A transcriptional signature of hub connectivity in the mouse connectome

Ben D. Fulcher[a,1] and Alex Fornito[a]

[a]Brain and Mental Health Laboratory, Monash Institute of Cognitive and Clinical Neurosciences, Monash University, Clayton, 3168 VIC, Australia

Edited by Danielle S. Bassett, University of Pennsylvania, Philadelphia, PA, and accepted by the Editorial Board December 11, 2015 (received for review July 11, 2015)

Connectivity is not distributed evenly throughout the brain. Instead, it is concentrated on a small number of highly connected neural elements that act as network hubs. Across different species and measurement scales, these hubs show dense interconnectivity, forming a core or "rich club" that integrates information across anatomically distributed neural systems. Here, we show that projections between connectivity hubs of the mouse brain are both central (i.e., they play an important role in neural communication) and costly (i.e., they extend over long anatomical distances) aspects of network organization that carry a distinctive genetic signature. Analyzing the neuronal connectivity of 213 brain regions and the transcriptional coupling, across 17,642 genes, between each pair of regions, we find that coupling is highest for pairs of connected hubs, intermediate for links between hubs and nonhubs, and lowest for connected pairs of nonhubs. The high transcriptional coupling associated with hub connectivity is driven by genes regulating the oxidative synthesis and metabolism of ATP—the primary energetic currency of neuronal communication. This genetic signature contrasts that identified for neuronal connectivity in general, which is driven by genes regulating neuronal, synaptic, and axonal structure and function. Our findings establish a direct link between molecular function and the large-scale topology of neuronal connectivity, showing that brain hubs display a tight coordination of gene expression, often over long anatomical distances, that is intimately related to the metabolic requirements of these highly active network elements.

connectome | complex networks | hub | rich club | metabolism

Certain neural elements possess an unusually high degree of connectivity, designating them as putative network hubs (1). Analyses of microscale, mesoscale, and macroscale connectomes of multiple species, constructed using a variety of methods, indicate that these hubs are strongly interconnected with each other, forming a so-called "rich club" of connectivity that mediates a large fraction of communication traffic in the brain and supports the efficient integration of otherwise segregated neural systems (2–8).

Hub connectivity is functionally advantageous, but it is also costly. Hub regions make more connections with other areas, and these connections often extend over long anatomical distances, thus requiring greater physical space, cellular material, and metabolic resources (3, 9). Accordingly, human neuroimaging studies have indicated that topologically central hub regions have a higher energetic demand than other brain areas (9–12), which may render them particularly vulnerable to the effects of damage or disease (10, 13). This hypothesis is supported by evidence that pathology in a broad range of disorders preferentially accumulates within highly connected brain regions (14).

Hub connectivity is thus a topologically central and costly aspect of brain network organization that is conserved across species and spatial scales. This conservation suggests that hub connectivity may be under tight genetic control. Growing evidence indicates that gene expression affects neuronal connectivity, with studies of worm, rat, and mouse nervous systems showing that the transcriptional profile of an individual neuron or neuronal population can predict its connectivity to other areas with greater than chance accuracy (15–19). Brain regions with similar transcriptional profiles display similar connectivity profiles (20, 21), and gene expression profiles are more correlated between pairs of structurally connected brain regions in the mouse/rat (20) and within functionally coupled networks of the human cortex (22). Functional neuroimaging of human twins indicates that the topological properties of hub connections are strongly heritable (23), but it is not known whether the topologically distinctive and functionally important connections between hub regions are associated with a unique transcriptional signature. Characterizing this relationship is critical for understanding the molecular basis of topological specialization in brain networks.

In this work, we show that the topologically central and costly connections involving hubs of the mouse brain are associated with a distinct transcriptional signature. Transcriptional coupling is greatest for pairs of connected hubs, intermediate for connections between hubs and nonhubs, and lowest for connected pairs of nonhubs, a trend that mirrors the signaling load that these connections are likely to carry (3, 4). The highly correlated gene expression profiles of connected hubs are not driven by the coupling of genes associated with structural connectivity in general (which we show are involved in neuronal connectivity and communication) but are driven by genes regulating oxidative metabolism. We thus identify a close interplay between gene transcription and large-scale brain network architecture and show that the primary genetic distinction between different classes of neuronal connections is intimately related to the metabolic demand of the regions that they interconnect.

Topological Centrality and Cost of Hub Connectivity

We first describe the topological properties of the mouse connectome from work by Oh et al. (24), represented here as a binary,

Significance

Some brain regions are highly connected with other areas, designating them as network hubs. These hubs are also heavily interconnected with each other, forming a dense core that integrates information across different neural systems. Here, we show that the functionally important projections linking hub areas of the mouse brain have a distinct genetic signature that is characterized by the tightly coupled expression of genes regulating the synthesis and metabolism of ATP, the primary energy source for neural activity. Our findings establish a direct link between molecular function and the large-scale organization of neuronal connectivity and suggest that coordinated gene expression between hub areas is closely related to the metabolic demands of these highly active and functionally important regions.

Author contributions: B.D.F. and A.F. designed research; B.D.F. performed research; B.D.F. analyzed data; and B.D.F. and A.F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. D.S.B. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: ben.fulcher@monash.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1513302113/-/DCSupplemental.

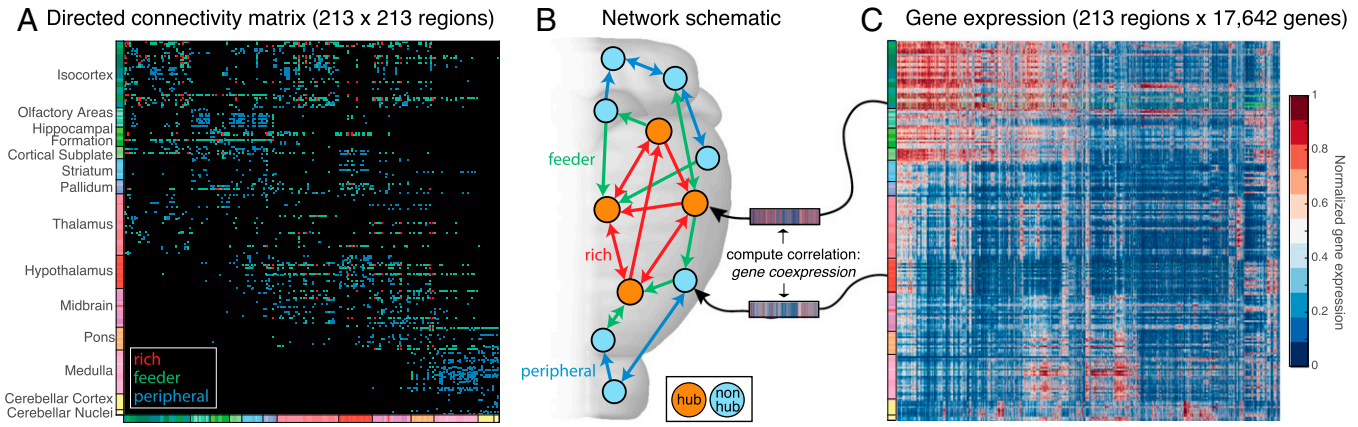

Fig. 1. Mapping the transcriptional signature of large-scale brain network topology. (*A*) Defining highly connected hub regions with connectivity degree $k > 44$, all neuronal connections between each of 213 brain regions were labeled as rich (hub → hub; red), feeder (hub → nonhub or nonhub → hub; green), or peripheral (nonhub → nonhub; blue). (*B*) Network schematic illustrating the different connection types in the mouse brain. (*C*) Normalized expression levels of 17,642 genes (columns) measured in each brain region (rows) visualized here using color from low (blue) to high (red) are used to compute the correlation in expression profiles or gene coexpression for each pair of brain regions. Missing data are shown as green, and columns of the matrix have been reordered using hierarchical clustering to place genes with correlated expression patterns close to one another.

directed adjacency matrix that encodes 3,063 anatomical connections between 213 brain regions in the right hemisphere (Fig. 1*A*). The total number of connections involving a given brain region is called its degree, k. The distribution of k across all regions of the mouse connectome, plotted in Fig. 2*A*, reveals an extended tail of highly connected hub regions. For each value of k, we quantified the tendency of nodes with degree $> k$ to preferentially connect to each other, forming a rich club, using the normalized rich club coefficient, $\Phi_{\text{norm}}(k)$. Values of $\Phi_{\text{norm}}(k) > 1$ indicate rich club organization of the network (7, 25). As shown in Fig. 2*B*, the mouse connectome displays rich club organization across the contiguous range $42 \le k \le 54$ ($P < 0.05$, shaded gray area in Fig. 2*B*), reflecting dense connectivity between these high-degree hub regions. This range of k is referred to as the "topological rich club regime" throughout this work.

Putative hub regions are distributed broadly across anatomical brain divisions in the topological rich club regime (Fig. S1*B*). For example, hubs with $k > 42$ are present in 9 of 13 broad anatomical divisions of the Allen Mouse Brain Atlas (24, 26). Relative to other types of network connections, connections between hubs show a greater mean connection distance (Fig. 2*B*), an increased proportion of reciprocal connections (Fig. S1*D*), and higher average connectivity weight (Fig. S1*E*). The high density, reciprocity, connection weight, and connection distance of hub–hub connections characterize the high-topological wiring cost of these links (3, 9, 10). These findings counter the general trend across the brain, where the probability of a connection between two brain areas decays exponentially with their physical separation, as does the probability that a connection will be reciprocal (Fig. S2). Hub–hub connections also play a topologically central role in network communication, as measured by their edge betweenness centrality and network communicability (Fig. S1*F*), suggesting that they are well-positioned to mediate a large proportion of signal traffic in the mouse brain. All of the above-mentioned properties of hub–hub connections display a similar increasing trend with k and a significant increase relative to all other connections across the topological rich club regime ($P < 0.05$). Thus, hubs of the mouse connectome are distributed broadly across anatomical divisions and show a rich club organization characterized by a high wiring cost and topological centrality, consistent with prior observations in other diverse species (2–7).

Gene Coexpression and Neuronal Connectivity

We next investigated how the connectivity of pairs of regions of the mouse brain relate to their transcriptional coupling, as illustrated in Fig. 1. Transcriptional data for 17,642 genes were obtained from the Allen Mouse Brain Atlas (26) and normalized across the brain for each gene, yielding an expression profile for each brain region (Fig. 1*C*, rows). To compare different classes of pairwise connections, we examined patterns of gene coexpression (transcriptional coupling) measured for each pair of brain regions as the Pearson correlation of their expression profiles. Gene coexpression values were corrected for strong spatial correlations in the data (Fig. S3), ensuring that our results reflect robust effects of connectivity and connection topology that cannot be explained simply by the spatial proximity of different pairs of brain regions (*Materials and Methods*).

We investigated the relationship between gene coexpression and neuronal connectivity by comparing three different classes of brain region pairs, i and j (excluding self-connections): (*i*) reciprocally connected ($i \leftrightarrow j$), (*ii*) unidirectionally connected ($i \rightarrow j$ or $j \rightarrow i$, but not both), and (*iii*) unconnected. Spatially corrected gene coexpression is greatest in reciprocally connected pairs of brain regions (mean ± SD = 0.10 ± 0.17) followed by unidirectionally connected pairs (0.06 ± 0.16) and lowest in unconnected

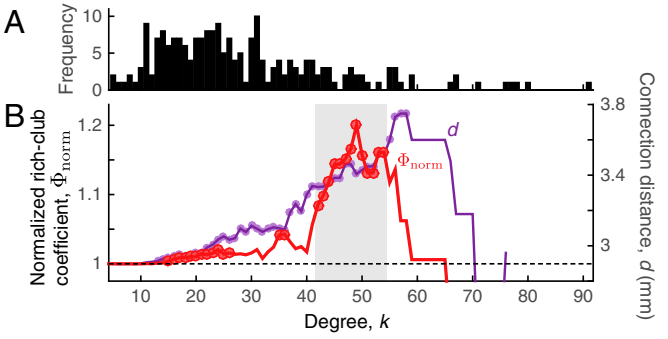

Fig. 2. The mouse connectome contains a costly and topologically central rich club of densely interconnected hub regions. (*A*) Degree distribution of the mouse connectome. (*B*) Normalized rich club coefficient, Φ_{norm} (red), and mean connection distance, d, of hub–hub links (purple) as a function of the degree, k, at which hubs (regions with degree $> k$) are defined. Red circles indicate values of Φ_{norm} that are significantly higher than an ensemble of 10,000 null networks (permutation test; $P < 0.05$); purple circles indicate where the mean connection distance of hub–hub links is significantly increased relative to all other network links (one-sided Welch's t test; $P < 0.05$). The topological rich regime ($42 \le k \le 54$) is shaded gray.

