

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.
¹⁸ He theorized that repeated activations induce the structural changes needed to group
¹⁹ neurons in these assemblies. Despite Hebb’s enduring influence, testing his predictions
²⁰ at relevant scales has been technically challenging. Here, we test the theory using a
²¹ novel, large-scale dataset featuring *in vivo* calcium fluorescence imaging of neural activity
²² with postmortem electron microscopy (EM) for detailed reconstruction of neurons
²³ from the same volume of mouse visual neocortex. A coregistration process matches
²⁴ EM-reconstructed neurons to their recorded fluorescence traces. From these traces, we
²⁵ extract neural assemblies from higher-order correlations in neural activity. We find
²⁶ multiple assemblies, many with overlapping neurons, and some neurons that do not
²⁷ participate in any assembly. We then show that these assemblies exhibit properties con-
²⁸ sistent with Hebb’s theory, including more reliable responses to repeated natural movie
²⁹ inputs than size-matched random ensembles and superior decoding of visual stimuli. Us-
³⁰ ing coregistration to probe structural correlates, we find that neurons that participate in
³¹ assemblies are significantly more integrated into the structural network than those that
³² do not. Contrary to Hebb’s original prediction, we do not observe a marked increase in
³³ the strength of monosynaptic excitatory connections between cells participating in the
³⁴ same assembly. However, we find significantly stronger indirect feed-forward inhibitory
³⁵ connections targeting cells in other assemblies. Intuitively, the delineation of assemblies
³⁶ can be realized either by internal excitation or external inhibition. Our findings support
³⁷ the latter mechanism. These results show the utility of assemblies in perception and
³⁸ provide a structural underpinning. They lay the foundation for future studies looking
³⁹ at the utility of assemblies in cognition and behavior, as well as the mechanisms for the
⁴⁰ formation and maintenance of such assemblies.

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

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

41 2 Introduction

42 Since Hebb's 1949 monograph, *The Organization of Behavior*[1], cell assemblies have
43 retained a persistent place in the imagination of the neuroscience community, both as
44 a prospective unit of functional organization and as a compellingly likely consequence
45 of simple rules of synaptic plasticity, such as the 'fire-together wire-together' synapses
46 dubbed 'Hebbian Synapses' by Yves Frégnac in 1986 [2]. The theory continues to stim-
47 ulate research on the activity-related aspects of assemblies, and the organization of
48 these processing modules is thought to operate through stable recurrent activity [3, 4].
49 Incorporating a behavioral approach, experimental evidence implies that the response
50 of discrete cell populations akin to assemblies may also have a causal link with motor
51 functions [5]. In particular, there has been confirmation of sequential activation within
52 neural assemblies of the primary visual cortex (V1) [6] as well as ongoing 'replay' of
53 coactivity in the absence of stimuli [7]. There has also been some limited evidence of
54 cell assemblies' stability over long periods of time [8].

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

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

78 To relate the correlated activity of a large number of neurons to their connections, we
79 used a novel large-scale multi-modal dataset: the Allen Brain Observatory V1 Deep Dive
80 (V1DD) [21] (Fig. 1A). V1DD offers a combination of Ca^{2+} fluorescent recordings and
81 detailed EM reconstruction of neurons and synapses, including post-synaptic density
82 volumes. Taking advantage of advances in optical imaging techniques [22], V1DD pro-
83 vides multiple scans of high-quality simultaneous two-photon calcium imaging (2PCI)
84 recordings of thousands of excitatory neurons within the mouse primary visual cortex
85 (Fig. 1B). In addition to its functional recordings, V1DD also contains electron mi-
86 croscopy (EM) of the same tissue volume, which has uncovered the fine-scale anatomy
87 of the cubic millimeter volume of the brain (Fig. 1C). EM has been employed exten-
88 sively in large-scale datasets to map drosophila [23, 24], worm [25], mouse [26, 27], and
89 even the human brain [28]. The combination of these two imaging modalities has been
90 applied in very few datasets [29, 30, 27], particularly at this scale.

91 These advances offer the opportunity to examine the structural correlates of Hebbian

V1 Deepdive: Large-Scale Multi-Modal Data

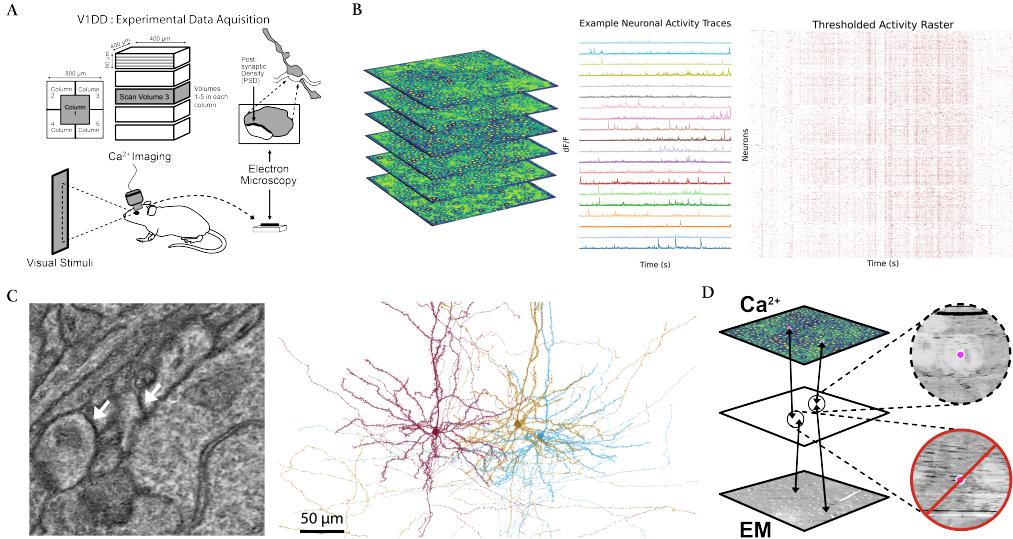


Figure 1: (A). Schematic of experimental data acquisition for the V1 Deepdive dataset (V1DD). V1DD consists of dense thousands of excitatory neurons in an $800 \times 800 \times 800 \mu\text{m}^3$ section of mouse V1, recorded during awake behaving imaging sessions, with no behavioral task. Our work 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.

92 cell assemblies at an unprecedented scale. We extract assemblies from a Ca^{2+} fluorescence
 93 imaging scan and examine the reliability of their activation and their functional
 94 significance in the encoding of visual stimuli. We then analyze the connectivity between
 95 neurons based on cells that were coregistered between fluorescence and structural EM
 96 scans (Fig. 1D). Deriving hypotheses directly from postulates advanced by Hebb, we
 97 test his predictions about synaptic sizes and connectivity within and across assemblies,
 98 producing results that suggest a significant role for inhibition in their formation and
 99 activation, different from what is traditionally assumed.

100 3 Results

101 3.1 Neuronal Organization of Hebbian Assemblies

102 We analyzed a scan of the optical imaging dataset consisting of 2708 excitatory
 103 neurons recorded in parallel. The Similarity Graph Clustering (SGC) algorithm (Fig.

104 2A) generated 15 assemblies, which we ordered by size with ‘A 1’ representing the largest
105 assembly ($n = 1016$) and ‘A 15’ the smallest ($n = 23$). A subset ($n = 748$, 27.6 percent
106 of 2708 total) of neurons was assigned to no assemblies.

107 Their spatial distributions are shown in (Fig. 2C, D). Although spatial organization
108 is not clearly visible, statistical analysis using the Kolmogorov-Smirnov (KS) [31] test
109 revealed significant differences in the spatial organization of assemblies compared to that
110 of the entire neural space (Fig. 2E). In particular, there was marked organization in
111 nearly every assembly along the X-Y plane, indicative of the retinotopic organization
112 observed in studies of the visual cortex [32].

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

117 3.2 Correlation and Sparsity

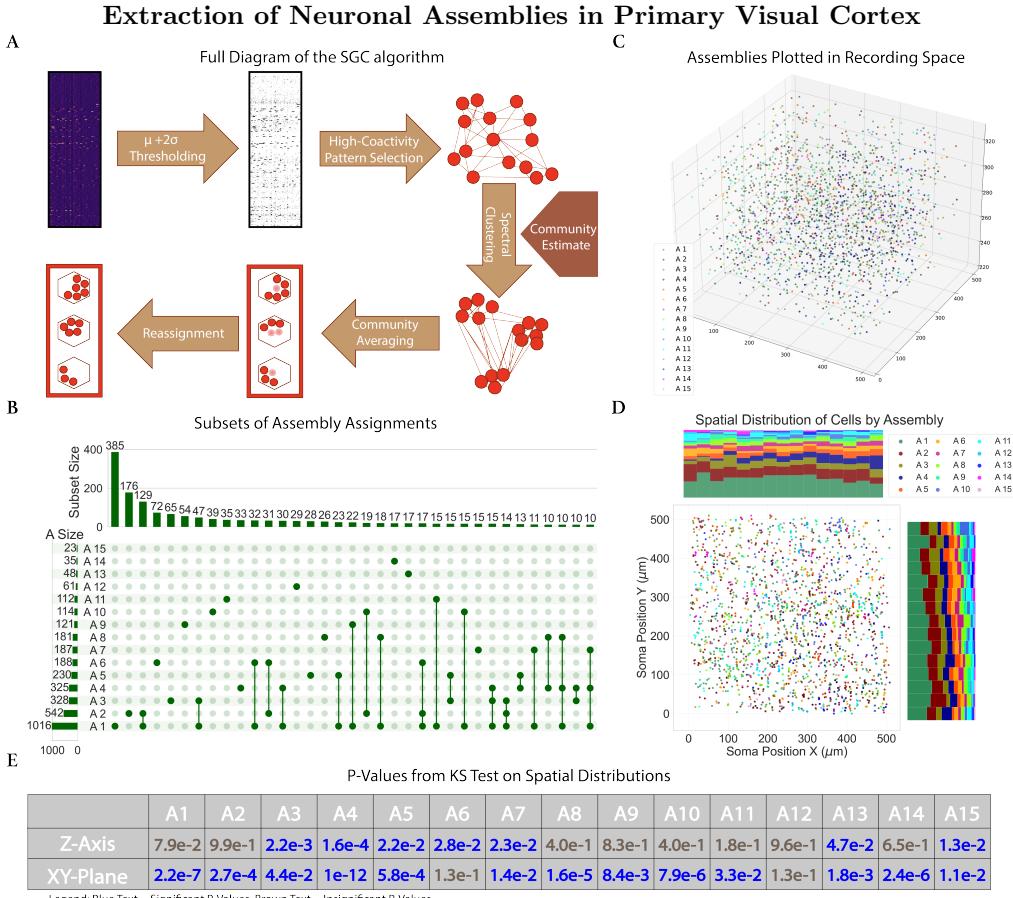
118 By construction, assemblies are expected to respond to distinct stimulus features,
119 with a coactivity correlation sufficiently low to prevent their merging into a single as-
120 sembly. Pearson correlation coefficients between assembly coactivity traces were signifi-
121 cantly lower than the coefficients between coactivity traces of random ensembles of the
122 same size distribution (see methods) (Fig. 3A). As expected from traces derived from
123 the average of population raster activity, both of these populations were significantly
124 higher in their correlation than individual cell activity raster correlations, regardless of
125 the subset of pyramidal cells being considered (assembly, non-assembly, or all individual
126 cells).

