Spatial transcriptomic data reveals pure cell types via the mosaic hypothesis

Yiliu Wang [1], **Christof Koch** [1,2], **and Uygar Sümbül** [1]

[1]Allen Institute, 615 Westlake Ave N, Seattle, WA, USA
[2]Tiny Blue Dot Foundation, Santa Monica, CA, USA

Neurons display remarkable diversity in their anatomical, molecular, and physiological properties. Although observed stereotypy in subsets of neurons is a pillar of neuroscience, clustering in high-dimensional feature spaces, such as those defined by single cell RNA-seq data, is often inconclusive and cells seemingly occupy continuous, rather than discrete, regions. In the retina, a layered structure, neurons of the same discrete type avoid spatial proximity with each other. While this principle, which is independent of clustering in feature space, has been a gold standard for retinal cell types, its applicability to the cortex has been only sparsely explored. Here, we provide evidence for such a mosaic hypothesis by developing a statistical point process analysis framework for spatial transcriptomic data. We demonstrate spatial avoidance across many excitatory and inhibitory neuronal types. Spatial avoidance disappears when cell types are merged, potentially offering a gold standard metric for evaluating the purity of putative cell types.

Keywords: cortical cell types, neuronal diversity, tiling, mosaic, mosaic hypothesis, cortex

Introduction

Anatomical stereotypy can facilitate the formation of canonical circuits. In the mammalian retina, a part of the central nervous system with a laminar structure like neocortex, different patches of the visual field implement the same transformations to achieve visual equivariance. Hence, neurons of a type should be roughly equally distributed across any central patch of the retina. This requires cells of the same type to *tile* the retinal surface by avoiding being too close to each other. This insight was confirmed by seminal studies,[1, 2, 3] and now forms a gold standard method of defining pure retinal cell types.[4] To the extent that neocortical canonical circuits (i.e., connectivity motifs repeating across the neocortex) exist, the same argument requires that cells of the same type should be approximately evenly distributed across the cortical sheet. Indeed, Francis Crick used this principle[5] to hypothesize the existence of on the order of 1,000 cell types in each cortical area (see also[6]).

Tiling refers to the approximately non-overlapping nature of the convex hulls of the dendritic arbors of neurons of the same type, like tiles on a bathroom floor. *Mosaic* refers to overlap of dendritic arbors but with cell bodies of the same type avoiding coming too close to each other. In the retina, neurons closer to the photoreceptors tile the surface whereas ganglion cells form a "mosaic".[4] Thus, the coverage factor of a type, which is the average number of neurons innervating a patch, is close to one at the level of the photoreceptors but increases deeper into the retinal circuit. When cell bodies are represented by points in space, both tiling and mosaic patterns bestow upon the resulting point set an organization different from complete spatial randomness (CSR) (Figure 1).

Retinal studies demonstrate that cell death, tangential migration, and dendritic avoidance contribute to the patterning of the spatial constellation.[7] Self-recognition and self-avoidance mechanisms help to regulate neuronal morphology by enabling recognition of sister branches of the same neuron via contact-dependent molecular recognition.[8, 9, 10, 11] These mechanisms can produce 3-dimensional signatures in the cortex even when the primary effect concerns 2-dimensional positioning in a layer because cortical neurons, unlike retinal ones, do not stratify to occupy thin layers within the cortex.

Astrocytes tile neocortical tissue by forming exclusive territories following migration,[12, 13, 14, 15] but less is known about the spatial organization of neurons. Maruoka *et al.*[16] showed microcolumnar organization of layer 5 intra- and extra-telencephalic neurons, demonstrating tangential repulsion of cells in these subclasses. The opposite effect, spatial clustering instead of avoidance, was observed in parvalbumin and somatostatin-expressing neurons.[17, 18] Importantly, these studies concern broad subclasses and not their subsets forming (pure) cell types – subpopulations whose members demonstrate similar phenotypes across all observation modalities.[19, 4, 20] Close approximations to this idealized definition exist in the retina[21] and the cortex.[22]

While single-cell transcriptomic technology has put forth multiple, putative pure types within those broad subclasses, their spatial organization, or lack thereof (i.e., spatial randomness), remains unknown. A confounding factor is that clustering in high-dimensional feature spaces is error-prone due to the "curse of dimensionality," sampling bias, and complicated noise processes, in addition to the ill-posed nature of the clustering objective.[23] Here, we study a recent whole-brain spatial transcriptomics dataset where over 1,100 genes are profiled in space and a transcriptomic identity (based on reference clustering of single-cell RNA sequencing data[24]) as well as a location within the common coordinate framework (CCF) are assigned to approximately 10 million cells in the mouse[25] (Figure 1). While this represents a monumental achievement, neither the initial clustering nor the identity assignment steps are immune to the problems listed above. Moreover, 2-d (instead of 3-d) imaging of brain sections further complicates the analysis of spatial distribution.

We develop a stochastic point process analysis framework to study apparent pairwise short-range (e.g., nearest neighbor) interactions between neurons. Focusing on cell types whose local organization can be studied with statistical rigor, we reveal self-avoidance of cell bodies within pure cell types and its loss in mixtures of cell types. Our findings elucidate a long-standing hypothesis on the organization of cortical circuitry and outline a method to evaluate and validate the purity of putative cell types.

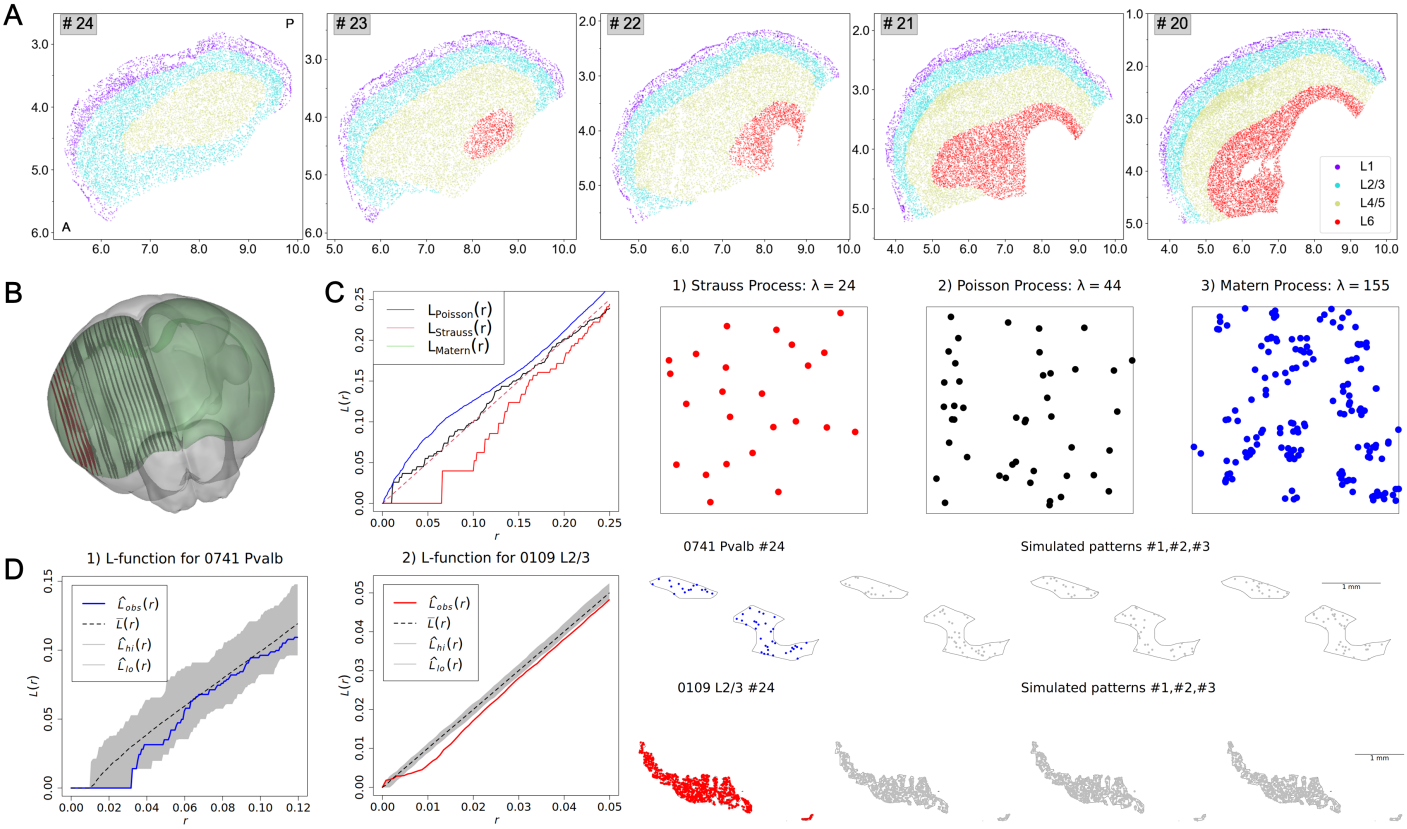


Figure 1. Point process analysis of spatial transcriptomic data with cell type annotations reveals spatial organization of neuronal cell types. A. Annotated, detected cells in lateral sections of the mouse brain used in this study, obtained from.[25] Each dot corresponds to a cell body and the depicted cortical layer annotations suggest the extent of the relevant surfaces for 2-d point process analysis. Top left: brain section number from.[25] A: anterior, P: posterior. **B.** 3D visualization of locations of the brain sections shown in (A) (red) and unused (grey) brain sections in the registered CCF[26, 27] brain. **C.** Illustration of the L-function for three different point processes with repulsion (L_{Strauss} in red), spatial randomness (L_{Poisson} in black), and clustering (L_{Matern} in blue). **D.** Illustration of the hypothesis testing procedure for two exemplar GABAergic (0741 Pvalb, $n = 50$) and Glutamatergic (0109 L2/3, $n = 743$) clusters. Left: L-functions of the observed patterns ($\hat{L}_{\text{obs}}$ in color) and their statistics for the corresponding 199 simulations ($\bar{L}, \hat{L}_{hi}, \hat{L}_{lo}$ are averaged with their upper and lower envelopes[28] respectively). See Table 1 for n-values. Center: observed point patterns in section 24 for the two cell types. Right: examplar simulated patterns for the two types. All coordinates are in mm.

Results

Selecting cell types for point process analysis

Three main sources of noise impede point process analysis of spatial transcriptomic datasets. First, reference cell type labels may not be accurate ("label noise"), which is the root cause of confusion in cell types research. This can be due to inaccuracies in initial single-cell RNA sequencing (scRNA-seq) based clustering,[24] incorrect mapping of spatial transcriptomic observations to those clusters,[25] or noise in image-based assignment of gene expression to cells. Second, imperfections in segmentation of cells and profiling their genes can lead to a decrease in the number of annotated cells, acting as another noise source that decreases the apparent intensities of the underlying point processes. We formally prove that label noise and thinning of the annotations bring the underlying process closer to spatial randomness (Theorem 1, Methods). Third, a single 2-d projection image is obtained from a slice of tissue that is $10\mu m$ thick. Thus, the distance between a pair of neurons may appear arbitrarily small if they are separated mostly along the slicing axis.