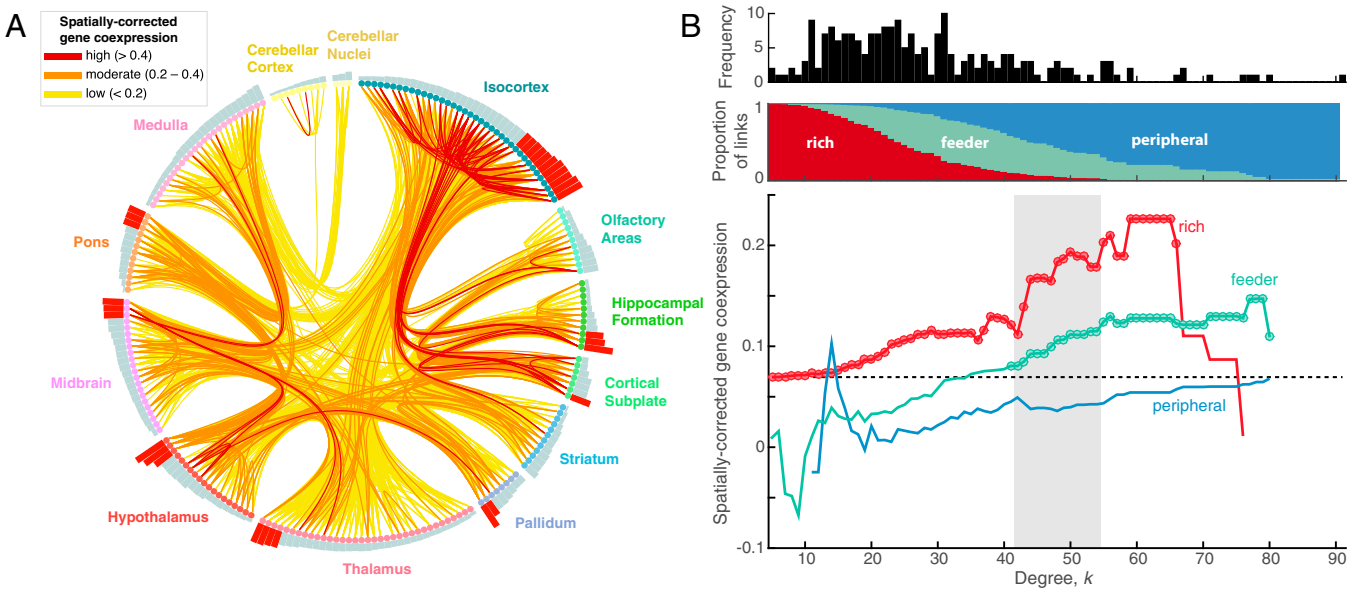

Fig. 3. Gene coexpression is elevated for connections involving brain network hubs. (*A*) Connectogram showing (spatially corrected) gene coexpression values across the mouse connectome. All neuronal connections (lines) between brain regions (circles) are colored according to the gene coexpression of the regions that they connect. Brain regions are organized by anatomical division and sorted by degree (shown as bars), with bars colored bright red for hubs ($k > 44$). A larger version of this connectogram with all regions labeled is in Fig. S4. (*B*, *Top*) Degree distribution. (*B*, *Middle*) Proportion of links classified as rich, feeder, and peripheral, where hub nodes have degree $> k$. (*B*, *Bottom*) Mean (spatially corrected) gene coexpression for rich, feeder, and peripheral connections as a function of k, with the mean across all network links shown as a dashed black line and the topological rich club regime shaded. Circles indicate a statistically significant increase in gene coexpression in a given link type relative to the rest of the network (one-sided Welch's t test; $P < 0.05$).

pairs (-0.01 ± 0.16; all differences are statistically significant; Welch's t test, $P < 10^{-6}$) (Fig. S3*C*).

To investigate which functional groups of genes contributed to this trend in transcriptional coupling, we developed a measure that quantifies the contribution of each gene to the overall correlation in expression levels between pairs of brain regions, referred to here as the gene coexpression contribution (GCC) score. These GCC scores were then used to perform a gene function analysis (*Materials and Methods*). At a false discovery rate of 0.05, 31 distinct functional groups of genes [using Gene Ontology (GO) annotations for biological processes (27)] show a significantly increased contribution to gene coexpression for connected pairs of brain regions relative to unconnected pairs ($P < 0.05$) (Table S1). The majority of these GO categories are related to neuronal connectivity and communication, including genes regulating synapse structure, function, and plasticity; neuronal membrane potentials; neurotransmitter signaling; dendritic spine morphogenesis; and axonogenesis. Similar categories were also selected when cellular components were included in the analysis (Table S2). Other categories are related to metabolism, such as those involved in the electron transport chain and mitochondrial function, suggesting an increased energy demand for connected pairs of brain regions over unconnected pairs, likely reflecting the metabolic cost of neuronal communication (28). Similar GO categories related to neuronal communication and connectivity were obtained when comparing separately (*i*) reciprocal vs. unconnected pairs and (*ii*) unidirectional vs. unconnected pairs, indicating a robust transcriptional signature of structural connectivity in the mouse brain that varies quantitatively (rather than qualitatively) as a function of connection presence and reciprocity.

Gene Coexpression and Hub Connectivity

Having characterized a distinctive transcriptional signature of neuronal connectivity in the mouse brain, we next investigated whether gene coexpression might also vary as a function of connection type, focusing particularly on different classes of connections involving hubs (Fig. 1*B*). At each k, we labeled each brain region as either a hub (nodes with degree $> k$) or a nonhub (otherwise), and then labeled each connection as rich (hub → hub), feeder (nonhub → hub or hub → nonhub), or peripheral (nonhub → nonhub) (3). The anatomical distribution of hubs, interregional connections, and gene coexpression values is shown in Fig. 3*A*.

Across the topological rich club regime, mean gene coexpression is significantly increased for connections involving hubs (i.e., rich and feeder connections) and is greatest for rich connections (Fig. 3*B*). Mean gene coexpression of rich connections increases sharply at the start of the topological rich club regime ($k = 42$) and continues to increase with k, indicating that transcriptional coupling is strongest for pairs of the most highly connected hubs. Across the topological rich club regime, gene coexpression is significantly greater in (*i*) rich links than feeder links and (*ii*) feeder links than peripheral links (Welch's t test; all $P < 0.01$). For example, at $k = 42$, (spatially corrected) gene coexpression is greatest for rich links (mean ± SD = 0.11 ± 0.17) followed by feeder links (0.08 ± 0.17) and peripheral links (0.05 ± 0.16). This same increase in gene coexpression for rich connections was reproduced using a range of different data processing methods [including variations in connectome density (Fig. S5) and spatial correction procedures (Fig. S6)], highlighting the robustness of this result.

To determine whether specific functional groups of genes drive this correlated gene expression signature of hub connectivity, we used our method of assigning GCC values to genes to compare connections involving hubs with peripheral connections between nonhubs (*Materials and Methods*). Hubs were defined as brain regions with $k > 44$, corresponding to 1 SD above the mean of the degree distribution (1, 7) (Table S3). The five biological process GO categories that show a significant increase in gene coexpression in rich and feeder connections over peripheral connections ($P < 0.05$) fall into two parent categories related to oxidative energy metabolism: (*i*) hydrogen ion transmembrane transport and (*ii*) citrate metabolic process (Table 1). When GO annotations for cellular components were also included in the

NEUROSCIENCE

Table 1. Genes regulating oxidative metabolism are implicated in hub connectivity

GO category	*P* value
Hydrogen ion transmembrane transport	0.04
Energy-coupled proton transmembrane transport against electrochemical gradient	0.0097
ATP hydrolysis-coupled proton transport	0.0097
Citrate metabolic process	0.045
Tricarboxylic acid cycle	0.014

GO annotated biological processes with significantly increased gene coexpression [measured using GCC scores; false discovery rate (FDR) corrected $P < 0.05$] in connections involving hubs compared with connections between nonhubs along with FDR corrected *P* values. Categories are organized into two parent categories in the GO hierarchy that contain nonoverlapping sets of genes (Fig. S7*B*).

analysis, in total, 25 functional groups of genes were selected ($P < 0.05$), including mitochondrial respiration, cellular respiration, mitochondrial membrane, and proton-transporting ATPase complex (Table S4). Importantly, similar functional groups of genes drive the increased transcriptional coupling of both (*i*) rich links compared with peripheral links and (*ii*) feeder links compared with peripheral links (Table S5), pointing to a robust and consistent transcriptional signature of connections involving hubs.

The 70 unique genes annotated to the metabolic processes implicated in hub connectivity (Table 1) show a strikingly selective increase in coexpression for connections involving hubs across the topological rich club regime (Fig. 4), being highest for rich connections followed by feeder connections and then peripheral connections. This result was reproduced when analyzing each GO category in Table 1 separately, despite the two parent categories containing nonoverlapping sets of genes (Fig. S7). Across these 70 genes, mean regional gene expression is also increased in hub regions over nonhub regions (Welch's *t* test; $P < 0.05$) (Fig. S8*C*). Thus, both the regional expression in hub regions and the interregional coexpression for pairs of brain regions involving hubs are increased for these metabolic genes. The result is robust, with similar results reproduced at less conservative significance thresholds and when cellular component annotations were included (Fig. S9).

Discussion

Connectivity is the substrate for neuronal communication. Here we show that connections between hub regions are topologically central and costly features of the mouse connectome, and that these connections are characterized by tightly coupled expression of genes regulating oxidative metabolism. This distinct transcriptional signature of hub connectivity is different to that of neuronal connectivity in general, which predominantly implicates genes involved in synaptic communication and plasticity, the regulation of membrane potentials, and neurite development and morphology. Our findings point to a molecular basis for the topological specialization of distinct classes of interregional connections in mesoscale brain networks and indicate that connections between hub regions in particular can be distinguished by the metabolic requirements of integrating large amounts of neural information over long distances.

Gene Coexpression and Neuronal Connectivity. The idea that connected neural elements should show coupled molecular function was suggested in the work of Ramón y Cajal (29) and later elaborated in Sperry's chemoaffinity hypothesis for how developing neural connections find their targets (30). Developmental processes involved in establishing and maintaining neuronal connectivity, such as neurite outgrowth and guidance, synapse formation, and synaptic transmission, are all under tight transcriptional regulation (31). We should therefore expect that genes involved in these processes will show coordinated expression in connected pairs of brain regions, as suggested previously using a combination of mouse gene expression and rat connectivity data (20).

Our findings, obtained using a novel methodology to combine connectivity and expression data, support this view. For both reciprocal and unidirectional connections, correlated gene expression is driven by the same types of functional gene groups, pointing to a uniform transcriptional profile of connectivity that increases with connection reciprocity. Similar functional categories of genes related to the development of neurons, neurites, and synapses, as well as the regulation of neuronal activity and synaptic plasticity, contribute to predicting the presence of a connection between neurons in *Caenorhabditis elegans* (15, 16) and larger-scale neuronal populations of the rat (17) and mouse brains (18, 19). The consistency of these findings across species, datasets, and analysis methods points to a robust transcriptional signature of neuronal connectivity characterized by the coordinated expression of genes involved in the development and ongoing function of neuronal networks.

Benefits and Cost of Hub Connectivity. The pressure to minimize network wiring costs can account for many diverse aspects of brain organization, suggesting that wiring cost minimization is an organizational imperative for brain networks (32). However, these wiring costs are not absolutely minimized, with some axons extending over long distances to interconnect spatially disparate brain areas (10). Our analysis indicates that these connections are often interposed between highly connected hub regions, which are dense, strong, reciprocal, and show rich club organization. These properties counter the general trend for interregional connectivity in the mouse brain, where the probability of a connection and the probability that a connection will be reciprocal both decay exponentially as a function of spatial separation (Fig. S2). Connections linking hub areas thus serve a topologically unique role, acting as a central but costly backbone that supports the integration of anatomically distributed and functionally segregated neural systems (3, 4, 8).

