PHILOSOPHICAL TRANSACTIONS B

Research

Gene transcription profiles associated with inter-modular hubs and connection distance in human functional magnetic resonance imaging networks

Petra E. Vértes[1], Timothy Rittman[2], Kirstie J. Whitaker[1], Rafael Romero-Garcia[1], František Váša[1], Manfred G. Kitzbichler[1], Konrad Wagstyl[1], Peter Fonagy[4], Raymond J. Dolan[5,6], Peter B. Jones[1,7], Ian M. Goodyer[1,7], the NSPN Consortium[†] and Edward T. Bullmore[1,3,7,8]

[1]Department of Psychiatry, [2]Department of Clinical Neurosciences, and [3]MRC/Wellcome Trust Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge CB2 0SZ, UK
[4]Research Department of Clinical, Educational and Health Psychology, University College London, London WC1E 6BT, UK
[5]Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology, University College London, London WC1N 3BG, UK
[6]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, London WC1B 5EH, UK
[7]Cambridgeshire and Peterborough NHS Foundation Trust, Huntingdon PE29 3RJ, UK
[8]Immuno-Psychiatry, Immuno-Inflammation Therapeutic Area Unit, GlaxoSmithKline R&D, Stevenage SG1 2NY, UK

MGK, 0000-0002-4494-0753; ETB, 0000-0002-8955-8283

Human functional magnetic resonance imaging (fMRI) brain networks have a complex topology comprising integrative components, e.g. long-distance inter-modular edges, that are theoretically associated with higher biological cost. Here, we estimated intra-modular degree, inter-modular degree and connection distance for each of 285 cortical nodes in multi-echo fMRI data from 38 healthy adults. We used the multivariate technique of partial least squares (PLS) to reduce the dimensionality of the relationships between these three nodal network parameters and prior microarray data on regional expression of 20 737 genes. The first PLS component defined a transcriptional profile associated with high intra-modular degree and short connection distance, whereas the second PLS component was associated with high inter-modular degree and long connection distance. Nodes in superior and lateral cortex with high inter-modular degree and long connection distance had local transcriptional profiles enriched for oxidative metabolism and mitochondria, and for genes specific to supragranular layers of human cortex. In contrast, primary and secondary sensory cortical nodes in posterior cortex with high intra-modular degree and short connection distance had transcriptional profiles enriched for RNA translation and nuclear components. We conclude that, as predicted, topologically integrative hubs, mediating long-distance connections between modules, are more costly in terms of mitochondrial glucose metabolism.

This article is part of the themed issue 'Interpreting BOLD: a dialogue between cognitive and cellular neuroscience'.

THE ROYAL SOCIETY PUBLISHING

1. Introduction

In many ways, functional magnetic resonance imaging (fMRI) is misaligned to the elementary scale of basic neuroscience. The neuron is the anatomical unit of nervous systems and the action potential is the physiological unit of communication between neurons. Functional MRI is incapable of

resolving cellular structures and processes on these microscopic scales of space and time in humans. A single fMRI voxel (around 1 cubic millimetre) represents approximately 1–2 million neurons and other cells. The sampling rate of fMRI (approx. $1/60 = 0.02$ Hz) is too slow to resolve rapid transients like action potentials, or the full spectrum of electrophysiological oscillations (0.1–1000 Hz). An fMRI time series will typically only resolve very low-frequency oscillations (approx. 0.1 Hz) and these 'resting-state' dynamics will represent local blood oxygenation changes coupled to neuronal activity rather than a direct measure of neuronal electrophysiology.

The mechanistic links between the blood oxygen-level dependent (BOLD) contrast measured by fMRI and underlying neuronal states of excitation and inhibition have been extensively investigated [1]. There is, for example, evidence that BOLD oscillations represent a hyperaemic response that is coupled to modulations in the amplitude envelope of higher frequency oscillations in local field potentials [2]. However, there remains a need for clearer explanatory connections between the molecular and cellular scale of neurovascular mechanisms of BOLD contrast in animals or other experimental models, and the whole-brain systems scale of cognitive and clinical neuroscientists using BOLD contrast for fMRI. Deeper biomechanistic interpretation of fMRI results from human experiments will inevitably be challenged by operational and ethical limits on what else can be measured in humans in an effort to control or explain the neurovascular factors of the BOLD signal. There are also a number of significant technical drawbacks to consider. BOLD contrast is not measured in SI units. A 'raw' human fMRI dataset is typically more noise than signal, and it can be challenging to control common and potentially severe sources of noise, such as transient micro-movements of the head (approx. 1 mm) during scanning [3,4].

To set against this list of arguments against fMRI, there are also two important advantages to consider. First and foremost, fMRI is remarkably safe and accessible for human participants including patients. Second, despite the caveats about its physiological origins, fMRI has turned out to be a highly reliable and plausible marker of local or regional brain (de-)activation by diverse cognitive tasks [5], as well as a robust signal of abnormal brain function in patients with clinical disorders [6].

In this context, it has been exciting to see new opportunities recently emerging to link fMRI phenomena to the genomic substrates of human brain organization. A pivotal role has been played by the Allen Institute for Brain Science (AIBS), which has measured expression of all approximately 20 000 genes in the human genome at each of approximately 500 locations in six post-mortem adult human brains, and publically released these data [7]. On this basis, it has been shown that human brain regions can be differentiated in terms of their transcriptional profiles of gene expression, e.g. cortex, cerebellum and thalamus have markedly different transcripts from each other. Conversely, brain regions can be aggregated with each other to constitute modules that demonstrate high co-expression of genes [7]. Gene expression profiles can also be mapped to the same anatomical space as human imaging data, enabling the first direct explorations of how the molecular mechanisms of whole-genome transcription might be related to fMRI dynamics and connectivity in humans [7,8]. For example, genes were more strongly co-expressed by functionally connected brain regions [9] and, more specifically, different classes of functional networks were distinguished by differing co-expression patterns in a set of 19 genes known to be enriched in human supra-granular cortex [10]. We were motivated by these and other results to explore genomic associations with functional MRI network parameters.

Human fMRI research is arguably negotiating a shift from an interventional paradigm, focused on local signal changes estimated in response to experimentally controlled changes of cognitive state, to a more naturalistic paradigm focused on estimating very low-frequency oscillations and their correlation or functional connectivity between pairs of brain regions or voxels [11]. Functional MRI connectivity has been analysed in many different ways, ranging from simple correlational analysis, through multivariate methods such as independent component analysis (ICA), to graph theoretical analysis of the topological properties of brain functional networks [12,13]. This work has discovered that there are spatially extensive systems or networks of correlated BOLD oscillation in the human brain. These functional networks representing anatomical patterns of fMRI time-series covariance are reliable [14], heritable [15], electrophysiologically explicable in terms of amplitude envelope coupling of underlying neuronal oscillations [2], related to normal cognitive functions [16] and implicated in the pathophysiology of many clinical disorders [17].

The complex topology of human fMRI networks—comprising small-worldness, hubs and modules among other non-random and non-regular features—has been accounted for theoretically by an economical model of competition between selection pressures for both low biological cost and high topological integration [18,19]. Low cost is certainly advantageous given inevitable constraints on the intra-cranial space, biological material and metabolic resources, available to build and sustain a human brain network [20]. Topological integration is thought to be advantageous because integrated networks or global workspaces are behaviourally valuable by conferring computational capacity for adaptive, 'higher order' cognitive functions [18,19,21]. However, network integration is typically expensive [22]. An integrated brain network with short topological paths between all possible pairs of spatially distributed nodes will cost more in terms of the length of axonal 'wiring' needed to connect network nodes over long anatomical distances. The wiring cost of brain networks is nearly minimized by the anatomical co-location of densely inter-connected clusters and modules of cortical areas that typically share a specialized information-processing function [23,24]. Modularity is biologically cost-saving: but if the network is not to decompose into isolated modules, and therefore to lose capacity for global integration, there must be some connections between nodes in different modules [23]. And these inter-modular edges will generally be longer distance (higher wiring cost) than the shorter distance intra-modular edges between anatomically concentrated nodes in the same module [24,25].

On this basis, we predicted that human fMRI network nodes with high inter-modular degree, mediating many long-distance connections between modules, would be associated with a distinctive gene expression profile compared to the transcriptional profile of nodes with high intra-modular degree, mediating many short-distance connections within the same module. To test this, we measured inter-modular

degree, intra-modular degree and connection distance, for each of 285 regional nodes, by graph theoretical analysis of resting-state, multi-echo echoplanar imaging data on 38 healthy young adults. We anatomically matched the fMRI network parameters at each node to detailed human brain maps of whole-genome expression provided by the AIBS (see the electronic supplementary material). Then, we used the multivariate technique of partial least squares (PLS) to identify combinations of approximately 20 000 genes whose regional expression profiles best predicted the fMRI network parameters [26]. Finally, we used a suite of recently developed gene enrichment algorithms to interpret the biological functions of genes relatively over- or under-expressed in association with specific network features [27–29].

2. Material and methods

(a) Sample, functional magnetic resonance imaging data and pre-processing

In total, 2500 healthy young people in the age range 14–24 years were recruited in north London and Cambridgeshire and provided details by postal questionnaire on socio-demographics and mental health. This primary cohort was stratified into five contiguous age-related strata: 14–15 years inclusive, 16–17 years, 18–19 years, 20–21 years and 22–24 years. Recruitment within each stratum was evenly balanced for sex and ethnicity. A demographically balanced cohort ($n = 300$) was sub-sampled from the primary cohort for structural and functional MRI assessments and more detailed cognitive testing. Here, we used fMRI data from 40 participants sampled from the top two age strata of the secondary cohort (20–24 years), with 10 men and 10 women in each of the two strata. Participants were excluded if they were currently being treated for a psychiatric disorder or for drug or alcohol dependence; had a current or past history of neurological disorders including epilepsy or head injury causing loss of consciousness; had a learning disability requiring specialist educational support and/or medical treatment; or had a safety contraindication prohibiting MRI.

