS, and an efficient Matlab-based software package, which can be downloaded from https://github.com/weikanggong/BWAS, has been

implemented. It takes about 2.5 hours to analyze the COBRE dataset (120 samples, 3mm resolution) and 3.5 hours to analyze the Cambridge datasets (197 samples, 3mm resolution) using a single core on a Linux workstation with Intel Xeon E5-2660 v3(2.60GHz) CPU and 128GB memory. For the permutation approach, 500 random permutations for the Cambridge dataset take about four days using 20 cores on the same workstation. The permutation approach is also implemented in BWAS software package.

The proposed framework in this paper is by no means a complete solution for the analysis of voxel-wise human connectome. Various statistical issues should be carefully addressed in the future. Perhaps the most important issue involves the high volume of available data. Under this circumstance, statistical methods for combining results from multiple imaging centers are needed. In GWAS, various meta-analysis approaches have been developed, and many novel genetic risk variations have been discovered [20]. In BWAS, integrating results from different datasets has also been shown to greatly reduce false positive rate and increase sensitivity [12, 13, 14]. However, the design of statistical methods for BWAS meta-analysis is challenging, because of sample heterogeneity introduced by different sources, such as different data acquisition pipelines, population stratification, and genetic background, etc.

A statistical testing framework based on random field theory can also be generalized to other fields, such as three-dimensional face studies, GWAS, and image-genetics studies. For example, similar to the human brain, the genome is also a complex three-dimensional structure, and newly developed chromosome conformation capture techniques make it possible to explore its three-dimensional organization [16]. We believe that the application of the random field theory to GWAS will not only provide deeper insights into the understanding of human genome functions, but also make GWAS more powerful in the future.

5 Acknowledgements

JF is supported by the National High Technology Research and Development Program of China (No. 2015AA020507), the Key Program of National Natural Science Foundation of China (No. 91230201), International (Regional) Collaborative and Exchange Program of National Natural Science (No. 71661167002), the Key Project of Shanghai Science and Technology Innovation Plan (No. 15JC1400101), and the Shanghai Soft Science Research Program (No. 15692106604), and the National Centre for Mathematics and Interdisciplinary Sciences (NCMIS) of the Chinese Academy of Sciences. JF is a Royal Society Wolfson Research Merit Award holder. LW is supported by the NSFC grants (No. 11571349 and No. 11201460), the NCMIS of the CAS, the Youth Innovation Promotion Association of the CAS, and the Strategic Priority Research Program of the CAS (XDB13040600).

6 Appendix

6.1 Image acquisition and preprocessing

Resting-state fMRI data were collected from three imaging sites: (1) 197 samples (normal people) from the Cambridge dataset in the 1000 FCP [8] (http://fcon_1000.projects.nitrc.org/fcpClassic/FcpTable.html); (2) 120 samples (67 normal people and 53 chronic schizophrenia patients) from the

COBRE dataset (http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html); and (3) 387 samples (241 normal people and 146 chronic schizophrenia patients) from the Taiwan dataset [32].

The detailed data acquisition methods may be found in the respective websites and papers. The data were preprocessed using SPM12 [36] and Data Processing and Analysis for Brain Imaging (DPABI) [45]. For each individual, the preprocessing steps included discarding the first 10 time points, slice timing correction, motion correction, coregistering to individual T1 structure image, segmenting structure images and DARTEL registration [3], regressing out nuisance covariates including white matter, cerebrospinal fluid and global signal using Friston 24-parameter model [23], filtering the fMRI time series with high-pass temporal filter (0.01 Hz) and low-pass temporal filter (0.1 Hz), normalizing to standard space of voxel size $3 \times 3 \times 3$ mm by DARTEL, and smoothing by a Gaussian kernel. Finally, all the images were manually checked to ensure successful preprocessing.

The ABIDE1 and ABIDE2 datasets [18] (http://fcon_1000.projects.nitrc.org/indi/abide/) are pooled together and preprocessed using SPM12 and DPARSF. The preprocessing pipeline is exactly the same as the dparsf pipeline used by the Preprocessed Connectomes Project (http://preprocessed-connectomes-project.org/abide/dparsf.html) with the global signal regressed out. The samples were discarded if 30% of their Framewise Displacement values were larger than 0.5 or they were preprocessed by error. A total of 1653 samples from 25 sites were used in our analysis. A summary of ABIDE demographic information is shown in Table 1.

6.2 The intrinsic volume and Gaussian EC-density

To perform peak-level and cluster-level inference, we should calculate the 0- to 3-dimensional intrinsic volume and the 0- to 6-dimensional EC-densities for the Gaussian random field.

Let P be the number of voxels, E_x (or E_y, E_z) be number of x (or y, z)-direction edges (two adjacent voxels), F_{xy} (or F_{yz}, F_{xz}) be number of xy (or yz, xz)-direction surface (four adjacent voxels), and C be the number of cubes (eight adjacent voxels). The r_x (or r_y, r_z) be the resel size of x (or y, z)-direction, which is defined as the voxel size divided by FWHM. The 0 to 3 dimensional intrinsic volume of S can be calculated as:

$$
\begin{aligned}
u_0(S) &= P - (E_x + E_y + E_z) + (F_{yz} + F_{xz} + F_{xy}) - C \\
u_1(S) &= (E_x - F_{xy} - F_{xz} + C)r_x + (E_y - F_{xy} - F_{yz} + C)r_y + (E_z - F_{xz} - F_{yz} + C)r_z \\
u_2(S) &= (F_{xy} - C)r_x r_y + (F_{xz} - C)r_x r_z + (F_{yz} - C)r_y r_z \\
u_3(S) &= C r_x r_y r_z
\end{aligned}
$$

The above calculation has been implement in SPM package as *spm_resels_vol* function. Two other methods also work well in practice. One is to replace the original space with a equal volume ball, as implement in the *fmristat* package, the other is to use a linear regression model [5], which do not need the knowledge of spatial smoothness.