The high wiring cost of hub connectivity is coupled with an increased demand for metabolic resources (9–12). Our results

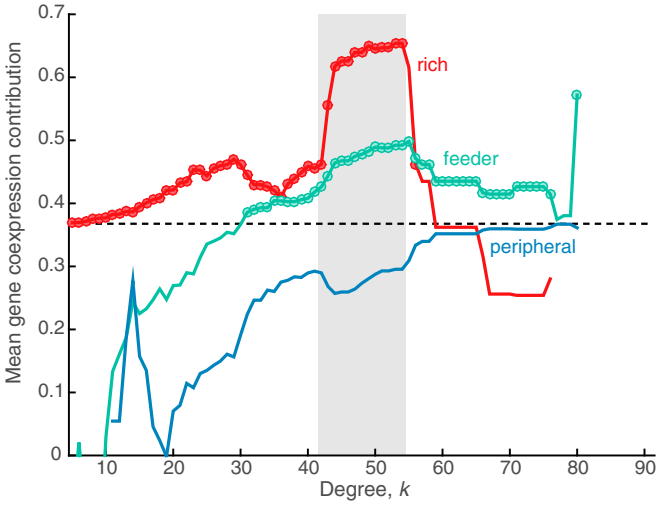

Fig. 4. Transcriptional coupling of metabolic genes is selectively increased for connections involving hubs. Mean GCC scores across 70 unique genes implicated in the transcriptional signature of hub connectivity (Table 1) in rich, feeder, and peripheral links as a function of the degree *k* at which hubs are defined. Circles indicate a significant enrichment in a given link type over all other links (one-sided Welch's *t* test; $P < 0.05$). There is a strikingly specific increase in the coexpression of these genes for rich connections in the topological rich club regime (shaded gray).

indicate that this demand defines the transcriptional signature of hub connectivity. Specifically, we report that gene coexpression is highest for rich connections (pairs of connected hubs) followed by feeder connections (connected hubs and nonhubs), with both rich and feeder connections showing significantly increased gene coexpression relative to peripheral connections (pairs of connected nonhubs). This result is robust to variations in processing and analysis procedures (Figs. S5 and S6) and is striking when one considers the broad anatomical distribution and functional diversity of hubs (Fig. 3*A*). Indeed, the strong transcriptional coupling of connections involving hubs persists despite spanning distinct neural systems and extending over long anatomical distances. Importantly, the same types of functional gene groups involved in oxidative metabolism drive the increase in gene coexpression for both rich and feeder connections relative to peripheral connections, indicating that the transcriptional distinction between rich and feeder connections is quantitative rather than qualitative. The increasing gradient of gene coexpression from peripheral to feeder to rich links follows the expected signal traffic that these different connection classes are thought to mediate, as indicated by both topological analysis (3) and computational models of interregional communication (4). This convergence suggests that the transcriptional signature of different types of interregional connections may be determined by the metabolic resources required to meet their differential signaling load, consistent with evidence that the energetic requirements of neuronal signaling scale with action potential frequency (28).

Genes driving the correlated gene expression signature of hub connectivity are involved in the synthesis and breakdown of ATP. Increased transcriptional coupling of these genes was highly specific to rich and feeder connections across the topological rich-club regime of the mouse connectome. ATP is the energetic currency of neuronal signaling (28) and is predominantly supplied by oxidative phosphorylation, with ∼10–12% of energy supplied by nonoxidative metabolism in the form of aerobic glycolysis (33). Human functional and metabolic imaging has shown that brain regions with high-degree and topological centrality consume more glucose (12) and have higher regional blood flow (11) and glycolytic activity (9, 10) than other areas, suggesting a role for both oxidative and nonoxidative pathways in meeting the energetic requirements of hub areas.

Our findings emphasize the role of oxidative phosphorylation in supporting high-cost communication between hub areas. Functional groups of genes showing elevated coexpression for rich and feeder connections (Table 1) include nine genes encoding different subunits of the mitochondrial *H^+-ATP synthase subunit 5*, which catalyzes ATP synthesis by oxidative phosphorylation, as well as four genes encoding subunits of cytochrome oxidase c, which is the terminal enzyme in the mitochondrial electron transport chain and which has activity levels that are tightly coupled with neuronal signaling (34). The categories also include a cluster of genes coding proteins involved in citrate metabolism (*Sdh*, *Mdh*, *Idh*, and *Pdh*). Notably, *Pdh* acts as a molecular bridge between glycolysis and oxidative phosphorylation by catalyzing the conversion of pyruvate to acetyl-CoA, further underlining the role of oxidative metabolism in the transcriptional signature of hub connectivity.

Our analysis of the adult mouse brain reflects the functional requirements of supporting neuronal connectivity in a mature neural system. Although many aspects of gene expression in the brain show a developmentally persistent profile (35), it is unclear whether the same transcriptional signature of hub connectivity would be apparent throughout development. Although rich club connectivity seems to be established early in development (2, 36), it also undergoes significant remodeling later in life (37). Interestingly, recent evidence indicates that aerobic glycolysis plays a prominent role in biosynthesis and growth and that it accounts for a larger fraction of the brain's energetic needs earlier in development, peaking in early childhood when levels of synaptic development are highest (35). This work also found that areas of the adult human brain with high levels of glycolytic activity show increased expression of genes regulating synapse formation and growth, whereas brain regions with high glucose metabolism show elevated expression of genes regulating mitochondria and synaptic transmission (35). Collectively, these findings suggest that the development and remodeling of synaptic networks is associated with the expression of genes regulating aerobic glycolysis. On the other hand, signaling across established or mature networks, particularly along links involving hub nodes, may be supported by the coordinated expression of genes regulating oxidative phosphorylation.

Implications for Disease. Many complex diseases of the brain can be construed as disorders of neuronal connectivity, and the high metabolic demand of hub regions may render these areas particularly vulnerable to the effects of injury or disease (10, 13, 14, 38). It is well-known that metabolic abnormalities (mitochondrial dysfunction in particular) play a key role in the pathophysiology of many neurological disorders, including Alzheimer's and Parkinson's diseases (38–41), schizophrenia (42), and others. Although the exact causes of these disorders are no doubt complex, our results point to a close interplay between the topological organization of hub connections and the transcription of metabolic genes. This link suggests that a closer investigation of how brain network topology relates to the energetic requirements of neuronal signaling may help elucidate the pathogenesis of these disorders.

Materials and Methods

A summary of our analysis methods is provided here, with additional detail provided in *SI Materials and Methods*. Mouse brain connectivity data were obtained from the Allen Mouse Brain Connectivity Atlas (24), and expression data were obtained from the Allen Mouse Brain Atlas (26). Because the magnitudes of in situ hybridization-measured expression levels are not directly comparable across genes (43), they were normalized across the brain for each gene using a scaled sigmoidal transformation. This choice of normalization did not drive our qualitative results, which were reproduced using a range of normalizing transformations (*SI Materials and Methods*). The gene coexpression value for a pair of brain regions is defined as the Pearson correlation between the normalized expression levels across all genes. Gene coexpression values display strong spatial correlations that decay exponentially with separation distance (Fig. S3*B*). We corrected for this exponential trend, analyzing spatially corrected gene coexpression data as the residuals of an exponential fit to the data. This correction allowed us to analyze patterns of gene coexpression beyond what would be expected purely based on the spatial proximity of brain regions. The contribution of each individual gene to the spatially corrected gene coexpression value for each interregion pair was measured as a GCC score using the definition of the Pearson correlation. Each gene was assigned a *t* statistic measuring the increase in GCC values (and thus, a more correlated pattern of gene expression) in one class of interregion pairs over another. Gene function analysis was performed as a gene score resampling analysis on these *t* statistics using ermineJ (44).

ACKNOWLEDGMENTS. The authors thank Dr. Beth Johnson, Dr. N. S. Jones, and Dr. N. Rogasch for helpful discussions and feedback on the manuscript. B.D.F. and A.F. are supported by National Health and Medical Research Council Grants 1089718, 1050504, and 1066779. A.F. is also supported by Australian Research Council Grant FT130100589.

1. Sporns O, Honey CJ, Kötter R (2007) Identification and classification of hubs in brain networks. *PLoS One* 2(10):e1049.
2. Towlson EK, Vértes PE, Ahnert SE, Schafer WR, Bullmore ET (2013) The rich club of the C. elegans neuronal connectome. *J Neurosci* 33(15):6380–6387.
3. van den Heuvel MP, Kahn RS, Goñi J, Sporns O (2012) High-cost, high-capacity backbone for global brain communication. *Proc Natl Acad Sci USA* 109(28):11372–11377.
4. Mišić B, Sporns O, McIntosh AR (2014) Communication efficiency and congestion of signal traffic in large-scale brain networks. *PLoS Comput Biol* 10(1):e1003427.
5. 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(9):e46497.
6. Stafford JM, et al. (2014) Large-scale topology and the default mode network in the mouse connectome. *Proc Natl Acad Sci USA* 111(52):18745–18750.

7. van den Heuvel MP, Sporns O (2011) Rich-club organization of the human connectome. *J Neurosci* 31(44):15775–15786.
8. van den Heuvel MP, Sporns O (2013) An anatomical substrate for integration among functional networks in human cortex. *J Neurosci* 33(36):14489–14500.
9. Collin G, Sporns O, Mandl RCW, van den Heuvel MP (2014) Structural and functional aspects relating to cost and benefit of rich club organization in the human cerebral cortex. *Cereb Cortex* 24(9):2258–2267.
10. Bullmore E, Sporns O (2012) The economy of brain network organization. *Nat Rev Neurosci* 13(5):336–349.
11. 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(5):1929–1934.
12. Tomasi D, Wang GJ, Volkow ND (2013) Energetic cost of brain functional connectivity. *Proc Natl Acad Sci USA* 110(33):13642–13647.
13. Fornito A, Zalesky A, Breakspear M (2015) The connectomics of brain disorders. *Nat Rev Neurosci* 16(3):159–172.
14. Crossley NA, et al. (2014) The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain* 137(8):2382–2395.
15. Kaufman A, Dror G, Meilijson I, Ruppin E (2006) Gene expression of Caenorhabditis elegans neurons carries information on their synaptic connectivity. *PLoS Comput Biol* 2(12):e167.
16. Baruch L, Itzkovitz S, Golan-Mashiach M, Shapiro E, Segal E (2008) Using expression profiles of Caenorhabditis elegans neurons to identify genes that mediate synaptic connectivity. *PLoS Comput Biol* 4(7):e1000120.
17. Wolf L, Goldberg C, Manor N, Sharan R, Ruppin E (2011) Gene expression in the rodent brain is associated with its regional connectivity. *PLoS Comput Biol* 7(5): e1002040.
18. Ji S, Fakhry A, Deng H (2014) Integrative analysis of the connectivity and gene expression atlases in the mouse brain. *Neuroimage* 84:245–253.
19. Fakhry A, Ji S (2015) High-resolution prediction of mouse brain connectivity using gene expression patterns. *Methods* 73:71–78.
20. French L, Pavlidis P (2011) Relationships between gene expression and brain wiring in the adult rodent brain. *PLoS Comput Biol* 7(1):e1001049.
21. Fakhry A, Zeng T, Peng H, Ji S (2015) Global analysis of gene expression and projection target correlations in the mouse brain. *Brain Informatics* 2(2):107–117.
22. Richiardi J, et al. (2015) Correlated gene expression supports synchronous activity in brain networks. *Science* 348(6240):1241–1244.
23. Fornito A, et al. (2011) Genetic influences on cost-efficient organization of human cortical functional networks. *J Neurosci* 31(9):3261–3270.
24. Oh SW, et al. (2014) A mesoscale connectome of the mouse brain. *Nature* 508(7495): 207–214.
25. Colizza V, Flammini A, Serrano M, Vespignani A (2006) Detecting rich-club ordering in complex networks. *Nat Phys* 2(2):110–115.
26. Lein ES, et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445(7124):168–176.
27. Ashburner M, et al.; The Gene Ontology Consortium (2000) Gene ontology: Tool for the unification of biology. *Nat Genet* 25(1):25–29.
28. Attwell D, Laughlin SB (2001) An energy budget for signaling in the grey matter of the brain. *J Cereb Blood Flow Metab* 21(10):1133–1145.
29. Ramón y Cajal S (1960) *Studies on Vertebrate Neurogenesis* (Thomas, Springfield, IL).
30. Sperry RW (1963) Chemoaffinity in the orderly growth of nerve fiber patterns and connections. *Proc Natl Acad Sci USA* 50(4):703–710.
31. Polleux F, Ince-Dunn G, Ghosh A (2007) Transcriptional regulation of vertebrate axon guidance and synapse formation. *Nat Rev Neurosci* 8(5):331–340.
32. Chklovskii DB, Koulakov AA (2004) Maps in the brain: What can we learn from them? *Annu Rev Neurosci* 27(1):369–392.
33. Raichle ME, Mintun MA (2006) Brain work and brain imaging. *Annu Rev Neurosci* 29(1):449–476.
34. Wong-Riley MTT (1989) Cytochrome oxidase: An endogenous metabolic marker for neuronal activity. *Trends Neurosci* 12(3):94–101.
35. 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(1):49–57.
36. Ball G, et al. (2014) Rich-club organization of the newborn human brain. *Proc Natl Acad Sci USA* 111(20):7456–7461.
37. Baker STE, et al. (2015) Developmental changes in brain network hub connectivity in late adolescence. *J Neurosci* 35(24):9078–9087.
38. Buckner RL, et al. (2009) Cortical hubs revealed by intrinsic functional connectivity: Mapping, assessment of stability, and relation to Alzheimer's disease. *J Neurosci* 29(6):1860–1873.
39. Vlassenko AG, et al. (2010) Spatial correlation between brain aerobic glycolysis and amyloid-β (Aβ) deposition. *Proc Natl Acad Sci USA* 107(41):17763–17767.
40. Kapogiannis D, Mattson MP (2011) Disrupted energy metabolism and neuronal circuit dysfunction in cognitive impairment and Alzheimer's disease. *Lancet Neurol* 10(2): 187–198.
41. Schapira AHV (2012) Mitochondrial diseases. *Lancet* 379(9828):1825–1834.
42. Prabakaran S, et al. (2004) Mitochondrial dysfunction in schizophrenia: Evidence for compromised brain metabolism and oxidative stress. *Mol Psychiatry* 9(7):684–697, 643.
43. Lee CK, et al. (2008) Quantitative methods for genome-scale analysis of in situ hybridization and correlation with microarray data. *Genome Biol* 9(1):R23.
44. Gillis J, Mistry M, Pavlidis P (2010) Gene function analysis in complex data sets using ErmineJ. *Nat Protoc* 5(6):1148–1159.
45. Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* 52(3):1059–1069.
46. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proc Natl Acad Sci USA* 99(12):7821–7826.
47. Estrada E, Hatano N (2008) Communicability in complex networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 77(3):036111.
48. de Reus MA, van den Heuvel MP (2013) Rich club organization and intermodule communication in the cat connectome. *J Neurosci* 33(32):12929–12939.
49. Ng L, et al. (2009) An anatomic gene expression atlas of the adult mouse brain. *Nat Neurosci* 12(3):356–362.
50. Zoubarev A, et al. (2012) Gemma: A resource for the reuse, sharing and meta-analysis of expression profiling data. *Bioinformatics* 28(17):2272–2273.

