

¹ Hebb’s Vision: The Structural Underpinnings of Hebbian Assemblies

² J Wagner-Carena^{1,2,3}, S Kate^{*1,5}, T Riordan^{*1}, R Abbasi-Asl^{1,4}, J Aman¹, A Amster¹, AL
³ Bodor¹, D Brittain¹, JA Buchanan¹, MA Buice¹, DJ Bumbarger¹, F Collman¹, NM da
⁴ Costa¹, DJ Denman¹, SEJ de Vries¹, E Joyce¹, D Kapner¹, CW King¹, JD Larkin¹, J Lecoq¹,
⁵ G Mahalingam¹, D Millman¹, J Mölter⁶, C Morrison¹, RC Reid¹, CM Schneider-Mizell¹, S
⁶ Daniel¹, S Suckow¹, KT Takasaki¹, M Takeno¹, R Torres¹, D Vumbaco¹, J Waters¹, DG
⁷ Wyrick¹, W Yin¹, J Zhuang¹, S Mihalas^{†1}, and S Berteau^{†1}

⁸ ¹*Allen Institute, Seattle, WA, USA*

⁹ ²*Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA*

¹⁰ ³*Neuroscience Graduate Program, University of California, San Francisco, San Francisco, CA, USA*

¹¹ ⁴*Department of Neurology, University of California, San Francisco, San Francisco, CA, USA*

¹² ⁵*Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA*

¹³ ⁶*Department of Mathematics, School of Computation, Information and Technology, Technical
University of Munich, Germany*

¹⁵ **1 Abstract**

¹⁶ In 1949, Donald Hebb proposed that groups of neurons that activate stereotypically
¹⁷ form the organizational building blocks of perception, cognition, and behavior. Finding
¹⁸ the structural underpinning of such assemblies has been technically challenging, due to a
¹⁹ lack of large-scale structure-activity maps. Here, we analyze this relation using a novel
²⁰ dataset that links *in vivo* optical physiology to connectivity using postmortem elec-
²¹ tron microscopy (EM). From the fluorescence traces, we extract neural assemblies from
²² higher-order correlations in neural activity. Physiologically, we show that these assem-
²³ blies exhibit properties consistent with Hebb’s theory, including more reliable responses
²⁴ to repeated natural movie inputs than size-matched random ensembles and superior
²⁵ decoding of visual stimuli. Structurally, we find that neurons that participate in assem-
²⁶ blies are significantly more integrated into the structural network than those that do
²⁷ not. Contrary to Hebb’s original prediction, we do not observe a marked increase in
²⁸ the strength of monosynaptic excitatory connections between cells participating in the
²⁹ same assembly. However, we find significantly stronger indirect feed-forward inhibitory
³⁰ connections targeting cells in other assemblies. These results show that assemblies can
³¹ be useful components of perception, and, surprisingly, they are delineated by mutual
³² inhibition.

^{*}These individuals contributed equally to this work as second authors

[†]These individuals contributed equally to this work as senior authors

33 2 Introduction

34 Since Hebb's 1949 monograph, *The Organization of Behavior*[1], cell assemblies have
35 retained a persistent place in the imagination of the neuroscience community, both as
36 a prospective unit of functional organization and as a compellingly likely consequence
37 of simple rules of synaptic plasticity, such as the “fire-together wire-together” synapses
38 dubbed ‘Hebbian Synapses’ by Yves Frégnac in 1986 [2]. Throughout this work, we
39 follow Hebb in using the term assembly to refer to overlapping sets of neurons that acti-
40 vate in a reproducible pattern with high fidelity; in much of the contemporary literature,
41 such groups are also called cell ensembles or simply ensembles. The theory continues to
42 stimulate research on the activity-related aspects of assemblies, and the organization of
43 these processing modules is thought to operate through stable recurrent activity [3, 4].
44 Incorporating a behavioral approach, experimental evidence implies that the response
45 of discrete cell populations akin to assemblies may also have a causal link with motor
46 functions [5]. In particular, there has been confirmation of sequential activation within
47 neural assemblies of the primary visual cortex (V1) [6] as well as ongoing ‘replay’ of
48 coactivity in the absence of stimuli [7]. There has also been some limited evidence of
49 cell assemblies’ stability over long periods of time [8].

50 However, research examining the structural aspects of Hebbian cell assemblies has
51 primarily focused on the potentiation of synapses between excitatory neurons. This em-
52 phasis is perhaps not surprising, given that this potentiation has a place of prominence
53 in Hebb’s original formulation of the theory [1]. In 1963 [9], Hebb himself acknowledged
54 that his excitation-only formulation of assembly theory was a concession to the state
55 of research on inhibitory synapses at that point in time; synaptic inhibition of neural
56 activity was not confirmed until 1952 [10]. Over the following decades, the association
57 between Hebbian assemblies and excitatory potentiation has itself been reinforced by
58 the computational plausibility of excitatory plasticity as a mechanism of assembly for-
59 mation [11, 12, 13], and the discovery of Hebbian synapses. However, such changes were
60 never a necessary precondition for assembly formation. Mechanistically, there exists a
61 broad range of plausible solutions to assembly formation, ranging from excitatory mod-
62 ulation against a backdrop of relatively stable inhibitory strengths to the opposite, in
63 which formation relies solely on inhibitory modulation contrasting with stable excitatory
64 connections [14].

65 The approach of this work, using large-scale recordings of individual neurons to
66 identify and analyze assembly function, enters a developing tradition in the literature of
67 population dynamics [15, 16, 17, 18, 19, 20]. However, due to limitations in structural
68 analysis, few studies can relate the structure of assemblies to their function, as reviewed
69 by [14]. Electrophysiological datasets can allow for a highly reliable inference of connec-
70 tivity in the case of multi-patch recordings, but with very small numbers of cells in any
71 given study. Extracellular recordings can overcome this limitation, but produce biased
72 connectivity estimates, and therefore cannot be considered a gold standard.

73 To relate the correlated activity of a large number of neurons to their connections,
74 we used a novel large-scale multi-modal dataset: the Allen Brain Observatory V1 Deep
75 Dive (V1DD) (Fig. 1A). V1DD offers a combination of Ca^{2+} fluorescent recordings and
76 detailed EM reconstruction of neurons and synapses, including post-synaptic density
77 volumes. Taking advantage of advances in optical imaging techniques [21], V1DD pro-
78 vides multiple scans of high-quality simultaneous two-photon calcium imaging (2PCI)
79 recordings of thousands of excitatory neurons within the mouse primary visual cortex
80 (Fig. 1B). In addition to its functional recordings, V1DD also contains electron mi-
81 croscopy (EM) of the same tissue volume, which has uncovered the fine-scale anatomy
82 of the cubic millimeter volume of the brain (Fig. 1C). EM has been employed exten-
83 sively in large-scale datasets to map drosophila [22, 23], worm [24], mouse [25, 26], and

V1 Deep Dive: Large-Scale Multi-Modal Data

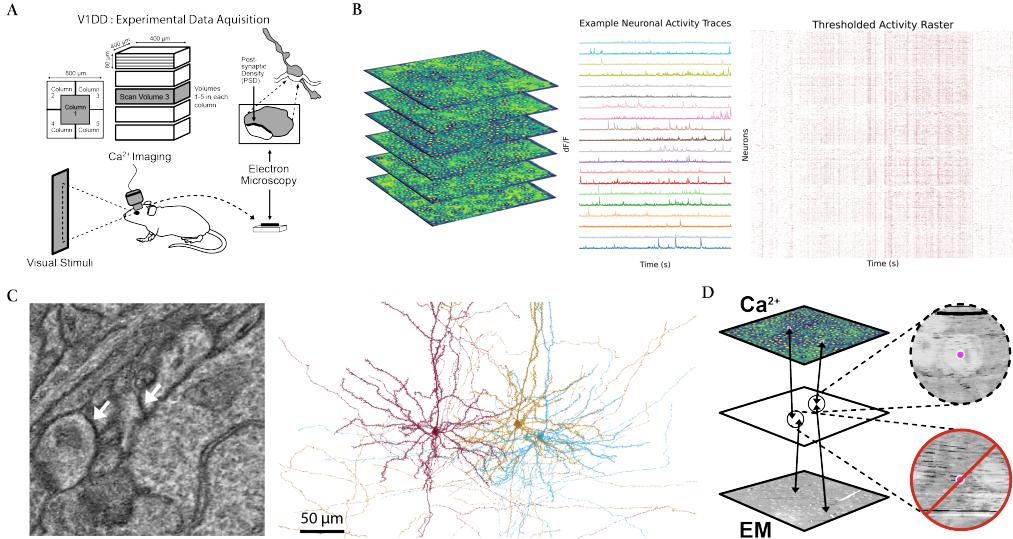


Figure 1: **(A)**. Schematic of experimental data acquisition for the V1 Deep Dive dataset (V1DD). V1DD consists of dense thousands of excitatory neurons in an $800 \times 800 \times 800 \mu\text{m}^3$ section of mouse V1, recorded during awake behaving imaging sessions, with no behavioral task. Our work focuses on the center column, particularly the third scan volume (pictured here in grey). Postmortem, the same tissue volume was fixed and imaged via transmission electron microscopy, allowing for reconstruction of synaptic connectivity, including post-synaptic density (PSD) volumes. **(B)** Each scan volume for in vivo imaging consisted of six stacked scan planes. Dense calcium activity allowed for the extraction of individual neuronal traces, with 20 example traces shown in addition to a raster plot of thresholded normalized activity for all 2708 neurons in Scan Volume 3 of Column 1. **(C)** Example of a microscopy view of connected neurons, with white arrows pointing to PSD. Reconstructed pyramidal cells corresponding to the left microscopy view are shown on the right. **(D)** Schematic showing the framework for coregistration of cells between the calcium recordings and the electron microscopy. Identified ROIs were mapped to an interstitial space (see Methods 5.1.4), where the correspondences were manually inspected.

even the human brain [27]. The combination of these two imaging modalities has been applied in very few datasets [28, 29, 26], particularly at this scale.

These advances offer the opportunity to examine the structural correlates of Hebbian cell assemblies at an unprecedented scale. We extract assemblies from a Ca²⁺ fluorescence imaging scan and examine the reliability of their activation and their functional significance in the encoding of visual stimuli. We then analyze the connectivity between neurons based on cells that were coregistered between fluorescence and structural EM scans (Fig. 1D). Deriving hypotheses directly from postulates advanced by Hebb, we test his predictions about synaptic sizes and connectivity within and across assemblies, producing results that suggest a significant role for inhibition in their formation and activation, different from what is traditionally assumed.

95 **3 Results**

96 **3.1 Neuronal Organization of Hebbian Assemblies**

97 We analyzed a scan of the optical imaging dataset consisting of 2708 excitatory
98 neurons recorded in parallel. The Similarity Graph Clustering (SGC) algorithm (Fig.
99 2A) generated 15 assemblies, which we ordered by size with ‘A 1’ representing the largest
100 assembly ($n = 1016$) and ‘A 15’ the smallest ($n = 23$). A subset ($n = 748$, 27.6 percent
101 of 2708 total) of neurons was assigned to no assemblies.

102 Their spatial distributions are shown in (Fig. 2C, D). Although spatial organization
103 is not clearly visible, statistical analysis using the Kolmogorov-Smirnov (KS) [30] test
104 revealed significant differences in the spatial organization of assemblies compared to that
105 of the entire neural space (Fig. 2E). In particular, there was marked organization in
106 most assemblies along the x-y plane, indicative of the retinotopic organization observed
107 in studies of the visual cortex [31].

108 Most, but not all, assemblies shared members with other assemblies (Fig. 2B).
109 For example, neurons assigned to ‘A 1’ were also assigned to ten other assemblies,
110 highlighting Hebb’s proposal that precisely timed phase sequences allow a set of shared
111 member neurons to participate in multiple assemblies.

112 **3.2 Correlation and Sparsity**

113 By construction, assemblies are expected to respond to distinct stimulus features,
114 with a coactivity correlation sufficiently low to prevent their merging into a single assem-
115 bly. Pearson correlation coefficients between assembly coactivity traces were significantly
116 lower than the coefficients between coactivity traces of random ensembles of the same
117 size distribution (see Methods) (Fig. 3A). As expected from traces derived from the av-
118 erage of population raster activity, both population groupings held significantly higher
119 correlations than individual cell activity raster correlations, regardless of the subset of
120 pyramidal cells being considered (assembly, non-assembly, or all individual cells).

121 To further characterize the functional properties of assemblies, we computed the Gini
122 coefficient [34] for each assembly’s activity trace. The Gini coefficient, a statistical mea-
123 sure of the ‘inequality’ of signal activity throughout the optical recording, revealed that
124 assemblies exhibited highly sparse activity patterns (Fig. 3B), exhibiting significantly
125 higher Gini coefficients than random ensembles (p -value : $6.23e - 5$). The coefficient for
126 each assembly ranged from 0.55 to 0.83, with particular assemblies with extreme spar-
127 sity, such as ‘A 13’ (0.79) and ‘A 15’ (0.83), exemplifying a high degree of functional
128 selectivity. Interestingly, this sparsity metric was not solely dependent on assembly size,
129 as intermediate-sized assemblies (e.g., ‘A 4’ through ‘A 12’) all exhibited similar Gini
130 coefficients of around 0.70.

131 **3.3 Assemblies Reliability in Stimuli Response**

132 A section of the visual stimuli presented to the mouse consisted of natural movies.
133 Across the hour scan time, twelve unique 15-second movie clips were repeatedly shown
134 eight times. To evaluate the functional reliability of neuronal assemblies, we analyzed
135 their responses to these stimuli. Reliability was quantified using an Oracle score, a
136 leave-one-out correlation metric (see Methods 5.3) that measures the consistency of an
137 activity trace across repeated presentations of the same stimulus. Oracle scores are used
138 to measure the reliability of neuronal responses [35]. To provide a baseline with which
139 to compare the functionality of these assemblies to tune to visual stimuli, the oracle
140 scores of assemblies’ coactivity traces were compared to the oracle scores of all neuronal
141 traces. In addition, we provide a population-level comparison with random ensembles.

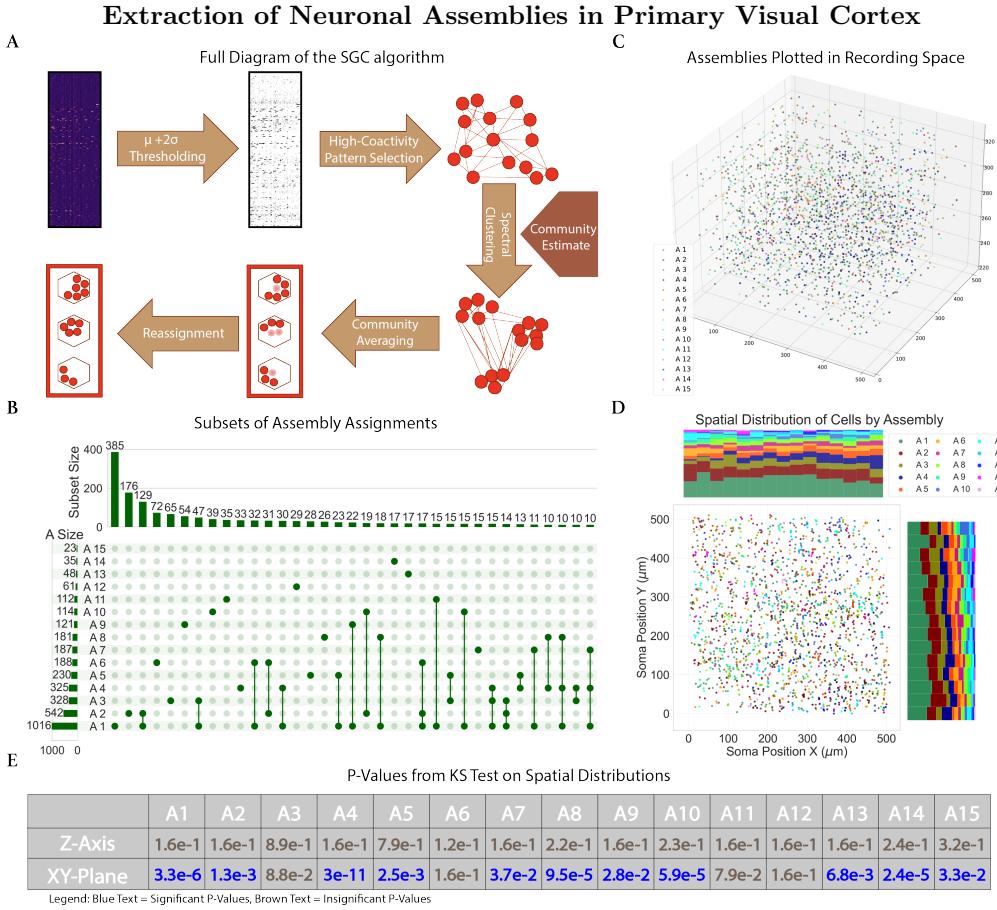


Figure 2: (A). SGC, an extraction algorithm for assemblies uniquely designed for calcium imaging data, groups frames of the calcium fluorescence input to determine when neurons in assemblies are coactivated. Figure adapted from Mölter et al. [32] (B). An UpSet [33] visualization of the subsets formed between assembly assignments. The histogram on the left represents the size of each individual assembly. The top histogram represents the size of the subsets between assemblies. Only subsets of ten neurons or greater were visualized. (C). Spatial positions of fifteen extracted assemblies in the three-dimensional recording field. There are 1960 neurons visualized, including neurons assigned to multiple assemblies (plotted only once) but not including the 748 neurons that were assigned to no assemblies. (D). Spatial distribution of assembly cells projected onto the x-y plane. Histograms for each axis are normalized to provide a per-bin proportional stack of the assembly distributions. (E). A table presenting the KS test results on each assembly's spatial distributions. Values colored in blue represent significant results ($p\text{-value} < 0.05$), while brown signifies insignificant results. All values have been corrected for false discovery under the Benjamini-Hochberg Procedure.

