PNAS
www.pnas.org

Supplementary Information for

Subsystem organization of axonal connections within and between the right and left cerebral cortex and cerebral nuclei (endbrain)

Larry W. Swanson, Joel D. Hahn, Lucas G.S. Jeub, Santo Fortunato, and Olaf Sporns

Larry W. Swanson
Email: larryswanson10@gmail.com

This PDF file includes:

Supplementary text
Figs. S1 to S7

Caption for movie S1
Captions for databases S1 and S2
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Movie S1
Datasets S1 and S2

SI Materials and Methods

Cerebral Hemisphere Histological Parcellation Granularity. To facilitate comparative analysis, all rat connection data were collated with respect to one standard rat brain atlas (1) where central nervous system (CNS) gray matter regions are arranged according to a hierarchical nomenclature with a strictly topographic ordering and the CNS hierarchical level of gray matter regions and subregions are explicitly recognized as comparable (respectively) to the species and subspecies levels in animal taxonomy (2). This nomenclature scheme recognizes 77 gray matter regions in the rat cerebral cortex of each hemisphere and 45 gray matter regions in the rat cerebral nuclei of each hemisphere, all of which were included in the present analysis (for both hemispheres).

Connection Report Collation and Selection for Network Analysis. Our methodology for expertly collating connectional data from the primary neuroanatomical research literature follows previous descriptions (3, 4) and is elaborated here. First, the primary literature was searched to find the best available connection data, from which connection reports were generated (see Dataset S1 for a list of all connection reports). Several criteria were used to assess the quality of connection data: these included the validity of the experimental pathway tracing method used, restriction of the pathway tracer injection site to the gray matter region of interest for positive data (very large injections sites are useful for negative data), injection site coverage of the region of interest, and thoroughness of description of the connection.

The validity of the experimental pathway tracing methods (that is, their ability to generate valid data) was assessed in relation to three primary criteria: (a) *reach* (mono- or

polysynaptic), (b) *sensitivity* (clarity of labeling), and (c) *uptake* (by axons-of-passage). Based on these main criteria an overall validity rating was assigned to each pathway tracer on a 7-point ordinal scale (lowest to highest validity). The basis of validity rating assignment to pathway tracers is described in ref. 4. However, we elaborate here on the 7-point scale used for numerical assignment of validity rating, and consider how this should apply to data obtained from intracellular pathway tracing methods. For the latter methods (including the juxtacellular labeling method), the potential confounding influence of axons-of-passage labeling is avoided if the neurons of origin are identified and their labeled axons traced. This provides a substantial potential advantage for intracellular pathway tracing methods over methods involving extracellular deposition of pathway tracer molecules that exposes them to possible uptake by axons-of-passage. Accordingly, we have applied up to a 2-point increase to the pathway tracer validity rating for data so obtained (except for tracers with an existing rating of 6 or 7, for which a 1 point or no increase was applied respectively).

Regarding the numerical assignment of validity rating for pathway tracing methods involving the extracellular deposition of pathway tracer molecules, the highest rating (highest / high validity) assigned was 7 (currently assigned only to PHAL). Conversely, the lowest rating (lowest / very low validity) assigned was 1. The second to lowest rating (a value of 2) was applied to pathway tracers of unknown validity (due to a lack of experimental evidence). However, no data used in the network analysis for connections determined to exist in the current (or previous) work was derived from pathway tracers of "unknown validity", and so this did not contribute to the average validity calculation. In addition to lowest (1), highest (7), and unknown validity (2), the following general

descriptors for validity, with their corresponding ordinal value, apply to pathway tracers rated numerically between 3 and 6: low (3), low-moderate (4), moderate (5), moderate-high (6). For connections confirmed to not exist (that is, those that are absent), the validity rating is interpreted differently because, as noted previously (4), negative evidence obtained from pathway tracers of poor general validity may be highly valid; for example, see retrograde tracer evidence for hippocampal projections (in Dataset S1). Finally, it should be noted that all assignment of validity rating is evidence-based and subject to possible revision in the light of future evidence. For example, it is likely that more sensitive pathway tracing methods will be developed and reveal previously undetected connections. The extent to which new data will alter the results of network analysis of current data will depend on the how different the old and new connection matricies are, and on possible advances in network analysis methods. Thus, current connection reports should be considered provisional, and they are essentially heuristic because very few pathway tracers have been subjected to systematic investigation of their validity (general information on pathway tracing methods and detailed information on numerous pathway tracers is available at neuromeproject.org).

After reviewing and assessing the collated data as described, a single connection report was selected as best representative for each node in the connection matrix, and this single connection report used for network analysis. If more than one connection report was recorded for a given connection (depending on the availability of data in the primary literature; see Dataset S1) then, all else being equal, the connection report based on the most valid pathway tracing method was selected for network analysis. Lastly, the weight

of each selected connection report was used to populate a connection matrix (Dataset S2, worksheet "EB BM4 bins") that was used for subsequent network analysis.

The collation process was considerably aided by a dedicated data entry platform (Axiome M; created by J.D.H.) designed as a worksheet template for use with Microsoft Excel. The template facilitates speed and accuracy of data entry by utilizing data validation, formula calculations, conditional formatting rules, and a highly structured and guided user-friendly interface.