Supporting Information

Fulcher and Fornito 10.1073/pnas.1513302113

SI Materials and Methods

Connectivity Data. Mouse brain connectivity data are based on 469 viral microinjection experiments in C57BL/6J male mice at postnatal age P56 obtained from the Allen Mouse Brain Connectivity Atlas (AMBCA) (24). The AMBCA data were derived from imaging EGFP-labeled axonal projections at high resolution (0.35 μm) across 140 coronal sections (with an intersection spacing of 100 μm). These images were then registered to the Allen Mouse Brain Atlas (24, 26), resulting in a summary of projection data at a resolution of 100 μm^3 (24).

In this work, we analyzed data summarizing the interconnectivity between each of 213 brain regions derived from the computational mesoscale connectivity model of Oh et al. (24) in the form of a weighted, directed connectivity matrix. The data include a normalized connection strength (or weight) and a P value for every pairwise connection. The 213 brain regions analyzed are a subset of 295 nonoverlapping anatomical regions defined using the Allen Mouse Brain Atlas (26), where the subset was selected according to two inclusion criteria: (i) each region contained at least one injection experiment infecting at least 50 voxels in the region (this criterion excluded 80 regions), and (ii) the set of regions selected was sufficiently linearly separable under the connectivity model given the full dataset (this criterion excluded 2 additional regions) (24). The connectivity data, anatomical division labels, and spatial Euclidean distances between all pairs of regions were obtained from the supplemental material in the work by Oh et al. (24). Assignment of each brain region to 1 of 13 anatomical divisions follows the work by Oh et al. (24): isocortex (contains 38 brain regions), olfactory areas (11), hippocampal formation (11), cortical subplate (7), striatum (12), pallidum (8), thalamus (35), hypothalamus (19), midbrain (22), pons (13), medulla (25), cerebellar cortex (9), and cerebellar nuclei (3).

We analyzed the 213×213 connectivity matrix corresponding to ipsilateral connectivity between regions in the right hemisphere of the mouse brain. We focused principally on the binary matrix, retaining links with $P < 0.05$ (excluding self-connections), resulting in a link density of 6.9%. Our results are not sensitive to this P value threshold of 0.05 for including network links: a spatially distributed rich club of hubs is observed across a range of significance thresholds, as is the enrichment of gene coexpression in rich and feeder links (Fig. S5). By analyzing binary rather than weighted connectivity data, we did not have to commit to any particular definition of edge weight from the computational connectivity model by Oh et al. (24) when distinguishing between different connection types (i.e., rich, feeder, and peripheral connections). Weights assigned to links of the connectome were only used for the topological analysis of hub–hub link weights (Fig. S1*E*), where we used normalized connection strengths, $w_{X,Y}$ (24), as described below. Because anterograde tracer injections were in the right hemisphere of the mouse brain, AMBCA data allow for the computation of a full connectome only in the right hemisphere. Our analysis focuses on this connectivity data, but it will be important to verify our results in the left hemisphere and across both hemispheres as data become available.

Network Measures. Node degree, k, was defined as the sum of in degree and out degree as $k = k_{in} + k_{out}$. At a given k, all regions were classified as either hub (degree $> k$) or nonhub (degree $\leq k$), and all edges were classified as rich (hub → hub), feeder (hub → nonhub or nonhub → hub), or peripheral (nonhub → nonhub). We used the label "peripheral" in favor of the previously used term, "local" (3), to refer to links between nonhubs to avoid any spatial connotations of our link classification. Hub → nonhub and nonhub → hub connections were grouped as a single feeder category for simplicity, because gene coexpression patterns were similar for both.

The extent to which high-degree nodes link preferentially to each other was measured as the rich club coefficient, $\phi(k)$ (7, 25):

$$\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k} - 1)}, \quad \text{[S1]}$$

where $E_{>k}$ represents the number of edges in the subgraph containing nodes with degree $> k$, and $N_{>k}$ is the number of nodes with degree $> k$. The rich club coefficient, $\phi(k)$, thus measures the link density in the subgraph containing nodes with degree $> k$. The coefficient, $\phi(k)$, can be expected to increase with k, because retaining nodes with higher degree will yield a higher expected link density in the subgraph containing nodes with degree $> k$. Consequently, $\phi(k)$ was compared with the rich club coefficient obtained from an ensemble of 10,000 randomized null networks, $\phi_{rand}(k)$ (rewiring each edge an average of 50 times per null network), constructed by shuffling the links in the empirical network while retaining the degree distribution of the network. Null networks were generated using the randmio_dir function from the Brain Connectivity Toolbox (45). The normalized rich club coefficient, $\Phi_{norm}(k)$, was computed as the ratio of the rich club coefficient of the empirical network to the mean rich club coefficient of the ensemble of randomized networks: $\Phi_{norm}(k) = \phi(k) / \langle \phi_{rand}(k) \rangle$ (25). Statistical significance was assessed by computing a P value directly from the empirical null distribution, $\phi_{rand}(k)$, as a permutation test under the null hypothesis $\phi(k) \leq \phi_{rand}(k)$ (7). Note that, although the sharp increase in ϕ_{norm} at $k = 42$ coincided with the exclusion of the piriform area, the last remaining region of the olfactory cortex (Fig. S1*B*), the network's rich club organization was not driven by the olfactory areas; similar rich club organization was also observed when all olfactory regions were excluded from the analysis.

Hubs are broadly distributed across anatomical divisions, yet hub–hub connections show a similar proportion of interdivision links to those in the rest of the network (two-sided Welch's t test; $P > 0.3$ for all $42 \leq k \leq 54$). To analyze whether hub–hub links display different properties to other types of network links, we evaluated a range of link-based properties as shown in Fig. S1 *C–F*. For a given property, we assigned a value to each network connection, (i,j), after which the distribution of these values in hub–hub connections was compared with that of all other connections in the network. Statistically significant increases in hub–hub links were evaluated relative to all other network links using a one-sided Welch's t test (at $P < 0.05$).

To investigate the cost of network connections, we analyzed their connectivity weights, their reciprocity, and their connection distances. Link weight was assigned to network links as the normalized connection strength estimated from the model by Oh et al. (24). To compute the proportion of reciprocal connections, each connection (i,j) was assigned a value of one if $i \rightarrow j$ is accompanied by $j \rightarrow i$ and zero otherwise. Connection distance was estimated as the Euclidean distance between brain region centers using data from the work by Oh et al. (24). Connection distance is widely used as an estimate of brain network wiring cost in diverse species (2, 3). Longer connections occupy greater physical space in the brain, require more cellular material, and thus consume more metabolic resources. The Euclidean distance may underestimate wiring costs for some regions linked by

curved fibers, and therefore, this measure can be considered a conservative estimate of the true regionwise wiring cost of a network link.

Two different measures were used to investigate the topological centrality of rich links. We first computed edge betweenness centrality as the number of shortest paths between pairs of nodes that traverse a given edge (46). Edges with high betweenness centrality are thought to represent potential information-processing bottlenecks, because they mediate a large proportion of network traffic, assuming that such traffic travels along the shortest topological path. Given that these neuronal signals may not necessarily propagate along the shortest topological paths, we also computed a second measure of centrality called communicability (47), which takes into account all possible path lengths between a pair of nodes, i and j, and thus, may provide a more appropriate measure of centrality in brain networks. Weighting topological paths of length l as $1/l!$, the communicability, C_{ij}, between nodes i and j is defined as

$$C_{ij} = \sum_{l=0}^{\infty} \frac{(A^l)_{ij}}{l!} = (e^A)_{ij} \qquad \textbf{[S2]}$$

for a binary adjacency matrix, A.

The selection of a degree threshold with which to define brain network hubs has not been systematic in the existing literature, with existing studies including anywhere from the top 10% of nodes by degree (36) (here, $k > 49$) to the top 16% (3) (here, $k > 43$) or the top 23% (48) (here, $k > 38$). As shown in Fig. 2*B*, these hub definitions mostly lie within the topological rich club regime of the mouse connectome, $42 \leq k \leq 54$, and therefore yield similar results for choices of k in this range. Where possible, we display our results across the full range of k to show the robustness of our findings to this parameter. However, gene function analysis requires a hard threshold to define hubs, which was set at $k > 44$, corresponding to the mean + 1 SD of the degree distribution or the top 14% of nodes by degree. Using this hub definition, the connectome contains 228 rich links, 1,310 feeder links, and 1,525 peripheral links.

Gene Expression Data. Gene expression data measured using in situ hybridization (ISH) from the adult C57BL/6J male mouse at age P56 were obtained from the Allen Mouse Brain Atlas (26). Allen Mouse Brain Atlas gene expression data were retrieved for the same set of 213 anatomical brain regions as reported for the mesoscale mouse connectome (24) by querying the Allen API (api.brain-map.org/api/v2/data). All 22,157 section datasets were retrieved (in JSON format) using the following API query: api.brain-map.org/api/v2/data/query.json?criteria=model::SectionDataSet,rma::criteria,[failed$eq'false'][expression$eq'true'],products[id$eq1]. For each section dataset retrieved, gene metadata were obtained using a query of the following form: api.brain-map.org/api/v2/data/query.json?criteria=model::Gene,rma::criteria,data_sets[id$eqXXX] for each dataset identification (XXX). To get the identifications of structures (brain regions) used in the connectivity analysis, we first downloaded all structures in the Allen Mouse Brain Atlas using the following query: api.brain-map.org/api/v2/data/query.json?criteria=model::Structure,rma::criteria,[graph_id$eq1] and then matched them to the 213 structures used in the connectivity analysis [matching on region acronyms provided in the work by Oh et al. (24)]. We then iterated over these 213 structures and all of the section datasets retrieved above to retrieve measures of gene expression energy and density for each brain region using queries of the following form, api.brain-map.org/api/v2/data/query.json?criteria=model::StructureUnionize,rma::criteria,section_data_set[id$eqXXX],structure[id$eqYYY], for each section dataset identification (XXX) and each structure identification (YYY). In this way, we obtained measures of gene expression density and energy (defined below) for each of 213 brain regions and 22,157 section datasets.

We analyzed the full set of 22,157 experimental section datasets spanning 17,642 unique genes. ISH data were obtained from either sagittal or coronal sections (intersection spacing of 200 μm), which are registered to the Allen Mouse Brain Atlas using an algorithm that results in ISH data in the atlas space at 100-μm^3 resolution [in supplemental methods 2 in the work by Lein et al. (26)]. Each 100-μm^3 "quadrat" is labeled with the anatomical structures that it intersects, allowing quantification of expression statistics for a given brain region. Gene expression for a brain region was quantified in two ways: (*i*) expression density, which refers to the proportion of expressed voxels in an anatomical division, and (*ii*) expression energy, which measures the mean pixel intensity in a region (26, 49). We followed previous studies and used expression energy (20), but note that energy and density measurements are similar and that the main qualitative results of this paper are also reproduced using expression density. Genes measured in multiple experiments were represented by their average expression level in each region over those experiments, as per previous work (17). Because of potential differences in data quality between expression measurements derived from coronal and sagittal sections, we checked that the qualitative results of this paper were not sensitive to our use of both coronal and sagittal section data. Indeed, the main results were reproduced when computing coexpression values using data from 3,191 genes measured from coronal sections, including the exponential distance dependence of gene coexpression; the trends in coexpression across reciprocally connected, unidirectionally connected, and unconnected pairs of brain regions; and the trend across topological connection type, such that coexpression is highest for rich links followed by feeder links and then peripheral links within the topological rich club regime.

