Mesoscale correlation structure with single cell resolution during visual coding

Yiyi Yu, Jeffrey N. Stirman, Christopher R. Dorsett, Spencer L. Smith

Dept. of Electrical & Computer Engineering
Center for BioEngineering
Neuroscience Research Institute
University of California Santa Barbara
Santa Barbara, CA

Abstract

Neural circuitry represents sensory input with patterns of spiking activity. Across brain regions, initial representations are transformed to ultimately drive adaptive behavior. In mammalian neocortex, visual information is processed by primary visual cortex (**V1**) and multiple higher visual areas (**HVAs**). The interconnections of these brain regions, over which transformations can occur, span millimeters or more. Shared variability in spiking responses between neurons, called “noise correlations” (**NCs**), can be due to shared input and/or direct or indirect connectivity. Thus, NCs provide insight into the functional connectivity of neuronal circuits. In this study, we used subcellular resolution, mesoscale field-of-view two-photon calcium imaging to systematically characterize the NCs for pairs of layer 2/3 neurons across V1 and four HVAs (areas LM, LI, AL and PM) of mice. The average NCs for pairs of neurons within or across cortical areas were orders of magnitude larger than trial-shuffled control values. We characterized the modulation of NCs by neuron distance, tuning similarity, receptive field overlap, and stimulus type over millimeter scale distances in mouse visual cortex, within and across V1 and multiple HVAs. NCs were positively correlated with shared tuning and receptive field overlap, even across cortical areas and millimeter length scales. We compared the structure of these NCs to that of hypothetical networks to determine what network types can account for the results. We found that to reproduce the NC networks, neuron connectivity was regulated by both feature similarities and hub mechanism. Overall, these results revealed principles for the functional organization and correlation structure at the individual neuron level across multiple cortical areas, which can inform and constrain computational theories of cortical networks.

Introduction

Navigating the environment, detecting predators, and recognizing mates are natural behaviors of mammals that can depend upon sophisticated visual perception. Visual information is encoded and abstracted through multiple interconnected brain areas, which span millimeters in the mouse[1]. These interconnections, both within and between brain areas, are critical components underlying brain functions. Anatomical tracing studies of flies[2] and mice[3], have provided insights into basic neuronal wiring principles. However, anatomical connectivity is insufficient for understanding the circuit interactions of an active brain, and the relationships between function and connectivity. Thus, in this study we investigated the functional circuit connectivity structure at millimeter scale.

Neuronal responses vary on nominally identical trials. The trial-to-trial variability of responses is correlated in a neuronal population. This shared variability between pairs of neurons can be quantified, and is called noise correlation (**NC**)[4]. The structure of NCs is rooted in the architecture of the neuronal circuit, e.g., shared bottom up and top down input, as well as recurrent connections[5–11]. The dynamics of NCs during stimulus onset[12], locomotion[13], circuit plasticity[14] and brain state change[9,15,16] can provide a window into neuronal connectivity principles, which are an important component of the functional properties of neural circuits[11].

The structures of NCs have provided important constraints on neuronal network models of stimulus encoding[9,17–19], motor behavior[18], attention[20] and decision making[21]. For example, a recurrent network model with spatially structured connectivity[22] was inspired by the distance-dependent structure of NCs[23,24], which was uncovered in an attention modulation study in the monkey visual cortical system[25,26]. In most studies, NCs were measured only in pairs of relatively closely positioned pairs of neurons in a population, often within a single cortical region, due to technical limitations of recording methods. Thus, the structure of longer distance, inter-area correlations remains unclear. Some studies have managed to record neuronal activity at single neuron resolution in two or more areas simultaneously by multielectrode recording[26–29], and these studies have also provided important insights into the structure of NCs. However, multielectrode recordings are not well suited for densely sampling a large population of neighboring neurons with little to no cross contamination of signals between neurons (i.e., the spike sorting problem)[30]. A complementary technique is two-photon calcium imaging, but it is commonly limited to relatively small fields of view that do not encompass multiple visual cortical areas in the mouse. These technical limitations have left larger scale principles of NCs between single neurons uninvestigated. To address these issues, we have developed a large field-of-view two-photon microscope with temporally multiplexed imaging beams[31], which can measure neuronal activity in hundreds to thousands of neurons at subcellular resolution, across millimeters of cortex[31].

In mouse visual cortex, the primary visual area (**V1**) and multiple higher visual areas (**HVAs**) are involved in visual processing[1,32–34]. Neurons in V1 and HVAs respond to diverse visual stimuli[35]. HVAs vary in spatial and temporal frequency selectivity[32,33,35], and receive specialized input from V1[36,37]. V1 and HVAs are densely interconnected[37,38], and form segregated inter-area feedforward and feedback connections[39–41], indicating hierarchical structure between V1 and HVAs[42,43].

To assess the functional interactions among neurons in V1 and HVAs, we systematically characterized the structure of NCs between pairs of neurons, both within individual visual areas and between visual areas. NCs were modulated by multiple similarity matrix, including tuning, cortical location, receptive location and stimulus response similarity. We found a strong relationship between tuning and NCs, even when neurons were over a millimeter from each other, and this relationship could overcome the distance-dependent decline in NCs in cases of inter-area pairs of neurons. In addition, using a Gaussian graphic model we found that the network structure of neuron populations was modulated by drifting gratings and naturalistic videos. To reproduce these network structures, neuron connectivity was regularted by both similarity mechanism and hub mechanism. Together, these findings provide insights

into the functional circuit structure of mouse visual cortex at millimeter scale, within and across cortical areas.

Results

1. Multi-region two-photon calcium imaging from V1 and HVAs

To measure neuronal activity, we used multi-region population calcium imaging of L2/3 neurons in V1 and four HVAs (lateromedial, **LM**; laterointermediate, **LI**; anterolateral, **AL**; and posteromedial, **PM**) using a multiplexing, large field-of-view two-photon microscope with subcellular resolution developed in-house[31] (**Fig. 1a**). Mice expressed the genetically encoded calcium indicator GCaMP6s[44,45] in cortical neurons. We located V1 and HVAs of each mouse using retinotopic maps obtained by intrinsic signal optical imaging[33,46] (**Fig. 1b**). We imaged neurons in 2 – 4 cortical visual areas simultaneously (**Fig. 1a, d**), in mice viewing stimuli on a display system. We typically imaged neurons in V1 and one or more HVAs. Up to 500 neurons (V1: 129 ± 92; HVAs: 94 ± 72; mean ± SD.) were recorded per imaging region (500 x 500 μm^2). The imaging regions were matched for retinotopy, so that the neurons in the simultaneously imaged areas had overlapping receptive fields (**RFs**). Calcium signals were used to infer probable spike trains for each neuron (*Methods*; **Supplementary Fig. 1a**). Neurons in V1 and HVAs exhibited similar instantaneous firing rates (**Supplementary Fig. 1b**). We mapped RFs for individual neurons and populations using small patches of drifting gratings (**Supplementary Fig. 1c**). Neurons in HVAs (LM, AL, PM and LI) had significantly larger RFs than V1 neurons (**Supplementary Fig. 1d**). Population RFs for a 500 x 500 μm^2 imaging region of HVAs covered significantly larger portions of visual space than that of V1 (**Supplementary Fig. 1e**), as expected given their differing magnification factors[47]. The overlap of population RFs confirmed that simultaneously imaged cortical areas (V1 and HVAs), each containing ~10^2 neurons, responded to stimuli in the same region of the stimulus display system (**Fig. 1c**).

Using the multi-region imaging paradigm, we recorded reliably responding neurons from V1 and HVAs to both drifting gratings (8 direction at 0.05 cycles/° and 2 Hz) and naturalistic videos (a mouse eye-level view navigating around a mouse home cage) (**Fig. 1e**). The firing reliability (trial-to-trial correlation) to drifting gratings was highest in AL (0.11 ± 0.1; mean ± S.D.), while slightly lower in V1 (0.08 ± 0.09; mean ± S.D.) and LM (0.06 ± 0.09; mean ± S.D.), and lowest in PM (0.036 ± 0.05; mean ± S.D.) (**Supplementary Fig. 2a**). The firing reliabilities for responses to naturalistic videos were similar in V1 and AL (V1, 0.052 ± 0.09), and were lower in LM and PM (0.029 ± 0.05; mean ± S.D.) (**Supplementary Fig. 2a**). Consistent with a prior report[35], we found that neuron responsiveness differed between areas and depended on the stimulus. Among neurons that fired reliably to stimulus (average reliability > 0.08), over half of the neurons responded to drifting gratings (V1, 54%; LM, 69%; AL, 67%; PM, 61%), and a partially overlapping group of neurons responded to a naturalistic video (V1, 87%; LM, 71%; AL, 62%; PM, 70%) (**Supplementary Fig. 2b**). Some neurons responded to both stimulus types, and this portion was smallest in AL and largest in PM (V1, 40%; LM, 39%; AL, 29%; PM, 50%; $p = 5.6 \times 10^{-11}$, One-way ANOVA with Bonferroni correction; **Supplementary Fig. 2b**). These reliable responses provide a basis for further analysis.

In a separate set of experiments, we examined the response selectivity of V1 and HVAs to drifting gratings of various parameters and naturalistic video using the metric lifetime and population sparseness. V1 and HVAs responded sparsely to both sine wave drifting gratings (72 conditions: 8 directions x 3 spatial frequencies x 3 temporal frequencies) and naturalistic videos (**Supplementary Fig. 2c, d**). In all tested areas, the lifetime sparseness of more than 70% of neurons were larger than 0.8. The mean values were >0.8 for both gratings and naturalistic video. The lifetime and population sparseness of responses to drifting gratings was slightly higher in V1 than HVAs (LM, AL, PM and LI), and the sparseness of responses in AL was the lowest ($p = 2.5 \times 10^{-35}$; ANOVA with Bonferroni correction; **Supplementary Fig.**

2c, d). The lifetime sparseness of responses to naturalistic videos were similar in V1, LM and AL, and slightly lower in PM ($p = 0.0004$; One-way ANOVA with Bonferroni correction; **Supplementary Fig. 2c**). In summary, neurons in simultaneously imaged areas (V1 and HVAs) in this dataset encode stimuli in similar regions of visual space, can respond reliably to stimuli, and exhibit high lifetime sparseness.

2. Within-area and inter-area noise correlation

Next, we systematically characterized NCs of pairs of neurons within and across V1 and multiple HVAs (LM, LI, AL, and PM). We computed the NCs of pairs of neurons within individual cortical areas (within-area NC), and NCs for pairs of neurons where the two neurons are in different cortical areas (inter-area NC) (**Fig. 2a**). NCs were computed using spike counts within 1 s bins, similar to previous work with electrophysiology[4,23]. Both within- and inter-area NCs had wide distributions (range: -0.2 – 0.6). The mean NCs across a population was positive and at least five times larger than NCs computed after shuffling the trials (5 – 20-fold, 25 – 75% quantile; **Fig. 2b**). Although the NC estimation of individual neuron pairs was noisy (**Supplementary Fig. 3a**), both simulation and experimental data suggested that the estimation of the population mean NC converges with increasing numbers of neurons (95% confidence interval is ~0.002 with 50 neuron pairs, **Supplementary Fig. 3b, c**). Thus, NCs at the population level are roughly an order of magnitude greater than trial-shuffled control values and are reliably estimated.