142 Assemblies exhibited significantly higher Oracle scores compared to the average re-
143 liability of individual neurons (p -values < 0.0001), indicating that the assemblies, as
144 populations, respond more consistently to visual stimuli (Fig. 3C). To ensure this
145 result was not merely due to the inclusion of highly reliable neurons within assem-
146 blies, we separately calculated Oracle scores for neurons within assemblies and those
147 assigned to no assemblies. We observed no significant difference between the cellular
148 sets (p -values > 0.25). This result suggests that the reliability of these assemblies is
149 derived from their collective activity rather than from the reliability of individual mem-
150 bers.

151 Since population coactivity is expected to be more reliable in the general case than
152 individual neurons, we compared our assemblies to random ensembles (defined as in
153 Methods 5.2). Assembly coactivity traces demonstrated higher Oracle scores than coac-
154 tivity traces computed in the same way as the assemblies for size-matched random
155 ensembles. Complementary results are seen in the reliability of responses to orienta-
156 tion gratings (see Supp. Fig. 6). Population activity during the presentation of these
157 gratings was used to characterize reliability as a function of orientation [36].

158 Further analysis revealed consistent patterns of high assembly coactivity during spe-
159 cific visual frames of these natural movies. We defined ‘trigger frames’ as moments when
160 assembly activity exceeded a baseline threshold (see Methods 5.2) and found that these
161 frames were highly consistent across repeated stimulus presentations. Example mean
162 trigger frame pixel values are presented in (Fig. 3F), along with those of the corre-
163 sponding size-matched random ensemble, and the squared difference. Visualizations of
164 these frames suggest that assembly activity responds to complex features in the natu-
165 ral movies. These collective results provide evidence of assemblies’ ability to serve as
166 functional populations with reliable and specific responses to visual stimuli.

167 3.4 Decoding Responses from Acute Visual Stimuli

168 We also assessed the ability of assemblies to decode visual stimuli by implementing
169 a classification framework comparing assemblies to random ensembles (Fig. 3D, E). We
170 employed a Multi-Layer Perceptron Classifier (MLPClassifier) to evaluate how well each
171 grouping could decode the identities of the twelve natural movie clips. These classifiers
172 have been shown to be effective in academic and clinical settings, with high levels of
173 accuracy and flexibility in available hyperparameters [37, 38].

174 The results of our classifier revealed that assemblies significantly outperformed ran-
175 dom ensembles in accuracy. Heatmaps of classification accuracy for natural movie clip
176 identities demonstrated that assemblies provided more reliable decoding across repeated
177 trials. A Mann-Whitney U-test confirmed this finding, with assembly accuracy signif-
178 icantly exceeding random ensembles for both overall performance (u -stat: 7546.0, p -
179 value: $6.07e - 5$) and diagonal elements (one-sided, u -stat: 131.5, p -value: $3.25e - 4$),
180 indicating enhanced stimulus-specific decoding.

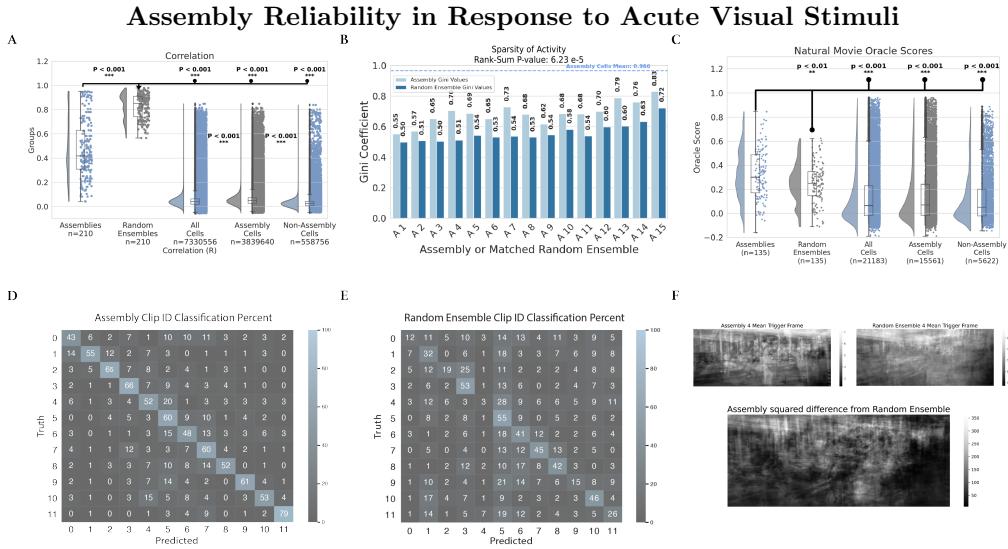


Figure 3: **(A)** Raincloud plot of pairwise correlations of coactivation between assemblies, size-matched random ensembles, and sets of individual cells. Coactivation for an individual cell is equivalent to a binary thresholded activity raster. **(B)**. Grouped bar plot of sparsity (measured by the Gini Coefficient) of coactivity over time in cell assemblies and the null-grouping of size-matched random ensembles. The random ensembles' coefficients are significantly smaller than the set of assembly coefficients (Wilcoxon Rank-Sum p-value: $6.234e-5$). The average sparsity of individual assembly cells and non-assembly cells is also plotted as nearly equal horizontal dashed lines. **(C)**. Raincloud plot illustrating the reliability of activity from assemblies and general neuronal populations in response to natural movies. Oracle scores of each assembly and random ensemble coactivity trace were plotted, as well as the scores of sets of individual cells. These oracle scores are computed for the concatenation of natural movie clips and their responses, rather than individual clips, in order to reduce the likelihood of sparse responses causing an artificially high reliability score. **(D, E)**. Heatmap illustrating decoding accuracy of natural movie clips with Assemblies and Random Ensembles. Heatmap values indicate the accuracy of clip decoding by the percentage of presentation. Clip IDs, indicating a unique natural movie clip, are balanced such that each clip has an equal frequency of presentation. Values in the assembly heatmap are significantly greater than the random ensemble heatmap (Mann-Whitney u-stat: 7546.0, p-value: $6.07e-5$; one-sided on diagonal elements u-stat: 131.5, p-value: $3.25e-4$). **(F)**. Example plots of the mean ‘trigger frame’ of assemblies and random ensemble during natural movies. Frames were generated by averaging the frames associated with peak coactivity. The natural movie frame was visually better reconstructed by the assembly activity than that of the random ensemble, as signified by the plotted squared difference.

181 3.5 Structural Organization of Assemblies

182 Coregistration between recorded activity and EM data provides us with a unique
 183 opportunity to explore the structural underpinnings of assemblies in the visual cortex.
 184 By mapping neural structure and connectivity at a micrometer resolution, we investi-
 185 gated the anatomical communication and organization of neurons with at least one
 186 shared assembly membership (shared assembly cells) compared to those with disjoint
 187 membership.

188 The strength of connections between the two groups was measured by performing a

189 Wilcoxon Rank Sum test to compare the connection weights (defined as the sum of PSD
190 volumes for all synapses between two cells). A Chi-squared test of Independence was
191 performed to determine whether the pairwise frequency of connections differed between
192 cells with shared assembly membership and those with disjoint assembly membership.
193 To investigate higher-order structural patterns, we performed similar tests on sets of
194 inbound and outbound disynaptic chains, as well as subdividing based on whether the
195 intermediate cell in each chain was inhibitory or excitatory. 5.11.2

196 Our initial analysis of first-order connectivity surprisingly revealed no significant
197 differences in the probability of direct monosynaptic connections between shared and
198 disjoint sets or the strengths of those connections (Fig. 4D,E). This finding is incongruous
199 with predictions that assemblies are defined by densely interconnected excitatory
200 neurons [1] and suggested that the defining structural characteristics of assemblies might
201 lie beyond simple pairwise connectivity metrics.

202 To investigate higher-order structural patterns of these assemblies, we conducted a
203 motif analysis. In a neural network, when a significant number of subgraphs containing
204 a small number of interconnected cells repeat a particular pattern (e.g., Cell Type A →
205 Cell Type B → Cell Type A), they are classified as a motif [39, 40], with each subgraph
206 counted as a motif instance. The frequency with which these motifs persist has been
207 predictive of correlation in similar neural networks [41].

208 Our analysis concentrated on second-order chain motifs, or structures consisting of
209 three neurons connected by two synaptic links (schematic shown in Fig. 4F,H). Second-
210 order neural motifs are divisible into various types based on the arrangement of their
211 synaptic connections. For this study, we prioritized chain motifs, as they allowed us to
212 utilize aspects of the EM dataset that are not yet coregistered to activity recordings
213 for the central elements in the chain, so long as the first and last cells in the chain are
214 coregistered.

215 Notably, while our excitatory chain analysis revealed no significant differences be-
216 tween shared and disjoint assembly memberships (Fig. 4F,G), an analysis of inhibitory
217 motifs convey a different story: disjoint assembly memberships exhibited significantly
218 stronger feed-forward inhibitory connections than shared memberships (Fig. 4H,I). Log-
219 scaled plotting of the chain weights (Fig. 4I, inset) revealed a complete lack of overlap
220 between SEM-based confidence intervals, suggesting that this is a result with both reliabil-
221 ity and a non-trivial effect size. Notably, this result was insignificant when restricting
222 our analysis to only the inhibitory synapse or only the excitatory synapse within the
223 feed-forward inhibitory chain, suggesting that both connections may play a role.

224 The significantly stronger feed-forward inhibitory connections suggest a mechanism
225 of mutual inhibition that may regulate the interaction between distinct assemblies, pre-
226 venting excessive coactivation and ensuring the discriminability of assembly responses
227 to inbound information. Indeed, we find a negative Pearson r correlation coefficient be-
228 tween the disjoint feed-forward inhibitory chain weights and the correlation of the two
229 disjoint assemblies' coactivity traces (r statistic : -0.19, p -value : 0.0048; see Methods
230 5.11.4).

231 All together, our structural results reveal that assemblies are not anatomically de-
232 fined solely by a fundamental increase in local connectivity, but through their higher-
233 order patterns of organization. Stronger feed-forward inhibitory connections in disjoint
234 memberships suggest a possible mechanism for maintaining functional segregation be-
235 tween assemblies. These structural insights provide intuition on the mechanistic basis
236 that drives the functional properties of assemblies, reinforcing their contribution as mod-
237 ular units of sensory processing.

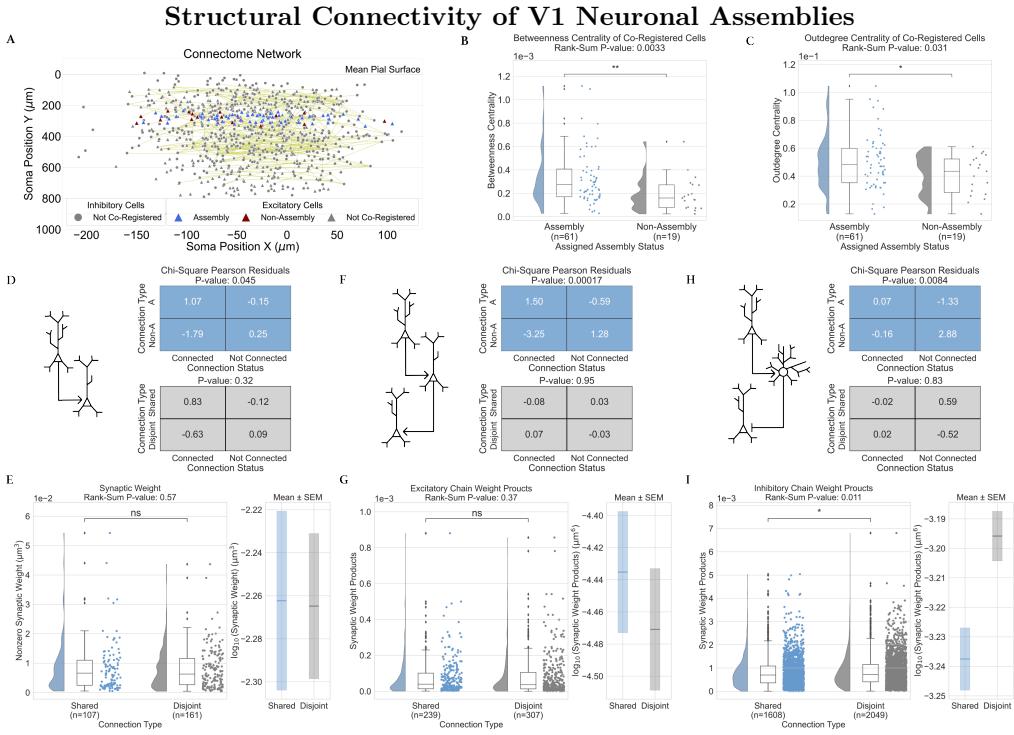


Figure 4: **(A)**. Visualization of the network being analyzed, showing the soma position of cells in the connectome, colored by cell-type and assembly assignment. All coregistered reconstructed neurons were found in layer 2/3 or layer 4 of V1. **(B)** A raincloud plot of betweenness centrality, demonstrating a higher centrality for assembly neurons than those not in assemblies(Wilcoxon Rank-Sum: p-value < 0.01). **(C)** A raincloud plot of outdegree centrality, a mathematical proxy for probability of connection, demonstrating a higher centrality for assembly neurons (Wilcoxon Rank-Sum: p-value < 0.05). **(D,F,H)**. Chi-squared analysis of the likelihood of monosynaptic (D), disynaptic excitatory (F), and disynaptic inhibitory (H) connections (schematic on the left). This comparison was made between assembly neurons and non-assembly neurons (top-right) as well as between neurons that share an assembly membership and neurons that participate in disjoint assemblies (bottom-right). **(E, G, I)** Raincloud plots showing the combined synaptic PSD volume per extant monosynaptic (E), disynaptic excitatory (G), and disynaptic inhibitory (I) connection, each divided between origin and terminus cell pairs which share assemblies and those which participate in disjoint assemblies. Log-scaled plotting of the chain weights with SEM-based confidence intervals is included as an inset plot to the right of each panel.

238 3.6 Non-Assembly Cells

239 The same monosynaptic and disynaptic analysis run on pairs of cells with shared
 240 assembly memberships and cells with disjoint memberships was also run on pairs of
 241 assembly cells (cells that both participate in at least one assembly) and pairs of non-
 242 assembly cells (cells that had no assembly memberships for either cell).