127 To further characterize the functional properties of assemblies, we computed the Gini
128 coefficient [35] for each assembly’s activity trace. The Gini coefficient, a statistical mea-
129 sure of the ‘inequality’ of signal activity throughout the optical recording, revealed that
130 assemblies exhibited highly sparse activity patterns (Fig. 3B), exhibiting significantly
131 higher Gini coefficients than random ensembles (p -value : $6.23e^{-5}$). The coefficient for
132 each assembly ranged from 0.55 to 0.83, with particular assemblies with extreme spar-
133 sity, such as ‘A 13’ (0.79) and ‘A 15’ (0.83), exemplifying a high degree of functional
134 selectivity. Interestingly, this sparsity metric was not solely dependent on assembly size,
135 as intermediate-sized assemblies (e.g., ‘A 4’ through ‘A 12’) all exhibited similar Gini
136 coefficients of around 0.70.

137 3.3 Assemblies Reliability in Stimuli Response

138 A section of the visual stimuli presented to the mouse consisted of natural movies.
139 Across the hour scan time, twelve unique 15-second movie clips were repeatedly shown
140 eight times. To evaluate the functional reliability of neuronal assemblies, we analyzed
141 their responses to these stimuli. Reliability was quantified using an Oracle score, a leave-
142 one-out correlation metric (see Methods) that measures the consistency of an activity
143 trace across repeated presentations of the same stimulus. Oracle scores have been used
144 as a measure of the reliability of neuronal response [36]. To provide a baseline with
145 which to compare the functionality of these assemblies to tune to visual stimuli, the
146 oracle scores of assemblies’ coactivity traces were compared to the oracle scores of all
147 neuronal traces. In addition, we provide a population-level comparison with random
148 ensembles.

149 Assemblies exhibited significantly higher Oracle scores compared to the average re-
150 liability of individual neurons (p -values < 0.0001), indicating that the assemblies, as
151 populations, respond more consistently to visual stimuli (Fig. 3C). To ensure this



152 result was not merely due to the inclusion of highly reliable neurons within assem-
153 blies, we separately calculated Oracle scores for neurons within assemblies and those
154 assigned to no assemblies. We observed no significant difference between the cellular
155 sets (p -values > 0.25). This result suggests that the reliability of these assemblies is
156 derived from their collective activity rather than from the reliability of individual mem-
157 bers.

158 Since population coactivity is expected to be more reliable in the general case than
159 individual neurons, we compared our assemblies to random ensembles (defined as in
160 Methods 5.2). Assembly coactivity traces demonstrated higher Oracle scores than coac-
161 tivity traces computed in the same way as the assemblies for size-matched random
162 ensembles. Complementary results are seen in the reliability of responses to orientation
163 gratings (S. ??). Population activity during the presentation of these gratings was used
164 to characterize reliability as a function of orientation [37].

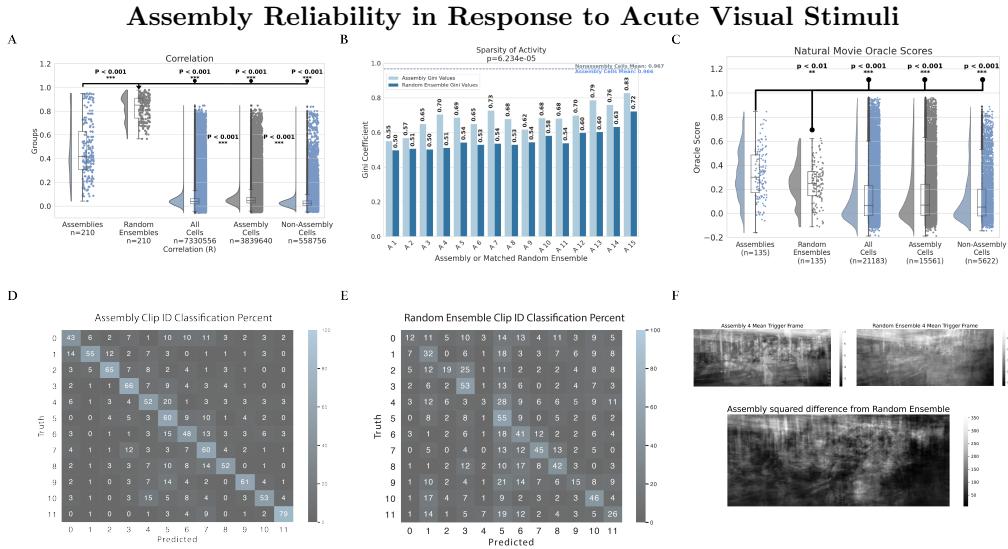


Figure 3: (A) Raincloud plot of pairwise correlations of coactivation between assemblies, size-matched random ensembles, and sets of individual cells. Coactivation for an individual cell is equivalent to a binary thresholded activity raster. (B). Grouped bar plot of sparsity (measured by the Gini Coefficient) 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: $6.06e^{-5}$, one-sided on diagonal elements u-stat: 131.5 p-value: 0.00032). (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.

165 Further analysis revealed consistent patterns of high assembly coactivity during specific visual frames of these natural movies. We defined ‘trigger frames’ as moments
 166 when assembly activity exceeded a baseline threshold (see methods) and found that
 167 these frames were highly consistent across repeated stimulus presentations. Example
 168 mean trigger frame pixel values are presented in (Fig. 3F), along with those of the cor-
 169 responding size-matched random ensemble, and the squared difference. Visualizations
 170 of these frames suggest that assembly activity responds to complex features in the nat-
 171 ural movies. These collective results provide evidence of assemblies’ ability to serve as
 172 functional populations with reliable and specific responses to visual stimuli.
 173

174 **3.4 Decoding Responses from Acute Visual Stimuli**

175 We also assessed the ability of assemblies to decode visual stimuli by implementing
176 a classification framework comparing assemblies to random ensembles (Fig. 3D, E). We
177 employed a Multi-Layer Perceptron Classifier (MLPClassifier) to evaluate how well each
178 grouping could decode the identities of the twelve natural movie clips. These classifiers
179 have been shown to be effective in academic and clinical settings, with high levels of
180 accuracy and flexibility in available hyperparameters [38, 39].

181 The results of our classifier revealed that assemblies significantly outperformed ran-
182 dom ensembles in accuracy. Heatmaps of classification accuracy for natural movie clip
183 identities demonstrated that assemblies provided more reliable decoding across repeated
184 trials. A Mann-Whitney U-test confirmed this finding, with assembly accuracy signifi-
185 cantly exceeding random ensembles for both overall performance (u-stat: 6742.5 p-value:
186 $2.89e^{-7}$) and diagonal elements (one-sided, u-stat: 139.0 p-value: $6.17e^{-5}$), indicating
187 enhanced stimulus-specific decoding.

188 **3.5 Structural Organization of Assemblies**

189 Coregistration between recorded activity and EM data provides us with a unique
190 opportunity to explore the structural underpinnings of assemblies in the visual cortex.
191 By mapping neural structure and connectivity at a micrometer resolution, we inves-
192 tigated the anatomical communication and organization of neurons with at least one
193 shared assembly membership (shared assembly cells) compared to those with disjoint
194 membership.

195 The strength of connections between the two groups was measured by performing
196 a Wilcoxon Rank Sum test to compare the connection weights (defined as the sum of
197 PSD volumes for all synapses between two cells). Chi-squared test of Independence was
198 performed to determine whether the pairwise frequency of connections differed between
199 cells with shared assembly membership and those with disjoint assembly membership.
200 To investigate higher-order structural patterns, we performed similar tests on sets of
201 inbound and outbound disynaptic chains, as well as subdividing based on whether the
202 intermediate cell in each chain was inhibitory or excitatory. 5.11.2

203 Our initial analysis of first-order connectivity surprisingly revealed no significant
204 differences in the probability of direct monosynaptic connections between shared and
205 disjoint sets or the strengths of those connections (Fig. 4C, F). This finding is incongruous
206 with predictions that assemblies are defined by densely interconnected excitatory
207 neurons [1] and suggested that the defining structural characteristics of assemblies might
208 lie beyond simple pairwise connectivity metrics.

209 To investigate higher-order structural patterns of these assemblies, we conducted a
210 motif analysis. In a neural network, when a significant number of subgraphs containing
211 a small number of interconnected cells repeat a particular pattern (e.g., Cell Type A →
212 Cell Type B → Cell Type A), they are classified as a motif [40, 41], with each subgraph
213 counted as a motif instance. The frequency with which these motifs persist has been
214 predictive of correlation in similar neural networks [42].

215 Our analysis concentrated on second-order chain motifs, or structures consisting of
216 three neurons connected by two synaptic links (Fig. 4D). Second-order neural motifs
217 are divisible into various types based on the arrangement of their synaptic connections.
218 For this study, we prioritized chain motifs, as they allowed us to utilize aspects of the
219 EM dataset that are not yet coregistered to activity recordings for the central elements
220 in the chain, so long as the first and last cells in the chain are coregistered.

