

RESEARCH ARTICLE

RichMind: A Tool for Improved Inference from Large-Scale Neuroimaging Results

Adi Maron-Katz^{1,2*}, David Amar³, Eti Ben Simon^{1,2}, Talma Hendler^{1,2,4,5}, Ron Shamir³

1 Functional Brain Center, Wohl Institute for Advanced Imaging Tel-Aviv Sourasky Medical Center, Tel Aviv, Israel, **2** Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel, **3** Blavatnik School of Computer Science, Tel-Aviv University, Tel Aviv 69978, Israel, **4** School of Psychological Sciences, Tel Aviv University, Tel-Aviv 69978, Israel, **5** Sagol school of Neuroscience, Tel Aviv University, Tel Aviv, Israel

* adimaron@gmail.com



OPEN ACCESS

Citation: Maron-Katz A, Amar D, Simon EB, Hendler T, Shamir R (2016) RichMind: A Tool for Improved Inference from Large-Scale Neuroimaging Results. PLoS ONE 11(7): e0159643. doi:10.1371/journal.pone.0159643

Editor: Satoru Hayasaka, University of Texas at Austin, UNITED STATES

Received: November 13, 2015

Accepted: July 6, 2016

Published: July 25, 2016

Copyright: © 2016 Maron-Katz et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data underlying case study 2 is present as Supporting Information. The data underlying case study 1 is owned by the study's authors (Nummenmaa Lauri et al), who will provide the data upon request to interested readers (contact information: Dr. Nummenmaa Lauri email: lauri.nummenmaa@aalto.fi).

Funding: This work was supported by the Ministry of Science, Technology and Space, Israel, to Adi Maron-Katz & Talma Hendler; Raymond and Beverly Sackler Chair in Bioinformatics, Tel-Aviv University to Ron Shamir; Israel Science Foundation, 317/13, to Ron Shamir; Azrieli Foundation to David Amar; and

Abstract

As the use of large-scale data-driven analysis becomes increasingly common, the need for robust methods for interpreting a large number of results increases. To date, neuroimaging attempts to interpret large-scale activity or connectivity results often turn to existing neural mapping based on previous literature. In case of a large number of results, manual selection or percent of overlap with existing maps is frequently used to facilitate interpretation, often without a clear statistical justification. Such methodology holds the risk of reporting false positive results and overlooking additional results. Here, we propose using *enrichment analysis* for improving the interpretation of large-scale neuroimaging results. We focus on two possible cases: *position group analysis*, where the identified results are a set of neural positions; and *connection group analysis*, where the identified results are a set of neural position-pairs (i.e. neural connections). We explore different models for detecting significant overrepresentation of known functional brain annotations using simulated and real data. We implemented our methods in a tool called RichMind, which provides both statistical significance reports and brain visualization. We demonstrate the abilities of RichMind by revisiting two previous fMRI studies. In both studies RichMind automatically highlighted most of the findings that were reported in the original studies as well as several additional findings that were overlooked. Hence, RichMind is a valuable new tool for rigorous inference from neuroimaging results.

Introduction

To date, functional interpretation of large-scale neuroimaging findings is often done by associating the identified regions to known *classes* (e.g., anatomic structures or functional networks). This process of using previous knowledge to ascribe functional meaning to findings is commonly based on a subjective visual inspection or on percent of overlap with existing maps [1–5]. Such methodology, which is not based on statistical justification, holds the risk of reporting false positive results and overlooking additional results. For example, Nummenmaa et al. (2012) analyzed fMRI signals recorded from 16 healthy participants, while viewing film clips depicting unpleasant, neutral, and pleasant emotions. They identified cerebral regions where

Edmond J. Safra Center for Bioinformatics at Tel Aviv University to David Amar.

Competing Interests: The authors have declared that no competing interests exist.

inter-subject correlations (ISC) were significantly correlated with subjective reports of valence and arousal provided by the participants. In order to interpret the findings, the authors subjectively associated the identified regions to known functional networks. They reported that arousal was mostly associated with ISC in regions of the sensori-motor network (SMN), visual network (VN) and dorsal attention network (DAN) while valence was negatively associated with ISC in regions of the default mode network (DMN) as well as regions known to be involved in emotional processing [2]. However, as in many neuroimaging studies, no quantitative statistical measure was presented to support this association of findings to functional networks. An even more complex case is the case where the identified findings are a collection of neural position pairs (i.e. connections). When this collection is very large, as may occur in data-driven studies [6–8], interpretation becomes challenging. In some cases, this challenge is faced by filtering the results either by manual selection or by repeating the analysis using a stricter statistical threshold. For example Wang et. al. (2013) reported a set of 363 functional connections that differed between a group of amnesic mild cognitive impairment (aMCI) patients and healthy controls. These connections were identified using the network-based statistic approach [9] with a predefined 1024 functional parcellation [1, 10]. In order to avoid the complexity of interpreting such a large set of connections, the network-based statistic analysis was repeated using a stricter statistical threshold, producing a set of 87 connections that was subsequently interpreted.

An alternative approach to interpreting such a large set of findings is to test whether the results contain significantly more elements with a specific class than expected by chance. For instance, one can examine whether an identified set of weakened connections in terms of DMN-SMN connectivity, and explore whether aMCI is associated with a significantly large number of weakened connections linking the DMN with the SMN. If the answer is positive, we say that the corresponding class (i.e. DMN-SMN) is *enriched* in the identified collection. Such enrichment (or over-representation) can be assigned with a statistical significance value under an appropriate null hypothesis [11].

In this study we propose using *enrichment analysis* to facilitate and improve the interpretation process of large-scale results obtained from fMRI studies. Notably, the term *enrichment analysis* in this case refers to the analysis of significance of class overrepresentation within sets (i.e. overlap) [11–13], and not to the analysis of significance of class distribution across a ranked set [14]. We address two cases: In the first case the identified results are a set of neural positions. For example, these positions could be a set of voxels that demonstrate increased activation under a specific condition. In this case, we call the enrichment analysis *position group analysis*. In the second case, the identified results are a set of neural position pairs (i.e. neural connections). For example, they can be pairs of neural positions demonstrating increased functional connectivity under a specific experimental condition [6–8]. In this case, we call the enrichment analysis *connection group analysis*. In both cases, in addition to the study results we are given an annotation of the brain that maps neural positions to classes representing known neural functions (e.g. [15–17]), or anatomic structures (e.g. [18–20]). We examined different models for detecting significant overrepresentation of known functional brain annotation using simulated and real data.