To account for these noise sources, we implement two filtering steps. First, we calculate the relative abundance of high confidence cells in each type, which are defined as cells whose reference mapping confidence exceeds a threshold,[25, 24] and exclude those cell types with low confidence ratios to filter against label noise ("label noise filter").

Our second filter is motivated by the fact that a test may fail to detect self-avoidance (*i.e.*, reject the null hypothesis of spatial randomness) if the process has low intensity. Concretely, if only a few points exist in a large area, random placement creates patterns, with high probability, without any points close to each other. To understand the effect of the prevalence of a cell type on the ability to reject the null hypothesis, we simulated 2-d point processes with different densities. We infer the intensity threshold below which the power of our chosen statistical test is insufficient to identify self-avoidance (Methods). For each type, we only consider brain sections with sufficient number and intensity (average density of cells) of high confidence cells to act as a filter against low intensity patterns ("minimal intensity filter").

Statistical testing for spatial organization of cell bodies

To analyze the local organization in the point pattern of cell bodies in 2-d brain sections, we first identify a set of regions tightly covering the points belonging to the transcriptionally defined cell type of interest, to avoid falsely assigning significance to global patterns of cell abundance. We calculate the average normalized abundance of points from a given point across these regions as a function of distance from this point, known as the L-function[29] (Fig. 1C,D).

For each cell type, we simulate the null hypothesis of a CSR process with the same number of points on the same regions multiple times. Noting that the L-function, expressing normalized intensity as a function of distance (Methods), of a Poisson process concentrates around the identity function (i.e., $L(x) = x$ as in Fig. 1C), we use the Diggle-Cressie-Loosmore-Ford (DCLF) test[29, 30] to measure the significance of one-sided deviations from CSR based on the area between the measured and simulated L-functions. While a deviation below the $L(x) = x$ line at x_0 implies avoidance of neighbors up to a distance x_0, a deviation above it implies spatial clustering (Fig. 1C).

The CSR null hypothesis can be trivially rejected due to the fact that cell bodies are non-penetrating bodies of non-zero size so that points cannot be arbitrarily close to each other in 3-d. We address this issue by applying the DCLF test only across a range of interactions that excludes this shortest scale that we term the *hardcore distance*. We obtain an unbiased distribution of soma sizes from a volume electron microscopy dataset[31](Figure S1). Moreover, we show that the signature of the hardcore separation between GABAergic cells can extend beyond the hardcore distance, but the projection onto two dimensions effectively destroys the hardcore separation for the Glutamatergic subpopulations, which are more numerous (Figure 1D). Therefore, for GABAergic types, we generate the benchmark 2-d simulations from a process with the finite minimal distance, known as the Strauss hardcore process, instead of a canonical Poisson point process (Methods).

We take advantage of the fact that multiple brain sections can qualify through the noise filtering steps for a given cell type by applying a meta analysis of the DCLF test results using Fisher's method.[32] Thus, we formally test the null hypothesis that the point pattern formed by cells display spatial randomness in *all* eligible brain sections (Figure 2, blue disks).

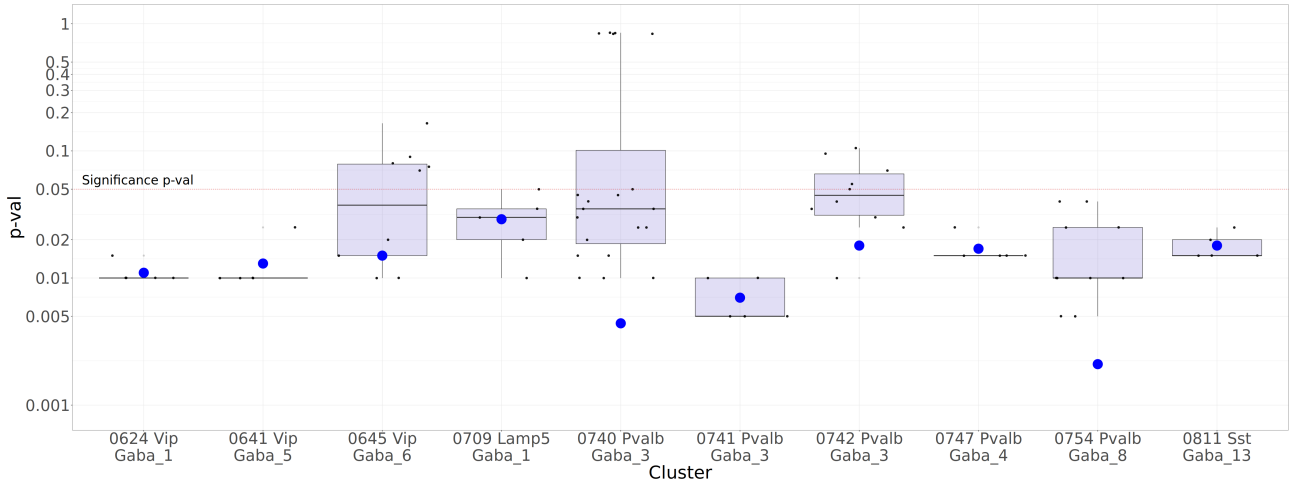


Figure 2. Statistical testing rejects the null hypothesis of complete spatial randomness (CSR) for all ten GABAergic types passing the filtering steps, at a significance level of 0.05 (1-sided DCLF test). Small, black dots indicate p-values over individual sections for the transcriptomic types. The test is repeated over multiple brain sections using multiple Monte-Carlo simulations with 5 different seeds. Large, blue dots indicate the aggregated p-values calculated according to Fisher's method. All aggregated p-values remain significant following correction for multiple testing at a false discovery rate of 0.05, using the Benjamini-Hochberg procedure (Supplementary Table S1 and Table S2). See Table 1 for n-values.

Cells of the same transcriptomic type display self-avoidance in neocortex

Our selection procedure picks 28 neuronal cell types across 5 sagittal brain sections (out of 23; Fig. 1B) for formal statistical testing. Table 1 displays the results of the DCLF test on these cell types, corrected for multiple testing by the Benjamini-Hochberg procedure[33] (this controls the expected proportion of false discoveries within all tested cell types, different from the Fisher's method that performs a meta-analysis over the different brain sections).

To study the noise processes introduced earlier, we also calculate three statistics for each cell type. (i) The average percentile of nearest neighbor distances that are $< 10\mu m$ relates to the severity of distortions due to observing a 3-D brain section projecting onto a 2-D plane. (ii) The average intensity across qualified sections is informative concerning cell types for which the DCLF test may fail to detect self-avoidance due to either a low coverage factor or missing cells. (iii) Lastly, the classification confidence refers to the average confidence score of cell type label by the reference study.[25] A lower average classification confidence implies that the cluster may suffer more from label noise. We also report the sections, the total hull area and the total number of points (summed over qualified sections) of the underlying point pattern, and the upper limit of the interaction range we use to conduct the DCLF test. Self-avoidance behavior, when it exists, should be observable over short distances for pairwise interactions. Therefore, we define the upper limit as the median nearest neighborhood distance across sections where the cluster passes the noise and minimum intensity filters.

Glutamatergic clusters. 18 excitatory neuronal clusters pass our filtering steps across 40 sections (Table 1). Of these, the null hypothesis of CSR is rejected in 16 clusters, in favor of self-avoidance at a significance value of $p = 0.05$. We note that the two clusters that fail to reach this level of statistical significance (0453 L6 CT CTX Glut 5 and 0004 CLA-EPd-CTX Car3 Glut 1) have relatively low intensities in their selected regions.

Considering that ∼11,500 non-overlapping disks with $\sim 10\mu m$ diameter can fit in a $1mm^2$ square,[34] the intensities listed in Table 1, peaking at 1060 cells/mm^2, suggests that the self-avoidance interaction ranges of all clusters are greater than the hardcore separation due to the soma size; that is, the interaction range of all clusters for which CSR is rejected is greater than the hardcore separation due to non-penetrating cell bodies.

GABAergic clusters. Table 1 includes the results of the same analysis for the 10 GABAergic clusters that pass the noise filtering step across 20 sections. Complete spatial randomness (CSR) can be rejected for all 10 (at $p = 0.05$): indeed, they all show self-avoidance (Fig. 2).

Since there are fewer GABAergic than Glutamatergic cells, the noise introduced by 2d projection imaging of the $10\mu m$-thick brain sections, which can distort the interpoint distances, is less of a concern for point process analysis of GABAergic clusters (Table 1). On the other hand, this makes them more vulnerable to thinning of the population (e.g., due to exclusion of cell annotations failing quality control). As discussed above, the power of the statistical test in identifying patterns diminishes when the underlying intensity decreases.

Cluster	Section	p-value	Area(mm^2)	R(μm)	%	λ(1/mm^2)	conf	n
0109 L2/3 IT CTX Glut_2	20–24	7.5e-08	4.04	18	8.53	1053	0.88	4261
0074 L4/5 IT CTX Glut_1	20–24	1.1e-06	1.52	27	3.96	476	0.75	649
0079 L4/5 IT CTX Glut_2	20,21,24	4.1e-05	1.27	16	11.68	1060	0.74	1404
0455 L6 CT CTX Glut_5	20,21	3.9e-04	0.61	24	2.68	506	0.77	332
0116 L2/3 IT CTX Glut_3	21,23,24	1.1e-03	0.30	20	9.59	816	0.78	268
0092 L4/5 IT CTX Glut_4	23,24	3.0e-03	0.38	50	0.00	243	0.71	88
0061 L5 IT CTX Glut_3	20	5.0e-03	0.44	38	0.00	271	0.78	120
0114 L2/3 IT CTX Glut_3	24	5.0e-03	0.10	29	3.08	643	0.79	65
0443 L6 CT CTX Glut_2	21	5.0e-03	0.31	29	1.40	456	0.77	143
0086 L4/5 IT CTX Glut_3	23,24	6.6e-03	0.48	53	0.00	112	0.84	44
0375 L5 ET CTX Glut_5	23	0.011	0.44	29	2.70	340	0.80	148
0026 L5/6 IT TPE-ENT Glut_2	23	0.015	0.44	49	0.00	133	0.70	59
0439 L6 CT CTX Glut_1	21	0.024	0.36	32	9.52	347	0.75	126
0007 IT EP-CLA Glut_1	20	0.027	0.10	28	3.77	518	0.90	53
0359 L5 ET CTX Glut_2	23	0.039	0.10	41	0.00	165	0.83	16
0003 CLA-EPd-CTX Car3 Glut_1	20,21	0.044	1.05	40	9.25	198	0.78	207
0453 L6 CT CTX Glut_5	20	0.061	0.05	22	0.00	275	0.80	15
0004 CLA-EPd-CTX Car3 Glut_1	23, 21	0.14	1.17	51	3.75	103	0.84	122
0754 Pvalb Gaba_8	20, 21	0.002	2.58	92	1.67	43	0.86	100
0740 Pvalb Gaba_3	20,21,23,24	0.004	1.85	59	2.94	84	0.81	158
0741 Pvalb Gaba_3	24	0.007	0.76	68	0.00	66	0.84	50
0624 Vip Gaba_1	24	0.011	0.94	120	0.00	12	0.86	11
0641 Vip Gaba_5	22	0.013	0.23	120	0.00	17	0.72	4
0645 Vip Gaba_6	20,24	0.015	1.79	120	0.00	14	0.82	26
0747 Pvalb Gaba_4	20	0.017	0.76	120	0.00	20	0.82	15
0742 Pvalb Gaba_3	23,24	0.018	1.14	93	0.00	39	0.89	44
0811 Sst Gaba_13	21	0.018	1.06	120	0.00	7	0.74	7
0709 Lamp5 Gaba_1	22	0.029	0.28	82	0.00	40	0.76	11

