NeuroImage xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Impact of the resolution of brain parcels on connectome-wide association studies in fMRI

Pierre Bellec [a,b,*], Yassine Benhajali [a,c], Felix Carbonell [d], Christian Dansereau [a,b], Geneviève Albouy [a,e], Maxime Pelland [e], Cameron Craddock [f,g], Oliver Collignon [e], Julien Doyon [a,e], Emmanuel Stip [h,i], Pierre Orban [a,h]

[a] Functional Neuroimaging Unit, Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Canada
[b] Department of Computer Science and Operations Research, University of Montreal, Montreal, QC, Canada
[c] Department of Anthropology, University of Montreal, Montreal, QC, Canada
[d] Biospective Incorporated, Montreal, QC, Canada
[e] Department of Psychology, University of Montreal, Montreal, QC, Canada
[f] Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, USA
[g] Center for the Developing Brain, Child Mind Institute, New York, NY, USA
[h] Department of Psychiatry, University of Montreal, Montreal, QC, Canada
[i] Centre Hospitalier de l'Université de Montréal, Montreal, QC, Canada

ARTICLE INFO

Article history:
Received 27 January 2015
Accepted 23 July 2015
Available online xxxx

Keywords:
fMRI
General linear model
Functional brain parcellation
Multiple comparison
False discovery rate
Multiresolution analysis
Connectome

ABSTRACT

A recent trend in functional magnetic resonance imaging is to test for association of clinical disorders with every possible connection between selected brain parcels. We investigated the impact of the resolution of functional brain parcels, ranging from large-scale networks to local regions, on a mass univariate general linear model (GLM) of connectomes. For each resolution taken independently, the Benjamini–Hochberg procedure controlled the false-discovery rate (FDR) at nominal level on realistic simulations. However, the FDR for tests pooled across all resolutions could be inflated compared to the FDR within resolution. This inflation was severe in the presence of no or weak effects, but became negligible for strong effects. We thus developed an omnibus test to establish the overall presence of true discoveries across all resolutions. Although not a guarantee to control the FDR across resolutions, the omnibus test may be used for descriptive analysis of the impact of resolution on a GLM analysis, in complement to a primary analysis at a predefined single resolution. On three real datasets with significant omnibus test (schizophrenia, congenital blindness, motor practice), markedly higher rate of discovery were obtained at low resolutions, below 50, in line with simulations showing increase in sensitivity at such resolutions. This increase in discovery rate came at the cost of a lower ability to localize effects, as low resolution parcels merged many different brain regions together. However, with 30 or more parcels, the statistical effect maps were biologically plausible and very consistent across resolutions. These results show that resolution is a key parameter for GLM-connectome analysis with FDR control, and that a functional brain parcellation with 30 to 50 parcels may lead to an accurate summary of full connectome effects with good sensitivity in many situations.

Introduction

Context

Brain connectivity in resting-state functional magnetic resonance imaging (fMRI) has been found to be associated with a wide variety of clinical disorders (Fox and Greicius, 2010; Castellanos et al., 2013; Barkhof et al., 2014). Rather than focusing on a limited set of a priori regions of interest, a recent trend is to perform statistical tests of association across the whole connectome, i.e. at every possible brain connection (Shehzad et al., 2014). Such connectome-wide association studies (CWAS) critically depend on the choice of the brain parcels that are used to estimate the connections. Analyses have been performed at different *resolutions* in the literature (Meskaldji et al., 2013), e.g. voxels (Shehzad et al., 2014), regions (Wang et al., 2007), or distributed networks (Jafri et al., 2008; Marrelec et al., 2008). The main objective of this work was to study the impact of the spatial resolution on the results of a CWAS.

Mass-univariate connectome-wide association studies

The mass-univariate approach to CWAS (Worsley et al., 1998) consists of independently estimating a GLM at every connection. In the GLM, a series of equations are solved to find a linear mixture of explanatory variables (called covariates) that best fit the connectivity values observed across the many subjects. A p value is generated for each

* Corresponding author at: Functional Neuroimaging Unit, Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Canada.
E-mail address: pierre.bellec@criugm.qc.ca (P. Bellec).

http://dx.doi.org/10.1016/j.neuroimage.2015.07.071
1053-8119/

connection to quantify the probability that the estimated strength of association between this connection and a covariate of interest could have arisen randomly in the absence of a true association (Worsley and Friston, 1995). The significance level of each test needs to be corrected for the total number of tests, i.e. the number of brain connections, using for example random field theory (Worsley et al., 1998) or FDR (Benjamini and Hochberg, 1995). Correction for multiple comparisons however generally comes at the cost of a sharp decrease in sensitivity.

Multiresolution parcellations and testing

A straightforward way to mitigate the impact of multiple comparisons on statistical power is to reduce the number of brain parcels. For example, the AAL template (Tzourio-Mazoyer et al., 2002) includes 116 brain parcels based on anatomical landmarks. Data-driven algorithms can also generate functional brain parcels (Bellec et al., 2006; Thirion et al., 2006, 2014; Craddock et al., 2012; Blumensath et al., 2013; Gordon et al., 2014). Few investigators have examined how resolution impacts the results of a CWAS. Abou Elseoud et al. (2011) explored the impact of the number of components in a dual-regression independent component analysis on the difference between patients suffering from non-medicated seasonal affective disorder and normal healthy controls. The authors concluded that the number of significant findings was maximized at resolution 45 (in this case, 45 independent components). The impact of the number of brain parcels was also investigated using spatially-constrained spectral clustering (Craddock et al., 2012) at much higher resolutions (from 50 to 3000+) by Shehzad et al. (2014). The authors concluded that the association between resting-state connectivity and intelligence quotient was consistent across resolutions. It should be noted that, in the above-mentioned studies (Abou Elseoud et al., 2011; Shehzad et al., 2014), the authors did not investigate the implications that the replication of statistical tests at multiple resolutions may have in terms of the control of false positives. We are thus not currently aware of a valid statistical framework to examine the results of a CWAS with data-driven brain parcellations at multiple resolutions.

Objectives

In this paper, we investigated empirically the impact of the number of brain parcels (resolution) on a mass univariate GLM analysis of connectomes (GLM-connectome). Our first objective was to empirically assess if the Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995) controlled appropriately the FDR with a single brain parcellation, and generated biologically plausible results. Our second objective was to assess if repeating a GLM-connectome analysis independently using multiple parcellations at different resolutions would inflate the overall FDR, pooling tests across all resolutions. Anticipating a lack of control in the absence of any true association, we developed an omnibus test checking for the overall presence of significant associations across all resolutions. Our third objective was to use the omnibus test to evaluate how the resolution impacted the rate of discovery in a GLM analysis, and if the associations derived with the GLM would be consistent across resolutions. We conducted a series of experiments involving both simulated and real datasets to address these three objectives, which have been summarized, along with the main findings, in Table 1.

Statistical testing procedures

Functional connectome

The first step to build a connectome is to select a parcellation of the brain, with R parcels. In this work, we relied on a "Bootstrap Analysis of Stable Clusters" (BASC), which can identify consistent functional parcels for a group of subjects (Bellec et al., 2010), using a hierarchical cluster with Ward's criterion both at the individual and the group levels. The functional parcels can be generated at any arbitrary resolution (within the range of the fMRI resolution), and we considered only parcels generated at the group level, which were non-overlapping and not necessarily spatially contiguous. For each resolution, and each pair of distinct parcels i and j at this resolution, the between-parcel connectivity $y_{i,j}$ is measured by the Fisher transform of the Pearson's correlation between the average time series of the parcels. Note that other measures can be used to quantify interactions between parcels, such as partial correlations (Marrelec et al., 2006). We used correlation as it is the simplest, most popular and still fairly accurate (Smith et al., 2011) measure of interaction in fMRI. The statistical framework presented here could still be applied to many other measures. The within-parcel connectivity $y_{i,i}$ is the Fisher transform of the average correlation between time series of every pair of distinct voxels inside parcel i. The connectome $\mathbf{Y} = (y_{i,j})_{i,j=1}^{R}$ is thus a $R \times R$ matrix. Each column j (or row, as the matrix is symmetric) codes for the connectivity between parcel j and all other brain parcels, or in other word is a full brain functional connectivity map. See Fig. 1a–b for a representation of a parcellation and associated

Table 1
Summary of the specific objectives, experiments and findings of the paper.

Specific objectives	Experiment(s)	Finding(s)
1a. Check the validity of the FDR-BH algorithm for GLM tests at a fixed resolution.	The homoscedasticity and normality assumptions of the GLM were tested on four real data samples at high resolution (300+). Simulations of group differences were implemented at multiple resolutions with and without dependencies between tests.	No significant departure from normality and homoscedasticity were observed (Results Section). In the simulations, the BH procedure controlled the FDR below or at the prescribed level (Figs. 3, 5).
1b. Assess the biological plausibility of the results identified with GLM-connectome with real data.	GLM-connectome analyses in three real datasets at multiple resolutions.	The GLM-connectome identified biologically plausible changes in connectivity in all three analyses (Figs. 9, 10).
2a. Assess the specificity of GLM-connectome when combining multiple resolutions.	Simulations of group differences were implemented at multiple resolutions with and without dependencies between tests.	The FDR across resolutions was controlled in simulations with strong or widespread signal, but was too liberal when no or weak signal was simulated (Figs. 3, 5).
2b. Assess the specificity of the omnibus test across resolutions in the absence of signal ("global null").	Test for differences in average connectivity between random subgroups of a large demographically homogeneous sample.	The FWE was controlled at nominal level by the omnibus test under the global null (Fig. 5). Departures from nominal levels of the FDR across resolutions could still be observed when the omnibus was significant (Fig. 5).
3a. Assess the sensitivity of the FDR-BH across resolutions.	GLM-connectome analyses of simulated and real datasets at multiple resolutions.	On simulations, the sensitivity varied substantially across resolutions and was higher at low resolutions, below 50. (Figs. 4, 6). On real data, the discovery rate was markedly higher at resolutions below 50 (Figs. Fig. 7, 9).
3b. Assess the consistency and differences of GLM-connectome results at different resolutions.	GLM-connectome analyses in three real datasets at multiple resolutions.	Statistical parametric maps were very consistent across resolutions (Fig. 11), although some effects associated with specific structures were better seen at resolutions above 50 (Figs. 9, 10).

Please cite this article as: Bellec, P., et al., Impact of the resolution of brain parcels on connectome-wide association studies in fMRI, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.07.071

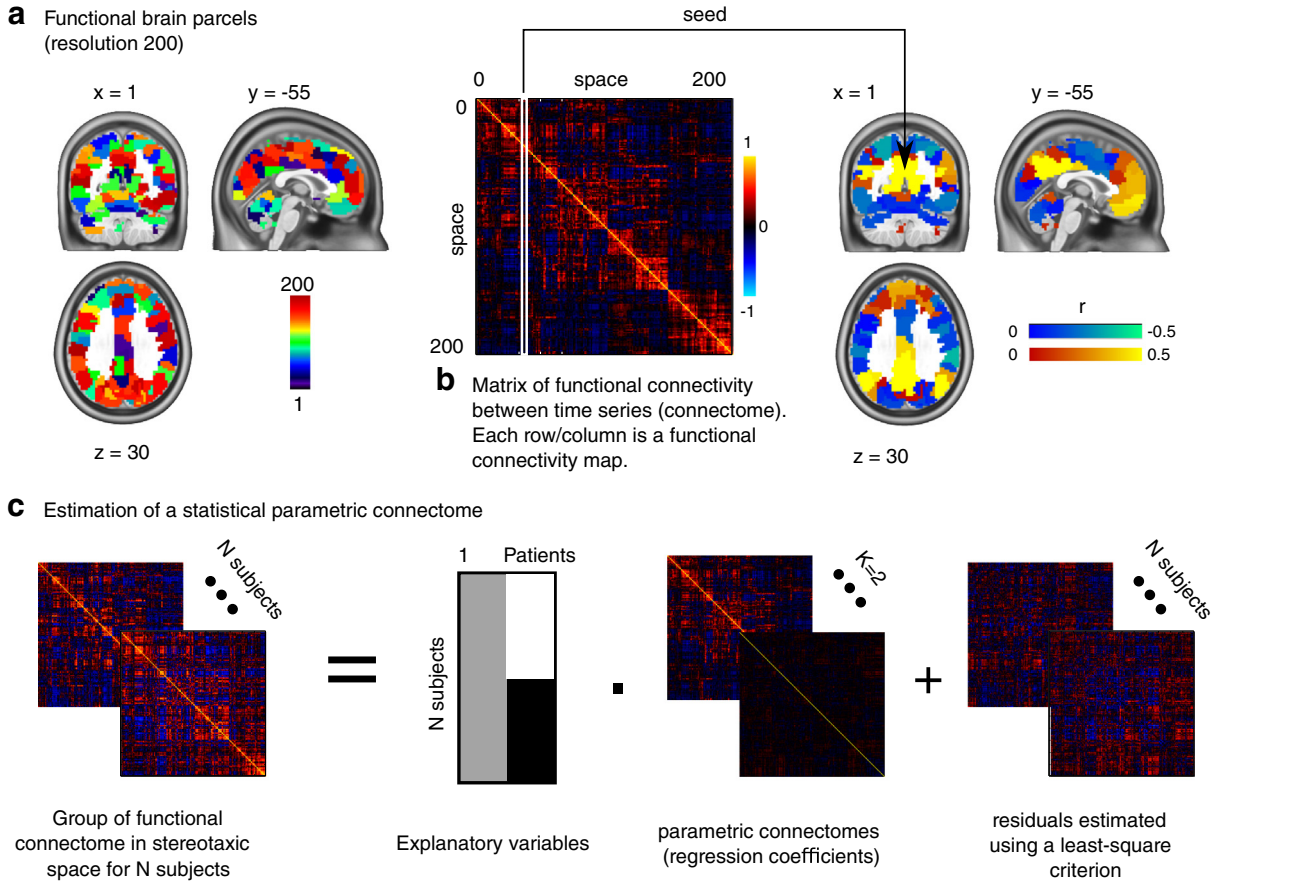

Fig. 1. General linear model applied to connectomes. The connectivity is measured between R brain parcels generated through a clustering algorithm (panel a). The connectome is a $R \times R$ matrix measuring functional connectivity between- and within-parcels (panel b). The association between phenotypes and connectomes is tested independently at each connection using a general linear model at the group level (panel c). The results presented here are for illustration purpose only, and not related to the results presented in the application sections of the manuscript.

connectome. Connectomes are generated independently at each resolution. See Supplementary Material S1 for a more formal description of the connectome generation.