243 We show assembly cells exhibiting a pattern of connectivity distinct from non-
 244 assembly cells, a requirement of Hebb's theory. We assessed the higher-order integration
 245 of assembly neurons into the broader structural network using centrality metrics.
 246 In particular, we found significantly lower betweenness centrality (Fig. 4B), a measure

of a node's importance in mediating communication within a network, and outdegree centrality (Fig. 4C), an analog for probability of outbound connectivity, in neurons outside assemblies compared to neurons within assemblies. Furthermore, we found that the pairwise probability of monosynaptic connections, disynaptic excitatory chains, and disynaptic inhibitory chains was significantly greater for cells sharing assembly membership than for non-assembly cells (p -values : 0.045, $1.7e-4$, and 0.0084, respectively). No significant differences in PSD Volumes between assembly and non-assembly cells were observed. Importantly, these results were not due to differences in the spatial locations of assembly and non-assembly cells with respect to the centroid of the network (Supp Fig. 9). These findings confirm that cells in the connectome participating in at least one assembly are more interconnected than those not participating in any.

Overall, these results indicate reduced participation in the structural framework of the primary visual cortex for non-assembly cells and imply an organizational role of assembly neurons as hubs for information flow. In contrast, non-assembly neurons may play a more peripheral, secondary role in the network, such as noise filtering.

4 Discussion

4.1 Inhibition as the Delineating Basis of Hebbian Assemblies

This work provides the first demonstration of the structural underpinnings of Hebbian assemblies and validates Hebb's assembly theory in a surprising way; while there is strong agreement on functional predictions, our structural findings run counter to Hebb's original proposal. We do not observe a marked increase in the strength or probability of excitatory connections between cells that participate in the same assembly (Fig. 4D-G). Instead, we find significantly stronger feed-forward inhibitory connections targeting cells in other assemblies (Fig. 4I). Hebb and others [42, 9, 3] have hypothesized that targeted inhibition could play a crucial role in the formation and delineation of assemblies. Our findings offer the first clear empirical evidence for this mechanism.

Despite this novel result, our findings remain a validation of Hebb's structural postulates and are consistent with what would be required for assemblies to be delineated by inhibition. Assembly cells were substantially more integrated into the connectome, displaying significantly higher betweenness and outdegree centrality (Fig. 4B,C). This pattern demonstrates that cells participating in our activity-defined assemblies possess markedly different patterns of connectivity from cells that do not, establishing the necessary structure-function relationship postulated by Hebb.

Intuitively, the delineation of assemblies can be realized on a spectrum between internal (shared-assembly) excitation and external (disjoint-assembly) inhibition. Shared-assembly excitation would rely on broad inhibitory connectivity across the entire population of assembly neurons, and disjoint-assembly inhibition would similarly rely on broad excitatory connectivity. We find a significantly greater probability of excitatory connections between cells belonging to any assembly (Fig. 4D,F), however we did not observe significantly stronger or more frequent excitatory connections between neurons with shared assembly membership (Fig. 4D-G). Combined with our finding of significantly greater feed-forward inhibitory weights between neurons with disjoint assembly membership (Fig. 4I), the broad excitation strongly suggests that the extracted assemblies are delineated predominantly via inhibition.

4.2 Functional Consistency of Assemblies

Functionally, ensembles of cells classified as assembly cells behave in ways consistent with Hebbian assemblies. The assemblies demonstrated markedly higher reliability in

294 their responses to naturalistic visual stimuli compared to individual cellular responses
295 of all three sub-populations of cells and size-matched random ensembles, reflected by
296 significantly elevated Oracle scores (Fig. 3C). This consistent coactivity of the assemblies
297 has been hypothesized to offer resilience to representational drift, allowing them to
298 serve as a substrate for long-lasting representation [8]. Representational drift reflects a
299 functional characterization of fluctuations at the cellular level, such as synaptic turnover
300 [43, 44], even under stable stimulus conditions [45]. This phenomenon has been observed
301 in areas other than V1, such as the piriform cortex [46]. By maintaining consistent and
302 coherent patterns of coactivity, assemblies may offer a general cortical mechanism of
303 stable perceptual representations despite such turnover.

304 Beyond their reliability in responses, assemblies were also superior in our decoding
305 of natural movie presentations (Fig. 3D), underscoring their efficacy in extracting
306 higher-order visual information. In addition, the sparsity of assembly coactivity was
307 substantially greater than that of random ensembles (Fig. 3B), which, when combined
308 with a lower average correlation (3A), is consistent with assemblies employing a cost-
309 effective encoding strategy [47]. Such a strategy would enhance the capacity for distinct
310 representation of sensory inputs, translating more flexible individual responses into reli-
311 able population-level encoding. Finally, cells assigned to assemblies exhibit significantly
312 higher pairwise correlations than non-assembly cells, corroborating the experimental
313 work of Harris and Carandini [48] defining 'choristers' and 'soloists'. However, we found
314 no difference in their average response reliability to natural movies or in their overall
315 signal sparsity. This result implies that a population perspective is required for the
316 encoding of reliable perceptually relevant stimulus features.

317 4.3 Bridging Structure and Activity

318 Some examined ensemble features bridge structure and activity, a crucial aspect of
319 this study, which has not previously been achievable at scale. First, the extensive overlap
320 of neurons across multiple assemblies (Fig. 2D) is consistent with both Hebb's proposal
321 of sub-assemblies [9] and Yuste and colleagues' review of definitions put forward for
322 ensembles [49], with individual neurons contributing to multiple functional modules.
323 Second, our observation of greater chain weights in feed-forward inhibitory chains offers
324 a mechanistic explanation of how assemblies can retain distinct responses to stimuli
325 without becoming so correlated in their activity that they merge into a single functional
326 ensemble. The plausibility of this explanation is supported by the significant observed
327 negative correlation between inhibitory chain connection strength between disjoint as-
328 semblies and the r coefficient of their paired assembly coactivity traces, showing that
329 the greater the inhibitory weight of the connections observed, the less correlated the
330 activation of the two assemblies is 3.5.

331 4.4 Further Directions

332 We observe that the assemblies exhibit notable variability in size (Fig. 2D). SGC
333 minimizes overestimation of the number of assemblies or neuron assignments, suggest-
334 ing that these size differences reflect intrinsic properties rather than methodological
335 artifacts. There is therefore an opportunity for future studies to delineate the distinct
336 roles potentially served by larger assemblies, such as 'A 1', compared to their smaller
337 counterparts. In addition, while novel in its scale and multifaceted nature, we see op-
338 portunities to reduce some of the current limitations within this dataset. Continued
339 reconstruction, proofreading, and coregistration will allow analyses of more neurons and
340 subsequent new lines of inquiry. For instance, the SGC algorithm additionally flags fre-
341 quent patterns of activity, potentially analogous to Hebbian phase sequences [6, 7]. We

342 currently do not have sufficiently many coregistered and fully reconstructed cells to allow
343 examination of connections that bridge one pattern to the next. Exploring how these
344 patterns interact could reveal mechanisms of large-scale neural coordination. Other
345 limitations will be more difficult to overcome without the availability of next-generation
346 multi-modal datasets. Most importantly, the temporal resolution of the scans leaves us
347 with no ability to examine activity on a tens-of-ms timescale, which is highly relevant
348 for synaptic plasticity.

349 These results suggest several potential implications for the field at large. Hebb pro-
350 posed assemblies as a universal building block, simultaneously addressing the functional
351 and structural sides of perception, cognition, and behavior. A basic unit that bridges
352 structure and function allows one to derive structural predictions from functional char-
353 acterizations and vice versa. Beyond validating Hebb's theory, the evidence presented
354 here for an underlying inhibitory mechanism provides further support for the analysis of
355 inhibition in cognitive and sensory disorders studied to date primarily through the lens
356 of excitation [50, 51]. In the future, the integration of cell-type-specific genetic informa-
357 tion with functional assembly data could also provide deeper insights into the molecular
358 foundations of assembly formation and maintenance in health and disease [52]. Hebb
359 postulated that assemblies form the atoms of cognition, and it has not escaped our
360 notice that the cross-inhibitory mechanism we here demonstrate might be a universal
361 feature of brain-wide assembly organization.

362 **5 Methods**

363 **5.1 V1 Deep Dive Dataset**

364 **5.1.1 Stimuli**

365 Visual stimuli were presented using the same monitor configuration as in de Vries et
366 al. [53]. Imaging sessions were one hour long and offered a wide variety of visual stimuli.
367 Assembly extraction was performed on the fluorescence data from the full session. The
368 remainder of our analysis utilizes only the natural movie clips and the full-field drifting
369 gratings, details of which are provided below. The other stimuli presented during the
370 session included natural images, windowed drifting gratings, and locally sparse noise.
371 The natural movies stimulus consisted of a series of clips concatenated into 3,600 frames
372 (with 30 Hz frame rate), presented 8 times.

373 The full-field drifting gratings stimulus consisted of a drifting sinusoidal grating at a
374 1 Hz temporal frequency and 80 percent contrast, presented at 12 different orientations
375 (multiples of 30°) and at 2 spatial frequencies (0.04 and 0.08 cycles per degree). Each
376 condition was presented eight times, in randomized order, with one second of mean
377 luminance grey between presentations. The windowed drifting grating stimulus matched
378 the full-field stimulus, but the stimulus was restricted to a 30° diameter window. For
379 each column, the position of the window was determined separately to align with the
380 population receptive fields of imaged neurons. The locally sparse noise stimulus consisted
381 of white and dark spots on a mean luminance gray background. Each spot was a square
382 9° on a side. Each frame of this stimulus was presented at 3 Hz.

383 Two natural scene stimuli were presented. One consisted of 12 images (selected
384 from those used in the Brain Observatory pipeline) presented in a frozen sequence, and
385 repeated 40 times. The other consisted of the full 118 images from the Brain Observatory
386 pipeline, presented 8 times total. The images were presented in a random order, but
387 fixed with two different seeds. Each of the two seeded sequences was presented 4 times.
388 The images were presented at 3 Hz.

389 **5.1.2 2/3 Photon Microscopy and Activity Data Processing**

390 The full $800 \times 800 \times 800 \mu\text{m}^3$ volume was divided into five columns, each imaged via ei-
391 ther 2-photon (2P) or 3-photon (3P) microscopy, depending on depth. The volumes were
392 scanned over the course of several sessions. Each session consisted of the full set of visual
393 stimuli (see below), presented in the same order and with the same timing. Our anal-
394 ysis is concerned only with the central column, centered within the $800 \times 800 \times 800 \mu\text{m}^3$
395 volume, where the largest set of reconstructed, coregistered, and proofread neurons is
396 currently available. Figure 1A shows the arrangement of scan volumes (5 volumes span-
397 ning 75 μm to 620 μm depth, each $400 \times 400 \times 80 \mu\text{m}^3$) and imaging planes (6 planes
398 within each scan volume, separated by 16 μm at thirty-seven frames per second so that
399 each plane is imaged at 6 Hz.) We selected volume three of the central column for our
400 primary analyses and validated our work with volume four. Both were imaged with 2P
401 microscopy.

402 The fluorescence data was preprocessed using the standard LIMS pipeline as used
403 for the Allen Institute’s Visual Coding 2P dataset [53], including motion correction,
404 segmentation, demixing, neuropil subtraction, ROI filtering, and df/f calculation. Iden-
405 tified regions of interest were run through a classifier trained on manual labeling data
406 meant to reduce the false classification of artifact ROIs as neuronal somas, and only
407 those with a high confidence score (at least 0.5) were included in our analysis.

408 **5.1.3 Electron Microscopy and Reconstruction**

409 The mouse was transcardially perfused with a fixative mixture of paraformaldehyde
410 and glutaraldehyde. All procedures were carried out in accordance with the Institutional
411 Animal Care and Use Committee at the Allen Institute for Brain Science. The large-
412 volume staining protocol was adapted from [54]. After dissection, the neurophysiological
413 recording site was identified by mapping the brain surface vasculature. A thick (1200 μm)
414 slice was cut with a vibratome and post-fixed in perfusate solution for 12 to 48 h. The
415 tissue was then infused with heavy metals, dehydrated, and embedded in EMS Hard
416 Plus resin. After curing, the samples were epoxy cured to a stub. They were then
417 sliced and placed onto continuous tape by a ATUMtome Automated Tape Collecting
418 Ultramicrotome.

419 The continuous tape was fed into an automated high-throughput transmission elec-
420 tron microscopy pipeline[55]. Transmission electron microscopy is particularly well
421 suited for automated imaging and preserves very good x-y resolution at the expense
422 of some resolution on the vertical z-axis, and so specialized methodology was deployed
423 during reconstruction[56, 57]. Serial section alignment was performed through a contract
424 with Zetta A.I, followed by stitching [58, 55], segmentation, and automated reconstruc-
425 tion. Proofreading of a subset of cells was performed under contract by Ariadne.ai.

426 Cell-type predictions were made for single-nucleus objects within 175 microns of the
427 centerline and with a nucleus volume greater than 218, based on dendritic skeleton fea-
428 tures adapted from [59]. Segmentation and annotation were stored in a CAVE database
429 for access via CAVEClient [60].

430 **5.1.4 Coregistration**

431 Manual coregistration was performed using the Fiji plugin BigWarp [61, 62]. A
432 structural scan of the vasculature was aligned with the two-photon imaging planes (max
433 intensity projection). Next, a downsampled EM image was aligned to this composite
434 using a thin-plate-spline transform based on manual landmarks. After initial alignment,
435 the transform was used to predict additional correspondences between two-photon ROI
436 centroids and segmented EM cells. Four hundred verified correspondences between a flu-
437 orescence imaging ROI with a high classifier score and a corresponding morphologically
438 typed EM-reconstructed cell passed manual inspection by two independent reviewers
439 and were included in this study. 315 of these correspond to cells whose fluorescence was
440 recorded in the scan we analyzed (volume three of the central column). Of these 315,
441 80 had axon reconstructions that were verified accurate by trained experts ('proofread')
442 to their maximal extension within the scan volume. Coregistration and reconstruction
443 are ongoing at the time of this writing, but the extant data already allow for a relation
444 of the physiology of neural data to its exact anatomy at an unprecedented level.

445 **5.2 Graph Clustering for Assembly Extraction**

446 To extract assemblies, we use the Similarity-Graph-Clustering (SGC) algorithm that
447 has been originally proposed for the detection of neural assemblies during spontaneous
448 activity in the zebrafish optic tectum [63]. The SGC algorithm identifies neuronal assem-
449 blies using ideas from graph theory by transforming the problem into one of community
450 detection on some graph in which assemblies correspond to distinct (graph) communi-
451 ties. Unlike traditional methods that rely on pairwise correlations between cells, SGC
452 groups frames of fluorescence indicative of significant coactivity [32]. These moments of
453 higher-order correlation are referred to as potential 'phase sequences', representing when
454 populations of neurons act cohesively as a closed circuit during assembly activation [64].

455 Importantly, by design, SGC allows cells to participate in multiple assemblies. Assembly overlap has been integral to assembly studies in the past [65]. Hebb originally
456 postulated that the shift between active assembly states could be what is colloquially
457 referred to as a ‘train of thought’ [1]. Our results revealed a sizable degree of overlap
458 between assemblies, which was visualized through an UpSet plot (Fig. 2D) [33], showing
459 neurons frequently assigned to multiple assemblies (Fig. 2D).

460 Compared with several other prominent assembly extraction algorithms upon application to both simulated and biological calcium imaging datasets under different conditions, this algorithm was shown to perform best overall [32]. Although it did not yield a perfect reconstruction of the assemblies in the biological dataset, SGC was able to recover the assemblies with higher accuracy than all other algorithms. In part, this performance has been attributed to the computational effort that SGC places in estimating the number of assemblies before defining them.