The 0- to 6-dimensional EC-densities for Gaussian random field at t are:

$$\begin{aligned}
\rho_0(t) &= 1 - \Phi(t) \\
\rho_1(t) &= (4\ln 2)^{\frac{1}{2}}(2\pi)^{-1}e^{\frac{-t^2}{2}} \\
\rho_2(t) &= (4\ln 2)(2\pi)^{-\frac{3}{2}}te^{\frac{-t^2}{2}} \\
\rho_3(t) &= (4\ln 2)^{\frac{3}{2}}(2\pi)^{-2}(t^2-1)e^{\frac{-t^2}{2}} \\
\rho_4(t) &= (4\ln 2)^{2}(2\pi)^{-\frac{5}{2}}(t^3-3t)e^{\frac{-t^2}{2}} \\
\rho_5(t) &= (4\ln 2)^{\frac{5}{2}}(2\pi)^{-3}(t^4-6t^2+3)e^{\frac{-t^2}{2}} \\
\rho_6(t) &= (4\ln 2)^{3}(2\pi)^{-\frac{7}{2}}(t^5-10t^3+15t)e^{\frac{-t^2}{2}}
\end{aligned}$$

where $\Phi(\bullet)$ is the cumulative distribution function of standard normal distribution.

6.3 Proof of formula (1)

Using the property of d-dimensional intrinsic volume [40]

$$u_d(\mathcal{P}\times\mathcal{Q}) = \sum_{k=0}^{d} u_k(\mathcal{P})u_{d-k}(\mathcal{Q})$$

When $d > P$, $u_d(\mathcal{P}) = 0$ and $d > Q$, $u_d(\mathcal{Q}) = 0$. It is easy to conclude that

$$\begin{aligned}
\sum_{d=0}^{P+Q}\mu_d(\mathcal{P}\times\mathcal{Q}) &= \sum_{d=0}^{P+Q}\sum_{k=0}^{d}u_k(\mathcal{P})u_{d-k}(\mathcal{Q}) \\
&= \sum_{i=0}^{P}\sum_{j=0}^{Q}\mu_i(\mathcal{P})\mu_j(\mathcal{Q})
\end{aligned}$$

In our case, we have $P = Q = 3$.

6.4 Proof of formula (4)

Let $Var(\dot{M}^{(s)}(p)) = \Lambda_{M_s}$ and $Var(\dot{N}^{(s)}(q)) = \Lambda_{N_s}$, then according to the Lemma 4.2 in [10],

$$\frac{\partial R^{(s)}(p,q)}{\partial p} \overset{D}{=} (1-R^{(s)}(p,q)^2)^{\frac{1}{2}}a_s^{-\frac{1}{2}}(\Lambda_{M_s})^{\frac{1}{2}}z_M^{(s)}$$

and

$$\frac{\partial R^{(s)}(p,q)}{\partial q} \overset{D}{=} (1-R^{(s)}(p,q)^2)^{\frac{1}{2}}a_s^{-\frac{1}{2}}(\Lambda_{N_s})^{\frac{1}{2}}z_N^{(s)}$$

where $a_s \sim \chi^2_{v^{(s)}}$, $z_M^{(s)}, z_N^{(s)} \sim N(0, I_{3,3})$ and independent of $R^{(s)}(p,q)$, and $\overset{D}{=}$ means equal in distribution. Then, after the Fisher's Z transformation, we have

$$\frac{\partial Z^{(s)}(p,q)}{\partial p} \overset{D}{=} (1-R^{(s)}(p,q)^2)^{-\frac{1}{2}}a_s^{-\frac{1}{2}}(\Lambda_{M_s})^{\frac{1}{2}}z_M^{(s)}$$

and

$$\frac{\partial Z^{(s)}(p,q)}{\partial q} \overset{D}{=} (1-R^{(s)}(p,q)^2)^{-\frac{1}{2}} a_s^{-\frac{1}{2}} (\Lambda_{N_s})^{\frac{1}{2}} z_N^{(s)}$$

Then,

$$\frac{\partial Z(p,q)}{\partial p} \overset{D}{=} \sum_{s=1}^{n} w^{(s)} (1-R^{(s)}(p,q)^2)^{-\frac{1}{2}} a_s^{-\frac{1}{2}} (\Lambda_{M_s})^{\frac{1}{2}} z_M^i$$

and

$$\frac{\partial Z(p,q)}{\partial q} \overset{D}{=} \sum_{s=1}^{n} w^{(s)} (1-R^{(s)}(p,q)^2)^{-\frac{1}{2}} a_s^{-\frac{1}{2}} (\Lambda_{N_s})^{\frac{1}{2}} z_N^i$$