Structural Connectivity of V1 Neuronal Assemblies

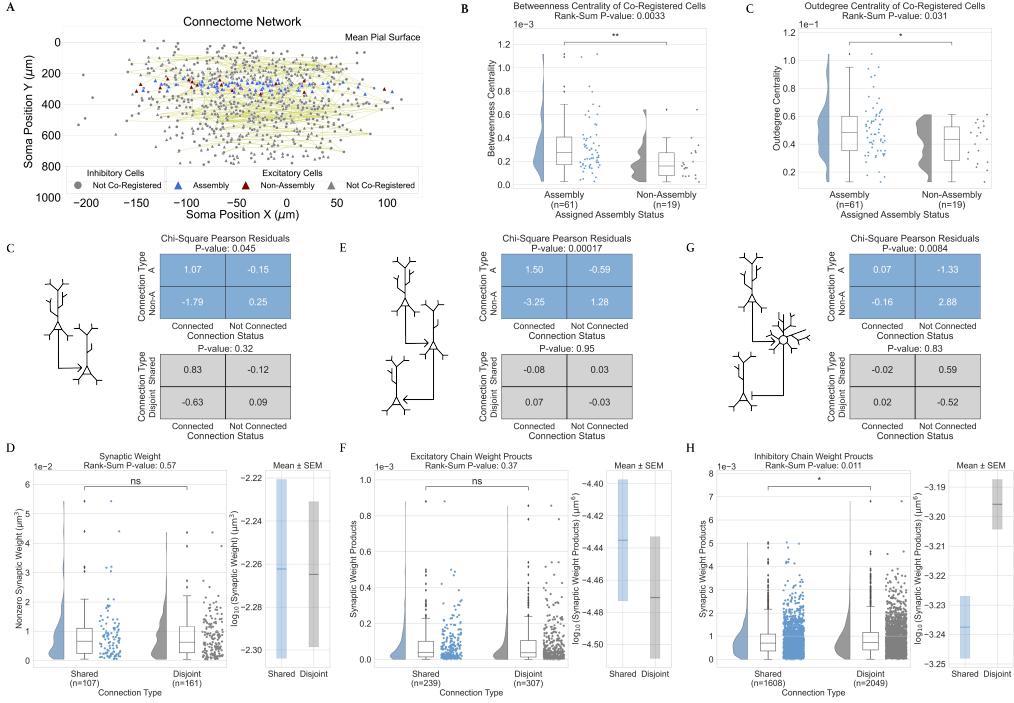


Figure 4: (A). Visualization of the network being analyzed, showing the soma position of cells in the connectome, colored by cell-type and assembly assignment. All coregistered reconstructed neurons were found in layer 2/3 or layer 4 of V1. (B) A raincloud plot of betweenness centrality, demonstrating 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). (C,E, G). Chi-squared analysis of the likelihood of monosynaptic (C), disynaptic excitatory (E), and disynaptic inhibitory (G) 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). (D, F, H) Raincloud plots showing the combined synaptic PSD volume per extant monosynaptic (F), disynaptic excitatory (G), and disynaptic inhibitory (H) connection, each divided between origin and terminus cell pairs which share assemblies and those which participate in disjoint assemblies.

Notably, while our excitatory chain analysis revealed no significant differences between shared and disjoint assembly memberships (Fig. 4D,G), an analysis of inhibitory motifs convey a different story: disjoint assembly memberships exhibited significantly stronger feed-forward inhibitory connections than shared memberships (Fig. 4E,H). Log-scaled plotting of the chain weights (Fig. 4H, inset) revealed a complete lack of overlap between SEM-based confidence intervals, suggesting that this is a result with both reliability and a non-trivial effect size. Notably, this result was not significant when restricting our analysis to only the inhibitory synapse or only the excitatory synapse within the feed-forward inhibitory chain, suggesting that both connections may play a role.

This suggests a mechanism of mutual inhibition that may regulate the interaction between distinct assemblies, preventing excessive coactivation and ensuring the dis-

233 criminability of assembly responses to inbound information. Indeed, we find a negative
234 Pearson r correlation coefficient between the disjoint feed-forward inhibitory chain
235 weights and the correlation of the two disjoint assemblies' coactivity traces: r statistic =
236 -0.19, p-value = 0.0048 (See methods 5.11.2).

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

244 3.6 Non-Assembly Cells

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

249 We show assembly cells exhibiting a pattern of connectivity distinct from non-
250 assembly cells, a requirement of Hebb's theory. We assessed the higher-order integra-
251 tion of assembly neurons into the broader structural network using centrality metrics.
252 In particular, we found significantly lower betweenness centrality (Fig. 4B), a measure
253 of a node's importance in mediating communication within a network, and outdegree
254 centrality (Fig. 4C), an analog for probability of outbound connectivity, in neurons
255 outside assemblies compared to neurons within assemblies. Furthermore, we found that
256 the pairwise probability of monosynaptic connections, disynaptic excitatory chains, and
257 disynaptic inhibitory chains was significantly greater for cells sharing assembly member-
258 ship than for non-assembly cells (p-values = 0.045, 0.0002, and 0.0084, respectively). No
259 significant differences in PSD Volumes between assembly and non-assembly cells were
260 observed. These findings confirm that cells in the connectome participating in at least
261 one assembly are more interconnected than those not participating in any.

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

266 4 Discussion

267 4.1 Inhibition as the Delineating Basis of Hebbian Assemblies

268 This work provides the first demonstration of the structural underpinnings of Hebb-
269 ian assemblies and validates Hebb's assembly theory in a surprising way; while there
270 is strong agreement on functional predictions, our structural findings run counter to
271 Hebb's original proposal. We do not observe a marked increase in the strength or prob-
272 ability of excitatory connections between cells that participate in the same assembly
273 (Fig. 4 C-F). Instead, we find significantly stronger feed-forward inhibitory connections
274 targeting cells in other assemblies (Fig. 4 H). Hebb and others [43, 9, 3] have hypothe-
275 sized that targeted inhibition could play a crucial role in the formation and delineation
276 of assemblies. Our findings offer the first clear empirical evidence for this mechanism.

277 Despite this novel result, our findings remain a validation of Hebb's structural pos-
278 tulates and are consistent with what would be required for assemblies to be delineated
279 by inhibition. Assembly cells were substantially more integrated into the connectome,

280 displaying significantly higher betweenness and outdegree centrality (Fig. 4 B, C). This
281 pattern demonstrates that cells participating in our activity-defined assemblies possess
282 markedly different patterns of connectivity from cells that do not, establishing the nec-
283 essary structure-function relationship postulated by Hebb.

284 Intuitively, the delineation of assemblies can be realized on a spectrum between in-
285 ternal (shared-assembly) excitation and external (disjoint-assembly) inhibition. Shared-
286 assembly excitation would rely on broad inhibitory connectivity across the entire pop-
287 ulation of assembly neurons, and disjoint-assembly inhibition would similarly rely on
288 broad excitatory connectivity. We find a significantly greater probability of excitatory
289 connections between cells belonging to any assembly (Fig. 4 C, E), however we did not
290 observe significantly stronger or more frequent excitatory connections between neurons
291 with shared assembly membership (Fig. 4 C-F). Combined with our finding of signifi-
292 cantly greater feed-forward inhibitory weights between neurons with disjoint assembly
293 membership (Fig. 4 H), the broad excitation strongly suggests that the extracted as-
294 semblies are delineated predominantly via inhibition.

295 **4.2 Functional Consistency of Assemblies**

296 Functionally, ensembles of cells classified as assembly cells behave in ways consistent
297 with Hebbian assemblies. The assemblies demonstrated markedly higher reliability in
298 their responses to naturalistic visual stimuli compared to individual cellular responses
299 of all three sub-populations of cells and size-matched random ensembles, reflected by
300 significantly elevated Oracle scores (Fig. 3C). This consistent coactivity of the assemblies
301 has been hypothesized to offer resilience to representational drift, allowing them to
302 serve as a substrate for long-lasting representation [8]. Representational drift reflects a
303 functional characterization of fluctuations at the cellular level, such as synaptic turnover
304 [44, 45], even under stable stimulus conditions [46]. This phenomenon has been observed
305 in areas other than V1, such as the piriform cortex [47]. By maintaining consistent and
306 coherent patterns of coactivity, assemblies may offer a general cortical mechanism of
307 stable perceptual representations despite such turnover.

308 Beyond their reliability in responses, assemblies were also superior in our decod-
309 ing of natural movie presentations (Fig. 3D), underscoring their efficacy in extracting
310 higher-order visual information. In addition, the sparsity of assembly coactivity was
311 substantially greater than that of random ensembles (Fig. 3B), which, when combined
312 with a lower average correlation 3A), is consistent with assemblies employing a cost-
313 effective encoding strategy [48]. Such a strategy would enhance the capacity for distinct
314 representation of sensory inputs, translating more flexible individual responses into reli-
315 able population-level encoding. Finally, cells assigned to assemblies exhibit significantly
316 higher pairwise correlations than non-assembly cells, corroborating the experimental
317 work of Harris and Carandini [49] defining ‘choristers’ and ‘soloists’. However, we found
318 no difference in their average response reliability to natural movies or in their overall
319 signal sparsity. This result implies that a population perspective is required for the
320 encoding of reliable perceptually relevant stimulus features.

321 **4.3 Connective Tissue of Structure and Activity**

322 Some examined ensemble features bridge structure and activity, a crucial aspect
323 of this study, which has not previously been achievable at scale. First, the extensive
324 overlap of neurons across multiple assemblies (Fig. 2D) is consistent with both Hebb’s
325 proposal of sub-assemblies [9] and Rafael Yuste’s review of definitions put forward for
326 ensembles [50], with individual neurons contributing to multiple functional modules.
327 Second, our observation of greater chain weights in feed-forward inhibitory chains offers

328 a mechanistic explanation of how assemblies can retain distinct responses to stimuli
329 without becoming so correlated in their activity that they merge into a single functional
330 ensemble. The plausibility of this explanation is supported by the significant observed
331 negative correlation between inhibitory chain connection strength between disjoint as-
332 sembles and the r coefficient of their paired assembly coactivity traces, showing that
333 the greater the inhibitory weight of the connections observed, the less correlated the
334 activation of the two assemblies is 3.5.

335 4.4 Further Directions

336 Looking ahead, assemblies exhibit notable variability in size (Fig. 2D). SGC min-
337 imizes overestimation of the number of assemblies or neuron assignments, suggesting
338 that these size differences reflect intrinsic properties rather than methodological arti-
339 facts. There is therefore an opportunity for future studies to delineate the distinct roles
340 potentially served by larger assemblies, such as 'A 1', compared to their smaller counter-
341 parts. In addition, while novel in its scale and multifaceted nature, we see opportunities
342 to reduce some of the current limitations within this dataset. Continued reconstruction,
343 proofreading, and coregistration will allow analyses of more neurons and subsequent
344 new lines of inquiry. For instance, the SGC algorithm flags frequent patterns of ac-
345 tivity, potentially analogous to Hebbian phase sequences [6, 7]. We currently do not
346 have sufficiently many coregistered and fully reconstructed cells to allow examination
347 of connections that bridge one pattern to the next. Exploring how these patterns inter-
348 act could reveal mechanisms of large-scale neural coordination. Other limitations will
349 be more difficult to overcome without the availability of next-generation multi-modal
350 datasets. Most importantly, the temporal resolution of the scans leaves us with no ability
351 to examine activity on some of the most relevant time scales for synaptic plasticity.

352 These results suggest several potential implications for the field at large. Hebb pro-
353 posed assemblies as a universal building block, simultaneously addressing the functional
354 and structural sides of perception, cognition, and behavior. A basic unit that bridges
355 structure and function allows one to derive structural predictions from functional char-
356 acterizations and vice versa. Beyond validating Hebb's theory, the evidence presented
357 here for an underlying inhibitory mechanism provides further support for the analysis of
358 inhibition in cognitive and sensory disorders studied to date primarily through the lens
359 of excitation [51, 52]. In the future, the integration of cell-type-specific genetic informa-
360 tion with functional assembly data could also provide deeper insights into the molecular
361 foundations of assembly formation and maintenance in health and disease [53]. Hebb
362 postulated that assemblies form the atoms of cognition, and it has not escaped our
363 notice that the cross-inhibitory mechanism we here demonstrate might be a universal
364 feature of brain-wide assembly organization.