461 We used a recent implementation of the SGC algorithm in Python [66]. For completeness, we briefly recall the main steps here: The algorithm begins by thresholding
462 the calcium fluorescence signals (df/f) of the ROIs (neurons) by two standard deviations
463 above the mean to minimize noise (Fig. 2B). Afterward, activity patterns are
464 selected where the coactivity level of neurons exceeds the significance threshold. This
465 threshold that determines the set of high-activity patterns is based on a null model of
466 coactivity obtained by shuffled activity signals (significance value: 0.05, rounds: 1000).
467 A k -nearest-neighbor graph is then constructed from the set of high-coactivity patterns
468 based on the similarity between the patterns in the cosine distance. For that, the number
469 of neighbors k is automatically chosen such that the resulting graph is connected. In
470 the next step, the number of communities in this graph is estimated using a statistical
471 inference procedure. For our study, we fixed the hyperparameters with five independent
472 Monte Carlo rounds of 150,000 steps each. With an estimate for the most likely number
473 of communities, spectral clustering is applied. These clusters of high-coactivity patterns
474 are the first prospective selection of coactivity patterns corresponding to the assemblies.
475 However, the final step is a combination of averaging and reassignment to reject groups
476 that may have been erroneously defined because of a high level of noise in the original
477 signal. This minimizes the likelihood of overestimating the number of assemblies or the
478 neurons that should be assigned to those assemblies. Neurons are assigned to assemblies
479 based on their affinity, the probability that they were active in any of the assemblies’
480 activity patterns (affinity: 0.4). To establish our hyperparameters, we first performed
481 a grid search on an independent functional scan volume (Scan Volume 4, acquired under
482 identical conditions). Candidate values of Monte Carlo rounds for our estimation
483 step (50000, 100000, 150000) and affinity cut-offs (0.2, 0.3, ...0.9) were evaluated by how
484 well the resulting assembly assignments could fit to a low-dimensional embedding of
485 pairwise activity correlations (see Supp. Fig. 7). To compare the regression fits, we
486 applied the Akaike Information Criterion, which is calculated from the number of independent
487 variables and the maximum likelihood estimate of the model. The parameters
488 that maximized the Akaike Information Criterion in the regression fits of scan volume 4
489 were 150000 Monte Carlo rounds and an affinity cut-off of 0.4. These parameter values
490 were then locked and applied to scan volume 3.

499 5.3 Random Ensembles

500 Random ensembles served as a null model and were defined as randomly selected
501 sets of neurons drawn with equal probability and no replacement from the population
502 of all recorded pyramidal cells within scan volume 3. Each set was size-matched to its
503 corresponding assembly, with the same number of neurons.

504 **5.4 Oracle Scores**

505 Oracle scores are a measure of the reliability of the trace response of cell activity
506 to repeated visual stimuli computed through a jackknife mean, or leave-one-out mean
507 [67, 68], of correlations between the activity trace across the repeated visual stimuli. In
508 this work, we treat the entire stereotyped stimulus sequence as a single stimulus, con-
509 catenating natural movie clips and their responses, rather than calculating the score for
510 individual clips. This should reduce the risk of sparse responses leading to an artificially
511 high reliability score.

512 Let $S_i(t)$ denote the activity trace of a neuron or an assembly during the i -th pre-
513 sentation of a stimulus at time t . Let $\bar{S}_{-i}(t)$ represent the mean activity trace of all
514 activity traces excluding the i -th presentation, out of n total presentations of the same
515 stimulus. From this, we can calculate the Oracle score O as

$$O = \frac{1}{n} \sum_{i=1}^n \frac{\text{Cov}(S_i(t), \bar{S}_{-i}(t))}{\sqrt{\text{Var}(S_i(t)) \cdot \text{Var}(\bar{S}_{-i}(t))}}.$$

516 In this formulation, the numerator will calculate the covariance between the activity
517 trace at each presentation to the mean activity of all other repeats, while the denomi-
518 nator scales the magnitude of this covariance by the product of the standard deviations
519 of $S_i(t)$ and $\bar{S}_{-i}(t)$.

520 **5.5 Trigger Frames**

521 To assess the visual stimuli associated with high activity in neuronal assemblies,
522 we computed trigger frames by identifying peak activity times and extracting the corre-
523 sponding images from a natural movie presentation. From a coactivity trace for assembly
524 k over time $S_k(t)$, we can detect peaks P_k using the `scipy` signal package to define local
525 maxima. We then define a mean trigger frame $\mu_k(x, y)$ for a pixel (x, y) as

$$\frac{1}{|P_k|} \sum_{t \in P_k} I_t(x, y)$$

526 where $I_t(x, y)$ is the natural movie frame at time t .

527 This allowed us to visualize the average triggering frame for each assembly. The
528 same process was then repeated for the coactivity traces of the size-matched random
529 ensembles, and the squared difference between the average frame for each assembly and
530 the average frame for its corresponding random ensemble. These computations allowed
531 us to determine which frame of visual stimuli and consistent features in those stimuli
532 were most strongly associated with high assembly coactivity in particular, not merely
533 broad neuronal activation.

534 **5.6 Decoder**

535 To evaluate the ability of assemblies to decode visual stimuli, we implemented a
536 Multi-Layer Perceptron Classifier (MLPClassifier) from the `scikit-learn` library. This
537 classifier was used to differentiate between 15 natural movie clips based on assembly
538 coactivity time traces. Random ensembles of neurons with the same size distribution as
539 the assemblies were used as a null model for comparison.

540 The MLP is a classical feed-forward neural network composed of an input layer, one
541 or more hidden layers, and an output layer. Each node in a layer is fully connected to
542 every node in the subsequent layer through weighted connections. The final output of
543 the function is determined by a non-linear activation function applied to the weighted
544 sum of its inputs plus a bias term.

545 The MLPClassifier from `scikit-learn` is an implementation of a Multi-Layer Per-
 546 ceptron (MLP), a type of feed-forward neural network. It operates by mapping input
 547 features \mathbf{x} to outputs $\hat{\mathbf{y}}$ through a series of hidden layers. Each layer consists of neurons
 548 that perform a weighted sum of their inputs, followed by the application of a non-linear
 549 activation function. Defining $\mathbf{W}^{(k)}$ as the weight matrix for the k -th layer, $\mathbf{h}^{(k)}$ as
 550 the input, and $\mathbf{b}^{(k)}$ as the bias vector for the k -th layer, we can define with activation
 551 function σ

$$\mathbf{h}^{(k)} = \sigma \left(\mathbf{W}^{(k)} \mathbf{h}^{(k-1)} + \mathbf{b}^{(k)} \right).$$

552 For classification, the final layer uses the softmax activation function to output prob-
 553 abilities for each class. With L denoting the number of layers, we can define these output
 554 probabilities as

$$\hat{\mathbf{y}} = \text{softmax}(\mathbf{h}^{(L)}) = \frac{\exp(\mathbf{h}^{(L)})}{\sum_j \exp(\mathbf{h}_j^{(L)})}.$$

555 The MLPClassifier is trained using backpropagation, optimizing the weights and
 556 biases via stochastic gradient descent or adaptive solvers such as Adam. Regularization
 557 can be applied through an ℓ_2 -penalty term, controlled by a hyperparameter.

558 The assembly coactivity time traces were paired with corresponding natural movie
 559 clip IDs. The data was split into training and test sets using an 80-20 split, and features
 560 were scaled to normalize the input. A cross-validated grid search was used to optimize
 561 the hyperparameters.

562 5.7 Gini coefficient

563 The Gini coefficient [34], a statistical measure that exemplifies the state of inequality
 564 within a population. While often applied in economics to evaluate income inequality,
 565 this metric has been applied as a valid approximation for signal sparsity [69, 70]. This
 566 rendition of the application has been shown to serve as a relatively simple and robust
 567 measure [71]. For our study, the coefficient is employed to quantify assembly signal
 568 heterogeneity.

569 The Gini coefficient, G is often calculated with respect to the Lorenz curve, which
 570 plots the cumulative distribution of a set (e.g., assembly coactivity trace) against its
 571 rank in ascending order. For a given assembly A with coactivity trace $S = [s_1, s_2, \dots, s_T]$
 572 where s_t is the proportion of active neurons at time point t , the coefficient for that
 573 assembly is then calculated as

$$G_A = \frac{\sum_{i=1}^{T-1} \sum_{j=i+1}^T |s_i - s_j|}{T^2 \cdot \bar{s}}.$$

574 This computation is performed independently for each assembly, providing a metric
 575 of signal inequality. A value of 0 for G_A implies all values are identical, while a value
 576 of 1 indicates perfect inequality *i.e. in a time series a single time point contains all the*
 577 *activity*. A high value of G_A indicates that coactivity is dominated by a small number
 578 of time points, reflecting the temporal sparsity.

579 5.8 Correlation

580 We examine coactivity correlations for each pair of assemblies, each pair of ran-
 581 dom ensembles, and each pair of neurons within each of our three sets of neurons (see
 582 Fig. 3A-C). To do so, we computed the Pearson's correlation coefficient r between the

583 two coactivity traces (where the coactivity trace of a single neuron is mathematically
584 equivalent to its thresholded raster activity, per [32]).

585 5.9 Motif Extraction

586 Motifs were extracted with the DotMotif Python package which detects subgraphs
587 within a graph based on the principle of subgraph monomorphism. For every graph
588 $H = (N_1, E_1)$ given by the user, the Dotmotif algorithm detects subgraphs G' within
589 the graph $G = (N_2, E_2)$ such that there exists a mapping $f : N_1 \rightarrow N'_2$ where $N'_2 \subseteq N_2$
590 and for every edge $(a, b) \in E_1$, the corresponding edge $(f(a), f(b)) \in E'_2$. The matched
591 subgraphs may also contain additional edges within the graph G . This algorithm was
592 used to detect disynaptic chains within the connectome.

593 To further refine our analysis, chain motifs were classified according to the type of
594 intermediary neuron, distinguishing between excitatory and inhibitory connections, with
595 the latter providing insight into feed-forward inhibition mechanisms.

596 5.10 Motivating Postulates

597 Our statistical analyses involve tests of the following postulates. First, excitatory
598 connections between cells that share at least one assembly ('shared' connections) will
599 be stronger than connections between cells which do not participate in any of the same
600 assemblies ('disjoint' connections), due to Hebbian plasticity [1]. This was examined
601 both with regard to the post-synaptic density volume of monosynaptic connections be-
602 tween known coregistered cells within the dataset, and separately in the form of the
603 product of connection PSD volumes in disynaptic excitatory chains which originated
604 and terminated with shared or disjoint cells, allowing us to evaluate indirect excitatory
605 connections where the middle cell had not yet been coregistered.

606 Second, that excitatory connections will be more frequent within assemblies than
607 between assemblies, due to a combination of Hebbian plasticity and pruning of synapses.
608 A number of computational studies have shown that long-term potentiation [72] and
609 pruning [73] play important roles in effective Hebbian assembly formation. This was
610 examined in the form of per-connection targeting statistics, as well as per-cell inbound
611 and outbound probability of both monosynaptic and disynaptic shared and disjoint
612 connections.

613 Third, as discussed in the introduction, Hebb himself suggested that the assem-
614 blies could have been formulated via modulation of inhibition. Additionally, a number
615 of computational models [74] have relied on inhibition between assemblies to restrict
616 simultaneous activation, enabling competition between assemblies. It was, therefore,
617 taken as an established hypothesis that one would expect inhibition between assemblies
618 to be greater than within a given assembly. As monosynaptic connections between ex-
619 citatory cells cannot be used to evaluate inhibition, this postulate was examined only in
620 disynaptic chains, where the middle cell was morphologically classified as an inhibitory
621 interneuron. Our examination involved the product of connection PSD volumes in such
622 disynaptic inhibitory chains bridging shared and disjoint assembly neurons, along with
623 a per-cell evaluation of the inbound and outbound probabilities of disynaptic inhibitory
624 chain connections.

625 Fourth, and finally, we acknowledge that Hebb discusses the reinforcement of sparse
626 connectivity in his accounts of the emergence of cell assemblies [42], and thus this re-
627 enforcement might have a significant effect on a sparse subset of connections while pro-
628 ducing a minimal difference in the central tendency of the overall set of connections.
629 Dorkenwald et al. [75] demonstrated a bimodality of the log PSD volume of excitatory-
630 excitatory synapses in the similar MICrONS mm³-dataset [26] (see also Supp. Fig.8 for

631 result in V1DD), suggesting that a subset of such connections is impacted differently
 632 by processes determining PSD volume. Combining these two, we decided to test the
 633 hypothesis that the larger of the two-component distributions found in [75] would be
 634 more likely than chance to involve connections between shared assembly neurons.

635 5.11 Statistical Methods

636 In this section, we detail the motivations and specifics of our analysis and methods.

637 We analyzed differences in connectivity metrics between shared-assembly and disjoint-
 638 assembly connection types, focusing on both the probability and strength of connections.
 639 Our analysis considered both direct monosynaptic and disynaptic connections between
 640 neurons (“by connection”), as well as the sets of inbound and outbound connections
 641 grouped by cell (“by cell”). For disynaptic chains, we grouped by whether the interme-
 642 diate (middle) cell was excitatory or inhibitory.

643 To evaluate connectivity metrics, we first defined the connection types based on
 644 assembly membership. Let A be the set of all assemblies, with A_i denoting the subset
 645 of assemblies that include cell i . Formally,

$$A_i = \{a \in A \mid i \in a\}.$$

646 Using these subsets, we defined the following binary indicators to capture the assem-
 647 bly relationship between pre-cell j and post-cell i :

- 648 • $\text{Shared}_{ij} = 1$ if $A_j \cap A_i \neq \emptyset$
- 649 • $\text{Disjoint}_{ij} = 1$ if $A_j \cap A_i = \emptyset$
- 650 • $\text{Assembly}_{ij} = 1$ if $A_j \neq \emptyset$ and $A_i \neq \emptyset$
- 651 • $\text{Non-Assembly}_{ij} = 1$ if $A_j = \emptyset$ and $A_i = \emptyset$

652 5.11.1 Monosynaptic Metrics

653 We defined w_{ij} as the pairwise summed post-synaptic density (PSD) between pre-
 654 cell j and post-cell i , and b_{ij} as an indicator variable that takes the value 1 if at least
 655 one synapse exists between j and i and 0 otherwise. All metrics exclude autapses, as
 656 these were not reliably represented in the dataset ($j \neq i$).

657 The probability of monosynaptic outbound connection for a pre-cell j was calcu-
 658 lated as the proportion of realized connections under a given connection type $C \in$
 659 $\{\text{Shared}, \text{Disjoint}, \text{Assembly}, \text{Non-Assembly}\}$, normalized by the total number of poten-
 660 tial post-cell partners for that connection type,

$$b_{\text{out}_j,C} = \frac{\sum_{i|C_{ij}=1} b_{ij}}{|\{i \mid C_{ij} = 1, i \neq j\}|}.$$

661 Similarly, the probability of a monosynaptic inbound connection for a post-cell i was
 662 defined as

$$b_{\text{in}_i,C} = \frac{\sum_{j|C_{ij}=1} b_{ij}}{|\{j \mid C_{ij} = 1, j \neq i\}|}.$$

663 For connection strength, we computed the average realized summed monosynaptic
 664 outbound PSD for a pre-cell j as the total PSD across all post-cells satisfying the
 665 connection type C , normalized by the number of realized ($b_{ij} = 1$) connection under C ,

$$w_{\text{out}_{j,C}} = \frac{\sum_{i|C_{ij}=1} w_{ij}}{\sum_{i|C_{ij}=1} b_{ij}}.$$

666 Similarly, the average realized summed monosynaptic inbound PSD for a post-cell i
 667 is given by

$$w_{\text{in}_{i,C}} = \frac{\sum_{j|C_{ij}=1} w_{ij}}{\sum_{j|C_{ij}=1} b_{ij}}.$$

668 5.11.2 Disynaptic Metrics

669 In examining the inhibition in sets of cells that share assembly membership, and
 670 in sets that do not, we were primarily interested in describing inhibition driven by
 671 the excitatory activity of an assembly's member cells. This moved us from the realm
 672 of monosynaptic connection analysis into an analysis of chains. Many aspects of our
 673 definition remained unaltered. A remained the set of all assemblies, with A_j the subset
 674 of assemblies that included pre-cell j , and A_i the subset of assemblies that included
 675 post-cell i . The binary indicators indicating the assembly relationship between cells i
 676 and j remained unaltered.

677 But rather than simply using the monosynaptic weight between neurons j and i
 678 where $j \neq i$, we defined w_{ikj} as the product of the pairwise summed post-synaptic
 679 densities w in a three-cell chain motif with j as the first cell, an interneuron k as the
 680 second cell, and i as the third cell. Thus,

$$w_{ikj} = w_{ik} w_{kj}.$$

681 And similar to our monosynaptic analysis, b_{ikj} was an indicator variable that took
 682 the value 1 if at least one disynaptic chain existed between j , k , and i and 0 otherwise.