GLM-connectome analysis

For R parcels, there are exactly $L = R(R+1)/2$ distinct elements in an individual connectome $\mathbf{Y}$. This connectome can be stored as a $1 \times L$ vector, where the brain connections have been ordered arbitrarily along one dimension. When functional data is available on N subjects, the group of connectomes is then assembled into a $N \times L$ array $\mathbf{Y} = (y_{n,l})$, where $n = 1,\ldots,N$ each code for one subject and $l = 1,\ldots,L$ each code for one connection. A general linear model (GLM) can then be used to test the association between brain connectivity and a trait of interest, such as the age or sex of participants. All of these C explanatory variables are entered in a $N \times C$ matrix $\mathbf{X}$. The variables are typically corrected to have a zero mean across subjects, and an intercept (i.e. a column filled with 1) is added to $\mathbf{X}$. The GLM relies on the following generative model:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}, \quad (1)$$

- $\mathbf{Y}$ is a $N \times L$ matrix where each row codes for a subject, and each column codes for a connection,
- $\mathbf{X}$ is a $N \times C$ matrix of explanatory variables (or covariates) where each row codes for a subject and each column codes for a covariate,
- $\mathbf{B}$ is an unknown $C \times L$ matrix of linear regression coefficients where each row codes for a covariate and each column codes for a connection,
- $\mathbf{E}$ is a $N \times L$ random (noise) variable, with similar coding as $\mathbf{Y}$.

We relied on the following parametric assumptions on the noise $\mathbf{E}$ are (1) that its rows are independent; (2) that each element follows a normal distribution with zero mean, and (3) that the variance of all elements are constant within a column, also called the homoscedasticity assumption. As the data generated from different subjects are statistically independent the first assumption is reasonable. We tested the normality and homoscedasticity assumptions on real datasets. Under these parametric assumptions, the regression coefficients $\mathbf{B}$ can be estimated with ordinary least squares and, for a given "contrast" vector $\mathbf{c}$ of size $1 \times C$, the significance of $\mathbf{c}\hat{\mathbf{B}}$ can be tested with a connectome of t-test $(t_l)_{l=1}^{L}$, with associated p-values $(p_l)_{l=1}^{L}$. The quantity p_l controls for the risk of false positive findings at each connection l. The GLM applied on connectomes is illustrated in Fig. 1c. See Supplementary material S2 for the equations related to the estimation and testing of regression coefficients in GLM-connectome analysis.

The Benjamini–Hochberg FDR procedure

The number L of tests $(p_l)_{l=1}^{L}$ grows quadratically with the resolution K. The significance value applied on p_l within a resolution thus needs to be adjusted for this multiple comparison problem. We implemented the Benjamini–Hochberg (BH) procedure (Benjamini and Hochberg, 1995) to control the FDR at a specified level α within resolution (Supplementary Material S3). The idea of the FDR is not to strictly

Please cite this article as: Bellec, P., et al., Impact of the resolution of brain parcels on connectome-wide association studies in fMRI, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.07.071

control the probability to observe at least one false positive (a quantity know as family-wise error, FWE), but rather to control, on average, the proportion of false positive amongst the findings. Note that controlling for the FDR is not necessarily a more liberal attitude than controlling for the FWE: if the global null hypothesis is verified, i.e. all discoveries are false positive, then the FDR is exactly the FWE. The BH procedure was designed for independent tests, yet it was shown to have a satisfactory behavior even in the presence of positive correlation between the tests p_i (Benjamini and Yekutieli, 2001). On simulations, the specificity of the FDR-BH algorithm was assessed in the presence of realistic correlations between tests.

Multiresolution GLM-connectome analysis

It is possible to assess how resolution impacts a GLM-connectome by replicating the analysis with different numbers of clusters (Fig. 2). A systematic approach would consist of a regular grid, e.g. from 10 to 300 brain parcels, with a step of 10. The GLM results at such resolutions may however be highly redundant, as some parcels may be found identically at different resolutions if those are close. An alternative strategy would be to select a limited number of non-redundant resolutions that span a given range (e.g. 10 to 300). For this purpose, we used the multiresolution stepwise selection (MSTEPS) algorithm recently proposed by Bellec (2013) to select a subset of resolutions that provides an accurate summary of the stable features of brain clusters observed across all possible resolutions. We evaluated both strategies (regular grid and sparse subset) in this work, both on simulations and real data.

FDR within and across resolutions

Testing GLM on connectomes at multiple resolutions introduces a new level of multiple comparisons, this time across resolutions rather than across connections. The FDR across resolutions is the FDR for the overall family of tests including all employed connections and resolutions. For example, if two resolutions were used, $K \in \{10{,}100\}$, there would be $K(K+1)/2$ tests at each resolution (i.e. 55 and 5050). If there were 10 discoveries at resolution 10, 1 of which was a false positive, and 200 discoveries at resolution 100, 10 of which were false positive, the ratio of false discoveries for this simulation would be $1/10 = 0.1$ within resolution 10, and $10/200 = 0.05$ within resolution 100. The ratio of false discoveries across resolutions would be $(1+10)/(10+200) = 0.052$. The FDR within resolution and across resolutions would be the average of the corresponding false discovery proportions over many replications of the testing procedure.

In the absence of any true association at any resolution (global null hypothesis), the FDR matches with the FWE, and is inflated when multiple resolutions of tests are combined. The FDR across resolutions is thus expected to be inflated compared to the FDR within resolution under the global null hypothesis. By contrast, in the presence of a substantial amount of true discoveries, Efron (2008) hypothesized based on simulations that the FDR across many resolutions of tests would match the FDR within each resolution. The rationale for this hypothesis is that, in the presence of signal, the FDR controls for a proportion, which behaves well when multiple resolutions are combined. We assessed this hypothesis on realistic simulations of multiresolution GLM-connectome analysis.

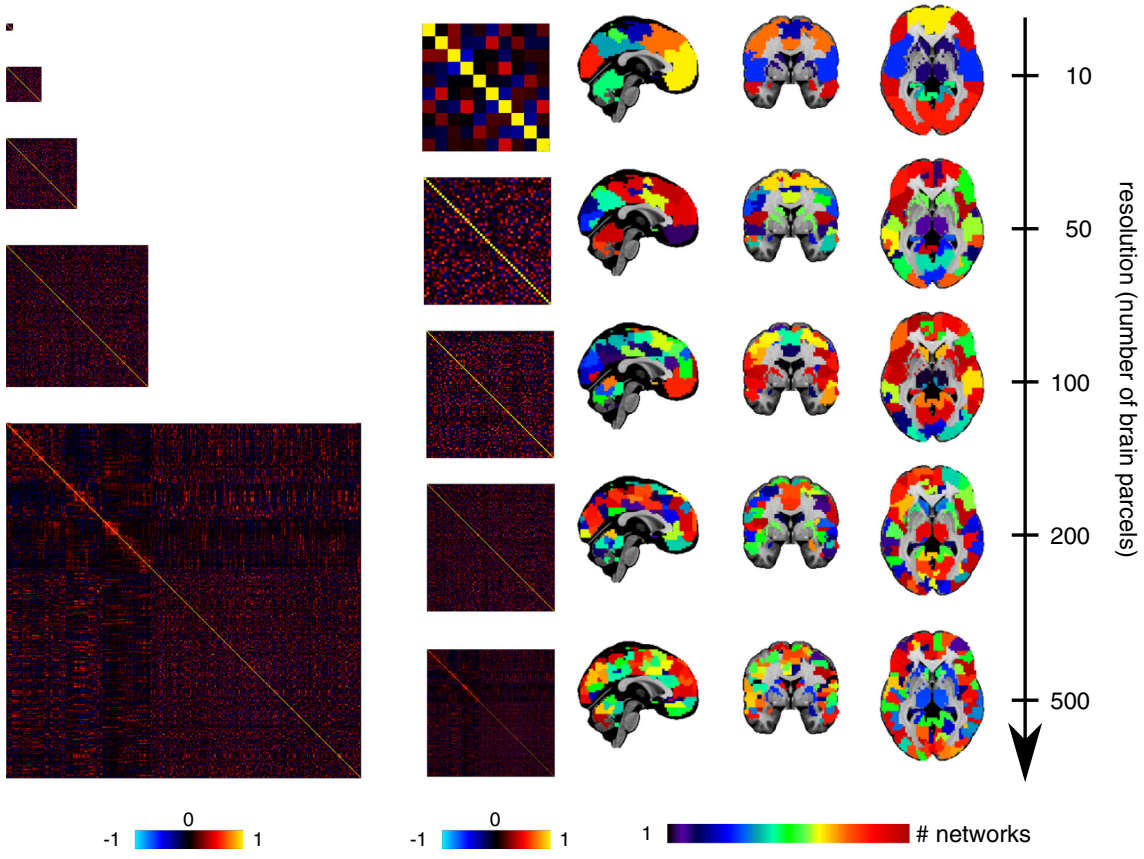

Fig. 2. General linear model applied to connectomes at multiple spatial resolutions. The generation of data-driven brain parcels is iterated at different resolutions (number of brain parcels), using the bootstrap analysis of stable clustered (BASC), with a hierarchical clustering using Ward's criterion. The statistical parametric connectomes are represented using both their real size (left column) and after rescaling to fit identical size (middle column) to illustrate the quadratic increase in the number of connections (multiple comparisons) that comes with an increase in the number of parcels. The results presented here are for illustration purpose only, and not related to the results presented in the application sections of the manuscript.

Please cite this article as: Bellec, P., et al., Impact of the resolution of brain parcels on connectome-wide association studies in fMRI, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.07.071

Multiresolution omnibus test

As outlined above, the most problematic scenario when exploring a GLM analysis at multiple resolutions is the global null hypothesis. To address this issue, we developed an omnibus test of the overall presence of true associations in the GLM, pooling FDR discoveries across all resolutions. At a given resolution K, V_K is the percentage of discoveries, i.e. the number of significant tests as identified by FDR-BH at a given level α, divided by the total number of tests. The overall volume of discoveries V is defined as the average of V_K across all resolutions K. The omnibus test is based on the probability that V could be observed under the global null hypothesis $(\mathcal{G}_0)$, i.e. *no non-null effect at any connection and any resolution*. This test proceeds by comparing the volume of discoveries V observed empirically in the group sample against the volume of discoveries V^* that could be observed under $(\mathcal{G}_0)$. The following steps are used to generate replications of V^* under $(\mathcal{G}_0)$:

- the GLM is first applied with a reduced model where the explanatory variable of interest (as selected through the contrast) is removed.
- A permutation of the residuals is generated as described in Anderson (2002), see Appendix A. In order to respect the dependencies between connectivity estimates within and across resolutions, the same permutation of the subjects is applied to all of the connections and resolutions.
- A replication of connectomes is generated under $(\mathcal{G}_0)$ by adding the permuted residuals to the estimated mixture of reduced explanatory variables.
- The detection procedure is applied, under $(\mathcal{G}_0)$, and the total volume of discoveries V^* is derived.

A Monte-Carlo approximation, with typically 10,000 permutation samples on real data, is used to estimate a false-positive rate p when testing against the global null hypothesis. Note that a single omnibus test is derived, controlling for the FWE of the experiment as a whole. If this test passed significance, each resolution is examined with a control of the FDR at $\alpha = 0.05$, uncorrected for multiple comparisons across resolutions. If the omnibus test does not reach significance, then no connection at any resolution is deemed significant.

Evaluation on simulated datasets with independent tests

Methods

Data-generating procedure

We started by simple simulations of independent tests, to assess to which extent the hypothesis of Efron (2008) was robust to different scenarios, and if the omnibus test would systematically ensure that the FDR across resolutions would be well controlled. A number K of test resolutions were generated independently, each one composed of L_k tests, $k = 1,\ldots,K$. Each resolution included a set proportion of true non-null hypotheses π_1, identical for all resolutions. If $\pi_1 L_k$ was not an integer, the number of true positives n_k was set to either $\lfloor \pi_1 L_k \rfloor$ or $\lfloor \pi_1 L_k \rfloor + 1$, with probabilities such that on average over many simulations $\mathbb{E}(n_1) = \pi_1 L_k$. For a non-null test l, the associated p-value was simulated as:

$$y_l = \theta + z_l, z_l \sim \mathcal{N}(0,1), \tag{2}$$

$$p_l = \Pr(x \geq y_l | x \sim \mathcal{N}(0,1)), \tag{3}$$

where $\mathcal{N}(0,,1)$ was a Gaussian distribution with zero mean and unit variance, and $\theta > 0$ was a simulation parameter (further called effect size). The null tests were generated the same way, but with an effect size $\theta = 0$.