365 **5 Methods**

366 **5.1 V1 Deepdive Dataset**

367 **5.1.1 Stimuli**

368 Visual stimuli were presented using the same monitor configuration as in de Vries
369 et al. [54]. Imaging sessions were one hour long and offered a wide variety of visual
370 stimuli. Assembly extraction was performed on the fluorescence data from the full
371 session. The remainder of our analysis utilizes only the natural movie clips and the full-
372 field drifting gratings, details of which are provided below. The other stimuli presented
373 during the session included natural images, windowed drifting gratings, and locally
374 sparse noise. The details of these other classes of stimuli can be found in [21]. The
375 natural movies stimulus consisted of 3,600 frames (with 30 Hz frame rate), presented 8
376 times. The full-field drifting gratings stimulus consisted of a drifting sinusoidal grating at
377 a 1 Hz temporal frequency and 80 percent contrast, presented at 12 different orientations
378 (multiples of 30°) and at 2 spatial frequencies (0.04 and 0.08 cycles per degree). Each
379 condition was presented eight times, in randomized order, with one second of mean
380 luminance grey between presentations.

381 **5.1.2 2/3 Photon Microscopy and Activity Data Processing**

382 The full $800 \times 800 \times 800 \mu\text{m}^3$ volume was divided into five columns, each imaged via ei-
383 ther 2-photon (2P) or 3-photon (3P) microscopy, depending on depth. The volumes were
384 scanned over the course of several sessions. Each session consisted of the full set of visual
385 stimuli (see below), presented in the same order and with the same timing. Our anal-
386 ysis is concerned only with the central column, centered within the $800 \times 800 \times 800 \mu\text{m}^3$
387 volume, where the largest set of reconstructed, coregistered, and proofread neurons is
388 currently available. Figure 1A shows the arrangement of scan volumes (5 volumes span-
389 ning 75 μm to 620 μm depth, each $400 \times 400 \times 80 \mu\text{m}^3$) and imaging planes (6 planes
390 within each scan volume, separated by 16 μm at thirty-seven frames per second so that
391 each plane is imaged at 6 Hz.) We selected volume three of the central column for our
392 primary analyses and validated our work with volume four. Both were imaged with 2P
393 microscopy. Details on the other columns and the 3P imaging can be found in [21].

394 The fluorescence data was preprocessed using the standard LIMS pipeline as used
395 for the Allen Institute’s Visual Coding 2P dataset [54], including motion correction,
396 segmentation, demixing, neuropil subtraction, ROI filtering, and df/f calculation. Iden-
397 tified regions of interest were run through a classifier trained on manual labeling data
398 meant to reduce the false classification of artifact ROIs as neuronal somas, and only
399 those with a high confidence score (at least 0.5) were included in our analysis.

400 **5.1.3 Electron Microscopy and Reconstruction**

401 The mouse was transcardially perfused with a fixative mixture of paraformaldehyde
402 and glutaraldehyde. All procedures were carried out in accordance with the Institutional
403 Animal Care and Use Committee at the Allen Institute for Brain Science. The large-
404 volume staining protocol was adapted from [55]. After dissection, the neurophysiological
405 recording site was identified by mapping the brain surface vasculature. A thick (1200-
406 μm) slice was cut with a vibratome and post-fixed in perfusate solution for 12 to 48h.
407 The tissue was then infused with heavy metals, dehydrated, and embedded in EMS
408 Hard Plus resin. After curing, the samples were epoxy cured to a stub. They were then
409 sliced and placed onto continuous tape by a ATUMtome Automated Tape Collecting
410 Ultramicrotome.

411 The continuous tape was fed into an automated high-throughput transmission elec-
412 tron microscopy pipeline[56]. Transmission electron microscopy is particularly well
413 suited for automated imaging and preserves very good x-y resolution at the expense
414 of some resolution on the vertical z-axis, and so specialized methodology was deployed
415 during reconstruction[57, 58]. Serial section alignment was performed through a contract
416 with Zetta A.I, followed by stitching [59, 56], segmentation, and automated reconstruc-
417 tion. Proofreading of a subset of cells was performed under contract by Ariadne.ai.

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

422 5.1.4 Coregistration

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

437 5.2 Graph Clustering for Assembly Extraction

438 To extract assemblies, we use the Similarity-Graph-Clustering (SGC) algorithm that
439 has been originally proposed for the detection of neural assemblies during spontaneous
440 activity in the zebrafish optic tectum [63]. The Similarity-Graph-Clustering (SGC) al-
441 gorithm identifies neuronal assemblies using ideas from graph theory by transforming
442 the problem into one of community detection on some graph in which assemblies corre-
443 spond to distinct (graph) communities. Unlike traditional methods that rely on pairwise
444 correlations between cells, SGC groups frames of fluorescence indicative of significant
445 coactivity [33]. These moments of higher-order correlation are referred to as potential
446 'phase sequences', representing when populations of neurons act cohesively as a closed
447 circuit during assembly activation [64].

448 Importantly, by design, SGC allows cells to participate in multiple assemblies. As-
449 sembly overlap has been integral to assembly studies in the past [65]. Hebb originally
450 postulated that the shift between active assembly states could be what is colloquially
451 referred to as a 'train of thought' [1]. Our results revealed a sizable degree of overlap
452 between assemblies, which was visualized through an UpSet plot (Fig. 2D) [34], showing
453 neurons frequently assigned to multiple assemblies (Fig. 2D).

454 Compared with several other prominent assembly extraction algorithms upon appli-
455 cation to both simulated and biological calcium imaging datasets under different condi-
456 tions, this algorithm was shown to perform best overall [33]. Although it did not yield
457 a perfect reconstruction of the assemblies in the biological dataset, SGC was able to
458 recover the assemblies with higher accuracy than all other algorithms. In part, this per-

459 formance has been attributed to the computational effort that SGC places in estimating
460 the number of assemblies before defining them.

461 We used a recent implementation of the SGC algorithm in Python [66]. For completeness,
462 we briefly recall the main steps here: The algorithm commences by thresholding
463 the calcium fluorescence signals (df/f) of the ROIs (neurons) by two standard deviations
464 above the mean to minimize noise (Fig. 2B). Afterwards, activity patterns are selected
465 where the coactivity level of neurons exceeds the significance threshold. This threshold
466 that determines the set of “high-activity patterns” is based on a null model of coactivity
467 obtained from shuffling the thresholded, activity signals (significance value: 0.05, rounds:
468 1000). A k -nearest-neighbor graph is then constructed from the set of high-activity
469 activity patterns based on the similarity between the patterns in the cosine distance. For
470 that, the number of neighbors k is chosen automatically such that the resulting graph
471 is connected. In the next step the number of communities in this graph is estimated
472 using a statistical inference procedure. For our study, we fixed the hyperparameters
473 with five independent Monte Carlo rounds of 150,000 steps each. With an estimate for
474 the most likely number of communities, spectral clustering is applied. These clusters of
475 high-activity patterns obtained from that are the first prospective selection of activity
476 patterns corresponding to the assemblies. However, the final step is a combination of
477 averaging and reassignment to reject groups that may have been erroneously defined
478 because of a high level of noise in the original signal. This minimizes overestimating
479 the number of assemblies or the neurons that should be assigned to those assemblies.
480 Finally, neurons are assigned to assemblies based on their affinity, the probability that
481 they were active in any of the assemblies’ activity patterns (affinity: 0.4).

482 5.3 Random Ensembles

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

487 5.4 Oracle Scores

488 Oracle scores are a measure of the reliability of the trace response of cell activity
489 to repeated visual stimuli computed through a jackknife mean, or leave-one-out mean
490 [67, 68], of correlations between the activity trace across the repeated visual stimuli.

491 Let $S_i(t)$ denote the activity trace of a neuron or an assembly during the i -th pre-
492 sentation of a stimulus at time t . Let $\bar{S}_{-i}(t)$ represent the mean activity trace of all
493 activity traces excluding the i -th presentation. From this, we can calculate the Oracle
494 score O as:

$$495 O = \frac{1}{n} \sum_{i=1}^n \frac{\text{Cov}(S_i(t), \bar{S}_{-i}(t))}{\sqrt{\text{Var}(S_i(t)) \cdot \text{Var}(\bar{S}_{-i}(t))}}$$

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

499 5.5 Trigger Frames

500 To assess the visual stimuli associated with high activity in neuronal assemblies,
501 we computed trigger frames by identifying peak activity times and extracting the corre-

502 sponding images from a natural movie presentation. From a coactivity trace for assembly
 503 k over time $S_k(t)$, we can detect peaks P_k using the **scipy** signal package to define local
 504 maxima. We then define a mean trigger frame $\mu_k(x, y)$ for a pixel (x, y) as:

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

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

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

513 5.6 Decoder

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

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

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

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

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

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

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

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

541 **5.7 Gini coefficient**

542 The Gini coefficient [35], a statistical measure that exemplifies the state of inequality
543 within a population. While often applied in economics to evaluate income inequality,
544 this metric has been applied as a valid approximation for signal sparsity [69, 70]. This
545 rendition of the application has been shown to serve as a relatively simple and robust
546 measure [71]. For our study, the coefficient is employed to quantify assembly signal
547 heterogeneity.

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

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

553 This computation is performed independently for each assembly, providing a metric
554 of signal inequality. A G_A of 0 implies all values are identical, while a value of 1
555 indicates perfect inequality. A high G_A indicates that the coactivity is dominated by a
556 small number of time points, reflecting temporal sparsity.

557 **5.8 Correlation**

558 To examine coactivity correlations, for each pair of assemblies, each pair of random
559 ensembles, and each pair of neurons in the set of all neurons, the set of neurons that
560 participate in at least one assembly, and the set of neurons that participate in no assem-
561 blies, we computed the Pearson's Correlation Coefficient r between the two coactivity
562 traces (where the coactivity trace of a single neuron is mathematically equivalent to its
563 thresholded raster activity, per [33]).

564 **5.9 Motif Extraction**

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

572 To further refine our analysis, chain motifs were categorized based on the type of
573 intermediary neuron, distinguishing between excitatory and inhibitory connections, with
574 the latter providing insights into feed-forward inhibition mechanisms.

575 **5.10 Motivating Postulates**

576 Our statistical analyses involve tests of the following postulates. First, that excita-
577 tory connections between cells which share at least one assembly ('shared' connections)
578 will be stronger than connections between cells which do not participate in any of the
579 same assemblies ('disjoint' connections), due to Hebbian plasticity [1]. This was exam-
580 ined both with regard to the post-synaptic density volume of monosynaptic connections
581 between known coregistered cells within the dataset, and separately in the form of the
582 product of connection PSD volumes in disynaptic excitatory chains which originated

583 and terminated with shared or disjoint cells, allowing us to evaluate indirect excitatory
584 connections where the middle cell had not yet been coregistered.