Connection Weight Scaling Methodology for Network Analysis. There is almost no quantitative data available in the literature regarding the density or weight of the rat macroconnections used in this analysis. Therefore, ranked qualitative connection weights from the literature were divided into 12 value categories. They are: no data, unclear, absent, axons-of-passage (in a gray matter region of interest; white matter tracts are not considered in this analysis), very weak, weak, weak to moderate, exists (present, but weight unreported), moderate, moderate to strong, strong, and very strong. For the purposes of our network analysis some of these values were binned. Reports of axons-of-passage were assigned a weight of "weak", and connections for which the reported value was entered as "exists" (present, but weight unreported) were assigned a weight of "moderate", and the category values of "unclear" and "no data" were assigned to the "absent" category. Thus, the set of ranked qualitative values used for network analysis included 8 values (7 weights, and 0 for absent) that were considered for our purposes to form an ordinal scale (see Dataset S2 worksheet "Key"). As justified previously (5), the

ranked qualitative connection weights were then transformed to approximately logarithmically spaced weights for network analysis using a 10^4 exponential scale.

Network Analysis Methods. Network analyses were carried out on the directed and log-weighted rat endbrain macroconnection matrix (Fig. 2) using tools collected in the Brain Connectivity Toolbox (www.brain-connectivity-toolbox.net). Multiresolution consensus clustering (MRCC) was performed using previously published methods and analysis tools (6). Detailed descriptions of most network measures and analysis procedures can be found in refs. 6 and 7. The complete set of 244 right and left rat cerebral cortical and cerebral nuclei gray matter regions are referred to as the nodes of the endbrain (telencephalic) network (EB2).

Multiresolution consensus clustering was carried out on weighted connection matrices, computing 100,000 event samples and applying a significance criterion of α=0.05. Modular partitions are derived by the Louvain algorithm, with the full set of samples covering a range of the resolution parameter between a minimal setting (where the number of modules is equal to 1) and a maximal setting (where the number of modules is equal to the number of nodes). An event sampling strategy is employed to ensure approximately equal coverage to all scales present in the network. The resulting multiresolution ensemble of partitions is then subjected to a divisive hierarchical clustering algorithm based on assessing significance of pairwise co-classification of nodes that delivers a tree (dendrogram) of nested partitions. The set of nested partitions is taken to represent the multiscale modular structure of the original network. We refer to the unique cuts in the resulting hierarchy as “levels”, with each level corresponding to a

unique partition. We note that this definition of the term "level" deviates from its common usage in the mathematics of hierarchical clustering.

We compared MRCC to our earlier approach to detecting modules in rat connectivity data while accounting for different spatial scales by varying the value of the resolution parameter (7). Using that approach, optimal module partitions were detected through repeated runs of the Louvain algorithm (8) for modularity maximization (9, 10). As in previous work (7) we used an objective function that included a resolution parameter γ (11) designed to detect modules that range over several spatial scales (or levels of resolution). Here, we varied γ over a range of $\gamma = [0.1\text{-}4.0]$, an interval that results in numbers of modules that range from very few (one or two modules) to very many (dozens), approximating the range of solutions delivered by MRCC. We optimized modularity 1,000 times for each setting (0.01 steps, yielding 391 distinct levels) of γ and encountered very little degeneracy among the partitions that were identified. Hence, at each level of γ, we selected the partition with the optimal value of the objective function without performing consensus clustering. Modular solutions obtained with this earlier approach were compared to modular solutions delivered by MRCC, by computing the distance between partitions as the variation of information.

Analyses of global network metrics such as clustering, path length and efficiency, and rich club organization were statistically evaluated by comparison to a degree-sequence preserving distribution of null models, as in previous work (3-5). Rewiring of the networks comprising the random null model followed a commonly used procedure equivalent to a Markov switching algorithm (12) that preserves the number of incoming and outgoing connections on all nodes. Spatial (for example, regional center-of-mass)

coordinates for the 244 regions in our analysis were unavailable and hence could not be incorporated into any of the analyses or null models employed here.

As in previous work (3-5, 13, 14), network hubs were determined based on aggregated rankings across several distinct nodal centrality measures: the node degree, node strength, node betweenness centrality, and closeness centrality. After ranking nodes on each of the four metrics, an aggregate "hub score" was determined for each node, expressing the number of centrality metrics for which each node appeared in the top 20%. Rich club organization (15) refers to the propensity of highly connected nodes (with high degree) also to be densely connected to each other, more so than expected by chance. Chance is defined by configured distribution of randomized networks corresponding to a null model, here generated with a degree-sequence preserving rewiring algorithm (described above). In all respects, our analysis proceeded as described in previous work (3-5).