The magnitudes of ISH-measured expression levels are not directly comparable across genes but rather, reflect the relative amount of signal, arising from limitations of high-throughput, nonradioactive ISH (namely tyramide amplification for detecting low transcript concentrations, variations in probe permeability into the cell, variability in cell volume, and probe accessibility to mRNA) (43). To facilitate a meaningful comparison of ISH measurements across different genes, we required a transformation that put all genes on a comparable scale and which also accounted for the presence of outliers in the data (which often represent artifacts). Accordingly, we normalized the expression levels across the brain for each gene using a sigmoidal transformation:

$$S(x) = \frac{1}{1 + \exp\left(-\frac{x - \langle x \rangle}{\sigma_x}\right)}, \qquad \textbf{[S3]}$$

where $S(x)$ is the normalized expression value of a given gene, x is the raw expression value of that gene, and $\langle x \rangle$ and σ_x are the mean and SD of the expression values for that gene across the brain, respectively. After normalization, each gene was linearly rescaled to the unit interval, yielding a normalized set of expression values for each gene (shown for all 17,642 genes across all 213 brain regions in Fig. 1*C*). Normalized gene expression levels can be interpreted as the relative expression of that gene across the brain: from low values for that gene (blue in Fig. 1*C*) to high values for that gene (red in Fig. 1*C*).

Unnormalized expression values used in other work (17, 18, 26, 49) or monotonic transformations of these values, such as the logarithmic transformation (20), do not take into account the particular distribution of each gene's expression across the brain, do not saturate outlying expression data (e.g., because of potential artifacts in these data), and allow genes with high overall expression to dominate computed coexpression values.

Robust normalizing transformations, such as the Hampel hyperbolic tangent transformation, could more directly account for outliers in the data, but here we used the standard sigmoid for simplicity. We note, however, that the main results reported here are not a consequence of using sigmoidal normalization; we found similar differences in gene coexpression using unnormalized data, a linear rescaling to the unit interval, and the Hampel hyperbolic tangent transformation. The low values of coexpression reported here relative to other studies (26, 49) are because of the normalization of gene expression and the spatial correction applied to gene coexpression values. As explained above, spatial correction of gene coexpression allows us to be confident that our results represent robust effects of connectivity and connection topology that cannot be explained simply by the spatial proximity of different pairs of brain regions. We note that our qualitative results are not caused by spatial correction; similar qualitative results were obtained when no spatial correction was applied (see Fig. S3*A* and Fig. S6*A*).

Expression data are relatively complete, with only 293 of 17,642 genes displaying more than 10% missing values across 213 brain regions analyzed here (missing values are plotted green in Fig. 1*C*). Only 6 of 213 brain regions had more than 10% gene expression data missing: perirhinal area (PERI, isocortex, 48.6% missing), primary auditory area (AUDp, isocortex, 35.0% missing), ventral auditory area (AUDv, isocortex, 34.1% missing), nucleus raphe magnus (RM, medulla, 24.7% missing), periventricular hypothalamic nucleus, preoptic part (PVpo, hypothalamus, 19.5% missing), and dorsal auditory area (AUDd, isocortex, 10.1% missing). The treatment of missing values in gene coexpression calculations is explained below.

Gene Coexpression. In this work, we investigated how transcriptional coupling varies between different types of connected pairs of brain regions, and between connected and unconnected pairs of brain regions. This analysis is inherently bivariate, being concerned with the relationship between different pairs of brain areas. Accordingly, our genetic analysis focused on patterns of coupled gene expression rather than the expression levels of genes within any single brain region on its own. To this end, we quantified the similarity in gene expression profiles between a pair of brain regions, (i,j), as a gene coexpression value, G_{ij}, (or transcriptional coupling) defined as the Pearson correlation between the normalized expression levels across genes. In performing this calculation, we included only genes with less than 10% missing values (excluding 293 of 17,642 genes or 1.7%) and required that at least 70% of the remaining 17,349 genes have a valid gene expression measurement for both brain regions (excluding 635 of 22,578 distinct pairs of brain regions or 2.8%).

As shown in Fig. S3*A*, gene coexpression is significantly higher in the 419 pairs of brain regions that are reciprocally connected (0.46 ± 0.22; mean ± SD) than the 2,225 pairs that have a unidirectional connection between them (0.31 ± 0.22; Welch's t test; $P < 10^{-31}$), which themselves show significantly higher gene coexpression than 19,934 pairs of brain regions that are not connected (0.14 ± 0.20; Welch's t test; $P < 10^{-201}$). However, as shown in Fig. S3*B*, gene coexpression displays strong spatial correlations (21, 49), with an approximately exponential relationship between gene coexpression and spatial distance. Because connection probability also decreases exponentially with spatial separation (Fig. S2*A*), connected pairs of brain regions are more likely to be separated by shorter physical distances and thus can be expected to display increased gene coexpression on the basis of their spatial separation alone.

Here, we are interested in understanding the relationship between gene coexpression and brain network connectivity beyond the effects of spatial correlations. We therefore corrected for this effect by fitting a regression line of the form $r_g(d) = \exp(-\eta d)$ to the relationship between distance, d_{ij}, and gene coexpression, G_{ij}, for all unique pairs of brain regions, (i,j), as shown in Fig. S3*B*. This functional form captures the trend well with a single parameter, $\eta = 0.533$, and enforces the constraints of $r_g = 1$ at $d = 0$ (i.e., gene expression shows maximal correlation with itself) and $r_g \to 0$ as $d \to \infty$ (i.e., gene coexpression decreases to zero with increasing distance of separation). Retaining the residuals from this exponential fit as $\hat{G}_{ij} = G_{ij} - r_g(d_{ij})$ allowed us to understand patterns of gene coexpression beyond spatial correlation effects.

Spatially corrected distributions of gene coexpression values for each class of interregion pairs (reciprocally connected, unidirectionally connected, and unconnected) are shown in Fig. S3*C*. The same trend in coexpression values is retained after correction, with reciprocally connected pairs of brain regions showing the highest corrected gene coexpression values (0.10 ± 0.17; mean ± SD), significantly higher than unidirectionally connected pairs (0.06 ± 0.16; $P < 10^{-6}$), which themselves show significantly higher gene coexpression values than unconnected pairs (-0.01 ± 0.16; $P < 10^{-64}$). Correcting for spatial correlation decreased mean gene coexpression in unconnected pairs of brain regions to $\langle \hat{G}_{ij} \rangle \approx 0$, indicating an unbiased fit that successfully captured the spatial correlation in the data. Previous analyses have corrected for spatial correlation in transcriptional measurements using partial Mantel tests (18, 20, 21) or by excluding pairs of samples in the same specific anatomical division (88 tissue classes in the human analysis in ref. 22). By contrast, here, we show that the spatial dependence is exponential and correct for it explicitly.

The spatial correction of gene coexpression values described above helps to overcome a strong effect of brain division on gene expression data (49). For example, as shown in Fig. 1*C*, a large cluster of genes is highly expressed in the isocortex and cortical subplate (in the top left corner of the region × gene expression matrix) compared with other brain regions. We found that additional corrections that explicitly take into account the effect of differential division-based gene expression [e.g., by fitting the coexpression–distance relationship separately for each of 13 anatomical divisions (Fig. S6)] produced similar results, and here, we favored the simpler global distance correction described above, which makes fewer assumptions of the data.

To generate the gene coexpression plot in Fig. 3*B*, we assigned a spatially corrected gene coexpression value, $\hat{G}_{ij}$, to each edge in the network. At each k, we computed the mean gene coexpression in each link type (i.e., rich, feeder, and peripheral). To evaluate significant increases in gene coexpression values assigned to a given link type against all other network links, we used a one-sided Welch's t test ($P < 0.05$).

In this work, we focused on patterns of gene coexpression between pairs of brain regions, which allowed us to properly correct for spatial correlations in the data and investigate differential coexpression patterns in different connection types. By contrast, analysis of gene expression at the level of individual brain regions (i.e., at the node level) is conflated with these spatial effects, including the selective expression of many genes in particular brain divisions (noted above). Despite this methodological limitation, analysis performed at the node level (that is, scoring each gene using a t statistic measuring increased expression in hub regions over nonhub regions) yielded meaningful categories related to neuronal communication, connectivity, and energy metabolism that mostly concurred with our analysis of pairwise interregional gene coexpression.

Gene Function Analysis. In this work, it was important to determine which specific functional groups of genes contribute to observed increases in gene coexpression values in different types of interregion pairs (e.g., connected vs. unconnected, connected hubs vs. connected nonhubs, etc.). We first measured the contribution of each individual gene to the spatially corrected gene coexpression value, $\hat{G}_{ij}$, described above using the definition of the Pearson correlation:

$$\hat{G}_{ij} = G_{ij} - r_g(d_{ij}) = \frac{1}{N}\sum_{a=1}^{N}\left[\tilde{g}_i^{(a)}\tilde{g}_j^{(a)} - r_g(d_{ij})\right], \qquad \textbf{[S4]}$$

where the sum is taken across all $N = 17{,}642$ genes, G_{ij} is the Pearson correlation of expression levels across all genes, and $\tilde{g}_i^{(a)}\tilde{g}_j^{(a)}$ is the product of z-scored, normalized gene expression values for gene a in regions i and j (with the z score taken relative to all genes in each brain region). Spatial correlations in the data, described above, were corrected using the fitted spatial correlation function $r_g(d_{ij}) = \exp(-0.533 d_{ij})$. Gene coexpression values assigned to each interregion pair (i,j) for a given gene a are referred to as GCC scores: $\mathrm{GCC}_{ij}^{(a)} = \tilde{g}_i^{(a)}\tilde{g}_j^{(a)} - r_g(d_{ij})$.

In this work, GCC scores were used to analyze the contribution of individual genes to gene coexpression values computed for different types of interregion pairs (connected pairs vs. unconnected pairs) and different types of network edges, such as rich, feeder, and peripheral connections. Each gene was assigned a t-statistic score measuring the increase in GCC values of a given gene in one such group over another using a one-sided Welch's t test; a high t-statistic score reflects a more correlated pattern of gene expression in a given class of interregion pairs over another. After scoring genes in this way, the computed t-statistic scores were then related to meaningful categories of biological processes using GO annotations (27). This ability to score all genes in a principled way by delineating their contribution to the measured Pearson correlation value (computed across all genes) circumvented the need for a greedy iterative removal method used in previous work (15, 20, 21), which is computationally expensive and produces enriched gene sets rather than providing a quantitative score for each gene.

Gene function analysis was performed using version 3.0.2 of ErmineJ (44). Gene annotations were assigned from GO (27) using an annotation file from GEMMA (50): Generic_mouse_ncbiIds_noParents.an was downloaded from www.chibi.ubc.ca/Gemma/arrays/showArrayDesign.html?id=736 on January 6, 2015. GO terms and definitions were obtained as an RDF XML file downloaded from archive.geneontology.org/latest-termdb/go_daily-termdb.rdf-xml.gz on January 6, 2015. We performed gene score resampling on the t-statistic scores assigned to genes (described above) considering gene set sizes in the range 5–100, taking the mean score in a GO group to summarize it, and using full resampling with 10^7 iterations. Annotations were present for 11,789 genes when analyzing 4,968 biological process categories (or 11,830 genes when analyzing 5,593 annotation categories for both biological processes and cellular components). Significance levels were adjusted for multiple comparison testing using the method by Benjamini and Hochberg.

To understand the behavior of a subset of genes (such as the 70 unique genes annotated to the GO-annotated biological process categories with increased coexpression in rich and feeder links relative to peripheral links), we took the mean in Eq. **S4** across just these genes, allowing us to compute their contribution to gene coexpression for each connection (i,j). The mean of these values across different connection types is plotted as a function of k in Fig. 4 (with statistically significant increases at $P < 0.05$ measured as before: using a one-sided Welch's t test).