We investigated how NCs within individual cortical areas were modulated by firing rate, RF overlap, the physical distance between the neurons (laterally, across the cortex), signal correlation (**SC**; similarity between two neurons of their average responses to stimuli), the stimulus type (gratings or naturalist videos), and anesthesia[9,48]. We tested how within- and inter-area noise correlations were jointly modulated by the firing rate, RF overlap, neuron distance, and SC by fitting a multiple linear regression model. Overall, about 10% of variance of within-area NC, and about 4% of variance of inter-area NC were explained by the aforementioned factors jointly (**Supplementary Fig. 4a**). SC is the most pronounced factor that explained about 4.3% variance and 2.7% variance of within- and inter-area NC, respectively. In all tested areas, within- and inter-area NCs were significantly positively related to SCs (**Fig. 2d, Supplementary Fig. 4a**). RF overlap is the second largest factor for within-area NC in most areas, which explained about 3.3% variance. It only explained 0.5% of variance for inter-area NC. Within- and inter-area NCs linearly increased with neuron spike count to both drifting gratings and naturalistic videos (**Supplementary Fig. 4a-c**). In addition, NCs were also positively modulated by neuron reliability to both stimulus types (**Supplementary Fig. 4d, e**). The relationship of NCs with neuron distance was variable, and this will be discussed in more detail later in this report. Given the strong influence of firing rate and reliability on NCs, we controlled for their influence in subsequent analysis.

We next compared population mean NCs and SCs between V1 and HVAs using subpopulations of reliably firing neurons with matched spike count distribution across visual areas. We found that lateral areas (LM and LI) exhibited lower within-area NCs and SCs than V1 to drifting gratings, while neurons within area PM were not different from V1 (**Supplementary Fig. 5a**). Inter-area NCs of V1-LM, V1-AL, V1-LI, and LM-LI were lower ($p < 0.05$) than those of V1-V1 (simultaneously imaged, non-overlapping regions of V1); while V1-PM was not detected to be different from V1-V1 (**Fig. 2c**). Similarly, the inter-area SCs of V1-V1 was larger ($p < 0.05$) than the inter-area SCs between V1 and HVAs, except PM and AL (**Fig. 2c**). The differences in inter-area correlations between V1-V1 and V1-HVAs were significant over a broad range of reliabilities and spike counts (**Supplementary Fig. 4a, c**). Thus, within-area and inter-area NCs and SCs exhibit area-specific structure.

The area-specific NCs and SCs were modulated by the statistics of the visual stimuli. V1, LM, and AL exhibited similar NCs during stimulation with naturalistic videos, as well as similar SCs (**Supplementary Fig. 5b, c**). Among these areas, LM, but not V1 and AL, exhibited larger NCs to naturalistic videos than to drifting gratings (gratings: 0.012 ± 0.007; naturalistic: 0.029 ± 0.01, mean ± S.D.; $p = 0.011$; Wilcoxon

rank sum test). The SC of V1 was lower to naturalistic videos, compared to drifting gratings (gratings: 0.12 ± 0.075 naturalistic: 0.088 ± 0.07, mean ± S.D.; $p = 1.6 \times 10^{-5}$; Wilcoxon rank sum test). On the other hand, during naturalistic video stimulation, the inter-area NCs of V1-AL was increased (gratings: 0.01 ± 0.005; naturalistic: 0.024 ± 0.016, mean ± S.D.; p = 0.017; Wilcoxon rank sum test), and it was not different from the NCs of V1-V1 and V1-LM (**Supplementary, Fig. 5c**). Overall, multiple features of correlation structure can be modulated by the visual stimulus.

To investigate the effects of anesthesia on the structure of NCs, we recorded from anesthetized mice with isoflurane (1% in O_2) in separate batch of experiments Within-area NCs increased when mice were anesthetized (**Supplementary Fig. 5c**), consistent with prior reports[9,49]. Here, we extended this finding to inter-area NCs, which also increased (**Supplementary Fig. 5c)**.

In summary, we observed that the structure of NCs for pairs of neurons within and across visual cortical areas exhibits area-specific features and can be modulated by the visual stimulus. This suggests that there are neural circuits whose efficacy at driving postsynaptic firing depends on the cortical area and stimulus.

3. Noise correlation is modulated by tuning similarity over millimeter distance scales

For neurons near each other (e.g., 50 microns or less), within a cortical region, the connectivity between neurons is positively modulated by their tuning similarity[48,50]. However, it is unclear whether that relationship holds over longer cortical distances, or across cortical areas. We addressed this question using the large field-of-view imaging system. We examined whether NCs are modulated by orientation tuning properties over millimeter cortical distance.

Neurons in V1 and HVAs exhibited orientation and direction selectivity to drifting grating stimuli. About 46 – 64% among all neurons in V1 and HVAs exhibited strong orientation selectivity (orientation selective index (OSI) > 0.8; V1 64%, LM 61%, AL 49%, PM 50%, LI 46%). And about 28 – 54% of neurons in V1 and HVAs exhibited strong direction selectivity (direction selective index (DSI) > 0.8; V1 54%, LM 38%, AL 29%, PM 28%, LI 49%). V1 exhibited strongest orientation and direction selectivity, while areas AL and PM exhibited lower direction selectivity.

In the subset of orientation selective neurons both within- and inter-area NCs were significantly modulated by the orientation tuning selectivity: neuron pairs that shared the same preferred orientation exhibited higher NCs (**Fig. 3a, b**). As NCs are higher when RF overlap is high (**Fig. 3c; Supplementary Fig 6**), we wondered whether RF overlap could account for the higher NCs of pairs of neurons with similar orientation tuning. To address this question, we examined the dependence of NC on tuning similarity in neuron groups with non-overlapping RFs in V1, LM, and AL (there were insufficient data for PM and LI). We found that for both within-area and inter-area neuron pairs, NC was significantly higher when the neurons shared same preferred orientation (**Fig. 3d**). This result suggests that the connectivity between neurons is modulated by tuning similarity independent of RF overlap, over millimeter distance scales.

4. Distance-dependence of noise correlation over millimeter scale

Prior studies have provided differing accounts of how NCs vary with the distance between neurons. Studies that reported low mean NCs found no distance-dependence of noise correlations[10,51], while studies that reported higher mean NCs found that noise correlations declined with neuron distance[23,24]. In the current study, we characterized the distance-dependence of NCs within and between multiple visual cortical areas. Note that all distances are lateral distances across cortex. In general, the NCs of neurons within a visual area declined with increasing distance for both drifting gratings and naturalistic video stimuli (**Fig 4a; Supplementary Fig 7a, b**). The distance-dependent decrease of within-area NCs was not different among V1, LM, AL and PM, for either drifting gratings or naturalistic videos (linear decreasing rate of NC = -0.02 ± 0.009 mm^{-1} (mean ± S.D.); **Supplementary Fig. 7c**). In these areas,

SCs also exhibited significant distance-dependent decreases (**Fig. 4a**). However, the distance-dependence of within-area NCs were not completely accounted for by the decrease in SCs, as the partial correlation between NCs and neuron distance, even when conditioned on both SC and firing rate, was significant within V1, LM, AL, and PM (Drifting gratings partial correlation: V1, $r = -0.03$, $p < 0.0001$; LM, $r = -0.0173$, $p = 0.025$; AL $r = -0.050$, $p < 0.0001$; PM, $r = -0.048$, $p = 0.0019$). We did not observe a distance-dependent pattern of NCs in LI (**Fig. 4a; Supplementary Fig 7a, b**). The distance-dependence of NCs, independent of tuning similarity and firing rate, suggested distance-dependent recurrent connections between neurons within visual areas (V1, LM, AL and PM). When mice were anesthetized, NCs decayed rapidly with neuron distance in V1 and AL (**Supplementary Fig.7d**).

In contrast to largely monotonic changes of within-area NCs within areas, inter-area NCs exhibited region-specific distance-dependence patterns that were often non-monotonic (**Fig. 4b; Supplementary Fig 7a, b**). To drifting gratings, inter-area SC of V1-LM, V1-AL, and V1-LI pairs increased with distance; and V1-PM and LM-LI pairs declined with neuron distance (**Fig. 4b**). The distance-dependent trend of V1-LM, V1-PM and LM-LI persisted after controlling for SCs and firing rate (partial correlation: V1-LM, $r = 0.045$, $p = 8.2 \times 10^{-26}$; LM-LI, $r = -0.048$, $p = 6.4 \times 10^{-6}$; V1-PM, -0.05, $p = 9 \times 10^{-15}$). We did not find a distance-dependent change for V1-AL and V1-LI after controlling for SCs and firing rate. The distance dependence of V1-LM was significant in individual experiments to drifting gratings, but not to naturalistic videos (**Supplementary Fig. 7a, b**). In summary, the distance-dependence of inter-area NCs were complex and diverse.

These diverse relationships between distance and NCs for inter-area neuron pairs could be due to the topography of retinotopic maps. In mouse visual cortex, the orientation of retinotopic maps vary by HVA. To examine whether the distance-dependent increase of V1-LM NCs to drifting gratings could be explained by retinotopy (**Fig. 1b**), we aligned V1 and LM retinotopy by affine transformations of intrinsic signal retinotopic maps of V1 to match those of LM (*Methods*, **Fig. 4c**). After retinotopic warping, V1 and LM were treated as two layers with aligned retinotopies (**Fig. 4d**), such that neurons with nearby receptive field locations in visual space also have small inter-area neuronal distances. We found the distance-dependent increase of V1-LM NCs to drifting gratings was eliminated after retinotopy alignment (**Fig. 4e**). After warping, the inter-area NCs between V1 and LM exhibited a distance-dependent decrease (linear correlation, $r = -0.26$, $p = 3 \times 10^{-6}$), similar to the trends found for within-area NCs in V1 and LM (**Fig. 4f**). In addition, after retinotopic alignment the inter-area NCs for V1-LM to naturalistic videos did not exhibit distance-dependence (Linear correlation, $p = 0.47 \pm 0.35$; mean ± SD.). We also investigated the distance relationship for V1-AL inter-area NCs, in which NCs increased with distance to drifting gratings (**Fig. 4b**). Using the same warping approach, alignment of V1 retinotopy to match that of AL resulted in a distance-dependent decline of inter-area NCs for V1-AL (**Supplementary Fig. 8a, b**). Overall, these results show that a major driver of inter-area NCs is retinotopic alignment, which can overcome simple distance-dependent decreases.

5. Functional network structure of mouse visual cortex favors information encoding specificity

The analysis presented thus far focused on measurements from individual pairs of neurons, then averaged over a population of neuron pairs. Next, we characterized the full NC matrix of simultaneously recorded neurons from multi-region imaging. This analysis revealed structural features of the network, both within and across visual cortical areas.

NCs reflect both direct and indirect correlations between neurons. When two neurons are positively correlated with a third neuron, these two neurons could exhibit positive correlations regardless of their actual connectivity. Thus, it is ambiguous to characterize neuron network structure using the original noise correlation matrix (**Fig. 5a**). To address this, we first converted the NCs to partial NCs, which are the residual correlations between a neuron pair after conditioning on all the other neurons in the data set

(**Fig. 5a, b**; *Methods*). The partial NC matrix was generated using a previously established method[52] (**Supplementary Fig. 9a-d**; *Methods*). The partial NC matrix is sparser than the original NC matrix, but still contains more positive elements compared to trial shuffled controls (mean partial NC, 0.0039; trial shuffled: 10^{-5}; t-test, $p < 10^{-4}$, **Fig. 5b; Supplementary Fig. 9e-g**). The partial NCs also still captured important features of the NCs discussed earlier, including the area-specific distance-dependence and tuning similarity-dependence structures (**Supplementary Fig. 10**).

We used the partial NC matrices to analyze the network structure. Individual neurons were treated as nodes, and two nodes were connected with an undirected edge when the partial correlation between them was above a threshold. The threshold and edge density were negatively related (**Supplementary Fig. 11a**). For the remaining analysis, a threshold of 0.015 was used if not otherwise stated, but qualitatively results were not dependent on thresholding (**Supplementary Figure 11g**). The edge density was slightly higher in data sets where naturalistic video stimuli were presented (0.18), than drifting grating stimuli (0.11). This was true over a wide range of thresholds (0.005 – 0.02), suggesting that neurons were slightly more densely connected during stimulation with naturalistic videos (**Supplementary Fig. 11a**). Neuron networks exhibited small world property that neurons were connected through short path over a wide range of edge density (path length/network size is about 2/134; **Supplementary Fig. 11a**).