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

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

604 Fourth, and finally, we acknowledge that Hebb discusses the reinforcement of sparse
605 connectivity in his accounts of the emergence of cell assemblies [43], and thus this re-
606 enforcement might have a significant effect on a sparse subset of connections while pro-
607 ducing a minimal difference in the central tendency of the overall set of connections.
608 Dorkenwald et al. [75] demonstrated a bimodality of the log PSD volume of excitatory-
609 excitatory synapses in the similar MICrONS mm³-dataset [27], suggesting that a subset
610 of such connections is impacted differently by processes determining PSD volume. Com-
611 bining these two we decided to test the hypothesis that the larger of the two-component
612 distributions found in [75] would be more likely than chance to involve connections
613 between shared assembly neurons.

614 5.11 Statistical Methods

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

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

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

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

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

- 627 • $\text{Shared}_{ij} = 1$ if $A_j \cap A_i \neq \emptyset$
- 628 • $\text{Disjoint}_{ij} = 1$ if $A_j \cap A_i = \emptyset$

- 629 • Assembly_{*ij*} = 1 if $A_j \neq \emptyset$ and $A_i \neq \emptyset$
 630 • Non-assembly_{*ij*} = 1 if $A_j = \emptyset$ and $A_i = \emptyset$

631 **5.11.1 Monosynaptic Metrics**

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

636 The probability of monosynaptic outbound connection for a pre-cell j was calcu-
 637 lated as the proportion of realized connections under a given connection type $C \in$
 638 {Shared, Disjoint}, normalized by the total number of potential post-cell partners for
 639 that connection type:

$$b_{\text{out}_{j,C}} = \frac{\sum_{i|C_{ij}=1} b_{ij}}{|\{i \mid C_{ij} = 1, i \neq j\}|} \quad (1)$$

640 Similarly, the probability of a monosynaptic inbound connection for a post-cell i was
 641 defined as:

$$b_{\text{in}_{i,C}} = \frac{\sum_{j|C_{ij}=1} b_{ij}}{|\{j \mid C_{ij} = 1, j \neq i\}|} \quad (2)$$

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

$$w_{\text{out}_{j,C}} = \frac{\sum_{i|C_{ij}=1} w_{ij}}{\sum_{i|C_{ij}=1} b_{ij}} \quad (3)$$

645 Similarly, the average realized summed monosynaptic inbound PSD for a post-cell i
 646 is given by:

$$w_{\text{in}_{i,C}} = \frac{\sum_{j|C_{ij}=1} w_{ij}}{\sum_{j|C_{ij}=1} b_{ij}} \quad (4)$$

647 **5.11.2 Disynaptic Metrics**

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

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

$$w_{ikj} = w_{ik} w_{kj} \quad (5)$$

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

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

668 Let n_e represent the number of excitatory cells and n_i represent the number of
 669 inhibitory cells from the set of all cells in the all-all connectome. Define $|k|$ be the
 670 number of potential middle partners. For inhibitory chains, $|k| = n_e$ as the middle cell
 671 is inhibitory. For excitatory chains, the middle cell cannot be the first or final cell, so
 672 $|k| = n_e - 2$.

673 The probability of disynaptic outbound connection for a pre-cell j was calculated
 674 as the proportion of realized disynaptic connections under a given connection type $C \in$
 675 $\{\text{Shared}, \text{Disjoint}\}$, normalized by the total number of potential chains satisfying the
 676 connection type:

$$b_{\text{out}_{j,C}} = \frac{\sum_k \sum_{i|C_{ij}=1} b_{ikj}}{|k| |\{i \mid C_{ij} = 1, i \neq j\}|} \quad (6)$$

677 Similarly, the probability of disynaptic inbound connection for a post-cell i was
 678 defined as:

$$b_{\text{in}_{i,C}} = \frac{\sum_k \sum_{j|C_{ij}=1} b_{ikj}}{|k| |\{j \mid C_{ij} = 1, j \neq i\}|} \quad (7)$$

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

$$w_{\text{out}_{j,C}} = \frac{\sum_k \sum_{i|C_{ij}=1} w_{ikj}}{\sum_k \sum_{i|C_{ij}=1} b_{ikj}} \quad (8)$$

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

$$w_{\text{in}_{i,C}} = \frac{\sum_k \sum_{j|C_{ij}=1} w_{ikj}}{\sum_k \sum_{j|C_{ij}=1} b_{ikj}} \quad (9)$$

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

687 **Set Definitions**

688 The following sets were defined to evaluate connectivity metrics:

- 689 • The set of nonzero pairwise connection strengths under a given connection type
690 C:
691

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

- 692 • The set of outbound probabilities of connection for each pre-cell under a given
693 connection type C:
694

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

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

- 698 • For inbound or outbound metrics where pre- or post-cells appear in both the
699 Shared and Disjoint groups, paired sets were constructed as:
700

$$\{b_{\text{out}_{j,C}}\}_{\text{paired}} = \{b_{\text{out}_{j,C}} \mid j \in \{j \mid b_{\text{out}_{j,\text{Shared}}} \} \cap \{j \mid b_{\text{out}_{j,\text{Disjoint}}} \}\}.$$

701 Similar paired collections were defined for inbound probabilities and both inbound
702 and outbound nonzero average connection strengths.

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

705 **Statistical Tests**

706 We performed one-way statistical tests at $\alpha = 0.05$ to compare the Shared and
707 Disjoint groups. All alternative hypotheses predict Shared > Disjoint, except for tests
708 involving di-synaptic Inhibitory chain sets, in which the alternative hypotheses predict
709 Shared < Disjoint. We ran the following tests:

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

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

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

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

717 Finally, to examine the functional correlates of the di-synaptic inhibitory chain find-
718 ings, we calculated the Pearson's Correlation Coefficient between the summed feed-
719 forward inhibitory weights of disjoint cells connecting an assembly pair (see methods
720 5.11.2) and the correlation scores between the assembly pair's coactivity traces (see
721 methods 5.8), and examined significance.

719 **5.11.3 Tail Analysis**

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

723 To identify “tail” connections, we modeled the distribution of connection strengths
724 using a Gaussian Mixture Model (GMM) with $k = 2$ components. The model was ini-
725 tialized via k-means clustering to estimate the weights, means, and standard deviations
726 of each component. The decision boundary separating the two Gaussian components
727 was calculated as the intersection of their weighted probability density functions, derived
728 using a quadratic equation based on the GMM parameters. Connections with values
729 greater than or equal to the decision boundary were classified as “tail” connections. We
730 present the model fit and evaluation as well as the tail boundary in the supplemental
731 figures section.

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

737 **5.11.4 Centrality analysis**

738 Centrality analysis was used to quantify whether in a given network, assembly cells
739 were more likely to be central to the network than non-assembly cells. This analysis
740 gave further insights into the role of assembly cells in higher-order connectivity. To
741 do this, we measured different centrality metrics for assembly and non-assembly cells,
742 namely, in-degree centrality, outdegree centrality, closeness centrality, and betweenness
743 centrality.

744 In a graph, the centrality of a node refers to its tendency to connect and generally in-
745 fluence other nodes within the network [76]. We developed a directed graph $G = (N, E)$,
746 using the binary connectome such that N represents all the cells in the connectome and
747 E represents the binary, directed connections between all cell pairs. The total number
748 of outbound synaptic connections are given by $\sum_n \deg^+(n)$ and inbound synaptic con-
749 nections are given by $\sum_n \deg^-(n)$. Normalizing these connections, in-degree centrality
750 was calculated as:

$$I = \frac{\sum_n \deg^-(n)}{(N - 1)} \quad (10)$$

751 and outdegree centrality was calculated as:

$$O = \frac{\sum_n \deg^+(n)}{(N - 1)} \quad (11)$$

752 where $n \in N$.

753 A monosynaptic connection, $e = (j, i) \in E$ between a pre-cell j and post-cell i , will
754 have i as the head and j as the tail end of the connection, where j and $i \in N$. The
755 path between a pre-cell j acting as a source neuron and a post-cell i acting as a target
756 neuron, is the alternating sequence of cells and connections starting from j and ending
757 at i , with each cell before a connection in the sequence being a pre-cell and each cell
758 after a connection being a post-cell. The number of monosynaptic connections within
759 the path indicates the length of the path. The shortest path between cells j and i is the
760 minimum length between the two cells. The shortest path, σ_{ij} , is also referred to as
761 the geodesic path.

762 Based on this, the closeness centrality for a given cell i was calculated as the reciprocal
763 of the summation of the shortest paths, or distances between the post-cell i and all other
764 pre-cells j in the graph.

$$V(i) = \frac{N - 1}{\sum_{j=1}^{N-1} \sigma_{ij}} \quad (12)$$

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

$$B_i(k) = \sum_E \frac{\sigma(k)_{ij}}{\sigma_{ij}} \quad (13)$$

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

768 **6 Code Availability**

769 All code is available at this GitHub link: <https://github.com/AllenInstitute/HebbsVision>

770 It should be noted that, for all code of our own generation, a fixed random seed is
771 used for reproducibility. Libraries used analysis and plotting of results include: numpy
772 [78], pandas [79], scikit-learn [80], upsetplot [34], networkx [81], dotmotif [82], matplotlib
773 [83], seaborn [84], Raincloud [85].
774

775 **7 Acknowledgements**

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

780 **8 List of Contributions:**

Name	Contributions
J Wagner-Carena	Conception and Design, Functional and Structural Data analysis, Data Interpretation, Writing
S Kate	Data Interpretation, Chain and Graph Theoretic Data Analysis, Writing
T Riordan	Data Interpretation, Statistical Data Analysis, Writing
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
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
S Berteau	Conception and Design, Data analysis, Data Interpretation, Writing

781 References

- 782 [1] Hebb, D. *The Organization of Behavior* (John Wiley & Sons, New York, 1949).
- 783 [2] Frégnac, Y. Aplysia: Hebbian or not? *Trends in Neurosciences* **9**, 410 (1986).