Table 1. A majority of Glutamatergic clusters and all GABAergic clusters display self-avoidance. Sixteen out of 18 excitatory and all 10 inhibitory cell types display statistically significant self-avoidance. p-values across different brain sections are combined by Fisher's method and corrected for multiple testing. `'p-value'` denotes the fraction of CSR simulations that display as or more extreme self-avoidance according to the one-sided DCLF test. `'Area'` denotes the total hull area of the underlying point pattern. `'R'` is the approximate interaction range that serves as the upper interaction range for the DCLF test. `'%'`: the average percentile of nearest neighbor distances that are $< 10\mu m$, `'$\lambda$'`: the average intensity across qualified sections (i.e., cell density), `'conf'`: the average classification confidence based on the reference study,[25] `'n'`: the total number of cell bodies. See Supplementary Tables S1 and S2 for the p-values for individual cluster-section combinations, and Supplementary Figures S2 and S3 for the pairwise correlation functions for individual clusters.

Spatial clustering of GABAergic neurons at the subclass level

Given that many cell types defined at the finest transcriptional level display self-avoidance (Tables 1), we next study whether this is true at coarser levels of the transcriptional hierarchy[24] as well. Indeed, it has previously been reported that GABAergic neurons belonging to the same subclass (e.g., all Pvalb neurons), show the opposite effect and spatially cluster, possibly to improve the effectiveness of their inhibitory function.[17, 18]

We formally test spatial clustering at the subclass level for Vip, Lamp5, Pvalb, and Sst. For each subclass, we consider sections where at least one cluster belonging to the subclass passes the intensity filter. (We exclude the 0641 Vip type in Section 22 from this analysis because it has too few samples (four) to impact the mixture statistics.) We use the 1-sided DCLF test for spatial clustering (e.g., $L(x) > x$) at the subclass level with the same benchmark process for spatial randomness as before, in the range between the hardcore separation distance and $20\mu m$.[17]

Consistent with previous findings,[17, 18] all inhibitory subclasses cluster at near distances (ca. $20\mu m$) with high statistical significance and are distributed randomly for larger distances (Supplementary Tables S3, S4 and Figure S4).

Segregation of Glutamatergic neurons at the subclass level

While GABAergic neurons cluster at the subclass level, this is not the case for excitatory, Glutamatergic cells. Figures 3A and 3B show the cell distribution of L2/3 IT neurons and Pvalb neurons in sagittal section 24, highlighting the dominant clusters for both excitatory and inhibitory cells that are non-randomly distributed (p=0.05). Glutamatergic types do not mix as well as GABAergic types. We quantify this observation by defining a segregation index for each cell type as the unnormalized k-nearest-neighbor correlation,[35] which is the mean fraction of cells among the k-nearest neighbours having the same cell type label as the reference cell. Hence, an index close to 1 indicates a well-separated cluster while an index close to 0 indicates complete spatial mixing. Cells of the three L2/3 IT types (Fig. 3B) occupy different regions at the subclass level (Segregation index: 0.77), while the segregation is much weaker for Pvalb clusters (Segregation index: 0.40). Indeed, in general, GABAergic cell types mix spatially with each other while Glutamatergic cell types are typically more segregated *within the context of their subclass* (Figure 3C). Thus, the anatomical location of the cluster does not offer a way to verify fine-scale GABAergic types.

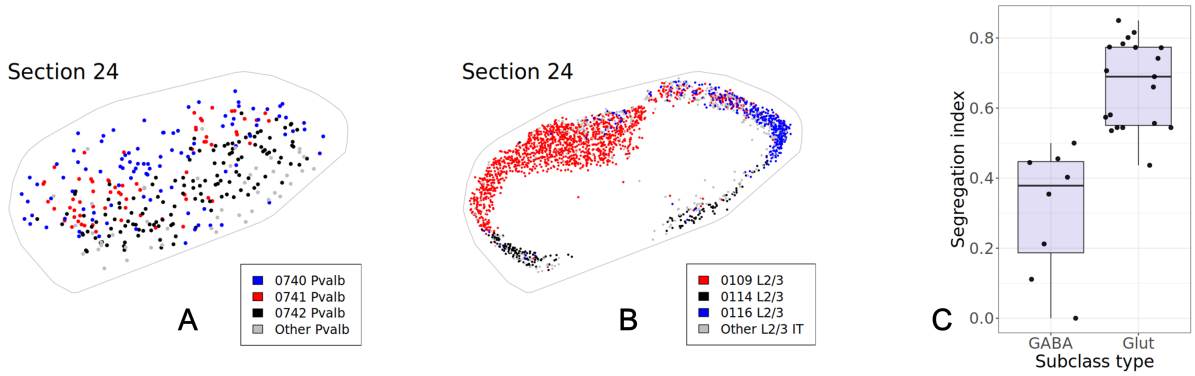


Figure 3. Segregation effect at the subclass level. A. Distribution of Pvalb clusters that mix spatially well with each other. **B.** Distribution of L2/3 IT clusters that predominate separate regions in L2/3. **C.** Segregation indices for Glutamatergic and GABAergic subclasses over the five sagittal brain sections 20 to 24.

Self-avoidance disappears in mixtures of transcriptomically similar cell types

We formally test self-avoidance at the subclass level for the four GABAergic subclasses. We conduct the DCLF test between the hardcore separation distance and the upper interaction range derived for each subclass and find that none of the subclasses display statistically significant self-avoidance for any of the five sections (Table S3).

This result puts forth the interesting possibility that, as in the retina, self-avoidance disappears as one moves up the transcriptomic hierarchy, offering a gold standard test of what defines a cortical neuronal cell type, sidestepping some of the difficulties associated with clustering in high-dimensional feature spaces. To elucidate the transition from statistically significant self-avoidance at the level of transcriptomic types to its lack at the subclass level, we merge groups along the reference transcriptomic hierarchy below the subclass level, referred to as supertypes.

A confounding factor is that many types are not spatially intermixed at the supertype level (e.g., 0741 Pvalb and 0742 Pvalb in Figure 3). An exception is 0740 Pvalb and 0742 Pvalb, which strongly co-exist in two regions (SSs4 and VISC5) in section 23 (Supplementary Table S5): indeed, CSR is rejected for both clusters at a significance level of 0.05 (Table S2); yet merging them into a single group destroys self-avoidance, supporting the hypothesis that they constitute genuine cell types (Supplementary Figure S5b).

We expand this analysis to all supertypes and subclasses in the filtered dataset. For example, 0740 Pvalb, 0741 Pvalb and 0742 Pvalb all belong to the same supertype, 0207 Pvalb. Yet CSR is rejected (at a Fisher-aggregated p-value of 0.13) when applying the same region selection method and hypothesis test to 0207. To facilitate the interpretation, we define the concept of effective cluster count (ECC) as the exponential of the entropy of the mixture pattern, which calculates an information-theoretic estimate for the number of clusters present in the sample pattern (Methods). For a mixture consisting of samples from N clusters, the ECC value has a minimum of 1, denoting the mixture has samples from a single cluster, and a maximum of N, denoting all clusters are *equally* represented in the mixture. The higher the ECC value, the more mixed the sample population is (Figure 4A). For instance, the supertype 0207 Pvalb has an ECC value of 2.84, implying that it is roughly dominated by three clusters.

Figure 4 shows the statistical results and ECC values for all supertypes and subclasses that are parents to the GABAergic types that reject CSR (1st level: subclasses, 2nd level: supertypes, leaf level: clusters). We note that some supertypes have ECC around 1 (0178 Vip, 0212 Pvalb and 0226 Sst). This implies that the supertype is mostly dominated by one cluster in the selected study regions. On the other hand, all subclasses have ECC values much greater than 1. Importantly, as we go up the tree hierarchy, we see an increase in p-values whenever there is an increase in ECC. This suggests a potential rule expressing the purity of a cell type: the test does not reject CSR for any subclass and rejects it for all transcriptomic types in Figure 4. Moreover, CSR is not rejected for supertypes when the corresponding ECC is 2 or higher. The only exception is the supertype '0173 Vip' (DCLF one-sided $p = 0.013$, ECC $= 2.54$). While noise sources thinning the underlying process can explain this exception where the test is performed on only 11 '0173 Vip' cells, an alternative explanation is that the reference hierarchy oversplits a single genuine cell type (Supplementary Figure S6a and S6b).