Simulations scenarios

For each experiment, all combinations of effect size in the grid {2,3,5} and π_1 in the grid {0 %,1 %,2 %,5 %,10 %} were considered. We implemented a series of experiments:

- We first checked how the FDR across resolutions behaved as a function of the number of resolutions K, with K in {2,5,10} and L_k equals to 1000 (corresponding approximately to the number of tests at resolution 45).
- We then checked how the FDR across resolutions behaved as a function of the number of tests per resolution L_k with L_k identical for all k, and in the grid {100,1000,10000} (corresponding roughly to resolutions 14, 45 and 141), and $K = 5$ resolutions.
- We checked how the FDR across resolutions behaved for a number of resolutions and a number of tests per family that would be comparable to situations encountered in a multiresolution GLM-connectome analysis.
  - We first tested the resolutions selected by MSTEPS on the SCHIZO dataset (see Section Application to real datasets), i.e. $K = 7$ and L_k in (28, 136, 325, 1540, 6555, 19900, 53956), corresponding to the number of tests at resolutions (7, 16, 25, 55, 114, 199, 328).
  - We then tested the procedure on $K = 30$ and L_k ranging from 55 to 45150, which would be equal to the number of tests associated with a regular grid covering resolutions 10 to 300 with a step of 10.
  - We finally tested the behavior of smaller grids, with a number of tests equivalent to GLM tests over resolutions ranging from 10 to either 50, 100 or 300 (with a step of 10).

Computational environment

All the experiments reported in the paper were performed using the NeuroImaging Analysis Kit (NIAK[1]) version 0.12.18, under CentOS version 6.3 with Octave[2] version 3.8.1 and the Minc toolkit[3] version 0.3.18. Analyses were executed in parallel on the "Guillimin" supercomputer[4], using the pipeline system for Octave and Matlab (Bellec et al., 2012), version 1.0.2. The scripts used for processing can be found on Github[5].

Statistical testing procedure

For each simulation scenario, the FDR-BH procedure was applied to each resolution independently, with a significance level α in the grid {0.01,0.05,0.1,0.2}. To estimate the distribution of the volume of discovery under the global null, 1000 samples were generated with the parameters θ and π_1 set to zero, for each choice of K and L_k. These replications under the global null were used to generate the p-values of the omnibus test for all simulations with identical K and L_k.

Effective FDR, sensitivity and omnibus test

For each resolution, the effective FDR and sensitivity were evaluated for tests at a single parcellation with the specified number of parcels. The effective FDR was computed as the number of false discoveries divided by the total number of discoveries, averaged across 1000 replications of each simulation scenario. The effective sensitivity was computed as the number of true discoveries, divided by the number of true non-null hypotheses present at this resolution, and averaged across the 1000 replications. To compute the FDR across resolutions, the same procedure to estimate the effective FDR was applied to the combination of tests pooled across all resolutions. Finally, we also derived a modified

[1] http://www.nitrc.org/projects/niak/.
[2] http://gnu.octave.org.
[3] http://www.bic.mni.mcgill.ca/ServicesSoftware/ServicesSoftwareMincToolKit.
[4] http://www.calculquebec.ca/en/resources/compute-servers/guillimin.
[5] https://github.com/SIMEXP/glm_connectome.

Please cite this article as: Bellec, P., et al., Impact of the resolution of brain parcels on connectome-wide association studies in fMRI, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.07.071

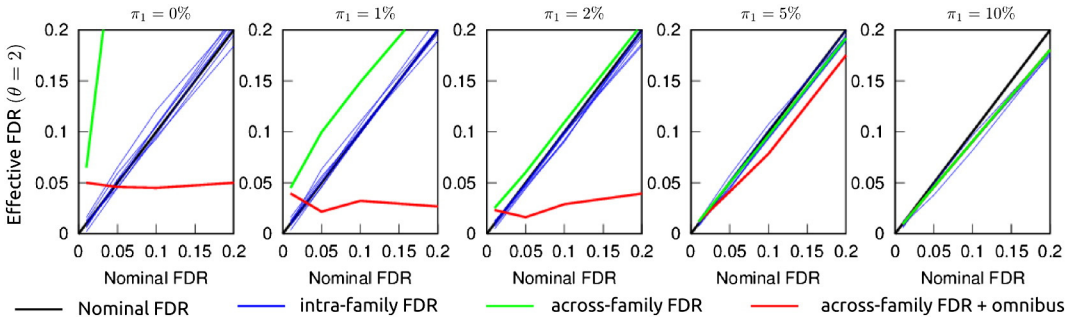

Fig. 3. Nominal vs effective FDR on simulations with independent tests ($K = 7$, L_k in (28,136,325,1540,6555,19900,53956), corresponding to the MSTEPS resolutions in the SCHIZO dataset). The effective FDR within resolution (blue), across resolutions (green) and across resolutions with omnibus (red) is plotted against the nominal FDR (black), for four levels: 0.01,0.05,0.1,0.2. Plots are presented for different proportions of non-null hypothesis per resolution π_1 (0 %,1 %,2 %,5 %,10 %), and an effect size $\theta = 2$. For large θ and/or π_1, the omnibus test is always rejected, and the green plot matches perfectly the red plot, which becomes invisible.

FDR and sensitivity, where the BH-FDR procedure was combined with the omnibus permutation test, at a significance level of $p < 0.05$.

Results

FDR within and across resolutions

The effective FDR within each resolution followed closely $(1 - \pi_1)\alpha$ (Fig. 3), which replicated a well-established result: the BH-FDR procedure is conservative for independent tests. For example, for $\alpha = 0.2$ and a π_1 of 10 %, the effective FDR was approximately 0.18. The FDR across resolutions followed a smooth transition between two regimes. In the first regime, called "liberal" ($\pi_1 = 0$ %, Fig. 3), the FDR matched with the FWE, i.e. the probability to have one or more false positive. As expected for a FWE, the effective FDR across resolutions was largely superior to the prescribed level α. In the second regime, called "exact", the FDR across resolutions precisely followed the FDR within resolution (e.g. $\pi_1 = 10$ %, Fig. 3). The transition between these two regimes (liberal and exact) was smooth, and in situations that resembled the global null hypothesis (i.e. at low π_1 or effect size), the FDR across resolutions was more liberal than the nominal α, sometimes by a wide margin, e.g. $L_k = 1000$, $\pi_1 = 1$ % and $\theta = 2$ (Supplementary Fig. S1). Increasing the effect size, or increasing π_1 both pushed the FDR across resolutions towards the "exact" regime (Supplementary Fig. S1).

FDR across resolutions, with omnibus test

The effect of the omnibus test on the FDR across resolutions was particularly apparent under the global null hypothesis in all simulations: the FDR across resolutions matched the FWE, which was less than, or equal to, the $p < 0.05$ threshold of the omnibus test, as expected ($\pi_1 = 0$ %, Fig. 3). More generally, for simulation scenarios that represented a transition between the liberal and exact regimes of the FDR across resolutions, the application of the omnibus test tended to make the FDR across resolutions more conservative. Note that for a nominal FDR lower than the threshold of the omnibus test (e.g. 0.01) and $\pi_1 = 0$ %, the effective FDR departed from the nominal level as the omnibus test only controlled the FWE at $p < 0.05$. Even for a nominal FDR larger than the threshold of the omnibus test, there was still no guarantee that the FDR across resolutions conformed to the specified α level, as can be seen for $L_k = 1000$, $\pi_1 = 1$ % and $\theta = 2$ in Supplementary Fig. S1, where the effective FDR was larger than 0.1 for a nominal FDR of 0.05.

Influence of the number of resolutions K

By varying K in {2,5,10} for a fixed number of tests per resolution ($L_k = 1000$ for all k), we found that the transition between the liberal and exact regime of the FDR across resolutions took longer when the number of resolutions increased. For $\pi_1 = 2$ % and $\theta = 2$, and a nominal FDR $\alpha = 0.05$, the effective FDR was about 0.07 with $K = 2$, while it increased to almost 0.1 for $K = 10$ (Supplementary Fig. S2). This was expected as more families K also mean a more severe multiple comparison problem under the global null, where the FDR matches the FWE and thus increases with more independent tests.

Influence of the number of tests per resolution L

By varying L_k in {100,1000,10000} (with L_k identical for all k) for a fixed number of resolutions $K = 5$, we found that the transition between the liberal and exact regime of the FDR was quicker when the number of tests per resolution increased. In other words, the exact regime appears as an asymptotic behavior of the FDR across resolutions, when the number of tests per resolution becomes large. For example, for $\pi_1 = 2$ % and $\theta = 2$, and a nominal FDR $\alpha = 0.05$, the effective FDR across resolutions went from above 0.1 with 100 tests per resolution to below 0.06 with 10000 tests per resolution (Supplementary Fig. S3).

Sensitivity

Increasing either π_1 or θ increased the overall sensitivity of the tests. The sensitivity peaked at very low resolutions, and decreased exponentially to reach a plateau around resolution 10 to 50. After this initial loss in sensitivity, the sensitivity was uniform across resolutions (Fig. 4). Identical conclusions were reached with resolutions akin to connectome testing on a regular grid of resolutions ranging from 10 to either 50, 100, or 300 parcels (with a step of 10). See Supplementary Figs. S4–S6 for sensitivity and effective FDR results in all scenarios.

Evaluation on simulated datasets with dependent tests

Methods

Data-generating procedure

We designed a simulation framework for multiresolution GLM-connectome analysis in the presence of dependencies between tests, both within resolution and across resolutions. To ensure that these dependencies would be as realistic as possible, semi-synthetic datasets were generated starting from a large real sample (Cambridge) released as part of the 1000 functional connectome project[6] (Biswal et al., 2010). This sample (Liu et al., 2009) included resting-state fMRI time series (eyes opened, TR of 3 s, 119 volumes per subject) collected with a 3 T scanner on 198 healthy subjects (75 males), with an age ranging from 18 to 30 yrs. All the datasets were preprocessed and resampled in stereotaxic space, as described in the Methods Section. A region growing algorithm was used to extract 483 regions, common to all subjects, as described in Bellec et al. (2010). For each subject, the average

[6] http://fcon_1000.projects.nitrc.org/fcpClassic/FcpTable.html.

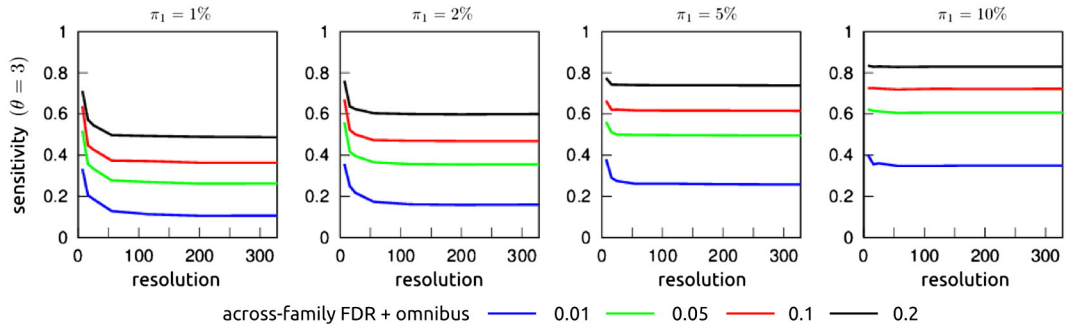

Fig. 4. Sensitivity on simulations with independent tests, $K = 7$, L_k in (28,136,325,1540,6555,19900,53956), corresponding to the number of connections associated with the resolutions selected by MSTEPS on the SCHIZO dataset. The sensitivity is plotted as a function of resolutions (number of brain parcels) at four tested (within-resolution) FDR levels: 0.01,0.05,0.1,0.2. For each resolution, the sensitivity was evaluated for tests at a single parcellation with the specified number of parcels. A test is only considered as significant if in addition an omnibus test against the global null hypothesis across resolutions as been rejected at $p < 0.05$. Each column corresponds to a certain proportion of non-null hypothesis per resolution π_1 with an effect size $\theta = 3$.

functional time series and associated connectomes were generated using these regions[7] (see Section Functional connectome). The average connectome across all subjects was derived, and a hierarchical clustering procedure (with Ward's criterion) was applied to derive a hierarchy of brain parcels at all possible resolutions, ranging from 1 to 483. The simulation procedure relied on the manual selection of a critical resolution K and a particular cluster k at this resolution. For each simulation, two non-overlapping subgroups of subject (N subjects per group) were randomly selected. A circular block bootstrap (CBB) procedure was applied to resample the individual time series, using identical time blocks within each cluster, and independent time blocks in different clusters. This resampling scheme ensured that within-cluster correlations were preserved, while between-cluster correlations had a value of zero on average. Finally, for the subjects selected to be in the first group, a single realization of an independent and identically distributed Gaussian variable, where each time point had a zero mean and a variance of a^2, was added to the time series of the regions inside cluster k, after the time series were corrected to a zero temporal mean and a variance of $(1 - a^2)$. The addition of this signal increased the intra-parcel connectivity of the cluster including cluster k for all resolutions smaller or equal to K, and increased the within- as well as between-parcel connectivity for all clusters included in cluster k for resolutions strictly larger than K. Because of the absence of correlations between parcels at resolution K (due to the CBB resampling), all other connections within- or between clusters at every resolution were left unchanged by this procedure. It was thus possible to know exactly which connections were true or false null hypothesis in the group difference at every resolution. Supplementary Figs. S7 and S8 outline the procedure of multiresolution connectome simulation.

Effect size and proportion of non-null hypothesis

A number of clusters of reference were handpicked such that the proportion of non-null hypothesis $\pi_1(k)$ would be about 1 %, 2 %, 5 % and 10 % at all resolutions k. Note that these reference clusters were used to set true non-null hypotheses at all the resolutions of analysis, yet the subdivisions (or merging) associated with these clusters represented a varying proportion of the number of clusters at any given resolution. As a consequence, and unlike simulations of independent tests, $\pi_1(k)$ was dependent on the resolution k. Two values for a^2 were selected: 0.1 and 0.2. The effect size associated with a given a^2 actually depended on the within-cluster correlations, between-subject variance in connectivity as well as the resolution of analysis. Two sample sizes were investigated: $N = 40$ (20 subject per group), and $N = 100$ (50 subjects per group). See Supplementary Material Fig. S9 for plots of the effect size and percentage of true non-null hypotheses as a function of resolution. For each simulation scenarios, 1000 Monte-Carlo samples were generated and subjected to a GLM-connectome analysis, with FDR levels of 0.01, 0.05, 0.1 and 0.2, as well as an omnibus test at $p < 0.05$. The same resolutions were tested here as in the simulations for independent tests: the resolutions selected by MSTEPS on the SCHIZO experiment, and a regular grid from 10 to 300 clusters (with a step of 10).