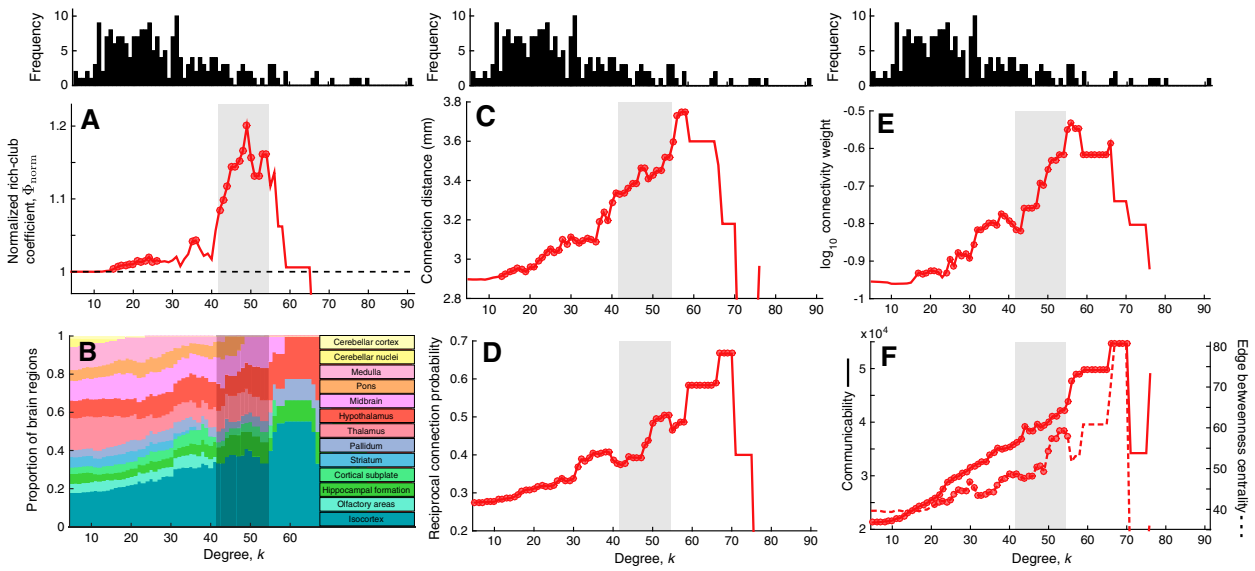

Fig. S1. Hub connectivity is topologically central and costly. Each plot shows the variation of a network property as a function of the number of connections (or degree k) at which hubs (regions with degree $> k$) are defined. *A*, *Upper*, *C*, *Upper*, and *E*, *Upper* show the degree distribution of the mouse connectome for ease of comparison with each plot. (*A*) Normalized rich club coefficient, Φ_{norm}, as in Fig. 2*B* reproduced here for ease of comparison. Circles indicate values of Φ_{norm} that are statistically higher than an ensemble of 10,000 null networks (permutation test; $P < 0.05$). The topological rich regime ($42 \leq k \leq 54$) is shaded gray in all figures. (*B*) The proportion of brain regions with degree $> k$ in each of 13 broad anatomical divisions from the Allen Mouse Brain Atlas (24, 26) as a function of k. (*C*–*F*) The remaining plots summarize other mean properties of links connecting regions with degree $> k$, with circles indicating statistically significant increases relative to values assigned to all other network links (Welch's t test; $P < 0.05$): (*C*) connection distance (millimeters), (*D*) reciprocal connectivity, (*E*) $\log_{10}$ connectivity weight, and (*F*) communicability (line) (47) and edge betweenness centrality (dashed line) (46). Each of these properties increases with k and is elevated across the topological rich club regime, indicating a high centrality and cost of connections between hub regions.

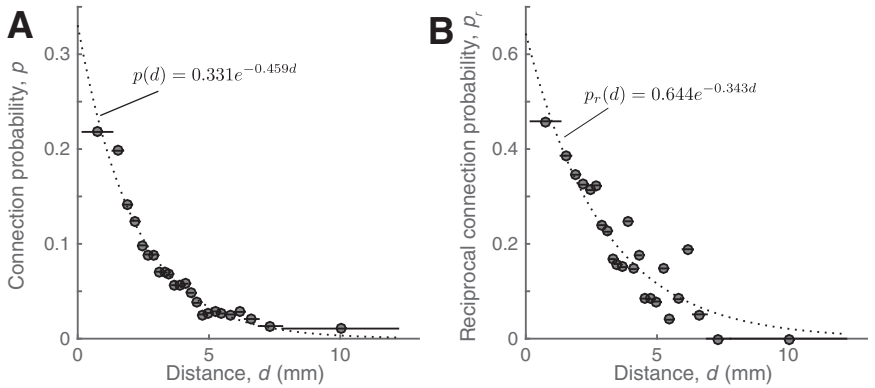

Fig. S2. Probability that (*A*) a connection exists, $p(d)$, and (*B*) an existing connection has a reciprocal match, $p_r(d)$, estimated in 25 equiprobable bins as a function of the separation distance, *d*, for all connected pairs of brain regions. The mean of each bin is shown with a circle, and its extent is shown with a horizontal line. Exponential fits are plotted as dotted lines and labeled.

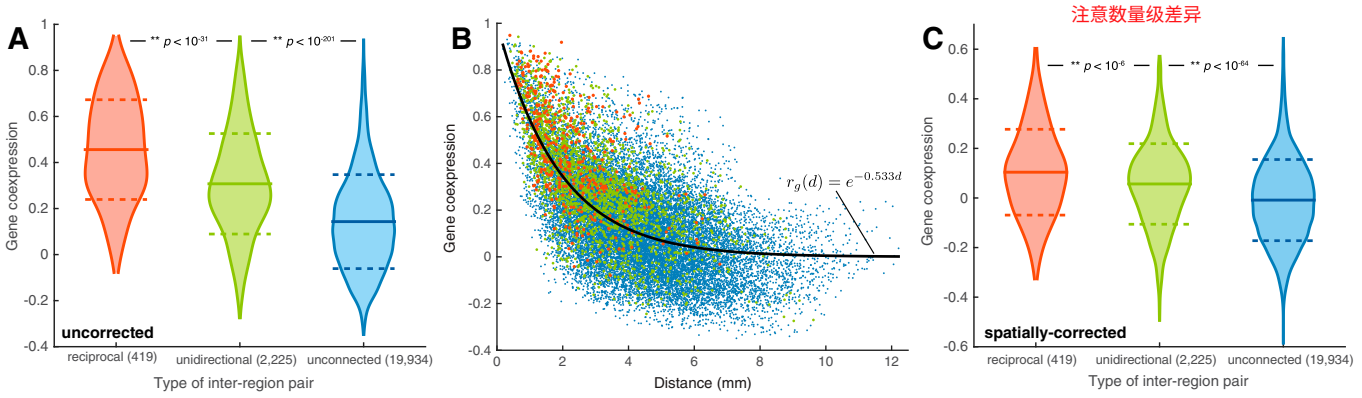

Fig. S3. Relationship between gene coexpression, connectivity, and separation distance in the mouse connectome. All pairs of brain regions *i* and *j* (excluding self-connections, $i \neq j$) were classed as (*i*) reciprocally connected pairs of brain regions if both $i \rightarrow j$ and $j \rightarrow i$ (orange), (*ii*) unidirectionally connected pairs of brain regions if either $i \rightarrow j$ or $j \rightarrow i$ (but not both; green), or (*iii*) unconnected pairs if neither connection is present (blue). (*A*) Distributions of gene coexpression for each of the above types of interregion pairs. (*B*) Gene coexpression, G_{ij}, as a function of Euclidean distance of separation, d_{ij}, for all interregion pairs, with a fitted exponential decay as labeled. (*C*) Distributions of gene coexpression for all classes of interregion pairs after correcting for the exponential distance relationship shown in *B*. Both before and after correcting for spatial correlation in the data, pairs of reciprocally connected brain regions have the highest gene coexpression followed by unidirectionally connected brain regions and unconnected brain regions; *P* values from Welch's *t* tests are annotated to *A* and *C*.

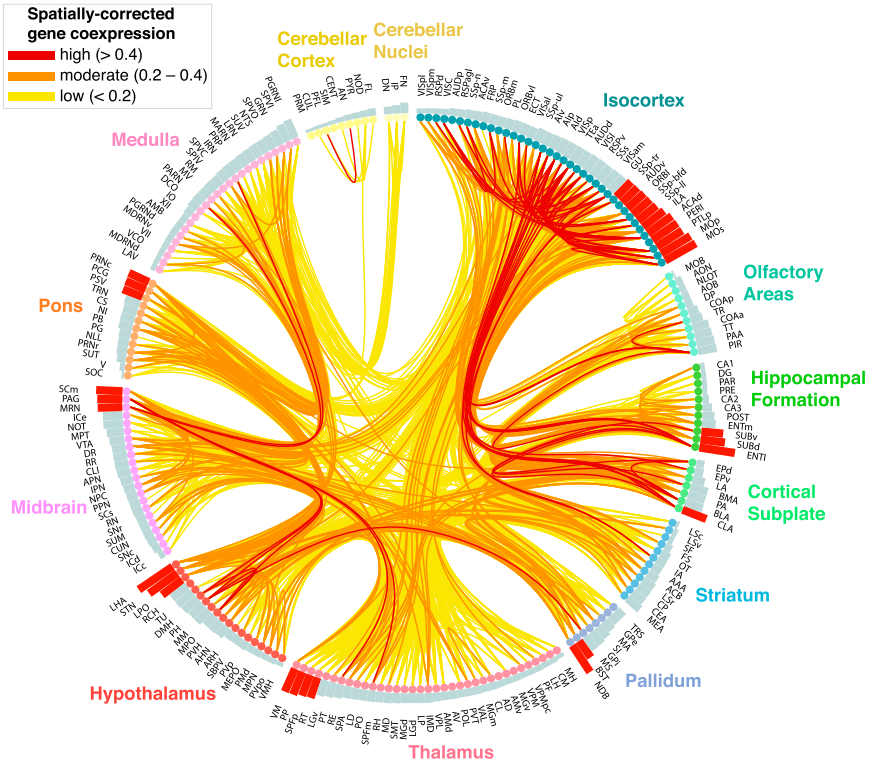

Fig. S4. A larger version of Fig. 3*A* with all regions labeled. All neuronal connections (lines) between brain regions (circles) are colored according to the gene coexpression of the regions that they connect. Brain regions are organized by anatomical division and sorted by degree (shown as bars), with bars colored bright red for hubs ($k > 44$). Abbreviations of brain regions are from the Allen Mouse Brain Atlas (26).

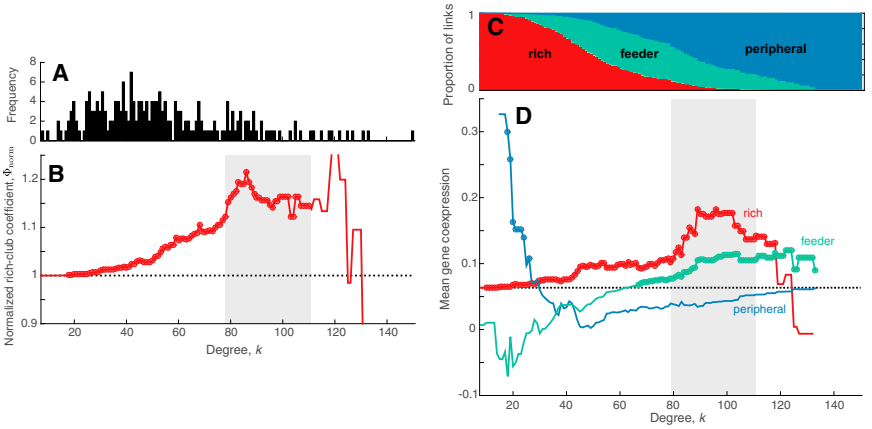

Fig. S5. Rich club organization and gene coexpression results are robust to the significance threshold used to retain connectome links. In our primary analyses, we retained connectome links with $P < 0.05$ in the computational model of the mouse connectome (24), producing a link density of 6.9%. Here, the data are reanalyzed using the more lenient threshold of $P < 0.5$, yielding a connectome with a link density of 12.9%. (*A*) Degree distribution. (*B*) Normalized rich club coefficient, $\Phi_{norm}(k)$, computed relative to 10,000 randomized null networks. The topological rich regime from the point where Φ_{norm} increases sharply (at $k = 78$) to $k = 110$ is shaded. (*C*) Proportion of each link type (rich, feeder, and peripheral) as a function of k. (*D*) Mean gene coexpression in each link type as a function of k, with statistically significant enrichment over other types of links indicated with circles (Welch's t test; $P < 0.05$) and an analogous topological rich club regime for this network shaded gray. The same qualitative results are reproduced with this denser connectome, including the coexpression increase for rich and feeder links across the topological rich club regime. The increase in gene coexpression for peripheral links at very low k is not meaningful for the hub connectivity analyzed here (at this k, 85% of nodes are labeled as hub, and less than 1% of links are labeled as peripheral). Qualitatively similar rich club curves and gene coexpression patterns were also found at link thresholds $P < 0.25$ (link density, 9.7%) and $P < 0.75$ (link density, 17.2%).