Since

$$Var(\dot{Z}(p,q)) = \begin{pmatrix} \mathbb{E}[\frac{\partial Z^{(s)}(p,q)}{\partial p}\frac{\partial Z^{(s)\prime}(p,q)}{\partial p}] & \mathbb{E}[\frac{\partial Z^{(s)}(p,q)}{\partial p}\frac{\partial Z^{(s)\prime}(p,q)}{\partial q}] \\ \mathbb{E}[\frac{\partial Z^{(s)}(p,q)}{\partial q}\frac{\partial Z^{(s)\prime}(p,q)}{\partial p}] & \mathbb{E}[\frac{\partial Z^{(s)}(p,q)}{\partial q}\frac{\partial Z^{(s)\prime}(p,q)}{\partial q}] \end{pmatrix}$$

and

$$\mathbb{E}[\frac{\partial Z^{(s)}(p,q)}{\partial p}\frac{\partial Z^{(s)\prime}(p,q)}{\partial q}] = \mathbb{E}[\frac{\partial Z^{(s)}(p,q)}{\partial q}\frac{\partial Z^{(s)\prime}(p,q)}{\partial p}] = 0$$

and

$$\mathbb{E}[\frac{\partial Z^{(s)}(p,q)}{\partial p}\frac{\partial Z^{(s)\prime}(p,q)}{\partial p}] = \sum_{s=1}^{n}(w^{(s)})^2 \mathbb{E}[(1-R^{(s)}(p,q)^2)^{-1}]\mathbb{E}[a_s^{-1}]\Lambda_{M_s}$$

$$\mathbb{E}[\frac{\partial Z^{(s)}(p,q)}{\partial q}\frac{\partial Z^{(s)\prime}(p,q)}{\partial q}] = \sum_{s=1}^{n}(w^{(s)})^2 \mathbb{E}[(1-R^{(s)}(p,q)^2)^{-1}]\mathbb{E}[a_s^{-1}]\Lambda_{N_s}$$

The expectations in the above equations are

$$\mathbb{E}[a_s^{-1}] = \frac{1}{v^{(s)}-2}$$

$$\mathbb{E}[(1-R^{(s)}(p,q)^2)^{-1}] = \frac{v^{(s)}-2}{v^{(s)}-3}$$

Finally we get

$$Var(\dot{Z}(p,q)) = \begin{pmatrix} \sum_{s=1}^{n}\frac{(w^{(s)})^2}{v^{(s)}-3}\Lambda_{M_s} & 0 \\ 0 & \sum_{s=1}^{n}\frac{(w^{(s)})^2}{v^{(s)}-3}\Lambda_{N_s} \end{pmatrix}_{6\times 6}$$

Substituting the variance covariance matrix of partial derivative of the random field by the FWHM using (3), we could get (4) by:

$$\left(\frac{\mathrm{FWHM}_Z}{(4\ln 2)^{\frac{1}{2}}}\right)^{-2(3+3)} = \left(\sum_{s=1}^{n}\frac{(w^{(s)})^2}{v^{(s)}-3}\frac{\mathrm{FWHM}_{M^{(s)}}^{-2}}{(4ln2)^{\frac{1}{2}}}\right)^3 \left(\sum_{s=1}^{n}\frac{(w^{(s)})^2}{v^{(s)}-3}\frac{\mathrm{FWHM}_{N^{(s)}}^{-2}}{(4ln2)^{\frac{1}{2}}}\right)^3$$

thus

$$\mathrm{FWHM}_Z = \left(\sum_{s=1}^{n}\frac{(w^{(s)})^2}{v^{(s)}-3}\mathrm{FWHM}_{M^{(s)}}^{-2}\right)^{-\frac{1}{4}} \left(\sum_{s=1}^{n}\frac{(w^{(s)})^2}{v^{(s)}-3}\mathrm{FWHM}_{N^{(s)}}^{-2}\right)^{-\frac{1}{4}}$$

References

[1] R. J. Adler. *The geometry of random fields*, volume 62. Siam, 1981.

[2] R. J. Adler and J. E. Taylor. *Random fields and geometry*. Springer Science & Business Media, 2009.

[3] J. Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113, 2007.

[4] J. Ashburner and K. J. Friston. Voxel-based morphometry—the methods. *Neuroimage*, 11(6):805–821, 2000.

[5] K. Bartz, S. Kou, and R. J. Adler. Estimating thresholding levels for random fields via euler characteristics. 2011.

[6] C. F. Beckmann and S. M. Smith. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE transactions on medical imaging*, 23(2):137–152, 2004.

[7] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995.

[8] B. B. Biswal, M. Mennes, X.-N. Zuo, S. Gohel, C. Kelly, S. M. Smith, C. F. Beckmann, J. S. Adelstein, R. L. Buckner, S. Colcombe, et al. Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences*, 107(10):4734–4739, 2010.

[9] J. Cao. The size of the connected components of excursion sets of χ 2, t and f fields. *Advances in Applied Probability*, pages 579–595, 1999.

[10] J. Cao, K. Worsley, et al. The geometry of correlation fields with an application to functional connectivity of the brain. *The Annals of Applied Probability*, 9(4):1021–1057, 1999.

[11] F. Carbonell, K. Worsley, and N. Trujillo-Barreto. On the fisher's z transformation of correlation random fields. *Statistics & Probability Letters*, 79(6):780–788, 2009.

[12] W. Cheng, L. Palaniyappan, M. Li, K. M. Kendrick, J. Zhang, Q. Luo, Z. Liu, R. Yu, W. Deng, Q. Wang, et al. Voxel-based, brain-wide association study of aberrant functional connectivity in schizophrenia implicates thalamocortical circuitry. *npj Schizophrenia*, 1, 2015.

[13] W. Cheng, E. T. Rolls, H. Gu, J. Zhang, and J. Feng. Autism: reduced connectivity between cortical areas involved in face expression, theory of mind, and the sense of self. *Brain*, page awv051, 2015.

[14] W. Cheng, E. T. Rolls, J. Qiu, W. Liu, Y. Tang, C.-C. Huang, X. Wang, J. Zhang, W. Lin, L. Zheng, et al. Medial reward and lateral non-reward orbitofrontal cortex circuits change in opposite directions in depression. *Brain*, page aww255, 2016.