Versioning connectomes. This study complements and extends our two previous studies of the intrinsic connections of the cerebral nuclei (3) and cerebral cortex (4). The data in one of the endbrain subconnectome datasets used in the current study (version 1.1 of the connectome dataset for the intrinsic connections of the cerebral nuclei—CNU) differs very slightly from the previously published dataset (version 1.0). Given that the differences are very minor they are not described here in detail (any differences in terms of the results of the network analysis would be negligible). Nevertheless, considering only the main basic connection (and related) numbers, and representative of their minor nature, differences between the current (version 1.1) and previous (version 1.0) (3)

intrinsic CNU connectome datasets include the following (the first number of each pair is the previous dataset): connections reported as present: 827 vs 830; connections reported as absent: 2806 vs 2810; total number of connection reports: 4067 vs 4259; connections for which no sufficient data was found: 372 vs 365; number of source articles cited: 40 vs 45. It is also worth noting that the intrinsic cerebral cortex (CTX) subconnectome used here does not differ significantly (in terms of the connection data) with that previous published (version 1.0) (4). Nevertheless, it has been edited for typographical errors and updated to the current version of our Axiome M workbook, accordingly it is referred to as version 1.1. Lastly, for the purposes of internal record keeping, we assign an identification (ID) number to each connectome dataset file and include that in the file name. This allows multiple updates to a file to be recorded simply by changing the associated file name ID number.

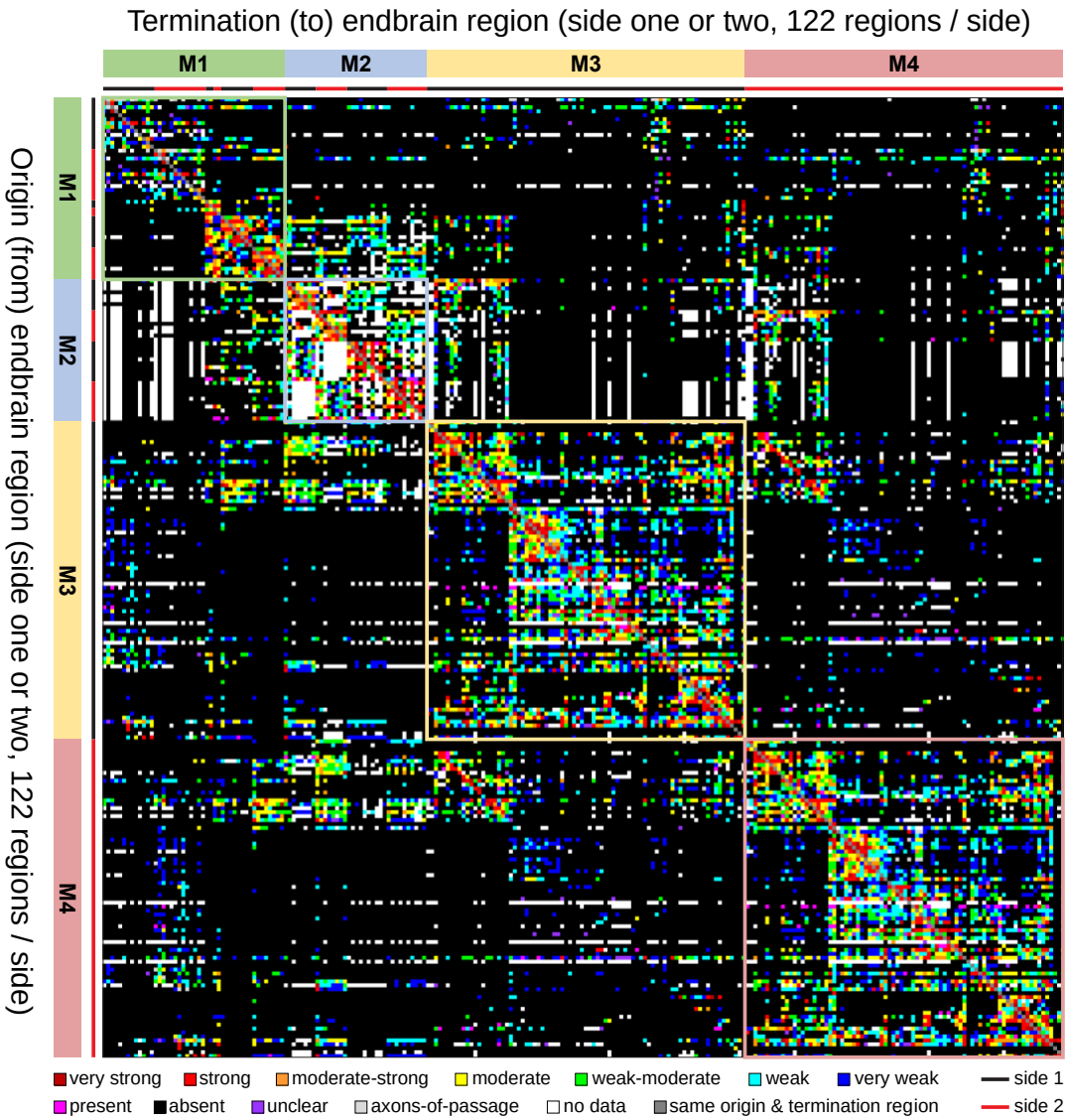


Fig. S1. Bilateral rat cerebral hemisphere (endbrain) macroconnectome. Directed and weighted monosynaptic macroconnection matrix with gray matter region sequence in a

modular or subsystem arrangement derived from a MRCC strategy (see Fig. 2). Key at bottom provides color codes for sides (1 & 2), and connection weights. For further information, including topographic (subsystem) and additional representations of the data, see Dataset S2 (for individual connection reports see Dataset S1).

