doi:10.1038/nature26159

Supplementary Notes

Fluorescence-based single neuron reconstructions

Supplementary Note 1: Abrupt terminations in GFP filled cells

The mapping of axonal processes by fluorescence-based single neuron reconstructions relies on the complete and roughly homogeneous filling of all axonal processes with GFP protein. If one branch is not completely filled, its full reach will not be detected, and will thus cause a false negative projection.

To minimize such false negatives in our dataset, we took a number of precautions both during the experiment and during data analysis. First, we used a strong promoter to express GFP protein within each electroporated cell, thus maximizing the amount of fluorescent protein expressed in each cell. We then monitored the expression of GFP by epifluorescence imaging through a cranial window, to determine the optimal time of GFP expression on a per cell basis, aiming for high GFP expression but avoiding the cytotoxic effects of too high transgene expression. After fixation and imaging, we assessed the health of every cell by the integrity of the axonal processes – a very sensitive indicator of cell health – and discarded any cell with fragmented axonal processes before tracing. Only 7/71 traced cells exhibit minor blebbing of the axon, such that we can conclude that axonal retraction due to poor health of a cell is not a contaminating factor in our dataset. Finally, we stringently screened for incomplete fills in our imaging datasets in cells that appeared healthy. Incomplete fills are generally accepted to show a gradual fading of fluorescence along the length of a process, as is intuitive for an incomplete fill by a diffusing fluorescent molecule. Such gradual terminations are very rare in our dataset (5 fading terminations detected in total from 52 cells with approximately 100 terminations each), suggesting a low false positive rate. In addition to the gradual terminations, we also observed abrupt terminations, where a process is brightly labeled and simply stops without branching or fading fluorescence either in the white or gray matter (Extended Data Fig. 2**a,b**). Such abrupt terminations are a common occurrence in single cell fills, irrespective of the exact method used for single cell filling, and may thus reflect the true extent of the axon (see for example refs[1,2]). We find that the vast majority of abrupt terminations observed in our dataset are more proximal to their cell body than the farthest filled process of the same cell (Extended Data Fig. 2**d**). These data suggest that abrupt terminations in our data are not the result of a distance dependent failure to fill an axonal branch, and are in agreement with the prevalent interpretation of abrupt terminations as real. Nonetheless, we decided to exclude all cells with abrupt terminations from our analysis as a cautionary measure to produce the best possible dataset to act as a gold standard for a comparison to our MAPseq dataset. Note, that all conclusions presented in this work, in particular the high number of target areas per V1 neuron, are robust to the ex/inclusion of abrupt terminations (Extended Data Fig. 2**e**).

MAPseq

Like any other method, MAPseq is subject to false positives (detection of an extra, artefactual projection) and false negatives (failing to identify a real connection). Please refer to ref[3], for a detailed discussion of the effect of fibers of passage, co-infections, infection of more than one cell with the same barcode sequence, and various other sources of false negatives and false positives. Below, we briefly discuss the most important considerations.

Supplementary Note 2: Unique labeling of cells by viral infection

As described in more detail previously[3], in MAPseq we deliver barcodes to cells by random viral infection. In the simplest scenario we aim to deliver one unique barcode per labeled cell, such that each cell can unequivocally be identified by a single sequence. In practice however, there are two scenarios that deviate from this simple model.

On the one hand, we might infect cells with more than one virus particle and thus label each cell with more than one barcode sequence. Such multiple labeling results in MAPseq overestimating the total number of traced cells, but does not result in a false projection pattern and maintains the relative proportions of projection types. Such multiple labeling will therefore not lead to any false positive results.

On the other hand, degenerate labeling, that is labeling more than one cell with the same barcode sequence, produces artificial projection patterns that result from the merged projection pattern of all the cells labeled with the same barcode. We avoid degenerate labeling in MAPseq by infecting cells with a very large virus library. The fraction of uniquely labeled cells for a given size of virus library with an empirically determined barcode probability distribution can be described by

$$F = 1 - \sum_{i=1}^{N} p_i (1 - (1 - p_i)^{k-1}) ,$$

where p_i is the probability of barcode i=1..N to be chosen from the virus library, k is the number of labeled cells and N is the total number of barcodes in the library[3]. Given the size of the library used in this study (~10^7 distinct barcode sequences) and the number of cells traced per brain (~300), the vast majority of cells (>99 %) will be uniquely labeled. In previous work[3] we validated these theoretical predictions by multiple independent experimental methods.

Supplementary Note 3: Other sources of false positives and negatives.