We implemented our methods in RichMind, an open-source Matlab-based computational tool that provides both statistical significance reports based on our suggested enrichment analysis methods, as well as brain visualizations. We demonstrate the abilities of RichMind by reanalyzing two previous fMRI studies: the first of subjects viewing emotion-inducing film clips [2], and the second of subjects suffering from aMCI [1]. We show that by using enrichment analysis, we were able to provide statistical validation to most of the conclusions drawn in the original studies, while revealing additional statistically significant results. In addition we show how enrichment analysis allows interpreting a large set of results without having to apply additional filters, as often applied in studies, thus, allowing more accurate interpretation of the results.

Methods

Neural annotation

A *neural annotation* is a mapping of neural positions to known classes. The annotation is based on previous knowledge, and can contain anatomic structure (E.g. Anatomic atlas labeling [21]), known functional mapping (E.g. functional networks identified in previous studies [15, 22]), previously known pathology association, etc.

In the current study we used two sets of annotations that are based on functional neural mapping. The first was used in [2] and it consists of 6 functional networks, and the second annotation was used in [1] and it consists of 5 functional brain networks. In our simulation, we used a made-up dummy annotation.

The Hypergeometric test

In this study, we use the hypergeometric (HG) test to calculate the significance of the overlap of two sets. The test is demonstrated in Fig 1. We are given a set S of M items, which we call the *background set*, and two subsets $A \subseteq S$ and $B \subseteq S$, of sizes N and K , respectively. (For example, S can be the set of brain positions considered, A is the set of positions identified in the experiment, and B is a particular subset of the positions annotated with a known structure or function, e.g. the cerebellum.) Let x be the size of the intersection between A and B . Is that intersection size meaningful or random? Here, our null hypothesis is that the N items in A were sampled randomly and independently from S without replacement. Therefore, the significance of the observed value x is the probability of having x or more elements in the intersection, which can be calculated using the HG cumulative distribution function, given in Eq 1. Notably, the Hypergeometric test is equivalent to a one-tailed Fisher's exact test. Equation variables are listed in Fig 1.

$$p = F(x|M, K, N) = \sum_{i=x}^{\min(N,K)} \frac{\binom{K}{i} \binom{M-K}{N-i}}{\binom{M}{N}} \quad (1)$$

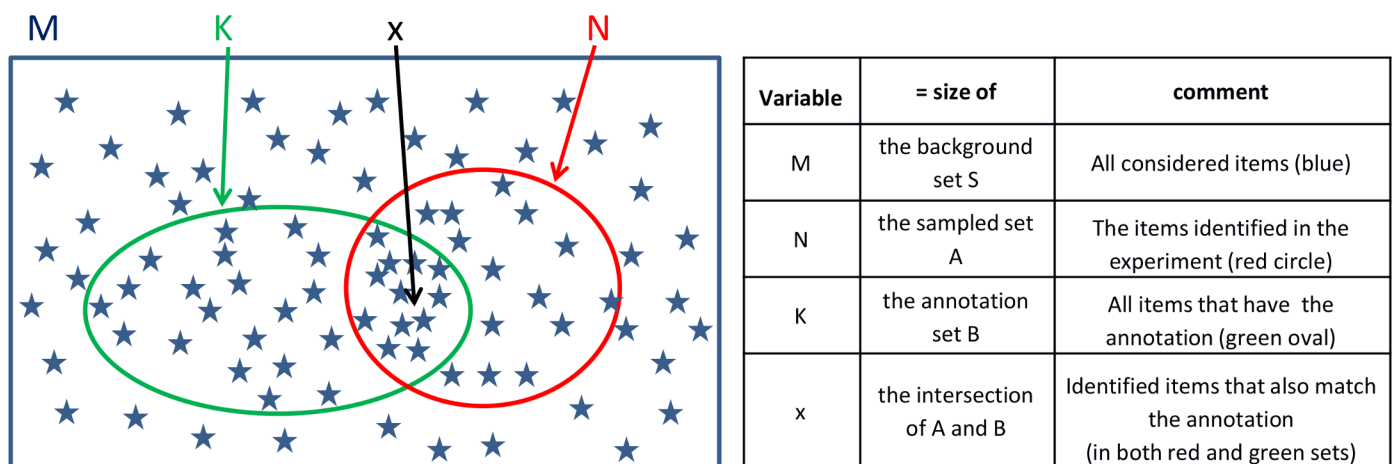


Fig 1. A demonstration of HG test setting: given a background set S , which contains M elements (stars), a specific annotation set B with K elements, and a randomly sampled set A with N elements, HG test estimates the probability of having x or more elements of B in A .

doi:10.1371/journal.pone.0159643.g001

Position group enrichment analysis using the HG test

Here, the study identified a group of neural positions (e.g., voxels or parcels) A^P out of a background set S . We wish to consider the significance of the overlap of A^P with another subset of neural positions B from S , which constitutes a known annotation class (e.g., an anatomically defined brain region or a set of regions that share a specific function or form a particular network). Our goal is to decide whether the number x of positions that are both in group A^P and class B is larger than expected by chance.

Connections group enrichment analysis

Here the study identified a group of connections (i.e. pairs of neural positions) $A^c = \{(x_1, y_1), \dots, (x_n, y_n)\}$, where each x_i and each y_i represents a neural position (e.g. voxel or parcel). This set can also be viewed as the set of edges in a graph $G(S, A^c)$, where S is the set of neural positions. In addition we are given two disjoint subsets of S , C and D , each of which constitutes a known annotation class (e.g. C is the precuneus and D is the basal ganglia). Our goal is to decide whether the number of connections between C and D that are also in A^c , denoted as $a(A^c, C, D)$, is larger than expected by chance. In this work, we test two approaches for this task: (1) a parametric approach that uses the HG test; and (2) a non-parametric test based on permutations. While the first test is relatively simple and easy to compute, it ignores the degree distribution in the graph represented by the neural connections. This is addressed by the second test, which uses permutations to create a large set of random graphs with the same node degrees. These two tests have been validated on simulated data (See [S1](#), [S2](#) and [S3](#) Figs), and the difference between them is illustrated in [S1](#) and [S4](#) Figs. In the next two sections we use the same notation as above.

Parametric connection group analysis using the HG test

Here, we use the HG test with the following parameters. N is the number of pairs in A^c . M (i.e., the background set size) is the number of all possible neural pairs: $M = |S|(|S|-1)/2$. $K = |C|*|D|$ is the number of possible pairs between C and D . Finally, x is the observed number of pairs between C and D — $a(E, C, D)$. The underlying null hypothesis of this test is that the graph G was randomly selected from the set of all graphs with the same number of edges N over the same set of nodes.