Simulations under the global null

To assess the behavior of the testing procedures in the absence of any signal, we also ran experiments under the global null. In that case real connectomes were generated for randomly selected and non-overlapping groups of subjects, and then a testing procedure was implemented to assess the significance of group differences. In these experiments, no bootstrap was performed on individual time series nor any signal was added. The experiments simply consisted in comparing real connectomes between random groups of subjects sampled from identical populations, using real dependencies between tests.

Robustness to the choice of clusters

Finally, we also investigated how the procedure behaved when the clusters used in the testing procedures did not match exactly with the clusters that were used to generate the simulations. For this purpose, for each simulation, no structured signal was generated in 30% of arbitrarily selected regions in the cluster of reference, but the same structured signal was instead added to an equivalent number of arbitrarily selected regions from other clusters. The same regions were selected across all simulations to simulate a systematic departure of the test clusters from the ground truth clusters. The multiresolution clusters without perturbations were used in the statistical testing procedures. In this setting, many connections outside of the cluster of reference ended up with very small effects, and we did not investigate the specificity given the very large number of true non-null hypotheses and large variations in effect size. However, we did investigate the sensitivity of the FDR-BH procedure, using the same definition of true non-null hypothesis as with the simulations without perturbation.

Results

Effective FDR within resolution

Fig. 5 represents the effective FDR as a function of resolution for the GLM-connectome procedure, in the case of a regular grid of resolutions covering 10 to 300 brain parcels and a perfect match between the true and test clusters. The effective FDR was conservative within resolution

[7] The average time series have been publicly released at http://figshare.com/articles/Cambridge_resting_state_fMRI_time_series_preprocessed_with_NIAK_0_12_4/1159331.

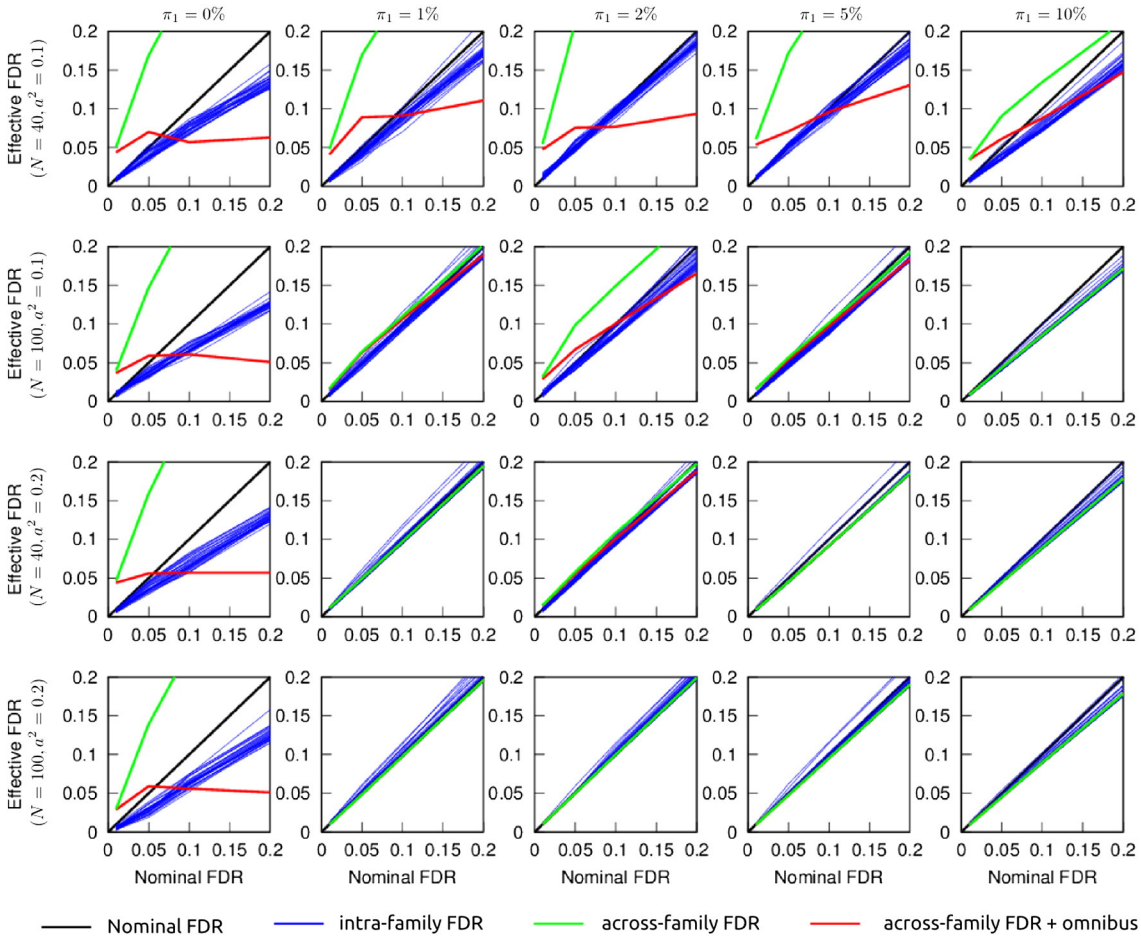

Fig. 5. Nominal vs effective FDR on simulations with dependent tests ($K = 30$, L_k ranging from 55 to 45150, corresponding to the number of connections associated with a regular grid of resolutions covering 10 to 300 with a step of 10). The effective FDR is plotted against the nominal FDR within each resolution (blue plots), across all resolutions (green plots) and across all resolutions, combined with an omnibus test for rejection of the global null hypothesis (red plot). The expected (nominal) values are represented in black plots, corresponding to the four tested FDR levels: 0.01,0.05,0.1,0.2. Each column corresponds to a certain proportion of non-null hypothesis per resolution π_1 (0 %,1 %,2 %,5 %,10 %), and each row corresponds to a different combination of effect and sample size N in {40,100}, a^2 in {0.1,0.2}, see text for details. Please note that in the presence of strong signal (large θ and/or π_1), the omnibus test is always rejected, and the green plot matches perfectly the red plot, which becomes invisible.

on the simulations with dependent tests, e.g. the effective FDR was about 0.15 for a nominal FDR of 0.2. This is in contrast with the independent tests, where the control of the FDR within resolution was exact under the global null hypothesis. Our interpretation was that the large positive correlations present in fMRI time series caused the FDR-BH procedure to become conservative. In the presence of signal, the FDR within resolution was still well controlled, with the same $(1 - \pi_1)$ factor on the effective FDR as was observed with independent tests.

Effective FDR across resolutions

As was observed on independent tests, the FDR across resolutions transitioned between a "liberal" regime, in simulation scenarios close to the global null hypothesis, to an exact regime, where the FDR across resolutions matched the FDR within resolution (Fig. 5). The transition between regimes happened quite fast, with either $a^2 = 0.2$ or $N = 100$, as soon as π_1 was larger than 5 %. When combined with the omnibus test at $p < 0.05$, the FWE under the global null hypothesis became exact or conservative for a FDR level above 0.05. Note that for a nominal FDR lower than 0.05 (e.g. 0.01) and $\pi_1 = 0$ %, the effective FDR departed from the nominal level as the omnibus test only controlled the FWE at $p < 0.05$. Importantly, the omnibus test also made the procedure either conservative (for $\alpha \geq 0.1$) or only slightly liberal in the scenarios were the FDR across resolutions transitioned between the "liberal" and "exact" regimes, with the effective FDR in the range 0.06 to 0.09 for a nominal level of 0.05 in the worst cases (i.e. $N = 40$, $a^2 = 0.1$ and $\pi_1 = 1$ %). The conclusions were identical when using a regular grid of $K = 30$ resolutions ranging from 10 to 300 parcels (with a step of 10), or $K = 7$ resolutions identical to those selected by MSTEPS on the SCHIZO dataset (Supplementary Fig. S10).

Sensitivity

When the true and test clusters perfectly matched, the sensitivity across resolutions followed a similar pattern in all scenarios: a decrease in sensitivity with increasing resolutions, although not as sharp as what was observed on simulations with independent tests (Fig. 6, see Supplementary Figs. S11 and S12 for all tested scenarios). This closely mirrored the large increase in effect size at low resolutions, due to averaging on clusters that perfectly matched the simulated signal (Supplementary Fig. S9). We noted that the simulation settings where departure from nominal levels were observed were also characterized by very low rate of discoveries, notably at high resolution, with sensitivity below 0.1 for resolutions higher than 50 and falling to zero for resolutions higher than 150 (Supplementary Fig. S11, first row). By contrast, when introducing a 30% mismatch between the true and test clusters, increases in sensitivity were observed across a wider range of low resolutions, e.g. $N = 100$, $a^2 = 0.1$ and $\pi_1 = 10$ %, or even at high resolutions, e.g. $\pi_1 = 2$ % in Fig. 6. This again reflected the more variable profiles of effect size as a function of resolutions across scenarios after the introduction of a mismatch between the true and test clusters. These simulations demonstrated the possibility to have increase in sensitivity as a

Please cite this article as: Bellec, P., et al., Impact of the resolution of brain parcels on connectome-wide association studies in fMRI, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.07.071

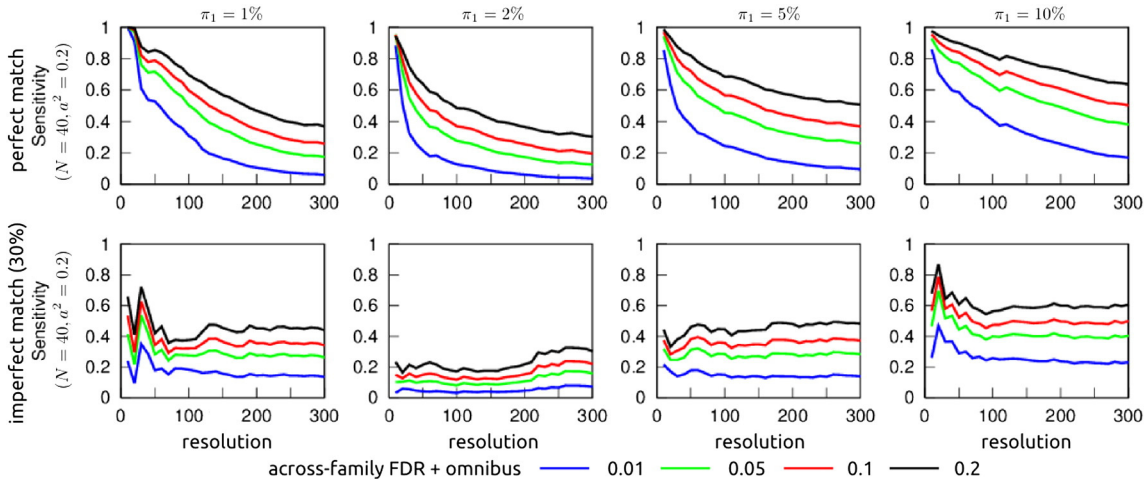

Fig. 6. Sensitivity on simulations with dependent tests ($K = 30$, L_k ranging from 55 to 45150, corresponding to the number of connections associated with a regular grid of resolutions covering 10 to 300 with a step of 10). The sensitivity is represented as a function of resolution, for four FDR levels: 0.01,0.05,0.1,0.2, with either no mismatch or a 30% mismatch between the true and test clusters. For each resolution, the sensitivity was evaluated for tests at a single parcellation with the specified number of parcels. In addition to FDR control within resolution, an omnibus test at $p < 0.05$ was performed. Each column corresponds to a certain proportion of non-null hypothesis per resolution π_1 (0 %,1 %,2 %,5 %,10 %), with a sample size $N = 40$ and $a^2 = 0.2$.

function of resolution, and that these gains would potentially be dependent on the effect size, the mismatch between the true/test clusters, as well as the sample size. These observations were made for a regular grid of resolutions, but were identical using the MSTEPS resolutions from the SCHIZO dataset (not shown).

Application to real datasets

Methods

Participants

We evaluated the GLM-connectome on three real datasets: (1) a study (SCHIZO) comparing patients suffering from schizophrenia with healthy control subjects; (2) a study (BLIND) on patients suffering of congenital blindness, compared to sighted controls; and (3) a study (MOTOR) where resting-state data connectivity was compared before and after learning of a motor task. The SCHIZO dataset was contributed by the Center for Biomedical Research Excellence (COBRE) to the 1000 functional connectome project[8] (Biswal et al., 2010). The sample comprised 72 patients diagnosed with schizophrenia (58 males, age range = 18–65 yrs) and 74 healthy controls (51 males, age range = 18–65 yrs). The BLIND (Collignon et al., 2011) and MOTOR (Albouy et al., 2015) datasets were acquired at the Functional NeuroImaging Unit, at the Institut Universitaire de Gériatrie de Montréal, Canada. Participants gave their written informed consent to take part in the studies, which were approved by the research ethics board of the Quebec BioImaging Network (BLIND, MOTOR), as well as the ethics board of the Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal (BLIND). The BLIND dataset was composed of 14 congenitally blind volunteers recruited through the Nazareth and Louis Braille Institute (10 males, age range = 26–61 yrs) and 17 sighted controls (8 males, age range = 23–60 yrs). The MOTOR sample included 54 healthy young participants (33 males, age range = 19–33 yrs).