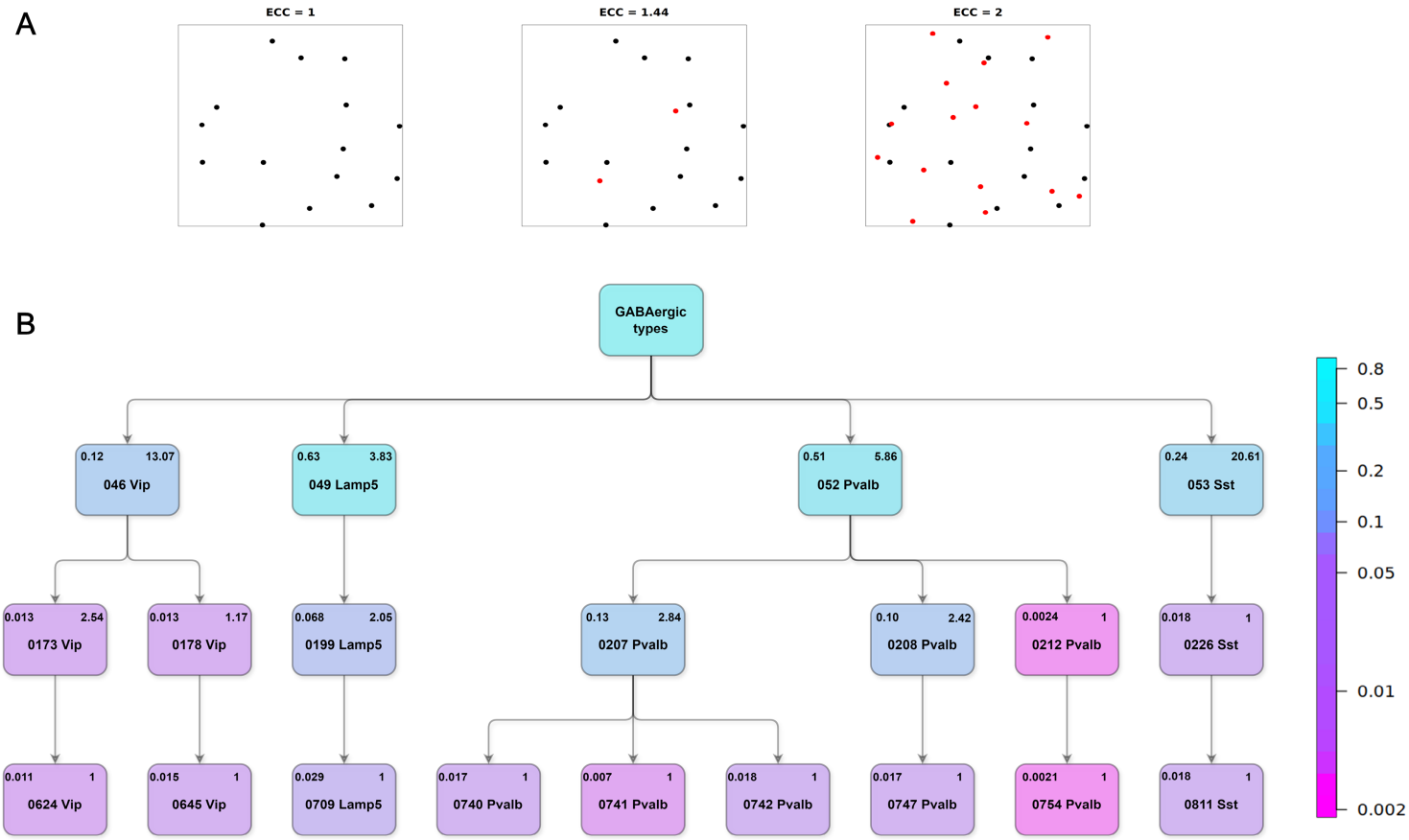


Figure 4. Increase of p-values in mixture types. **A.** Different effective cluster count (ECC) values for a pair of point patterns. The ECC has a minimum of 1 when there is a single type and a maximum of N when N types have the same number of points. In this illustration, $N = 2$. The larger ECC, the more mixed the underlying population is. **B.** Tree hierarchy diagram for GABAergic types. Nodes are color colored according to the p-values of the CSR test, with their numerical values in the top left and their ECC values in the top right corner. When moving up the hierarchy, pure cell types merge into supertypes and subclasses (at the top), and spatial avoidance disappears. The transition is governed by the ECC value of the population.

Discussion

Morphologically defined neuronal cell types formed the basis of Cajal's exposition of the nervous system.[36] While a cell-type-centric view gradually fell out of fashion, the last few decades witnessed a forceful return of this concept. Identifying and accessing cell types has received much attention and funding.[37] Indeed, studying "identical" cells using transcriptional, morphological, developmental, and functional modalities enables operational access and control as well as conceptual leaps in our understanding of nervous system function in health and disease.[38]

However, this promise can only be realized when the underlying cell type definitions are pure and not fragmented or lumped. In this paper, we provided evidence across multiple cell types for the long-standing hypothesis that, like the retina, neocortical neurons of the same type display a mosaic pattern and that this property disappears for mixtures of cell types. This bestows upon the point pattern of cell bodies a distinct statistical signature which can be used to assess the purity of cell types.

Our study mitigates against multiple noise sources and artifacts: (i) While we are primarily interested in imaging a tangential surface (*i.e.*, along a cortical layer), the available datasets cut through this surface with wide gaps between consecutive brain sections. (ii) Our data consists of 2-d images of individual brain sections, with a thickness of $10\mu m$. This dramatically increases throughput; yet the ensuing projection distorts the distances between cells, including the hardcore distances, here given by the diameter of the soma. (iii) If the underlying cell type definitions are inaccurate, the test will not reject the null hypothesis in case of a mixture population or will exclude the cell type due to low intensity when oversplitting. (iv) Similarly, while the cell type definitions may be correct, the mapping from the limited set of spatial transcriptomic genes to type labels may be noisy. While we designed the study to mitigate the effects of these noise sources, their suppression in future datasets will improve accuracy and enable the study of more transcriptomically-defined types. Alternatives, such as volumetric electron microscopy datasets, offer impeccable spatial characterization, but their field of view remains limited and detailed cell type proposals are not yet available for these datasets.

Our findings support the existence of cortical columns that implement a canonical circuit capable of a wide range of computations.[39] This requires the availability of all circuit components (i.e., cell types) in each column. Therefore, efficient placement suggests that cells of the same type avoid each other to maintain a relatively uniform density. This argument has led to proposals that there may be on the order of 1,000 cell types in cortex,[5] in agreement with recent transcriptomic studies.[24] Discovering and verifying the cellular components will be a huge step forward towards revealing such canonical connectivity motifs.[40] Experience constantly modifies the phenotype, overlaying continuous factors of variability onto discrete identities.[41] In addition, the "curse of dimensionality" makes it harder to uncover the discrete factors as the feature dimensionality grows to 10,000 or more of expressed genes or RNA transcripts. The statistical procedure put forth here can verify the purity of a proposed cell type without relying on the features used by the clustering algorithm [42] as it only depends on the theory of point processes in the plane. That is, if two distinct population of cells are truly discrete cell types, then they should display repulsion and violate CSR when treated separately, but not when lumped into a single class, except for clustering at short distances for GABAergic cells.

The extent to which the methodology of this study can be fruitfully applied to subcortical regions that have a nuclear structure, such as the hypothalamus, is an open question. However, the implications of self-avoidance are clearer in the case of the cortical sheet that is spatially extended but relatively thin.

Data and code availability

We study a recent whole-brain spatial transcriptomics dataset by (25). The dataset is publicly available at https://alleninstitute.github.io/abc_atlas_access/. The code for producing figures in the main text and supplementary material is available at https://github.com/yiliuw/mosaics.

Acknowledgements

We wish to thank the founder of the *Allen Institute for Brain Science*, P. G. Allen, for his vision, encouragement and support. The original motivation to pursue this question arose from discussions between Francis Crick and C.K. many decades earlier. We wish to thank Zizhen Yao, Michael Kunst, and Kelly Jin for useful discussions, Forrest Collman and Leila Elabbady for guidance on the MICrONS dataset, Forrest Collman and Stephen Smith for valuable feedback on the manuscript. Y.W. is a *Shanahan Foundation Fellow*.

Bibliography

[1] H. Wässle and H. Riemann, "The mosaic of nerve cells in the mammalian retina," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 200, no. 1141, pp. 441–461, 1978.

[2] S. H. Devries and D. A. Baylor, "Mosaic arrangement of ganglion cell receptive fields in rabbit retina," *Journal of neurophysiology*, vol. 78, no. 4, pp. 2048–2060, 1997.

[3] R. L. Rockhill, T. Euler, and R. H. Masland, "Spatial order within but not between types of retinal neurons," *Proceedings of the National Academy of Sciences*, vol. 97, no. 5, pp. 2303–2307, 2000.

[4] H. S. Seung and U. Sümbül, "Neuronal cell types and connectivity: lessons from the retina," *Neuron*, vol. 83, no. 6, pp. 1262–1272, 2014.

[5] F. Crick, "Cell types in the neocortex: The tiling hypothesis," in *Sydney Brenner Collection*, Wellcome Collection, 1983.

[6] C. F. Stevens, "Neuronal diversity: too many cell types for comfort?," *Current Biology*, vol. 8, no. 20, pp. R708–R710, 1998.

[7] K. T. Nguyen-Ba-Charvet and A. Chédotal, "Development of retinal layers," *Comptes rendus biologies*, vol. 337, no. 3, pp. 153–159, 2014.

[8] B. J. Matthews, M. E. Kim, J. J. Flanagan, D. Hattori, J. C. Clemens, S. L. Zipursky, and W. B. Grueber, "Dendrite self-avoidance is controlled by dscam," *Cell*, vol. 129, no. 3, pp. 593–604, 2007.

[9] J. L. Lefebvre, D. Kostadinov, W. V. Chen, T. Maniatis, and J. R. Sanes, "Protocadherins mediate dendritic self-avoidance in the mammalian nervous system," *Nature*, vol. 488, no. 7412, pp. 517–521, 2012.

[10] J. L. Lefebvre, J. R. Sanes, and J. N. Kay, "Development of dendritic form and function," *Annual review of cell and developmental biology*, vol. 31, pp. 741–777, 2015.

[11] A. M. Garrett, A. Khalil, D. O. Walton, and R. W. Burgess, "Dscam promotes self-avoidance in the developing mouse retina by masking the functions of cadherin superfamily members," *Proceedings of the National Academy of Sciences*, vol. 115, no. 43, pp. E10216–E10224, 2018.

[12] E. A. Bushong, M. E. Martone, Y. Z. Jones, and M. H. Ellisman, "Protoplasmic astrocytes in ca1 stratum radiatum occupy separate anatomical domains," *Journal of Neuroscience*, vol. 22, no. 1, pp. 183–192, 2002.

[13] B. A. Barres, "The mystery and magic of glia: a perspective on their roles in health and disease," *Neuron*, vol. 60, no. 3, pp. 430–440, 2008.

[14] H. M. Barber, M. F. Ali, and S. Kucenas, "Glial patchwork: oligodendrocyte progenitor cells and astrocytes blanket the central nervous system," *Frontiers in Cellular Neuroscience*, vol. 15, p. 803057, 2022.

[15] M. K. Hayashi, K. Sato, and Y. Sekino, "Neurons induce tiled astrocytes with branches that avoid each other," *International Journal of Molecular Sciences*, vol. 23, no. 8, p. 4161, 2022.

[16] H. Maruoka, N. Nakagawa, S. Tsuruno, S. Sakai, T. Yoneda, and T. Hosoya, "Lattice system of functionally distinct cell types in the neocortex," *Science*, vol. 358, no. 6363, pp. 610–615, 2017.

[17] T. Ebina, K. Sohya, I. Imayoshi, S.-T. Yin, R. Kimura, Y. Yanagawa, H. Kameda, H. Hioki, T. Kaneko, and T. Tsumoto, "3d clustering of gabaergic neurons enhances inhibitory actions on excitatory neurons in the mouse visual cortex," *Cell Reports*, vol. 9, no. 5, pp. 1896–1907, 2014.