- 784 URL https://www.academia.edu/71978970/Aplysia_Hebbian_or_not.
- 785 [3] Buzsáki, G. Neural Syntax: Cell Assemblies, Synapsemsembles, and Readers. *Neuron* 786 **68**, 362–385 (2010). URL <https://www.sciencedirect.com/science/article/pii/S0896627310007658>.
- 787 [4] Miehl, C., Onasch, S., Festa, D. & Gjorgjieva, J. Formation and computational
788 implications of assemblies in neural circuits. *The Journal of Physiology* **601**,
789 3071–3090 (2023). URL <https://onlinelibrary.wiley.com/doi/abs/10.1113/JP282750>.
790 _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1113/JP282750>.
- 791 [5] Matheus Gauy, M. *et al.* A Hippocampal Model for Behavioral Time Acquisition
792 and Fast Bidirectional Replay of Spatio-Temporal Memory Sequences. *Frontiers in Neuroscience* 793 **12**, 961 (2018). URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6306028/>.
- 794 [6] Carrillo-Reid, L., Miller, J.-E. K., Hamm, J. P., Jackson, J. & Yuste, R. Endogenous
795 sequential cortical activity evoked by visual stimuli. *The Journal of Neuroscience: The Official
796 Journal of the Society for Neuroscience* **35**, 8813–8828 (2015).
- 797 [7] MacLean, J. N., Watson, B. O., Aaron, G. B. & Yuste, R. Internal Dynamics
798 Determine the Cortical Response to Thalamic Stimulation. *Neuron* **48**,
800 811–823 (2005). URL <https://www.sciencedirect.com/science/article/pii/S0896627305010019>.
- 801 [8] Pérez-Ortega, J., Alejandre-García, T. & Yuste, R. Long-term stability of cortical
802 ensembles. *eLife* **10**, e64449 (2021). URL <https://doi.org/10.7554/eLife.64449>. Publisher: eLife Sciences Publications, Ltd.
- 803 [9] Hebb, D. O. The semiautonomous process: Its nature and nurture. *American
804 Psychologist* **18**, 16–27 (1963). Place: US Publisher: American Psychological Association.
- 805 [10] Brock, L. G., Coombs, J. S. & Eccles, J. C. The recording of potentials from motoneurons with an intracellular electrode. *The Journal of Physiology* 806 **117**, 431–460 (1952). URL <https://onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1952.sp004759>.
807 _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1952.sp004759>.
- 808 [11] Marr, D. & Brindley, G. S. Simple memory: a theory for archicortex. *Philosophical
809 Transactions of the Royal Society of London. B, Biological Sciences* **262**, 23–81
810 (1997). URL <https://royalsocietypublishing.org/doi/10.1098/rstb.1971.0078>. Publisher: Royal Society.
- 811 [12] Traub, R. D., Tu, Y. & Whittington, M. A. Cell assembly formation and structure
812 in a piriform cortex model. *Reviews in the Neurosciences* **33**, 111–132 (2022).
- 813 [13] Gastaldi, C., Schwalger, T., Falco, E. D., Quiroga, R. Q. & Gerstner, W. When
814 shared concept cells support associations: Theory of overlapping memory engrams.
815 *PLOS Computational Biology* **17**, e1009691 (2021). URL <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009691>.
816 Publisher: Public Library of Science.
- 817 [14] Holtmaat, A. & Caroni, P. Functional and structural underpinnings of neuronal
818 assembly formation in learning. *Nature Neuroscience* **19**, 1553–1562 (2016).

- 827 [15] Grinvald, A., Arieli, A., Tsodyks, M. & Kenet, T. Neuronal assemblies: Single cortical neurons are obedient members of a huge orchestra. *Biopolymers* **68**, 422–436
828 (2003). URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.10273>.
829 eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/bip.10273>.
- 831 [16] Harris, K. D. Neural signatures of cell assembly organization. *Nature Reviews Neuroscience* **6**, 399–407 (2005). URL <https://www.nature.com/articles/nrn1669>.
832 Number: 5 Publisher: Nature Publishing Group.
- 834 [17] Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G. & Buzsáki, G. Organization of
835 cell assemblies in the hippocampus. *Nature* **424**, 552–556 (2003).
- 836 [18] Kamiński, J. *et al.* Persistently active neurons in human medial frontal and medial
837 temporal lobe support working memory. *Nature Neuroscience* **20**, 590–601 (2017).
- 838 [19] Wilson, M. A. & McNaughton, B. L. Reactivation of hippocampal ensemble mem-
839 ories during sleep. *Science (New York, N.Y.)* **265**, 676–679 (1994).
- 840 [20] Yuste, R., Nelson, D. A., Rubin, W. W. & Katz, L. C. Neuronal domains in
841 developing neocortex: mechanisms of coactivation. *Neuron* **14**, 7–17 (1995).
- 842 [21] Sherafati, A. *et al.* Depth-dependant heterogeneity of neuronal responses in mouse
843 primary visual cortex}. *BioRxiv* (2025).
- 844 [22] Grienberger, C. & Konnerth, A. Imaging Calcium in Neurons. *Neuron* **73**, 862–
845 885 (2012). URL [https://www.cell.com/neuron/abstract/S0896-6273\(12\)00172-9](https://www.cell.com/neuron/abstract/S0896-6273(12)00172-9). Publisher: Elsevier.
- 847 [23] Scheffer, L. K. *et al.* A connectome and analysis of the adult Drosophila central
848 brain. *eLife* **9**, e57443 (2020). URL <https://doi.org/10.7554/eLife.57443>.
849 Publisher: eLife Sciences Publications, Ltd.
- 850 [24] Dorkenwald, S. *et al.* Neuronal wiring diagram of an adult brain. *Nature* **634**, 124–
851 138 (2024). URL <https://www.nature.com/articles/s41586-024-07558-y>.
852 Publisher: Nature Publishing Group.
- 853 [25] Witvliet, D. *et al.* Connectomes across development reveal principles of brain mat-
854 uration. *Nature* **596**, 257–261 (2021). URL <https://www.nature.com/articles/s41586-021-03778-8>. Number: 7871 Publisher: Nature Publishing Group.
- 856 [26] Gour, A. *et al.* Postnatal connectomic development of inhibition in mouse barrel
857 cortex. *Science (New York, N.Y.)* **371**, eabb4534 (2021).
- 858 [27] Bae, J. A. *et al.* Functional connectomics spanning multiple areas of mouse visual
859 cortex. *Nature* **640**, 435–447 (2025). URL <https://www.nature.com/articles/s41586-025-08790-w>. Publisher: Nature Publishing Group.
- 861 [28] Shapson-Coe, A. *et al.* A petavoxel fragment of human cerebral cortex recon-
862 structed at nanoscale resolution. *Science* **384**, eadk4858 (2024). URL <https://www.science.org/doi/10.1126/science.adk4858>. Publisher: American Asso-
863 ciation for the Advancement of Science.
- 865 [29] Bock, D. D. *et al.* Network anatomy and in vivo physiology of visual cortical
866 neurons. *Nature* **471**, 177–182 (2011).
- 867 [30] Lee, W.-C. A. *et al.* Anatomy and function of an excitatory network in the visual
868 cortex. *Nature* **532**, 370–374 (2016). URL <https://www.nature.com/articles/nature17192>. Number: 7599 Publisher: Nature Publishing Group.

- 870 [31] Massey, F. J. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the*
871 *American Statistical Association* **46**, 68–78 (1951). URL <https://www.jstor.org/stable/2280095>. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].
- 874 [32] Tootell, R. B. H. *et al.* Functional analysis of primary visual cortex (V1) in humans.
875 *Proceedings of the National Academy of Sciences* **95**, 811–817 (1998). URL <https://www.pnas.org/doi/full/10.1073/pnas.95.3.811>. Publisher: Proceedings of
876 the National Academy of Sciences.
- 878 [33] Möller, J., Avitan, L. & Goodhill, G. J. Detecting neural assemblies in calcium
879 imaging data. *BMC biology* **16**, 143 (2018).
- 880 [34] Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. & Pfister, H. UpSet: Visu-
881 alization of Intersecting Sets. *IEEE Transactions on Visualization and Computer*
882 *Graphics* **20**, 1983–1992 (2014). URL <https://ieeexplore.ieee.org/document/6876017>. Conference Name: IEEE Transactions on Visualization and Computer
883 Graphics.
- 885 [35] Dorfman, R. A Formula for the Gini Coefficient. *The Review of Economics and*
886 *Statistics* **61**, 146–149 (1979). URL <https://www.jstor.org/stable/1924845>.
887 Publisher: The MIT Press.
- 888 [36] Walker, E. Y. *et al.* Inception loops discover what excites neurons most using deep
889 predictive models. *Nature Neuroscience* **22**, 2060–2065 (2019). URL <https://www.nature.com/articles/s41593-019-0517-x>. Number: 12 Publisher: Nature
890 Publishing Group.
- 892 [37] Green, M. Contrast detection and direction discrimination of drifting gratings.
893 *Vision Research* **23**, 281–289 (1983). URL <https://www.sciencedirect.com/science/article/pii/0042698983901177>.
- 895 [38] Abdel-aziem, A. H. & Soliman, T. H. M. A Multi-Layer Perceptron (MLP) Neu-
896 ral Networks for Stellar Classification: A Review of Methods and Results. *Inter-
897 national Journal of Advances in Applied Computational Intelligence* **Volume 3**,
898 29–37 (2023). URL <https://www.americasp.org/articleinfo/31/show/2002>.
899 Publisher: American Scientific Publishing Group (ASPG).
- 900 [39] Bhatkar, A. P. & Kharat, G. Detection of Diabetic Retinopathy in Retinal Images
901 Using MLP Classifier. In *2015 IEEE International Symposium on Nanoelectronic*
902 *and Information Systems*, 331–335 (2015). URL <https://ieeexplore.ieee.org/abstract/document/7434449>.
- 904 [40] Alon, U. Network motifs: theory and experimental approaches. *Nature Reviews*
905 *Genetics* **8**, 450–461 (2007). URL <https://www.nature.com/articles/nrg2102>.
906 Number: 6 Publisher: Nature Publishing Group.
- 907 [41] Sporns, O. & Kötter, R. Motifs in Brain Networks. *PLoS Biology* **2**, e369 (2004).
908 URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC524253/>.
- 909 [42] Hu, Y., Trousdale, J., Josić, K. & Shea-Brown, E. Motif statistics and spike cor-
910 relations in neuronal networks. *Journal of Statistical Mechanics: Theory and Ex-*
911 *periment* **2013**, P03012 (2013). URL <https://dx.doi.org/10.1088/1742-5468/2013/03/P03012>. Publisher: IOP Publishing and SISSA.
- 913 [43] Hebb, D. O. Intelligence, Brain Function and The Theory of Mind. *Brain* **82**,
914 260–275 (1959). URL <https://doi.org/10.1093/brain/82.2.260>.