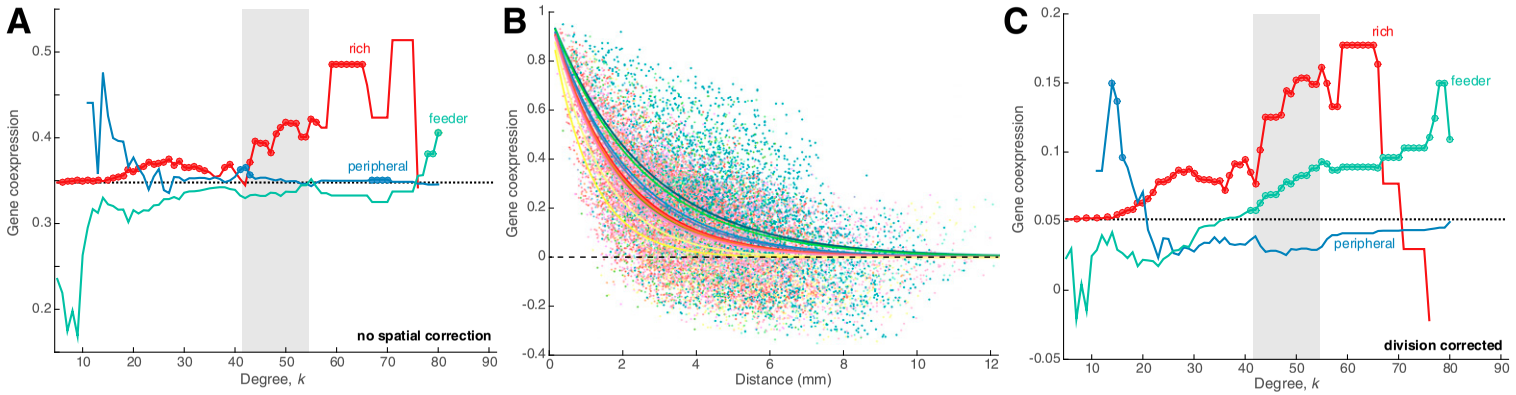

Fig. S6. The relationship between gene coexpression and hub connectivity is robust to different processing methods. *A* and *C* are the same as in Fig. 3*B* but with different corrections applied. (*A*) When spatial correlations in gene coexpression values are not corrected for, gene coexpression remains increased for rich connections in the topological rich regime but does not remain increased for feeder links. (*B*) An alternative to the global spatial correction for gene coexpression values applied here is to correct the effect in each division separately. Exponential fits, $r_g(d) = \exp(-\eta d)$, are shown (where links $i \to j$ are labeled as division i), with every pair of regions shown as a point in the plot colored by their division (using the same color labels as in Fig. S1*B*). (*C*) Applying this division-based spatial correction shown in *B* yields similar results to that of the global spatial correction.

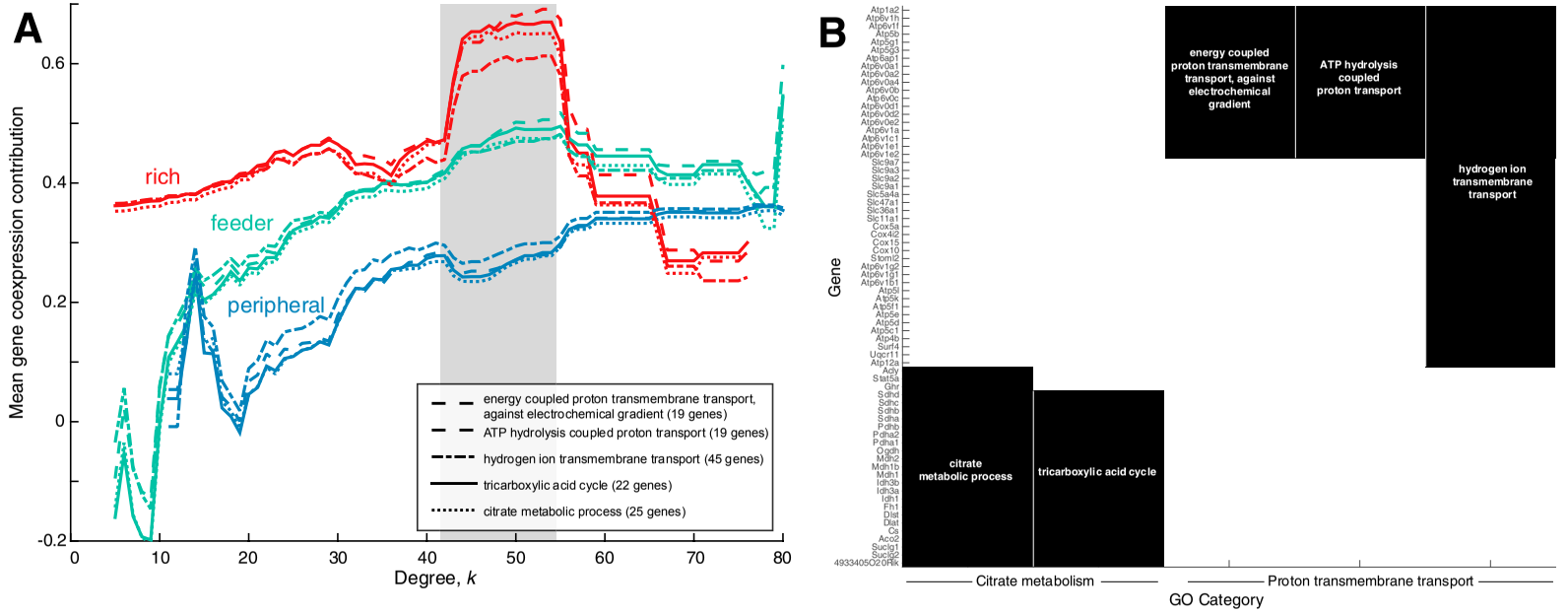

Fig. S7. Mean gene coexpression in rich, feeder, and peripheral links for the five GO-annotated biological process categories showing a significant increase in gene coexpression for connections involving hubs computed as the mean GCC value (Table 1). Gene coexpression curves are highly consistent across these GO categories, despite containing different sets of genes. (*A*) Mean GCC score computed across genes in each category (labeled in the legend) for rich, feeder, and peripheral links as a function of the degree, k, at which hubs are defined (degree > k). (*B*) Table showing the membership of 70 unique genes (rows) annotated to five GO categories (columns); membership is indicated in black. The citrate metabolic process and tricarboxylic acid cycle categories contain many similar genes, which are completely different from the genes annotated to the other categories. The category energy-coupled proton transmembrane transport against electrochemical gradient contains the same annotations as the ATP hydrolysis-coupled proton transport category, and all genes annotated to these categories are also annotated to hydrogen ion transmembrane transport, which includes 26 additional genes.

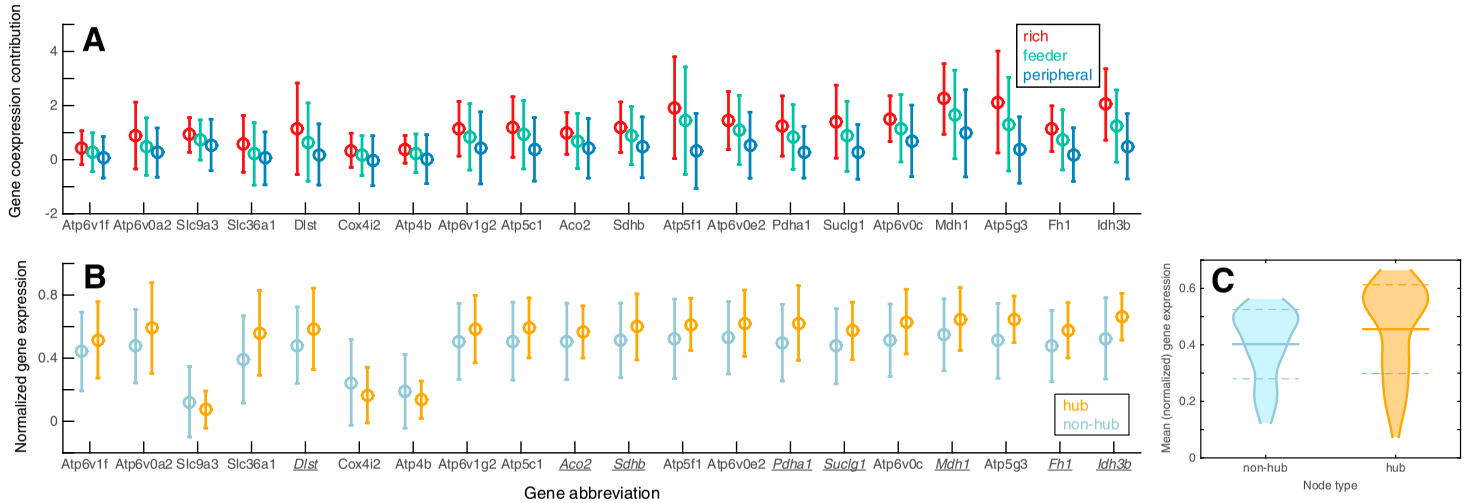

Fig. S8. Interregional gene coexpression and regional gene expression of 20 exemplary metabolic genes related to hub connectivity. The 20 genes plotted here (of 70 unique genes annotated to the processes listed in Table 1) are those with the greatest increases in coexpression in rich and feeder connections relative to peripheral connections. Of the 70 genes annotated to these biological processes, 64 (or 91%) show increased expression in hubs over nonhubs [of which 12 are significantly increased: $P < 0.05$; false discovery rate (FDR) -corrected across 70 genes; one-sided Welch's t test; none showed significant decreases], and 57 (or 81%) show increased GCC scores in rich links over peripheral links (of which 46 are significantly increased: $P < 0.05$; FDR-corrected; one-sided Welch's t test). (*A*) Distributions (mean ± SD) of coexpression values for each individual gene across rich (red), feeder (green), and peripheral (blue) connections. (*B*) Distribution of normalized regional gene expression for hubs (orange) and nonhubs (blue). Genes related to citrate metabolism are shown italicized and underlined to distinguish them from genes related to proton transport. (*C*) Distributions of mean gene expression in nonhub regions and hub regions across all 70 genes. Mean gene expression is significantly increased in hubs (orange) over nonhubs (blue; $P = 0.028$; Welch's t test).

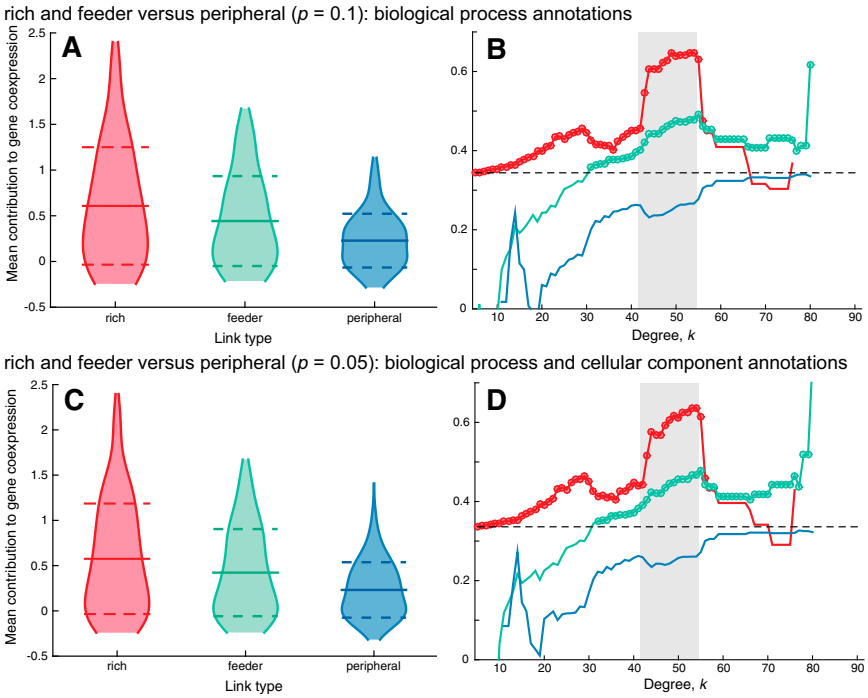

Fig. S9. The association between topological connection type and mean GCC scores is robust to changes in significance threshold and gene annotation sets. Plots are the same as in Fig. 4, but use (*A* and *B*) all 112 genes annotated to 12 biological process GO categories showing increased gene coexpression in rich and feeder links over peripheral links at a false discovery rate (FDR) of 0.1 (Table S5) or (*C* and *D*) all 335 genes annotated to 25 biological process and cellular component categories showing increased gene coexpression in rich and feeder links over peripheral links at an FDR of 0.05 (Table S4).