[15] Y. Choi, J. Taylor, and R. Tibshirani. Selecting the number of principal components: Estimation of the true rank of a noisy matrix. *arXiv preprint arXiv:1410.8260*, 2014.

[16] J. Dekker, M. A. Marti-Renom, and L. A. Mirny. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature Reviews Genetics*, 14(6):390–403, 2013.

[17] J. E. Desmond and G. H. Glover. Estimating sample size in functional mri (fmri) neuroimaging studies: statistical power analyses. *Journal of neuroscience methods*, 118(2):115–128, 2002.

[18] A. Di Martino, C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014.

[19] A. Eklund, T. E. Nichols, and H. Knutsson. Cluster failure: Why fmri inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, page 201602413, 2016.

[20] E. Evangelou and J. P. Ioannidis. Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics*, 14(6):379–389, 2013.

[21] K. J. Friston, A. Holmes, J.-B. Poline, C. J. Price, and C. D. Frith. Detecting activations in pet and fmri: levels of inference and power. *Neuroimage*, 4(3):223–235, 1996.

[22] K. J. Friston, A. P. Holmes, and K. J. Worsley. How many subjects constitute a study? *Neuroimage*, 10(1):1–5, 1999.

[23] K. J. Friston, S. Williams, R. Howard, R. S. Frackowiak, and R. Turner. Movement-related effects in fmri time-series. *Magnetic resonance in medicine*, 35(3):346–355, 1996.

[24] K. J. Friston, K. J. Worsley, R. Frackowiak, J. C. Mazziotta, and A. C. Evans. Assessing the significance of focal activations using their spatial extent. *Human brain mapping*, 1(3):210–220, 1994.

[25] S. Hayasaka and T. E. Nichols. Validating cluster size inference: random field and permutation methods. *Neuroimage*, 20(4):2343–2356, 2003.

[26] S. Hayasaka, A. M. Peiffer, C. E. Hugenschmidt, and P. J. Laurienti. Power and sample size calculation for neuroimaging studies by non-central random field theory. *Neuroimage*, 37(3):721–730, 2007.

[27] K. E. Joyce and S. Hayasaka. Development of powermap: a software package for statistical power calculation in neuroimaging studies. *Neuroinformatics*, 10(4):351–365, 2012.

[28] J. F. J. F. Kenney. Mathematics of statistics. Technical report, 1939.

[29] J. Kim, M. Pignatelli, S. Xu, S. Itohara, and S. Tonegawa. Antagonistic negative and positive neurons of the basolateral amygdala. *Nature Neuroscience*, 19(12):1636–1646, 2016.

[30] J. Kim, J. R. Wozniak, B. A. Mueller, X. Shen, and W. Pan. Comparison of statistical tests for group differences in brain functional networks. *NeuroImage*, 101:681–694, 2014.

[31] D. Lamparter, D. Marbach, R. Rueedi, Z. Kutalik, and S. Bergmann. Fast and rigorous computation of gene and pathway scores from snp-based summary statistics. *PLoS Comput Biol*, 12(1):e1004714, 2016.

[32] C.-Y. Z. Lo, T.-W. Su, C.-C. Huang, C.-C. Hung, W.-L. Chen, T.-H. Lan, C.-P. Lin, and E. T. Bullmore. Randomization and resilience of brain functional networks as systems-level endophenotypes of schizophrenia. *Proceedings of the National Academy of Sciences*, 112(29):9123–9128, 2015.

[33] J. A. Mumford and T. E. Nichols. Power calculation for group fmri studies accounting for arbitrary design and temporal autocorrelation. *Neuroimage*, 39(1):261–268, 2008.

[34] T. E. Nichols and A. P. Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, 15(1):1–25, 2002.

[35] V. Nosko. Local structure of gaussian random fields in the neighborhood of high-level shines. In *Dokl. Akad. Nauk SSSR*, volume 189, pages 714–717, 1969.

[36] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols. *Statistical parametric mapping: the analysis of functional brain images*. Academic press, 2011.

[37] E. T. Rolls, M. Joliot, and N. Tzourio-Mazoyer. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage*, 122:1–5, 2015.

[38] Z. Shehzad, C. Kelly, P. T. Reiss, R. C. Craddock, J. W. Emerson, K. McMahon, D. A. Copland, F. X. Castellanos, and M. P. Milham. A multivariate distance-based analytic framework for connectome-wide association studies. *Neuroimage*, 93:74–94, 2014.

[39] S. M. Smith and T. E. Nichols. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, 44(1):83–98, 2009.

[40] J. Taylor and K. Worsley. Random fields of multivariate test statistics, with applications to shape analysis. *The Annals of Statistics*, pages 1–27, 2008.

[41] K. J. Worsley. Local maxima and the expected euler characteristic of excursion sets of χ 2, f and t fields. *Advances in Applied Probability*, pages 13–42, 1994.

[42] K. J. Worsley, A. C. Evans, S. Marrett, P. Neelin, et al. A three-dimensional statistical analysis for cbf activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, 12:900–900, 1992.

[43] K. J. Worsley, S. Marrett, P. Neelin, and A. Evans. Searching scale space for activation in pet images. *Human brain mapping*, 4(1):74–90, 1996.

[44] K. J. Worsley, S. Marrett, P. Neelin, A. C. Vandal, K. J. Friston, A. C. Evans, et al. A unified statistical approach for determining significant signals in images of cerebral activation. *Human brain mapping*, 4(1):58–73, 1996.