We characterized the degree of each neuron in the network, which is the number of edges originating from a node (**Fig. 5a**). The degree of nodes was positively related with response strength of neurons (**Supplementary Fig. 11c**). Also, higher degree nodes tend to carry higher information about the input stimuli, but have lower reliability (**Supplementary Fig. 11c**). Neurons that were high degree nodes for gratings, were typically not high degree nodes for naturalistic videos, and vice versa (**Supplementary Fig. 11d**). As a control, we also generated Erdös–Rėnyi (ER) random networks with the same edge density. There was a larger fraction of high degree nodes in the observed data from visual cortex, compared to the matched, randomly generated networks (**Fig. 5c**). This was true with either stimulus, gratings or naturalistic videos, but the degree distribution of naturalistic video-evoked networks was right-shifted compared to that of drifting grating-evoked networks. This stimulus-evoked difference in degree distribution was consistent across individual experiments and across various HVAs (**Supplementary Fig. 11e, f**).

In addition, neurons formed local clusters in the network such that neighboring nodes tend to connect to form closed clique (**Fig. 5a**). To characterize neuron clustering, we computed the clustering coefficients of each neuron, which was the fraction of neighbors of a node that connect to each other, forming a clique (*Methods*). The clustering coefficients of neurons were gradually decreasing in an increasingly sparse network (**Supplementary Fig. 11b**). At the selected threshold (0.015, edge density 0.11-0.18), neuron networks exhibited relatively high clustering coefficients (Drifting gratings: 0.36; Naturalistic video: 0.43), compare to an ER random network with the same edge density (**Fig. 5d, e**). The clustering coefficients of neuron networks were slightly lower than regular networks with or without the same degree distribution as actual neuron networks (**Fig. 5d, e**; *Methods*). The regular network preserving the degree distribution of actual neuron networks set the upper bound of local clustering.

The clustering coefficients of individual nodes was negatively related with the degree of nodes, thus neurons connected with many other neurons tend to have low clustering coefficients, i.e., their neighbors are less likely to be interconnected as fully as those of lower degree nodes. This was true both for reference networks and neuron networks based on our partial NC matrices (**Fig. 5f**). The degree-dependent decay of clustering coefficients was fit by a power law: $clustering\ coefficients \propto degree^{\gamma}$ (*Methods*). The exponent γ is an important feature of neuron networks[53]. The exponents of the neuron networks at selected threshold (0.015, edge density 0.11-0.18) were approximately -0.33 for both drifting gratings and naturalistic video (**Fig. 5f**). The absolute exponent values of actual neuron networks intermediate between those for the reference random and regular networks (**Fig. 5f**). A random network

with or without preserving the degree distribution of actual neuron network had flatter joint distribution of clustering coefficients and node degrees. A regular network with or without preserving the degree distribution had steeper joint distribution of clustering coefficients and node degrees.

Together, these results show that the visual cortical neuron network inferred by partial NC matrices has right skewed degree distribution with more high degree nodes than expected for random networks. Moreover, neuron networks exhibited small world properties with short path length and high local clustering. The clustering coefficients were monotonically decreasing with the degree distribution at a power law with exponent -0.33.

6. Reproducing the observed network structure using simulations (Discussion)

So far, we have showed that a subset of visual cortical neurons excited by a stimulus formed a functional network, characterized by the partial noise correlation network, with very specific structure. The functional neuron network has right skew degree distribution with heavy tail, high local clustering and short path length (**Fig. 5, Supplementary Fig. 11**). A natural question we had next was the emergence of the functional network structure. The structural feature of the functional network of mouse visual cortex resembled that of the mesoscale connectome network of mouse cortex[53]. However, the anatomical structure alone did not explain the stimulus dependent change of the degree distribution of functional neuron networks (**Fig. 5**).

On the other hand, we have identified multiple feature similarity matrices, including tuning property, RF location (or retinotopy), neuron distance and response similarity, which partially explained the variance of noise correlation and partial noise correlation (**Supplementary Fig. 4 & 10**). Overall, noise correlations and signal correlations tend to increase together. RF overlap was a major component of signal correlations that was associated with higher noise correlations. Orientation tuning also played a role, even when RFs did not overlap. These trends persisted at the mm length scale, and across cortical area boundaries, suggesting feature similarity being a universal mechanism underlying the correlation network. We tested whether a feature similarity mechanism would reproduce the functional network structure of visual cortex (**Fig. 6a**). We simulated nodes in feature space and computed feature similarity index (*SI*) between nodes using a design function. The design function regulated the distribution of node similarity (**Supplementary Fig. 12a, b; *Methods***). The probability of node i and node j being connected was determined by rectified feature similarity index ($p(k_{i,j} = 1) = f(SI_{i,j}), f: rectifier$) (**Supplementary Fig. 12c**). We have tested a wide range of design functions and rectifiers (**Supplementary Fig. 12a-c**). In general, the similarity mechanism generated networks with weak left-skewed or bell shape degree distribution at the experimental level of edge density (0.11-0.18) (**Supplementary Fig. 12d**). Only when the simulated network was very sparse (edge density < 0.05), it had right skewed degree distribution. This was true for all tested design functions and rectifiers (**Supplementary Fig. 12e**). Also, we found that both local clustering and path length of simulated networks were modulated by the shape of the rectifier, while not sensitive to the distribution of feature similarity (**Supplementary Fig. 12e, f**). A sharp super-linear rectifier generated networks with high local clustering (0.76) and long path length (5.2); while sub-linear rectifier generated networks with low local clustering (0.18) and short path length (2.1). Simulation with a linear rectifier and sharp design function generated networks with experimental-level clustering coefficients (0.38) and path length (2.1), and also reproduced the joint distribution of node degree and clustering coefficients (power law exponent of simulation = -0.32) (**Supplementary Fig. 12e, g)**. These simulation results suggested that a feature similarity mechanism favored establishing local clustering, however it was insufficient to fully recover the degree distribution of functional network structure. Indeed we found that feature similarity matrices explained no more than 20% of the variance of the noise correlation matrix in combined, suggesting additional mechanism (**Supplementary Fig. 4a**).

It was suggested that a network with heavy-tailed degree distribution could be generated by preferentially attaching new nodes to high-degree nodes in the network (preferentially attachment mechanism)[57] (**Fig. 6a**). Preferential attachment mechanism generated scale-free networks and favored the establishment of hubs. However, the preferential attachment mechanism alone generated networks with flat joint distribution of clustering coefficients and node degree (**Fig. 6a-c**). We wondered whether combining the preferentially attachment mechanism and feature similarity mechanism (PAFS) would reproduce the functional neuron network structure (**Fig. 6a-c**). In the PAFS simulation, a network was initialized by feature similarity mechanism, and then a subset of nodes were further attracted to form connections preferentially with high degree nodes such that the probability of node l being connected to node i was determined by the connectivity of node i ($p(k_{l,i} = 1) = \frac{k_i}{\sum_n k_n}$; $k_i = \sum_j k_{i,j}$). The strength of high degree attraction was parameterized by the size of the attracted subset and the number of preferential attachment made by each nodes. In general, the PAFS mechanism generated networks with short distance, right-skewed degree distribution, high clustering coefficients and the negative relationship between clustering coefficient and node degree (**Supplementary Fig. 13a-d**). The degree distribution of PAFS networks were modulated by the strength of attraction: smaller subset (0.2-0.5) of attracted nodes led to sparser network with right-skewed degree distribution, while large subset (0.6-0.8) of attracted nodes led to denser network with bell-shaped degree distribution. Thus, the PAFS mechanism reproduced the stimulus-specific degree distribution of functional neuron networks (**Fig. 6d; Supplementary Fig. 13a, c**).

Although the preferentially attachment mechanism sounds arbitrary, it can be achieved with a rank-one global modulation. We simulated a network that the probability of two nodes being connected was defined by a low-rank weight matrix (**Fig. 6e**). The simulated network had right-skewed degree distribution with heavy tail and non-decaying joint distribution of clustering coefficient and node degree (**Fig. 6e**). Increasing the rank of the weight matrix, the degree distribution shifted from right-skewed towards bell-shaped, while kept the flat joint distribution of clustering coefficient and node degree (**Fig. 6e**). Previous studies showed that sensory cortical neurons received low-rank global modulation on brain state, which led to low rank trial-to-trial co-variation of population neurons[6,9]. Also, a recent study showed that contrast, adaptation, attention and task switching affected monkey visual cortical neurons through a similar low-rank normalization mechanism[58]. We hypothesized that during sensory encoding, mouse visual cortical neurons receive a low-rank global modulation that drives the establishing of right skewed degree distribution of neuron network to fulfill the function of sensory representation.

Over all, in the current study we showed that population neurons in mouse visual cortex exhibited trial-to-trial covariation, and the covariation of neurons formed a functional neuron network which exhibited specific structure with right-skewed degree distribution, high local clustering, short distance and decaying joint distribution of cluster coefficient and node degree. We found that both feature similarity mechanism and preferential attachment mechanism were required for establishing the functional neuron network structure. We have observed multiple feature similarity matrices in mouse visual cortex, including tuning preference, RF location, neuron location and response selectivity, which were the candidates leading the feature similarity mechanism. On the other hand, we proposed that the preferential attachment was achieved by a low-rank global modulation process.

Author contributions

All experiments and analysis were performed by YY. The imaging system was built by JNS. Analysis was assisted by CRD. Study design and supervision by SLS. YY and SLS wrote the paper.

Acknowledgements

Funding was provided by grants from the NIH (R01EY024294, R01NS091335), the NSF (1707287, 1450824), the Simons Foundation (SCGB325407), and the McKnight Foundation to SLS; a Helen Lyng White Fellowship to YY; a career award from Burroughs Wellcome to JNS; and training grant support for CRD (T32NS007431).

Online Methods

Animal and surgery

All animal procedures and experiments were approved by the institutional Animal Care and Use Committee of University of North Carolina at Chapel Hill and carried out in accordance with the regulation of the US Department of Health and Human Services. GCaMP6s expressing transgenic adult mice of both sex were used in this study. Mice were 110~300 days old for data collection. GCaMP6s expressing were induced by triple crossing of TITL-GCaMP6s line (Allen Institute Ai94), Emx1-Cre line (Jackson Labs #005628) and ROSA:LNL:tTA line (Jackson Labs #011008)[44]. Mice were housed under a 12-h/12-h light-dark cycle, and experiments were performed during the dark cycle of mice. Mice were anesthetized with isoflurane (1.5~1.8%) and acepromazine (1.5~1.8 mg/kg body weight) when perform visual cortex craniotomy. Carpofen (5 mg/kg body weight) was administered prior to surgery. Mice body temperature were maintained using physically activated heat packs during surgery. Mouse eyes were kept moist with ointment during surgery. The scalp overlaying the right visual cortex was removed, and a custom head-fixing imaging chamber with 5-mm diameter opening was mounted to the skull with cyanoacrylate-based glue (Oasis Medical) and dental acrylic (Lang Dental). A 4-mm diameter craniotomy was performed over visual cortex and covered with #1 thickness coverslip.