[18] X. Wang, W. E. Allen, M. A. Wright, E. L. Sylwestrak, N. Samusik, S. Vesuna, K. Evans, C. Liu, C. Ramakrishnan, J. Liu, *et al.*, "Three-dimensional intact-tissue sequencing of single-cell transcriptional states," *Science*, vol. 361, no. 6400, p. eaat5691, 2018.

[19] B. Cleland, W. Levick, and H. Wässle, "Physiological identification of a morphological class of cat retinal ganglion cells.," *The Journal of physiology*, vol. 248, no. 1, pp. 151–171, 1975.

[20] J. R. Sanes and R. H. Masland, "The types of retinal ganglion cells: current status and implications for neuronal classification," *Annual review of neuroscience*, vol. 38, pp. 221–246, 2015.

[21] I.-J. Kim, Y. Zhang, M. Yamagata, M. Meister, and J. R. Sanes, "Molecular identification of a retinal cell type that responds to upward motion," *Nature*, vol. 452, no. 7186, pp. 478–482, 2008.

[22] E. Boldog, T. E. Bakken, R. D. Hodge, M. Novotny, B. D. Aevermann, J. Baka, S. Bordé, J. L. Close, F. Diez-Fuertes, S.-L. Ding, *et al.*, "Transcriptomic and morphophysiological evidence for a specialized human cortical gabaergic cell type," *Nature neuroscience*, vol. 21, no. 9, pp. 1185–1195, 2018.

[23] J. Kleinberg, "An impossibility theorem for clustering," *Advances in neural information processing systems*, vol. 15, 2002.

[24] Z. Yao, C. T. van Velthoven, M. Kunst, M. Zhang, D. McMillen, C. Lee, W. Jung, J. Goldy, A. Abdelhak, M. Aitken, *et al.*, "A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain," *Nature*, vol. 624, no. 7991, pp. 317–332, 2023.

[25] M. Zhang, X. Pan, W. Jung, A. R. Halpern, S. W. Eichhorn, Z. Lei, L. Cohen, K. A. Smith, B. Tasic, Z. Yao, *et al.*, "Molecularly defined and spatially resolved cell atlas of the whole mouse brain," *Nature*, vol. 624, no. 7991, pp. 343–354, 2023.

[26] Q. Wang, S.-L. Ding, Y. Li, J. Royall, D. Feng, P. Lesnar, N. Graddis, M. Naeemi, B. Facer, A. Ho, T. Dolbeare, B. Blanchard, N. Dee, W. Wakeman, K. E. Hirokawa, A. Szafer, S. M. Sunkin, S. W. Oh, A. Bernard, J. W. Phillips, M. Hawrylycz, C. Koch, H. Zeng, J. A. Harris, and L. Ng, "The allen mouse brain common coordinate framework: A 3d reference atlas," *Cell*, vol. 181, no. 4, pp. 936–953.e20, 2020.

[27] F. Claudi, A. L. Tyson, L. Petrucco, T. W. Margrie, R. Portugues, and T. Branco, "Visualizing anatomically registered data with brainrender," *eLife*, vol. 10, pp. 1–16, 2021.

[28] A. Baddeley, E. Rubak, and R. Turner, *Spatial Point Patterns: Methodology and Applications with R*. London: Chapman and Hall/CRC Press, 2015.

[29] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan, *Statistical analysis and modelling of spatial point patterns*. John Wiley & Sons, 2008.

[30] A. Baddeley, P. J. Diggle, A. Hardegen, T. Lawrence, R. K. Milne, and G. Nair, "On tests of spatial pattern based on simulation

envelopes," *Ecological Monographs*, vol. 84, no. 3, pp. 477–489, 2014.

[31] T. M. Consortium, J. A. Bae, M. Baptiste, C. A. Bishop, A. L. Bodor, D. Brittain, J. Buchanan, D. J. Bumbarger, M. A. Castro, B. Celii, E. Cobos, F. Collman, N. M. da Costa, S. Dorkenwald, L. Elabbady, P. G. Fahey, T. Fliss, E. Froudarakis, J. Gager, C. Gamlin, W. Gray-Roncal, A. Halageri, J. Hebditch, Z. Jia, E. Joyce, J. Joyce, C. Jordan, D. Kapner, N. Kemnitz, S. Kinn, L. M. Kitchell, S. Koolman, K. Kuehner, K. Lee, K. Li, R. Lu, T. Macrina, G. Mahalingam, J. Matelsky, S. McReynolds, E. Miranda, E. Mitchell, S. S. Mondal, M. Moore, S. Mu, T. Muhammad, B. Nehoran, O. Ogedengbe, C. Papadopoulos, S. Papadopoulos, S. Patel, X. Pitkow, S. Popovych, A. Ramos, R. Clay Reid, J. Reimer, P. K. Rivlin, V. Rose, C. M. Schneider-Mizell, H. S. Seung, B. Silverman, W. Silversmith, A. Sterling, F. H. Sinz, C. L. Smith, S. Suckow, M. Takeno, Z. H. Tan, A. S. Tolias, R. Torres, N. L. Turner, E. Y. Walker, T. Wang, A. Wanner, B. A. Wester, G. Williams, S. Williams, K. Willie, R. Willie, W. Wong, J. Wu, C. Xu, R. Yang, D. Yatsenko, F. Ye, W. Yin, R. Young, S.-c. Yu, D. Xenes, and C. Zhang, "Functional connectomics spanning multiple areas of mouse visual cortex," *bioRxiv*, 2023.

[32] R. A. Fisher, "Statistical methods for research workers," in *Breakthroughs in statistics: Methodology and distribution*, pp. 66–70, Springer, 1970.

[33] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[34] H.-C. Chang and L.-C. Wang, "A simple proof of thue's theorem on circle packing," *arXiv preprint arXiv:1009.4322*, 2010.

[35] D. Stoyan and H. Stoyan, *Fractals, random shapes and point fields: methods of geometrical statistics*. John Wiley and Sons, 1994.

[36] R. y Cajal, *La rétine des vertébrés*. La Cellule, 1893.

[37] B. I. C. C. N. (BICCN), "A multimodal cell census and atlas of the mammalian primary motor cortex," *Nature*, vol. 598, no. 7879, pp. 86–102, 2021.

[38] H. Zeng, "What is a cell type and how to define it?," *Cell*, vol. 185, no. 15, pp. 2739–2755, 2022.

[39] R. J. Douglas, K. A. Martin, and D. Whitteridge, "A canonical microcircuit for neocortex," *Neural computation*, vol. 1, no. 4, pp. 480–488, 1989.

[40] E. Jonas and K. Kording, "Automatic discovery of cell types and microcircuitry from neural connectomics," *Elife*, vol. 4, p. e04250, 2015.

[41] Y. Marghi, R. Gala, F. Baftizadeh, and U. Sümbül, "Joint inference of discrete cell types and continuous type-specific variability in single-cell datasets with mmidas," *bioRxiv*, 2023.

[42] M. Mahajan, P. Nimbhorkar, and K. Varadarajan, "The planar k-means problem is np-hard," *Theoretical Computer Science*, vol. 442, pp. 13–21, 2012.

[43] J. Besag and P. J. Diggle, "Simple monte carlo tests for spatial pattern," *Journal of the Royal Statistical Society Series C: Applied Statistics*, vol. 26, no. 3, pp. 327–333, 1977.

[44] A. Baddeley and G. Nair, "Approximating the moments of a spatial point process," *Stat*, vol. 1, no. 1, pp. 18–30, 2012.

[45] C. Loader, *Local regression and likelihood*. Springer Science & Business Media, 2006.

[46] M. Thomas and A. T. Joy, *Elements of information theory*. Wiley-Interscience, 2006.

[47] V. Isham, "Multitype markov point processes: some approximations," *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, vol. 391, no. 1800, pp. 39–53, 1984.

Methods

Finite Poisson and Strauss point processes

We consider a finite point pattern $\boldsymbol{x} = (x_1, \ldots, x_n)$ observed in a bounded window W. The probability density function of finite point processes are defined with respect to a Poisson process of unit intensity.[29]

Poisson process The homogeneuous Poisson process with intensity β has the following probability density,

$$f(x_1, \ldots, x_n) = c\beta^n$$

where c is a normalizing constant such that $c = \exp((1-\beta)|W|)$. Note that the probability density does not depend on the locations of the points, so that different points are independently and uniformly distributed. The conditional Poisson process with n points is the binomial process, with probability density $f_n(x_1, \ldots, x_n) = 1/|W|^n$.

Strauss process The Strauss process has the following probability density function,

$$f(x_1, \ldots, x_n) = \beta^n \exp\left(-\left(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \phi(||x_i - x_j||)\right)\right)/Z$$

where

$$\phi(r) = \begin{cases} \gamma, & \text{for } r \leq R \\ 0, & \text{for } r > R \end{cases}$$

and Z is an intractable normalizing constant. The Strauss process is called a hardcore process if $\gamma = \infty$ for $r \leq R$, implying that no pairs of points that are closer than the distance R is allowed. The Strauss process has three important parameters, β, γ and R. Note that for Strauss processes, β is called the self-potential or chemical activity, not the intensity of the model. The parameter γ such that $\gamma \geq 0$ is the interaction parameter controlling the strength of inhibition and R is the range of interaction.

Due to the intractable normalization factor, it is hard to work directly with the density function. Instead, we consider the Papangelou conditional intensity function, defined as $\lambda(u|\boldsymbol{x}) = f(\boldsymbol{x} \cup \{u\})/f(\boldsymbol{x})$.

Let

$$s(\boldsymbol{x}, R) = \frac{1}{2} \sum_{x_i, x_j \in \boldsymbol{x}, x_i \neq x_j} \mathbb{1}(||x_i - x_j|| \leq R)$$

denote the number of pairs of points with an interpoint distance of R or less and $t(u, \boldsymbol{x}, R) = s(\boldsymbol{x} \cup \{u\}, R) - s(\boldsymbol{x}, R)$ denote the number of points with an interpoint distance of R or less to a typical point u. We can write the Papangelou conditional intensity function of the Strauss process above as

$$\lambda(u|x_1, \ldots, x_n) = \beta \exp(-\gamma t(u, \boldsymbol{x}, R)).$$

Mixture process We consider a superposition of a Strauss process with interaction strength γ and a homogeneous Poisson process. The following theorem shows that the resulting mixture point process can be approximated by another Strauss process with a weaker interaction strength $\eta < \gamma$.

Theorem 1: Let $\boldsymbol{x_1}$ be a realization of a Strauss process with interaction strength $\gamma \geq 0$, interaction range R and self-potential β in a study region $W \in \mathbb{R}^d$. Let $\boldsymbol{x_0}$ be another point pattern in the same study region. We assume $\boldsymbol{x_0}$ is a realization of a Poisson process with intensity $a\beta$ where $a > 0$. The mixture process, defined as $x = \boldsymbol{x_1} \cup \boldsymbol{x_0}$ is also a Strauss process with the same self-potential β and interaction range R, but with a weaker interaction strength η.

