High-throughput mapping of mesoscale connectomes in individual mice

Longwen Huang[a,*], Justus M Kebschull[a,b,*], Daniel Furth[a], Simon Musall[a], Matthew T Kaufman[a], Anne K Churchland[a], and Anthony M Zador[a]

[a] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA
[b] Watson School of Biological Sciences, Cold Spring Harbor, NY 11724, USA

*These authors contributed equally to the work as first authors.

Correspondence and requests should be addressed to Anthony Zador, zador@cshl.edu.

Abstract

Brain function is determined by connectivity among brain areas, and disruption of this connectivity leads to neuropsychiatric disorders. Understanding connectivity is essential to modern neuroscience, but mesoscale connectivity atlases are currently slow and expensive to generate, exist for few model systems, and require pooling across many brains. Here we present a method, muMAPseq (multisource Multiplexed Analysis of Projections by sequencing), which leverages barcoding and high-throughput sequencing to generate atlases from single animals rapidly and at low cost. We apply muMAPseq to tracing the neocortical connectome of individual mice, and demonstrate high reproducibility, and accuracy. Applying muMAPseq to the mutant BTBR mouse strain, which lacks a corpus callosum, we recapitulate its known connectopathies, and also uncover novel deficits. muMAPseq allows individual laboratories to generate atlases tailored to individuals, disease models, and new model species, and will facilitate quantitative comparative connectomics, permitting examination of how age, sex, environment, genetics and species affect neuronal wiring.

Keywords: connectomics, barcoding, sequencing, neuroanatomy, connectopathies, BTBR, comparative connectomics

Main

Neocortical function is determined by the connections among brain areas. Miswiring of these area-to-area connections leads to dysfunction, andmay be central to neuropsychiatric disorders such as autism and schizophrenia (*1, 2*). To understand normal and disrupted function in cortical and other networks, modern neuroscience depends on high-resolution anatomical atlases, in the same way that molecular biology depends on high-quality genome sequences and annotations. In particular, systematic mesoscale (area-to-area) connectivity atlases have transformed systems neuroscience (*3-5*), allowing researchers to simply look up anatomical pathways of interest, instead of performing time-consuming experiments.

Historically, atlases were compiled manually from results generated by many individual laboratories, each using somewhat different approaches and methods, and each presenting data relating to one or a few brain areas of interest in idiosyncratic formats (*6, 7*). Recent studies avoid the confounds inherent in inferring connectivity across techniques and laboratories by relying on a standardized set of tracing techniques (*3-5*). Unfortunately, high-quality mesoscale atlases still exist only for a small number of model organisms. Moreover, even the atlases that do exist typically only document mesoscale connectivity in a single genetic background, in a single sex, and at a single developmental time point. The reason that atlases remain scarce is that they are expensive and labor-intensive to generate, so fundamental questions of how connectivity varies and drives behavior—in different species, genotypes, sexes and across developmental stages, as well as in animal models of neuropsychiatric disorders—remain unanswered.

We therefore sought to develop a high throughput, low cost method to enable generation of a mesoscale atlas in a single laboratory over the course of a few weeks. In classical area-to-area connectivity tracing, a single tracer—for example, a virus encoding green fluorescent protein (GFP)—is injected into a "source" brain area. The brain is then dissected and imaged, and any region in which GFP-labeled axonal projections are observed is a projection "target". Fluorescence intensity at the target is interpreted as the strength of the projection. This procedure must be performed in a separate specimen for each source region of interest, since multiple injections within a single specimen would lead to ambiguity about which injection was the source of the observed fluorescence (Fig. 1A). Although multi-color tract tracing methods (*4*) achieve some multiplexing by increasing the number of fluorophores, the increase in throughput is modest because only a small number of colors can be reliably distinguished. Thus to obtain a mesoscale atlas, data must be pooled across many animals. To develop a higher throughput tool for mapping mesoscale connectivity, we therefore focused on multiplexing tracers for multiple source areas as a way of reducing the requisite number of animals, and associated costs, labor and animal-to-animal variability.