[45] C.-G. Yan, X.-D. Wang, X.-N. Zuo, and Y.-F. Zang. Dpabi: Data processing & analysis for (resting-state) brain imaging. *Neuroinformatics*, pages 1–13, 2016.

[46] A. Zalesky, L. Cocchi, A. Fornito, M. M. Murray, and E. Bullmore. Connectivity differences in brain networks. *Neuroimage*, 60(2):1055–1062, 2012.

[47] J. Zhang, W. Cheng, Z. Liu, K. Zhang, X. Lei, Y. Yao, B. Becker, Y. Liu, K. M. Kendrick, G. Lu, et al. Neural, electrophysiological and anatomical basis of brain-network variability and its characteristic changes in mental disorders. *Brain*, 139(8):2307–2321, 2016.

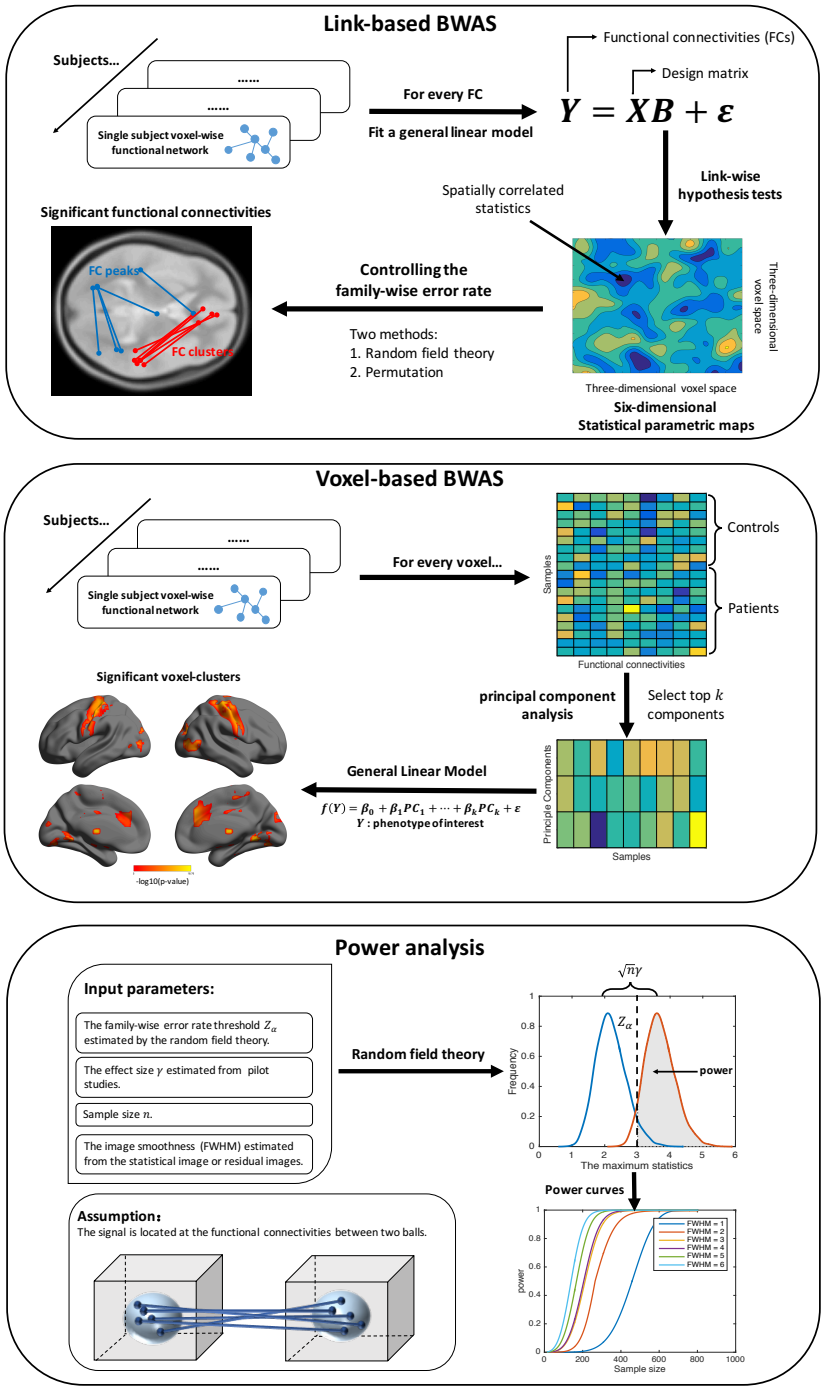

Figure 1: A flow chart of link-based BWAS, voxel-based BWAS and the power analysis.

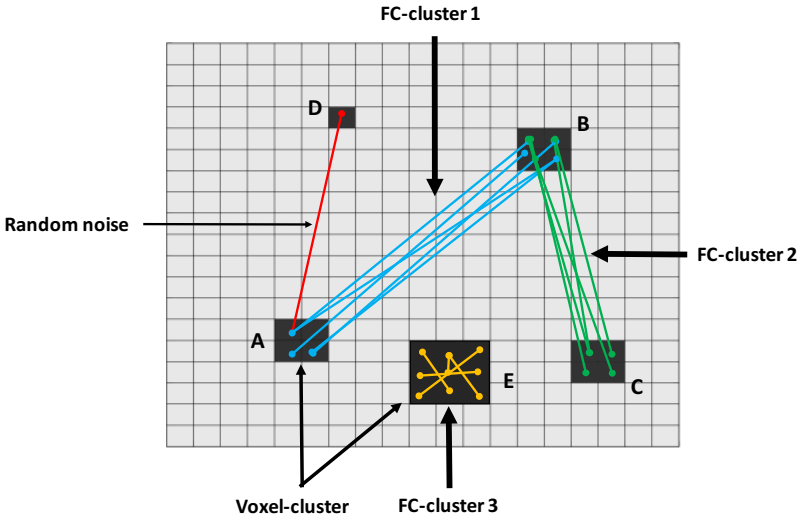

Figure 2: A two-dimensional diagram of FC-cluster. Based on the definition of SPC statistic, FCs between AB, BC and AD are FC-clusters, and FCs within voxel-cluster E also form a FC-cluster.