Locating visual areas with intrinsic signal optical imaging (ISOI)

ISOI experiments were carried out similarly as previously[31,46,59]. Briefly, the pial vasculature images and intrinsic signal images were collected using CCD camera (Teledyne DALSA 1M30) at the craniotomy window. A 4.7 × 4.7 mm^2 cortical area was imaged at 9.2 μm/pixel spatial resolution and at 30 Hz frame rate. The pial vasculature was illuminated and captured through green filters (550 ± 50 nm and 560 ± 5 nm, Edmund Optics). The ISO image was collected by focusing 600 μm down from the pial surface. The intrinsic signals were illuminated and captured through red filters (700 ± 38 nm, Chroma and 700 ± 5 nm, Edmund Optics). Custom ISOI instrument were adapted from Kalatsky and Stryker[60]. Custom acquisition software for ISOI imaging collection was adapted from David Ferster[31]. During ISOI, mice were 20 cm from a flat monitor (60 × 34 cm^2), which covers the visual field (110° x 75°) of the left eye. Mice were lightly anesthetized with isoflurane (0.5%) and acepromazine (1.5–3 mg/kg). The body temperature was maintained at 37 °C using custom electric heat pad[31]. Intrinsic signal response to vertical and horizontal drifting bar was used to generate azimuth and elevation retinotopic maps (Fig. 1b). The retinotopic maps were then used to locate V1 and HVAs (Fig. 1b). Borders between these areas were drawn at the meridian of elevation and azimuth retinotopy manually[33,46]. The vasculature map then provided landmarks to identify visual areas in two-photon imaging.

In vivo two-photon imaging

Two-photon imaging was carried out using custom Trepan2p microscope controlled by costume LabView software[31]. Two-region imaging was achieved by laser beam split and temporal multiplexing[31]. Briefly, two-photon excitation light from an 80 MHz Newport Spectra-Physics Mai Tai DeepSee laser was split into two beams through polarized optics, and one path is delayed 6.25 ns relative to the other. The two beams were imparted by custom-motorized steering mirrors and tunable lenses independently, such that the X, Y, Z plane of the two paths can be independently positioned in the full field (4.4 mm diameter) positioned during imaging. Both beams were scanned by resonant scanner (4 kHz, Cambridge Technologies), and simple photon signal was collected by PMT (H7422P-40, Hamamatsu). In the current study, two-photon imaging of 500 x 500 $μm^2$ was collected at 13.3 Hz for two-region imaging, or 6.67 Hz for quartic-region imaging. Imaging was performed with about <80 mW excitation 910 nm laser out of the front of the objective (0.45 NA). Mice were recovered in home cage for at least 2 days before acquiring two-photon imaging, after surgery. Mice were head-fixed at about 11 cm from a flat monitor, with their left eye facing the monitor, during imaging. A 70° x 45° the left visual field was covered. If not otherwise

stated, two-photon images were recorded from quite awake mice. For anesthetized experiments in supplementary fig. 3c, d, mice were lightly anesthetized under 1% isoflurane. During two-photon imaging, we monitor mouse pupil position and diameter using custom-controlled CCD camera (GigE, National Instruments) at 20~25 fps. No additional light stimulation was used for pupil imaging.

Visual stimuli

Visual stimulation was displayed on 60 Hz LCD monitor (9.2 x 15 cm^2). All stimuli were displayed in full contrast, and in red color cannel of the monitor to minimize the cross talk with the two-photon imaging. For course population RF and single neuron RF mapping (Fig 1, supplementary fig 1), a rectangular (7.5° x 8.8°) bright moving patch containing vertical drifting grating (2Hz, 0.05 cpd) on a dark background was displayed. The moving patch was appeared and disappear on a random position of the full monitor in pseudo random order without interruption by a gray screen, and presented on each position for 5 sec.

To characterized the value and structure of correlation, as well as tuning properties of V1 and HVAs (**Fig. 2-4**), we showed mice full screen sine-wave drifting grating stimuli is 8 directions (0 – 315°, in 45° step), with a of 0.02, 0.05 or 0.19 cpd, and a TF of 1, 2 or 8 Hz (72 conditions in total). Each of the sine-wave drifting grating stimuli was presented for 2 s in pseudo random order. Stimuli with the same SF and TF was presented successively without interruption. A gray screen was presented for 3 s when changing the SF and TF of stimuli. We also characterized the noise correlation structure to naturalistic videos (**Fig. 2, 4**). Two naturalistic videos of navigating a mouse home cage was shown. Each lasted for 32 secs.

For estimation correlation with regularized models (**Fig. 5**, **Supplementary Fig. 11**), we used two following stimuli: (1) drifting gratings with 3 directions (60°, 80°, 90°) at 0.05 cpd and 2 Hz (3 conditions); (2) sine-wave upward motion grating at a SF of 0.02, 0.05 or 0.19 cpd, and a TF of 1, 2 or 8 Hz (9 conditions). For both stimuli, gratings were presented for 1.5 sec in pseudo random order. For stimulus (1), 1 s gray screen was shown after every 9 grating stimuli. For stimulus (2), the contrast of the sine-wave grating stimulus ramp up from 0% to 100% within 0.3 s, and then ramp down from 100% to 0% within 0.3 s.

Calcium imaging processing

Calcium imaging processing was carried out using custom MATLAB codes. Two-photon calcium imaging was motion corrected using Suit2p subpixel registration module[61]. Neuron ROIs and cellular calcium traces were extracted from imaging stacks using custom code adapted from Suit2p modules[61]. Neuropil contamination was corrected by subtracting the common time series (1st principal component) of a spherical surrounding mask of each neuron from the cellular calcium traces[30,32]. Neuropil contamination corrected calcium traces were then deconvolved using a Markov chain Monte Carlo (**MCMC**) methods[30,62]. For each calcium trace, we repeated the MCMC simulation for 400 times, and measured the signal-to-noise of MCMC spike train inference for each cell (**Supplementary Fig. 1a**). For all subsequent analysis, only cells that reliable spike train inference results were included. Neurons with low responsiveness was excluded for subsequent analysis (trial averaged spike count to preferred spatiotemporal frequency summed over all orientations < 1; or trial averaged spike count to a 32 s naturalistic video < 1).

Receptive field, sparseness and reliability

We mapped RFs by reverse correlation of neuronal responses with the locations of the moving patch of drifting grating stimulus. For population RF mapping, population neuronal responses of simultaneously recorded neurons from a 500 x 500 μm^2 imaging window were reverse correlated with the stimulus locations.

To quantify reliability, we computed trial-to-trial Pearson correlations for neuronal responses (500 ms time bins) to each stimulus type. The sparseness was computed as equation 1 (ref. [63]). For lifetime

sparseness, r_i is the response to the ith movie frame and averaged across trials and N is the number of movie length. For population sparseness, r_i is the response of the ith neuron and averaged across trials and N is the number of neurons.

$$S = \frac{1 - \frac{1}{N} * \frac{(\sum_i r_i)^2}{\sum_i r_i^2}}{1 - \frac{1}{N}} \qquad (eq.1)$$

$$r_i : average\ response\ to\ ith\ stimulus;\ N: total\ number\ of\ stimulu$$

Sparseness to sine-wave gratings was computed using 2 s time bins (**Fig. 2a**). While the sparseness to naturalistic videos was computed using 500 ms bin (**Fig. 2a**). We quantified the sparseness of reliably firing neurons (reliability > 0.08). The binning size did not affect the qualitative results of reliability and sparseness.

Correlation calculation

Noise correlation was defined as the trial-to-trial correlation of residual spike count (1 s time window, if not otherwise stated) after subtracting the mean response to each stimulus. Residual spike count to all stimuli (eg. gratings with different directions and SFs and TFs), and all trials were concatenated into one column vector per neuron (u_i, i = 1,2…N, neuron identity). The noise correlation r_{sc} was computed as the Pearson correlation of u_i and u_j.

$$r_{sc} = \frac{cov(u_i, u_j)}{\sqrt{var(u_i) * var(u_i)}} \qquad i,j: neuron\ identity \qquad (eq.2)$$

In **fig. 3e**, we computed the normalized noise correlation by subtracting the mean of the population noise correlation and divided the standard deviation of the population noise correlation.

For characterizing noise correlation structure to sine-wave grating (72 conditions) (**Fig. 2-4**, supplementary fig. 2-10) and naturalistic videos, we collected 10~20 trials for each condition. For modeling correlation and estimation partial correlation (**Fig 5, Supplementary Fig 11**), we collected about 240~300 trials/condition for the 3 directions drifting grating stimuli (3 conditions) and collected 80~100 trials/condition for the upward motion drifting grating stimuli (9 conditions).

Signal correlation was defined as the neuron-to-neuron Pearson correlation of mean responses. Mean response was a 72-element column vector, computed by trial averaging responses to sine-wave gratings with 72 conditions (**Fig 2-4, Supplementary Fig. 4**).

To examine the relation between noise correlation and joint firing rate between a pair of neurons. We computed the mean joint spike count (geometric mean spike count average over all stimuli (eg. 72 conditions for sine-wave drifting gratings) and the peak joint spike count (maximal geometric mean spike count) of individual neuron pairs.

To compare the noise correlation between areas fairly, we computed the population mean noise and signal correlations with a subset of neurons from each areas so that the distribution of the mean joint spike count of neuron pairs of one visual areas matches with another[12], after excluding a subset of neuron pairs with extremely low joint spike count (mean joint spike count < 0.02 sp/s) and low reliability (< 0.08). Population with fewer than 50 neuron pairs after mean matching were excluded. Also, for computing the noise correlation of a distance bin, or a signal correlation bin, data points with fewer than 50 neuron pairs were excluded.

We computed inter-area NCs with simultaneously recorded regions that shared greater than 40% of population RF. We kept this criterion even though we did not detect a relationship between the inter-area NC and the fraction of population RF overlap within the tested range (p = 0.37; **Supplementary Fig. 1e**).

Significance of noise correlation

Since the value of noise correlations was small, we tested whether these values were significantly above zero. We compared the noise correlation with trial-shuffled noise correlation, the latter was computed using trial-shuffled data (the order of trials were randomized for each neuron independently). The population mean noise correlation computed with trial-aligned data was significantly higher than that of the trial-shuffled data with the size of experimental population (**Fig. 2b, c**).

Accuracy of noise correlation

We investigated the accuracy of noise correlation estimation with both data and model. The individual noise correlations of the same set of neurons varied when computing using different random sub set of trials (**Supplementary Fig. 2a**). We computed the population mean value of noise correlation of random subset of neuron pairs and calculated the confidence interval for estimating the population mean noise correlation. The accuracy of population mean estimation increases with the number of neurons, even with limited number of trials (**Supplementary fig. 2b**). We further characterized the estimation accuracy by simulating correlated neuron population[64], which allows arbitrary number of trials. The expected firing rate and expected population mean correlation matches our experimental data. To achieve an accurate estimation (1/10 standard error/mean value) of population mean correlation converges with >100 neurons even using experimental level trial numbers (**Supplementary Fig. 2d**).

Orientation and direction selectivity

The direction and orientation selectivity of each neuron was computed using neuron response to its optimal spatial-temporal frequency of drifting grating stimuli. Direction selectivity index and orientation selectivity index were computed using following equations.

$$DSI\ or\ OSI\ = \frac{R_{prefer} - R_{null}}{R_{prefer} + R_{null}} \qquad (eq.3)$$

ISOI warping

We spatially registered ISOI map of V1 to align with that of LM or AL. We first segmented the ISOI map by color segmentation using kmeans clustering, and then determined the center of each color segment. Then we performed the affine transformation of color band centers of V1 to match that of LM or AL. The transformation matrix M was determined by minimizing the distance between transformed V1 centers and LM or AL centers using Matlab function *fminsearch*.

Information about stimuli

We computed mutual information about stimulus carried by individual neuron by subtracting noise entropy from the total entropy ($I(x;s) = H(x) - H(x|s)$). The total entropy was computed from the empirical probability distribution of response from all time of individual neuron. The noise entropy was computed from the conditional probability distribution of neuronal response to the same stimulus over multiple trials.