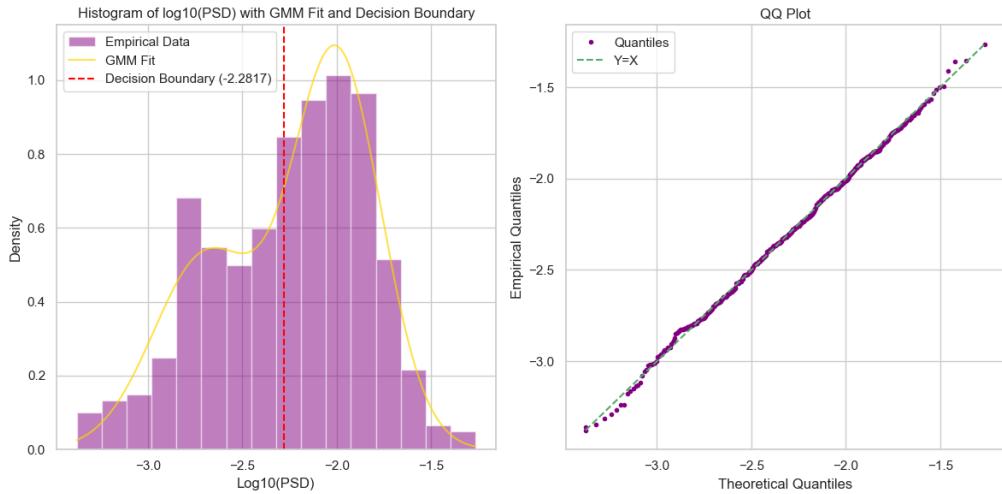
- 915 [44] Holtmaat, A. & Svoboda, K. Experience-dependent structural synaptic plasticity in the mammalian brain. *Nature Reviews Neuroscience* **10**, 647–658 (2009).
916 URL <https://www.nature.com/articles/nrn2699>. Number: 9 Publisher: Nature
917 Publishing Group.
- 919 [45] Mongillo, G., Rumpel, S. & Loewenstein, Y. Intrinsic volatility of synaptic connec-
920 tions — a challenge to the synaptic trace theory of memory. *Current Opinion in*
921 *Neurobiology* **46**, 7–13 (2017). URL <https://www.sciencedirect.com/science/article/pii/S0959438817300673>.
- 923 [46] Aitken, K., Garrett, M., Olsen, S. & Mihalas, S. The geometry of representa-
924 tional drift in natural and artificial neural networks. *PLOS Computational Bi-
925 ology* **18**, e1010716 (2022). URL [https://journals.plos.org/ploscompbiol/
926 article?id=10.1371/journal.pcbi.1010716](https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010716). Publisher: Public Library of Sci-
927 ence.
- 928 [47] Schoonover, C. E., Ohashi, S. N., Axel, R. & Fink, A. J. P. Representational drift
929 in primary olfactory cortex. *Nature* **594**, 541–546 (2021). URL <https://www.nature.com/articles/s41586-021-03628-7>. Number: 7864 Publisher: Nature
930 Publishing Group.
- 932 [48] Olshausen, B. A. & Field, D. J. Sparse coding of sensory inputs. *Current Opin-
933 ion in Neurobiology* **14**, 481–487 (2004). URL [https://www.sciencedirect.com/
934 science/article/pii/S0959438804001035](https://www.sciencedirect.com/science/article/pii/S0959438804001035).
- 935 [49] Okun, M. *et al.* Diverse coupling of neurons to populations in sensory cortex. *Nature*
936 **521**, 511–515 (2015). URL <https://www.nature.com/articles/nature14273>.
937 Publisher: Nature Publishing Group.
- 938 [50] Yuste, R., Cossart, R. & Yaksi, E. Neuronal ensembles: Building blocks of neural
939 circuits. *Neuron* S0896–6273(23)00967–4 (2024).
- 940 [51] Vogels, T. P. & Abbott, L. F. Gating multiple signals through detailed balance
941 of excitation and inhibition in spiking networks. *Nature Neuroscience* **12**, 483–491
942 (2009). URL <https://www.nature.com/articles/nn.2276>. Publisher: Nature
943 Publishing Group.
- 944 [52] Vogels, T. P., Sprakler, H., Zenke, F., Clopath, C. & Gerstner, W. Inhibitory
945 Plasticity Balances Excitation and Inhibition in Sensory Pathways and Memory
946 Networks. *Science* **334**, 1569–1573 (2011). URL [https://www.science.org/doi/
947 full/10.1126/science.1211095](https://www.science.org/doi/full/10.1126/science.1211095). Publisher: American Association for the Ad-
948 vancement of Science.
- 949 [53] Xu, S. *et al.* Behavioral state coding by molecularly defined paraventricular hy-
950 pothalamic cell type ensembles. *Science* **370**, eabb2494 (2020). URL <https://www.science.org/doi/10.1126/science.abb2494>. Publisher: American As-
951 sociation for the Advancement of Science.
- 953 [54] de Vries, S. E. J. *et al.* A large-scale standardized physiological survey reveals
954 functional organization of the mouse visual cortex. *Nature Neuroscience* **23**, 138–
955 151 (2020). URL <https://www.nature.com/articles/s41593-019-0550-9>. Pub-
956 lisher: Nature Publishing Group.
- 957 [55] Hua, Y., Laserstein, P. & Helmstaedter, M. Large-volume en-bloc staining for elec-
958 tron microscopy-based connectomics. *Nature Communications* **6**, 7923 (2015). URL
959 <https://www.nature.com/articles/ncomms8923>. Publisher: Nature Publishing
960 Group.

- 961 [56] Yin, W. *et al.* A petascale automated imaging pipeline for mapping neuronal circuits with high-throughput transmission electron microscopy. *Nature Communications* **11**, 4949 (2020). URL <https://www.nature.com/articles/s41467-020-18659-3>. Publisher: Nature Publishing Group.
- 962
963
964
- 965 [57] Macrina, T. *et al.* Petascale neural circuit reconstruction: automated methods (2021). URL <https://www.biorxiv.org/content/10.1101/2021.08.04.455162v1>. Pages: 2021.08.04.455162 Section: New Results.
- 966
967
- 968 [58] Phelps, J. S. *et al.* Reconstruction of motor control circuits in adult Drosophila using automated transmission electron microscopy. *Cell* **184**, 759–774.e18 (2021). URL [https://www.cell.com/cell/abstract/S0092-8674\(20\)31683-4](https://www.cell.com/cell/abstract/S0092-8674(20)31683-4). Publisher: Elsevier.
- 969
970
971
- 972 [59] Mahalingam, G. *et al.* A scalable and modular automated pipeline for stitching of large electron microscopy datasets. *eLife* **11**, e76534 (2022). URL <https://doi.org/10.7554/eLife.76534>. Publisher: eLife Sciences Publications, Ltd.
- 973
974
- 975 [60] Schneider-Mizell, C. M. *et al.* Inhibitory specificity from a connectomic census of mouse visual cortex. *Nature* **640**, 448–458 (2025). URL <https://www.nature.com/articles/s41586-024-07780-8>. Publisher: Nature Publishing Group.
- 976
977
- 978 [61] Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682 (2012). URL <https://www.nature.com/articles/nmeth.2019>. Publisher: Nature Publishing Group.
- 979
980
- 981 [62] Bogovic, J. A., Hanslovsky, P., Wong, A. & Saalfeld, S. Robust registration of calcium images by learned contrast synthesis. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 1123–1126 (2016). URL <https://ieeexplore.ieee.org/document/7493463>. ISSN: 1945-8452.
- 982
983
984
- 985 [63] Avitan, L. *et al.* Spontaneous Activity in the Zebrafish Tectum Reorganizes over Development and Is Influenced by Visual Experience. *Current Biology* **27**, 2407–2419.e4 (2017). URL [https://www.cell.com/current-biology/abstract/S0960-9822\(17\)30793-5](https://www.cell.com/current-biology/abstract/S0960-9822(17)30793-5). Publisher: Elsevier.
- 986
987
988
- 989 [64] Almeida-Filho, D. G. *et al.* An investigation of Hebbian phase sequences as assembly graphs. *Frontiers in Neural Circuits* **8**, 34 (2014). URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3986516/>.
- 990
991
- 992 [65] Kursin, A., Húsek, D. & Neruda, R. Faster Learning with Overlapping Neural Assemblies. In Kolllias, S. D., Stafylopatis, A., Duch, W. & Oja, E. (eds.) *Artificial Neural Networks – ICANN 2006*, Lecture Notes in Computer Science, 226–233 (Springer, Berlin, Heidelberg, 2006).
- 993
994
995
- 996 [66] Möller, J. & Goodhill, G. J. Detecting Neural Assemblies Through Similarity Graph Clustering. In Carrillo-Reid, L. (ed.) *Identification, Characterization, and Manipulation of Neuronal Ensembles*, 167–176 (Springer US, New York, NY, 2025). URL https://doi.org/10.1007/978-1-0716-4208-5_7.
- 997
998
999
- 1000 [67] Miller, R. G. A Trustworthy Jackknife. *The Annals of Mathematical Statistics* **35**, 1594–1605 (1964). URL <https://www.jstor.org/stable/2238296>. Publisher: Institute of Mathematical Statistics.
- 1001
1002
- 1003 [68] Richter, C. G., Thompson, W. H., Bosman, C. A. & Fries, P. A jackknife approach to quantifying single-trial correlation between covariance-based metrics undefined on a single-trial basis. *NeuroImage* **114**, 57–70 (2015). URL <https://www.sciencedirect.com/science/article/pii/S1053811915003316>.
- 1004
1005
1006