Non-parametric degree-preserving analysis

The HG approach for connection groups does not account for the degree distribution in the graph G . The importance of this distribution has been previously observed in brain networks [[23](#), [24](#)]. We therefore propose an additional non-parametric test. Here, our null hypothesis is that the graph G was randomly selected from the set Σ of all graphs with the same node degrees. In other words, Σ is the set of graphs with node set S such that each node has the same number of neighbors as in G .

We calculate the p-value empirically by drawing graphs from S using a heuristic rewiring step: remove two disjoint edges in the current graph and replace them by two others so that node degrees remain unchanged. A long chain of such steps leads to a near-random sampling from S [[25](#)]. The method has been successfully used in multiple bioinformatics applications [[26](#), [27](#)]. Given a set of graphs generated using this process we calculate for each one the number of observed edges between C and D . This step produces a vector of scores $a = a_1, \dots, a_m$ (by default we generate $m = 1000$ randomized graphs). The final empirical p-value is the fraction of scores in a greater than or equal to $a(E, C, D)$. We call this process degree preserving

permutation (DPP), as the edges in each random graph generated form a permutation of the original edges that also preserves the degree constraints.

The DPP and the HG tests use different null models and hence can differ in their sensitivity depending on the data. The HG test assumes that all possible graphs with N edges are equally likely to be selected, and the DPP test assumes that only graphs with N edges and with the same number of neighbors for each node are equally likely to be selected. As a result, the former is more sensitive when the number of identified connections between C and D is relatively high compared to M , while the latter is more sensitive when the identified connections between C and D is relatively high compared to the number of identified connections involving C or D . See [S1 Text](#) and [S1](#) and [S4 Figs](#) for an example and simulations that demonstrate the difference.

Multiple testing correction

Since enrichment is tested for each combination of an identified collection and a class, the output contains multiple p-values. Therefore, we correct for multiple testing at a false discovery rate (FDR) of 0.05 using the procedure suggested by Benjamini and Hochberg (BH) [28]. The FDR approach controls for the rate of false discovery, and is considered to be more powerful than methods that control for the family-wise error rate such as Bonferroni approach [28–30]. The FDR-adjusted p-value of each enrichment is called its q-value [31].

Results

We implemented the two statistical tests for enrichment in results of fMRI studies in a Matlab-based tool called RichMind. Below, we first give a brief explanation on the input and output of the tool. Next, we show two case studies in which we apply RichMind to real fMRI data.

RichMind—a toolbox for analysis of enrichment of fMRI results

RichMind receives as input a collection of items identified in a neuroimaging study. These can be either (1) one or more sets of neural positions for position group analysis, or (2) a set of neural position-pairs for connection group analysis. Each set of positions in (1) can be input using (a) a 3D NIFTI file, containing at each position a numerical group identifier (the number of the set to which that position belongs), (b) a tabular text file, in which each line contains a numerical position identifier followed by a numerical group identifier, or (c) a Matlab data file, in which each line contains a numerical position identifier followed by a numerical group identifier. Sets of position-pairs in (2) can be input using similar formats as (b) or (c) with three columns, where the first two columns contain position identifiers and the third column contains the group identifier. In addition to the groups file, the collection of all positions considered in the experiment is required. This collection is used to define the background set, and should preferably include all positions used in the initial analysis that yielded the interest groups. The file can be a 3D NIFTI file, a tabular text file or a Matlab file. In both cases each line should contain the MNI coordinates of a single position, and the line number is used as the position identifier in (1) and (2).

RichMind uses an established neural annotation, attributing neural positions to meaningful terms. These classes reflect prior knowledge of brain function or anatomic structure, so they can be anatomic labels, functions, pathology association, etc. By default, RichMind uses the functional neural mapping provided in [15] as the annotation. Alternatively, it provides an option to use the anatomic mapping provided in [32], or any other mapping provided by the user. In each type of analysis RichMind calculates the p-values for over-representation of the classes (see [Methods](#) for details). When using the DPP test, if k is the number of graphs

sampled, then the minimum empirical p-value that can be obtained is $1/k$. Thus, a larger k increases the maximum significance achievable, but also the test runtime. For example, in case study 2, running DPP with 1000 graphs gave a maximum significance of 10^{-3} and took 212 seconds on a quad core Pentium i7 processor, with 8Gb RAM. Since many groups are tested, the use of nominal p-values is inadequate and correction for multiple testing is needed. Hence, all p-values are corrected for multiple testing using the false discovery rate (FDR) method [28] and q-values are reported. Alternatively, the user can choose the more stringent Bonferroni correction.

Finally, RichMind reports a list with all significant enrichments (0.05 FDR by default), and also produces bar plots that display the p-value and an additional measure of enrichment level called the “frequency ratio” result (see Figs 2A and 3A for examples). The frequency ratio is the ratio between fraction of class members in the tested set and in the background set. This measure evaluates how frequent is the class representation in the set vs. the background. High ratios indicate over-representation, but do not have direct statistical significance as p-values, and should be used cautiously when comparing results that involve classes of very different sizes. On the other hand, when class or sets are too small to provide significant p-values, high ratios can call attention to particular results. Ratios should be used cautiously when comparing results that involve classes of very different sizes. For each reported result, brain 2D and 3D views overlaying the neural positions (or connections) are available by clicking on the result (see Figs 2B and 3B for examples). In addition, one can export these overlay graphs into files that can be loaded to the BrainNet viewer [33] (see Figs 2B and 3B for examples).

Case Study 1: Inter-subject correlation identified while watching emotion inducing film clips

Here we revisited findings reported by Nummenmaa et al. (2012), who identified ISC maps that were correlated with self-reported valence and arousal scores. ISCs were based on fMRI data recorded from 16 healthy participants, while viewing film clips depicting unpleasant, neutral, and pleasant emotions. ISCs were derived by calculating, for each voxel, the Pearson correlation coefficient of the BOLD time series recorded in each pair of subjects. This was done both for the entire time frame and for sliding windows of 17 time points. Ongoing measures of self-reported valence and arousal provided by participants were used as regressors in a general linear model (GLM), to identify significantly associated ISCs. Results were interpreted by the authors in the context of six functional networks extracted using seed-based functional connectivity analysis on the same data—the VN, SMN, AN, DMN, DAN and the executive control network (ECN). The authors reported that arousal was mostly associated with ISC in the SMN, VN and DAN while valence was negatively associated with ISC in the DMN as well as in regions involved in emotional processing, such as midbrain, thalamus, ventral striatum, insula, and anterior cingulate cortex [2]. No quantitative statistical measure was presented to support this interpretation.