Acquisition

Resting-state fMRI scans were acquired on a 3 T Siemens TrioTim for all datasets. One single run was obtained per subject for either the SCHIZO or BLIND dataset while two runs were acquired in each subject for the MOTOR dataset, one immediately preceding and one following the practice on a motor task. For the SCHIZO dataset, 150 EPI blood-oxygenation level dependent (BOLD) volumes were obtained in 5 min (TR = 2 s, TE = 29 ms, FA = 75°, 32 slices, voxel size $3 \times 3 \times 4$ mm^3, matrix size 64×64), and a structural image was acquired using a multi-echo MPRAGE sequence (TR = 2.53 s, TE = 1.64, 3.5, 5.36, 7.22, 9.08 ms, FA = 7°, 176 slices, voxel size $1 \times 1 \times 1$ mm^3, matrix size 256×256). For the BLIND dataset, 136 EPI BOLD volumes were acquired in 5 min (TR = 2.2 s, TE = 30 ms, FA = 90°, 35 slices, voxel size = $3 \times 3 \times 3.2$ mm^3, gap = 25%, matrix size = 64×64), and a structural image was acquired using a MPRAGE sequence (TR = 2.3 s, TE = 2.91 ms, FA = 9°, 160 slices, voxel size = $1 \times 1 \times 1.2$ mm^3, matrix size = 240×256). For the MOTOR dataset, 150 EPI volumes were recorded in 6 min 40 s (TR = 2.65 s, TE = 30 ms, FA = 90°, 43 slices, voxel size = $3.4 \times 3.4 \times 3$ mm^3, gap = 10%, matrix size = 64×64), and a structural image was acquired using a MPRAGE sequence (TR = 2.3 s, TE = 2.98 ms, FA = 9°, 176 slices, voxel size = $1 \times 1 \times 1$ mm^3, matrix size = 256×256).

Motor task

Between the two rest runs of the MOTOR experiment, subjects were scanned while performing a motor sequence learning task with their left non-dominant hand. 14 blocks of motor practice were interspersed with 15 s rest epochs. Motor blocks required subjects to perform 60 finger movements, ideally corresponding to 12 correct five-element finger sequences. The duration of the practice blocks decreased as learning progressed. It should be noted that the effect of motor learning per se on the subsequent rest run could not be distinguished from that of a mere motor practice/fatigue effect in the present experimental setting.

Preprocessing

Each fMRI dataset was corrected for inter-slice difference in acquisition time and the parameters of a rigid-body motion were estimated for each time frame. Rigid-body motion was estimated within as well as between runs, using the median volume of the first run as a target. The median volume of one selected fMRI run for each subject was coregistered with a T1 individual scan using Minctracc (Collins and Evans, 1997), which was itself non-linearly transformed to the Montreal Neurological Institute (MNI) template (Fonov et al., 2011) using the CIVET pipeline (Ad-Dab'bagh et al., 2006). The MNI symmetric template was generated from the ICBM152 sample of 152 young adults, after 40 iterations of non-linear coregistration. The rigid-body transform, fMRI-to-T1 transform and T1-to-stereotaxic transform were all combined, and the functional volumes were resampled in the MNI space at a 3 mm isotropic resolution. The scrubbing method of Power et al.

[8] http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html.

(2012), was used to remove the volumes with excessive motion (frame displacement greater than 0.5 mm). A minimum number of 60 unscrubbed volumes per run, corresponding to ~ 180 s of acquisition, was then required for further analysis. For this reason, some subjects were rejected from the subsequent analyses: 16 controls and 29 schizophrenia patients in the SCHIZO dataset (none in either the BLIND or MOTOR datasets). The following nuisance parameters were regressed out from the time series at each voxel: slow time drifts (basis of discrete cosines with a 0.01 Hz high-pass cut-off), average signals in conservative masks of the white matter and the lateral ventricles as well as the first principal components (95% energy) of the six rigid-body motion parameters and their squares (Giove et al., 2009). The number of confounds regressed from the individual time series ranged from 12 to 18 for the MOTOR sample, from 11 to 15 for the BLIND sample, and from 10 to 17 for the SCHIZO sample. The fMRI volumes were finally spatially smoothed with a 6 mm isotropic Gaussian blurring kernel. Note that the preprocessed fMRI time series for the COBRE experiment have been made publicly available[9].

Multiresolution parcellation

Brain parcellations were derived using BASC separately for each dataset, while pooling the patient and control groups in the SCHIZO and BLIND datasets, and runs in the MOTOR dataset. The BASC used 100 replications of the clustering of each individual time series, using circular block bootstrap, and 500 replications of the group clustering, using regular bootstrap. The functional group clusters were first generated on a fixed regular grid, from 10 to 300 clusters with a step of 10, identical for all three real datasets. The MSTEPS procedure was then implemented to select a data-driven subset of resolutions approximating the group stability matrices up to 5% residual energy, through linear interpolation over selected resolutions.

General linear model

For all GLM analyses, the covariates included an intercept, the age and sex of participants as well the average frame displacement of the runs involved in the analysis (two covariates of frame displacement were used in the MOTOR dataset, one per run). The contrast of interest was on a dummy covariate coding for the difference in average connectivity between the two groups for SCHIZO and BLIND, and on the intercept (average of the difference in connectivity pre- and post-training) for the MOTOR dataset. Note that for the motor dataset the difference in connectivity between the second run and the first run was entered in the group-level GLM, in place of the individual connectome. All covariates except the intercept were corrected to a zero mean.

Modeling assumptions

The parametric GLM relies on a series of assumptions, most critically the normality of distribution of the residuals of the tests, and the homoscedasticity of residuals, i.e. equal variance across subjects. For each connection and each contrast, the normality of distribution for the residuals of the regression was tested with a composite test[10]: Shapiro–Francia for platykurtic distributions and Shapiro–Wilk for leptokurtic distributions (Royston, 1993). A test for homoscedastic residuals was also implemented using the procedure of White (1980), where all variables as well as their two-way interactions (including squared variables) were regressed against the square of the residuals, and an F test was performed on the combination of all exploratory variables. A p value was generated at each connection, both for the normality and the homoscedasticity tests, for the highest resolution selected by MSTEPS, and multiple comparisons across all connections were corrected with the FDR-BH procedure ($q < 0.05$). In addition to the MOTOR, BLIND and SCHIZO datasets, the Cambridge dataset previously used in the simulations was also employed here. The GLM only included an intercept and an arbitrary group difference, for different sample sizes ($N \in \{40,100,180\}$), in order to investigate how the testing of assumptions behaved for different sample sizes.

Results

Modeling assumptions

No tests of departure from normality passed correction for multiple comparison using FDR-BH at $q < 0.05$. However, some trends towards significance were observed in all datasets, in particular with a large sample size. For a threshold of $p < 0.05$, uncorrected for multiple comparisons, the normality hypothesis was rejected for 9 %, 6.8 % and 11 % of connections, for the MOTOR, BLIND and SCHIZO experiments, respectively (Supplementary Fig. S13).

No test for heteroscedasticity survived a correction for multiple comparisons with FDR-BH at $q < 0.05$, and there was no apparent trend. At $p < 0.05$, the homoscedasticity hypothesis was rejected for 3.4 %, 4.2 % and 7.4 % of connections in the MOTOR, BLIND and SCHIZO experiments, respectively. The trends observed for heteroscedasticity testing were similar to those observed in the Cambridge dataset, using random subgroups that are thus in fact homoscedastic (Supplementary Fig. S14).

Multiresolution discoveries

The MSTEPS procedure selected 6 resolutions for the MOTOR and BLIND samples, and 7 on the SCHIZO sample, ranging from 7 to 300+, see Table S1 for multi-level resolution parameters. The GLM-connectome detection generated maximal percentages of discoveries at low resolutions for the three datasets (Fig. 7). Using a grid from 10 to 300 resolutions with a step of 10, peak discoveries were detected at resolution 10 for the SCHIZO and MOTOR contrasts, and resolution 20 for the BLIND contrast. Peak percentages of discoveries were 30%, 2.3% and 15%, for the SCHIZO, BLIND and MOTOR contrasts, respectively. The omnibus test was significant ($p < 0.05$) for all three contrasts, whether using a large grid of 30 resolutions or the 6–7 resolutions identified with MSTEPS. The overall trend was that the rate of discoveries decreased as the number of parcels increased, with the largest discovery rate found below resolution 50, followed by a notable plateau from 50 to 100 clusters. These relationships between discovery rate and resolutions shared similarities with the sensitivity plots observed on simulations (Figs. 3, 6). While the absolute percentages of discoveries were quite different for the three datasets, the relative changes in discovery rate as a function of resolution were thus rather similar.

Spatial distribution of significant discoveries

Discovery percentage maps revealed which parcels were associated with the largest proportion of significant connections for any given parcel, see Fig. 8 for a representation of the BASC multiresolution parcels and associated discovery percentage maps for the SCHIZO analysis. For each contrast, results were shown for all 6–7 resolutions extracted with the MSTEPS procedure. The areas showing maximal percentage of discoveries were quite different for the three datasets (Fig. 9). Widespread effects were observed for the SCHIZO dataset at both cortical and subcortical levels (see also Fig. 8, for a volumetric representation). Parcels with the highest discovery rate were found in the temporal cortex, the medial temporal lobe, the anterior cingulate cortex and the basal ganglia. The BLIND contrast revealed more localized effects, in the occipital cortex and to a lesser extent in the temporal and frontal cortices. Finally, the MOTOR contrast identified significant effects within an extended visuomotor cortico-subcortical network.

Despite the highest rate of discoveries being observed at very low resolutions (10 and 20), the spatial distributions of discoveries were fairly consistent across resolutions. It was also interesting to note that

[9] http://figshare.com/articles/COBRE_preprocessed_with_NIAK_0_12_4/1160600.

[10] As implemented in the swtest.m procedure http://www.mathworks.com/matlabcentral/fileexchange/13964-shapiro-wilk-and-shapiro-francia-normality-tests/content/swtest.m, retrieved on 12/2014.

Please cite this article as: Bellec, P., et al., Impact of the resolution of brain parcels on connectome-wide association studies in fMRI, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.07.071

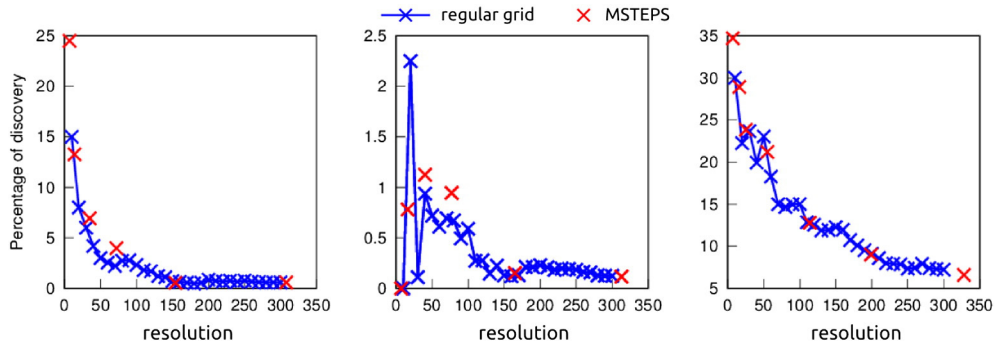

Fig. 7. Percentages of discovery as a function of resolutions (number of brain parcels). Plots show the percentage of discovery for the MOTOR, BLIND and SCHIZO contrasts. For each resolution, the percentage of discovery was evaluated for tests at a single parcellation with the specified number of parcels. The blue curve represents the resolutions selected on a regular grid, from 10 to 300 with a step of 10, and the red crosses show the resolutions selected by the MSTEPS procedure (see text for details).

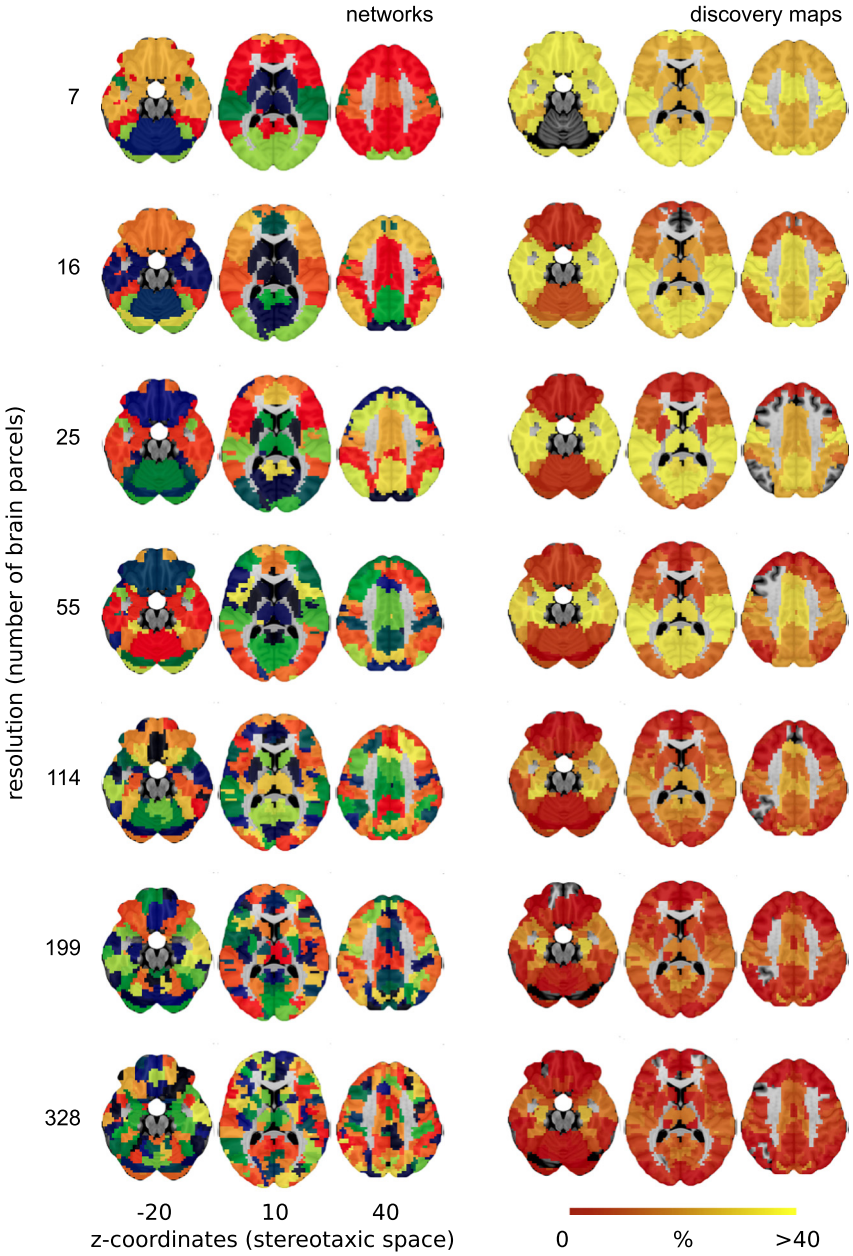

Fig. 8. MSTEPS parcels and percentage of discovery maps in the SCHIZO contrast, in volumetric space. Networks show the functional brain parcellations for the 7 MSTEPS resolutions. Corresponding percentage discovery maps show the percentage of connections with a significant effect, for each brain parcel. MNI coordinates are given for representative slices superimposed onto the MNI 152 non-linear template.