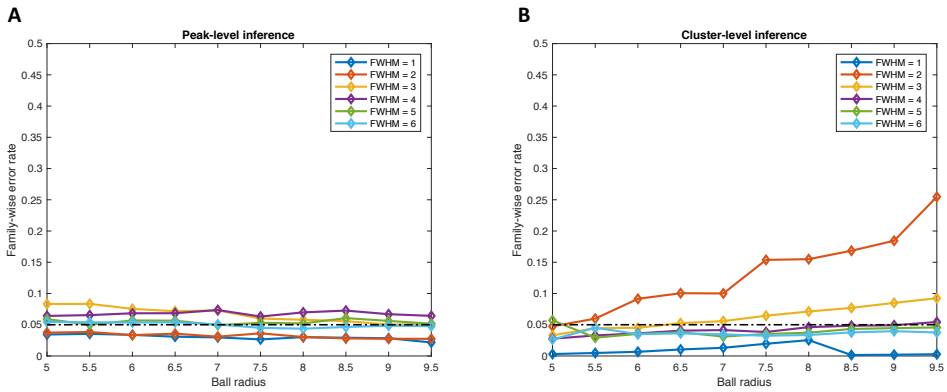

Figure 3: Comparing the random field theory with simulation approach at FWER=0.05 under different region shapes and smoothness. The estimated FWER is the proportion of simulations with any significant signals found by the random field theory. (A) Results for peak-level inference. (B) Results for cluster-level inference at CDT=4.5.

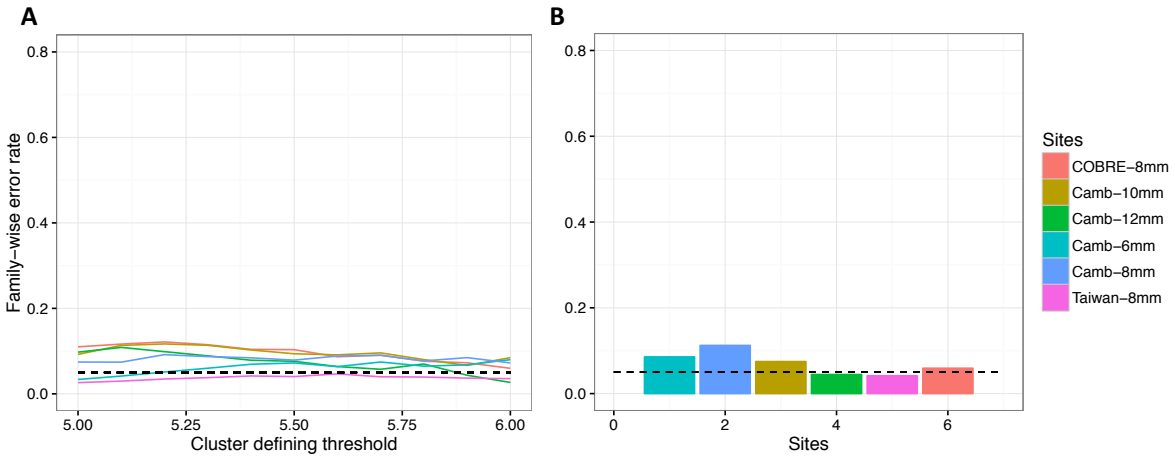

Figure 4: Comparing the random field theory with permutation approach at FWER=0.05 in the Cambridge, COBRE and Taiwan datasets under different smoothness. The estimated FWER is the proportion of permutations with any significant signals found by the random field theory. (A) Results for cluster-level inference at different CDTs; (B) Results for peak-level inference.

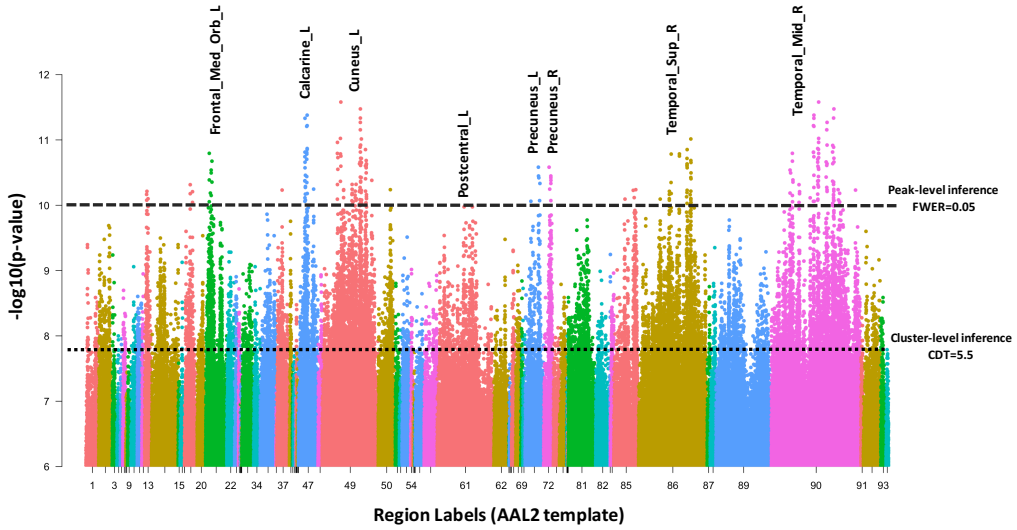

Figure 5: Manhattan plot for link-based BWAS results of ABIDE dataset ($p < 10^{-6}$ only). Each point represents a functional connectivity grouped by the 94 cerebrum regions of AAL2 template. Note that each functional connectivity can appear once or twice in this figure, depending on its endpoints.