MRI scanning was conducted at the following three sites: (i) the Wellcome Trust Centre for Neuroimaging, London, (ii) the Wolfson Brain Imaging Centre, Cambridge, and (iii) the Medical Research Council Cognition and Brain Sciences Unit, Cambridge. All sites were identically operating 3 T whole-body MRI systems (Magnetom TIM Trio, Siemens Healthcare, Erlangen, Germany; VB17 software version) with standard 32-channel radio-frequency (RF) receive head coil and RF body coil for transmission.

Resting-state fMRI data were acquired using a multi-echo echoplanar imaging sequence with online reconstruction [30]: repetition time (TR) = 2.42 s; GRAPPA with acceleration factor = 2; flip angle = 90°; matrix size = 64 × 64 × 34; FOV = 240 × 240 mm; in-plane resolution = 3.75 × 3.75 mm; slice thickness = 3.75 mm with 10% gap, sequential slice acquisition, 34 oblique slices; bandwidth = 2368 Hz/pixel; echo times (TE) = 13, 30.55 and 48.1 ms. For pre-processing of these data, we used multi-echo independent component analysis (ME-ICA) [3,30] to identify the sources of variance in the fMRI time series that scaled linearly with TE and could therefore be confidently regarded as representing BOLD contrast. Other sources of fMRI variance, such as head movement, which were not BOLD dependent, and therefore did not scale with TE, were identified by ME-ICA and discarded. The retained independent components, representing BOLD contrast, were optimally recomposed to generate a broadband denoised fMRI time series at each voxel [3]. We used a wavelet transform for estimating functional connectivity in these data because of prior evidence indicating that cortical fMRI time series often have

slowly decaying positive autocorrelation [31,32]. This approach also allowed us to focus on functional associations between brain regions based on a physiologically relevant frequency range or wavelet scale. We used a discrete wavelet transform (Daubechies 4 wavelet), resulting in a BOLD signal oscillating in the frequency range 0.025–0.111 Hz (scales 2 and 3) [33].

Pre-processing and ME-ICA was performed with the AFNI tool meica.py [3] which we slightly modified for a more stable ICA and more conservative component selection. The forked release is based on the original ME-ICA V2.5 and was released on GitHub (doi://10.5281/zenodo.50505). Wavelet decompositions were implemented using an open source, R-based software library: brainwaver v. 1.6, which is freely downloadable at: https://cran.r-project.org/web/packages/brainwaver/index.html

(b) Functional magnetic resonance imaging connectivity and network analysis

To define regional nodes or parcels of cortex for network analysis, we used a backtracking algorithm [34] to parcellate the Freesurfer average (fsaverage) brain, subdividing regions of the Desikan–Killiany surface-based anatomical atlas of the human brain [35] into 308 smaller contiguous regions (nodes) with approximately homogeneous sizes (500 mm^2 on the surface). This parcellation template image in standard space was transformed to the native space of each individual's fMRI dataset and regional BOLD time series were estimated by averaging the time series over all voxels in each of the 308 parcels. Some regions (particularly near the frontal and temporal poles) were excluded because of low regional mean signal in at least one subject, and two participants were excluded because of poor co-registration between functional and anatomical data. Two further regions were later excluded from the analysis due to outlier values of the corresponding gene expression data (see the electronic supplementary material, figure S1). The fMRI dataset available for analysis thus consisted of 38 individual matrices of regional mean BOLD oscillations at each of 285 cortical regions.

Functional connectivity was estimated by the pairwise wavelet correlations between each possible pair of regional mean fMRI time series. The resulting functional connectivity or association matrices were thresholded to construct binary (undirected and unweighted) adjacency matrices or graphs [36]. We used the minimum spanning tree to ensure that the graphs were node-connected even at the sparsest connection density [37]. Additional edges were then superimposed in order of decreasing inter-regional correlation to construct networks with arbitrary connection density, in the range 0–100% of the total number of possible pairwise connections.

We used standard network metrics to characterize the topology of the fMRI network (see the electronic supplementary material). To estimate the inter- and intra-modular degree of each node, we first had to define the modular community structure of each network. The modularity, $Q(G)$, of a graph is proportional to the number of intra-modular edges compared to the number of intra-modular edges expected in a random graph [38]. The prototypical modules of the healthy brain functional network were derived from the sample mean wavelet correlation matrix. We used consensus modular decomposition [39] over 100 runs of the Louvain modularity algorithm on the 10% density graph constructed from this matrix. The Louvain algorithm parameter γ defining the coarseness of the modular partition was set at $\gamma = 2$, yielding eight modules. We demonstrate in the electronic supplementary material that our key results are robust to this parameter setting.

For each subject's binary graph, the degree centrality of each node, $k(i)$, $i = 1,2,3 \ldots 285$, was calculated as the total number of edges connecting it to the rest of the network. Total degree was subdivided into the intra-modular degree $k_{intra}(i)$, i.e. the number of edges connecting the ith node to other nodes within

4

the same module, and the inter-modular degree $k_{\text{inter}}(i)$, i.e. the number of edges connecting the ith node to nodes in other modules. For each node, we also estimated the participation coefficient, PC(i), which has been widely used as a measure of nodal role in a modular community structure [40,41]:

$$PC(i) = 1 - \sum_{m=1}^{M} \left(\frac{k_i(m)}{k_i} \right)^2, \qquad (2.1)$$

where $k_i(m)$ is the number of edges between node i and all nodes in module m, and the sum is over all M modules [42]. The PC of the index node is therefore inversely related to its intra-modular degree and will be close to PC = 0 if it is mostly connected to other nodes in the same module and close to PC = 1 if it is more homogeneously connected to all modules. The nodal measures $k_{\text{intra}}(i)$, $k_{\text{inter}}(i)$ and PC(i) were estimated for each subject with respect to the prototypical modular structure estimated for the sample as a whole. For each edge, we also estimated its connection distance (in millimetres) as the Euclidean distance between the linked pair of regional nodes in standard anatomical space. Topological metrics were calculated using the Brain Connectivity Toolbox, v. 2014-04-05 [43].

(c) Microarray data and pre-processing

Microarray data for six donors (H0351.1009, H0351.1016, H0351.1015, H0351.2002, H0351.1012 and H0351.2001) were available from the AIBS (http://human.brain-map.org/static/download; [19]). Five of the donors were male and one was female; three were Caucasian, two African-American and one Hispanic; mean age = 42.5 years (see the electronic supplementary material for more information on the AIBS dataset).

We used the Maybrain package (see the electronic supplementary material) to match the centroids of the regions of the fMRI parcellation template to the closest regional gene expression profile. Microarray data were averaged across all samples from all donors in homologous regions in both hemispheres. The data were also averaged across probes corresponding to the same gene, excluding probes that were not matched to gene symbols in the AIBS database. We used the Z-transformation to normalize mean expression of each gene for variance in its expression (see the electronic supplementary material). The final output was a (20 737 × 285) matrix, T, of Z-scored expression values for each of 20 737 genes estimated in 285 fMRI regions. Gene expression data for individual genes or subsets, such as the 19 genes selectively over-expressed in human supragranular cortex [10], and the 162 genes specialized for aerobic glycolysis (AG) [44], were available by sub-sampling the appropriate rows of this whole-genome brain regional transcription matrix (T).

(d) Parcellation into cytoarchitectonic classes

Each of the 308 regions in the cortical parcellation scheme was assigned to a cytoarchitectural type according to the classification scheme of von Economo & Koskinas [45] (figure 1 and electronic supplementary material, figure S2). This atlas subdivided the cortex into five types according to the laminar structure of the cortex and roughly corresponding to functional cortical specializations. Briefly, regions with poor laminar differentiation, particularly the primary motor cortex/precentral gyrus are structural type 1, regions generally considered to be association cortices are structural types 2 and 3, while secondary and primary sensory areas are types 4 and 5, respectively. The original classification of structural types does not discriminate between true six-layered isocortex and mesocortex or allocortex, which have markedly different cytoarchitectures and ontogenies [46]. We therefore defined two additional subtypes: limbic cortex which included the entorhinal, retrosplenial, presubicular and cingulate cortices, and thus primarily constitutes allocortex;

and the insular cortex which contains granular, agranular and dysgranular regions, and is therefore not readily assigned a single structural type. Structural types were manually assigned to cortical regions based on visual comparison with von Economo & Koskinas's parcellation and anatomical landmarks.

(e) Partial least squares

To explore the associations between topological centrality and distance metrics at each node of the fMRI networks, and transcription across the whole genome, we used the multivariate technique of PLS.

PLS is an established and widely used multivariate method for identifying associations between a set of response variables and a set of predictor variables, especially when the number of predictor variables exceeds the number of observations, and when the predictor variables are highly interdependent or multi-collinear [26,47]. In this case, the (285 × 20 737) regional gene transcription matrix T comprised the predictor variables; the (285 × 3) matrix C comprised the response variables of intra-modular degree, inter-modular degree and mean connection distance for each regional node of the fMRI networks.