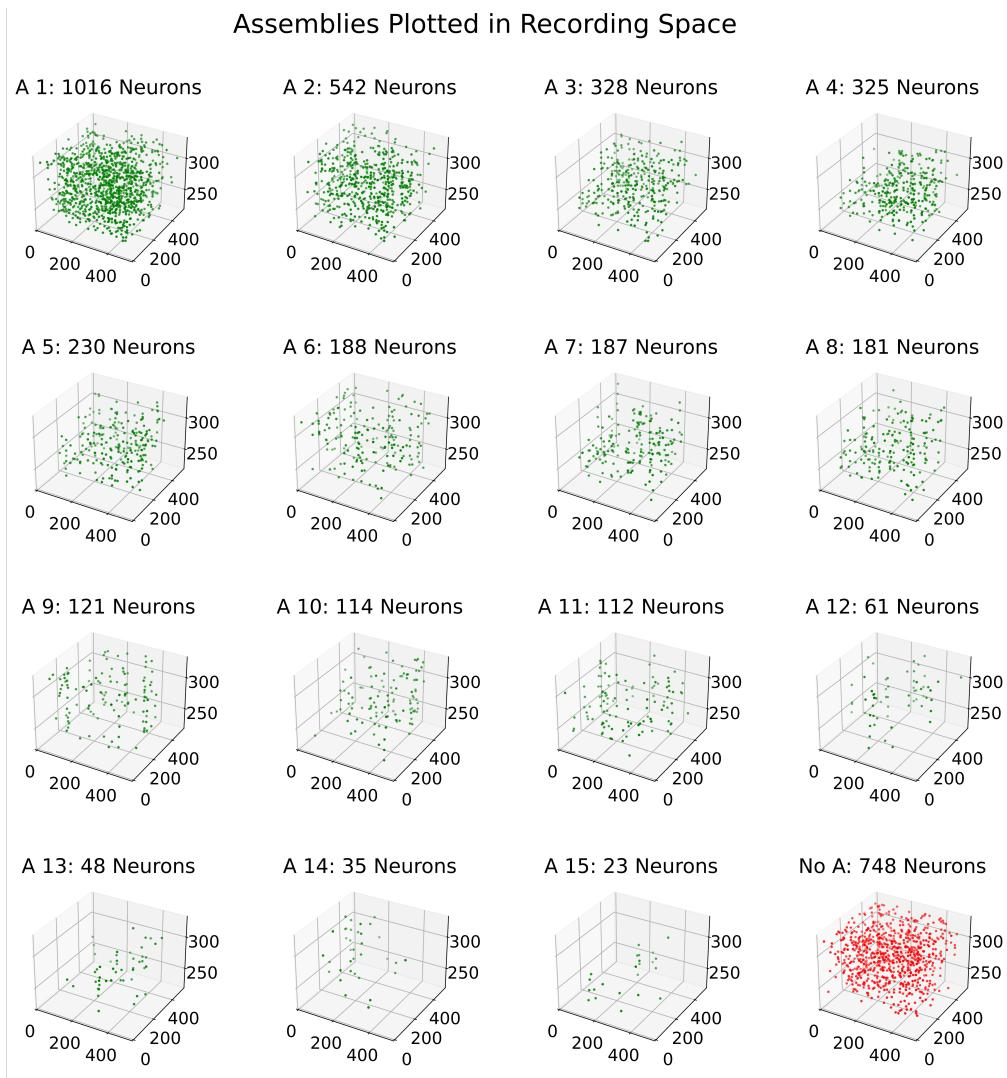
- 1007 [69] Hurley, N. & Rickard, S. Comparing Measures of Sparsity. *IEEE Transactions on*
 1008 *Information Theory* **55**, 4723–4741 (2009). URL <https://ieeexplore.ieee.org/abstract/document/5238742>. Conference Name: IEEE Transactions on Information
 1009 Theory.
- 1011 [70] Khosravy, M., Nitta, N., Gupta, N., Patel, N. & Babaguchi, N. Chapter 3 -
 1012 A descriptive review to sparsity measures. In Khosravy, M., Dey, N. & Duque,
 1013 C. A. (eds.) *Compressive Sensing in Healthcare*, Advances in ubiquitous sensing
 1014 applications for healthcare, 43–63 (Academic Press, 2020). URL <https://www.sciencedirect.com/science/article/pii/B9780128212479000081>.
- 1016 [71] Wright Muelas, M., Mughal, F., O'Hagan, S., Day, P. J. & Kell, D. B. The
 1017 role and robustness of the Gini coefficient as an unbiased tool for the selection
 1018 of Gini genes for normalising expression profiling data. *Scientific Reports* **9**, 17960
 1019 (2019). URL <https://www.nature.com/articles/s41598-019-54288-7>. Number:
 1020 1 Publisher: Nature Publishing Group.
- 1021 [72] Hetherington, P. A., & Shapiro, M. L. Simulating Hebb cell assemblies: the neces-
 1022 sity for partitioned dendritic trees and a post-not-pre LTD rule. *Network: Com-
 1023 putation in Neural Systems* **4**, 135–153 (1993). URL https://www.tandfonline.com/doi/abs/10.1088/0954-898X_4_2_001. Publisher: Taylor & Francis.
- 1025 [73] Garagnani, M., Wennekers, T. & Pulvermüller, F. Recruitment and Consolidation
 1026 of Cell Assemblies for Words by Way of Hebbian Learning and Competition in
 1027 a Multi-Layer Neural Network. *Cognitive Computation* **1**, 160–176 (2009). URL
 1028 <https://doi.org/10.1007/s12559-009-9011-1>.
- 1029 [74] Fransen, E., Lansner, A. & Liljenström, H. A Model of Cortical Associative Memory
 1030 Based on Hebbian Cell Assemblies. In Eeckman, F. H. & Bower, J. M. (eds.)
 1031 *Computation and Neural Systems*, 431–435 (Springer US, Boston, MA, 1993). URL
 1032 https://doi.org/10.1007/978-1-4615-3254-5_65.
- 1033 [75] Dorkenwald, S. *et al.* Binary and analog variation of synapses between cortical
 1034 pyramidal neurons. *eLife* **11**, e76120 (2022). URL <https://doi.org/10.7554/eLife.76120>. Publisher: eLife Sciences Publications, Ltd.
- 1036 [76] Bavelas, A. A Mathematical Model for Group Structures. *Human Organization* **7**,
 1037 16–30 (2009). URL <https://doi.org/10.17730/humo.7.3.f4033344851g1053>.
- 1038 [77] Freeman, L. C. A Set of Measures of Centrality Based on Betweenness. *Sociometry*
 1039 **40**, 35–41 (1977). URL <https://www.jstor.org/stable/3033543>. Publisher:
 1040 [American Sociological Association, Sage Publications, Inc.].
- 1041 [78] Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).
 1042 URL <https://www.nature.com/articles/s41586-020-2649-2>. Publisher: Na-
 1043 ture Publishing Group.
- 1044 [79] McKinney, W. Data Structures for Statistical Computing in Python. *scipy* (2010).
 1045 URL <https://proceedings.scipy.org/articles/Majora-92bf1922-00a>.
- 1046 [80] Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine*
 1047 *Learning Research* **12**, 2825–2830 (2011). URL <http://jmlr.org/papers/v12/pedregosa11a.html>.
- 1049 [81] Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring Network Structure, Dy-
 1050 namics, and Function using NetworkX. *scipy* (2008). URL <https://proceedings.scipy.org/articles/TCWV9851>.

- 1052 [82] Matelsky, J. K. *et al.* DotMotif: an open-source tool for connectome subgraph
1053 isomorphism search and graph queries. *Scientific Reports* **11**, 13045 (2021). URL
1054 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8219732/>.
- 1055 [83] Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science
1056 & Engineering* **9**, 90–95 (2007). URL <https://ieeexplore.ieee.org/document/4160265>. Conference Name: Computing in Science & Engineering.
- 1058 [84] Waskom, M. L. seaborn: statistical data visualization. *Journal of Open Source Software* **6**, 3021 (2021). URL <https://joss.theoj.org/papers/10.21105/joss.03021>.
- 1061 [85] Micah, A. *et al.* Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Research* (2021). URL <https://wellcomeopenresearch.org/articles/4-63/v2>.
- 1062
1063

1064 **9 Supplemental Figures**

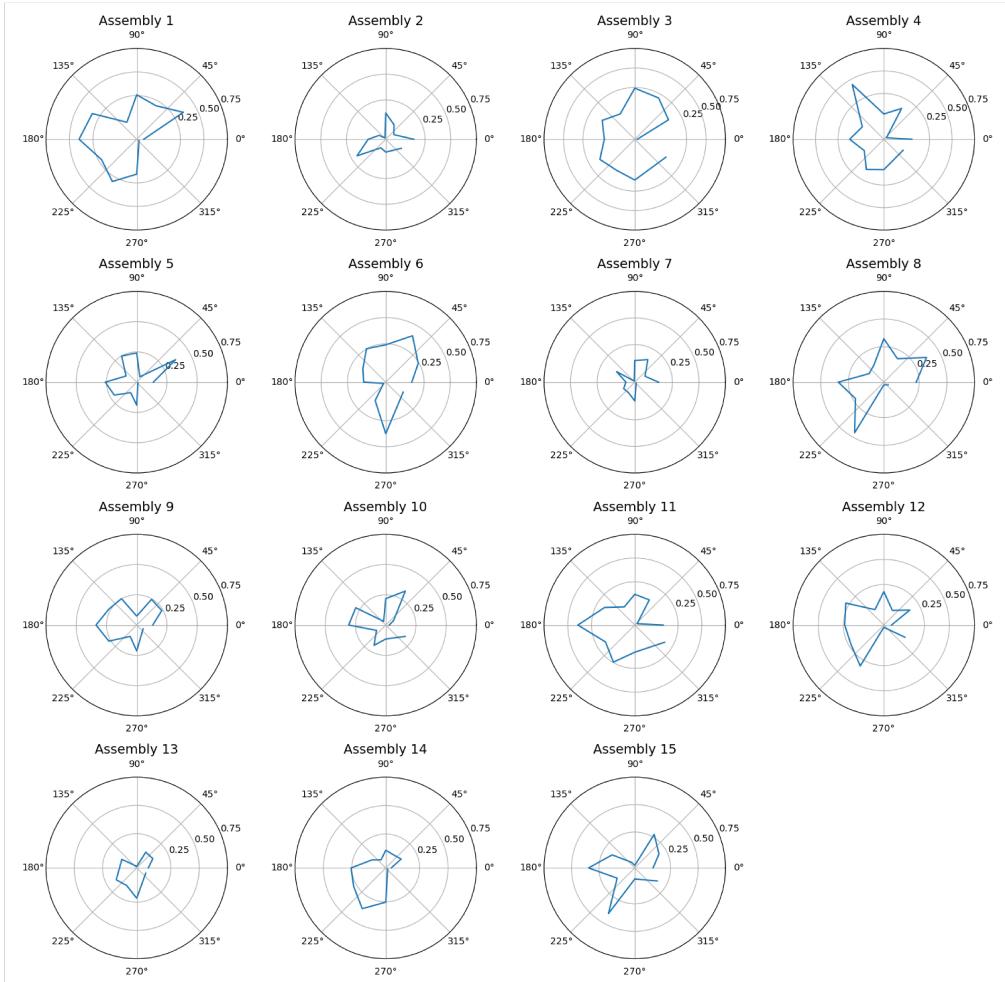


Supplementary figure 1. Histogram of connection strengths (log10-scaled) with GMM fit overlaid, and QQ plot evaluating the fit. The red dashed line indicates the decision boundary separating the two Gaussian components, used to classify “tail” connections.



Supplemental Figure 2. Individual Assemblies Plotted in 2PCI 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 in the 2PCI. 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, one thousand nine hundred and six neurons were assigned to assemblies and seven hundred forty-eight were not. Assemblies are typically spatially spread out throughout the recording field. This supplemental figure is meant to complement the visualization in Figure 5.

Assembly Oracle Score as a function of Fullscreen Grating Orientation



Supplemental Figure 3. 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 are represented by a polar plot. Reliability is measured through the Oracle score metric. The scores of assembly co-activity trace in response to gratings are typically lower than those seen in natural movies (Fig. 8), but the results are still indicative of tuning properties in these functional populatinos which may be a product of the shared tuning of their individual neural components. Notably, some assemblies trace response seem to be highly reliable to particular orientations, similar to the orientation receptive fields of simple cells in the primary visual cortex.