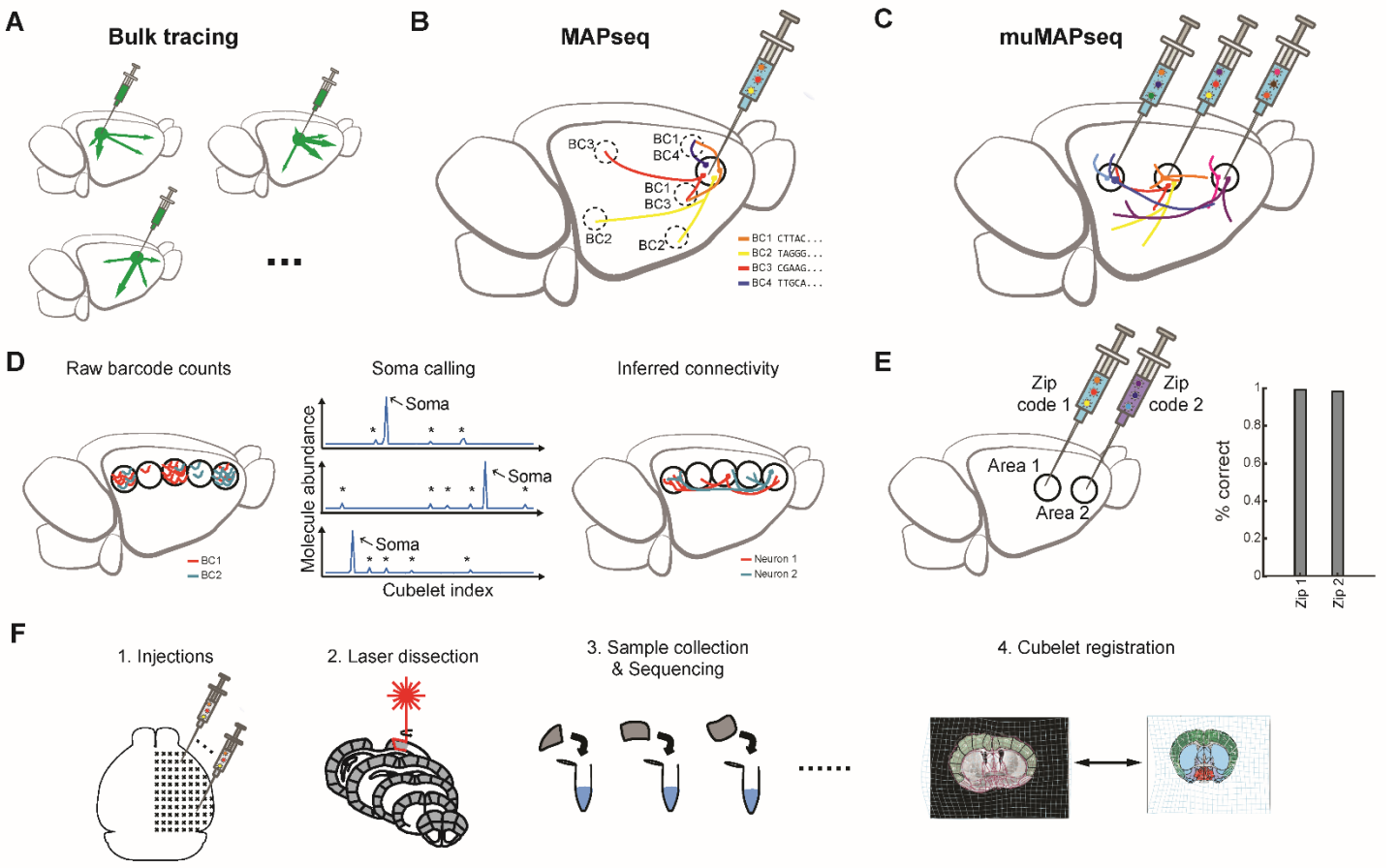

Fig. 1. Mapping mesoscale connectivity with muMAPseq. A Conventional bulk tracing method for connectivity atlas mapping. Individual brains are required for each source area. **B** Concept of MAPseq. Barcoded Sindbis virus is injected in an area of interest, and RNA barcodes from target areas of interest are extracted and sequenced to determine single cell projection patterns. BC, barcodes. **C** Concept of muMAPseq for connectivity atlas mapping. Whole-brain atlas can be derived from a single brain. **D**,**E** Soma calling strategy and its validation experiments. **F** muMAPseq pipeline.