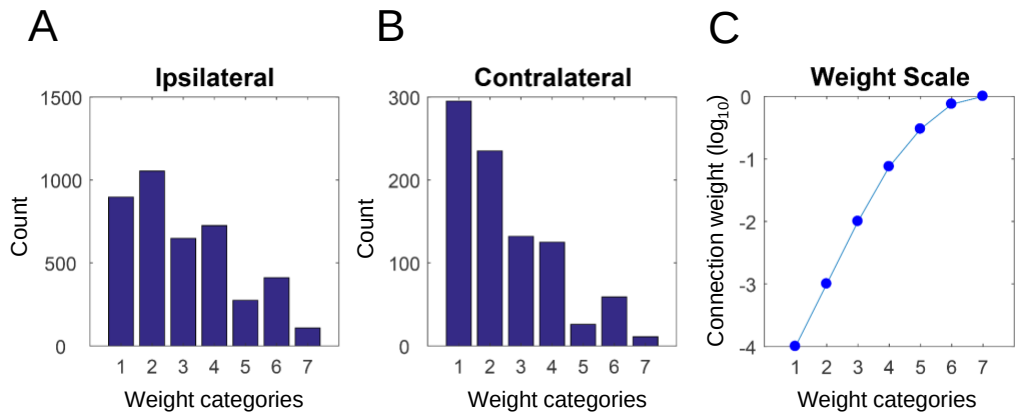


Fig. S2. (*A* and *B*) Distribution of weight categories for unilateral macroconnections (*A*) and for contralateral macroconnections (*B*) of the endbrain (as used in the current network analysis). (*C*) Weight scale used for weighted network analysis.

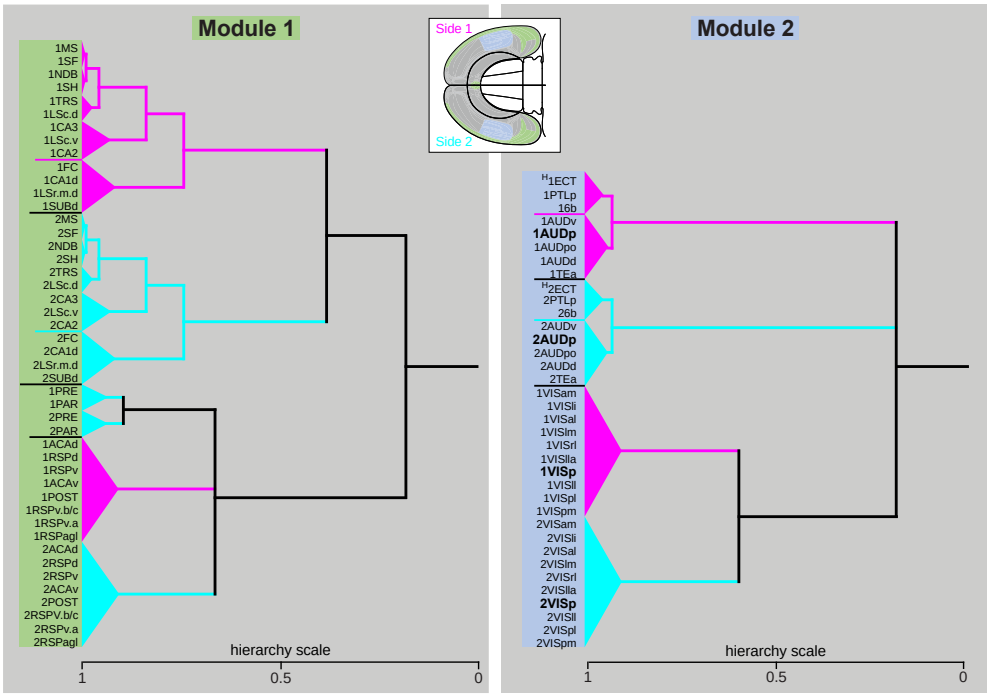


Fig. S3. The hierarchy branches for Module 1 (left) and Module 2 (right), expanded from the complete EB2 hierarchy tree shown in Fig. 3. For clarity, branches for parts of the hierarchy on Side 1 of each module are shown in magenta and parts of the hierarchy on Side 2 are shown in cyan. Module coloration follows Fig. 3 (see flatmap inset for location

of Module 1 (green) and Module 2 (blue), derived from Fig. 4). Putative hubs (H) are shown as superscripts on corresponding gray matter region abbreviations, which are defined in Dataset S2 and the flatmap representation of Fig. 6B. Primary sensory cortical area abbreviation text is larger and bold. For more on the hierarchy scale see Fig. 3 and ref. 6.