PLS is related to principal component analysis (PCA) and combines a PCA-style dimensionality reduction with linear regression. While PCA identifies the so-called principal components in the data that best explain the variance in the predictor variables T, PLS finds components from T (gene expression) that have maximum *covariance* with the response variables in C (fMRI network measures). The total number of components needed to exactly fit the response C is limited by the number of observations (in this case the number of brain regions). The PLS components are ranked by covariance between predictor and response variables, so the first few PLS components (PLS1, PLS2, PLS3, etc.) will provide the optimal low-dimensional representation of the covariance between the higher dimensional data matrices (see the electronic supplementary material).

We used non-parametric data resampling techniques for inferential analysis of PLS results. We tested the goodness of fit of low-dimensional PLS components by repeating the analysis 1000 times after shuffling the regional labels assigned to each set of three response variables. We note that the dependency between spatially neighbouring regions that exists in both fMRI and transcriptomic datasets may lead to an over-estimation of significance by this simple, spatially naive null model. In the electronic supplementary material, figure S4, we confirm that the PLS results were robust to the use of more sophisticated null models that account for spatial correlations expected due to homogeneous tissue type within anatomical regions. We used bootstrapping (resampling with replacement of the 285 cortical regions) to estimate the error on the PLS weights estimated for each gene. The ratio of the weight of each gene to its bootstrap standard error was used to rank the genes according to their contribution to each PLS component. The list of ranked genes for the first three PLS components can be found in the electronic supplementary material: Vertes_PLS_GOenrichment.xlsx

(f) Gene ontology and enrichment analysis

We used gene ontology (GO) tools for enrichment analysis of the ranked gene lists defined by the first three PLS components (Gorilla: http://cbl-gorilla.cs.technion.ac.il, version 30 January 2016) [27,28]. The GO terms are based on a large online database of gene annotations corresponding to 'biological processes' and 'cellular components' [48]. Enrichment analysis identified GO terms that were over-represented among the most positively and most negatively weighted genes on each PLS component. We further filtered the resulting list of enriched GO terms
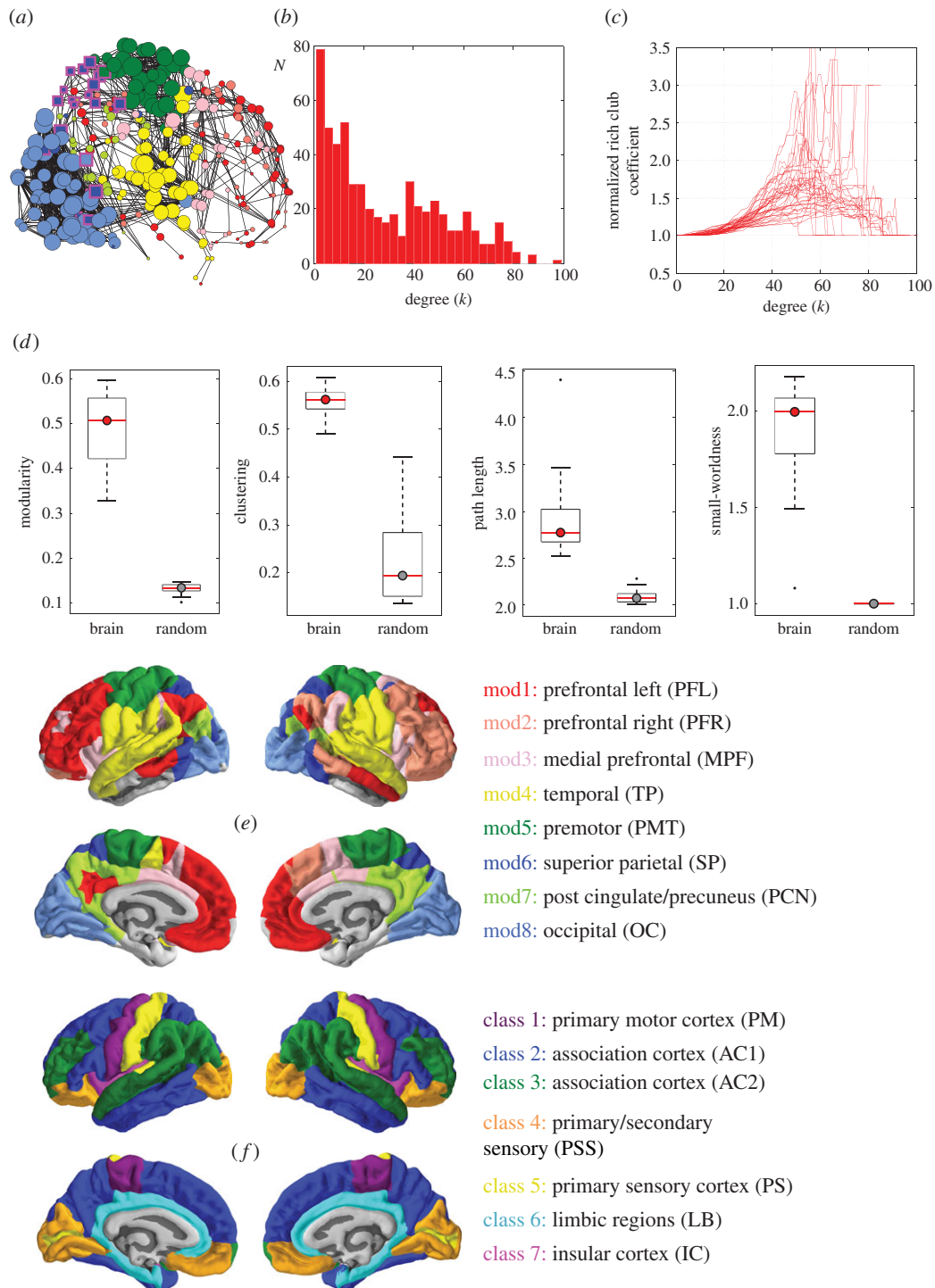
Figure 1. Complex topology of fMRI brain networks. (*a*) Network representation of brain functional connectivity. Colours represent eight distinct modules; the size of nodes is proportional to their degree; only the top 4% strongest connections are shown for clarity. (*b*) Degree distribution of the brain functional networks, pooled across subjects. (*c*) Normalized rich club curves of each participant's brain functional network. (*d*) Boxplots showing key network measures for the brain functional networks (in red) compared to randomized networks with preserved degree distribution (grey). From left to right, the metrics shown are: modularity Q, clustering C, path length L, and small-worldness σ. (*e*) Cortical surface map colour-coding brain regions according to fMRI modules, as in panel (*a*). The legend describes the approximate anatomical location of each module and defines the acronym with which each module (mod) is represented in figures 2 and 3. (*f*) Cortical surface map colour-coding brain regions according to von Economo & Koskinas's cytoarchitectonic classification [45]. Class 1 (purple): granular cortex, primary motor cortex. Classes 2 and 3 (blue and green): association cortex. Class 4 (orange): dysgranular cortex, primary/secondary sensory cortex. Class 5 (yellow): agranular cortex, primary sensory cortex. Class 6 (cyan): limbic regions, allocortex. Class 7 (magenta): insular cortex. The legend also defines the acronym with which each cytoarchitectonic class is represented in figures 2 and 3.