Estimate partial correlation matrix

The noise correlation measures the marginal linear correlations between neuron pairs without controlling for the activity of other neurons. Using the multi-neuronal recordings, we estimated the pairwise functional connectivity by conditioning on the activity of the rest of the network, which takes the form of partial correlation. Estimating partial correlation

Partial correlation may be defined through conditional distribution of data with multivariate normal distribution[65,66]. Assume neuronal responses were from multivariate normal distribution. Suppose we have a n p dimensional variables $X = \{x_i;\ i = 1,2,\dots,n\}$ from a multivariate normal distribution $X \sim N_p(\mu, \Sigma)$. By definition, the multivariate distribution function is

$$f(X) = \frac{1}{(2\pi)^{n/2}|\sigma|^{1/2}} e^{-\frac{1}{2}(X-\mu)^T\sigma^{-1}(X-\mu)} \qquad (eq.4)$$

Suppose we partition the dataset X into two subsets $X = \binom{X_1}{X_2}$, $X_1 = \{x_i;\ i = 1,2,\dots,l\}$ and $X_2 = \{x_j;\ j = 1,2,\dots,k\}$, $l + k = n$. Accordingly, mean and covariance matrix are partitioned as $\mu = \binom{\mu_1}{\mu_2}$ and $\Sigma = \begin{pmatrix} \sigma_{11}, & \sigma_{12} \\ \sigma_{21}, & \sigma_{22} \end{pmatrix}$, where the subset of variables X_1 have means μ_1 and variances σ_1 and the X_1 and X_2 have covariances σ_{12} and σ_{21}. The conditional distribution of X_1 given X_2 is a multivariate normal distribution with conditional mean $E(X_1|X_2)$ and covariance $cov(X_1|X_2)$.

$$E(X_1|X_2) = \mu_1 + \sigma_{12}\sigma_{22}^{-1}(X_2 - \mu_2) \qquad (eq.5)$$

$$cov(X_1|X_2) = E\{(X_1 - E(X_1|X_2))^2|X_2\} = \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21} \qquad (eq.6)$$

Also, we would like to fill in the theorem of the matrix inversion here. The inverse of a covariance matrix $\Sigma = \begin{pmatrix} \sigma_{11}, & \sigma_{12} \\ \sigma_{21}, & \sigma_{22} \end{pmatrix}$ is

$$\omega = \Sigma^{-1} = \begin{pmatrix} \omega_{11}, & \omega_{12} \\ \omega_{21}, & \omega_{22} \end{pmatrix} = \begin{pmatrix} (\sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21})^{-1}, & -(\sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21})^{-1}\sigma_{12}\sigma_{22}^{-1} \\ -(\sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{12})^{-1}\sigma_{21}\sigma_{11}^{-1}, & (\sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{12})^{-1} \end{pmatrix} \qquad (eq.7)$$

After establishing the conditional distribution of multivariate normal variables, we went on to define partial correlation between neuron x_l and x_k, $(X = \{x_l,\ x_k;\ l,k \in \{1,n\}\})$, given the rest of the population $(X^c = \{x_i;\ i = 1,2,\dots,n;\ i \cong k,l\})$. Suppose neuron data are from a multivariate normal distribution $N_p \sim (\binom{\mu_X}{\mu_{X^c}}, \begin{pmatrix} \sigma_X, & \sigma_{X,X^c} \\ \sigma_{X^c,X}, & \sigma_{X^c} \end{pmatrix})$.

The inverse of the covariance matrix is $\omega = \Sigma^{-1} = \begin{pmatrix} \omega_{11}, & \omega_{12} \\ \omega_{21}, & \omega_{22} \end{pmatrix}$ with $\omega_{11} = \begin{bmatrix} \omega_l & \omega_{lk} \\ \omega_{kl} & \omega_k \end{bmatrix}$

The conditional covariance between x_l and x_k is

$$cov(x_l,\ x_k|X^c) = \sigma_X - \sigma_{X,X^c}\sigma_{X^c}^{-1}\sigma_{X^c,X} \qquad (eq.8)$$

According to the matrix inversion theorem (eq.). The conditional covariance is the inversion of the first element in the inverse matrix of covariance.

$$cov(X|X^c) = \sigma_X - \sigma_{X,X^c}\sigma_{X^c}^{-1}\sigma_{X^c,X} = \omega_{11}^{-1} = \frac{1}{\det(\omega_{11})}\begin{bmatrix} \omega_k & -\omega_{lk} \\ -\omega_{kl} & \omega_l \end{bmatrix} \qquad (eq.9)$$

The conditional correlation or partial correlation is computed as

$$p_{l,k|X^c} = -\frac{\omega_{lk}}{\sqrt{\omega_k\omega_l}} \qquad (eq.10)$$

Thus, the partial correlation matrix P, which is the residual correlation between any pair of neurons condition on the rest of the population, was computed using the inverse of the covariance matrix (Σ).

$$P = -\left(diag(\Sigma^{-1})\right)^{-\frac{1}{2}}\Sigma^{-1}\left(diag(\Sigma^{-1})\right)^{-\frac{1}{2}} \qquad (eq.11)$$

As the experimental covariance matrix is noisy and may be not full rank, so that it is imprecise to compute the partial correlation matrix by inversing the experimental covariance matrix. To improve the estimation of partial correlation matrix, we used a previously developed method[52]. It estimates the

covariance matrix by imposing some regularized structures. The covariance matrix is described by linear combination of latent components, or a sparse partial correlation matrix that a large fraction of partial correlations are zeros, or both. We used cross-validation to determine the best regularization structure. The cross-validation loss function was adapted from Kullback-Leibler divergence ($L(\Sigma_{regularized}, \Sigma)$). Consistent with reference [52], regularization structure combining latent components and sparse components was the better than directly estimating covariance using sample data, or estimating with latent components or sparse components along (**Supplementary Fig. 9a**). In the following analysis, we estimated covariance matrix by imposing combined structure of latent and sparse components (latent + sparse) as in eq. 12 (**Supplementary Fig. 9b**). The hyper parameters were selected by minimizing the loss function as in eq. 13. Eq. 11-13 were adapted from ref [52].

$$\Sigma_{latent+sparse} = (S - L)^{-1}\ ; \quad S: sparse\ components;\ L: latent\ components; \qquad (eq. 12)$$

$$(S, L) = \underset{\hat{S},\hat{L}}{\operatorname{argmin}}\{L(\Sigma_{latent+sparse}, \Sigma) + \alpha \left|\left|\hat{S}\right|\right|_1 + \beta trace(\hat{L})\} \qquad (eq. 13)$$

In the latent + sparse model, the sparseness of sparse components was high (> 0.95) and the number of the latent components scaled linearly with the number of neurons in the population and entailed about 15 components per 100 neurons (**Supplementary Fig. 9c**). The latent + sparse model accurately recovered the population mean noise correlation (**Supplementary Fig. 9d**). The resulted partial correlation matrix is sparser than the original correlation matrix (**Supplementary Fig. 9e**), but significantly denser than the partial correlation matrix of trial shuffled data (**Supplementary Fig. 9f, g**). Dominant latent components captures the area-specific distance-dependence of partial noise correlation (**Supplementary Fig. 10a**).

Network structure

We characterized edge density, degree and clustering coefficients following standard equations. In a network with n nodes, we defined

$$edge\ density = \frac{number\ of\ edge}{number\ of\ all\ possible\ edge\ or\ \mathrm{C}(n, 2)} \qquad (eq. 14)$$

We defined the degree of node i in the network as

$$degree\ of\ node\ i = number\ of\ edges\ connected\ with\ node\ i \qquad (eq. 15)$$

$$normalized\ node\ degree = \frac{node\ degree}{n} \qquad (eq. 16)$$

Clustering coefficient of node k measures the fraction of closed triplets of its neighbors (nodes that are directly connected with k). When node k has m neighbors,

$$clustering\ coefficient\ of\ k = \frac{number\ of\ closed\ triplets}{number\ of\ all\ possible\ triplets\ or\ \mathrm{C}(m, 2)} \qquad (eq. 17)$$

We fitted a power law function to the relation of degree and clustering coefficients. The exponents of power law function were fitted by least mean square.

$$clustering\ coefficients \propto\ degree^{\gamma} \qquad (eq. 18)$$

Reference networks

We compared the network features of an actual partial correlation neuron network to that of reference networks generated as following.

Random network with binomial degree distribution

Random network was generated using Erdös–Rėnyi model[67]. A random network with binomial degree distribution was initialized with a set of isolated nodes, and then edges was added between them at an equal probability. The probability of the presence of an edge was matched to the edge density of an actual neuron network.

Random network preserve degree distribution

A random network preserving the degree distribution was generated as [50,51,71]. In this control network, nodes have the same degree distribution as the target actual neuron network, but randomly shuffling the link between interaction partners.

Regular network with binomial degree distribution

A regular network was generated as [68,69] through "Latticization" a ER network with binomial degree distribution. This reference network have the same edge density as the target actual neuron network. It and has non-zero entries of the connection matrix close to the main diagonal. Briefly, this network was generated by randomly rewiring the link between interaction partners when the rewired edge was located closer to the main diagonal.

Regular network preserve degree distribution

A regular network preserving degree distribution was generated as [68,69] through "Latticization". This reference network have the same degree distribution as the target actual neuron network and has non-zero entries of the connection matrix close to the main diagonal. Briefly, this network was generated by randomly rewiring the link between interaction partners when the rewired edge was located closer to the main diagonal.

Network simulation

We simulated networks following three mechanisms. Results were from simulations with 200 node, if not otherwise stated. Results hold true in larger simulation.

Similarity mechanism

We simulated nodes which were randomly distribution in a feature space. We computed similarity index (*SI*) between nodes i, j with exponential design functions. By modulating the shape of the design function, we were be able to simulate networks with various distributions of similarity index (**Supplementary Fig. 12a, b**).

$$SI = \frac{1}{a + \exp\left(b * dist_{i,j}\right)} + c \qquad (eq. 19)$$

$$dist_{i,j}: distance\ between\ node\ i, j\ in\ feature\ spece$$

In a similarity network, the connectivity between two nodes was defined by rectified similarity index following threshold (eq. 20). By changing the threshold, we can simulate networks with different edge density. We have tested rectifiers with various shapes by modulating the k value in eq. 21 (**Supplementary Fig. 12c**). By

$$p\left(k_{i,j} = 1\right) = \max\left\{0, f\left(SI_{i,j} - threshold\right)\right\} \qquad (eq. 20)$$

$$f(x) = \frac{x - k * x}{k - 2k * abs(x) + 1} \qquad (eq. 21)$$

Preferential attachment mechanism

In preferential attachment network, a small network was initialized randomly. Then a new node was added to the network on each step, and made m connections with the existing network. The probability of the new node l being connected with exist node i was defined by the connectivity of node i.

$$p(k_{l,i} = 1) = \frac{k_i}{\sum_n k_n};\ k_i = \sum_j k_{i,j} \qquad (eq.22)$$

PAFS

In the PAFS simulation, the connectivity between all simulated nodes were sparsely initialized following similarity mechanism. The similarity index was computed by $SI = \frac{1}{1+\exp(dist_{i,j})}$ and following a sharp super-linear rectifier (k=-0.999).

Then, we randomly pick a node in the network at a step. The node make about m new edges with the rest of the network following preferential attachment mechanism (eq. 22). We iterated this process so that a subset ($\alpha\%$) of nodes were involved in this process. We have tested network structure by modulating the number of new connections m made by each nodes and the subset of nodes being affected ($\alpha\%$).

Reference

1. Wang, Q. & Burkhalter, A. Area Map of Mouse Visual Cortex. *J. Comp. Neurol.* **502**, 339–357 (2007).

2. Chiang, A. S. *et al.* Three-dimensional reconstruction of brain-wide wiring networks in drosophila at single-cell resolution. *Curr. Biol.* **21**, 1–11 (2011).

3. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).

4. Cohen, M. R. & Kohn, A. Measuring and interpreting neuronal correlations. *Nat. Neurosci.* **14**, 811–819 (2011).

5. Shadlen, M. N. & Newsome, W. T. The Variable Discharge of Cortical Neurons : Implications for Connectivity , Computation , and Information Coding. *J. Neurosci.* **18**, 3870–3896 (1998).