Moreover, when $\boldsymbol{x_0}$ and $\boldsymbol{x_1}$ are finite processes with n_0 and n_1 points respectively and $\alpha = n_0/n_1$ denotes the noise/label ratio,

$$\eta = -\log\left(\frac{e^{-\gamma} + \alpha^2 + \alpha + O(\epsilon^2)}{(\alpha+1)^2}\right)$$

for some small $\epsilon > 0$ with $\epsilon = (1 - e^{-\gamma})/g_1$, $g_1 \geq 0$.

We show the proof in the Mixture Point Process section of Supplementary material.

Nearest neighbor function, L-function and Pairwise correlation function

Nearest neighbor function, L-function and Pairwise correlation function are useful summary characteristics for the spatial point processes.

The nearest-neighbor function $G(r)$ is defined as the distribution function of the random distance from a typical point to its nearest neighbor. It usually serves as the first step in analysis as it is easy to calculate. However, as it only considers the nearest neighbor (short-sighted), it may not be very useful in many cases.

The L-function[43] is a scaled version of the Ripley's K-function $K(r)$, that calculates the average number of points found within distance r from a typical point in the spatial point pattern. In the planar case the L-function is defined as $L(r) = \sqrt{K(r)/\pi}$ for $r \geq 0$. The function has a simple form in the homogeneous Poisson process case, such that $L(r) = r$.

The pairwise correlation function (PCF) characterizes the pair-distances of a spatial point process and is proportional to the derivative of the Ripley's K-function $K(r)$. In the planar case the PCF is defined as $g(r) = K'(r)/2\pi r$ for $r \geq 0$. For homogeneous Poisson processes, $g(r)$ always stay at the value 1. For small r, a value of $g(r)$ below 1 indicates repulsion. Note that for large r, $g(r)$ always takes the value 1.

K-function, L-function and PCF are important second order characteristics for spatial point patterns, and PCF is regarded as the most informative one among the three functions.[29]

Pre-hoc cell type selection

Label noise filter. To reduce the label noise, we require the cell type to have sufficient number and intensity of high confidence cells, defined as cells with classification confidence score at least 0.7 (lower quartile of cell confidence scores) according to the reference clustering.[24] For Glutamatergic cell types, we require the type to take at least 1% – average abundance in the presence of $\sim$100 cell types – of the total number of Glutamatergic cells of high confidence in the section of interest. For GABAergic cells, we further require the number of cells to be greater than 10 because there are fewer GABAergic cells per section. For each cluster, we study the sections where it has sufficient number of high confidence cells. Different sections are treated as replicated samples.

Minimal intensity filter. To remove cell types with low intensities from further analysis, we compare the intensity of high-confidence cells within individual types, as assigned by the reference clustering,[24] to that of hardcore processes with the same interaction range. For each candidate cell type, we take the 20% quantile of its nearest-neighbor distance distribution as its approximated interaction range R. This corresponds to allowing at most 20% noise in the distance distribution of the corresponding hardcore process. We use the hardcore process with low intensity as our comparison benchmark. (See "Benchmark process for the minimal intensity filter" below.) This corresponds to the hardcore process with an interaction range $R = 0.1$ and self potential $\lambda = 25$ (i.e., approximated intensity 15) on a unit square, which rejects CSR at the limit significance of $p = 0.05$ in the absence of independent, additive noise points. To make the benchmark intensity comparable to the cell type intensity at the individual interaction distances, we rescale the benchmark intensity by a factor of $(10R)^2$.

We focus on regions of high density for each cell type and compare its intensity to the benchmark. To avoid edge effects and region bias, we require reasonable homogeneity and sufficient area of the region under consideration. Specifically, we require that the area of the region should be at least $(10R)^2$, comparable to the benchmark window size adjusted for R. We implement this using kernel smoothing as follows. We first calculate the maximum density across the section of interest. Then, we grow the window size by decreasing a density threshold (starting from the maximum density) until the area of the region meets the area constraint. We reject the section for that cell type if the resulting density no longer satisfies the intensity threshold. Hence, we use this intensity filter independently for each section of each candidate cell type, and exclude those whose intensity falls below the threshold.

Secondary label noise filter. For each cluster that passes through the intensity filter, we calculate the ratio of the number of high confidence cells to the total number of cells of the cluster (HC ratio) in every qualified section. Considering that low HC ratios are indicative of label noise, we further exclude those cell types whose qualified sections all have low HC ratios (≤ 0.5), i.e., cells with confidence score greater or equal to 0.7 only take less than 50% of the total cells in the section. Importantly, the threshold is chosen such that HC$= 0.5$ corresponds to the noise/label ratio (Theorem 1) $\alpha = 1$, where even the hardcore process with high intensity decays to a CSR. We note that more restrictive parameters can be chosen and our filters are robust to different parameter choices.

Summary of filtering steps. We show the filtering results for Glutamatergic and GABAergic clusters in Supplementary Table S6. There are around 100 Glutamatergic and GABAergic clusters in each section, 10-20

of which have a sufficient number of high confidence cells. The intensity filter further excludes around 50% of the candidate cell types in each section. In total, there are 18 Glutamatergic cell types and 10 GABAergic cell types that are qualified in at least one section. We show the HC ratio for each qualified cell type in each section (Supplementary figure S6).

Benchmark process for the minimal intensity filter

We use benchmark simulations to identify a minimal intensity below which the power of statistical testing is not enough to detect structure in the point process. For this, we take the unit square as our study region and fix the hardcore interaction range $R = 0.1$. We note that this does not affect the generality of our study as rescaling is always possible: if we consider a different interaction range, we can choose a different window size and rescale the point pattern to window of unit size.

We simulate Strauss hardcore point patterns of high, medium and low intensities, such that the self-potential λ is equal to 100, 50 and 25, respectively. The corresponding true intensity is equal to 34, 25 and 15 approximately, using Poisson-saddlepoint approximation.[44] The generated instances have 32, 25 and 15 points (Supplementary Figure S7). We note that, if we approximate the coverage region of each point by a circle of radius $R = 0.1$, then the total coverage of the points is approximately 1 in the high intensity case, the area of the unit square, ignoring edge effects.

The noise points Y are assumed to follow a Poisson process throughout the study region. Since we are interested in the noise/label ratio, we condition on the number of noise points in the study region. We define the label noise ratio α to be the number of the noise points divided by the number of points in the baseline patterns. We gradually add noise points to the baseline point patterns, from zero to the equal number of points as the baseline. We would anticipate the mixture process to show a gradual decay to the completely spatial randomness (CSR).

We quantify the approach to CSR using the DCLF hypothesis test based on the L-function. We take 99 CSR point patterns at each noise proportion α (See Theorem 1) for a starting and incremental p-value of 0.01. To stabilize the outcome, we fix the reference CSR patterns to be the same for each α ratio.

We plot the results for the three cases in Supplementary Figure S8. First, in all cases, the increase in p-values is significant. This confirms our intuition that the mixture process gradually decays to CSR. Moreover, we note the significance level of 0.05 corresponds to different levels of noise proportion α for different cases. Specifically, it corresponds to $\alpha = 0.4$ in high intensity case, $\alpha = 0.2$ in medium intensity case, and around 0 in low intensity case. The implies that the pattern is more robust to noise as we increase intensities of the baseline hardcore process. Moreover, the ability to detect the low-intensity Strauss pattern disappears as soon as any noisy points are added.

We further explore the effect of the point process intensity by testing the p-value under the DCLF test for hardcore process realizations with different intensities. Specifically, we consider hardcore processes with intensity parameters ranging from 10 to 50. For each setting, we simulate 10 point processes and test for statistical significance for each of them. We take the average over these 10 samples and plot the corresponding results in Supplementary Figure S9, which shows that the p-value decreases with the intensity parameter. It starts at around $p = 0.2$ when the intensity parameter (self-potential) is 10, implying insignificance, and decreases to $p = 0.05$ at around 25, the intensity parameter corresponding to the low intensity case studied above (Supplementary Figure S8). It then stabilizes below 0.05 for intensity parameters above 40. We thus take the low intensity case with a self-potential of 25 as our benchmark intensity for cell type selection.

Selection of regions of interest

We perform statistical testing on tightly selected regions of interest to avoid assigning false significance to the fact that cells of a given type typically do not cover the whole brain. To avoid bias in region selection, we consider all cells of a given cell type, not just the cells annotated as “high-confidence”. For clusters and supertypes, we begin by considering anatomical parcellation substructures with non-zero counts in any section passing the intensity filter for that type as the initial window. For GABAergic cell types, this often corresponds to the whole section because GABAergic cells of a given type typically spread over the whole cortical section. For cell types with layer specification (e.g., L2/3 IT CTX Glut_1), we use all cells in the subclass they belong to (e.g., L2/3 IT) to identify the relevant parcellation substructures. We define the initial observation window for each cell type and section as the union of parcellation substructures with above-average abundance and above-median intensity per section. Hence, we include every parcellation substructure with reasonable density for the cell type of interest in each section.

We refine the initial observation window to relatively homogeneous regions using a kernel smoothing method. Specifically, we convolve the point pattern with an isotropic Gaussian kernel to smoothen the point pattern. We choose the standard deviation (bandwidth) of the kernel as half of the value suggested by likelihood cross-validation suggested in Ref.[45] Following kernel smoothing, we erode the selected region using the bandwidth of the kernel as the erosion parameter to remove the spurious regions introduced by smoothing.

Projecting a slice of 3-d tissue onto a 2-d surface ignores the thickness of the section $(10\mu m)$ and distorts the distances. This distortion can produce spurious spatial clustering for some cell types, especially the Glutamatergic clusters which are more abundant than the GABAergic ones in the cortex. To alleviate this issue, for Glutamatergic clusters with nearest neighbor distances less than $10\mu m$ (Table 1), we further impose an upper limit on the intensity filter at two standard deviations above the average intensities.

Supplementary Table S7 shows the corresponding statistics for Glutamatergic clusters before this extra step of region selection. Importantly, we note that the two clusters that fail to reject CSR, 0453 L6 CT CTX Glut 5 and 0004 CLA-EPd-CTX Car3 Glut 1, have higher values for both average percentiles (of nearest neighbor distances less than the section thickness) and point process intensities before this extra step of region selection (Table S7).

Simulation of 3D soma locations under the spatial randomness hypothesis

The dataset used in this study[25] comprises 2-d images of $10\mu m$-thick sections with $200\mu m$ gaps between consecutive sections. The cells in the same section are thus assumed to be on the same surface, distorting the actual distances between the cells. For instance, cells may appear to be right on top of each other in the 2-d section, destroying the underlying hardcore process. We note that this is more of a concern for Glutamatergic clusters as the Glutamatergic cells are more numerous compared to GABAergic cells in the cortex.