The multi-site mapping strategy we developed, muMAPseq, builds on MAPseq (*8*). In MAPseq (Fig. 1B), multiplexed single neuron tracing from a single source was achieved by labeling individual neurons with easily distinguishable nucleotide sequences, or "barcodes," which are expressed as mRNA and trafficked into axonal processes. Because the number of nucleotide sequences, and therefore distinct labels, is effectively infinite—a short (30 base) random oligonucleotide has a potential diversity $4^{30}\approx10^{18}$—MAPseq can be thought of as a kind of "infinite color Brainbow" (*9*). Brain regions representing potential projection targets are microdissected into "cubelets" and homogenized, and the barcodes within each cubelet are sequenced to act as readouts of single cell projection patterns. The contribution of potential artifacts, including those due to degenerate labeling, fibers of passage, or non-uniform barcode transport, have been extensively quantified in previous work ((*8, 10, 11*); Supplementary Note 1), and shown to be minimal. MAPseq has now been validated using several different methods, including single neuron reconstruction, in multiple brain circuits (*8, 10, 11*).

A simple generalization of MAPseq to determine the mesoscale connectivity of many source areas in the same experiment would be to tag neurons with an additional area-specific barcode sequence—a "zipcode"—which could be used to identify the source (somatic origin) of each projection. In this approach, the overall strength of the projection from area 1 to area 2 would be determined by averaging the number of single neuron projections between those areas. Such an approach would, however, still be very labor intensive, because it would require the production, standardization and injection of hundreds of uniquely zipcoded batches of virus.

For muMAPseq, we therefore pursued a simpler strategy, requiring only a single batch of virus (Fig. 1C). We hypothesized that we could reliably determine the source of each projection using only sequencing, by exploiting the higher abundance of RNA barcodes in the somata compared with the axon terminals. According to this 'soma-max' strategy, the cubelet with the highest abundance of a given barcode of interest is assumed to be the soma (Fig. 1D). To validate this soma-max strategy, we injected two distinct viral libraries, each labeled with a known zipcode, into two separate but densely connected cortical areas (primary motor area and secondary motor area). We dissected both injection sites, and sequenced the barcodes present in each. Compared to the ground truth determined by the zipcode, the soma-max strategy correctly identified the soma location for 98.5% of all cells (Fig. 1E). Widespread injection of a single barcoded library, in combination with the soma-max strategy, should therefore allow the reconstruction of mesoscale connectivity networks in a single experiment.