6. Rabinowitz, N. C., Goris, R. L., Cohen, M. & Simoncelli, E. P. Attention stabilizes the shared gain of V4 populations. *Elife* 1–24 (2015). doi:10.7554/eLife.08998

7. Hofer, S. B. *et al.* Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nat. Neurosci.* **14**, 1045–1052 (2011).

8. Middleton, J. W., Omar, C., Doiron, B. & Simons, D. J. Neural Correlation Is Stimulus Modulated by Feedforward Inhibitory Circuitry. *J. Neurosci.* **32**, 506–518 (2012).

9. Ecker, A. S. *et al.* State Dependence of Noise Correlations in Macaque Primary Visual Cortex. *Neuron* **82**, 235–248 (2014).

10. Renart, A. *et al.* The asynchronous state in cortical circuits. *Science (80-.).* **327**, 587–590 (2010).

11. Ocker, G. K. *et al.* From the statistics of connectivity to the statistics of spike times in neuronal networks. *Curr. Opin. Neurobiol.* **46**, 109–119 (2017).

12. Churchland, M. M. *et al.* Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* **13**, 369–378 (2010).

13. Dadarlat, M. C. & Stryker, M. P. Locomotion enhances neural encoding of visual stimuli in mouse V1. (2017).

14. Rose, T., Jaepel, J., Hübener, M. & Bonhoeffer, T. Cell-specific restoration of stimulus preference after monocular deprivation in the visual cortex. *Science* **352**, 1319–1322 (2016).

15. Reimer, J. *et al.* Pupil Fluctuations Track Fast Switching of Cortical States during Quiet Wakefulness. *Neuron* **84**, 355–362 (2014).

16. Poulet, J. F. A. & Petersen, C. C. H. Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. **454**, (2008).

17. Litwin-kumar, A. & Doiron, B. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat. Neurosci.* **15**, 1498–1505 (2012).

18. Stringer, C. *et al.* Inhibitory control of correlated intrinsic variability in cortical networks. *Elife* **5**, 1–33 (2016).

19. Lin, I. C., Okun, M., Carandini, M. & Harris, K. D. The Nature of Shared Cortical Variability. *Neuron* **87**, 645–657 (2015).

20. Ecker, A. S., Denfield, G. H., Bethge, M. & Tolias, A. S. On the Structure of Neuronal Population Activity under Fluctuations in Attentional State. *J. Neurosci.* **36**, 1775–1789 (2016).

21. Law, C.-T. & Gold, J. I. Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nat. Neurosci.* **12**, 655–663 (2009).

22. Rosenbaum, R., Smith, M. A., Kohn, A., Rubin, J. E. & Doiron, B. The spatial structure of correlated

neuronal variability. *Nat. Neurosci.* **20**, 107–114 (2017).

23. Smith, M. A. & Sommer, M. A. Spatial and Temporal Scales of Neuronal Correlation in Visual Area V4. *J. Neurosci.* **33**, 5422–5432 (2013).

24. Smith, M. A. & Kohn, A. Spatial and Temporal Scales of Neuronal Correlation in Primary Visual Cortex. *J. Neurosci.* **28**, 12591–12603 (2008).

25. Huang, C. *et al.* Circuit Models of Low-Dimensional Shared Variability in Cortical Networks. *Neuron* (2019). doi:10.1101/217976

26. Ruff, X. D. A. & Cohen, X. M. R. Attention Increases Spike Count Correlations between Visual Cortical Areas. *J. Neurosci.* **36**, 7523–7534 (2016).

27. Ruff, X. D. A. & Cohen, X. M. R. Stimulus Dependence of Correlated Variability across Cortical Areas. *J. Neurosci.* **36**, 7546–7556 (2016).

28. Ruff, D. A. & Cohen, M. R. A normalization model suggests that attention changes the weighting of inputs between visual areas. (2017). doi:10.1073/pnas.1619857114

29. Schaffelhofer, S. & Scherberger, H. Object vision to hand action in macaque parietal , premotor , and motor cortices. *Elife* 1–24 (2016).

30. Harris, K. D., Quiroga, R. Q., Freeman, J. & Smith, S. L. Improving data quality in neuronal population recordings. *Nat. Neurosci.* **19**, 1165–1174 (2016).

31. Jeffrey N. Stirman, Ikuko T. Smith, Michael W. Kudenov, S. L. S. Wide field-of-view, multi-region two-photon imaging of neuronal activity. *Nat. Biotechnol.* **34**, 857–862 (2016).

32. Andermann, M. L., Kerlin, A. M., Roumis, D. K., Glickfeld, L. L. & Reid, R. C. Functional Specialization of Mouse Higher Visual Cortical Areas. *Neuron* **72**, 1025–1039 (2011).

33. Marshel, J. H., Garrett, M. E., Nauhaus, I. & Callaway, E. M. Functional specialization of seven mouse visual cortical areas. *Neuron* **72**, 1040–1054 (2011).

34. Zhuang, J. *et al.* An extended retinotopic map of mouse cortex. *Elife* **6**, 1–29 (2017).

35. Vries, S. E. J. de *et al.* A large-scale, standardized physiological survey reveals higher order coding throughout the mouse visual cortex. *bioRxiv* 359513 (2018). doi:10.1101/359513

36. Glickfeld, L. L., Andermann, M. L., Bonin, V. & Reid, R. C. Cortico-cortical projections in mouse visual cortex are functionally target specific. *Nat. Neurosci.* **16**, 219–26 (2013).

37. Han, Y. *et al.* The logic of single-cell projections from visual cortex. *Nature* **556**, 51 (2018).

38. Wang, Q., Sporns, O. & Burkhalter, A. Network Analysis of Corticocortical Connections Reveals Ventral and Dorsal Processing Streams in Mouse Visual Cortex. *J. Neurosci.* **32**, 4386–4399 (2012).

39. Kim, M., Znamenskiy, P., Iacaruso, M. F. & Mrsic-flogel, T. D. Exclusive functional subnetworks of intracortical projections neurons in primary visual cortex. *bioRxiv* (2017).

40. Berezovskii, V. K., Nassi, J. J. & Born, R. T. Segregation of feedforward and feedback projections in mouse visual cortex. *J. Comp. Neurol.* **519**, 3672–3683 (2011).

41. D'Souza, R. D., Meier, A. M., Bista, P., Wang, Q. & Burkhalter, A. Recruitment of inhibition and excitation across mouse visual cortex depends on the hierarchy of interconnecting areas. *Elife* **5**, 1–19 (2016).

42. Harris, K. D. & Shepherd, G. M. G. The neocortical circuit: themes and variations. *Nat. Neurosci.* **18**, 170–181 (2015).

43. Seabrook, T. A., Burbridge, T. J., Crair, M. C. & Huberman, A. D. Architecture, Function, and Assembly of the Mouse Visual System. *Annu. Rev. Neurosci.* **40**, 499–538 (2017).

44. Madisen, L. *et al.* Transgenic Mice for Intersectional Targeting of Neural Sensors and Effectors with High Specificity and Performance. *Neuron* **85**, 942–958 (2015).

45. Chen, T. W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).

46. Smith, I. T., Townsend, L. B., Huh, R., Zhu, H. & Smith, S. L. Stream-dependent development of higher visual cortical areas. *Nat. Neurosci.* **20**, 200–208 (2017).

47. Garrett, M. E., Nauhaus, I., Marshel, J. H. & Callaway, E. M. Topography and Areal Organization of Mouse Visual Cortex. *J. Neurosci.* **34**, 12587–12600 (2014).

48. Schulz, D. P. A., Sahani, M. & Carandini, M. Five key factors determining pairwise correlations in visual cortex. *J. Neurophysiol.* **114**, 1022–1033 (2015).

49. Greenberg, D. S., Houweling, A. R. & Kerr, J. N. D. Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nat. Neurosci.* **11**, 749–751 (2008).

50. Ko, H. *et al.* Functional specificity of local synaptic connections in neocortical networks. *Nature* **473**, 87–91 (2011).

51. Ecker, A. S. *et al.* Decorrelated Neuronal Firing in Cortical Microcircuits. *Science (80-.).* **327**, 584–586 (2010).

52. Yatsenko, D. *et al.* Improved Estimation and Interpretation of Correlations in Neural Circuits. *PLoS Comput Biol* **11**, 1–28 (2015).

53. Henriksen, S., Pang, R. & Wronkiewicz, M. A simple generative model of the mouse mesoscale connectome. *Elife* **5**, 1–19 (2016).

54. Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M. & Harris, K. D. High-dimensional geometry of population responses in visual cortex. *Nature* (2019). doi:10.1038/s41586-019-1346-5

55. Lynn, C. W. & Bassett, D. S. The physics of brain network structure , function , and control. *Nat. Rev. Phys.* **1**, 318–332 (2019).

56. Bullmore, E. & Sporns, O. Complex brain networks : graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10**, 106–198 (2009).

57. Albert, R., Jeong, H. & Baraba´ si, A.-L. B. Error and attack tolerance of complex networks. *Science (80-.).* **406**, 378–382 (2000).

58. Ruff, D., Xue, C., Kramer, L., Baqai, F. & Cohen, R. M. Low rank mechanisms underlying flexible visual representations. *bioRxiv* (2019).

59. Smith, S. L. & Häusser, M. Parallel processing of visual space by neighboring neurons in mouse visual cortex. *Nat. Neurosci.* **13**, 1144–1149 (2010).

60. Kalatsky, V. A., Stryker, M. P. & Foundation, W. M. K. New Paradigm for Optical Imaging : Temporally Encoded Maps of Intrinsic Signal. *Neuron* **38**, 529–545 (2003).

61. Pachitariu, M. *et al.* Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv* 061507 (2017). doi:10.1101/061507

62. Pnevmatikakis, E. A., Merel, J., Pakman, A. & Paninski, L. Bayesian spike inference from calcium imaging data. 0–5

63. Froudarakis, E. *et al.* Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. *Nat. Publ. Gr.* **17**, 851–857 (2014).

64. Macke, J. H., Berens, P., Ecker, A. S., Tolias, A. S. & Bethge, M. Generating Spike Trains with Specified Correlation Coefficients. *Neural Comput.* **21**, 397–423 (2009).

65. Liang, F., Song, Q. & Qiu, P. An Equivalent Measure of Partial Correlation Coefficients for High-Dimensional Gaussian Graphical Models. *J. Am. Stat. Assoc.* **110**, 1248–1265 (2015).

66. Chandrasekaran, V., Parrilo, P. A. & Willsky, A. S. Latent variable graphical model selection via convex optimization. *Annu. Stat.* **40**, 1935–1967 (2012).

67. Erdös, P. & Rėnyi, A. . . ., P,, by selecting N edges from the ; i) possible edges Px (1 5. *Publ. Math.* **6**, 290–297 (1959).

68. Sporns, O. & Zwi, J. D. The Small World of the Cerebral Cortex. *Neuroinformatics* **2**, 145–162 (2004).

69. Rubinov, M. & Sporns, O. NeuroImage Complex network measures of brain connectivity : Uses and interpretations. *Neuroimage* **52**, 1059–1069 (2010).