As a first step towards simulating soma locations under the spatial randomness hypothesis, we obtain an unbiased estimate of the distribution of soma sizes from a volume election microscopy dataset.[31] We examine the effect of 3-d to 2-d projection by studying section cuts from Strauss hardcore processes with hardcore distances specified by estimates of the soma radius.

To simulate a section, we consider a cuboid of size $200\mu m \times 200\mu m \times 50\mu m$ (width $\times$ length $\times$ height). We define the benchmark as cells being separated by hardcore distances only. We consider hardcore interaction ranges corresponding to both the minimum $(7.4\mu m)$ and average soma diameters $(10\mu m)$ based on the electron microscopy dataset and simulate 10 point patterns that respect the hardcore constraint. We observe that projecting the section to 2-d completely destroys the hardcore distance of $7.4\mu m$ while the hardcore distance of $10\mu m$ is decreased to $\sim 5\mu m$ (Supplementary figure S10).

We further explore the effect of projection by studying 3-d point patterns of different intensities. Intuitively, as the intensity of the hardcore point pattern decreases, the hardcore distances will become less significant and the process will approach complete spatial randomness. We simulate 3D Strauss hardcore processes of different intensities with the measured minimum soma diameter as the hardcore distance. For each 3-d point patterns, we calculate the intensity and p-value of the projected 2-d pattern using the 1-sided DCLF test at the specified interaction range. We take 99 CSR point patterns for a starting and incremental p-value of 0.01. Supplementary Figure S10 shows that the projected point patterns become statistically indistinguishable from CSR as intensity drops below 1500 cells/mm^2.

In summary, results for clusters with large 3-d projection errors should be carefully interpreted to avoid falsely declaring a clustering effect due to the apparent overlap of cells. Moreover, statistically significant results obtained by our 2-d testing framework for clusters with small interaction ranges and low intensities compared to the benchmark process above cannot be solely attributed to hardcore separation (due to the physical space the soma occupies). Instead, other factors, such as softcore repulsion with a longer interaction range, should be present.

Testing of spatial randomness for cell populations

We consider the DCLF-test based on the L-function by comparing the theoretical L-function to the L-function of the observed point patterns of interests. For GABAergic cell types, we use the Strauss hardcore process with a hardcore distance of $10\mu m$ (average soma diameter) as the theoretical benchmark of CSR because GABAergic clusters are less affected by the projection error. On the other hand, Glutamatergic cells are more abundant compared to GABAergic cells and suffer more from projection to 2-d. As studied in the previous section, the hardcore distances could be destroyed due to projection so that the Strauss hardcore process is not suitable to serve as the CSR benchmark. Therefore, we use the homogeneous Poisson process as the CSR benchmark for Glutamatergic cell

types.

For each sample population, we conduct the 1-sided DCLF test based on L-functions of 199 CSR point patterns with the same number of points as the observed point pattern in the observation window. The DCLF test is sensitive to the interaction range chosen for the test. We consider distances from the hardcore soma separation ($10\mu m$) to an upper limit such that interactions beyond that limit are not of interest anymore. We set the upper limit as the median nearest neighborhood distance across sections where the cluster passes the noise and minimum intensity filters.

We calculate the minimum p-values at which we reject CSR for regularity over a range of interaction distances (every 1 μm from the hardcore separation distance to the upper limit). The p-values corresponding to individual sections are combined using Fisher's method and corrected for multiple testing using the Benjamini-Hochberg procedure. To stabilize the outcome, we fix the reference CSR patterns to be the same for all distances. If the corrected p-value is below 0.05, we say the point pattern *rejects CSR*.

Effective cluster count

Consider a population $\mathcal{X}$ in a study region W that consists of N unique cluster labels, i.e, $\{c_1, \ldots, c_N\}$. For each cluster c_i, we calculate its relative abundance $p_i = N(c_i|\mathcal{X})/\sum_j N(c_j|\mathcal{X})$ where $N(c_i|\mathcal{X})$ is the number of samples of cluster c_i inside the study region W. We define the effective cluster count (ECC) of $\mathcal{X}$ in W as

$$\mathrm{ECC}_W(\mathcal{X}) = \exp\left(-\sum_{i=1}^{N} p_i \log p_i\right)$$

If the cluster is studied on multiple brain sections $k = 1, \ldots, K$, we calculate its ECC for study region W_k in each section k and take a simple average over all sections, i.e,

$$\mathrm{ECC}(\mathcal{X}) = 1/K \sum_{k=1}^{K} \mathrm{ECC}_{W_k}(\mathcal{X})$$

The definition of ECC relates to the concept of types ((46), Section 11.1) in information theory. We can treat the population $\mathcal{X}$ in study region W as a marked point process of n points where the mark of each point corresponds to its cluster label $c \in \{c_1, \ldots, c_N\}$. We then tabulate the point process in a study window W as a sequence of n symbols from an alphabet $A = \{c_1, \ldots, c_N\}$ (arbitrary ordering). We denote the tabulated sequence as $\boldsymbol{x} = (x_1, \ldots x_n)$. The type $P_{\boldsymbol{x}}$ of the sequence $\boldsymbol{x}$ is defined as the relative proportion of occurrences of each symbol of A, which corresponds to the relative abundance of each cluster.

Note that the type does not uniquely define a sequence. Let $\mathcal{P}_n$ denote the set of sequences of length n. If $P \in \mathcal{P}_n$, the type class of P is defined as $T(P) = \{\boldsymbol{x} \in A^n | P_{\boldsymbol{x}} = P\}$. By Theorem 11.1.3 in (46), for any $P \in \mathcal{P}_n$,

$$\frac{1}{(n+1)^{|A|}} 2^{nH(P)} \leq |T(P)| \leq 2^{nH(P)}$$

where the entropy is measured in bits. The theorem states that there are an exponential number of sequences of each type, i.e, $|T(P)| \doteq 2^{nH(P)}$.

Now we introduce the concept of effective count by taking the n-th root of the size upper limit and using nats as our entropy measure. Note that the size of $T(P)$ is the number of ways of arranging $np_1, \ldots np_N$ objects in the sequence and there are n elements in total. If we assume there are effectively only $\tilde{N}$ symbols in the sequence, each component can only take $\tilde{N}$ values, so there are at most $\tilde{N}^n$ choices for arrangement. By equating the number of arrangements, we have

$$\tilde{N} = T(P)^{1/n},$$

defined as the ECC of population $\mathcal{X}$ in study region W.

Moreover, we assume that different brain sections are replicated samples of the cell type distribution. Therefore, the types in different study regions can be thought of as i.i.d. samples form the type distribution. We can then take a simple average over multiple brain sections to obtain the overall ECC of the cell type $\mathcal{C}$.

Supplementary Material

Distribution of the radii of cell bodies

We obtain a distribution of soma sizes by studying soma segmentations from multiple areas of mouse visual cortex in the MICrONS dataset.[31] The dataset consists of 94014 cells with extracted somatic measurements, including nucleus area and volume, nuclear folding metrics, soma area and volume and somatic synapse information. Each cell also has a nucleus id, soma id and a predicted cell type.

We approximate the soma by a sphere and estimate the soma radius from the soma volume measurement. Figure S1 shows distributions of the estimated soma radius for Glutamatergic and GABAergic cells.

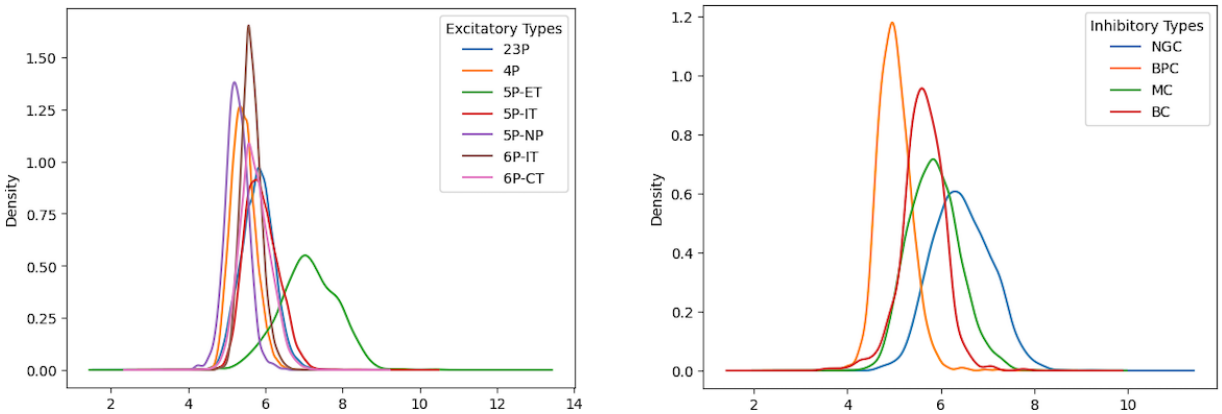


Figure S1. Distribution of soma radius. Estimated soma radius for Glutamatergic (left) and GABAergic (right) subclasses in the MICrONS dataset. The radius units (x-axis) are in μm.

Except for 5P-ET cells, the approximated soma radius is similar across Glutamatergic types. The minimum radius is $\sim 3.7\mu m$, and the average is $5\mu m$. The soma radius estimate for 5P-ET cells is significantly larger with a mean of $\sim 7\mu m$. The soma radius estimate for GABAergic cells varies by subpopulation. The minimum radius estimate is $3.5\mu m$, slightly smaller than the minimum estimate for Glutamatergic cells.

Pairwise correlation functions for all clusters

Table S1 and S2 show that inhibitory clusters generally have higher p-values per section compared to excitatory clusters. That is, inhibitory clusters have less significant deviations from CSR in general, based on the proposed testing paradigm. This is also hinted by Figure 1D, where the red line (observed L-function for 0109 L2/3) deviates more from the shaded upper and lower envelopes of the CSR simulations, compared to the blue line (observed L-function for 0741 Pvalb).

To further explore this difference between excitatory and inhibitory clusters, we plot the pairwise correlation functions of the cell bodies in every cluster that reject CSR and has at least 10 samples in a window of interest (Supplementary Figures S2 and S3). For each cluster, we pool the PCF over selected sections with $p < 0.05$ and plot over the range from $0\mu m$ to $50\mu m$ for excitatory clusters and to $120\mu m$ for inhibitory clusters (maximum of `'R'` in Table 1). We use the standard kernel estimator to compute the PCF by default.[35] However, for clusters with non-zero average percentiles (`'%'` in Table 1)), we use a modified estimator[28] to improve the distortion near zero due to the 3D projection error.

In *absolute* terms, inhibitory clusters generally show very clear hardcore distances and have much longer (2-3 times) radius of self-avoidance compared to excitatory clusters. However, we note that PCFs for most of the inhibitory clusters cross the value of 1 before its upper interaction range `'R'`, while the PCFs for most of the excitatory clusters stay below the value of 1 throughout the plotted range (excluding mild initial increases, presumably due to the projection error). Thus, excitatory clusters have longer *relative* interaction ranges and higher interaction strengths compared to inhibitory clusters.