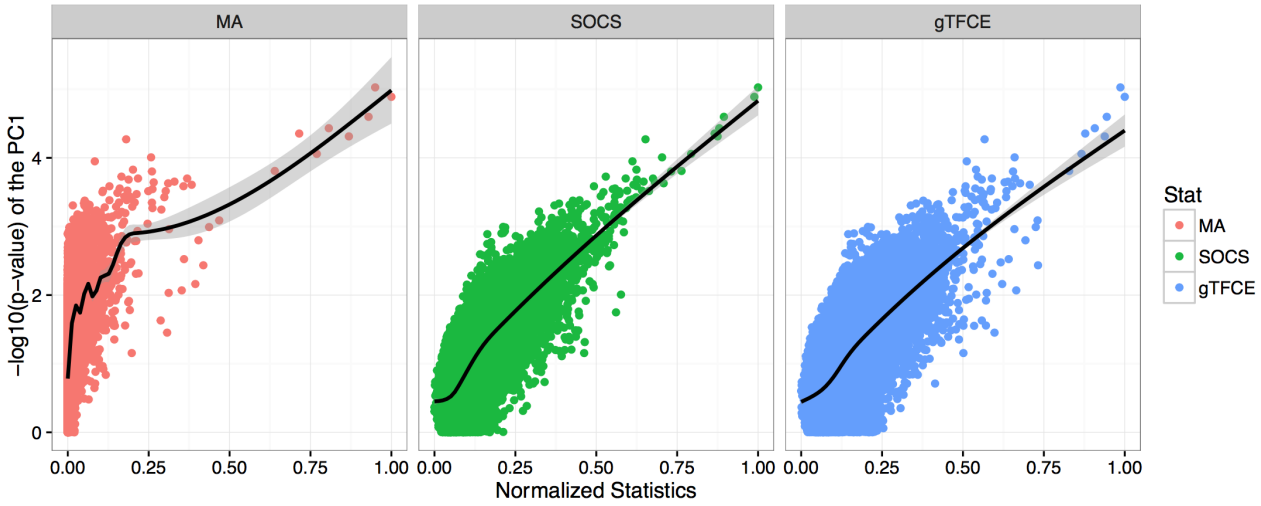

Figure 6: There is a positive relationship between the results of voxel-based BWAS using the first principal component and three summary statistics methods, including MA, SOCS, and gTFCE, in the ABIDE-KKI dataset. Each point represents the corresponding statistics on one voxel. The three summary statistics are rescaled to range between zero and one. For MA value, a threshold of $|z_0| = 4$ is used.

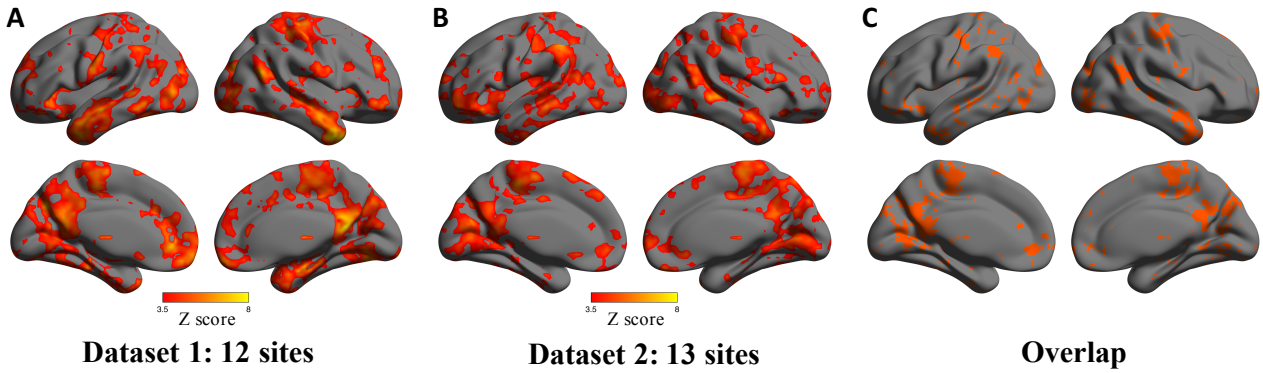

Figure 7: One of the 10000 random meta-analysies of voxel-based BWAS using the ABIDE dataset. The whole ABIDE dataset is divided into two small datasets. One contains 12 sites, and the other contains 13 sites. The Fisher's method is used to integrate the results, and the p-values are converted to positive Z scores. All significant voxel-clusters passed cluster-level inference 0.05 correction with CDT=3.5. (A) Significant regions of dataset 1. (B) Significant regions of dataset 2. (C) Overlapped regions.