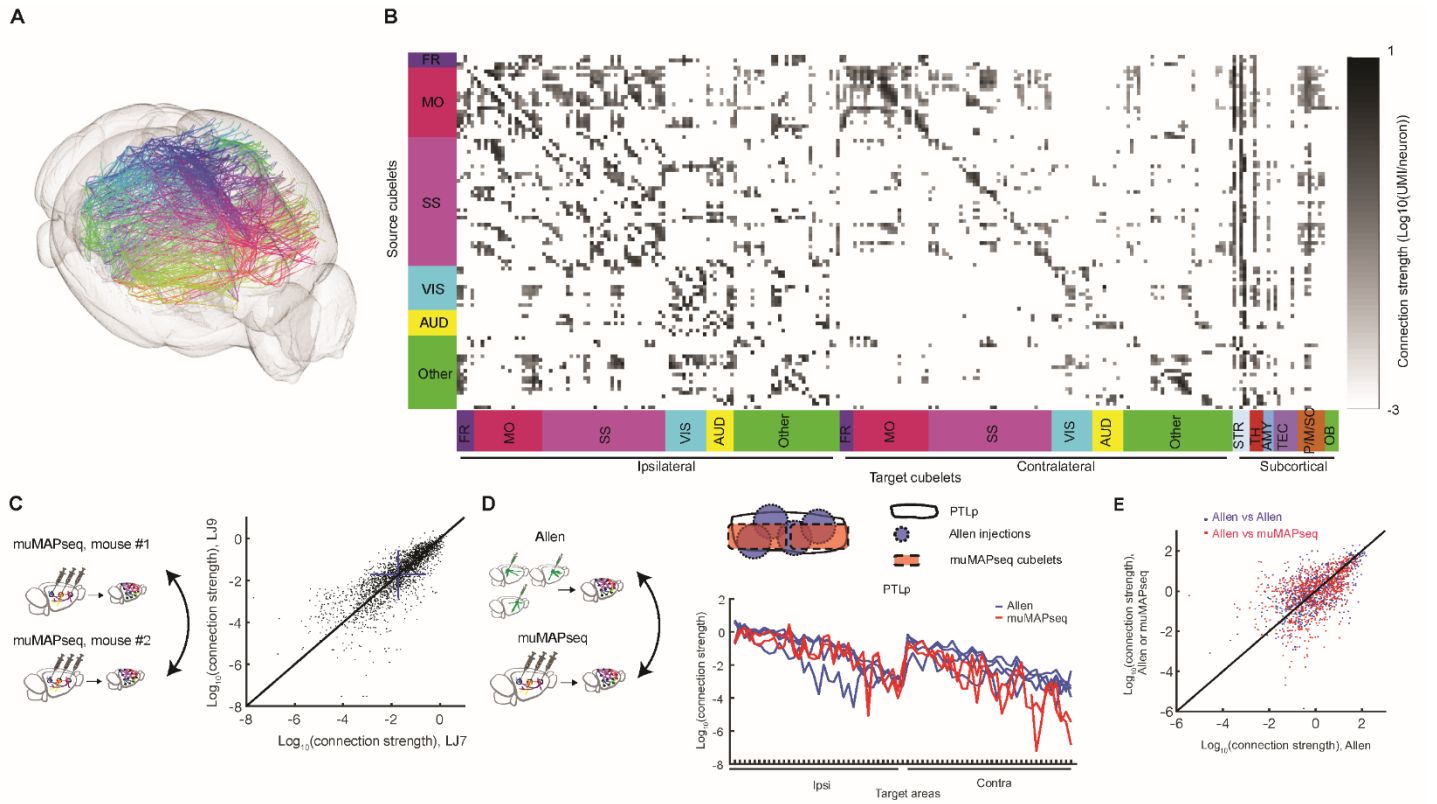

Fig. 2. Brain-wide corticocortical projectome mapped by muMAPseq and its validation. A,B Cubelet-to-cubelet connectivity atlas in mouse LJ7. **A** depicts connections in 3D space as lines following the most likely fiber tracts, coloured continuously to reflect the origin of each connection. See also Supplemental Video 1. In **B**, each row is a source cubelet, and each column is a target cubelet. Cubelets are assigned to their primary brain area. FR, frontal areas; MO, motor areas; SS, somatosensory areas; VIS, visual areas; AUD, auditory areas. **C** Reproducibility of brain area-to-brain area connection maps between LJ7 and LJ9. R = 0.8318, linear regression $p < 10^{-100}$. The unity line is in black. Blue bars show mean±s.d. **D,E** Agreement between muMAPseq and Allen Connectivity Atlas. **D**, an example comparison of PTLp; **E** the whole network. Z-scores are plotted for each axis; Allen vs Allen, R = 0.5719, linear regression $p < 10^{-100}$; Allen vs muMAPseq, R = 0.5161, linear regression $p < 10^{-100}$. The unity line is in black.

We first applied muMAPseq to generating a mesoscale atlas of the cortex of the adult male C57BL/6 mouse, for which there exist reference data sets (*3, 4*). To do so, we tiled the entire right hemicortex of each mouse with barcoded virus by making over 100 penetrations (3-6 injections/penetration at different depths) in a grid pattern with 500 µm edge length (Supplemental Table 2). Forty-four hours after viral injection, we cryosectioned the brain into 300 µm coronal slices, and used laser dissection to generate cortical (arc length ~ 1 mm) and subcortical cubelets (Fig. 1F, Fig. S1C,D). The locations of all cortical cubelets were registered to the Allen Reference Atlas (2011 version, Fig. 1F). We then quantified the number of each barcode sequence in each cubelet (Fig. S1D).