(i) by retaining only terms that were significantly enriched after controlling the *p*-value for significance of each term so that the false discovery rate (FDR) over all GO terms was $P_{FDR} < 0.001$ and (ii) for visualization purposes we also discarded enriched GO terms for biological processes associated with over 2500 genes, which typically correspond to general ontological terms near the top of the hierarchy, such as 'cellular process' or 'organelle organization'. All excluded terms are still listed and highlighted in grey in the electronic supplementary material: Vertes_PLS_GOenrichment.xlsx.
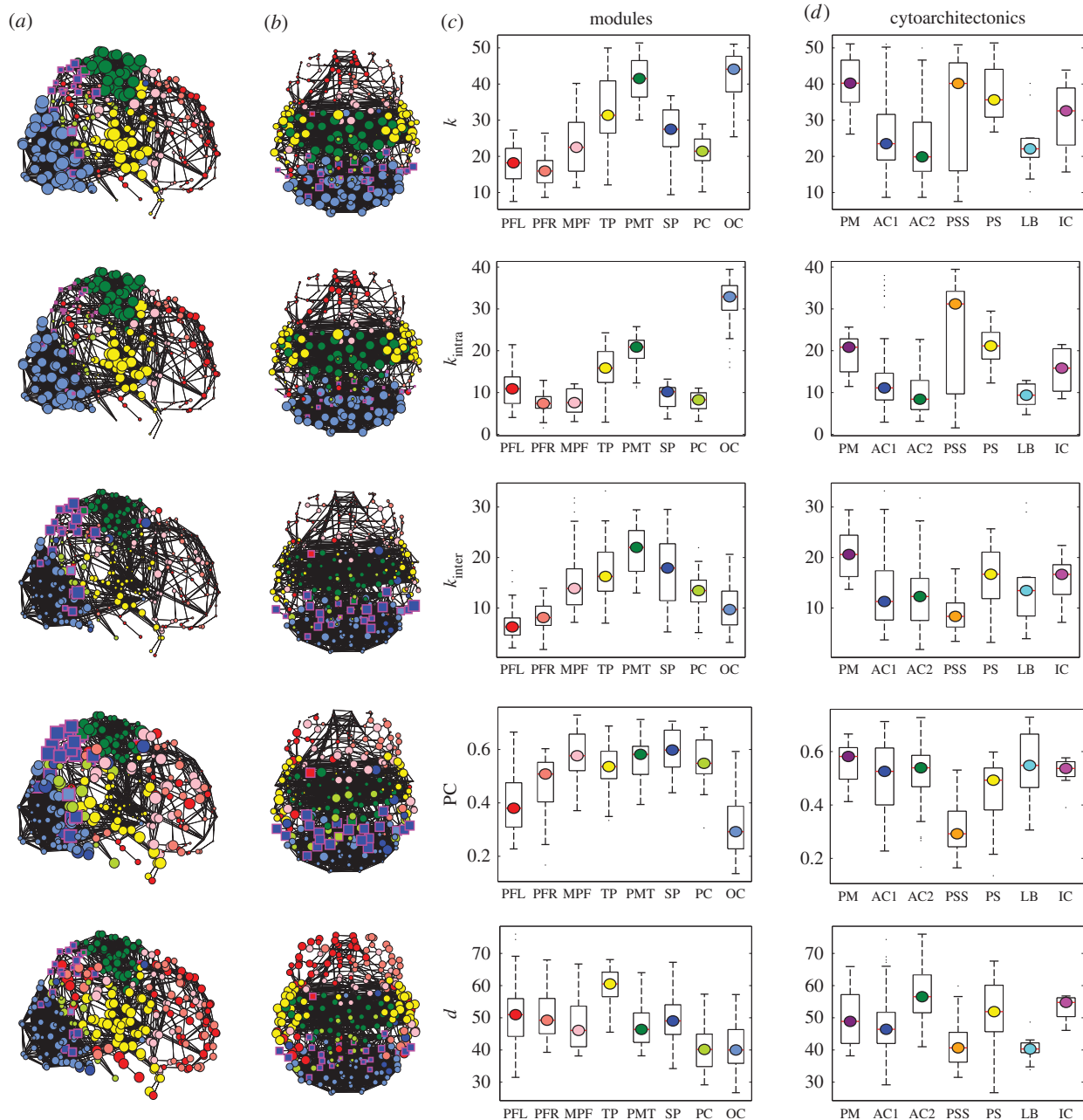
Figure 2. Anatomical and cytoarchitectonic patterning of fMRI network hubs. (a) Binary graphs constructed at 4% connection density (for clarity) showing a sagittal view of the brain. Nodal size was scaled by five nodal metrics: from top to bottom, total degree (k), intra-modular degree (k_{intra}), inter-modular degree (k_{inter}), participation coefficient (PC), and average nodal distance (d). Nodes are coloured by module, as defined in figure 1. Nodes with high PC (>0.5) are highlighted by square markers with a magenta outline. (b) Axial view of the brain networks in panel (a). (c) Boxplots showing the distribution of nodal distance and nodal topological metrics in each of the eight modules. Modules are colour-coded and named according to the scheme shown in figure 1. (d) Boxplots showing the distribution of nodal distance and nodal topological metrics in each of the seven cytoarchitectonic classes as defined by von Economo & Koskinas's [45] classification of cortical laminar patterns. Cytoarchitectonic classes are numbered and colour-coded according to the scheme in figure 1.

To visualize the results of whole-genome enrichment analysis, we used the online tool REViGO (http://revigo.irb.hr) to summarize the list of significant GO terms by selecting representative subsets of the terms using a simple clustering algorithm that relies on measures of semantic similarity between terms [29]. For example, the terms 'respiratory electron transport chain', 'electron transport chain' and 'mitochondrial electron transport, NADH to ubiquinone' will be clustered together by the algorithm and only some of the terms will be retained. To further facilitate interpretation, REViGO was used to plot the remaining significant GO terms in semantic space, where semantically similar terms are represented close to one another. Markers are scaled according to the $\log_{10}$ of the p-value for the significance of each term. Clusters in these plots therefore represent families of related GO terms, and the GO term associated with the largest marker in each cluster can be annotated to label the whole cluster in a representative manner.

For the more hypothetically driven enrichment analysis of gene lists associated *a priori* with (i) supragranular layers of human cortex (human supragranular enriched, HSE) [10] or (ii) aerobic glycolysis (AG) [44] we also used permutation testing for non-parametric inference. We estimated the PLS weightings of 1000 randomly drawn sets of 19 genes and compared the PLS weights of the HSE genes to this permutation distribution to estimate the probability of HSE gene enrichment of each PLS component under the null hypothesis. We note that this permutation procedure does not take into account the correlation between HSE genes or their average expression values or their exclusively cortical origin in sampling the null distribution. More sophisticated null models
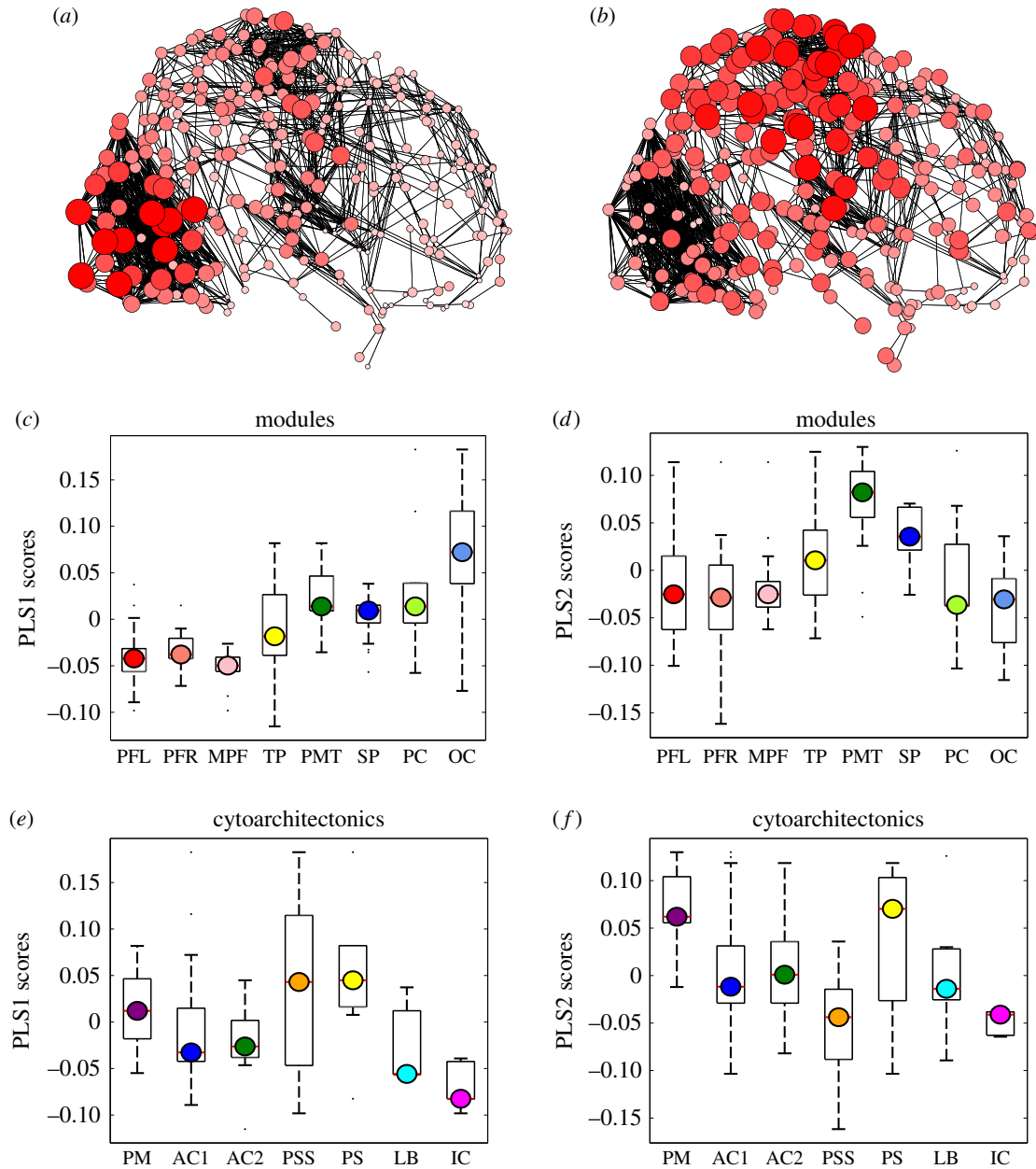
Figure 3. Gene expression profiles associated with fMRI network topology. (*a,b*) Binary graphs constructed at 4% connection density (for clarity) showing nodes with size and colour saturation scaled by regional scores on PLS1 (*a*) and PLS2 (*b*). Larger, darker nodes represent regions with higher PLS scores, i.e. higher expression levels of genes positively weighted on the corresponding PLS component. (*c,d*) Boxplots representing the distribution of regional scores on PLS1 (*c*) and PLS2 (*d*) in each of the eight network modules shown in figure 1. (*e,f*) Boxplots representing the distribution of regional scores on PLS1 (*e*) and PLS2 (*f*) in each of the seven cytoarchitectonic classes shown in figure 1.

for permutation testing that controlled for these or other characteristics of candidate genes will be important to develop for computational inference in future studies.

3. Results

(a) Human functional magnetic resonance imaging network parameters: spatial patterning and cytoarchitectonic differentiation

We constructed graphs of the human brain functional network by binary thresholding of the pairwise inter-regional wavelet correlation (functional connectivity) matrices estimated for $n = 38$ healthy volunteers aged 20–24 years. We focus most attention on the characteristics of the graph with 10%

connection density. This fMRI network had complex topological properties consistent with many prior studies [13], including a broad scale degree distribution, a rich club, a community structure comprising eight modules, high clustering, short path length and small-worldness (see figure 1*a–d* and electronic supplementary material for definitions of these widely used network metrics).