Please cite this article as: Bellec, P., et al., Impact of the resolution of brain parcels on connectome-wide association studies in fMRI, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.07.071

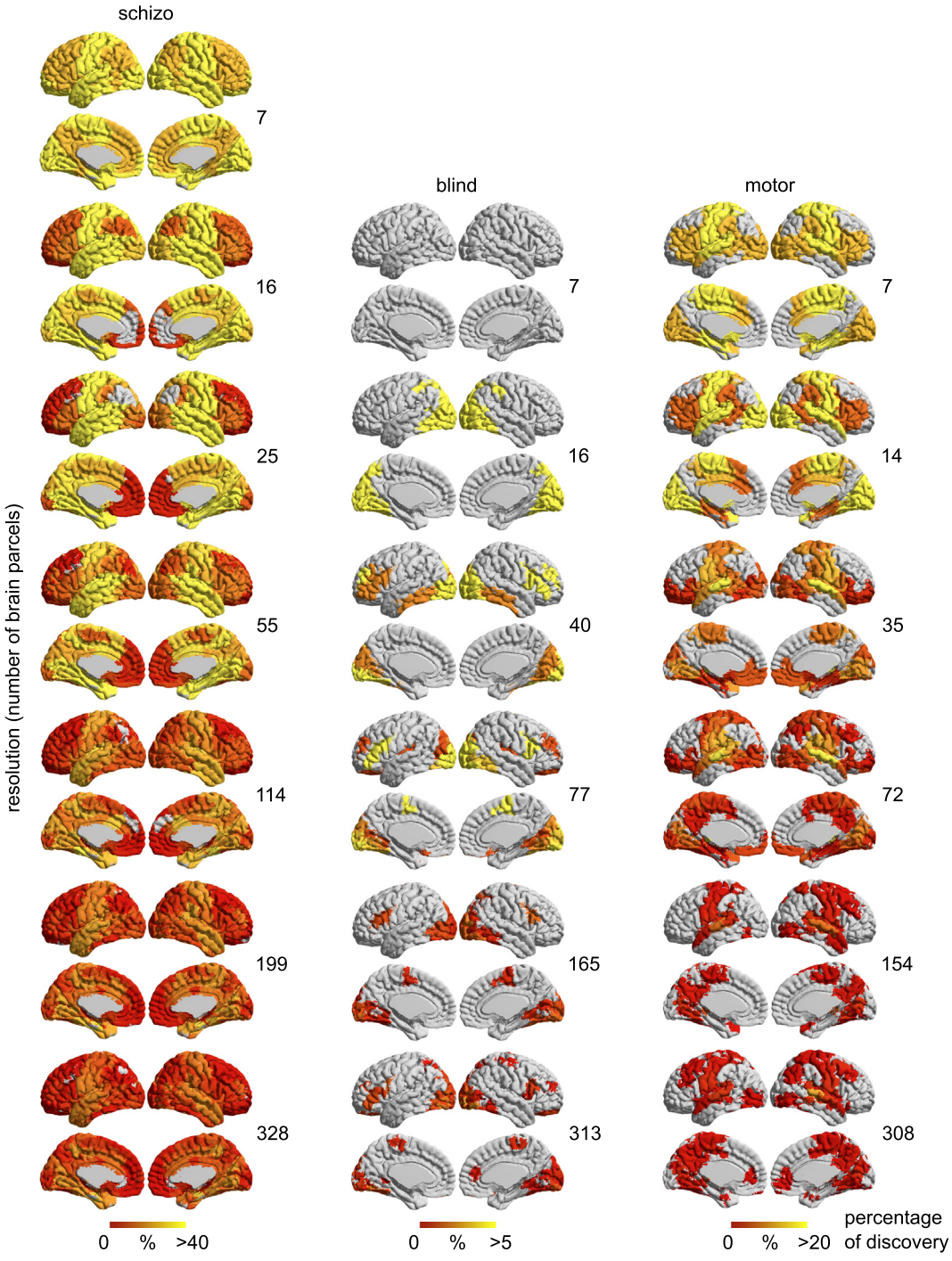

Fig. 9. Percentage of discovery maps in the three real datasets for all MSTEPS resolutions. Surface maps show the percentage of connections with a significant effect, for each brain parcel, in respectively the SCHIZO, BLIND and MOTOR contrasts. Maps are projected onto the MNI 2009 surface. See Fig. 8 for volumetric representations showing results at the subcortical level in the SCHIZO contrast.

the resolution with highest overall discovery rate did not always provide the highest discovery rate for a given brain parcel. For instance, the proportion of connections showing a significant effect in the basal ganglia for the SCHIZO contrast became maximal at resolution 55, once the thalami were isolated as a single parcel (Fig. 8). As another example, the dorsolateral prefrontal cortex only showed a significant effect in the BLIND contrast for functional brain parcellations above resolution 40 (Fig. 9).

Seed-based maps of t-statistics

The maps of discovery rate did not characterize which specific connections were identified as significant for each parcel, nor the direction of the effect (i.e. an increase vs a decrease in connectivity). We illustrated how these questions can be explored using the SCHIZO dataset, as it showed widespread changes in functional connectivity. The percentage of discovery maps were used to select a number of seed parcels of interest, i.e. showing maximal effects (Fig. 10). Parcels selected at the highest

Fig. 10. Group FDR-corrected *t*-test maps in the SCHIZO dataset, in volumetric space. *t*-Test maps showed significant alterations ($q < 0.05$ for FDR-BH) in functional connectivity (decreases and increases) in schizophrenia for the 7 MSTEPS resolutions and several seeds. The seed that included the hippocampus, the anterior cingulate and the thalami were shown as stroke white parcels at all resolutions. Intra-parcel changes in connectivity were thus not shown for seeds (e.g., decreased connectivity within the basal ganglia). The *z* MNI coordinates were given for representative slices superimposed onto the MNI 152 non-linear template.

resolutions corresponded to the hippocampus, anterior cingulate cortex and thalamus. Corresponding parcels for lower resolutions were selected based on their maximal overlap with the parcels chosen at the highest resolutions. For instance, the most distributed parcel encompassing the hippocampus at resolution 7 covered the whole medial temporal lobe, the temporal pole and ventral prefrontal cortex. For each brain parcel, a FDR-corrected *t*-test map associated with the contrast of interest was generated. These *t*-test maps revealed that the alterations in functional coupling in schizophrenia essentially took the form of a decrease in connectivity for the hippocampus and associated regions as well as for the anterior cingulate cortex and its associated parcel. By contrast, the thalamus and basal ganglia showed an increase in functional connectivity with the occipital cortex, beyond decreased connectivity within the basal ganglia.

Impact of resolution on statistical maps

While visual exploration of the *t*-test maps in the SCHIZO dataset revealed similarities of the effects across resolutions, it also highlighted some specificities. High resolutions indeed proved in some cases to be additionally informative compared to low resolutions, despite decreased overall detection rate. For instance, the parcel centered on the hippocampus was seen to be more positively connected with the thalamus and caudate nucleus in schizophrenia only when the ventral prefrontal cortex was not part of the parcel (Fig. 10). As another example, the thalamus showed increased connectivity with a large sensorimotor cortical parcel at resolution 25 and above only, when it was not part of the same parcel as the putamen. Furthermore, the thalamus only showed a significant decrease in connectivity with the dorsolateral prefrontal cortex at resolution 55 and above, when isolated as a single parcel rather than smoothed out inside the basal ganglia.

We more formally tested the level of correspondence of the effects across resolutions for the three seeds listed above in the SCHIZO dataset, as well as for seeds matching our a priori in the BLIND and MOTOR datasets, respectively located in the right primary visual cortex and the left primary motor cortex. Pairwise comparisons between spatial effect maps across resolutions mostly revealed positive correlation values in all three datasets and for all seeds (Fig. 11). Correlations for the three seeds investigated in the SCHIZO contrast were as follows: hippocampus (mean, standard deviation, minimum, maximum = 0.86, 0.06,

Please cite this article as: Bellec, P., et al., Impact of the resolution of brain parcels on connectome-wide association studies in fMRI, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.07.071

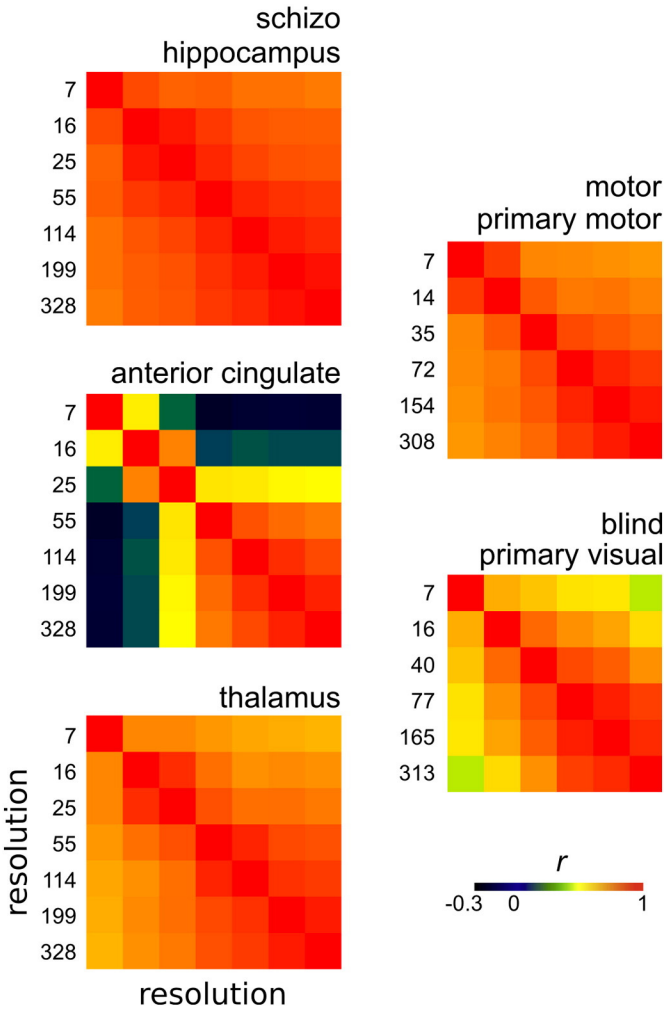

Fig. 11. Correspondence of effects maps across resolutions for the three real datasets. Correlation matrices show pairwise comparisons between 7 and 6 MSTEPS resolutions of the effects maps for three selected seeds in the SCHIZO dataset and one a priori seed in each of the BLIND and MOTOR dataset.

0.76, 0.97), anterior cingulate (0.41, 0.39, −0.20, 0.93), and thalamus (0.78, 0.09, 0.64, 0.94). High correlations were always observed when comparing high resolutions (above resolution 55) between them. Comparisons between low and high resolutions remained associated with high correlations values for two out of the three seeds, namely the hippocampus and thalamus. However, results for the anterior cingulate demonstrated that a low correspondence between low and high resolutions was possible. Results for the seeds in the BLIND (0.71, 0.15, 0.46, 0.93) and MOTOR (0.80, 0.08, 0.70, 0.95) datasets further supported the general conclusions drawn from the SCHIZO dataset.

Discussion

Specificity in multiresolution analysis

This work investigated empirically how the resolution impacts GLM analyses on connectomes, in particular in terms of specificity. We confirmed on realistic simulations the validity of a FDR control using the BH procedure at a single resolution. On three real datasets, there was no sign of substantial departure from the assumptions of a basic parametric GLM. The censoring of time frames with excessive motion will still very likely introduce some departure from the homoscedasticity assumption, albeit small. Future work may investigate the gains of more general GLM procedures able to accommodate heteroscedasticity of the residuals.

We also investigated the specificity of GLM-connectome analyses across resolutions. We notably tested empirically the hypothesis of Efron (2008) that the FDR would be controlled across resolutions in the presence of a strong signal. We did identify two regimes: a "liberal" regime where the FDR across resolutions is inflated compared to the FDR within resolution, and an "exact" regime where the FDR within and across resolutions precisely match. The "liberal" regime corresponded to situations where effects are either weak or present at very few connections.