683 Building on this definition, we outlined the method of normalization and metrics
 684 for disynaptic chain analysis, which accounted for both the intermediate and final (or
 685 first) cells in the chain. In this analysis, each pre-cell and post-cell was a coregistered
 686 excitatory neuron with an extended axon, consistent with the monosynaptic analysis.
 687 The intermediary cell in disynaptic chains, however, did not need to be coregistered or
 688 possess an extended axon.

689 Let n_e represent the number of excitatory cells and n_i represent the number of
 690 inhibitory cells from the set of all cells in the all-all connectome. Define $|k|$ be the
 691 number of potential middle partners. For inhibitory chains, $|k| = n_e$ as the middle cell
 692 is inhibitory. For excitatory chains, the middle cell cannot be the first or final cell, so
 693 $|k| = n_e - 2$.

694 The probability of disynaptic outbound connection for a pre-cell j was calculated
 695 as the proportion of realized disynaptic connections under a given connection type $C \in$
 696 {Shared, Disjoint}, normalized by the total number of potential chains satisfying the
 697 connection type,

$$b_{\text{out}_{j,C}} = \frac{\sum_k \sum_{i|C_{ij}=1} b_{ikj}}{|k| |\{i \mid C_{ij} = 1, i \neq j\}|}.$$

698 Similarly, the probability of disynaptic inbound connection for a post-cell i was
 699 defined as

$$b_{\text{in}_{i,C}} = \frac{\sum_k \sum_{j|C_{ij}=1} b_{ikj}}{|k| |\{j \mid C_{ij} = 1, j \neq i\}|}.$$

For nonzero strength of connection, we computed the average realized summed disynaptic outbound PSD for a pre-cell j as the total PSD across all chains satisfying the connection type C, normalized by the number of realized ($b_{ikj} = 1$) chains under C,

$$w_{\text{out}_{j,C}} = \frac{\sum_k \sum_{i|C_{ij}=1} w_{ikj}}{\sum_k \sum_{i|C_{ij}=1} b_{ikj}}.$$

Similarly, the average realized summed disynaptic inbound PSD for a post-cell i was given by

$$w_{\text{in}_{i,C}} = \frac{\sum_k \sum_{j|C_{ij}=1} w_{ikj}}{\sum_k \sum_{j|C_{ij}=1} b_{ikj}}.$$

To facilitate statistical testing, we defined collections of metrics based on connection type C for both monosynaptic and disynaptic analyses. For disynaptic sets, the indices ij are replaced with ijk appropriately.

5.11.3 Set Definitions

The following sets were defined to evaluate connectivity metrics:

- The set of nonzero pairwise connection strengths under a given connection type C,

$$\{w_{ij} \mid C_{ij} = 1\}.$$

- The set of outbound probabilities of connection for each pre-cell under a given connection type C,

$$\{b_{\text{out}_{j,C}}\}.$$

Similar collections were defined for inbound probability of connection, and for inbound and outbound nonzero average connection strengths by replacing the metric accordingly.

- For inbound or outbound metrics, paired sets were constructed by including only post- or pre-cells with at least one fulfilled connection under both the Shared and Disjoint connection types. Let

$$\mathcal{I}_{j,C} = \{i \mid b_{ij} = 1 \text{ and } C_{ij} = 1\}$$

be the set of post-cells connected to pre-cell j under connection type C . The set of pre-cells included in the paired comparison was then defined as:

$$\mathcal{J}_{\text{paired}} = \{j \mid \mathcal{I}_{j,\text{Shared}} \neq \emptyset \wedge \mathcal{I}_{j,\text{Disjoint}} \neq \emptyset\}$$

and the paired set of outbound metrics was:

$$\{(b_{\text{out}_{j,\text{Shared}}}, b_{\text{out}_{j,\text{Disjoint}}}) \mid j \in \mathcal{J}_{\text{paired}}\}$$

Similar paired collections were constructed for inbound probabilities and both inbound and outbound nonzero average connection strengths, as well as for metrics where pre- or post-cells appear in both the Assembly and the Non-Assembly groups.

These sets provided the basis for the statistical tests used to compare metrics across connection types.

729 **5.11.4 Statistical Tests**

730 We performed one-way statistical tests at $\alpha = 0.05$ to compare the Shared and
731 Disjoint groups, as well as to compare the Assembly and Non-Assembly groups. All
732 alternative hypotheses predict Shared > Disjoint or Assembly > Non-Assembly, except
733 for tests involving di-synaptic Inhibitory chain sets, in which the alternative hypotheses
734 predict Shared < Disjoint or Assembly < Non-Assembly. We ran the following tests:

- 735 • For unpaired sets, we use a one-sided Wilcoxon Rank-Sum test
736 • For paired sets, we use a one-sided paired Wilcoxon Signed-Rank test to compare
737 metrics within cells appearing in both groups.

738 For pairwise binary connectivity, we created a contingency table to compare the
739 frequencies of successful and failed connections across connection types:

Connection Type	Successful Connections ($b_{ij} = 1$)	Failed Connections ($b_{ij} = 0$)
Shared	$\sum_{ij Shared_{ij}=1} b_{ij}$	$\sum_{ij Shared_{ij}=1} (1 - b_{ij})$
Disjoint	$\sum_{ij Disjoint_{ij}=1} b_{ij}$	$\sum_{ij Disjoint_{ij}=1} (1 - b_{ij})$

740 Then, we performed a Chi-Squared Test of Independence at $\alpha = 0.05$ to determine
741 if pairwise connection frequency differs across connection types.

742 Finally, to examine the functional correlates of the di-synaptic inhibitory chain
743 findings, we calculated the Pearson's correlation coefficient between the summed feed-
744 forward inhibitory weights of disjoint cells connecting an assembly pair (see Methods
745 5.11.2) and the correlation scores between the assembly pair's coactivity traces (see
746 Methods 5.8), and examined significance.

747 **5.11.5 Tail Analysis**

748 In addition to the mono-synaptic and di-synaptic analyses, we performed a “tail”
749 analysis to investigate whether the proportion of Shared versus Disjoint connections
750 differs between all pairwise connections and those classified as “tail” connections.

751 To identify “tail” connections, we modeled the distribution of connection strengths
752 using a Gaussian Mixture Model (GMM) with $k = 2$ components. The model was ini-
753 tialized via k-means clustering to estimate the weights, means, and standard deviations
754 of each component. The decision boundary separating the two Gaussian components
755 was calculated as the intersection of their weighted probability density functions, derived
756 using a quadratic equation based on the GMM parameters. Connections with values
757 greater than or equal to the decision boundary were classified as “tail” connections. We
758 present the model fit and evaluation as well as the tail boundary in the supplemental
759 figure section (see Supp. Fig. 8).

760 Once the tail connections were identified, we compared the proportions of Shared
761 and Disjoint connections in this subset to their proportions in the full dataset using a
762 Chi-Squared Goodness-of-Fit Test at $\alpha = 0.05$. This test considered only the Shared
763 and Disjoint groups, with expected proportions calculated relative to the total counts
764 of these two groups in the full dataset.

765 **5.11.6 Centrality analysis**

766 Centrality analysis was used to quantify whether in a given network, assembly cells
767 were more likely to be central to the network than non-assembly cells. This analysis
768 gave further insights into the role of assembly cells in higher-order connectivity. To
769 do this, we measured different centrality metrics for assembly and non-assembly cells,

770 namely, indegree centrality, outdegree centrality, closeness centrality, and betweenness
771 centrality.

772 In a graph, the centrality of a node refers to its tendency to connect and generally
773 influence other nodes within the network [76]. We developed a directed graph $G =$
774 (N, E) , using the binary connectome such that $|N|$ represents the number of cells in the
775 connectome and E represents the binary, directed connections between all cell pairs. The
776 total number of outbound synaptic connections are given by $\sum_n \deg^+(n)$ and inbound
777 synaptic connections are given by $\sum_n \deg^-(n)$. Normalizing these connections, indegree
778 centrality was calculated as

$$I = \frac{\sum_n \deg^-(n)}{|N| - 1}$$

779 and outdegree centrality was calculated as

$$D = \frac{\sum_n \deg^+(n)}{|N| - 1}$$

780 where $n \in N$.

781 A monosynaptic connection, $e = (j, i) \in E$ between a pre-cell j and post-cell i , will
782 have i as the head and j as the tail end of the connection, where j and $i \in N$. The
783 path between a pre-cell j acting as a source neuron and a post-cell i acting as a target
784 neuron, is the alternating sequence of cells and connections starting from j and ending
785 at i , with each cell before a connection in the sequence being a pre-cell and each cell
786 after a connection being a post-cell. The number of monosynaptic connections within
787 the path indicates the length of the path. The shortest path between cells j and i is the
788 minimum length between the two cells. The shortest path, ξ_{ij} , is also referred to as the
789 geodesic path.

790 Based on this, the closeness centrality for a given cell i was calculated as the reciprocal
791 of the sum of the shortest paths, or distances between the post-cell i and all other pre-
792 cells j in the graph,

$$V(i) = \frac{|N| - 1}{\sum_{j=1}^N \xi_{ij}}.$$

793 For a given cell k , the betweenness centrality [77] was calculated as

$$B(k) = \sum_{(i,j)} \frac{\xi(k)_{ij}}{\xi_{ij}}$$

794 where $\xi(k)_{ij}$ is the number of shortest paths between pre-cell j and post-cell i that
795 pass through cell k . This value was also normalized to fall between 0 and 1.

796 6 Code Availability

797 All code and data are available at this GitHub link <https://github.com/AllenInstitute/HebbsVision>.

799 It should be noted that, for all code of our own generation, a fixed random seed
800 is used for reproducibility. Libraries used for analysis and plotting of results include:
801 numpy [78], pandas [79], scikit-learn [80], upsetplot [33], networkx [81], dotmotif [82],
802 matplotlib [83], seaborn [84], and Raincloud [85].

803 **7 Acknowledgements**

804 We wish to thank the Allen Institute for Brain Science founder, Paul G. Allen, for
805 his vision, encouragement, and support. S.B. has been supported by NIH R01EB029813.
806 S.M. has been in part supported by NSF 2223725, NIH R01EB029813, and RF1DA055669
807 grants.

808 **8 Declaration Of Interests**

809 The authors declare no competing interests.

810 9 List of Contributions

Name	Contributions
J Wagner-Carena	Conception and Design, Functional and Structural Data Analysis, Data Interpretation, Writing, Revision
S Kate	Data Interpretation, Chain and Graph Theoretic Data Analysis, Writing, Revision
T Riordan	Data Interpretation, Statistical Data Analysis, Writing, Revision
R Abbasi-Asl	Data processing, Revision, Writing
J Aman	Coregistration (Data Analysis), Writing
A Amster	Functional ROI extraction, team PIKA
AL Bodor	Data Acquisition: EM Histology/Sectioning; Revision
D Brittain	Data Acquisition: EM Imaging; Revision
JA Buchanan	Data Acquisition: EM Histology; Revision
MA Buice	Data processing: ROI extraction; Revision
DJ Bumbarger	Data Acquisition: EM Sectioning; Revision
F Collman	Interpretation, Dataset Resources and Infrastructure; Revision
NM da Costa	Data Processing; Interpretation; Writing, Revision
DJ Denman	Data Acquisition: Imaging, stimuli
SEJ deVries	Scientific leadership, optical physiology
E Joyce	Data Processing: Skeleton generation, layer definition, soma&nucleus feature extraction; Revision
D Kapner	Data Processing: EM alignment; Revision
CW King	Software development: allen_v1dd library
JD Larkin	Data Acquisition: Functional Imaging
D Liu	Software development: allen_v1dd library
G Mahalingam	Data Processing: EM Image Processing; Revision
D Millman	Stimulus Creation, Data Analysis
J Möller	SGC Algorithm, Writing, Revision
C Morrison	Functional ROI extraction, team PIKA
RC Reid	Interpretation; Dataset Resources and Infrastructure, Revision
CM Schneider-Mizell	Interpretation; Dataset Resources and Infrastructure, Revision
D Scott	Functional ROI extraction, team PIKA
S Suckow	Project Management
KT Takasaki	Data Acquisition: Functional Imaging
M Takeno	Data Acquisition: EM Histology/Sectioning; Data Processing: coRegistration; Revision
R Torres	Data Processing: EM Image Processing/Coregistration; Revision; Writing
D Vumbaco	Proofreading Project Management
J Waters	Microscopy team lead
DG Wyrick	Software development: allen_v1dd library
W Yin	Data Acquisition: EM Imaging; Revision
J Zhuang	Data processing
S Mihalas	Conception and Design, Data Interpretation, Writing, Revision
S Berteau	Conception and Design, Data Analysis, Data Interpretation, Writing, Revision

811 **References**

- 812 [1] Hebb, D. *The Organization of Behavior* (John Wiley & Sons, New York, 1949).
- 813 [2] Frégnac, Y. Aplysia: Hebbian or not? *Trends in Neurosciences* **9**, 410 (1986).
- 814 [3] Buzsáki, G. Neural Syntax: Cell Assemblies, Synapsembles, and Readers. *Neuron* **68**, 362–385 (2010).
- 815
- 816 [4] Miehl, C., Onasch, S., Festa, D. & Gjorgjieva, J. Formation and computational
817 implications of assemblies in neural circuits. *The Journal of Physiology* **601**, 3071–
818 3090 (2023).
- 819 [5] Matheus Gauy, M. *et al.* A Hippocampal Model for Behavioral Time Acquisition
820 and Fast Bidirectional Replay of Spatio-Temporal Memory Sequences. *Frontiers in*
821 *Neuroscience* **12**, 961 (2018).
- 822 [6] Carrillo-Reid, L., Miller, J.-E. K., Hamm, J. P., Jackson, J. & Yuste, R. Endogenous
823 sequential cortical activity evoked by visual stimuli. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* **35**, 8813–8828 (2015).
- 824
- 825 [7] MacLean, J. N., Watson, B. O., Aaron, G. B. & Yuste, R. Internal Dynamics
826 Determine the Cortical Response to Thalamic Stimulation. *Neuron* **48**, 811–823
827 (2005).
- 828 [8] Pérez-Ortega, J., Alejandre-García, T. & Yuste, R. Long-term stability of cortical
829 ensembles. *eLife* **10**, e64449 (2021). Publisher: eLife Sciences Publications, Ltd.
- 830 [9] Hebb, D. O. The semiautonomous process: Its nature and nurture. *American
831 Psychologist* **18**, 16–27 (1963). Place: US Publisher: American Psychological Association.
- 832
- 833 [10] Brock, L. G., Coombs, J. S. & Eccles, J. C. The recording of potentials from
834 motoneurones with an intracellular electrode. *The Journal of Physiology* **117**, 431–
835 460 (1952).
- 836 [11] Marr, D. & Brindley, G. S. Simple memory: a theory for archicortex. *Philosophical
837 Transactions of the Royal Society of London. B, Biological Sciences* **262**, 23–81
838 (1997). Publisher: Royal Society.
- 839 [12] Traub, R. D., Tu, Y. & Whittington, M. A. Cell assembly formation and structure
840 in a piriform cortex model. *Reviews in the Neurosciences* **33**, 111–132 (2022).
- 841 [13] Gastaldi, C., Schwalger, T., Falco, E. D., Quiroga, R. Q. & Gerstner, W. When
842 shared concept cells support associations: Theory of overlapping memory engrams.
843 *PLOS Computational Biology* **17**, e1009691 (2021). Publisher: Public Library of
844 Science.
- 845 [14] Holtmaat, A. & Caroni, P. Functional and structural underpinnings of neuronal
846 assembly formation in learning. *Nature Neuroscience* **19**, 1553–1562 (2016).
- 847 [15] Grinvald, A., Arieli, A., Tsodyks, M. & Kenet, T. Neuronal assemblies: Single
848 cortical neurons are obedient members of a huge orchestra. *Biopolymers* **68**, 422–
849 436 (2003).
- 850 [16] Harris, K. D. Neural signatures of cell assembly organization. *Nature Reviews
851 Neuroscience* **6**, 399–407 (2005). Number: 5 Publisher: Nature Publishing Group.