At each node in the fMRI network, we estimated four topological parameters: total degree (k), inter-modular degree (k_{inter}), intra-modular degree (k_{intra}) and PC. We note that definition of k_{inter}, k_{intra} and PC will depend on both the network connection density and the resolution parameter, γ, defining the coarseness of the modular partition. In coarser decompositions with fewer modules, for example, fewer edges will be classed as inter-modular and the inter-modular degree distribution will shift to the left. We show in the electronic supplementary material

Table 1. Correlations between nodal fMRI variables and spatial coordinates. Matrix showing the correlations (Pearson's r) between MRI variables of interest as well as spatial coordinates. The y-coordinate corresponds to the rostro-caudal axis and the z-coordinate corresponds to the dorsoventral axis. We use the absolute value of the x-coordinate (corresponding to the medio-lateral axis) due to the symmetry between brain hemispheres. Significant correlations and anti-correlations are highlighted in italic and bold, respectively; $p < 0.05$. k = total degree, PC = participation coefficient, k_{intra} = intra-modular degree, k_{inter} = inter-modular degree, d = mean Euclidean distance of edges connecting each node to the network.

	k	PC	k_{intra}	k_{inter}	d	$|x|$	y	z
k	1.00	0.05	*0.83*	*0.63*	−0.09	−0.11	−**0.49**	*0.23*
PC	0.05	1.00	−**0.45**	*0.73*	*0.27*	*0.20*	0.05	*0.51*
k_{intra}	*0.83*	−**0.45**	1.00	0.09	−**0.24**	−**0.19**	−**0.50**	−0.05
k_{inter}	*0.63*	*0.73*	0.09	1.00	*0.18*	0.07	−**0.19**	*0.48*
d	−0.09	*0.27*	−**0.24**	*0.18*	1.00	*0.81*	*0.16*	−0.06
$|x|$	−0.11	*0.20*	−**0.19**	0.07	*0.81*	1.00	0.00	−**0.19**
y	−**0.49**	0.05	−**0.50**	−**0.19**	*0.16*	0.00	1.00	−0.09
z	*0.23*	*0.51*	−0.05	*0.48*	−0.06	−**0.19**	−0.09	1.00

that similar results were obtained with a coarser decomposition defining only four modules, as well as with a range of thresholds on the functional connectivity matrix, yielding networks with connection densities in the range 10–30%. For each node, we also estimated four spatial parameters: the connection distance (mean Euclidean distance of nodal edges) and the three-dimensional (x, y, z) coordinates of each node's location in anatomical space (electronic supplementary material, table S1).

The topological parameters were correlated with each other by construction: total degree, $k = k_{inter} + k_{intra}$, was positively correlated with both inter-modular degree k_{inter} and intra-modular degree k_{intra}; participation coefficient PC $\sim 1/k_{intra}$ was positively correlated with inter-modular degree and negatively correlated with intra-modular degree. The spatial parameter of connection distance was positively correlated with inter-modular degree and PC, but negatively correlated with intra-modular degree (electronic supplementary material, table S1).

The three-dimensional location of nodes was correlated with their topological properties (table 1): intra-modular degree was negatively correlated with y location, meaning that intra-modular hubs were concentrated in posterior cortical regions; whereas inter-modular degree and PC were positively correlated with z location, meaning that inter-modular hubs were concentrated in superior cortical regions. The spatial patterning of these parameters can also be represented by anatomical maps of nodal topology and connection distance (figure 2a). Hubs defined by high intra-modular degree were concentrated in occipital and somatosensorimotor cortex, or spatially patterned on the rostro-caudal axis; whereas hubs defined by high inter-modular degree and PC were concentrated in somatosensorimotor and superior parietal cortex, or spatially patterned on the dorsoventral axis. Nodes with greater connection distance were concentrated in lateral cortex, reflecting the existence of long-distance, inter-hemispheric connections between bilaterally homologous cortical areas (figure 2a–c).

Following prior work relating local cytoarchitectonic characteristics to nodal properties of human structural connectomes [49], here we explored the relationships between cortical histology and fMRI network parameters. To this end, we mapped each regional node of the fMRI network to an existing atlas of cytoarchitectonic areas classified according to the scheme of von Economo & Koskinas (figure 1f, §2d) [45].

Both intra- and inter-modular degree were significantly different between cytoarchitectonic classes (k_{intra}: $F_{6,278} = 22.7$, $p < 0.001$; k_{inter}: $F_{6,278} = 11.1$, $p < 0.001$). Intra-modular degree was highest in primary and secondary sensory cortex and lowest in association cortex and limbic regions. Inter-modular degree was highest in primary motor cortex and lowest in secondary sensory cortex. Because both total degree and PC were constructed from intra- and inter-degree, they also showed significant differences between cytoarchitectonic classes (k: $F_{6,278} = 13.6$, $p < 0.001$; PC: $F_{6,278} = 22.7$, $p < 0.001$; figure 2d), with similar patterning to k_{intra} and k_{inter}, respectively. The average nodal connection distance was also significantly different between cytoarchitectonic classes (repeated measures ANOVA for main effect of class: $F_{6,278} = 21.3$, $p < 0.001$; figure 2d). The longest connection distances were in association cortex, whereas the shortest distances were in secondary sensory cortex and limbic regions (figure 1). We note that the association between inter-modular degree (as well as PC) and cytoarchitecture depended on the coarseness of the modular partition (see the electronic supplementary material).

(b) Partial least-squares analysis of functional magnetic resonance imaging network parameters and gene expression

To investigate how this spatially patterned set of fMRI network nodal topology (k_{intra} and k_{inter}) and distance parameters was related to local expression of approximately 20 000 genes, we co-registered the fMRI regional nodes in the same anatomical space as the AIBS dataset of human brain gene expression. For each fMRI node, we then estimated the mean regional expression of each of 20 737 genes (§2e). To explore a low-dimensional representation of the multivariate relationships between the matrix of response variables (fMRI network nodal parameters) and the matrix of predictor variables (gene expression profiles) we used PLS (§2e).

The first three PLS components accounted for about 37% of the total variance in nodal metrics and this measure of the goodness of fit was statistically significant ($p < 0.001$) for a spatially naive permutation test (see the electronic

Table 2. Correlations between gene expression profiles and fMRI network topology and geometry. The first three partial least-squares components (PLS1, PLS2 and PLS3) were differently correlated (Pearson's r) with: total degree (k), participation coefficient (PC), intra-modular degree (k_{intra}), inter-modular degree (k_{inter}), average nodal distance (d) and spatial locations in three dimensions ($|x|$, y, z). Significant correlations and anti-correlations are highlighted in italic and bold, respectively; $p < 0.05$.

	k	PC	k_{intra}	k_{inter}	d	$|x|$	y	z
PLS1	*0.50*	**−0.26**	*0.59*	0.07	**−0.34**	**−0.19**	**−0.66**	0.10
PLS2	0.26	*0.45*	−0.01	*0.48*	0.28	0.23	**−0.15**	*0.53*
PLS3	0.11	−0.05	0.10	0.06	**−0.55**	**−0.67**	0.17	0.15

supplementary material for additional data and for spatially constrained permutation tests).

The first PLS component (PLS1) was positively correlated with intra-modular degree and not significantly predictive of inter-modular degree (table 2). This means that genes positively weighted on this component were relatively over-expressed in intra-modular hubs (figures 2 and 3). Given the pattern of correlations already observed between the fMRI network parameters (table 1), it is not surprising that PLS1 was also negatively correlated with connection distance, meaning that positively weighted genes were relatively under-expressed in nodes with many long-distance connections. We also note that PLS1 results were coherent with nodal parameters (PC, x, y, z) that were not explicitly included in the PLS model but were known to be significantly correlated with intra-modular degree (table 1). Thus, PLS1 scores were negatively correlated with the nodal PC and were spatially patterned on the y-axis (rostro-caudal), with positive PLS1 scores located in posterior cortical nodes (figure 3a,c). Nodal PLS1 scores were also significantly different between cytoarchitectonic classes ($F_{6,278} = 14.4$, $p < 0.001$; figure 3e) with positive PLS1 scores located in primary and secondary sensory cortical areas.

By contrast, the second PLS component (PLS2) was predictive of inter-modular degree and not significantly predictive of intra-modular degree (table 2). This means that genes positively weighted on this component were relatively over-expressed in inter-modular hubs (figures 2 and 3). Convergently, PLS2 was positively correlated with connection distance, meaning that positively weighted genes were relatively over-expressed in nodes with many long-distance connections. PLS2 results were coherent with nodal parameters (PC, x, y, z) that were not explicitly included in the PLS model but were known to be significantly correlated with inter-modular degree (table 1). Thus, PLS2 scores were positively correlated with nodal PC and spatially patterned on the z (dorsoventral) and x (medio-lateral) axes, with high PLS2 scores located in superior and lateral cortex (figure 3b,d). Nodal PLS2 scores were also significantly different between cytoarchitectonic classes ($F_{6,278} = 19.4$, $p < 0.001$; figure 2f) with the highest PLS2 scores located in primary motor and primary sensory cortical areas and the lowest scores located in secondary sensory cortex. We note that the association between PLS2 and cytoarchitecture depended on the coarseness of the modular partition because coarser partitions by definition lead to longer inter-modular connections (see the electronic supplementary material).