As a partial remedy, the proposed omnibus test was found to appropriately control the overall FWE across resolutions. However, even when combined with the omnibus test, we still observed simulation settings where the FDR across resolutions departed from nominal levels (e.g. Fig. 5). This may be sufficient for a descriptive analysis that does not critically rely on significance testing, such as the quantification of similarities between statistical maps. In such cases, the omnibus test simply guarantees that the similarities between resolutions are at least in part attributable to true associations. For proper control of the FDR, GLM-connectome analysis should thus be implemented at a single resolution, selected ahead of times, or new methods would need to be developed to correct for multiple comparisons across resolutions.

Sensitivity across resolutions

We found some clear evidence on simulations for increased sensitivity at certain resolutions for the FDR-BH procedure in GLM-connectome analyses. For independent tests, the sensitivity decreased sharply with resolution, to reach a plateau around resolution 50. This behavior appeared to be a consequence of multiple testing in the FDR-BH procedure, as the proportion of true non-null hypothesis and the effect size were maintained strictly constant across resolutions. For dependent tests, the resolution directly impacted the effect size, as some test clusters matched better the underlying simulated signals than others. However, when the true and test clusters matched, the same trends as in the simulations for independent tests were observed. On real data, highest discovery rates were found below 50 parcels. This profile resembled most closely the sensitivity results from simulations with independent tests, and may reflect some intrinsic property of the FDR-BH procedure. The simulation experiments suggest that increased sensitivity of the FDR-BH procedure at low resolutions likely explain this increase in discovery rate. In particular, resolutions larger than 100, routinely used with the AAL template (Tzourio-Mazoyer et al., 2002), was systematically associated with a much smaller discovery rate than lower resolutions (below 50). There thus appears to be a trade-off to be made with the FDR-BH procedure between sensitivity and resolution, similar to what was observed with the NBS procedure (Zalesky et al., 2010a). In other words, our ability to detect an effect (increased at low resolution) seems to be competing with our ability to tell which particular brain connections is showing this effect (increased at high resolution). For example, the BASC procedure tends to merge homologous regions into a single parcel for low resolutions. An explicit testing of connectivity between homologous regions would require using fairly high resolutions (200+).

What is a good resolution of brain parcellation?

In the three real data experiments, we did not identify strong discrepancies between statistical maps generated at different resolutions, consistent with the observations of Shehzad et al. (2014). More specifically, on real data, statistical maps at resolutions above 30 matched closely the maps generated with several hundreds of parcels. Taken together with our evaluation of sensitivity across resolutions, our findings support the use of a single resolution, around 30, that will provide an accurate approximation of effect maps observed at higher resolutions,

Please cite this article as: Bellec, P., et al., Impact of the resolution of brain parcels on connectome-wide association studies in fMRI, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.07.071

while being associated with larger discovery rates and, likely, sensitivity. There may still be structures best observed at different resolutions. For example, the difference in thalamic connectivity in the SCHIZO analysis was better seen at resolution 55 and above, where the thalami were clustered in one parcel rather than aggregated with the putamen and caudate nuclei. The multiresolution may thus prove useful to identify resolutions tailored to specific brain structures or specific experimental conditions to be used in future, independent studies. Note that it would not be advisable to explore multiple resolutions, and then simply report results at the resolution with the highest discovery rate. There would be no guarantee that the FDR would be controlled for a resolution that was selected precisely because of a high associated discovery rate, a classic case of circular analysis.

Biological plausibility of effects on real datasets

The effects found on the real datasets were consistent with the existing literature. First, schizophrenia has been defined as a dysconnectivity syndrome, with aberrant functional interactions between brain regions being a core feature of this mental illness (for reviews, see Calhoun et al., 2009; Pettersson-Yeo et al., 2011; Fornito et al., 2012). As shown here for two out of three parcels, and observed for other unreported brain parcels, widespread decrease in connectivity was observed in patients, with the addition of more localized increases in connectivity. The prominence of decreases in connectivity in the temporal lobe, hippocampus and anterior cingulate cortex, amongst other regions, is well supported by previous studies (Williamson and Allman, 2012). Similarly, increased connectivity between the thalamus and sensorimotor cortex but decreased connectivity with striatal and prefrontal regions has been reported before (Anticevic et al., 2014). Second, resting-state fMRI studies have previously shown that congenital blindness is associated with a reorganization of the interactions between the occipital cortex and other parts of the brain, in particular the auditory and premotor cortices (Liu et al., 2007; Qin et al., 2013, 2015), consistent with our findings. Finally, our results are in agreement with the observation that brain activity at rest is modulated by previous intensive motor practice (Albert et al., 2009; Vahdat et al., 2011; Sami et al., 2014). Even in the absence of a definite ground truth on these real life applications, our findings thus had good face validity, and suggested that GLM-connectome analysis could be successfully applied to a variety of clinical or experimental conditions.

Impact of resolution on alternative statistical methods

We did observe a strong impact of resolution on the statistical power of the FDR-BH procedure, yet other statistical approaches may behave quite differently regarding resolution. Shehzad et al. (2014) developed a multivariate test that applies on a region-to-brain connectivity map, called multivariate distance matrix regression (MDMR). Because the test relies on the similarity between maps across subjects, and because statistical maps are well approximated even with 30 brain parcels, we expect this procedure to be relatively insensitive to the resolution of the parcellation. This procedure would be used to screen for promising seed-based connectivity maps worthy to explore in a subsequent, independent analysis. The MDMR approach effectively performs one test per parcel, instead of one test per connection, and thus greatly alleviates the multiple comparison problem. It does not however provide a control of statistical risk at the level of single connections. Zalesky et al. (2010a) proposed to use uncorrected threshold on the individual p-values, but then to identify to which extent the connections that survive the test are interconnected. This extent measure is compared against what could be observed under a null hypothesis of no association, implemented through permutation testing. This approach, called Network-Based Statistics (NBS), is the connectome equivalent to the "cluster-level statistics" used in SPMs. The NBS only offers a loose control of false-positive rate at the level of a single connection, but can be used to reject the possibility that a group of significant findings could be observed by chance in the FWE sense. The NBS has been found to outperform the FDR-BH when the connections with significant effects are indeed interconnected. The validity of this assumption may be quite dependent on the resolution of the parcellation. An interesting direction for future work is therefore to investigate how CWAS techniques such as FDR-BH, MDMR and NBS compare as a function of the resolution of parcels, while existing comparisons have been limited to a fixed resolution (Zalesky et al., 2012).

Of note is the recent work of Meskaldji et al. (2014), which combines two resolutions to perform connectome-wide testing: the low resolution is used to screen for promising groups of intra- or inter-parcel connections, and the tests at high resolution are re-weighted based on that screening. The weights can be adjusted to ensure control of the FDR across the connectome. This alternative approach to multiresolution testing is limited to two resolutions, but may provide additional statistical power compared to simply replicating the GLM analysis independently at two resolutions independently as was done here.

Beyond resolution selection: choice of the brain parcellation

We only briefly examined here how the choice of parcels, and not just their number, could impact sensitivity. We could, for example, have used random parcellations, like (Zalesky et al., 2010b), a parcellation based on anatomical landmarks such as the AAL atlas (Tzourio-Mazoyer et al., 2002), or a functional parcellation with spatial connexity constraints (Craddock et al., 2012). From our results on simulations, it seems clear that dramatic differences in statistical power can be achieved at a given spatial resolution, if a set of parcels is best adapted to the spatial distribution of an effect. The work of Craddock et al. (2012) suggested that functional brain parcels are more homogeneous than anatomical parcels. We believe that important improvement in sensitivity could be gained from the optimization of the parcellation scheme, rather than resolution, and this represents an important avenue for future research. Following an idea initially explored in Thirion et al. (2006), it may even be possible to relax the constraint of identical parcels across subjects, by matching different individual-specific parcels and use this correspondence to run group-level GLM-connectome analysis.

Conclusion

Our overall conclusion is that the GLM analysis of connectomes with control of the FDR using the BH procedure is statistically valid when used at a single resolution, and has the potential to identify biologically plausible associations in a variety of experimental conditions. Caution should be exercised when replicating a GLM-connectome analysis at different resolutions, as the FDR over the tests combined across all resolutions may depart from the FDR within resolution. We proposed a valid omnibus test, combining findings across all resolutions to establish the overall presence of true effects. This test can be used for exploratory analysis of the impact of resolution on the results of a GLM-connectome analysis. We observed that the statistical maps generated at resolution 30+ were highly similar on three real datasets, and that the rate of discovery decreased sharply after resolution 50. An analysis using a single resolution in the range of 30 to 50 brain parcels thus appears as a reasonable default option, likely to have a sensitivity superior to the common approach using 100+ brain parcels in many settings. A mulitresolution GLM-connectome pipeline is available in the NIAK package[11] (Bellec et al., in press), a free and open-source software that

[11] niak.simexp-lab.org.

[12] http://figshare.com/articles/Group_multiscale_functional_template_generated_with_BASC_on_the_Cambridge_sample/1285615.

Please cite this article as: Bellec, P., et al., Impact of the resolution of brain parcels on connectome-wide association studies in fMRI, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.07.071

runs in matlab and GNU octave, and we also publicly released a set of multiresolution functional brain parcellations [12].

Acknowledgments

Parts of this work were presented at the 2012 and 2013 annual meetings of the organization for human brain mapping, as well as the International Conference on Resting-State Connectivity 2012 (Magdeburg). The authors are grateful to the members of the 1000 functional connectome consortium for publicly releasing the "Cambridge" and "COBRE" data samples, as well as Dr Shehzad for valuable feedback. The computational resources used to perform the data analysis were provided by Compute Canada[13] and CLUMEQ[14], which is funded in part by NSERC (MRS), FQRNT, and McGill University. This project was funded by NSERC grant number RN000028, a salary award from "Fonds de recherche du Québec — Santé" to PB as well as a salary award by the Canadian Institute of Health Research to PO.

Appendix A. Generation of statistical parametric connectomes under the global null hypothesis

Let $\mathbf{Y}^{(K)}$ be the (subjects × connections) matrix of individual connectomes at resolution K. A replication of the connectome matrix under the global null hypothesis $(\mathcal{G}_0)$ is generated by recomposing the linear mixture while excluding the c-th covariate of interest, tested by the model. Formally, let $\mathbf{X}_{\bar{c}}$ be the reduced model where the c^{th} covariate has been removed from the (subjects × covariates) matrix $\mathbf{X}$. Let $\hat{\mathbf{B}}_{\bar{c}}^{(K)}$ be the ordinary least square estimate of the regression coefficients using the reduced model. Each permutation sample of the dataset is generated as described in (Anderson, 2002):

$$\mathbf{Y}^{(K,*)} = \mathbf{X}_{\bar{c}}\hat{\mathbf{B}}_{\bar{c}}^{(K)} + \hat{\mathbf{E}}^{(K,*)}. \tag{A.1}$$

where $\hat{\mathbf{E}}^{(K,*)}$ is a replication of the residuals of the regression of the reduced model, with permuted rows (subjects). The GLM procedure is then implemented with the $\mathbf{Y}^{(K,*)}$ and the full model $\mathbf{X}$ to generate a replication $V_K^{(*)}$ of the volume of discoveries at resolution K under $(\mathcal{G}_0)$.

Because the same dataset at voxel resolution is used to generate all the connectome datasets $(\mathbf{Y}^{(K)})_s$, the samples $V_s^{(*)}$ are not independent. In order to respect these dependencies, for any given replication, the same permutation of the subjects is used to generate to all of the $(\hat{\mathbf{E}}^{(K,*)})_K$. The replication of the total volume of discoveries $V^{(*)}$ is then simply the sum of $V_K^{(*)}$ for all K. This procedure is repeated B times in order to generate B replications $(V^{(*\,b)})_{b=1}^{B}$ of the total volume of discoveries under $(\mathcal{G}_0)$. The Monte-Carlo estimation of the probability to observe a greater total volume of discoveries under $(\mathcal{G}_0)$ than the actual total volume of discoveries V generated on the original (non-permuted) dataset is then:

$$\Pr\left(V^{(*)} \geq V | \mathcal{G}_0\right) \doteq \#\left\{b = 1, \ldots, B | V^{(*b)} \geq V\right\}/B \tag{A.2}$$

where $\doteq$ means that the two terms are asymptotically equal as B tends towards infinity.

Appendix B. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2015.07.071.

[13] https://computecanada.org/.
[14] http://www.clumeq.mcgill.ca/.

References

Abou Elseoud, A., Littow, H., Remes, J., Starck, T., Nikkinen, J., Nissilä, J., Timonen, M., Tervonen, O., Kiviniemi, V., 2011. Group-ICA model order highlights patterns of functional brain connectivity. Front. Syst. Neurosci. 5. http://dx.doi.org/10.3389/fnsys.2011.00037.

Ad-Dab'bagh, Y., Einarson, D., Lyttelton, O., Muehlboeck, J.S., Mok, K., Ivanov, O., Vincent, R.D., Lepage, C., Lerch, J., Fombonne, E., Evans, A.C., 2006. The CIVET image-processing environment: a fully automated comprehensive pipeline for anatomical neuroimaging research. In: Corbetta, M. (Ed.), Proceedings of the 12th Annual Meeting of the Human Brain Mapping OrganizationNeuroimage. Elsevier, Florence, Italy.

Albert, N.B., Robertson, E.M., Miall, R.C., 2009. The resting human brain and motor learning. Curr. Biol. 19 (12), 1023–1027. http://dx.doi.org/10.1016/j.cub.2009.04.028 (Jun).

Albouy, G., Fogel, S., King, B.R., Laventure, S., Benali, H., Karni, A., Carrier, J., Robertson, E.M., Doyon, J., 2015. Maintaining vs. enhancing motor sequence memories: respective roles of striatal and hippocampal systems. NeuroImage 108, 423–434 (URL http://view.ncbi.nlm.nih.gov/pubmed/25542533).

Anderson, C.W., 2002. Quantitative Methods for Current Environmental Issues. Springer (Mar. URL http://www.worldcat.org/isbn/1852332948).