Beyond errors introduced by degenerate labeling (see above), MAPseq is subject to false positives and negatives from other sources. False negatives are introduced into the dataset when the strength of a real projection falls below the detection floor of MAPseq. Conversely, MAPseq false positives

are introduced when barcodes are detected in areas in which they were not originally present in (e.g. by sample cross-contamination).

Several lines of evidence suggest that MAPseq false negative and positive rates are low. In previous experiments[3], we measured the efficiency of MAPseq to be very similar to that of Lumafluor retrobeads (91.4 ± 6 % (mean ± std error))[3], and therefore concluded that MAPseq false negative rates are comparable to those of other, well established methods. Similarly, we found MAPseq false positive rates to be low (1.4 ± 0.8 % (mean ± std error))[3].

In the present study we improve on these previous estimates by comparing MAPseq data directly to the gold standard of single neuron tracing. To do so, we first used a bootstrapping procedure to measure the minimum pairwise cosine distances between each member of a randomly sampled set of MAPseq neurons and the remaining MAPseq neurons. We then measured the minimum pairwise cosine distance between the fluorescence-based single neuron reconstructions and the remaining MAPseq neurons. As a negative control, we further measured the minimum pairwise distance between a set of random neurons (with their projection strengths drawn from a uniform distribution) and the remaining MAPseq neurons. We then compared the distribution of minimum distances for the three sets of measurement and found that while the MAPseq-to-MAPseq and fluorescence-based reconstructions-to-MAPseq distributions are statistically indistinguishable (two sample Kolmogorov-Smirnov test, $p=0.9439$), both the MAPseq-to-MAPseq and fluorescence-based reconstructions-to-MAPseq distributions are statistically different from the random neuron-to-MAPseq distribution (Fig. 2**d**; two sample Kolmogorov-Smirnov test, $p=2.75\text{x}10^{-4}$ and $p=8.76\text{x}10^{-5}$, respectively). Taken together, these results indicate that the statistics of projections inferred by MAPseq are indistinguishable from those obtained by fluorescence-based single neuron reconstructions.

Supplementary Note 4: False negatives cannot explain the observed structure in the projectional dataset.

In order to investigate the effect of the MAPseq false negative rate on our analysis of over- and underrepresented projection motifs, we simulated how our results changed by removing or adding projections to the MAPseq dataset, while keeping the average area projection probabilities fixed. As expected, randomly adding or removing connections did not produce structure in the dataset, consistently increasing p-values (i.e. making them less significant) for all six significantly over- or underrepresented projection motifs reported in Fig. 3 (Extended Data Fig. 2**f**).

Supplementary Tables

Area 1	Area 2	Area 1 projecting cells	Area 2 projecting cells	Area 1&2 projecting cells	% double labeled cells relative to total	% double labeled cells relative to Area 1	% double labeled cells relative to Area 2
Li	LM	43	323	34	10.2	79.1	10.5
Li	AL	43	207	4	1.6	9.3	1.9
Li	PM	43	228	20	8.0	46.5	8.8
Li	AM	43	30	0	0.0	0.0	0.0
Li	RL	43	18	0	0.0	0.0	0.0
LM	AL	323	207	132	33.2	40.9	63.8
LM	PM	323	228	71	14.8	22.0	31.1
LM	AM	323	30	10	2.9	3.1	33.3
LM	RL	323	18	6	1.8	1.9	33.3
AL	PM	207	228	31	7.7	15.0	13.6
AL	AM	207	30	7	3.0	3.4	23.3
AL	RL	207	18	7	3.2	3.4	38.9
PM	AM	228	30	22	9.3	9.6	73.3
PM	RL	228	18	7	2.9	3.1	38.9
AM	RL	30	18	2	4.3	6.7	11.1

Supplementary Table 1: Simulated double retrograde tracing based on MAPseq anterograde tracing data. We determined whether any given MAPseq neuron targeted any one area using the same projection criterion used for the analysis in Fig. 3. For any pair of the six higher visual areas analyzed using MAPseq, we then determined the number of cells that projected to either area in the pair, or to both — effectively simulating double retrograde tracing from the two areas in the pair to V1. We here show the raw counts of cell projecting to each area, and the percentages of cells that project to the indicated pairs of areas, i.e. “double labeled” cells.

References

1. Kita, T. & Kita, H. The subthalamic nucleus is one of multiple innervation sites for long-range corticofugal axons: a single-axon tracing study in the rat. *J. Neurosci.* **32,** 5990–9 (2012).
2. Yamashita, T. *et al.* Membrane Potential Dynamics of Neocortical Projection Neurons Driving Target-Specific Signals. *Neuron* **80,** 1477–1490 (2013).
3. Kebschull, J. M. *et al.* High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA. *Neuron* **91,** 975–987 (2016).