In two adult male C57BL/6J mice we mapped the connections from 106±5 (mean±S.D.) source cubelets to 261.5±0.7 target cubelets (230.5±0.7 cortical, 31±0 subcortical). All dissected cubelets were potential targets; source cubelets were defined as the subset of all cubelets containing barcoded somata. From each source cubelet we obtained the sequences of several hundred somata (625±683) located therein, as well of projections from several thousand ($1.5\times10^3\pm1.3\times10^3$) neurons with somata located elsewhere. We aggregated these single neuron data to calculate area-to-area axonal projection strengths (Fig. 2A,B, Fig. S4A). Thus the strength of the projection from source cubelet X to target cubelet Y was defined as the number of barcodes per soma in target Y originating from somata in source region X. We also estimated a confidence bound on our estimate of the strength of each connection (Fig. S3; Supplementary note 2). All self-self projection strengths were set to 0. In addition, all the neighbor-projection strengths were also set to 0 to minimize false positives due to dendritic innervation of neighboring cubelets. Although in principle muMAPseq data could have been used to determine single neuron projection patterns, in practice sequencing depth and other experimental details

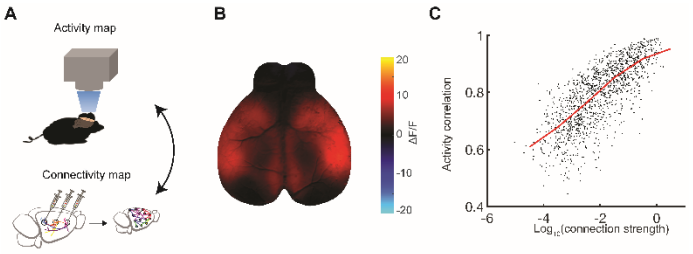

Fig. 3. MuMAPseq predicts functional connectivity. A muMAPseq connectivity compared with cortex-wide Ca^{2+} imaging. **B** A single frame example of cortex-wide wide-field calcium imaging in a behaving animal. **C** Activity correlation between pairs of cubelets vs. reciprocal connection strengths between them. The median line is in red. n = 7225 pairs, R = 0.7637, linear regression $p < 10^{-100}$.

precluded such an analysis for this dataset (Supplementary Note 4.1).

To fulfill its potential as a high-throughput method for determining mesoscale connectivity, muMAPseq must be both reproducible and accurate. The reproducibility of muMAPseq was high. Estimated connection strengths were similar for the two C57BL/6J brains tested (R = 0.83; Fig. 2C; Supplementary Note 3). Differences between the measured connections in the two samples arose from a combination of technical and biological variability. Major sources of technical variability likely include differences in injections and in dissection borders. We minimized biological variability by comparing subjects of the same age, sex and genetic background, but since the actual degree of animal-to-animal variability in cortical connections is unknown, these results represent an upper bound on the technical variability of muMAPseq.

To assess the accuracy of muMAPseq, we compared our results to the Allen Connectivity Atlas (Supplementary Table 2 in ref (*3*)), which was generated using conventional fluorophore-based techniques. The relationship between the ~100 cortical muMAPseq cubelets (defined by dissection) and cortical "areas" (defined by the Atlas) was not one-to-one: Each area typically spanned several cubelets, and each cubelet contributed to several areas. We therefore limited the comparison to the subset of cubelets that resided primarily (>70%) in a single source area (Supplementary Note 3). The agreement between muMAPseq and the Allen Atlas was excellent (R=0.51); indeed, the agreement was comparable to inter-experiment variability within the Allen Atlas (R=0.57, Fig. 2D,E). This not only confirms that potential MAPseq artifacts (from e.g. degenerate labeling, fibers of passage, non-uniform barcode trafficking) are as minimal in muMAPseq as indicated by other validations (*8, 10, 11*), but also suggests that muMAPseq could be used reliably for mapping mesoscale connectivity.

Anatomical connectivity results in correlations neural activity. To assess whether muMAPseq—a purely anatomical measure—could predict such "functional connectivity," we compared our results to a measure of correlated neural activity based on mesoscale wide-field imaging. We performed cortex-wide wide-field calcium imaging in awake Emx-Cre; Ai93; LSL-tTA mice engaged in a visual/auditory d ision task (Fig. S5A) (*12*). After brain registration, the activity of each cubelet was calculated as the mean activity over all its pixels. The connection strength determined by muMAPseq predicted the correlation of activity measured by calcium imaging (R = 0.7637, Fig. 3). The remarkable agreement between these two very different large-scale methods further indicates that muMAPseq provides a functionally relevant measure of cortical connectivity, and opens up the possibility of systematically relating structure and function in different animal