Anticevic, A., Cole, M.W., Repovs, G., Murray, J.D., Brumbaugh, M.S., Winkler, A.M., Savic, A., Krystal, J.H., Pearlson, G.D., Glahn, D.C., 2014. Characterizing thalamo-cortical disturbances in schizophrenia and bipolar illness. Cereb. Cortex 24 (12), 3116–3130. http://dx.doi.org/10.1093/cercor/bht165 (Dec).

Barkhof, F., Haller, S., Rombouts, S.A., 2014. Resting-state functional MR imaging: a new window to the brain. Radiology 272 (1), 29–49 (Jul., URL http://view.ncbi.nlm.nih.gov/pubmed/24956047).

Bellec, P., 2013. Mining the hierarchy of resting-state brain networks: selection of representative clusters in a multiscale structure. Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop on, pp. 54–57 (Jun).

Bellec, P., Perlbarg, V., Jbabdi, S., Pélégrini-Issac, M., Anton, J., Doyon, J., Benali, H., 2006. Identification of large-scale networks in the brain using fMRI. NeuroImage 29 (4), 1231–1243. http://dx.doi.org/10.1016/j.neuroimage.2005.08.044 (Feb).

Bellec, P., Rosa-Neto, P., Lyttelton, O.C., Benali, H., Evans, A.C., 2010. Multi-level bootstrap analysis of stable clusters in resting-state fMRI. NeuroImage 51 (3), 1126–1139. http://dx.doi.org/10.1016/j.neuroimage.2010.02.082 (Jul).

Bellec, P., Lavoie-Courchesne, S., Dickinson, P., Lerch, J.P., Zijdenbos, A.P., Evans, A.C., 2012. The pipeline system for Octave and Matlab (PSOM): a lightweight scripting framework and execution engine for scientific workflows. Front. Neuroinforma. 6. http://dx.doi.org/10.3389/fninf.2012.00007.

Bellec, P., Carbonell, F.M., Perlbarg, V., Lepage, C., Lyttelton, O., Fonov, V., Janke, A., Tohka, J., Evans, A.C., 2011. A neuroimaging analysis kit for Matlab and Octave. Proceedings of the 17th International Conference on Functional Mapping of the Human Brain (in press).

Benjamini, Y., Hochberg, Y., 1995. Controlling the false-discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300.

Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29 (4), 1165–1188. http://dx.doi.org/10.2307/2674075.

Biswal, B.B., Mennes, M., Zuo, X.-N.N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., Dogonowski, A.-M.M., Ernst, M., Fair, D., Hampson, M., Hoptman, M.J., Hyde, J.S., Kiviniemi, V.J., Kötter, R., Li, S.-J.J., Lin, C.-P.P., Lowe, M.J., Mackay, C., Madden, D.J., Madsen, K.H., Margulies, D.S., Mayberg, H.S., McMahon, K., Monk, C.S., Mostofsky, S.H., Nagel, B.J., Pekar, J.J., Peltier, S.J., Petersen, S.E., Riedl, V., Rombouts, S.A., Rypma, B., Schlaggar, B.L., Schmidt, S., Seidler, R.D., Siegle, G.J., Sorg, C., Teng, G.-J.J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X.-C.C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y.-F.F., Zhang, H.-Y.Y., Castellanos, F.X., Milham, M.P., 2010. Toward discovery science of human brain function. Proc. Natl. Acad. Sci. U. S. A. 107 (10), 4734–4739. http://dx.doi.org/10.1073/pnas.0911855107 (Mar).

Blumensath, T., Jbabdi, S., Glasser, M.F., Van Essen, D.C., Ugurbil, K., Behrens, T.E.J., Smith, S.M., 2013. Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. NeuroImage 76, 313–324. http://dx.doi.org/10.1016/j.neuroimage.2013.03.024 Aug.

Calhoun, V.D., Eichele, T., Pearlson, G., 2009. Functional brain networks in schizophrenia: a review. Front. Hum. Neurosci. 3. http://dx.doi.org/10.3389/neuro.09.017.2009.

Castellanos, F.X., Di Martino, A., Craddock, R.C., Mehta, A.D., Milham, M.P., 2013. Clinical applications of the functional connectome. NeuroImage 80, 527–540. http://dx.doi.org/10.1016/j.neuroimage.2013.04.083 (Oct).

Collignon, O., Vandewalle, G., Voss, P., Albouy, G., Charbonneau, G., Lassonde, M., Lepore, F., 2011. Functional specialization for auditory-spatial processing in the occipital cortex of congenitally blind humans. Proc. Natl. Acad. Sci. U. S. A. 108 (11), 4435–4440. http://dx.doi.org/10.1073/pnas.1013928108 (Mar).

Collins, D.L., Evans, A.C., 1997. Animal: validation and applications of nonlinear registration-based segmentation. Int. J. Pattern Recognit. Artif. Intell. 11, 1271–1294.

Craddock, R.C., James, G.A., Holtzheimer, P.E., Hu, X.P., Mayberg, H.S., 2012. A whole brain fMRI atlas generated via spatially constrained spectral clustering. Hum. Brain Mapp. 33 (8), 1914–1928. http://dx.doi.org/10.1002/hbm.21333 (Aug).

Efron, B., 2008. Simultaneous inference: when should hypothesis testing problems be combined? Ann. Appl. Stat. 2 (1), 197–223. http://dx.doi.org/10.1214/07-aoas141 (Mar).

Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., Collins, D.L., 2011. Unbiased average age-appropriate atlases for pediatric studies. NeuroImage 54 (1), 313–327. http://dx.doi.org/10.1016/j.neuroimage.2010.07.033 (Jan).

Fornito, A., Zalesky, A., Pantelis, C., Bullmore, E.T., 2012. Schizophrenia, neuroimaging and connectomics. NeuroImage 62 (4), 2296–2314 (Oct., URL http://view.ncbi.nlm.nih.gov/pubmed/22387165).

Fox, M.D., Greicius, M., 2010. Clinical applications of resting state functional connectivity. Front. Syst. Neurosci. 4. http://dx.doi.org/10.3389/fnsys.2010.00019.
Giove, F., Gili, T., Iacovella, V., Macaluso, E., Maraviglia, B., 2009. Images-based suppression of unwanted global signals in resting-state functional connectivity studies. Magn. Reson. Imaging 27 (8), 1058–1064. http://dx.doi.org/10.1016/j.mri.2009.06.004 (Oct).
Gordon, E.M., Laumann, T.O., Adeyemo, B., Huckins, J.F., Kelley, W.M., Petersen, S.E., 2014. Generation and evaluation of a cortical area parcellation from resting-state correlations. Cereb. Cortex http://dx.doi.org/10.1093/cercor/bhu239 (Oct., bhu239+).
Jafri, M.J., Pearlson, G.D., Stevens, M., Calhoun, V.D., 2008. A method for functional network connectivity among spatially independent resting-state components in schizophrenia. NeuroImage 39 (4), 1666–1681. http://dx.doi.org/10.1016/j.neuroimage.2007.11.001 (Feb).
Liu, Y., Yu, C., Liang, M., Li, J., Tian, L., Zhou, Y., Qin, W., Li, K., Jiang, T., 2007. Whole brain functional connectivity in the early blind. Brain 130 (8), 2085–2096. http://dx.doi.org/10.1093/brain/awm121 (Aug).
Liu, H., Stufflebeam, S.M., Sepulcre, J., Hedden, T., Buckner, R.L., 2009. Evidence from intrinsic activity that asymmetry of the human brain is controlled by multiple factors. Proc. Natl. Acad. Sci. 106 (48), 20499–20503. http://dx.doi.org/10.1073/pnas.0908073106 (Dec).
Marrelec, G., Krainik, A., Duffau, H., Pélégrini-Issac, M., Lehéricy, S., Doyon, J., Benali, H., 2006. Partial correlation for functional brain interactivity investigation in functional MRI. NeuroImage 32 (1), 228–237. http://dx.doi.org/10.1016/j.neuroimage.2005.12.057.
Marrelec, G., Bellec, P., Krainik, A., Duffau, H., Pélégrini-Issac, M., Lehéricy, S., Benali, H., Doyon, J., 2008. Regions, systems, and the brain: hierarchical measures of functional integration in fMRI. Med. Image Anal. http://dx.doi.org/10.1016/j.media.2008.02.002 (Feb).
Meskaldji, D.E., Fischi-Gomez, E., Griffa, A., Hagmann, P., Morgenthaler, S., Thiran, J.-P., 2013. Comparing connectomes across subjects and populations at different scales. NeuroImage 80, 416–425. http://dx.doi.org/10.1016/j.neuroimage.2013.04.084 (Oct).
Meskaldji, D.-E., Vasung, L., Romascano, D., Thiran, J.-P., Hagmann, P., Morgenthaler, S., Van De Ville, D., 2014. Improved statistical evaluation of group differences in connectomes by screeningfiltering strategy with application to study maturation of brain connections between childhood and adolescence. NeuroImage http://dx.doi.org/10.1016/j.neuroimage.2014.11.059 (Dec).
Pettersson-Yeo, W., Allen, P., Benetti, S., McGuire, P., Mechelli, A., 2011. Dysconnectivity in schizophrenia: where are we now? Neurosci. Biobehav. Rev. 35 (5), 1110–1124 (Apr., URL http://view.ncbi.nlm.nih.gov/pubmed/21115039).
Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage 59 (3), 2142–2154. http://dx.doi.org/10.1016/j.neuroimage.2011.10.018 (Feb).
Qin, W., Liu, Y., Jiang, T., Yu, C., 2013. The development of visual areas depends differently on visual experience. PLoS One 8 (1) (URL http://view.ncbi.nlm.nih.gov/pubmed/23308283).
Qin, W., Xuan, Y., Liu, Y., Jiang, T., Yu, C., 2015. Functional connectivity density in congenitally and late blind subjects. Cereb. Cortex 25 (9), 2507–2516 (URL http://view.ncbi.nlm.nih.gov/pubmed/24642421).
Royston, P., 1993. A toolkit for testing for non-normality in complete and censored samples. J. R. Stat. Soc. Ser. D 42 (1), 37–43.
Sami, S., Robertson, E.M., Miall, R.C., 2014. The time course of task-specific memory consolidation effects in resting state networks. J. Neurosci. 34 (11), 3982–3992. http://dx.doi.org/10.1523/jneurosci.4341-13.2014 (Mar).
Shehzad, Z., Kelly, C., Reiss, P.T., Cameron Craddock, R., Emerson, J.W., McMahon, K., Copland, D.A., Xavier Castellanos, F., Milham, M.P., 2014. A multivariate distance-based analytic framework for connectome-wide association studies. NeuroImage 93, 74–94. http://dx.doi.org/10.1016/j.neuroimage.2014.02.024 (Jun).
Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., Woolrich, M.W., 2011. Network modelling methods for FMRI. NeuroImage 54 (2), 875–891. http://dx.doi.org/10.1016/j.neuroimage.2010.08.063 (Jan).
Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J.-B.B., 2006. Dealing with the shortcomings of spatial normalization: multi-subject parcellation of fMRI datasets. Hum. Brain Mapp. 27 (8), 678–693. http://dx.doi.org/10.1002/hbm.20210 (Aug).
Thirion, B., Varoquaux, G., Dohmatob, E., Poline, J.-B.B., 2014. Which fMRI clustering gives good brain parcellations? Front. Neurosci. 8 (URL http://view.ncbi.nlm.nih.gov/pubmed/25071425).
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15 (1), 273–289. http://dx.doi.org/10.1006/nimg.2001.0978 (Jan).
Vahdat, S., Darainy, M., Milner, T.E., Ostry, D.J., 2011. Functionally specific changes in resting-state sensorimotor networks after motor learning. J. Neurosci. 31 (47), 16907–16915 (Nov., URL http://view.ncbi.nlm.nih.gov/pubmed/22114261).
Wang, K., Liang, M., Wang, L., Tian, L., Zhang, X., Li, K., Jiang, T., 2007. Altered functional connectivity in early Alzheimer's disease: a resting-state fMRI study. Hum. Brain Mapp. 28 (10), 967–978. http://dx.doi.org/10.1002/hbm.20324 (Oct).
White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 817–838.
Williamson, P.C., Allman, J.M., 2012. A framework for interpreting functional networks in schizophrenia. Front. Hum. Neurosci. 6. http://dx.doi.org/10.3389/fnhum.2012.00184.
Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited—again. NeuroImage 2 (3), 173–181. http://dx.doi.org/10.1006/nimg.1995.1023 (Sep).
Worsley, K.J., Cao, J., Paus, T., Petrides, M., Evans, A.C., 1998. Applications of random field theory to functional connectivity. Hum. Brain Mapp. 6 (5-6), 364–367 (URL http://view.ncbi.nlm.nih.gov/pubmed/9788073).
Zalesky, A., Fornito, A., Bullmore, E.T., 2010a. Network-based statistic: identifying differences in brain networks. NeuroImage 53 (4), 1197–1207. http://dx.doi.org/10.1016/j.neuroimage.2010.06.041 (Dec).
Zalesky, A., Fornito, A., Harding, I.H., Cocchi, L., Yücel, M., Pantelis, C., Bullmore, E.T., 2010b. Whole-brain anatomical networks: does the choice of nodes matter? NeuroImage 50 (3), 970–983. http://dx.doi.org/10.1016/j.neuroimage.2009.12.027 (Apr).
Zalesky, A., Cocchi, L., Fornito, A., Murray, M.M., Bullmore, E., 2012. Connectivity differences in brain networks. NeuroImage 60 (2), 1055–1062. http://dx.doi.org/10.1016/j.neuroimage.2012.01.068 (Apr).

Please cite this article as: Bellec, P., et al., Impact of the resolution of brain parcels on connectome-wide association studies in fMRI, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.07.071