We ran RichMind position group analysis on two sets of cerebral regions: one where ISC was inversely associated with self-reported valence, and another where ISC was positively associated with self-reported arousal. All gray matter voxels were used as background for enrichment test. The mapping of voxels to the six functional networks was taken from the original paper. The results are presented in Table 1 and Fig 2. RichMind identified arousal associated ISC to be enriched with regions involved in AN ($q = 1.28E-10$), SMN ($q = 1.85E-10$), DAN ($q = 9.28E-09$), and VN ($q = 9.26E-11$), and valence associated ISCs to be enriched with DMN ($q = 9.26E-11$), SMN ($<1.4E-37$) and ECN ($q = 6.59E-09$). These results recapitulate the results of the original paper. However, they add additional findings of AN enrichment within arousal

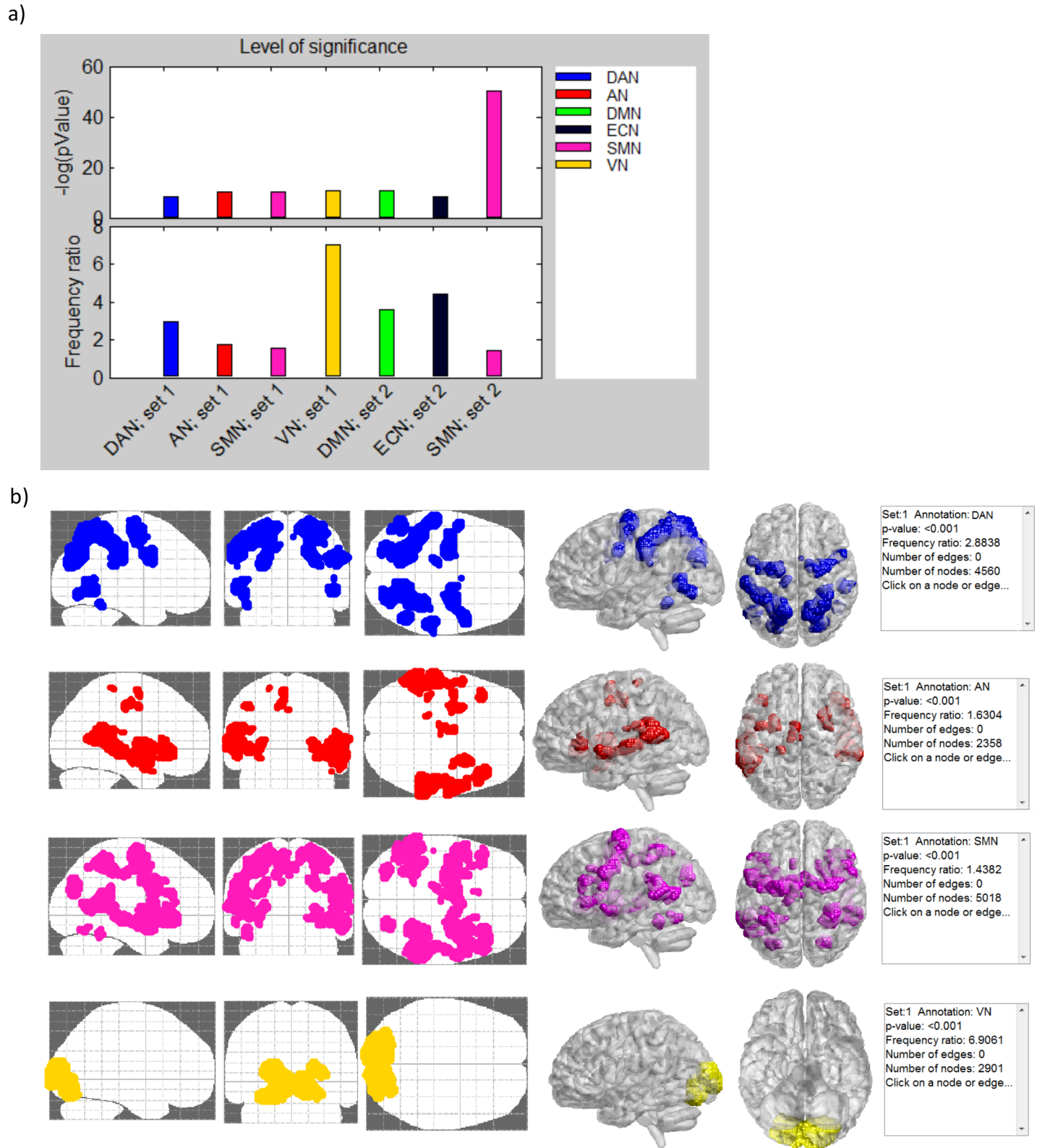
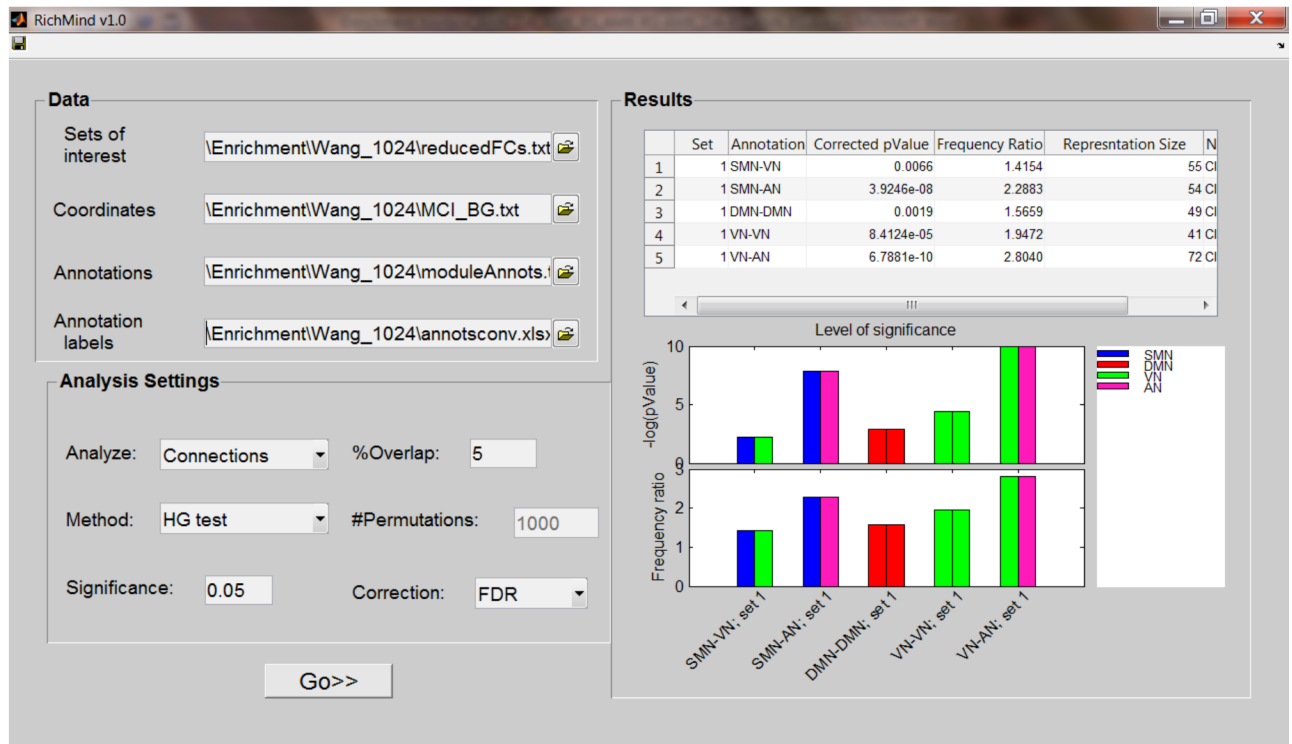