- 852 [17] Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G. & Buzsáki, G. Organization of
853 cell assemblies in the hippocampus. *Nature* **424**, 552–556 (2003).
- 854 [18] Kamiński, J. *et al.* Persistently active neurons in human medial frontal and medial
855 temporal lobe support working memory. *Nature Neuroscience* **20**, 590–601 (2017).
- 856 [19] Wilson, M. A. & McNaughton, B. L. Reactivation of hippocampal ensemble mem-
857 ories during sleep. *Science (New York, N.Y.)* **265**, 676–679 (1994).
- 858 [20] Yuste, R., Nelson, D. A., Rubin, W. W. & Katz, L. C. Neuronal domains in
859 developing neocortex: mechanisms of coactivation. *Neuron* **14**, 7–17 (1995).
- 860 [21] Grienberger, C. & Konnerth, A. Imaging Calcium in Neurons. *Neuron* **73**, 862–885
861 (2012). Publisher: Elsevier.
- 862 [22] Scheffer, L. K. *et al.* A connectome and analysis of the adult Drosophila central
863 brain. *eLife* **9**, e57443 (2020). Publisher: eLife Sciences Publications, Ltd.
- 864 [23] Dorkenwald, S. *et al.* Neuronal wiring diagram of an adult brain. *Nature* **634**,
865 124–138 (2024). Publisher: Nature Publishing Group.
- 866 [24] Witvliet, D. *et al.* Connectomes across development reveal principles of brain mat-
867 uration. *Nature* **596**, 257–261 (2021). Number: 7871 Publisher: Nature Publishing
868 Group.
- 869 [25] Gour, A. *et al.* Postnatal connectomic development of inhibition in mouse barrel
870 cortex. *Science (New York, N.Y.)* **371**, eabb4534 (2021).
- 871 [26] Bae, J. A. *et al.* Functional connectomics spanning multiple areas of mouse visual
872 cortex. *Nature* **640**, 435–447 (2025). Publisher: Nature Publishing Group.
- 873 [27] Shapson-Coe, A. *et al.* A petavoxel fragment of human cerebral cortex reconstructed
874 at nanoscale resolution. *Science* **384**, eadk4858 (2024). Publisher: American Asso-
875 ciation for the Advancement of Science.
- 876 [28] Bock, D. D. *et al.* Network anatomy and in vivo physiology of visual cortical
877 neurons. *Nature* **471**, 177–182 (2011).
- 878 [29] Lee, W.-C. A. *et al.* Anatomy and function of an excitatory network in the visual
879 cortex. *Nature* **532**, 370–374 (2016). Number: 7599 Publisher: Nature Publishing
880 Group.
- 881 [30] Massey, F. J. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the*
882 *American Statistical Association* **46**, 68–78 (1951). Publisher: [American Statistical
883 Association, Taylor & Francis, Ltd.].
- 884 [31] Tootell, R. B. H. *et al.* Functional analysis of primary visual cortex (V1) in humans.
885 *Proceedings of the National Academy of Sciences* **95**, 811–817 (1998). Publisher:
886 Proceedings of the National Academy of Sciences.
- 887 [32] Möller, J., Avitan, L. & Goodhill, G. J. Detecting neural assemblies in calcium
888 imaging data. *BMC biology* **16**, 143 (2018).
- 889 [33] Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. & Pfister, H. UpSet: Visu-
890 alization of Intersecting Sets. *IEEE Transactions on Visualization and Computer*
891 *Graphics* **20**, 1983–1992 (2014). Conference Name: IEEE Transactions on Visual-
892 ization and Computer Graphics.

- 893 [34] Dorfman, R. A Formula for the Gini Coefficient. *The Review of Economics and*
894 *Statistics* **61**, 146–149 (1979). Publisher: The MIT Press.
- 895 [35] Walker, E. Y. *et al.* Inception loops discover what excites neurons most using
896 deep predictive models. *Nature Neuroscience* **22**, 2060–2065 (2019). Number: 12
897 Publisher: Nature Publishing Group.
- 898 [36] Green, M. Contrast detection and direction discrimination of drifting gratings.
899 *Vision Research* **23**, 281–289 (1983).
- 900 [37] Abdel-aziem, A. H. & Soliman, T. H. M. A Multi-Layer Perceptron (MLP) Neu-
901 ral Networks for Stellar Classification: A Review of Methods and Results. *International*
902 *Journal of Advances in Applied Computational Intelligence* **Volume 3**, 29–37 (2023). Publisher:
903 American Scientific Publishing Group (ASPG).
- 904 [38] Bhatkar, A. P. & Kharat, G. Detection of Diabetic Retinopathy in Retinal Images
905 Using MLP Classifier. In *2015 IEEE International Symposium on Nanoelectronic*
906 *and Information Systems*, 331–335 (2015).
- 907 [39] Alon, U. Network motifs: theory and experimental approaches. *Nature Reviews*
908 *Genetics* **8**, 450–461 (2007). Number: 6 Publisher: Nature Publishing Group.
- 909 [40] Sporns, O. & Kötter, R. Motifs in Brain Networks. *PLoS Biology* **2**, e369 (2004).
- 910 [41] Hu, Y., Trousdale, J., Josić, K. & Shea-Brown, E. Motif statistics and spike corre-
911 lations in neuronal networks. *Journal of Statistical Mechanics: Theory and Experi-
912 ment* **2013**, P03012 (2013). Publisher: IOP Publishing and SISSA.
- 913 [42] Hebb, D. O. Intelligence, Brain Function and The Theory of Mind. *Brain* **82**,
914 260–275 (1959).
- 915 [43] Holtmaat, A. & Svoboda, K. Experience-dependent structural synaptic plasticity in
916 the mammalian brain. *Nature Reviews Neuroscience* **10**, 647–658 (2009). Number:
917 9 Publisher: Nature Publishing Group.
- 918 [44] Mongillo, G., Rumpel, S. & Loewenstein, Y. Intrinsic volatility of synaptic connec-
919 tions — a challenge to the synaptic trace theory of memory. *Current Opinion in*
920 *Neurobiology* **46**, 7–13 (2017).
- 921 [45] Aitken, K., Garrett, M., Olsen, S. & Mihalas, S. The geometry of representational
922 drift in natural and artificial neural networks. *PLOS Computational Biology* **18**,
923 e1010716 (2022). Publisher: Public Library of Science.
- 924 [46] Schoonover, C. E., Ohashi, S. N., Axel, R. & Fink, A. J. P. Representational drift
925 in primary olfactory cortex. *Nature* **594**, 541–546 (2021). Number: 7864 Publisher:
926 Nature Publishing Group.
- 927 [47] Olshausen, B. A. & Field, D. J. Sparse coding of sensory inputs. *Current Opinion*
928 *in Neurobiology* **14**, 481–487 (2004).
- 929 [48] Okun, M. *et al.* Diverse coupling of neurons to populations in sensory cortex.
930 *Nature* **521**, 511–515 (2015). Publisher: Nature Publishing Group.
- 931 [49] Yuste, R., Cossart, R. & Yaksi, E. Neuronal ensembles: Building blocks of neural
932 circuits. *Neuron* S0896–6273(23)00967–4 (2024).
- 933 [50] Vogels, T. P. & Abbott, L. F. Gating multiple signals through detailed balance
934 of excitation and inhibition in spiking networks. *Nature Neuroscience* **12**, 483–491
935 (2009). Publisher: Nature Publishing Group.

- 936 [51] Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C. & Gerstner, W. Inhibitory
937 Plasticity Balances Excitation and Inhibition in Sensory Pathways and Memory
938 Networks. *Science* **334**, 1569–1573 (2011). Publisher: American Association for
939 the Advancement of Science.
- 940 [52] Xu, S. *et al.* Behavioral state coding by molecularly defined paraventricular hy-
941 pothalamic cell type ensembles. *Science* **370**, eabb2494 (2020). Publisher: Ameri-
942 can Association for the Advancement of Science.
- 943 [53] de Vries, S. E. J. *et al.* A large-scale standardized physiological survey reveals
944 functional organization of the mouse visual cortex. *Nature Neuroscience* **23**, 138–
945 151 (2020). Publisher: Nature Publishing Group.
- 946 [54] Hua, Y., Laserstein, P. & Helmstaedter, M. Large-volume en-bloc staining for
947 electron microscopy-based connectomics. *Nature Communications* **6**, 7923 (2015).
948 Publisher: Nature Publishing Group.
- 949 [55] Yin, W. *et al.* A petascale automated imaging pipeline for mapping neuronal circuits
950 with high-throughput transmission electron microscopy. *Nature Communications*
951 **11**, 4949 (2020). Publisher: Nature Publishing Group.
- 952 [56] Macrina, T. *et al.* Petascale neural circuit reconstruction: automated methods
953 (2021). Pages: 2021.08.04.455162 Section: New Results.
- 954 [57] Phelps, J. S. *et al.* Reconstruction of motor control circuits in adult Drosophila
955 using automated transmission electron microscopy. *Cell* **184**, 759–774.e18 (2021).
956 Publisher: Elsevier.
- 957 [58] Mahalingam, G. *et al.* A scalable and modular automated pipeline for stitching
958 of large electron microscopy datasets. *eLife* **11**, e76534 (2022). Publisher: eLife
959 Sciences Publications, Ltd.
- 960 [59] Schneider-Mizell, C. M. *et al.* Inhibitory specificity from a connectomic census of
961 mouse visual cortex. *Nature* **640**, 448–458 (2025). Publisher: Nature Publishing
962 Group.
- 963 [60] Dorkenwald, S. *et al.* CAVE: Connectome Annotation Versioning Engine (2023).
964 Pages: 2023.07.26.550598 Section: New Results.
- 965 [61] Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis.
966 *Nature Methods* **9**, 676–682 (2012). Publisher: Nature Publishing Group.
- 967 [62] Bogovic, J. A., Hanslovsky, P., Wong, A. & Saalfeld, S. Robust registration of
968 calcium images by learned contrast synthesis. In *2016 IEEE 13th International*
969 *Symposium on Biomedical Imaging (ISBI)*, 1123–1126 (2016). ISSN: 1945-8452.
- 970 [63] Avitan, L. *et al.* Spontaneous Activity in the Zebrafish Tectum Reorganizes over
971 Development and Is Influenced by Visual Experience. *Current Biology* **27**, 2407–
972 2419.e4 (2017). Publisher: Elsevier.
- 973 [64] Almeida-Filho, D. G. *et al.* An investigation of Hebbian phase sequences as assembly
974 graphs. *Frontiers in Neural Circuits* **8**, 34 (2014).
- 975 [65] Kursin, A., Húsek, D. & Neruda, R. Faster Learning with Overlapping Neural
976 Assemblies. In Kollias, S. D., Stafylopatis, A., Duch, W. & Oja, E. (eds.) *Artificial*
977 *Neural Networks – ICANN 2006*, Lecture Notes in Computer Science, 226–233
978 (Springer, Berlin, Heidelberg, 2006).

- 979 [66] Möller, J. & Goodhill, G. J. Detecting Neural Assemblies Through Similarity
 980 Graph Clustering. In Carrillo-Reid, L. (ed.) *Identification, Characterization, and*
 981 *Manipulation of Neuronal Ensembles*, 167–176 (Springer US, New York, NY, 2025).
- 982 [67] Miller, R. G. A Trustworthy Jackknife. *The Annals of Mathematical Statistics* **35**,
 983 1594–1605 (1964). Publisher: Institute of Mathematical Statistics.
- 984 [68] Richter, C. G., Thompson, W. H., Bosman, C. A. & Fries, P. A jackknife approach
 985 to quantifying single-trial correlation between covariance-based metrics undefined
 986 on a single-trial basis. *NeuroImage* **114**, 57–70 (2015).
- 987 [69] Hurley, N. & Rickard, S. Comparing Measures of Sparsity. *IEEE Transactions on*
 988 *Information Theory* **55**, 4723–4741 (2009). Conference Name: IEEE Transactions
 989 on Information Theory.
- 990 [70] Khosravy, M., Nitta, N., Gupta, N., Patel, N. & Babaguchi, N. Chapter 3 - A
 991 descriptive review to sparsity measures. In Khosravy, M., Dey, N. & Duque, C. A.
 992 (eds.) *Compressive Sensing in Healthcare*, Advances in ubiquitous sensing applica-
 993 tions for healthcare, 43–63 (Academic Press, 2020).
- 994 [71] Wright Muelas, M., Mughal, F., O'Hagan, S., Day, P. J. & Kell, D. B. The role
 995 and robustness of the Gini coefficient as an unbiased tool for the selection of Gini
 996 genes for normalising expression profiling data. *Scientific Reports* **9**, 17960 (2019).
 997 Number: 1 Publisher: Nature Publishing Group.
- 998 [72] Hetherington, P. A., & Shapiro, M. L. Simulating Hebb cell assemblies: the neces-
 999 sity for partitioned dendritic trees and a post-not-pre LTD rule. *Network: Compu-*
 1000 *tation in Neural Systems* **4**, 135–153 (1993). Publisher: Taylor & Francis.
- 1001 [73] Garagnani, M., Wennekers, T. & Pulvermüller, F. Recruitment and Consolidation
 1002 of Cell Assemblies for Words by Way of Hebbian Learning and Competition in a
 1003 Multi-Layer Neural Network. *Cognitive Computation* **1**, 160–176 (2009).
- 1004 [74] Fransen, E., Lansner, A. & Liljenström, H. A Model of Cortical Associative Memory
 1005 Based on Hebbian Cell Assemblies. In Eeckman, F. H. & Bower, J. M. (eds.)
 1006 *Computation and Neural Systems*, 431–435 (Springer US, Boston, MA, 1993).
- 1007 [75] Dorkenwald, S. *et al.* Binary and analog variation of synapses between cortical
 1008 pyramidal neurons. *eLife* **11**, e76120 (2022). Publisher: eLife Sciences Publications,
 1009 Ltd.
- 1010 [76] Bavelas, A. A Mathematical Model for Group Structures. *Human Organization* **7**,
 1011 16–30 (2009).
- 1012 [77] Freeman, L. C. A Set of Measures of Centrality Based on Betweenness. *Sociometry*
 1013 **40**, 35–41 (1977). Publisher: [American Sociological Association, Sage Publications,
 1014 Inc.].
- 1015 [78] Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).
 1016 Publisher: Nature Publishing Group.
- 1017 [79] McKinney, W. Data Structures for Statistical Computing in Python. *scipy* (2010).
- 1018 [80] Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine*
 1019 *Learning Research* **12**, 2825–2830 (2011).
- 1020 [81] Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring Network Structure, Dy-
 1021 namics, and Function using NetworkX. *scipy* (2008).

- 1022 [82] Matelsky, J. K. *et al.* DotMotif: an open-source tool for connectome subgraph
1023 isomorphism search and graph queries. *Scientific Reports* **11**, 13045 (2021).
- 1024 [83] Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science &*
1025 *Engineering* **9**, 90–95 (2007). Conference Name: Computing in Science & Engi-
1026 neering.
- 1027 [84] Waskom, M. L. seaborn: statistical data visualization. *Journal of Open Source*
1028 *Software* **6**, 3021 (2021).
- 1029 [85] Micah, A. *et al.* Raincloud plots: a multi-platform tool for robust data visualization.
1030 *Wellcome Open Research* (2021).