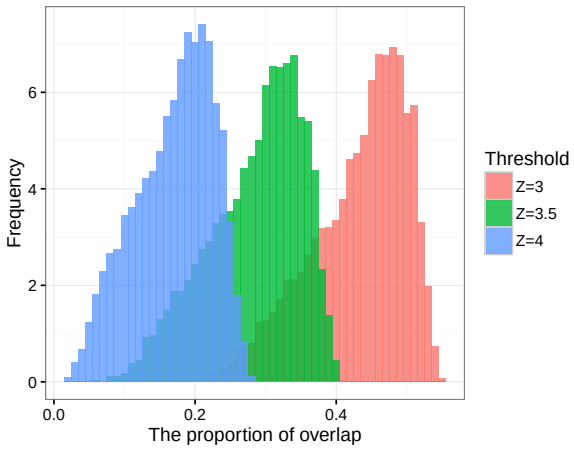

Figure 8: Reproducibility of voxel-based BWAS. The distributions of proportion of overlap under three different CDTs in the ABIDE dataset. Each of the distribution is generated by 10000 random combinations of the meta-analysies.

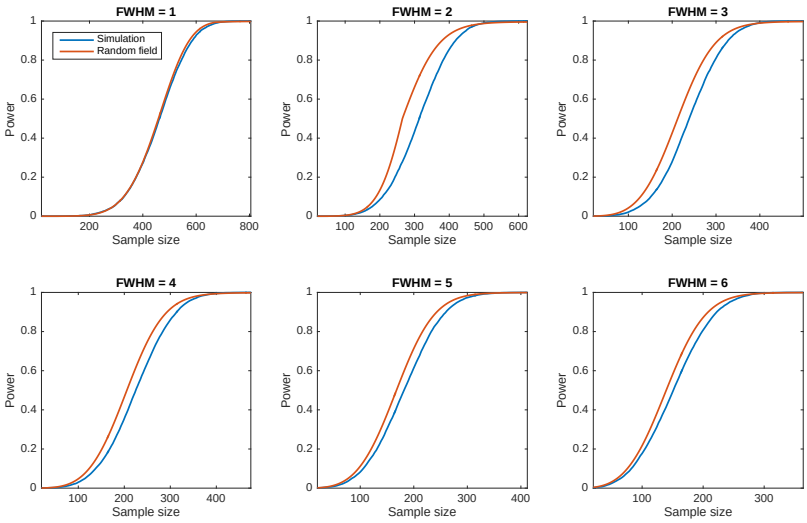

Figure 9: Using the simulation (blue line) to evaluate the validity of the modified random field theory (red line) for power analysis. The effect size is set to 0.2, and FWHM= 1 to 6 voxels.

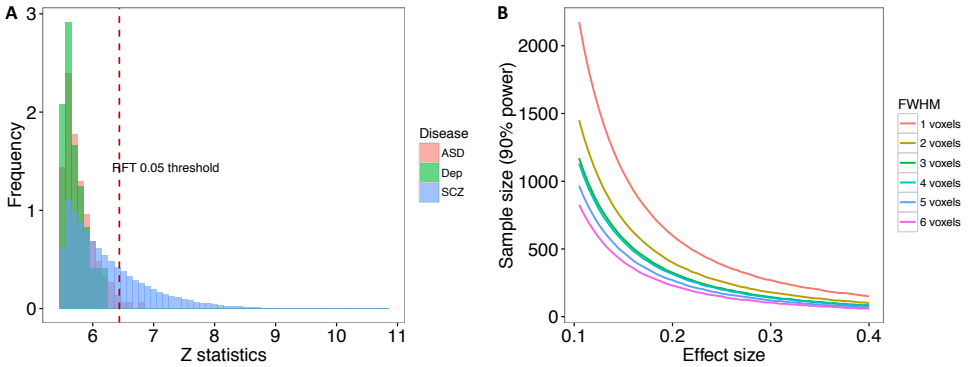

Figure 10: Power analysis results of our previous BWAS studies. (A) Histograms of absolute Z statistics obtained by three previous brain-wide association studies; absolute Z-values larger than 5.5 are shown. Red dashed line is the threshold estimated by the RFT at FWER=0.05. (B) Power curves between the effect size γ and the sample size n at 90% power under different smoothness. Effect sizes are around 0.2 to 0.23 at RFT 0.05 threshold in previous studies.

Table 1: A summary of the demographic information of ABIDE dataset

Site	N normal	N patient	Mean age	N male	N right handed	Mean IQ	Mean FD
BNI	22	22	36.5	44	44	108.77	0.13
CALTECH	5	10	26.8	10	13	107.71	0.06
EMC	16	18	8.4	28	26	NaN	0.10
ETH	6	23	23.4	29	29	116.34	0.10
GU	36	47	10.8	53	72	119.22	0.11
IP	15	25	21.9	18	34	94.50	0.06
IU	17	20	24.6	29	30	116.27	0.08
KKI	52	164	10.3	144	183	110.84	0.12
LEUVEN	23	30	18.6	47	46	112.43	0.08
MAXMUN	19	26	26.9	41	43	112.02	0.09
NYU	117	127	13.8	203	50	109.47	0.07
OHSU	43	41	10.9	65	81	112.15	0.10
OILH	22	34	22.9	38	47	111.36	0.08
OLIN	16	13	17.1	24	24	113.04	0.13
PITT	20	22	18.8	36	37	110.19	0.12
SBL	11	12	33.7	23	0	106.50	0.14
SDSU	40	45	13.7	70	71	104.44	0.07
STANFORD	14	20	10.0	27	28	113.18	0.09
TCD	14	18	15.9	32	32	115.03	0.12
TRINITY	21	23	17.3	44	44	109.64	0.08
UCD	14	14	15.0	21	27	108.04	0.09
UCLA	52	51	12.8	89	95	105.45	0.10
UM	47	68	14.5	89	92	108.28	0.08
USM	50	37	23.0	82	12	106.58	0.10
YALE	25	26	12.9	38	41	99.67	0.10