Fig 2. RichMind results visualization for case study 1: (a) Bar plots displaying the p-values and frequency ratios of enrichment analysis results. Each bar is colored according to the enriched class. (b) 2D and high-resolution 3D brain visualization, which shows, for each enriched class, all neural positions that are both in the SOI and in the class. Positions are colored according to the corresponding classes. High resolution 3D images were generated using BrainNet viewer [33].

doi:10.1371/journal.pone.0159643.g002

a)



b)

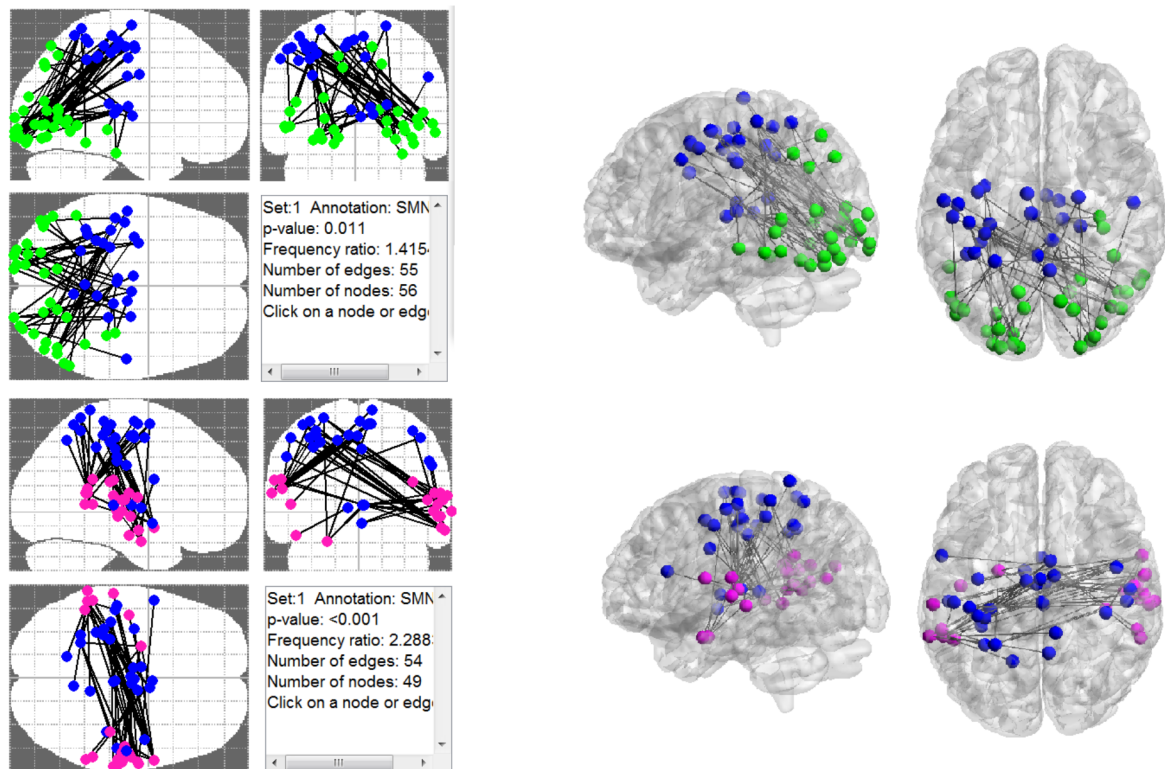


Fig 3. RichMind results visualization for case study 2: (a) Bar plots displaying the p-values and frequency ratios of enrichment analysis results. Each bar is composed of two rectangles colored by the two classes that constitute the enriched class. (b) 2D and high-resolution 3D brain

visualization, showing, for each enriched class, all neural connections that are both in CC363 and in the class. Parcels are colored according to the corresponding classes. High resolution 3D images were generated using BrainNet viewer [33].

doi:10.1371/journal.pone.0159643.g003

associated ISCs, and ECN and SMN enrichment within valence associated ISCs. These findings reinforce the claim made in the original study, by which high arousal serves to direct individuals' attention to features of the environment. Identifying ECN and SMN enrichment within valence associated ISCs, is in line with the authors' suggestion by which negative valence synchronizes brain circuitries, supporting emotional sensations across individuals.

Case Study 2: fMRI functional connectivity differences identified in cases of amnesic mild cognitive impairment

Wang et al. (2013) analyzed resting state fMRI data recorded from 37 subjects with aMCI, and 47 healthy controls. The analysis produced functional connections that differed between the groups. These connections were identified using the network-based statistic approach [9] on a predefined functional parcellation containing 1024 parcels [1, 10]. The approach identified connected components (CCs) that are composed of FCs for which the inter-group difference exceeded a pre-defined threshold. Component significance was estimated using a permutation test. This analysis detected a single CC of 363 reduced FCs when using a p-value threshold of 5×10^{-4} . We call this set CC363. In addition, two CCs of 65 and 22 reduced FCs were discovered using a p-value threshold of 10^{-4} , denoted as CC65 and CC22, respectively.

In the original study, due to the large number of connections in CC363, only CC65 and CC22 were further interpreted. This was done in the context of a modular architecture derived from the control group, which includes five modules corresponding to the VN, the SMN, the DMN, the AN and the ventral attention network (VAN). CC65 was reported as comprised mainly of inter-module connections (46/65, 70.8%), which linked regions in the SMN module, the VN module, and the AN module. CC22 was reported to contain predominantly intra-module connections (15/22, 68.2%) within the DMN module [1].