Table S1. Functional gene groups showing significantly increased transcriptional coupling in pairs of connected brain regions (relative to unconnected pairs) are mainly related to neuronal connectivity and communication

GO category	Description	No. of genes	*P* value
GO:0048167	Regulation of synaptic plasticity	84	0.0014
GO:0007270	Neuron–neuron synaptic transmission	45	0.0039
GO:0048168	Regulation of neuronal synaptic plasticity	37	0.0062
GO:0045837	Negative regulation of membrane potential	7	0.0063
GO:0035235	Ionotropic glutamate receptor signaling pathway	19	0.014
GO:0060078	Regulation of postsynaptic membrane potential	35	0.018
GO:0006414	Translational elongation	29	0.021
GO:0050806	Positive regulation of synaptic transmission	63	0.023
GO:0050808	Synapse organization	77	0.024
GO:0008344	Adult locomotory behavior	66	0.026
GO:0060079	Regulation of excitatory postsynaptic membrane potential	31	0.031
GO:0097090	Presynaptic membrane organization	7	0.034
GO:0051650	Establishment of vesicle localization	74	0.036
GO:0035249	Synaptic transmission, glutamatergic	28	0.036
GO:0050685	Positive regulation of mRNA processing	16	0.036
GO:0051648	Vesicle localization	80	0.039
GO:0045454	Cell redox homeostasis	46	0.04
GO:0061001	Regulation of dendritic spine morphogenesis	16	0.04
GO:0050770	Regulation of axonogenesis	77	0.041
GO:1903313	Positive regulation of mRNA metabolic process	19	0.041
GO:2000463	Positive regulation of excitatory postsynaptic membrane potential	12	0.041
GO:0022900	Electron transport chain	32	0.042
GO:0042775	Mitochondrial ATP synthesis-coupled electron transport	15	0.042
GO:0051899	Membrane depolarization	67	0.043
GO:0007628	Adult walking behavior	33	0.043
GO:0007416	Synapse assembly	27	0.043
GO:0022904	Respiratory electron transport chain	28	0.044
GO:0006626	Protein targeting to mitochondrion	24	0.045
GO:0021885	Forebrain cell migration	32	0.047
GO:0007015	Actin filament organization	88	0.048
GO:0097105	Presynaptic membrane assembly	5	0.049

GO biological process categories are listed in increasing order of their false discovery rate-corrected *P* values. Repeating the analysis including annotations for cellular components revealed similar categories related to neurites and synapses as well as the respiratory chain and oxidoreductase complex (Table S2).

Table S2. Biological process and cellular component GO categories of genes showing significantly increased coexpression in connected pairs of brain regions over unconnected pairs

GO category	Description	No. of genes	FDR-corrected *P* value
GO:0022626	Cytosolic ribosome	63	2.6×10^{-9}
GO:0005746	Mitochondrial respiratory chain	41	5.2×10^{-9}
GO:0070469	Respiratory chain	46	0.00034
GO:0044391	Ribosomal subunit	97	0.00039
GO:0044455	Mitochondrial membrane part	99	0.00041
GO:0048167	Regulation of synaptic plasticity	84	0.00043
GO:0022627	Cytosolic small ribosomal subunit	27	0.00052
GO:0007270	Neuron–neuron synaptic transmission	45	0.00097
GO:1990204	Oxidoreductase complex	59	0.0011
GO:0005747	Mitochondrial respiratory chain complex I	30	0.0018
GO:0030964	NADH dehydrogenase complex	30	0.0018
GO:0045271	Respiratory chain complex I	30	0.0018
GO:0045837	Negative regulation of membrane potential	7	0.0018
GO:0044304	Main axon	36	0.0019
GO:0048168	Regulation of neuronal synaptic plasticity	37	0.0021
GO:0015935	Small ribosomal subunit	46	0.0025
GO:0032839	Dendrite cytoplasm	17	0.0032
GO:0035235	Ionotropic glutamate receptor signaling pathway	19	0.0038
GO:0060078	Regulation of postsynaptic membrane potential	35	0.0066
GO:0005798	Golgi-associated vesicle	45	0.0096
GO:0050806	Positive regulation of synaptic transmission	63	0.0097
GO:0006414	Translational elongation	29	0.0098
GO:0050808	Synapse organization	77	0.012
GO:0008344	Adult locomotory behavior	66	0.013
GO:0008021	Synaptic vesicle	74	0.014
GO:0014069	Postsynaptic density	88	0.014
GO:0060079	Regulation of excitatory postsynaptic membrane potential	31	0.015
GO:0097090	Presynaptic membrane organization	7	0.016
GO:0044306	Neuron projection terminus	81	0.016
GO:0051650	Establishment of vesicle localization	74	0.019
GO:0035249	Synaptic transmission, glutamatergic	28	0.02
GO:0016469	Proton-transporting two-sector ATPase complex	32	0.02
GO:0050685	Positive regulation of mRNA processing	16	0.02
GO:0042775	Mitochondrial ATP synthesis-coupled electron transport	15	0.024
GO:0045454	Cell redox homeostasis	46	0.024
GO:0022625	Cytosolic large ribosomal subunit	32	0.024
GO:0061001	Regulation of dendritic spine morphogenesis	16	0.024
GO:0007628	Adult walking behavior	33	0.025
GO:1903313	Positive regulation of mRNA metabolic process	19	0.025
GO:0051648	Vesicle localization	80	0.027
GO:0006626	Protein targeting to mitochondrion	24	0.028
GO:0007416	Synapse assembly	27	0.028
GO:0050770	Regulation of axonogenesis	77	0.028
GO:0022900	Electron transport chain	32	0.028
GO:0051899	Membrane depolarization	67	0.028
GO:2000463	Positive regulation of excitatory postsynaptic membrane potential	12	0.029
GO:0022904	Respiratory electron transport chain	28	0.03
GO:0043679	Axon terminus	74	0.031
GO:0021885	Forebrain cell migration	32	0.033
GO:0007015	Actin filament organization	88	0.033
GO:0097105	Presynaptic membrane assembly	5	0.033
GO:0022029	Telencephalon cell migration	30	0.042
GO:0030426	Growth cone	69	0.042
GO:0030120	Vesicle coat	28	0.043
GO:0048169	Regulation of long-term neuronal synaptic plasticity	23	0.043
GO:0021795	Cerebral cortex cell migration	25	0.044
GO:0060076	Excitatory synapse	20	0.045
GO:0071229	Cellular response to acid chemical	63	0.047
GO:0045838	Positive regulation of membrane potential	18	0.049
GO:0044309	Neuron spine	82	0.049

FDR, false discovery rate.

Table S3. Hubs of the mouse connectome

Structure name (acronym)	Major division	Degree, k
Lateral hypothalamic area (LHA)	Hypothalamus	91
Entorhinal area, lateral part (ENTl)	Hippocampal formation	80
Secondary motor area (MOs)	Isocortex	78
Subthalamic nucleus (STN)	Hypothalamus	77
Primary motor area (MOp)	Isocortex	76
Diagonal band nucleus (NDB)	Pallidum	71
Posterior parietal association areas (PTLp)	Isocortex	67
Perirhinal area (PERI)	Isocortex	67
Anterior cingulate area, dorsal part (ACAd)	Isocortex	66
Infralimbic area (ILA)	Isocortex	59
Superior colliculus, motor-related (SCm)	Midbrain	59
Claustrum (CLA)	Cortical subplate	57
Ventral medial nucleus of the thalamus (VM)	Thalamus	56
Lateral preoptic area (LPO)	Hypothalamus	56
Periaqueductal gray (PAG)	Midbrain	56
Subiculum, dorsal part (SUBd)	Hippocampal formation	55
Peripeduncular nucleus (PP)	Thalamus	55
Midbrain reticular nucleus (MRN)	Midbrain	55
Primary somatosensory area, lower limb (SSp-ll)	Isocortex	53
Primary somatosensory area, barrel field (SSp-bfd)	Isocortex	51
Orbital area, lateral part (ORBl)	Isocortex	50
Subparafascicular nucleus, parvicellular part (SPFp)	Thalamus	50
Subiculum, ventral part (SUBv)	Hippocampal formation	49
Pontine reticular nucleus, caudal part (PRNc)	Pons	49
Ventral auditory area (AUDv)	Isocortex	48
Reticular nucleus of the thalamus (RT)	Thalamus	48
Pontine central gray (PCG)	Pons	48
Primary somatosensory area, trunk (SSp-tr)	Isocortex	47
Principal sensory nucleus of the trigeminal (PSV)	Pons	47
Bed nuclei of the stria terminalis (BST)	Pallidum	45

Hubs are defined as regions with $k > 44$ (corresponding to the mean + SD of the degree distribution). For each hub, we list the region name, abbreviation, major division (24), and degree, k, and entries have been sorted (descending) by degree.

Table S4. Biological function and cellular component GO categories showing increased gene coexpression for connections involving hubs (i.e., rich and feeder links) compared with connections between nonhubs (peripheral links) up to a maximum false discovery rate-corrected *P* value of 0.1

GO category	Description	No. of genes	FDR-corrected *P* value
GO:0005746	Mitochondrial respiratory chain	41	0.00052
GO:0070469	Respiratory chain	46	0.00078
GO:0016469	Proton-transporting two-sector ATPase complex	32	0.0019
GO:0044455	Mitochondrial membrane part	99	0.0026
GO:0006099	Tricarboxylic acid cycle	22	0.0036
GO:0015988	Energy-coupled proton transmembrane transport against electrochemical gradient	19	0.004
GO:0015991	ATP hydrolysis-coupled proton transport	19	0.004
GO:0005747	Mitochondrial respiratory chain complex I	30	0.014
GO:0030964	NADH dehydrogenase complex	30	0.014
GO:0045271	Respiratory chain complex I	30	0.014
GO:0072350	Tricarboxylic acid metabolic process	27	0.018
GO:0005868	Cytoplasmic dynein complex	18	0.019
GO:1902600	Hydrogen ion transmembrane transport	45	0.019
GO:0006101	Citrate metabolic process	25	0.021
GO:0005832	Chaperonin-containing T complex	7	0.023
GO:0000502	Proteasome complex	46	0.024
GO:0022626	Cytosolic ribosome	63	0.025
GO:0006103	2-Oxoglutarate metabolic process	10	0.025
GO:0006107	Oxaloacetate metabolic process	9	0.026
GO:0042026	Protein refolding	7	0.031
GO:0015992	Proton transport	57	0.031
GO:0033178	Proton-transporting two-sector ATPase complex, catalytic domain	11	0.032
GO:1990204	Oxidoreductase complex	59	0.038
GO:0006818	Hydrogen transport	58	0.04
GO:0072384	Organelle transport along microtubule	19	0.047
GO:0022625	Cytosolic large ribosomal subunit	32	0.06
GO:0030137	COPI-coated vesicle	15	0.062
GO:0030286	Dynein complex	23	0.065
GO:0045333	Cellular respiration	67	0.079
GO:0032886	Regulation of microtubule-based process	89	0.081
GO:0043648	Dicarboxylic acid metabolic process	53	0.086
GO:0033176	Proton-transporting V-type ATPase complex	17	0.091
GO:0045259	Proton-transporting ATP synthase complex	13	0.091

The $P = 0.05$ threshold is marked with a midrule. COPI, coat protein; FDR, false discovery rate.

Table S5. Biological process GO categories showing increased gene coexpression for rich links compared with peripheral links, feeder links compared with peripheral links, and rich and feeder links (as a group) compared with peripheral links

GO category	Description	No. of genes	FDR-corrected *P* value
	Rich vs. peripheral		
GO:0015988	Energy-coupled proton transmembrane transport against electrochemical gradient	19	0.065
GO:0015991	ATP hydrolysis-coupled proton transport	19	0.065
GO:0006099	Tricarboxylic acid cycle	22	0.08
	Feeder vs. peripheral		
GO:0006099	Tricarboxylic acid cycle	22	0.009
GO:0015988	Energy-coupled proton transmembrane transport against electrochemical gradient	19	0.015
GO:0015991	ATP hydrolysis-coupled proton transport	19	0.015
GO:0006101	Citrate metabolic process	25	0.031
GO:0072350	Tricarboxylic acid metabolic process	27	0.034
GO:0006107	Oxaloacetate metabolic process	9	0.039
GO:1902600	Hydrogen ion transmembrane transport	45	0.041
GO:0015992	Proton transport	57	0.045
GO:0006103	2-Oxoglutarate metabolic process	10	0.045
GO:0006818	Hydrogen transport	58	0.051
GO:0072384	Organelle transport along microtubule	19	0.083
GO:0032886	Regulation of microtubule-based process	89	0.088
GO:0042026	Protein refolding	7	0.089
	Rich and feeder vs. peripheral		
GO:0015988	Energy-coupled proton transmembrane transport against electrochemical gradient	19	0.0097
GO:0015991	ATP hydrolysis-coupled proton transport	19	0.0097
GO:0006099	Tricarboxylic acid cycle	22	0.014
GO:1902600	Hydrogen ion transmembrane transport	45	0.04
GO:0006101	Citrate metabolic process	25	0.045
GO:0072350	Tricarboxylic acid metabolic process	27	0.051
GO:0006103	2-Oxoglutarate metabolic process	10	0.052
GO:0006107	Oxaloacetate metabolic process	9	0.053
GO:0015992	Proton transport	57	0.058
GO:0042026	Protein refolding	7	0.065
GO:0006818	Hydrogen transport	58	0.07
GO:0072384	Organelle transport along microtubule	19	0.089

In each case, GO categories are ordered by their false discovery rate (FDR)-corrected *P* values up to 0.1. Statistical power is limited for analyses in which rich links are grouped on their own because of the relatively low number of connections in this category.