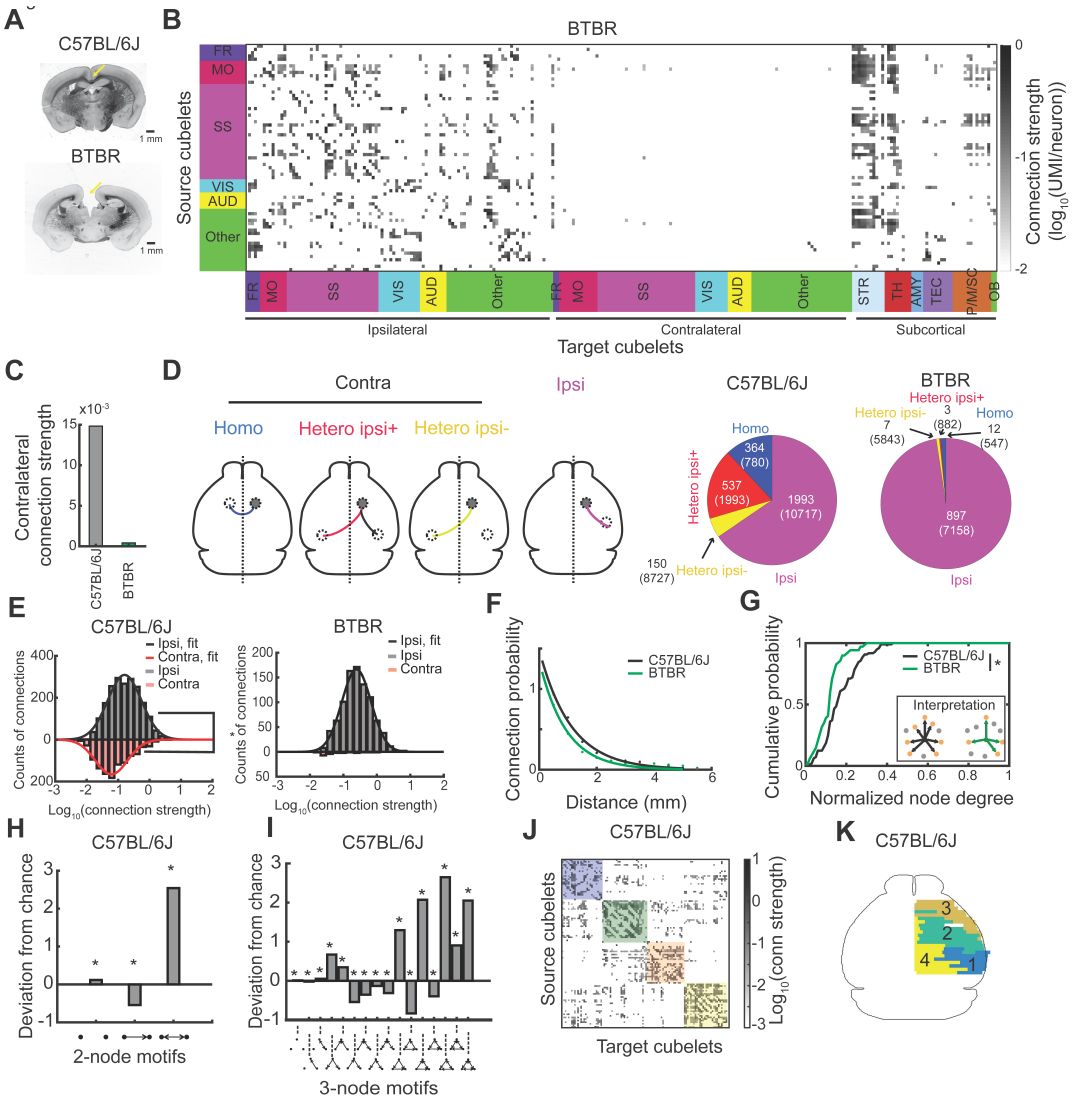

Fig. 4. Comparison of the BTBR and C57BL/6J cortical connectivity. A Bright field images of a C57BL/6J brain slice and a BTBR brain slice. **B** Cubelet-to-cubelet connection matrix showing connection strengths in the BTBR mouse. **C** Quantification of contralateral connection strengths in C57BL/6J (LJ7) and BTBR. **D** Nonzero connections in the C57BL/6J (LJ7) and BTBR. **E** Distributions of ipsilateral/contralateral corticocortical connection strengths in C57BL/6J (LJ7) and BTBR. *, $p < 0.001$, Kolmogorov-Smirnov test. **F** Distance-dependent ipsilateral connection probability in the C57BL/6J (LJ7) and BTBR. **G** Cumulative probability distributions of normalized node degree in C57BL/6J (LJ7) and BTBR. $p < 0.0001$; Kolmogorov-Smirnov test. **H,I** Abundance of 2-node and 3-node motifs in cortical network in C57BL/6J (LJ7) compared to randomly generated networks. *, $p < 0.001$. **J** Sorted cubelet-to-cubelet connection matrix based on modules in C57BL/6J (LJ7). **K** Connection-based modules in C57BL/6J (LJ7).

models.