We used RichMind to analyze CC363. All 1024 parcels were used to generate the background for the enrichment analyses. A mapping of nodes to functional modules was taken from the original paper. The results are presented in Table 2 and Fig 3. The HG-based analysis identified CC363 as enriched with inter-modular FCs that link regions of the SMN module, the VN module, and the AN module ($q_{(SMN-VN)} = 0.011$, $q_{(SMN-AN)} = 6.54E-08$; $q_{(AN-VN)} = 1.2E-09$), and with intra-modular connections within the DMN module ($q = 0.003$). These results reproduce the main conclusions of the original study, but were obtained on the larger CC, which was not discussed in the original study due to its size. In addition, the test revealed enrichment in connections within the VN module ($q = 0.00014$), which was not reported in

Table 1. RichMind results for case study 1. DAN = dorsal attention network, AN = auditory network, SMN = sensori-motor network, VN = visual network, DMN = default-mode network, ECN = executive control network.

Set	Enriched class	HG-based q-values	Frequency Ratio	# Voxels
Arousal ISCs	DAN	9.28E-09	2.9	4560
Arousal ISCs	AN	1.28E-10	1.6	2358
Arousal ISCs	SMN	1.85E-10	1.4	5018
Arousal ISCs	VN	9.26E-11	6.9	2901
Valence ICSS	DMN	9.26E-11	3.5	1357
Valence ICSS	ECN	6.59E-09	4.3	2684
Valence ICSS	SMN	<1.4E-37	1.3	2056

doi:10.1371/journal.pone.0159643.t001

Table 2. RichMind results for case study 2; Class abbreviations: VN = visual network, AN = auditory network, SMN = sensori-motor network, DMN = default-mode network.

Enriched inter-class connection	HG q-value	DPP q-value	Frequency ratio	# Connections
VN-AN	6.8E-10	0.0325	2.8	72
SMN-AN	3.9E-08	0.00075	2.33	54
VN-VN	8.4E-05	0.79	1.9	41
DMN-DMN	0.0019	<0.00075	1.66	49
SMN-VN	0.0066	0.79	1.4	55

doi:10.1371/journal.pone.0159643.t002

the original study. The degree preserving permutation test did not identify FC enrichment within the VN nor did it identify FC enrichment between the SMN and the VN. However, it recovered the other three inter-module links (see [Table 2](#)). Notably, the differences between HG and DPP in this case are primarily for connections that involve the VN. This is probably due to the relatively low degree of VN nodes in the CC363 component graph. While HG is not sensitive to the degree of nodes in each class, DPP does consider this information, and favors higher density links between classes (See also [S4 Fig](#) and [S1 Text](#)).

Repeating case analyses with an external annotation

The above analyses were conducted using the same annotations that were used by the authors of the original papers for interpretation. In both cases, these were functional brain networks that were extracted from the same experiment. However, to validate the results of a new experiment, it is preferable that enrichment analysis is conducted using an independent annotation. Another advantage of using such an external annotation is that it allows the results to be comparable across studies. Accordingly, we repeated both case analyses using an annotation reported in [\[19\]](#), which includes a partition of the cortex into seven functional brain networks. This annotation was selected because it is based on a thorough analysis of a very large cohort of 1000 subjects. When we repeated the analysis of RichMind using this annotation two of the 5 results in our previous analysis of case study 2 were identified (VN-SMN and VN-VN connectivity). In case study 1, the results remained similar to those obtained in our previous analysis, see supporting [S1 Table](#) and [S2 Table](#). However, slight differences were identified. For example, valence associated ISC was enriched with SMN using the original annotation but not the external one. This difference results from discrepancies in the mapping of voxels to networks.

Discussion

In this work we describe RichMind, a Matlab-based, easy-to-use computational tool that tests for enrichment of known classes in large-scale neural results. It provides both statistical reports and brain visualizations of the identified enrichments. Statistical reports are based on the probability of getting the observed representation of each annotation in the tested set by chance.

We applied RichMind to two case studies, and in both of them RichMind reinforced the main claims made in the original papers, while adding new findings. In case study 1, the involvement of the ACC in the valence-associated ISCs seems to contribute most to the identified ECN enrichment within that group ([Fig 2B](#), shown in black), in accordance with the statement in the original study. However, regions of the SMN were reported in the original study only in association with arousal ISCs and not with valence ISCs. Using RichMind we reveal SMN enrichment within valence associated ISCs, a finding with extremely low q-value, indicating that it is highly significant. Notably, this enrichment was not identified using the external annotation that was based on [\[15\]](#), due to differences between the mappings. This

inconsistency demonstrates the need for an established functional mapping of the brain that is acknowledged in the field as “common ground”.

While enrichment analysis is standard in genomic and genetic studies [11, 12, 34], few previous fMRI studies addressed the issue of large-scale interpretation by calculating the relative frequency of specific classes. For example, in case study 2, Wang et al. [1] used maps of known functional brain networks extracted from the set of healthy controls through modularity analysis, and then calculated the percent of the results that link each pair of networks [1]. However, such approach does not take into consideration the spatial coverage of each class, which has a major effect on the frequency of its representation in the results. Furthermore, it does not provide statistical significance of the reported findings.

Unlike the simple case where results contain sets of neural positions, when examining sets of neural connections, the null hypothesis of random independent sampling, which underlies the hyper-geometric test, may not be suitable. This is due to a non-uniform distribution of the degrees in the brain network [23, 24]. Instead, empirical p-values can be calculated using a permutation test, in which the random background model preserves the degrees of the nodes in the graph. Such degree-preserving permutation test has been previously used for analyzing enrichment within protein-protein interaction networks [26, 27]. In our tests, when comparing HG to the degree preserving permutation test, we observed that the latter was often much more stringent and produced less results.

Shortcomings and future plans

Using a data-driven approach, which considers all possible classes, while correcting for multiple tests, is very strict, and thus may increase the rate of false negative findings. In addition, the analysis is conducted under the assumption of specific null models, which, in some cases, may not hold. Other null models can be added to RichMind in the future based on user requests. Another possible future direction for improved large-scale analysis is ranking-based tests, which proved successful in many genomic applications [14].