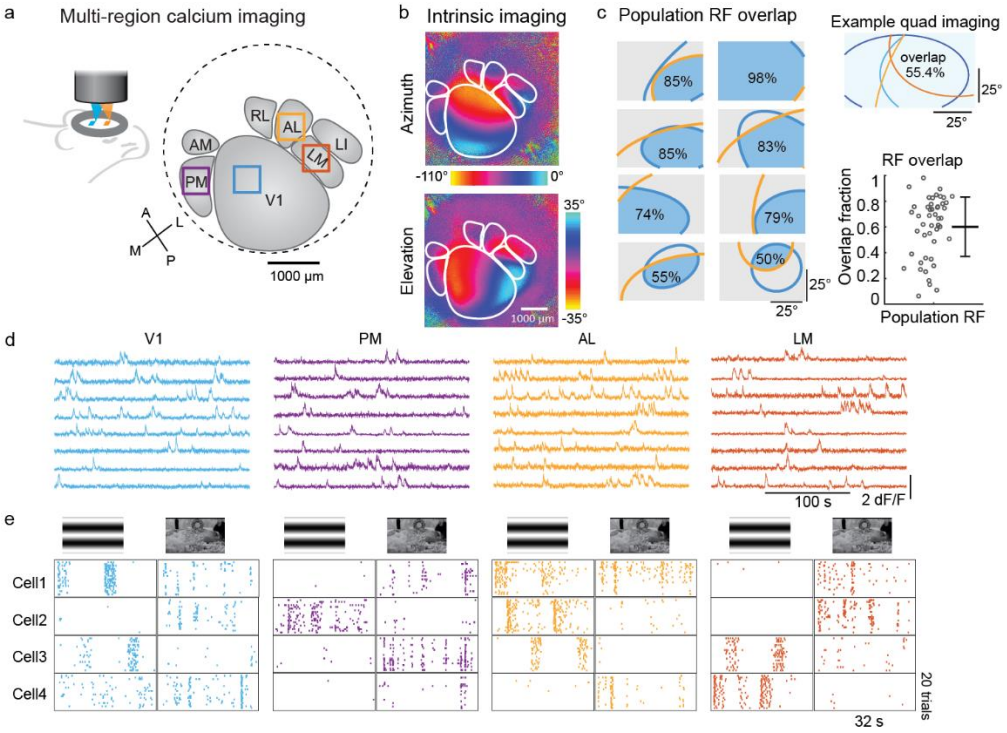

Figure 1. Multi-region two-photon calcium imaging. (**a**) Cartoon of multi-region two-photon imaging of mouse V1 and HVAs, using custom wide field-of-view microscope. Example imaging session of simultaneous recording session of V1, LM, AL and PM. Blue and orange square indicate a 500 μm^2 imaging window. (**b**) Example intrinsic signal imaging of mouse visual areas. (**c**) Left: example population RFs of simultaneously imaged populations. Blue and orange contours indicate Gaussian profile of population RF of neurons from different visual areas, and blue shade indicates the overlap region of population RF of two simultaneously imaging regions. Values indicate the fraction of overlap. Upper right: example population RF of the quartic-region imaging shown in (**a**). Lower right: summarize the fraction of population RF overlap of individual experiments (gray circle). Error bar indicate the mean and standard division. (**d**) Example neuropil corrected calcium traces of multi-region calcium imaging to drifting grating stimuli. (**e**) Spike raster plots of example neurons from V1, PM, AL and LM responding to drifting grating stimuli at 8 directions (SF: 0.04 cpd, TF: 2 Hz, directions: 0 – 315°) and to naturalistic video.

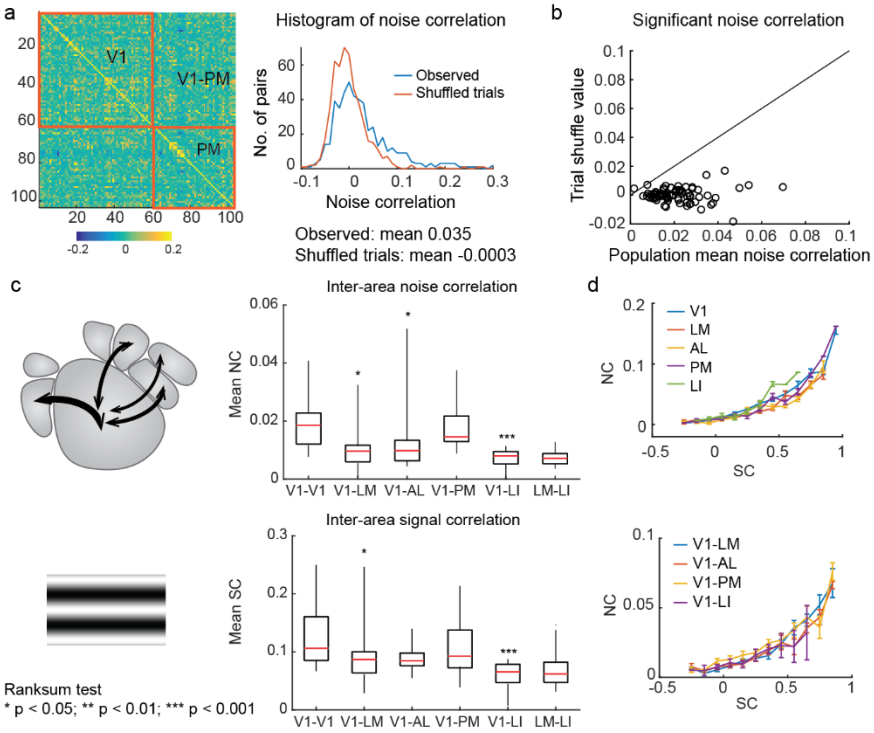

Figure 2. Within- and inter-area noise correlation. (**a**) Left: example covariance matrix of a quadratic recording of V1 and PM. Right: NC of an imaging session computed by trial aligned (blue) and trial shuffled data (red). (**b**) Population mean NC is significantly larger than the residual correlation after trial shuffling. Gray dots indicate the value of individual experiments. (**c**) Left: cartoon shows relative strength of inter-area correlations. Right: Boxplot of population mean inter-area NC (upper) and SC (lower) to gratings. (NC: $P_{V1\text{-}LM} = 0.016$, $P_{V1\text{-}AL} = 0.03$, $P_{V1\text{-}PM} = 0.75$, $P_{V1\text{-}LI} = 0.0008$, $P_{LM\text{-}LI} = 0.0016$; SC: $P_{V1\text{-}LM} = 0.039$, $P_{V1\text{-}AL} = 0.05$, $P_{V1\text{-}PM} = 0.26$, $P_{V1\text{-}LI} = 0.0008$, $P_{LM\text{-}LI} = 0.015$; number of experiments: $N_{V1\text{-}V1} = 13$, $N_{V1\text{-}LM} = 21$, $N_{V1\text{-}AL} = 16$, $N_{V1\text{-}PM} = 19$, $N_{V1\text{-}LI} = 11$, $N_{LM\text{-}LI} = 8$). (**d**) Within- and inter-area noise correlations are positively related with SC in V1 and HVAs to Sine wave drifting grating stimuli.

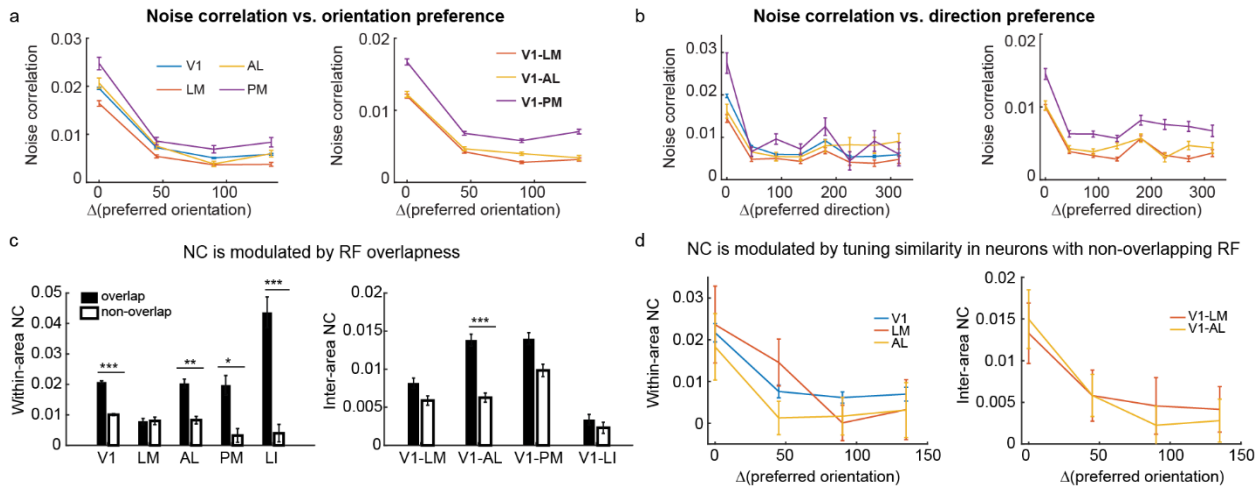

Figure 3: Noise correlation is modulated by tuning similarity over millimeter distance scales. (**a**) Plot within- and inter-area noise correlation as a function of difference in preferred orientation (**b**) Plot within- and inter-area noise correlation as a function of difference in preferred direction (**a, b**) Only orientation selective neurons (OSI > 0.5) were included in the analysis of noise correlation vs. preferred orientation difference. Only direction selective neurons (DSI > 0.3) were included in the analysis of noise correlation vs. preferred direction difference. (**c**) Within-area (left) and inter-area (right) NC is significantly higher in neuron pairs with shared RF (within-area, $P_{V1} = 5.6 \times 10^{-18}$, $P_{LM} = 0.98$, $P_{AL} = 0.0063$, $P_{PM} = 0.033$, $P_{LI} = 9.1 \times 10^{-4}$; $N_{V1} = 19358$, $N_{LM} = 1199$, $N_{AL} = 1753$, $N_{PM} = 447$, $N_{LI} = 303$; inter-area: $P_{V1\text{-}LM} = 0.33$, $P_{V1\text{-}AL} = 0.0007$, $P_{V1\text{-}PM} = 0.15$, $P_{V1\text{-}LI} = 0.82$; $N_{V1\text{-}LM} = 4834$, $N_{V1\text{-}AL} = 5442$, $N_{V1\text{-}PM} = 3253$, $N_{V1\text{-}LI} = 2889$; t-test). (**d**) Noise correlation of neurons with non-overlapping RF is modulated by orientation tuning similarity. (within-area, $P_{V1} = 7.4 \times 10^{-14}$, $P_{LM} = 0.03$, $P_{AL} = 0.019$; $N_{V1} = 3401$, $N_{LM} = 181$, $N_{AL} = 284$; inter-area: $P_{V1\text{-}LM} = 0.019$, $P_{V1\text{-}AL} = 0.0004$; $N_{V1\text{-}LM} = 650$, $N_{V1\text{-}AL} = 998$; t-test).