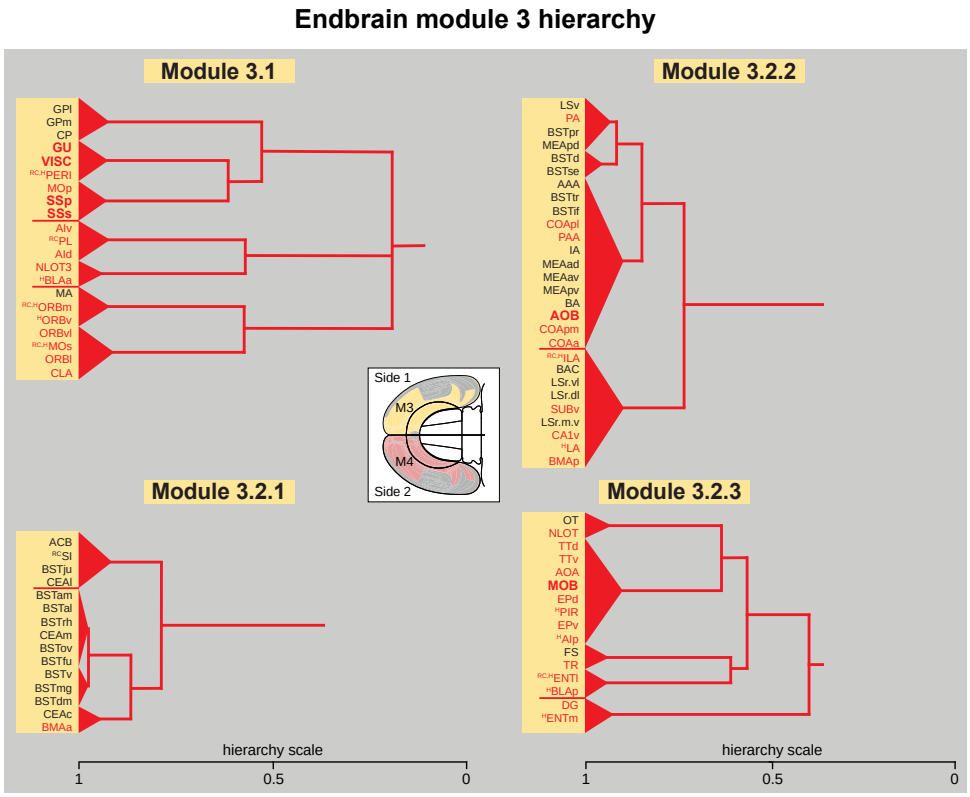
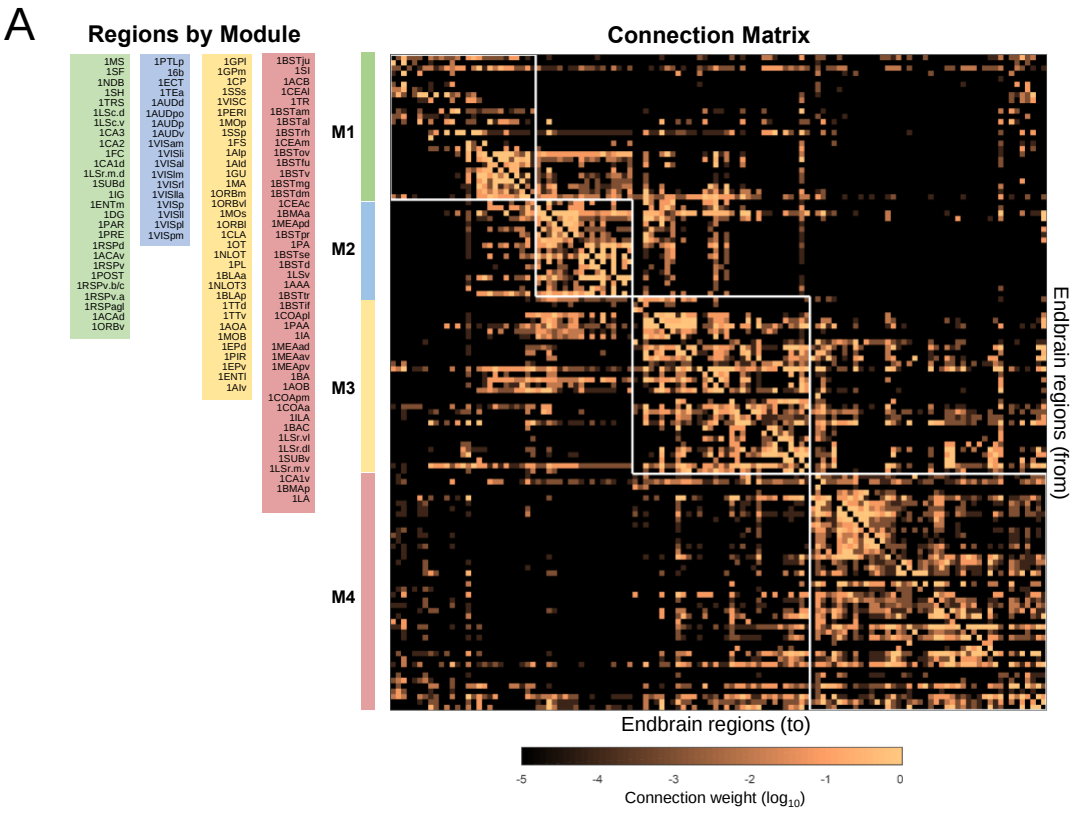


Fig. S4. The four main hierarchy branches of Module 3 (identical to its counterpart, M4, on the other side of the brain, indicated in pink on flatmap inset but not shown) expanded from the complete EB2 hierarchy tree shown in Fig. 3. Putative hubs (H) and central rich club members (RC) are shown as superscripts on corresponding gray matter region

abbreviations, which are defined in Dataset S2. Primary sensory cortical area abbreviations text is larger and bold; all cortical region abbreviations are red, and all cerebral nuclei abbreviations are black. The topological location of the two primary branches of M3 (M3.1 and M3.2) are shown in Fig. 6A. For more on the hierarchy scale see Fig. 3 and ref. 6.