1031 **10 Statistical Tests**

Figure	Variable	Units	N	Test	Test Statistic	P-Value
SPATIAL ANALYSIS						
Fig. 2E	Spatial Distribution of A1: XY Plane	N/A	A1: 1016, All Cells: 2708	Kolmogorov-Smirnov	0.1272	2.201e-07, BH corrected: 3.301e-06 ****
Fig. 2E	Spatial Distributions of A2: XY Plane	N/A	A2: 542, All Cells: 2708	Kolmogorov-Smirnov	0.1204	0.0002708, BH corrected: 0.001354 **
Fig. 2E	Spatial Distributions of A3: XY Plane	N/A	A3: 328, All Cells: 2708	Kolmogorov-Smirnov	0.09705	0.04408, BH corrected: 0.08816
Fig. 2E	Spatial Distributions of A4: XY Plane	N/A	A4: 325, All Cells: 2708	Kolmogorov-Smirnov	0.2657	1.092e-12, BH corrected: 3.277e-11 ****
Fig. 2E	Spatial Distributions of A5: XY Plane	N/A	A5: 230, All Cells: 2708	Kolmogorov-Smirnov	0.1656	0.0005795, BH corrected: 0.002483 **
Fig. 2E	Spatial Distributions of A6: XY Plane	N/A	A6: 188, All Cells: 2708	Kolmogorov-Smirnov	0.1045	0.1342, BH corrected: 0.1611
Fig. 2E	Spatial Distributions of A7: XY Plane	N/A	A7: 187, All Cells: 2708	Kolmogorov-Smirnov	0.1422	0.01374, BH corrected: 0.03748 *
Fig. 2E	Spatial Distributions of A8: XY Plane	N/A	A8: 181, All Cells: 2708	Kolmogorov-Smirnov	0.2217	1.588e-05, BH corrected: 9.526e-05 ****
Fig. 2E	Spatial Distributions of A9: XY Plane	N/A	A9: 121, All Cells: 2708	Kolmogorov-Smirnov	0.1813	0.008358, BH corrected: 0.02786 *
Fig. 2E	Spatial Distributions of A10: XY Plane	N/A	A10: 114, All Cells: 2708	Kolmogorov-Smirnov	0.2809	7.871e-06, BH corrected: 5.903e-05 ****
Fig. 2E	Spatial Distributions of A11: XY Plane	N/A	A11: 112, All Cells: 2708	Kolmogorov-Smirnov	0.1623	0.03342, BH corrected: 0.0791
Fig. 2E	Spatial Distributions of A12: XY Plane	N/A	A12: 61, All Cells: 2708	Kolmogorov-Smirnov	0.1748	0.1285, BH corrected: 0.1611
Fig. 2E	Spatial Distributions of A13: XY Plane	N/A	A13: 48, All Cells: 2708	Kolmogorov-Smirnov	0.3107	0.001807, BH corrected: 0.006776 **
Fig. 2E	Spatial Distributions of A14: XY Plane	N/A	A14: 35, All Cells: 2708	Kolmogorov-Smirnov	0.4971	2.426e-06, BH corrected: 2.426e-05 ****
Fig. 2E	Spatial Distributions of A15: XY Plane	N/A	A15: 23, All Cells: 2708	Kolmogorov-Smirnov	0.3682	0.0111, BH corrected: 0.03329 *
Fig. 2E	Spatial Distribution of A1: Z Axis	N/A	A1: 1016, All Cells: 2708	Kolmogorov-Smirnov	4.62e-10	0.1221, BH corrected: 0.1611
Fig. 2E	Spatial Distribution of A2: Z Axis	N/A	A2: 542, All Cells: 2708	Kolmogorov-Smirnov	6.00e-05	0.1069, BH corrected: 0.1587
Fig. 2E	Spatial Distribution of A3: Z Axis	N/A	A3: 328, All Cells: 2708	Kolmogorov-Smirnov	0.6303	0.04313, BH corrected: 0.08816
Fig. 2E	Spatial Distribution of A4: Z Axis	N/A	A4: 325, All Cells: 2708	Kolmogorov-Smirnov	2.10e-05	0.1397, BH corrected: 0.1612
Fig. 2E	Spatial Distribution of A5: Z Axis	N/A	A5: 230, All Cells: 2708	Kolmogorov-Smirnov	0.9566	0.03428, BH corrected: 0.0791
Fig. 2E	Spatial Distribution of A6: Z Axis	N/A	A6: 188, All Cells: 2708	Kolmogorov-Smirnov	0.4015	0.06642, BH corrected: 0.1245
Fig. 2E	Spatial Distribution of A7: Z Axis	N/A	A7: 187, All Cells: 2708	Kolmogorov-Smirnov	0.02452	0.1111, BH corrected: 0.1587
Fig. 2E	Spatial Distribution of A8: Z Axis	N/A	A8: 181, All Cells: 2708	Kolmogorov-Smirnov	2.65e-06	0.198, BH corrected: 0.22
Fig. 2E	Spatial Distribution of A9: Z Axis	N/A	A9: 121, All Cells: 2708	Kolmogorov-Smirnov	0.2182	0.09628, BH corrected: 0.1587

Continued on next page

Continued from previous page

Figure	Variable	Units	N	Test	Test Statistic	P-Value
Fig. 2E	Spatial Distribution of A10: Z Axis	N/A	A10: 114, All Cells: 2708	Kolmogorov-Smirnov	6.38E-05	0.2151, BH corrected: 0.2305
Fig. 2E	Spatial Distribution of A11: Z Axis	N/A	A11: 112, All Cells: 2708	Kolmogorov-Smirnov	0.1665	0.1059, BH corrected: 0.1587
Fig. 2E	Spatial Distribution of A12: Z Axis	N/A	A12: 61, All Cells: 2708	Kolmogorov-Smirnov	0.4322	0.1102, BH corrected: 0.1587
Fig. 2E	Spatial Distribution of A13: Z Axis	N/A	A13: 48, All Cells: 2708	Kolmogorov-Smirnov	0.3679	0.1303, BH corrected: 0.1611
Fig. 2E	Spatial Distribution of A14: Z Axis	N/A	A14: 35, All Cells: 2708	Kolmogorov-Smirnov	0.03531	0.2361, BH corrected: 0.2443
Fig. 2E	Spatial Distribution of A15: Z Axis	N/A	A15: 23, All Cells: 2708	Kolmogorov-Smirnov	0.01637	0.3151, BH corrected: 0.3151
FUNCTIONAL ANALYSIS						
Fig. 3A	Correlations of Activity: Assemblies vs Random Ensembles	N/A	Assemblies: 210, Random Ensembles: 210	Wilcoxon Rank-Sum	-13.2	5.534e-40, BH Corrected: 5.534e-40 ****
Fig. 3A	Correlations of Activity: Assemblies vs All Cells	N/A	Assemblies: 210, All Cells: 7330556	Wilcoxon Rank-Sum	24.7	1.457e-134, BH Corrected: 1.457e-134 ****
Fig. 3A	Correlations of Activity: Assemblies vs Assembly Cells	N/A	Assemblies: 210, Assembly Cells: 3839640	Wilcoxon Rank-Sum	24.5	4.941e-133, BH corrected: 5.929e-133 ****
Fig. 3A	Correlations of Activity: Assemblies vs Non-Assembly Cells	N/A	Assemblies: 210, Non-Assembly Cells: 558756	Wilcoxon Rank-Sum	24.9	7.997e-137, BH corrected: 1.599e-136 ****
Fig. 3A	Correlations of Activity: All Cells vs Assembly Cells	N/A	All Cells: 7330556, Assembly Cells: 3839640	Wilcoxon Rank-Sum	-396	0, BH corrected: 0 ****
Fig. 3A	Correlations of Activity: Assembly Cells vs Non-Assembly Cells	N/A	Assembly Cells: 3839640, Non-Assembly Cells: 558756	Wilcoxon Rank-Sum	515	0, BH corrected: 0 ****
Fig. 3B	Sparcity of Activity: Assemblies vs Random Ensembles	N/A	Assemblies: 15, Random Ensembles: 15	Wilcoxon Rank-Sum	-3.837	6.234e-05 ****
Fig. 3C	Natural Movie Oracle Scores: Assemblies vs Random Ensembles	N/A	Assemblies: 135, Random Ensembles: 135	Wilcoxon Rank-Sum	2.607	0.009134, BH corrected: 0.0137 *
Fig. 3C	Natural Movie Oracle Scores: Assemblies vs All Cells	N/A	Assemblies: 135, All Cells: 21183	Wilcoxon Rank-Sum	10.1	5.569e-24, BH corrected: 1.831e-23 ****
Fig. 3C	Natural Movie Oracle Scores: Assemblies vs Assembly Cells	N/A	Assemblies: 135, Assembly Cells: 15561	Wilcoxon Rank-Sum	10.07	7.264e-24, BH corrected: 1.831e-23 ****
Fig. 3C	Natural Movie Oracle Scores: Assemblies vs Non-Assembly Cells	N/A	Assemblies: 135, Non-Assembly Cells: 5622	Wilcoxon Rank-Sum	10.05	9.156e-24, BH corrected: 1.831e-23 ****
None	Natural Movie Oracle Scores: All Cells vs. Assembly Cells	N/A	All Cells: 21183, Assembly Cells: 15718	Wilcoxon Rank-Sum	-0.4177	0.6762, BH corrected: 0.6762
None	Natural Movie Oracle Scores: Assembly Cells vs. Non-Assembly Cells	N/A	Assembly Cells: 15718, Non-Assembly Cells: 5465	Wilcoxon Rank-Sum	1.085	0.2778, BH corrected: 0.3333
Fig. 3D/E	Classification Heatmap: Assemblies vs Random Ensembles	%	Assemblies: 225, Random Ensembles: 225	Mann-Whitney U-Test	7546	6.070e-05 ****
Fig. 3D/E	Classification Heatmap: Assemblies vs Random Ensembles	%	Assemblies: 225, Random Ensembles: 225	Wilcoxon Rank-Sum	-3.993	6.515e-05 ****
Fig. 3D/E	Classification Heatmap Diagonal: Assemblies vs Random Ensembles	%	Assemblies: 15, Random Ensembles: 15	Mann-Whitney U-Test	131.5	3.247e-04 ***
Fig. 3D/E	Classification Heatmap Diagonal: Assemblies vs Random Ensembles	%	Assemblies: 15, Random Ensembles: 15	Wilcoxon Rank-Sum	3.435	2.960e-04 ***

Continued on next page

Continued from previous page

Figure	Variable	Units	N	Test	Test Statistic	P-Value
None	Correlations of Activity vs Inhibitory Chain Post-Synaptic Density Product	N/A	Assemblies: 210, Random Ensembles: 210	Pearson's R	-0.19	0.0048 **
STRUCTURAL ANALYSIS: SHARED Vs. DISJOINT ASSEMBLIES						
Fig. 4D	Monosynaptic Probability of Connection	N/A	Shared: 4945, Disjoint: 8475	Chi-Squared Test of Independence	0.9821	0.3217
Fig. 4E	Monosynaptic Nonzero Summed Post-Synaptic Density	μm^3	Shared: 107, Disjoint: 161	Wilcoxon Rank-Sum	-0.1762	0.5699
Fig. 4F	Probability of Excitatory Chain Connection	N/A	Shared: 1610, Disjoint: 2050	Chi-Squared Test of Independence	0.004	0.9493
Fig. 4H	Probability of Inhibitory Chain Connection	N/A	Shared: 1610, Disjoint: 2050	Chi-Squared Test of Independence	0.044	0.8338
Fig. 4G	Excitatory Chain Post-Synaptic Density Product	μm^6	Shared: 239, Disjoint: 307	Wilcoxon Rank-Sum	0.3382	0.3676
Fig. 4I	Inhibitory Chain Post-Synaptic Density Product	μm^6	Shared: 1608, Disjoint: 2049	Wilcoxon Rank-Sum	-2.302	0.01066 *
None	Probability of Outbound Monosynaptic Connection	N/A	Shared: 61, Disjoint: 61	Wilcoxon Rank-Sum; Wilcoxon Signed-Rank	-0.7041; 916.5	0.7593; 0.4060
None	Average Monosynaptic Nonzero Outbound Summed Post-Synaptic Density	μm^3	Shared: 39, Disjoint: 53	Wilcoxon Rank-Sum; Wilcoxon Signed-Rank	-0.2647; 266.0	0.600; 0.6043
None	Probability of Inbound Monosynaptic Connection	N/A	Shared: 220, Disjoint: 221	Wilcoxon Rank-Sum; Wilcoxon Signed-Rank	-1.755; 4262	0.9603; 0.7186
None	Average Monosynaptic Nonzero Inbound Summed Post-Synaptic Density	μm^3	Shared: 71, Disjoint: 103	Wilcoxon Rank-Sum; Wilcoxon Signed-Rank	-0.1347; 431.0	0.5536; 0.2880
None	Probability of Outbound Excitatory Chain Connection	N/A	Shared: 60, Disjoint: 61	Wilcoxon Rank-Sum; Wilcoxon Signed-Rank	-0.3655; 829.0	0.6426; 0.5813
None	Probability of Outbound Inhibitory Chain Connection	N/A	Shared: 60, Disjoint: 61	Wilcoxon Rank-Sum; Wilcoxon Signed-Rank	0.4251; 1336	0.6646; 0.9990
None	Probability of Inbound Excitatory Chain Connection	N/A	Shared: 60, Disjoint: 61	Wilcoxon Rank-Sum; Wilcoxon Signed-Rank	-0.3396; 607.0	0.6329; 0.8783
None	Probability of Inbound Inhibitory Chain Connection	N/A	Shared: 60, Disjoint: 61	Wilcoxon Rank-Sum; Wilcoxon Signed-Rank	-0.4977; 619.0	0.3094; 0.01466
None	Average Outbound Post-Synaptic Density Excitatory Chain Product	μm^6	Shared: 51, Disjoint: 56	Wilcoxon Rank-Sum; Wilcoxon Signed-Rank	0.0499; 596.0	0.4801; 0.4696
None	Average Outbound Post-Synaptic Density Inhibitory Chain Product	μm^6	Shared: 60, Disjoint: 61	Wilcoxon Rank-Sum; Wilcoxon Signed-Rank	-1.337; 493.0	0.09053; 9.463e-04 ***
None	Average Inbound Post-Synaptic Density Excitatory Chain Product	μm^6	Shared: 46, Disjoint: 56	Wilcoxon Rank-Sum; Wilcoxon Signed-Rank	0.7263; 502.0	0.2338; 0.5710
None	Average Inbound Post-Synaptic Density Inhibitory Chain Product	μm^6	Shared: 60, Disjoint: 61	Wilcoxon Rank-Sum; Wilcoxon Signed-Rank	-0.6376; 606.0	0.2619; 0.01146
Supp. Fig. 4	Connection Type Frequency (Shared/Disjoint in tail vs. expected)	Count	Shared: 60, Disjoint: 93	Chi-Squared Goodness of Fit	0.0321	0.8577
STRUCTURAL ANALYSIS: ASSEMBLY Vs. NON-ASSEMBLY						
Fig. 4D	Monosynaptic Probability of Connection	N/A	Assembly: 4945, Non-Assembly: 1767	Chi-Squared Test of Independence	4	0.045 *