In short, the first two PLS components defined independent gene expression profiles that were specifically associated with (high) intra-modular degree and (short) connection distance (PLS1) or with (high) inter-modular degree and (long)
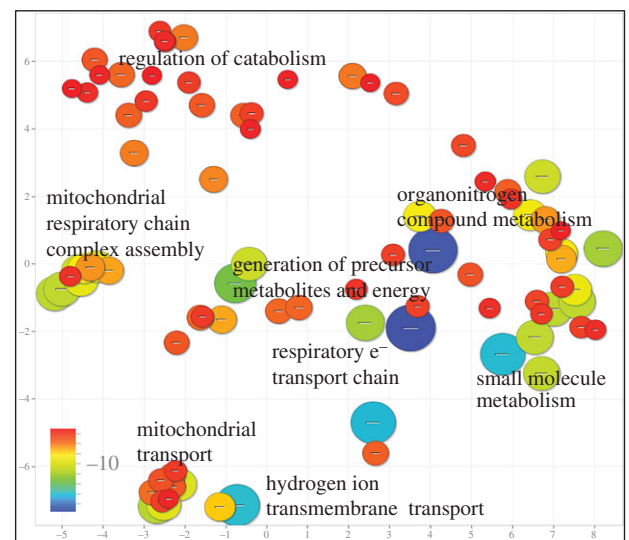

Figure 4. Enrichment analysis of PLS2 gene expression profile associated with long-distance connections and inter-modular hubs in fMRI networks. Significantly enriched GO terms are plotted in semantic space such that similar terms are represented close to one another. Markers are scaled and coloured according to the $\log_{10}$ of the p-value for the significance of each term. Large blue circles are highly significant ($P_{FDR} < 10^{-15}$), while small red circles are less so ($P_{FDR} < 0.001$).

connection distance (PLS2). The third PLS component defined an independent gene expression profile that was not significantly predictive of inter- or intra-modular degree (or PC), but was significantly predictive of connection distance, and was spatially patterned in all three dimensions (see the electronic supplementary material). Since we were hypothetically motivated to explore the relationships between fMRI network topology and gene expression, PLS3 was not as relevant as PLS1 or PLS2. The fourth and subsequent PLS components inevitably explained progressively smaller proportions of covariance between network metrics and gene expression, and will be less robust to noise. We therefore focus further attention only on the first two components of the PLS solution.

(c) Enrichment analysis of gene expression profiles (PLS1 and PLS2) associated with functional magnetic resonance imaging network topology and connection distance

Given the statistical independence of PLS components, these results indicate that there are specific or distinct gene

expression profiles associated with different nodal roles in fMRI networks. We used enrichment analysis to resolve the differences in transcriptional predictors of nodal topology and connection distance in more detail. We found that PLS1 and PLS2 components were significantly enriched for distinct biological functions.

Genes significantly over-expressed in association with high intra-modular degree (positively weighted on PLS1) were significantly enriched ($P_{FDR} < 0.001$) for GO terms related to transcriptional regulation in the nucleus. By contrast, genes significantly over-expressed in association with high inter-modular degree and longer connection distance (positively weighted on PLS2) were significantly enriched ($P_{FDR} < 0.001$) for GO terms related to oxidative metabolism and mitochondria (figure 4); see the electronic supplementary material Vertes2016_PLS_GOenrichment.xlsx and figure S5 for full GO enrichment results. We additionally show in the electronic supplementary material that these results were robust to a range of parameter settings in constructing fMRI networks and to the use of alternative PLS response variables. We also show that the key results were preserved in sensitivity analyses across AIBS post-mortem brain donors.

As well as these hypothesis-free analyses that tested for enrichment of all GO terms, we also conducted two more hypothesis-driven enrichment analyses: (i) to test each PLS component for over-expression of a set of genes anatomically specific to supragranular layers of human cortex (human supragranular enriched, HSE [10]) and (ii) to test each PLS component for over-expression of a set of genes functionally specialized for aerobic glycolysis [44].

It has been shown [10] that the transcriptional profile of 19 HSE genes over-expressed specifically in supragranular layers of human cortex (but not mouse cortex) was different between brain regions with predominantly local connectivity compared to association cortical regions with a higher proportion of long-range connectivity. This motivated us to test the hypothesis that HSE genes would be differentially enriched in the first two PLS components, with greater HSE enrichment expected for PLS2 because of its specific association with long-distance and inter-modular connections in these data. As predicted, we found that HSE genes were significantly enriched among the over-expressed genes positively weighted on PLS2 (permutation test, $p < 0.001$); but HSE genes were not significantly enriched among the over-expressed genes positively weighted on PLS1.

Separately, it has been shown [44] that a set of 116 genes were functionally specialized for AG in the brain. This corresponds to non-oxidative metabolism of glucose despite the presence of oxygen. Because PLS2 was specifically enriched for oxidative metabolism genes, we were motivated to investigate whether AG genes would also be enriched in the second PLS component. We found that AG genes were not over-represented among the genes positively weighted on PLS2 (permutation test, $p > 0.05$). We note, however, that this result depended on the coarseness of the modular partition in a predictable manner, with coarser partitions by definition constraining inter-modular links to span longer distances. As connection distance was weighted more strongly on PLS2, AG genes were correspondingly more enriched in PLS2 in the analysis of the coarser four-module community structure (see the electronic supplementary material).

4. Discussion

The biological validation of human fMRI networks has been challenging, not least because there are fundamental questions outstanding about the biological sources of the BOLD signal itself. We cannot yet securely reduce the basic fMRI observation of inter-regional correlation between pairs of functionally connected BOLD time series to a mechanistic explanation in terms of cellular processes, neuronal physiology, neurovascular coupling or anatomical connectivity. This has made it difficult to test (refute or validate) economical models of human fMRI networks by the classical reductionist logic of explanatory mechanistic coupling between cellular processes, at a micro scale, and network topology metrics, at a macro scale (table 2).

Most of the previous evidence for biological validity of human MRI connectomes has therefore rested on analogical rather than reductionist logic. Informative analogies make comparisons between a poorly understood system and a more certainly or completely understood system of the same type. There have been several studies recently supporting the value of this approach for comparative connectomics—comparing the topology of human neuroimaging networks to the topology of more biologically specified nervous systems [50]. For example, MRI networks generally comprise a rich club of densely inter-connected high-degree hubs. Human brain network-rich clubs are topologically integrative by mediating many of the shortest paths between pairs of more peripheral nodes in different modules [51]. Rich clubs of fMRI co-activation networks are associated with diverse cognitive functions including higher order executive tasks [16], and are biologically expensive in terms of wiring cost [52,53]. These MRI observations are suggestive of an economical trade-off, between minimizing biological cost and maximizing topological value, and have been affirmed by demonstration of analogous findings in more certainly known nervous systems. In particular, the anatomical network of axonal projections and synapses between the 302 neurons of *Caenorhabditis elegans* [53] includes a topologically integrative rich club that is expensively wired and comprises command interneurons known to be functionally important for coordinated movement and adaptive behaviours [54]. Indeed, high-cost, high-value rich clubs have now been demonstrated across a wide range of scales, species, experimental techniques for network mapping, and computational models of network generation [55–58]. It seems plausible, on this basis, that the topological properties of human fMRI networks are not idiosyncratic epiphenomena but are instead representative of a general class of brain networks that have been naturally selected by the same competitive pressures for relatively low biological cost and high topological integration [18].

Here, we have provided further evidence in support of this economical model of brain network organization by a more reductionist approach. We spatially co-registered fMRI network parameters with whole-genome expression data on anatomically corresponding brain regions. We used the multivariate method of PLS to define the gene expression profiles that were optimally predictive of fMRI network parameters. We found that the first two PLS components were specifically predictive of distinct nodal network phenotypes: PLS1 was predictive of high intra-modular degree and short connection distance; whereas PLS2 was predictive of high inter-modular degree and long connection distance (PLS3 was not predictive

of nodal topology; table 2). The genes most positively weighted on PLS2, and therefore relatively over-expressed in brain regions mediating many long-distance inter-modular connections, were enriched for oxidative metabolism and mitochondria. This result is clearly convergent with the expectation that integrative network features should be more metabolically expensive. It has previously been shown that hubs of human brain networks have greater rates of glucose metabolism and blood flow [59,60]; these observations have been regarded as consistent with prior knowledge of the brain's metabolic budget [61]. Synaptic transmission and maintenance of resting membrane potentials represent major demands on the brain's supply of ATP, generated mainly by mitochondrial metabolism of glucose [62]. Our data further suggest that hubs mediating more long-distance connections between modules face greater metabolic demands than hubs mediating more short-distance connections within modules. This may be because long-distance, inter-modular hubs must energetically restore and maintain electrical potentials across a greater surface area of axonal membrane. These results from human fMRI network analysis are analogous to recent results from tract-tracing and gene expression data in the mouse indicating that genes regulating oxidative metabolism were strongly co-expressed in pairs of brain regions that included a hub [63].

It was also notable that the genes associated with inter-modular degree and long distance (positively weighted on PLS2 in these data) were enriched for a set of 19 genes, expressed specifically in human supragranular cortex, that may have been necessary for the characteristically human evolution of cortico-cortical connectivity and associative cognitive processes [10]. The HSE genes were not significantly enriched in the genes associated with intra-modular degree and short distance (positively weighted on PLS1). Interestingly, the positively weighted genes on PLS2 were not significantly enriched for 116 genes specialized for AG, suggesting that the energetic resources of long-distance, inter-modular hubs are largely provided by oxidative metabolism rather than AG. However, in assessing these and other enrichment results, it is important to bear in mind that they are based on a small experimental sample ($n = 6$ post-mortem adult brains) and that individual differences in gene expression can have a marked effect on group statistics (see the electronic supplementary material for sensitivity analysis).