A
Regions by Module
Connection Matrix
M1
M2
M3
M4
Endbrain regions (from)
Endbrain regions (to)
-5
-4
-3
-2
-1
0
Connection weight (log10)

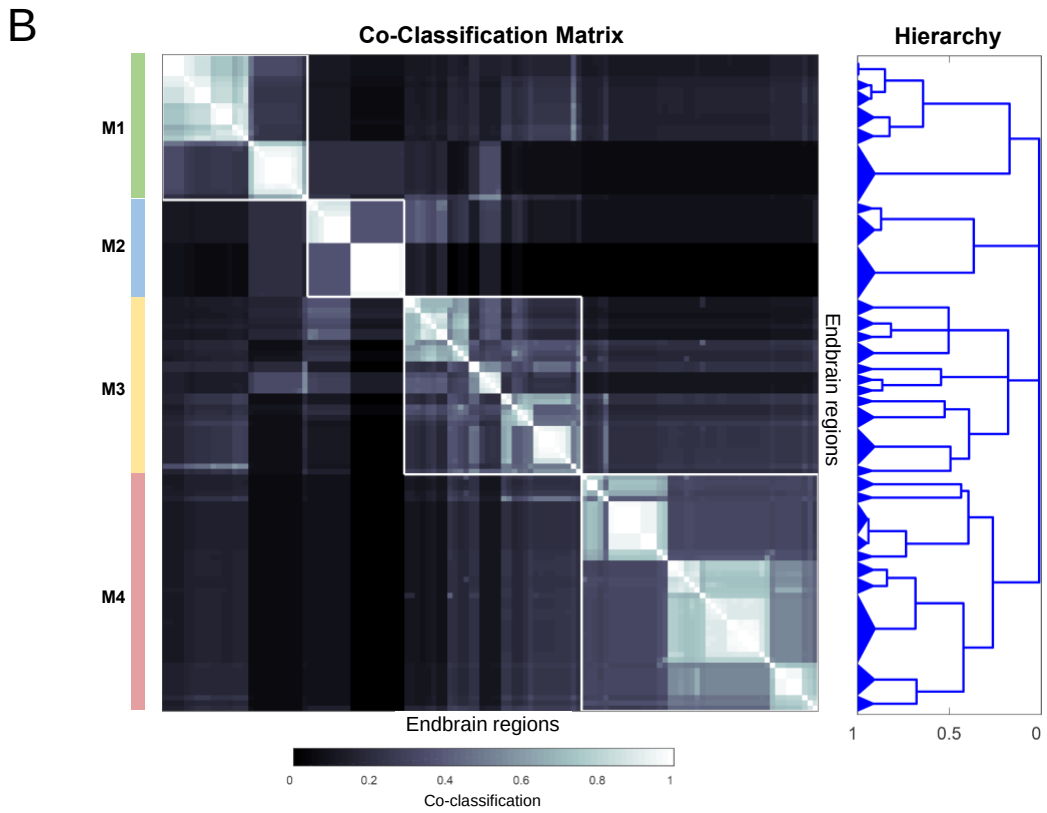

B
Co-Classification Matrix
Hierarchy
M1
M2
M3
M4
Endbrain regions
Endbrain regions
1
0.5
0
0
0.2
0.4
0.6
0.8
1
Co-classification

Fig. S5. Top-level solution (*A*), and complete MRCC hierarchy solution (*B*) for one side of the endbrain (EB1). Regions are arranged in the order they are assigned after MRCC analysis, performed with $\alpha = 0.05$ and 100,000 event samples (as in Figs. 2 and 3). Blades of the tree are arranged to most closely match modules derived from the EB2 MRCC analysis. The connectivity matrix (*A*) is displayed on a $\log_{10}$ scale. Co-classification levels in the co-classification matrix (*B*) are indicated by a linearly scaled co-classification index (bottom), that gives a range between 0 (no co-classification at any resolution) and 1 (perfect co-classification across all resolutions).