Continued on next page

Continued from previous page

Figure	Variable	Units	N	Test	Test Statistic	P-Value
Fig. 4E	Monosynaptic Nonzero Summed Post-Synaptic Density	μm^3	Assembly: 107, Non-Assembly: 24	Wilcoxon Rank-Sum	-0.8211	0.7942
Fig. 4F	Probability of Excitatory Chain Connection	N/A	Assembly: 1610, Non-Assembly: 342	Chi-Squared Test of Independence	14.16	1.700e-04 ***
Fig. 4H	Probability of Inhibitory Chain Connection	N/A	Assembly: 1610, Non-Assembly: 342	Chi-Squared Test of Independence	6.94	0.0084 **
Fig. 4G	Excitatory Chain Post-Synaptic Density Product	μm^6	Assembly: 239, Non-Assembly: 24	Wilcoxon Rank-Sum	1.422	0.07757
Fig. 4I	Inhibitory Chain Post-Synaptic Density Product	μm^6	Assembly: 1608, Non-Assembly: 338	Wilcoxon Rank-Sum	7.741	1
None	Probability of Outbound Monosynaptic Connection	N/A	Assembly: 61, Non-Assembly: 19	Wilcoxon Rank-Sum	-0.08479	0.5338
None	Average Monosynaptic Nonzero Outbound Summed Post-Synaptic Density	μm^3	Assembly: 39, Non-Assembly: 15	Wilcoxon Rank-Sum	-0.4345	0.668
None	Probability of Inbound Monosynaptic Connection	N/A	Assembly: 220, Non-Assembly: 94	Wilcoxon Rank-Sum	1.416	0.07844
None	Average Monosynaptic Nonzero Inbound Summed Post-Synaptic Density	μm^3	Assembly: 71, Non-Assembly: 20	Wilcoxon Rank-Sum	-0.8147	0.7924
None	Probability of Outbound Excitatory Chain Connection	N/A	Assembly: 60, Non-Assembly: 19	Wilcoxon Rank-Sum	2.61	0.004532 **
None	Probability of Outbound Inhibitory Chain Connection	N/A	Assembly: 60, Non-Assembly: 19	Wilcoxon Rank-Sum	3.195	0.9993
None	Probability of Inbound Excitatory Chain Connection	N/A	Assembly: 60, Non-Assembly: 19	Wilcoxon Rank-Sum	2.667	0.003827 **
None	Probability of Inbound Inhibitory Chain Connection	N/A	Assembly: 60, Non-Assembly: 19	Wilcoxon Rank-Sum	3.08	0.999
None	Average Outbound Post-Synaptic Density Excitatory Chain Product	μm^6	Assembly: 51, Non-Assembly: 13	Wilcoxon Rank-Sum	1.577	0.05741
None	Average Outbound Post-Synaptic Density Inhibitory Chain Product	μm^6	Assembly: 60, Non-Assembly: 19	Wilcoxon Rank-Sum	0.6538	0.7434
None	Average Inbound Post-Synaptic Density Excitatory Chain Product	μm^6	Assembly: 46, Non-Assembly: 10	Wilcoxon Rank-Sum	1.262	0.1034
None	Average Inbound Post-Synaptic Density Inhibitory Chain Product	μm^6	Assembly: 60, Non-Assembly: 19	Wilcoxon Rank-Sum	1.422	0.9225
Fig. 4C	Outdegree Centrality of coregistered cells in entire connectome network	N/A	Assembly: 61, Non-Assembly: 19	Wilcoxon Rank-Sum	1.86	0.03146 *
None	Indegree Centrality of coregistered cells in entire connectome network	N/A	Assembly: 61, Non-Assembly: 19	Wilcoxon Rank-Sum	1.747	0.04034 *
Fig. 4B	Betweenness Centrality of coregistered cells in entire connectome network	N/A	Assembly: 61, Non-Assembly: 19	Wilcoxon Rank-Sum	2.719	0.003273 **
None	Closeness Centrality of coregistered cells in entire connectome network	N/A	Assembly: 61, Non-Assembly: 19	Wilcoxon Rank-Sum	1.413	0.07879
None	Outdegree Centrality of coregistered cells with just pyramidal cells	N/A	Assembly: 61, Non-Assembly: 19	Wilcoxon Rank-Sum	1.899	0.02876 *

Continued on next page

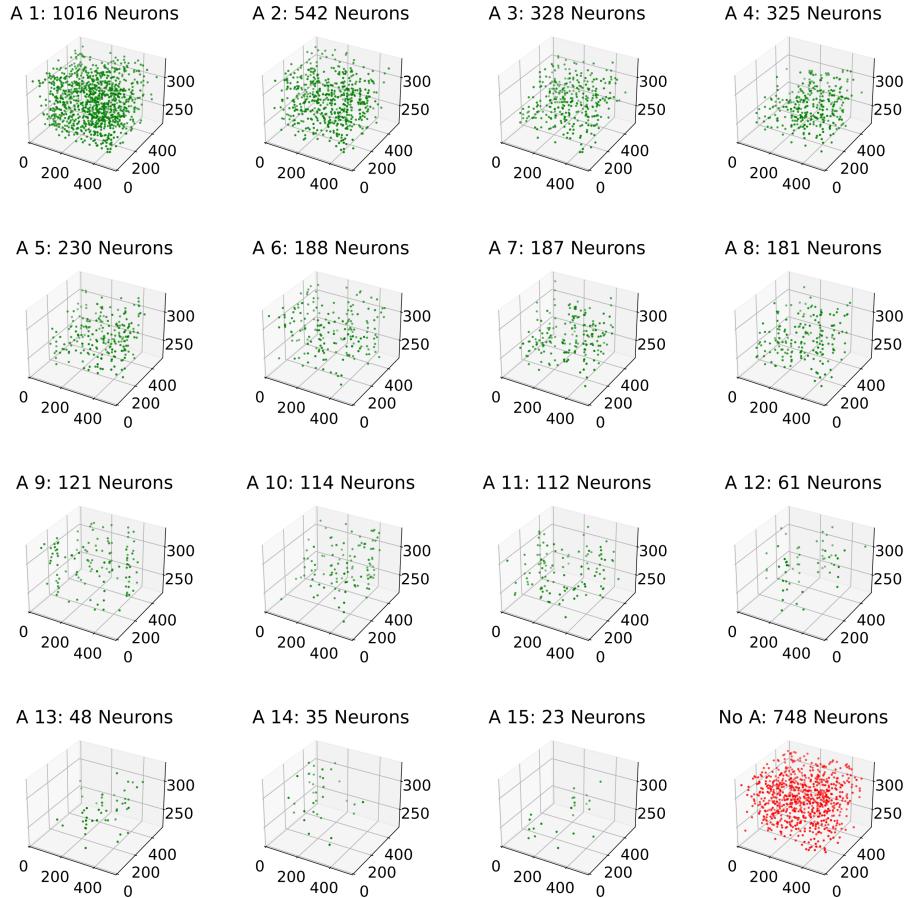
Continued from previous page

Figure	Variable	Units	N	Test	Test Statistic	P-Value
None	Indegree Centrality of coregistered cells with just pyramidal cells	N/A	Assembly: 61, Non-Assembly: 19	Wilcoxon Rank-Sum	1.538	0.06207
None	Betweenness Centrality of coregistered cells with just pyramidal cells	N/A	Assembly: 61, Non-Assembly: 19	Wilcoxon Rank-Sum	1.894	0.02913 *
None	Closeness Centrality of coregistered cells with just pyramidal cells	N/A	Assembly: 61, Non-Assembly: 19	Wilcoxon Rank-Sum	0.8423	0.1998
Supp. Fig. 5	Distance from Centroid of Connectome Network	μm	Assembly: 61, Non-Assembly: 19	Wilcoxon Rank-Sum	-0.7632	0.4454

1032

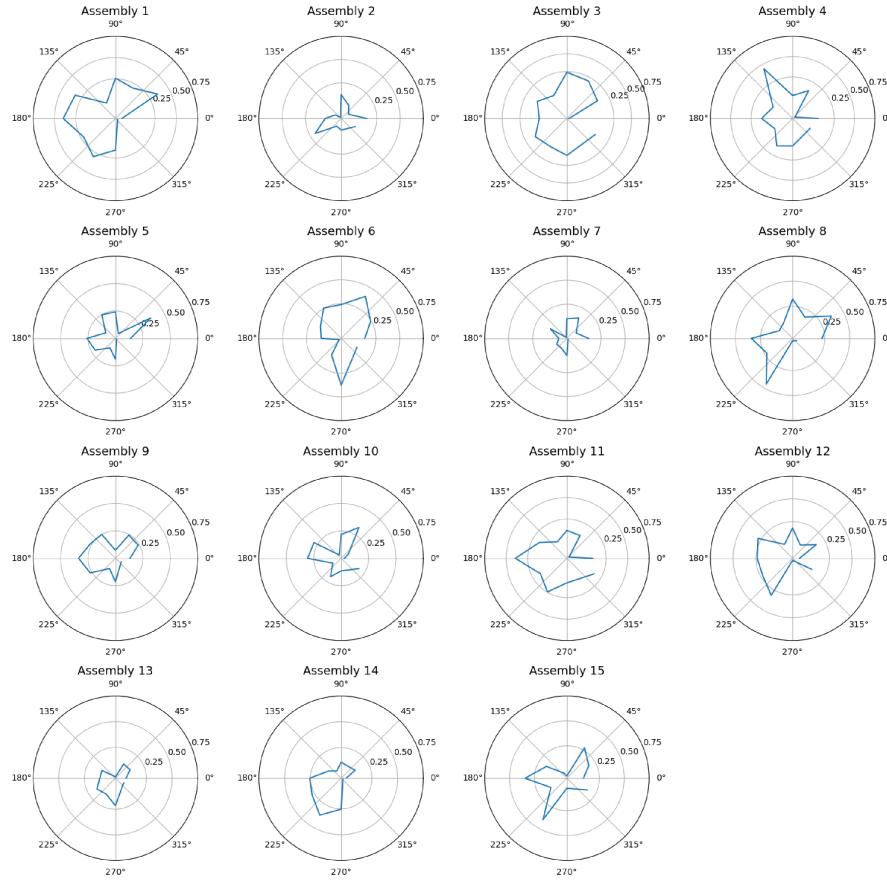
1033 11 Extended Data

Individual Assemblies Plotted in Recording Space



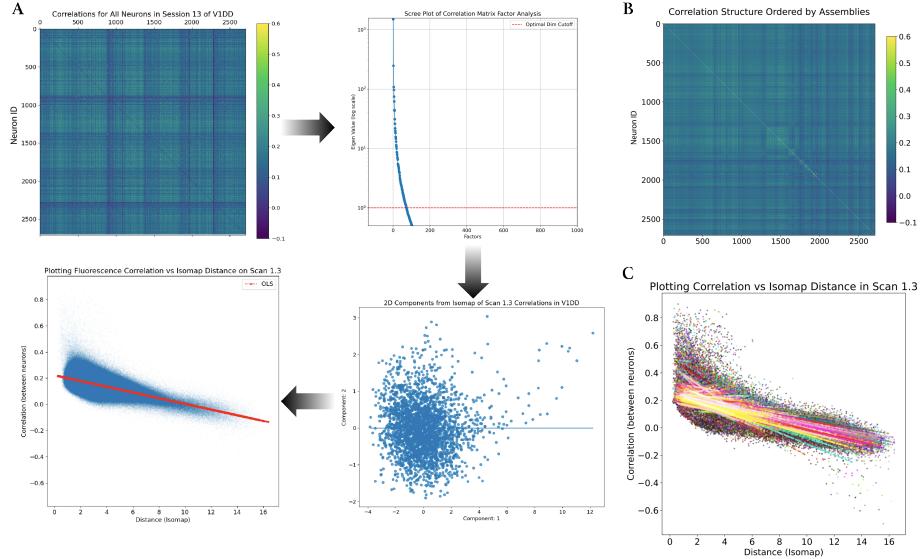
Supplementary Figure 1. Spatial positions of individual extracted assemblies in the three-dimensional recording field. Every subplot provides an isolated view of an assembly, ordered by size, visualized in the optical imaging recording space. Each individual point refers to an identified excitatory neuron. The plot of neurons assigned to no assemblies, ‘No A’, is also shown in red (bottom-right). Sub-plot axes refer to the three spatial dimensions of the recording field, with units in micrometers. Sub-plot titles also include the size of each assembly. In total, 1960 neurons were assigned to assemblies, and 748 were not. Assemblies are typically spatially spread out throughout the recording field.

Assembly Oracle Scores as a function of Fullscreen Grating Orientations



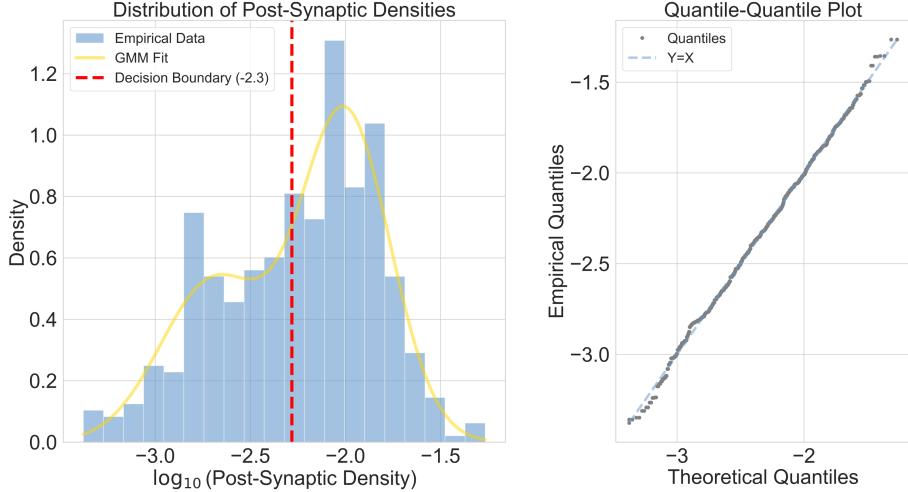
Supplementary Figure 2. Individual Assemblies Oracle Scores to Fullscreen Gratings. Every sub-plot provides an isolated visualization of the reliability in an assembly's response with respect to the orientation of fullscreen gratings. Orientation of gratings is represented by a polar plot. Reliability is measured through the Oracle score metric (see Methods 5.3). The scores of assembly coactivity trace in response to gratings are typically lower than those seen in natural movies (Fig. 3C), but the results are still indicative of tuning properties in these functional populations. Notably, some assembly traces seem to be highly reliable to particular orientations, similar to the orientation receptive fields of simple cells in the primary visual cortex.

Low-Dimensional Description of Pairwise Activity Correlations



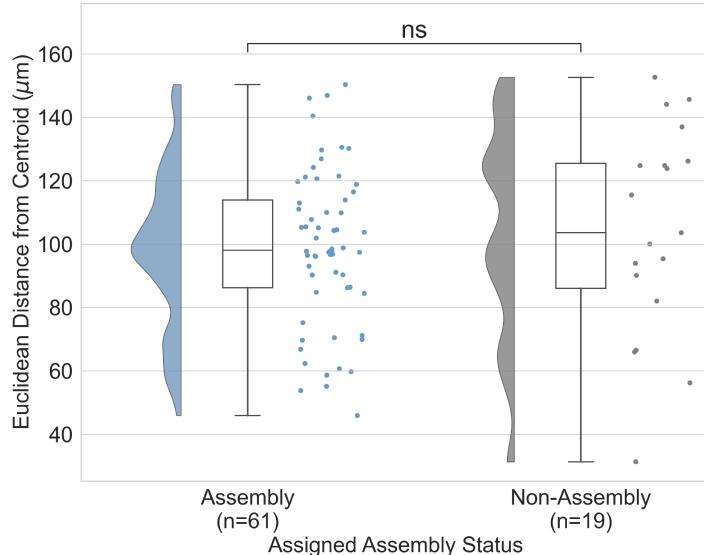
Supplementary Figure 3. Fitting Regressions to establish low-dimensional descriptions of pairwise activity correlations of cells. This framework was used to choose the optimal hyperparameters from Scan Volume 4, with those parameters later being chosen for Scan Volume 3. **(A)**. Pipeline to produce our regression fits to activity correlations. *Top-left*: The procedure begins with the pairwise correlation matrix of the whole scan throughout the hour recording. *Top-right*: Factor analysis reveals which dimensions of the original data matrix explain the most variance. This is done by plotting the eigenvalue of each factor, or dimension. The dotted red line is a horizontal marker of the eigenvalue of one, a threshold for considered dimensions justified by the Kaiser criterion. We ensure that at least ninety percent of the total variance is left explained in the final embedding. Less than fifty factors, or dimensions, are needed to do so. *Bottom-right*: With these dimensions in mind, we apply Isomap to produce a non-linear low-dimensional manifold of the original correlation matrix. This plot illustrates the first and second components from that embedding. *Bottom-right*: Finally, our low-dimensional description of the activity correlation, or our activity correlation space, is produced by plotting the distance of neurons across the Isomap embedding with respect to their pairwise correlations. The OLS fit (solid-red line) is a statistical model exemplifying how well the system continuum is able to describe this activity correlation space. **(B)** Correlation matrix of the whole scan with neurons ordered by assemblies. Cells on each axis were ordered by their largest corresponding assembly, from 'A 1' to 'A 15'. **(C)** OLS fits were developed that corresponded to assembly assignments with pairwise cells. This implies that multiple OLS fits were performed to account for all assignments. Regression fits to different assembly assignments are compared through the Akaike information criterion.

Identification of Tail Connections through Post-Synaptic Densities



Supplementary Figure 4. Histogram of connection strengths (\log_{10} -scaled) with a Gaussian Mixture Model fit overlaid (see Methods 5.11.5), and a Quantile-Quantile plot evaluating the fit. The red dashed line indicates the decision boundary separating the two Gaussian components, used to classify “tail” connections.

Distance from Centroid of Connectome Network Rank-Sum P-value: 0.45



Supplementary Figure 5. A raincloud plot comparing the distance from the centroid of the connectome network (Fig. 4A), demonstrating no difference between the distribution of distances with neurons assigned to an assembly and the distribution of those assigned to no assembly (Wilcoxon Rank-Sum p-value: 0.45).