The use of enrichment analysis is always based on some previously established mapping that is used as an annotation. The selection of a specific annotation system may have a crucial effect on the results of the analysis. For this reason, it would be ideal to use an established functional mapping of the brain that is accepted in the field as “common ground”. Such annotation systems exist in other fields for this type of analysis, e.g. the Gene Ontology initiative [35] or the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [36], which are used as standard gene annotations in computational genomics analysis. However, due to the lack of such a common ground in neuroscience, we adopted a functional annotation that was based on a previously published study, conducted on the 1000 connectomes data, and an anatomic annotation of lobe-laterality information that was based on the TD atlas. We believe that established mapping systems will be available in the near future, and will encourage and improve the use of enrichment analysis in the field. Alternatively, an annotation can be defined based on an independent analysis in order to allow comparing findings across studies or datasets.

Lastly, although RichMind was designed to address the issue of rigorous inference from large-scale findings of a single study, the use of enrichment analysis may potentially be extended into a framework for cross-study meta-analytic inference, as has been demonstrated for genetic data [33]. Such an extension is a possible direction for future work in the field.

Availability and requirements

RichMind runs on Matlab (version R2011a or later). The software, simulation data and sample data for case study 2 are freely available for download at <http://acgt.cs.tau.ac.il/RichMind>. The

dataset of case study 1 is available from the authors (lauri.nummenmaa@aalto.fi). A technical user manual is available at <http://acgt.cs.tau.ac.il/RichMind/help.html>.

Supporting Information

S1 Dataset. Dataset for case study 2.

(TAR)

S1 Fig. An illustration of the difference between the two significance scores in connection group analysis. Each of the graphs contains 20 nodes, of which three are labeled “green” and three are labeled “red”. The number of connections between green nodes and red nodes is 7 in both cases. However, the number of connections and consequently the degree of the nodes varies greatly between the two cases. Graph (a) was found to be enriched with red-green connections using HG-test (FDR $q = 4.4 \times 10^{-5}$) but not using DPP (FDR $q = 0.22$). On the other hand, graph (b) was found to be enriched with red-green connections using DPP (FDR $q < 0.001$) but not using HG test (FDR $q = 0.34$). (TIF)

S2 Fig. Degree distributions on real and simulated networks. The histograms show the degree distribution of the network CC363 analyzed in the study (a), and of instances of scale-free networks of same size randomly generated using power law with $\text{exp} = 2$ (b) and $\text{exp} = 3$ (c). (TIF)

S3 Fig. Simulation results. Graphs with 244 nodes and 363 edges were randomly generated using a power law wiring scheme with $\text{exp} = 2$ (in a,c) and $\text{exp} = 3$ (in b,d), and then complete 5x5 subgraphs were implanted and their edges were randomly removed according to noise level q . The $-\log$ of mean p-value over 1000 runs is plotted against level of introduced noise (q parameter). Error bars correspond to one show standard deviation. P-values were estimated by RichMind for the “real” implanted 5X5 bicluster A-B (blue) and for “dummy” 5X5 bicluster A'-B' (red), using HG test (in a,b) and DPP with 1000 randomized graphs (in c,d). (TIF)

S4 Fig. HG vs. DPP test. Graphs with 244 nodes and 1000 edges were randomly generated using a power law scheme with $\text{exp} = 2$, and then complete bipartite 5x5 subgraphs were planted in them and some of their edges were randomly and independently removed according to noise level q . The $-\log$ of mean p-value over 1000 runs is plotted against level of introduced noise (q parameter). Error bars correspond to one show standard deviation. P-values were estimated by RichMind for the “real” implanted 5X5 bicluster A-B (blue) and for “dummy” 5X5 bicluster A'-B' (red), with A' and B' selected to contain high degree nodes. HG test results are shown in a. DPP test results are shown in b. Each DPP test was run with 1000 randomized graphs. (TIF)

S1 Table. RichMind results for case study 1 using 7 functional brain network reported in [15] as annotations; Class abbreviations: VN = visual network, SMN = sensori-motor network, DMN = default-mode network, FPCN = fronto-parietal control network (PDF)

S2 Table. RichMind results for case study 2 using 7 functional brain network reported in [15] as annotations; Class abbreviations: VN = visual network, SMN = sensori-motor network, DAN = dorsal attention network, VAN = ventral attention network, DMN = default-mode network (PDF)

S1 Text. Method validation supplementary section.
(DOCX)

Acknowledgments

This research was supported in part by the Israel Science Foundation (grant 317/13), the Raymond and Beverly Sackler Chair in Bioinformatics (to RS) and by the Ministry of Science, Technology and Space, Israel (to AMK and TH). DA was supported in part by fellowships from the Azrieli Foundation and the Edmond J. Safra Center for Bioinformatics at Tel Aviv University. This work was done in part while RS and DA were visiting the Simons Institute for the Theory of Computing.

Author Contributions

Conceived and designed the experiments: AMK DA RS. Analyzed the data: AMK DA. Wrote the paper: AMK DA RS EBS TH. Wrote the software code: AMK.