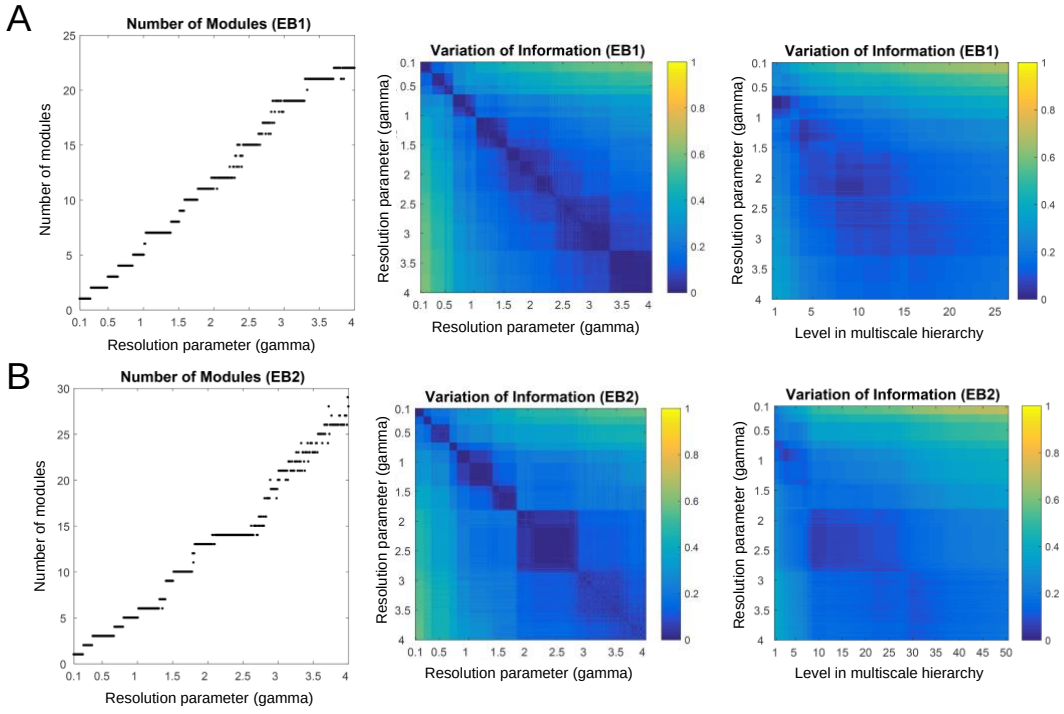


Fig. S6. Module detection across variations in the resolution parameter γ (as carried out in previous work; refs. 5-7), and comparison to MRCC modules as derived in the present paper, for unilateral (*A*) and bilateral (*B*) endbrain connection matrices. Left panels show optimal module numbers across variation in γ in the interval (0.1 - 4.0), in 0.01 increments, after modularity optimization was carried out 10,000 times. Middle panels show the variation of information (VI; a distance metric) across the entire γ range. Dark blue 'blocks' along the main diagonal indicate internally homogeneous and stable solutions that arise as γ increases. Right panels show the VI between modular partitions at all 391 levels of γ (y-axis) and all MRCC hierarchical levels (x-axis). Strong correspondence between modular partitions derived by the two approaches is indicated by low VI distances, especially along the main diagonal of the plot (the region where the

spatial resolution and number of modules match most closely). Note the convergence of solutions around $\gamma = 0.75$ and multiresolution consensus solution at level 1 (top level) for EB1 (*A*, right panel), and the convergence of solutions around $\gamma = 0.75$ and multiscale consensus solution at level 4 (top level) for EB2 (*B*, right panel).

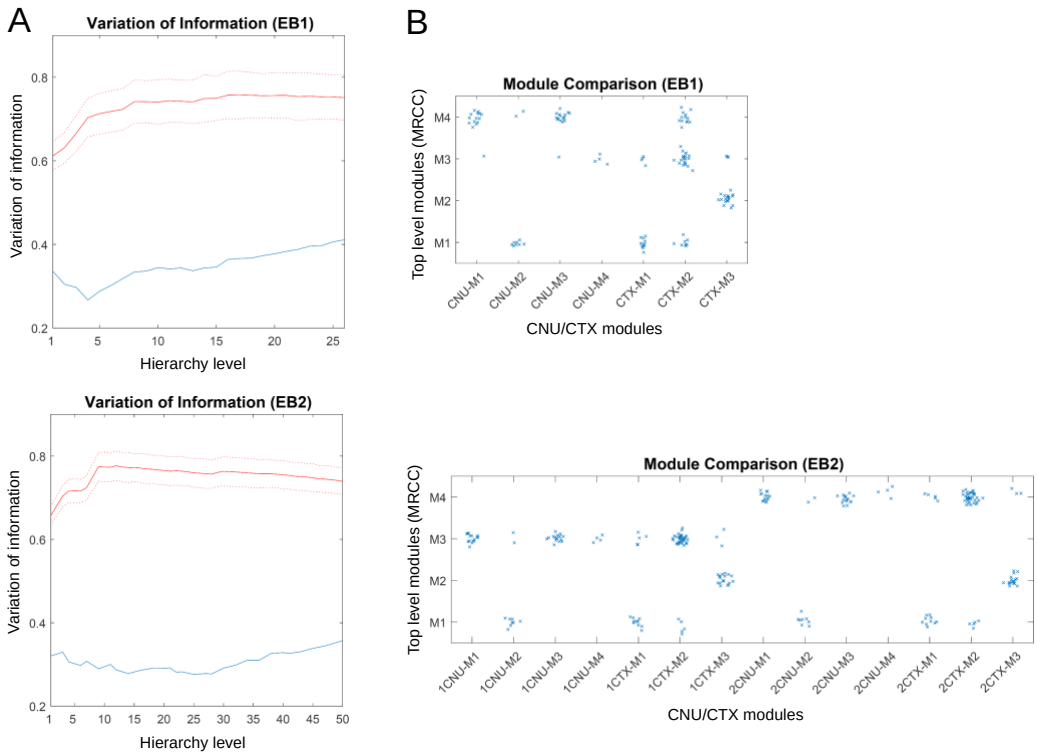


Fig. S7. Comparison of modules obtained from earlier analysis of ipsilateral subconnectomes CNU1 (4 modules) and CTX1 (3 modules) with those obtained by MRCC of EB1 and EB2. Plots in panels (*A*) show the VI between the set of 7 modules from CNU1/CTX1 on one side and the set of modules (at all hierarchical levels) obtained from MRCC of EB1 (top panel), and the set of 14 bilateral CNU2/CTX2 modules and the MRCC of EB2 (bottom panel). Note that the VI (blue line) is significantly below the levels of VI (mean and S.D.; red lines) obtained by chance (after 1,000 permutations of module assignments at each hierarchical level). Plots show minima for hierarchical MRCC solutions of 8 and 14 modules, respectively. Scatter plots in panels (*B*) show region-by-region module assignments to 7 (top) or 14 (bottom) CNU/CTX modules compared against the 4 modules (EB1 and EB2, top and bottom, respectively) obtained