Interpreting the relationship between gene expression and network topology is potentially complicated by the spatial patterning of both transcriptional and topological phenotypes. Regional co-expression, or sharing of the same gene expression profile between a pair of brain regions, is related to the spatial distance between them: regions that are closer to each other will have greater genomic co-expression. Functional connectivity between regions is also conditional on distance: regions that are closer to each other are more likely to have correlated fMRI time series and so to form clusters and modules of locally inter-connected nodes in fMRI brain graphs. The question arises: does an association between gene co-expression and fMRI connectivity trivially reflect the confounding effect of distance? Previous studies have been concerned that topological clusters and modules of nodes may have transcriptional profiles in common simply because they are spatial neighbours, or members of the same cytoarchitectonic class [7]. For example, in several studies of transcriptional similarity and anatomical connectivity in the rodent brain, a statistical correction for the distance of edges was used to show that greater genomic

co-expression between connected nodes was not entirely attributable to the shorter distance between them [63–68].

Our analysis of network topology has focused on nodes rather than edges and we have shown that different nodal properties have different spatial patterning, e.g. intra-modular degree was correlated with nodal location on the y-axis, whereas inter-modular degree was correlated with nodal location on the z-axis and connection distance was correlated with nodal location on the x-axis. PLS analysis enabled us simultaneously to explore the relationships between all three network parameters and the whole genome. We found two distinct components of gene expression that were specifically related to different aspects of nodal topology and connection distance, and had different spatial patterning as well as differential enrichment for biological processes. This pattern of results is not obviously attributable to a homogeneous and spatially isotropic effect of distance as a confounding factor of both genomic co-expression and functional connectivity [63,64,68]. Instead, these results suggest to us that multiple dimensions of spatially patterned gene expression define functionally connected systems of cytoarchitectonically similar nodes. By this account, the intertwined patterning of spatial location, nodal topology and gene expression is in fact the phenotype of interest and it would therefore not be appropriate to correct the data for spatial location prior to analysis. This is not to say that potentially problematic issues related to spatial location were entirely neglected in our analysis. For statistical inference on PLS results by permutation testing, for example, the significance of association between nodal topology and gene expression defined by the first two PLS components could be artefactually inflated by randomly permuting the regional network metrics regardless of their spatial proximity. This point is related to the more general issue of exchangeability in the proper design of permutation tests. If units of observation (regional nodes) are not expected to be independent of each other, e.g. due to their close proximity in space or time, they are not exchangeable under the null hypothesis and a permutation test based on the random permutation of individual units will not be valid. We addressed this technical concern with a block permutation algorithm that randomly permuted spatially contiguous subsets of regional nodes, rather than permuting each node individually (see the electronic supplementary material for details). This methodological refinement did not materially affect the statistical robustness of our findings although it will be important to investigate more sophisticated permutation testing methods in future.

There are a number of theoretical and methodological limitations to this work. First, the results reported here are correlational, not causative. More generally, there remains an explanatory gap in determining whether and how these specific gene profiles support short-range intra-modular connectivity (PLS1) or long-range inter-modular connectivity (PLS2). Second, the matching between fMRI and transcriptional data is imperfect as the transcriptional data are based on six human brains (mean age = 42.5 years) sampled at approximately 500 locations in each hemisphere, whereas the fMRI data are recorded from 38 healthy young adults (mean age = 22 years) sampled at 308 locations across the cerebral cortex. Third, the interpretability of enrichment analyses based on these microarray data is hampered by the fact that expression was not measured separately in different cell types. Up- or down-regulation of a gene may therefore equally represent variations in the density of certain cell

types enriched for that gene or variation of expression levels of the gene within a cell type. Fourth, we approximated the distance between connected nodes by the Euclidean (straight line) distance, which is anatomically unrealistic, especially for inter-hemispheric connections between posterior parietal and anterior prefrontal cortex which follow a nonlinear trajectory constrained by the contours of the corpus callosum. In addition, our nodal measure of connection distance averaged the distance of all edges, whereas it would be informative in future studies to distinguish between the nodal distance of intra-hemispheric and inter-hemispheric or trans-callosal connections.

Nonetheless, it is encouraging to see that reductionist strategies, linking macro network properties measured in human fMRI to the underlying micro organization of cortical cytoarchitectonics and gene expression, are increasingly tractable and informative. The fMRI/mRNA results reported here have provided mechanistic support for the economic model that highly integrative network nodes, mediating many long-distance connections between modules, are metabolically expensive.

References

1. Logothetis NK, Wandell BA. 2004 Interpreting the BOLD signal. *Annu. Rev. Physiol.* **66**, 735–769. (doi:10.1146/annurev.physiol.66.082602.092845)

2. Fox MD, Raichle ME. 2007 Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.* **8**, 700–711. (doi:10.1038/nrn2201)

3. Kundu P et al. 2013 Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proc. Natl Acad. Sci. USA* **110**, 16 187–16 192. (doi:10.1073/pnas.1301725110)

4. Patel AX, Kundu P, Rubinov M, Jones PS, Vértes PE, Ersche KD, Suckling J, Bullmore ET. 2014 A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. *NeuroImage* **95**, 287–304. (doi:10.1016/j.neuroimage.2014.03.012)

5. Raichle ME, Snyder AZ. 2007 A default mode of brain function: a brief history of an evolving idea. *NeuroImage* **37**, 1083–1090. (doi:10.1016/j.neuroimage.2007.02.041)

6. Greicius MD, Srivastava G, Reiss AL, Menon V. 2004 Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proc. Natl Acad. Sci. USA* **101**, 4637–4642. (doi:10.1073/pnas.0308627101)

7. Hawrylycz M et al. 2015 Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* **18**, 1832–1844. (doi:10.1038/nn.4171)

8. Wang G-Z et al. 2015 Correspondence between resting-state activity and brain gene expression. *Neuron* **88**, 659–666. (doi:10.1016/j.neuron.2015.10.022)

9. Richiardi J et al. 2015 Correlated gene expression supports synchronous activity in brain networks. *Science* **348**, 1241–1244. (doi:10.1126/science.1255905)

10. Krienen FM, Yeo BTY, Ge T, Buckner RL, Sherwood CC. 2016 Transcriptional profiles of supragranular-enriched genes associate with corticocortical network architecture in the human brain. *Proc. Natl Acad. Sci. USA* **113**, E469–E478. (doi:10.1073/pnas.1510903113)

11. Raichle ME. 2010 Two views of brain function. *Trends Cogn. Sci.* **14**, 180–190. (doi:10.1016/j.tics.2010.01.008)

12. Salvador R et al. 2005 Neurophysiological architecture of functional magnetic resonance images of human brain. *Cereb. Cortex* **15**, 1332–1342. (doi:10.1093/cercor/bhi016)

13. Achard S, Salvador R, Whitcher B, Suckling J, Bullmore E. 2006 A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* **26**, 63–72. (doi:10.1523/JNEUROSCI.3874-05.2006)

14. Damoiseaux JS, Rombouts SARB, Barkhof F, Scheltens P, Stam CJ, Smith SM, Beckmann CF. 2006 Consistent resting-state networks across healthy subjects. *Proc. Natl Acad. Sci. USA* **103**, 13 848–13 853. (doi:10.1073/pnas.0601417103)

15. Fornito A et al. 2011 Genetic influences on cost-efficient organization of human cortical functional networks. *J. Neurosci.* **31**, 3261–3270. (doi:10.1523/JNEUROSCI.4858-10.2011)

16. Crossley NA, Mechelli A, Vertes PE, Winton-Brown TT, Patel AX, Ginestet CE, McGuire P, Bullmore ET. 2013 Cognitive relevance of the community structure of the human brain functional coactivation network. *Proc. Natl Acad. Sci. USA* **110**, 11 583–11 588. (doi:10.1073/pnas.1220826110)

17. Crossley NA, Mechelli A, Scott J, Carletti F, Fox PT, McGuire P, Bullmore ET. 2014 The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain* **137**, 2382–2395. (doi:10.1093/brain/awu132)

18. Bullmore E, Sporns O. 2012 The economy of brain network organization. *Nat. Rev. Neurosci.* **13**, 336–349.

19. Vértes PE, Alexander-Bloch AF, Gogtay N, Giedd JN, Rapoport JL, Bullmore ET. 2012 Simple models of human brain functional networks. *Proc. Natl Acad. Sci. USA* **109**, 5868–5873. (doi:10.1073/pnas.1111738109)

20. Sterling P, Laughlin S. 2015 *Principles of neural design*. Cambridge, MA: MIT Press.

21. Dehaene S, Changeux J-P. 2011 Experimental and theoretical approaches to conscious processing. *Neuron* **70**, 200–227. (doi:10.1016/j.neuron.2011.03.018)

22. Kaiser M, Hilgetag CC. 2006 Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Comput. Biol.* **2**, e95. (doi:10.1371/journal.pcbi.0020095)

23. Meunier D, Lambiotte R, Bullmore ET. 2010 Modular and hierarchically modular organization of brain networks. *Front. Neurosci.* **4**, 200. (doi:10.3389/fnins.2010.00200)

24. Rubinov M, Ypma RJ, Watson C, Bullmore ET. 2015 Wiring cost and topological participation of the mouse brain connectome. *Proc. Natl Acad. Sci. USA* **112**, 10 032–10 037. (doi:10.1073/pnas.1420315112)

25. Alexander-Bloch AF, Vertes PE, Stidd R, Lalonde F, Clasen L, Rapoport J, Giedd J, Bullmore ET, Gogtay N. 2013 The anatomical distance of functional connections predicts brain network topology in health and schizophrenia. *Cereb. Cortex* **23**, 127–138. (doi:10.1093/cercor/bhr388)

26. Abdi H. 2010 Partial least squares regression and projection on latent structure regression (PLS

27. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009 GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinform.* **10**, 48. (doi:10.1186/1471-2105-10-48)