References

1. Wang J, Zuo X, Dai Z, Xia M, Zhao Z, Zhao X, et al. Disrupted functional brain connectome in individuals at risk for Alzheimer's disease. *Biological psychiatry*. 2013; 73(5):472–81. doi: [10.1016/j.biopsych.2012.03.026](https://doi.org/10.1016/j.biopsych.2012.03.026) PMID: [22537793](https://pubmed.ncbi.nlm.nih.gov/22537793/)
2. Nummenmaa L, Glerean E, Viinikainen M, Jääskeläinen IP, Hari R, Sams M. Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences*. 2012; 109(24):9599–604.
3. Lahnakoski JM, Glerean E, Jääskeläinen IP, Hyönä J, Hari R, Sams M, et al. Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage*. 2014; 100:316–24. doi: [10.1016/j.neuroimage.2014.06.022](https://doi.org/10.1016/j.neuroimage.2014.06.022) PMID: [24936687](https://pubmed.ncbi.nlm.nih.gov/24936687/)
4. Ames DL, Honey CJ, Chow MA, Todorov A, Hasson U. Contextual Alignment of Cognitive and Neural Dynamics. *Journal of cognitive neuroscience*. 2015.
5. Jola C, McAleer P, Grosbras M-H, Love SA, Morison G, Pollick FE. Uni- and multisensory brain areas are synchronised across spectators when watching unedited dance recordings. *i-Perception*. 2013; 4(4):265. doi: [10.1068/i0536](https://doi.org/10.1068/i0536) PMID: [24349687](https://pubmed.ncbi.nlm.nih.gov/24349687/)
6. Finn ES, Shen X, Holahan JM, Scheinost D, Lacadie C, Papademetris X, et al. Disruption of functional networks in dyslexia: a whole-brain, data-driven analysis of connectivity. *Biological psychiatry*. 2014; 76(5):397–404. doi: [10.1016/j.biopsych.2013.08.031](https://doi.org/10.1016/j.biopsych.2013.08.031) PMID: [24124929](https://pubmed.ncbi.nlm.nih.gov/24124929/)
7. Tyszka JM, Kennedy DP, Paul LK, Adolphs R. Largely typical patterns of resting-state functional connectivity in high-functioning adults with autism. *Cerebral Cortex*. 2014; 24(7):1894–905. doi: [10.1093/cercor/bht040](https://doi.org/10.1093/cercor/bht040) PMID: [23425893](https://pubmed.ncbi.nlm.nih.gov/23425893/)
8. Sripada C, Kessler D, Fang Y, Welsh RC, Prem Kumar K, Angstadt M. Disrupted network architecture of the resting brain in attention-deficit/hyperactivity disorder. *Human brain mapping*. 2014; 35(9):4693–705. doi: [10.1002/hbm.22504](https://doi.org/10.1002/hbm.22504) PMID: [24668728](https://pubmed.ncbi.nlm.nih.gov/24668728/)
9. Zalesky A, Fornito A, Bullmore ET. Network-based statistic: identifying differences in brain networks. *Neuroimage*. 2010; 53(4):1197–207. Epub 2010/07/06. doi: [10.1016/j.neuroimage.2010.06.041](https://doi.org/10.1016/j.neuroimage.2010.06.041) PMID: [20600983](https://pubmed.ncbi.nlm.nih.gov/20600983/).
10. Craddock RC, James GA, Holtzheimer PE, Hu XP, Mayberg HS. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*. 2012; 33(8):1914–28. doi: [10.1002/hbm.21333](https://doi.org/10.1002/hbm.21333) PMID: [21769991](https://pubmed.ncbi.nlm.nih.gov/21769991/)
11. Goeman JJ, Bühlmann P. Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics*. 2007; 23(8):980–7. PMID: [17303618](https://pubmed.ncbi.nlm.nih.gov/17303618/)
12. Ulitsky I, Maron-Katz A, Shavit S, Sagir D, Linhart C, Elkon R, et al. Expander: from expression microarrays to networks and functions. *nature protocols*. 2010; 5(2):303–22. doi: [10.1038/nprot.2009.230](https://doi.org/10.1038/nprot.2009.230) PMID: [20134430](https://pubmed.ncbi.nlm.nih.gov/20134430/)
13. Liu Y, Koyutürk M, Barnholtz-Sloan JS, Chance MR. Gene interaction enrichment and network analysis to identify dysregulated pathways and their interactions in complex diseases. *BMC systems biology*. 2012; 6(1):1.

14. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*. 2005; 102(43):15545–50. doi: [10.1073/pnas.0506580102](https://doi.org/10.1073/pnas.0506580102)
15. Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*. 2011; 106(3):1125–65. doi: [10.1152/jn.00338.2011](https://doi.org/10.1152/jn.00338.2011) PMID: [21653723](https://pubmed.ncbi.nlm.nih.gov/21653723/)
16. Shirer W, Ryali S, Rykhlevskaia E, Menon V, Greicius M. Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral Cortex*. 2012; 22(1):158–65. doi: [10.1093/cercor/bhr099](https://doi.org/10.1093/cercor/bhr099) PMID: [21616982](https://pubmed.ncbi.nlm.nih.gov/21616982/)
17. Damoiseaux J, Rombouts S, Barkhof F, Scheltens P, Stam C, Smith SM, et al. Consistent resting-state networks across healthy subjects. *Proceedings of the national academy of sciences*. 2006; 103(37):13848–53.
18. Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*. 2003; 19(3):1233–9. PMID: [12880848](https://pubmed.ncbi.nlm.nih.gov/12880848/)
19. Talairach J, Tournoux P. Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging. 1988.
20. Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, et al. Automated Talairach atlas labels for functional brain mapping. *Human brain mapping*. 2000; 10(3):120–31. PMID: [10912591](https://pubmed.ncbi.nlm.nih.gov/10912591/)
21. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*. 2002; 15(1):273–89. PMID: [11771995](https://pubmed.ncbi.nlm.nih.gov/11771995/)
22. Greicius MD, Krasnow B, Reiss AL, Menon V. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*. 2003; 100(1):253–8.
23. Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*. 2010; 52(3):1059–69. doi: [10.1016/j.neuroimage.2009.10.003](https://doi.org/10.1016/j.neuroimage.2009.10.003) PMID: [19819337](https://pubmed.ncbi.nlm.nih.gov/19819337/)
24. Sporns O. The non-random brain: efficiency, economy, and complex dynamics. *Frontiers in computational neuroscience*. 2011; 5.
25. Milo R, Kashtan N, Itzkovitz S, Newman ME, Alon U. Uniform generation of random graphs with arbitrary degree sequences. *arXiv preprint cond-mat/0312028*. 2003; 106:1–4.
26. Pradines JR, Farutin V, Rowley S, Dancik V. Analyzing protein lists with large networks: edge-count probabilities in random graphs with given expected degrees. *Journal of Computational Biology*. 2005; 12(2):113–28. PMID: [15767772](https://pubmed.ncbi.nlm.nih.gov/15767772/)
27. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research*. 2013; 41(D1):D808–D15.
28. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*. 1995:289–300.
29. Dudoit S, Van Der Laan MJ. Multiple testing procedures with applications to genomics: Springer Science & Business Media; 2007.
30. Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*. 2003; 19(3):368–75. PMID: [12584122](https://pubmed.ncbi.nlm.nih.gov/12584122/)
31. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*. 2003; 100(16):9440–5. doi: [10.1073/pnas.1530509100](https://doi.org/10.1073/pnas.1530509100)
32. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. 2002; 33(3):341–55. PMID: [11832223](https://pubmed.ncbi.nlm.nih.gov/11832223/)
33. Xia M, Wang J, He Y. BrainNet Viewer: a network visualization tool for human brain connectomics. *PLoS One*. 2013; 8(7):e68910. doi: [10.1371/journal.pone.0068910](https://doi.org/10.1371/journal.pone.0068910) PMID: [23861951](https://pubmed.ncbi.nlm.nih.gov/23861951/)
34. Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*. 2009; 37(1):1–13. doi: [10.1093/nar/gkn923](https://doi.org/10.1093/nar/gkn923) PMID: [19033363](https://pubmed.ncbi.nlm.nih.gov/19033363/)
35. Consortium GO. The Gene Ontology (GO) database and informatics resource. *Nucleic acids research*. 2004; 32(suppl 1):D258–D61.
36. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*. 2000; 28(1):27–30. PMID: [10592173](https://pubmed.ncbi.nlm.nih.gov/10592173/)