by MRCC. Each blue dot represents one brain region. Dense clouds of dots indicate good agreement between module assignments for sizeable sets of regions, that is, illustrate homologies between modules obtained from different subconnectomes.

Movie S1. Movie displaying the co-classification matrix, hierarchical tree, and all levels of hierarchically nested modular solutions derived from multiresolution consensus clustering. The first part of the movie shows results obtained from the EB1 dataset and the second part shows results obtained from the EB2 dataset. All plotting conventions and scales are as in Fig. 3 and Fig. S5. The sequence of modular solutions starts at the top-level (level 1) of the hierarchy and ends at the bottom level (level 26 for EB1; level 50 for EB2). The tree is traversed from the top (right margin of the hierarchy), with the level currently shown indicated by a red line; corresponding modules are framed in bright red, with the previous level superimposed in dark red. The tree is traversed at linear speed such that solutions that correspond to longer branches (more persistent) occupy more time than those that correspond to shorter branches (less persistent).

Additional data table S1 (separate file)

The complete collated connection report dataset (including all reports used for network analysis) [link] The sequence of tabulated connection reports follows the topographic arrangement of regions in standard rat brain atlas (11). When multiple connection reports for a connection of interest were found, one was selected for network analysis (as described in *SI Materials and Methods*). All reports for each top-level subconnectome of the endbrain are listed together; the arrangement of the subconnectomes is as follows: CNU to CNU, CTX to CTX, CNU to CTX, CTX to CNU.

Abbreviations for pathway tracers: ARGM, autoradiographic method; BDA-3K/10K, biotinylated dextran amine, MW 3,000/10,000; CTB, cholera toxin B subunit; HRP, horseradish peroxidase; PHAL, *Phaseolus vulgaris*-leucoagglutinin; WGA-HRP, horseradish peroxidase conjugated to wheat germ agglutinin.

Additional data table S2 (separate file)

Data matrices in Microsoft Excel worksheet (spreadsheet) format for the ipsilateral and contralateral intrinsic connections of the rat endbrain derived from analysis of connection reports collated from the primary literature. [link] The Microsoft Excel workbook has 7 worksheets, each presented on a different workbook sheet. From left, the first worksheet provides a key to the workbook sheets and interpretation of the data therein. Worksheets 2-7 provide rat endbrain macroconnection data in binned and raw format arranged in one of three ways: (a) according to a 4-module top-level solution with sides 1 & 2 collated (as depicted in Fig. 2), (b) the same 4-module top-level solution with the sides separated, and (c) a topographic arrangement (11), with module assignment also indicated by cell coloration. Region numbers indicated in worksheets 2-7 corresponding to gray matter regions listed in topographic sequence in the rat brain atlas used (11). The data is read from the *y*-axis to the *x*-axis.

References

1. Swanson LW (2018) Brain maps 4.0—Structure of the rat brain: an open access atlas with global nervous system nomenclature ontology and flatmaps. *J Comp Neurol* 526(6):935-943.

2. International Commission on Zoological Nomenclature (1999) *International Code of Zoological Nomenclature* (International Trust for Zoological Nomenclature, London).
3. Swanson LW, Sporns O, Hahn JD (2016) Network architecture of the cerebral nuclei (basal ganglia) association and commissural connectome. *Proc Natl Acad Sci USA* 113(40):E5972-E5981.
4. Swanson LW, Hahn JD, Sporns O (2017) Organizing principles for the cerebral cortex network of commissural and association connections. *Proc Natl Acad Sci USA* 114(45):E9692-9701.
5. Bota M, Sporns O, Swanson LW (2015) Architecture of the cerebral cortical association connectome underlying cognition. *Proc Natl Acad Sci USA* 112:E2093-E2101.
6. Jeub LGS, Sporns O, Fortunato S (2018) Multiresolution consensus clustering in networks. *Science Reports* 8(1):3259.
7. Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* 529(3):1059-1069.
8. Blondel V, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech* P10008.
9. Sporns O, Betzel RF (2016) Modular brain networks. *Ann Rev Psychol* 67:613-640.
10. Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E* 69:026113.
11. Fortunato S, Barthélemy M (2007) Resolution limit in community detection. *Proc Natl Acad Sci USA* 104:36-41.

12. Maslov S, Sneppen K (2002) Specificity and stability of protein networks. *Science* 296:910-913.
13. 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.
14. Sporns O, Honey CJ, Kötter R. (2007) Identification and classification of hubs in brain networks. *PLoS ONE* 2:e1049.
15. Colizza V, Flammini A, Serrano MA, Vespignani A (2006) Detecting rich-club ordering in complex networks. *Nature Physics* 2:110-115.