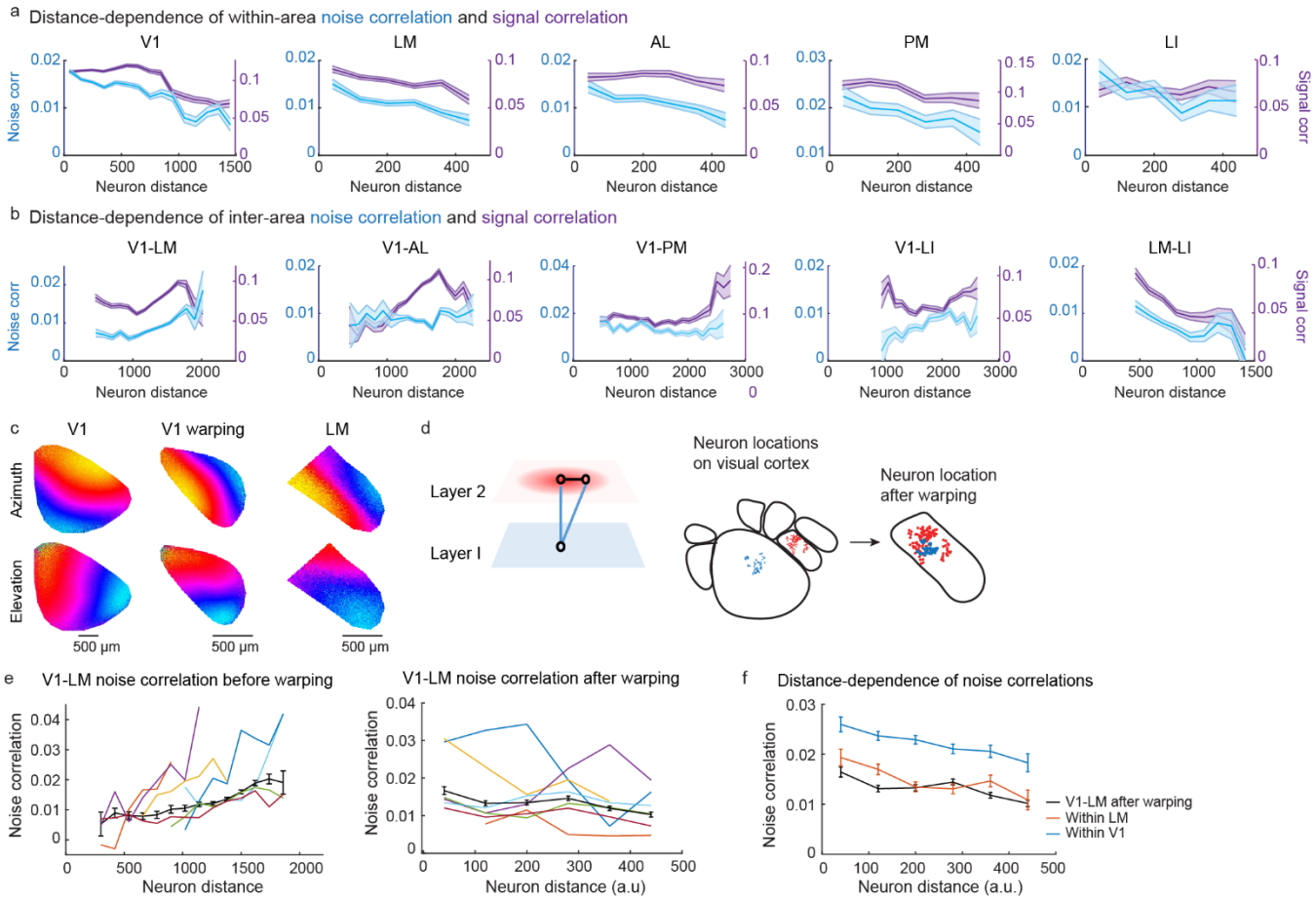

Figure 4. Distance-dependence of noise correlation. (**a**) Distance-dependence of within-area NC (blue) and SC (purple) to sine-wave drifting gratings (NC: V1, r = -0.044, p = 3.5 x 10^{-40}; LM, r = -0.026, p = 0.0009; AL, r = -0.05, p = 7.7 x 10^{-7}; PM, r = -0.048, p = 0.0026; LI, r = -0.025, p = 0.17. SC: V1, r = -0.03, p =9.8 x 10^{-24}; LM, r = -0.036, p = 4.4 x 10^{-6}; AL, r = -0.028, p = 0.0067; PM, r = -0.048, p = 0.005; LI, r = -0.037, p = 0.047; Pearson correlation). (**b**) Distance-dependence of inter-area NC (blue) and SC (purple) to sine-wave drifting gratings (NC: V1-LM, r = 0.058, p = 5.5 x 10^{-41}; V1-AL, r = 0.013, p = 0.02; V1-PM, r = -0.05, p = 2.8 x 10^{-14}; V1-LI, r = 0.028, p = 0.0007; LM-LI, r = -0.08, p = 3.2 x 10^{-15}; SC: V1-LM, r = 0.073, p = 2.6 x 10^{-64}; V1-AL, r = 0.073, p = 1.7 x 10^{-35}; V1-PM, r = -0.01, p = 0.047; V1-LI, r = 0.056, p = 8.2 x 10^{-12}; LM-LI, r = -0.084, p = 2.8 x 10^{-53}; Pearson correlation). (**a, b**) Solid lines indicate mean values and shaded areas indicate standard error of mean. Each distance bin contains >50 data points. (**c**) Example affine transformation of ISOI maps. The left shows the original V1 map, the middle is the V1 map after affine transformation and the right is the original LM map. (**d**) Left, cartoon of two recurrent layer with aligned retinotopic map. Right: neuron location on the visual cortex before and after warping. (**e**) Distance-dependent increasing of V1-LM NC to sine-wave drifting gratings before (left) and after retinotopic warping (right). Individual experiments with significant distance-dependence are in shown in colored curves. The black curve shows the population mean and standard error (Pearson correlation, before warping, r = 0.066, p = 1.3 x 10^{-36}; after warping, r = -0.026, p = 3.8 x 10^{-7}). (**f**) Distance-dependence of within-area NCs of paired recorded V1 and LM, and inter-area NC of V1-LM after retinotopic warping.

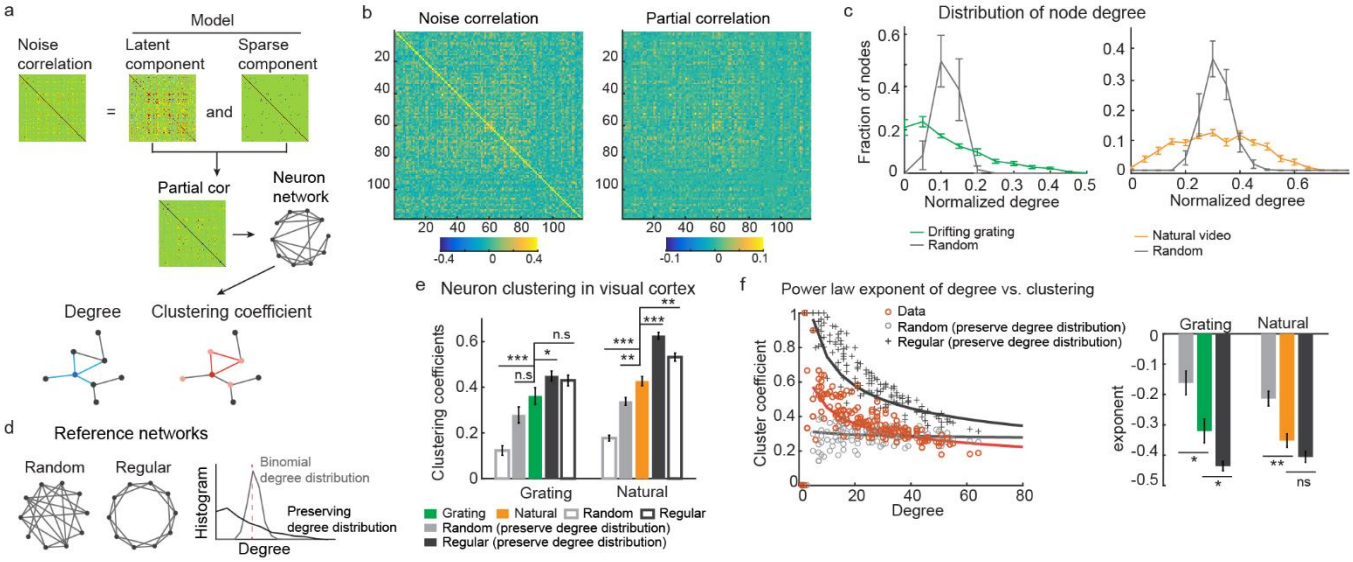

Figure 5. Network structure of the population noise correlation matrix (**a**) Schematic of characterizing the network structure of the partial correlation network of the neuron population. The partial correlation was computed by imposing a regularized model with latent factor and sparse components to the covariance matrix. (**b**) Example noise correlation matrix (left) and partial correlation matrix (right). (**c**) The normalized degree distribution of networks evoked by drifting gratings (left) and naturalistic movies (right), and random network with binomial degree distribution. Skewness of degree distribution is about 0.76 ± 0.13 for drifting grating and 0.5 ± 0.09 for natural video. (**d**) Reference random and regular networks w/o preserving the degree distribution of actual network. (**e**) Mean clustering coefficients of actual are higher than or not differed to that of random networks preserving degree distribution (grating: $p = 0.17$ ($n = 6$); natural: $p = 0.0036$ ($n = 7$)), or not preserving degree distribution (grating: $p = 0.17$; natural: $p = 0.002$). Mean clustering coefficients of actual are lower than that of regular networks preserving degree distribution (grating: $p = 0.01$; natural: $p = 8 \times 10^{-6}$), or not preserving degree distribution (grating: $p = 0.01$; natural: $p = 8 \times 10^{-6}$). (**f**) Clustering coefficients of individual nodes are negatively related with its degree, and is well fitted by power law. The absolute value of the power law exponent of actual neuron networks is larger than that of random networks preserving the degree distribution (grating: $p = 0.017$ ($n = 6$); natural: $p = 0.0013$ ($n = 7$)), and smaller than or similar to that of regular networks preserving the degree distribution (grating: $p = 0.02$ ($n = 6$); natural: $p = 0.08$ ($n = 7$)). (**e**, **f**) t-test.

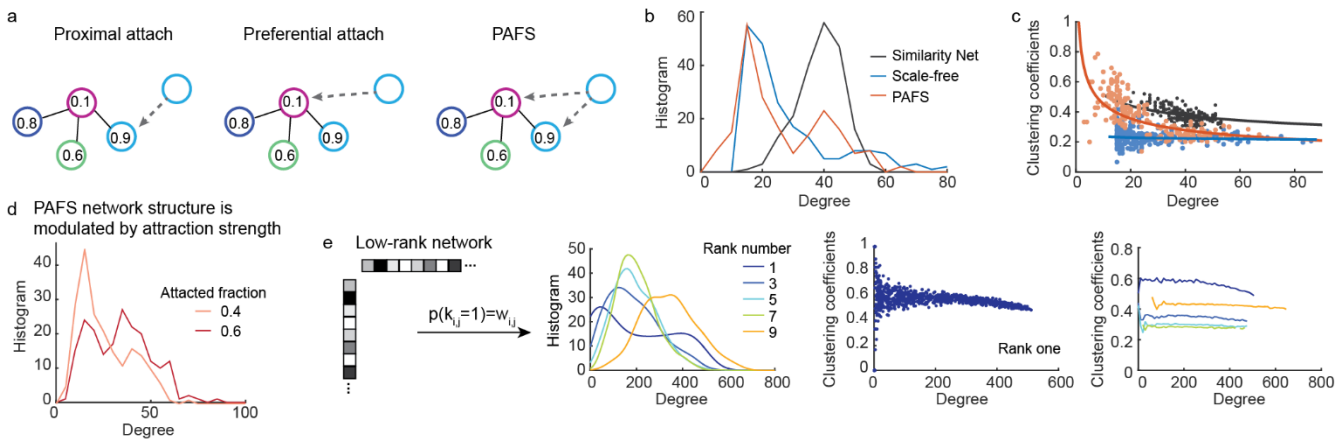

Figure 6. Reproduce functional neuron network structure. (**a**) Cartoons indicate three mechanisms for network simulation. Numbers indicate the feature similarity index. From left to right: feature similarity mechanism that a node tends to connect to nodes with similar feature; preferential attachment that a node tends to connect to nodes with high degree; PAFS that a node tens to connect to both nodes with similar feature and high degree. (**b, c**) Example simulation results with 200 nodes using the three mechanism in (**a**). (**b**) Shows the degree distribution (edge density is 0.14-0.19). (**c**) Shows the joint distribution of clustering coefficients and node degrees (mean clustering coefficients are 0.38, 0.23 and 0.37, and exponents are -0.22, 0 and -0.35 for feature similarity, preferential attachment and PAFS respectively). (**d**) Example degree distribution of PAFS networks with 200 nodes. The edge density is 0.12 and 0.17, and the mean clustering coefficient is 0.32 and 0.37 for the two simulations with 40% and 60% attracted subset. Also, **Supplementary Fig. 13**. (**e**) Schematic of low-rank networks. Right panels show the example degree distribution, and the joint distribution of clustering coefficients and node degrees of low-rank networks of 1000 nodes. No significant degree-dependent decay of clustering coefficients.