A key advantage of muMPAseq is that it allows for rapid and systematic comparison of brain connectivity between model systems. We applied muMAPseq to compare the reference cortical connectome (Fig. 2B) to that of a BTBR mouse, an inbred strain lacking the corpus callosum and displaying social deficits (*13-15*) (Fig. 4A; Fig. S6A). Most strikingly and as expected, muMAPseq revealed a nearly complete absence of commissural cortical connections (Fig. 4B,C). In the C57BL/6J, commissural connections constitute 35.2±1.0% of total connections, whereas in BTBR the percentage is 2.3% (Fig. 4D; Supplementary Note 4.6; the few remaining nonzero commissural connections in BTBR were found exclusively in target cubelets close to the midline, and likely represented dissection error and contamination from the ipsilateral hemisphere). Thus, the known connectopathies of the BTBR strain is recapitulated with muMAPseq.

We next sought to uncover novel wiring deficits in the BTBR mouse. To establish a baseline for comparison, we first applied a range of network analyses to the reference C57BL/6J brain in the cubelet coordinate system. Network analyses of muMAPseq-derived mesoscale connectivity differ from previous studies, as the natural coordinate frame is given by regularly spaced cubelets and all data were obtained from a single individual. They thus provide a new avenue for brain network anlysis.

Consistent with previous reports (3), in the C57BL/6J, connection strengths were well fit by a log-normal distribution (Fig. 4E, left; Fig. S8D). The decay of connection strength with distance (Fig. S8A,E) was fit with a double exponential (λ_1 = 0.23±0.06 mm, λ_2 = 3.99±3.91 mm, 95% confidence intervals), and connection probability (Fig. S8A,E) with a single exponential (λ = 0.92±0.16 mm, 95% confidence interval). The distribution of ipsilateral connection strengths in the BTBR was similarly fit by a log-normal distribution (Fig. 4E, right), and the inferred ipsilateral area-to-area connections were not grossly disrupted (Fig. S8H) (REF). However, the probability of ipsilateral connections between cubelets was significantly reduced relative to controls (Fig. 4F). Consistent with this, the ipsilateral node out-degree—the typical number of output connections per cubelet—was also reduced (Fig. 4G). The BTBR ipsilateral connectivity is therefore characterized by a smaller number of connections than the C57BL/6J.

We next analyzed the topological properties of the ipsilateral cortical networks. By decomposing the network into small motifs containing 2 or 3 cubelets, and quantitatively comparing the abundance of these motifs to randomly generated networks, we found that in the C57BL/6J, the fraction of 2-cubelet motif with a reciprocally connected pair was greater than the null model, and densely connected 3-cubelet motifs were also significantly overrepresented (Fig. 4H,I; Supplementary Note 4.4). Furthermore, four network modules—regions of the brain within which connections are dense, and which may reflect functional units—were revealed by connection-based clustering of cubelets in the C57BL/6J (Fig. 4J,K Supplementary Note 4.5). Surprisingly, similar results were found in the BTBR (Fig. S9E; Fig. S10J-L), suggesting that these high-order topological properties were largely maintained in the BTBR strain. Thus the BTBR ipsilateral connectivity is characterized by a smaller number of connections and reduced node-degree than the C57BL/6J, while nonetheless maintaining much of its higher-order structure.