28. Eden E, Lipson D, Yogev S, Yakhini Z. 2007 Discovering motifs in ranked lists of DNA sequences. *PLoS Comput. Biol.* **3**, e39. (doi:10.1371/journal.pcbi.0030039)

29. Supek F, Bošnjak M, Škunca N, Šmuc T. 2011 REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLoS ONE* **6**, e21800. (doi:10.1371/journal.pone.0021800)

30. Kundu P, Inati SJ, Evans JW, Luh WM, Bandettini PA. 2012 Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *NeuroImage* **60**, 1759–1770. (doi:10.1016/j.neuroimage.2011.12.028)

31. Achard S, Bassett DS, Meyer-Lindenberg A, Bullmore E. 2008 Fractal connectivity of long-memory networks. *Phys. Rev. E* **77**, 036104. (doi:10.1103/PhysRevE.77.036104)

32. Bullmore E, Long C, Suckling J, Fadili J, Calvert G, Zelaya F, Carpenter TA, Brammer M. 2001 Colored noise and computational inference in neurophysiological (fMRI) time series analysis: resampling methods in time and wavelet domains. *Hum. Brain Mapping* **12**, 61–78. (doi:10.1002/1097-0193(200102)12:2<61::AID-HBM1004>3.0.CO;2-W)

33. Bullmore E, Fadili J, Maxim V, Şendur L, Whitcher B, Suckling J, Brammer, M, Breakspear M. 2004 Wavelets and functional magnetic resonance imaging of the human brain. *NeuroImage* **23**, S234–S249. (doi:10.1016/j.neuroimage.2004.07.012)

34. Romero-Garcia R, Atienza M, Clemmensen LH, Cantero JL. 2012 Effects of network resolution on topological properties of human neocortex. *NeuroImage* **59**, 3522–3532. (doi:10.1016/j.neuroimage.2011.10.086)

35. Desikan RS *et al.* 2006 An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980. (doi:10.1016/j.neuroimage.2006.01.021)

36. Fornito A, Zalesky A, Bullmore ET. 2016 *Fundamentals of brain network analysis*. New York, NY: Academic Press.

37. Alexander-Bloch AF, Gogtay N, Meunier D, Birn R, Clasen L, Lalonde F, Lenroot R, Giedd J, Bullmore ET. 2010 Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia. *Front. Syst. Neurosci.* **4**, 147. (doi:10.3389/fnsys.2010.00147)

38. Newman ME. 2006 Modularity and community structure in networks. *Proc. Natl Acad. Sci. USA* **103**, 8577–8582. (doi:10.1073/pnas.0601602103)

39. Lancichinetti A, Fortunato S. 2012 Consensus clustering in complex networks. *Sci. Rep.* **2**, 336. (doi:10.1038/srep00336)

40. Meunier D, Achard S, Morcom A, Bullmore E. 2009 Age-related changes in modular organization of human brain functional networks. *Neuroimage* **44**, 715–723. (doi:10.1016/j.neuroimage.2008.09.062)

41. Power JD, Schlaggar BL, Lessov-Schlaggar CN, Petersen SE. 2013 Evidence for hubs in human functional brain networks. *Neuron* **79**, 798–813. (doi:10.1016/j.neuron.2013.07.035)

42. Guimera R, Mossa S, Turtschi A, Amaral LAN. 2005 The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proc. Natl Acad. Sci. USA* **102**, 7794–7799. (doi:10.1073/pnas.0407994102)

43. Rubinov M, Sporns O. 2010 Complex network measures of brain connectivity: uses and interpretations. *NeuroImage* **52**, 1059–1069. (doi:10.1016/j.neuroimage.2009.10.003)

44. Goyal MS, Hawrylycz M, Miller JA, Snyder AZ, Raichle ME. 2014 Aerobic glycolysis in the human brain is associated with development and neotenous gene expression. *Cell Metab.* **19**, 49–57. (doi:10.1016/j.cmet.2013.11.020)

45. von Economo CF, Koskinas GN. 2008 *Atlas of cytoarchitectonics of the adult human cerebral cortex* (ed. LC Triarhou). Basel, Switzerland: Karger.

46. Zilles K, Amunts K. 2012 Architecture of the cerebral cortex. In *The human nervous system* (eds J Mai, G Paxinos), pp. 836–895. New York, NY: Academic Press.

47. Krishnan A, Williams LJ, McIntosh AR, Abdi H. 2011 Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage* **56**, 455–475. (doi:10.1016/j.neuroimage.2010.07.034)

48. Ashburner M *et al.* 2000 Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29. (doi:10.1038/75556)

49. van den Heuvel MP, Scholtens LH, Barrett LF, Hilgetag CC, de Reus MA. 2015 Bridging cytoarchitectonics and connectomics in human cerebral cortex. *J. Neurosci.* **35**, 13 943–13 948. (doi:10.1523/JNEUROSCI.2630-15.2015)

50. van den Heuvel MP, Bullmore ET, Sporns O. 2016 Comparative connectomics. *Trends Cogn. Sci.* **20**, 345–361. (doi:10.1016/j.tics.2016.03.001)

51. van den Heuvel MP, Sporns O. 2011 Rich club organization of the human connectome. *J. Neurosci.* **31**, 15 775–15 786. (doi:10.1523/JNEUROSCI.3539-11.2011)

52. van den Heuvel MP, Kahn RS, Goni J, Sporns O. 2012 High-cost, high-capacity backbone for global brain communication. *Proc. Natl Acad. Sci. USA* **109**, 11 372–11 377. (doi:10.1073/pnas.1203593109)

53. White JG, Southgate E, Thomson JN, Brenner S. 1986 The structure of the nervous system of the nematode *Caenorhabditis elegans*: the mind of a worm. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340. (doi:10.1098/rstb.1986.0056)

54. Towlson EK, Vértes PE, Ahnert SE, Schafer WR, Bullmore ET. 2013 The rich club of the *C. elegans* neuronal connectome. *J. Neurosci.* **33**, 6380–6387. (doi:10.1523/JNEUROSCI.3784-12.2013)

55. van den Heuvel MP, Sporns O, Collin G, Scheewe T, Mandl RCW, Cahn W, Goñi J, Hulshoff Pol HE, Kahn RS. 2013 Abnormal rich club organization and functional brain dynamics in schizophrenia. *JAMA Psychiatry* **70**, 783–792. (doi:10.1001/jamapsychiatry.2013.1328)

56. Harriger L, Van Den Heuvel MP, Sporns O. 2012 Rich club organization of macaque cerebral cortex and its role in network communication. *PLoS ONE* **7**, e46497. (doi:10.1371/journal.pone.0046497)

57. Vértes PE, Alexander-Bloch A, Bullmore ET. 2014 Generative models of rich clubs in Hebbian neuronal networks and large-scale human brain networks. *Phil. Trans. R. Soc. B* **369**, 20130531. (doi:10.1098/rstb.2013.0531)

58. Schroeter MS, Charlesworth P, Kitzbichler MG, Paulsen O, Bullmore ET. 2015 Emergence of rich-club topology and coordinated dynamics in development of hippocampal functional networks in vitro. *J. Neurosci.* **35**, 5459–5470. (doi:10.1523/JNEUROSCI.4259-14.2015)

59. Tomasi D, Wang GJ, Volkow ND. 2013 Energetic cost of brain functional connectivity. *Proc. Natl Acad. Sci. USA* **110**, 13 642–13 647. (doi:10.1073/pnas.1303346110)

60. Liang X, Zou Q, He Y, Yang Y. 2013 Coupling of functional connectivity and regional cerebral blood flow reveals a physiological basis for network hubs of the human brain. *Proc. Natl Acad. Sci. USA* **110**, 1929–1934. (doi:10.1073/pnas.1214900110)

61. Attwell D, Laughlin SB. 2001 An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* **21**, 1133–1145. (doi:10.1097/00004647-200110000-00001)

62. Harris JJ, Jolivet R, Attwell D. 2012 Synaptic energy use and supply. *Neuron* **75**, 762–777. (doi:10.1016/j.neuron.2012.08.019)

63. Fulcher BD, Fornito A. 2016 Transcriptional signature of hub connectivity in the mouse connectome. *Proc. Natl Acad. Sci. USA* **113**, 1435–1440. (doi:10.1073/pnas.1513302113)

64. French L, Tan PPC, Pavlidis P. 2011 Large-scale analysis of gene expression and connectivity in the rodent brain: insights through data integration. *Front. Neuroinform.* **5**, 12. (doi:10.3389/fninf.2011.00012)

65. French L, Pavlidis P. 2011 Relationships between gene expression and brain wiring in the adult rodent brain. *PLoS Comp. Biol.* **7**, e1001049. (doi:10.1371/journal.pcbi.1001049)

66. Fakhry A, Ji S. 2015 High-resolution prediction of mouse brain connectivity using gene expression patterns. *Methods* **73**, 71–78. (doi:10.1016/j.ymeth.2014.07.011)

67. Ji S, Fakhry A, Deng H. 2014 Integrative analysis of the connectivity and gene expression atlases in the mouse brain. *NeuroImage* **84**, 245–253. (doi:10.1016/j.neuroimage.2013.08.049)

68. Fakhry A, Zeng T, Peng H, Ji S. 2015 Global analysis of gene expression and projection target correlations in the mouse brain. *Brain Inform.* **2**, 107–117. (doi:10.1007/s40708-015-0014-2)

Wiley Interdiscip. Rev. Comput. Stat **2**, 97–106. (doi:10.1002/wics.51)