We have described muMAPseq, a high-throughput method to determine mesoscale connectivity networks in individual animals, rather than in hundreds, and apply it to mapping the cortical network of both C57BL/6J animals and a mouse model with aberrant connectivity and behavioral phenotypes. Compared with conventional mapping methods, muMAPseq is faster and less expensive, but at the cost of the fixed spatial resolution imposed by dissection into cubelets. Single axon spatial resolution can be achieved with *in situ* sequencing (*10*), and may be scaled up to whole brain analysis as throughput increases. MuMAPseq allows individual labs to generate relevant mesoscale connectivity atlases for their particular model system, thus providing a systematic foundation for studying circuits in mouse models in which connectivity deviates from that of C57BL/6J males, as well as in a wide range of non-standard animal model systems, including peromyscus, voles, marmosets and others. Beyond the enabling impact of such atlases, we expect that—akin to the insights gained from comparative genomics—comparison of muMAPseq-derived connectivity networks across mouse models of neuropsychiatric disorders, strains, sexes and even species will reveal fundamental principles of brain architecture and information flow.

End Matter

Author Contributions and Notes

L.H. and J.M.K. performed the muMAPseq experiments. S.M., M.T.K. and A.K.C. designed and performed the Ca^{2+} imaging experiments. D.F. designed whole brain visualizations. L.H., J.M.K and A.M.Z. designed the study, analyzed the data and wrote the paper.

A.M.Z. is founder and equity owner in MapNeuro.

This article contains supporting information online.

Acknowledgments

We would like to thank Pavel Osten, Hongwei Dong and Liqun Luo for comments on the manuscript, and the Finkelstein Lab for providing the word Biorxiv-Template used to typeset this article.

Funding sources: National Institutes of Health (5RO1NS073129 to A.M.Z., 5RO1DA036913 to A.M.Z.); Brain Research Foundation (BRF-SIA-2014-03 to A.M.Z.); IARPA (MICrONS D16PC0008 to A.M.Z.); Simons Foundation (382793/SIMONS to A.M.Z.); Paul Allen Distinguished Investigator Award (to A.M.Z.); PhD fellowship from the Boehringer Ingelheim Fonds (to J.M.K.); PhD fellowship from the Genentech Foundation (to J.M.K); Swiss National Science Foundation (to S.M.); Pew Charitable Trusts (to A.K.C.); Simons Collaboration on the Global Brain (to A.K.C. and M.T.K.).

References

1. D. H. Geschwind, P. Levitt, Autism spectrum disorders: developmental disconnection syndromes. *Curr Opin Neurobiol* **17**, 103-111 (2007).
2. M. Kubicki *et al.*, A review of diffusion tensor imaging studies in schizophrenia. *J Psychiatr Res* **41**, 15-30 (2007).
3. S. W. Oh *et al.*, A mesoscale connectome of the mouse brain. *Nature* **508**, 207-214 (2014).
4. B. Zingg *et al.*, Neural networks of the mouse neocortex. *Cell* **156**, 1096-1111 (2014).
5. N. T. Markov *et al.*, A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cereb Cortex* **24**, 17-36 (2014).
6. D. J. Felleman, D. C. Van Essen, Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* **1**, 1-47 (1991).
7. M. Bota, O. Sporns, L. W. Swanson, Architecture of the cerebral cortical association connectome underlying cognition. *Proc Natl Acad Sci U S A* **112**, E2093-2101 (2015).

8. J. M. Kebschull *et al.*, High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA. *Neuron* **91**, 975-987 (2016).
9. J. Livet *et al.*, Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56-62 (2007).
10. X. Chen, J. M. Kebschull, H. Zhan, Y.-C. Sun, A. M. Zador, Spatial organization of projection neurons in the mouse auditory cortex identified by in situ barcode sequencing. *bioRxiv*, (2018).
11. Y. Han *et al.*, The logic of single-cell projections from visual cortex. *Nature* **556**, 51-56 (2018).
12. S. Musall, M. T. Kaufman, S. Gluf, A. K. Churchland, Movement-related activity dominates cortex during sensory-guided decision making. *bioRxiv*, (2018).
13. L. R. Fenlon *et al.*, Formation of functional areas in the cerebral cortex is disrupted in a mouse model of autism spectrum disorder. *Neural Dev* **10**, 10 (2015).
14. H. G. McFarlane *et al.*, Autism-like behavioral phenotypes in BTBR T+tf/J mice. *Genes Brain Behav* **7**, 152-163 (2008).
15. D. Wahlsten, P. Metten, J. C. Crabbe, Survey of 21 inbred mouse strains in two laboratories reveals that BTBR T/+ tf/tf has severely reduced hippocampal commissure and absent corpus callosum. *Brain Res* **971**, 47-54 (2003).