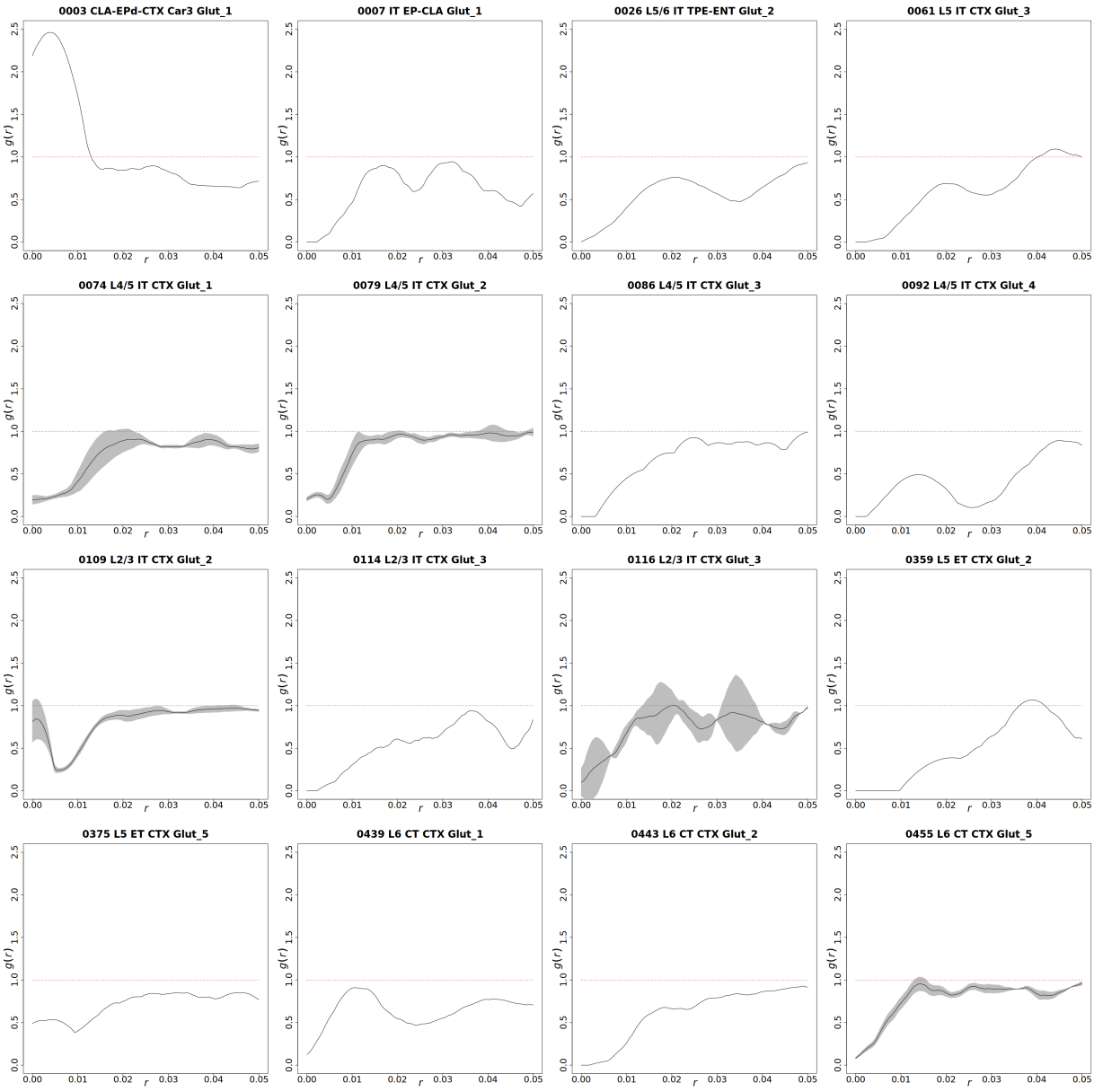


Figure S2. Pairwise correlation functions (PCF) for Glutamatergic clusters pooled over their corresponding sections that reject CSR. Solid lines are averaged PCFs of observed patterns plotted over the range of 50μm, maximum value of 'R' for all excitatory clusters. The shaded regions cover lower and upper values over multiple sections.

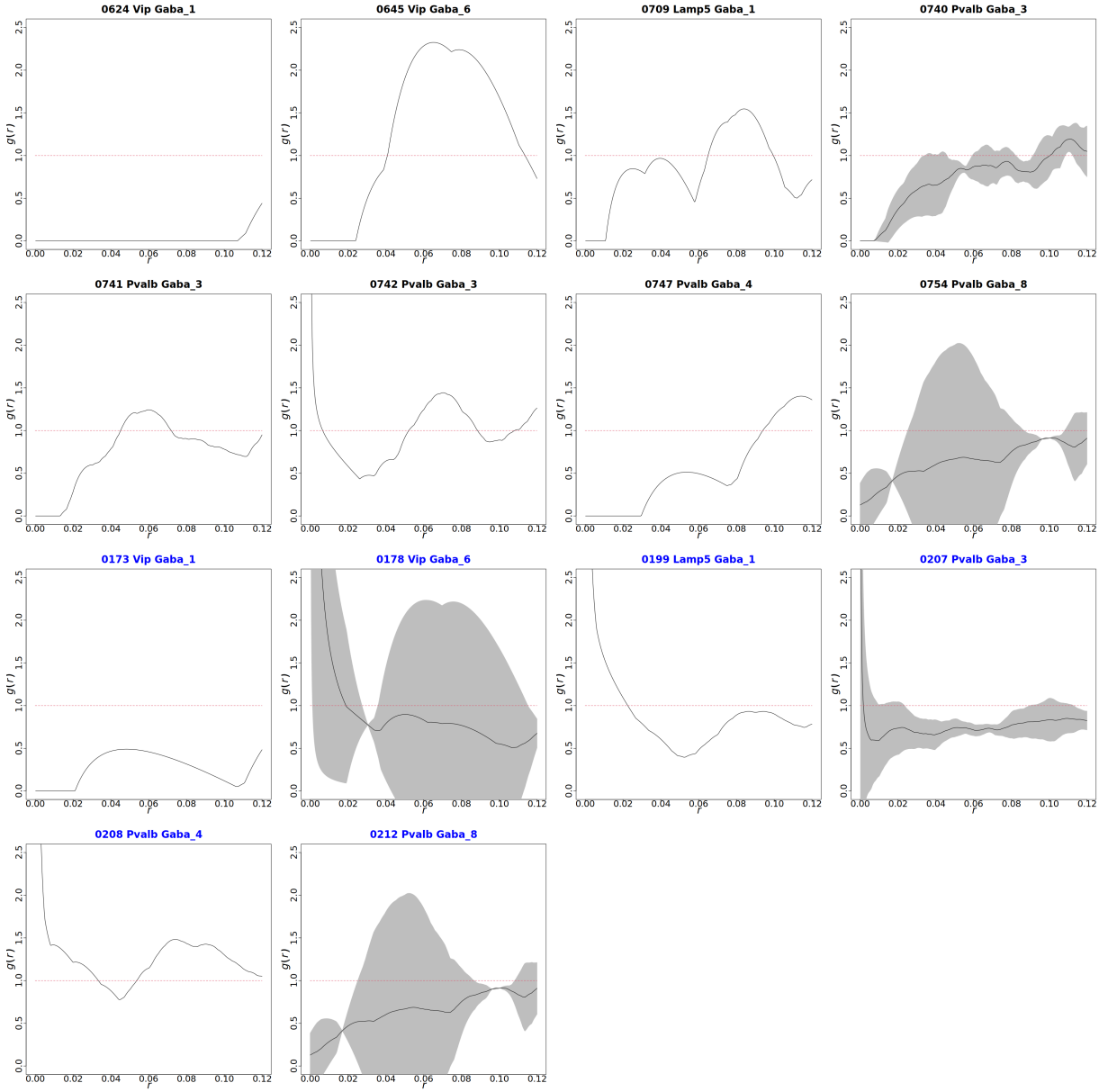


Figure S3. Pairwise correlation functions (PCF) for GABAergic clusters (black titles) and supertypes (blue titles) pooled over their corresponding sections that reject CSR. Solid lines are averaged PCFs of observed patterns plotted over the range of 120μm, maximum value of 'R' for all inhibitory clusters. The shaded regions cover lower and upper values over multiple sections.

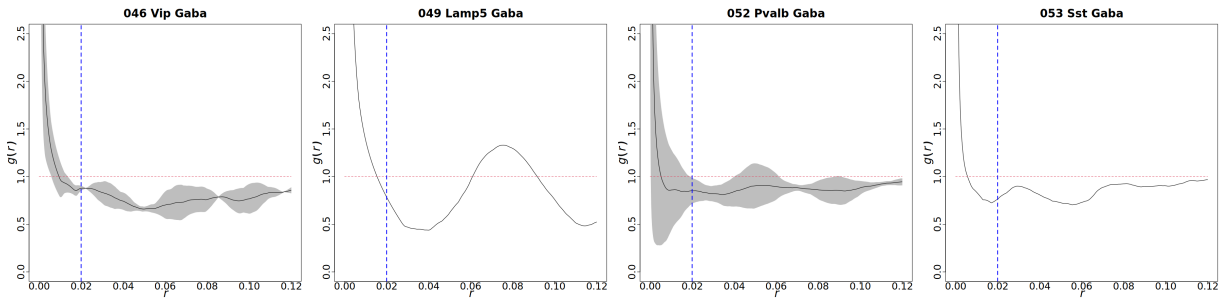


Figure S4. Pairwise correlation functions (PCF) for GABAergic subclasses pooled over their corresponding sections of interest. For details, see Figure S3. All subclasses show clustering effect for distances below approximately $20\mu m$ and spatial randomness for larger values of r.

Statistical testing for individual cluster-section pairs and its aggregation

In the main text, we report raw p-values aggregated over sections according to the Fisher's method. In this section, we show the p-values for individual sections and adjusted p-values using the Benjamini-Hochberg procedure for all types of interest (Tables S1 and S2).

For each type and the section where it passes the noise and minimum intensity filters, we conduct the 1-sided DCLF test with 5 replicates using different seeds. The DCLF test covers the interaction range from the hardcore separation of $10\mu m$ to the range defined by the median value of nearest neighborhood distributions. The minimum p-values over the interaction range are recorded per replicate and averaged. Then the aggregated p-value is calculated according to Fisher's method and corrected using the Benjamini-Hochberg procedure.

Table S1. p-values for Glutamatergic individual cluster-section combinations at cluster level. `'pv.adj'` denotes the Banjamini-Hochberg corrected p-values.

Cluster name	S24	S23	S22	S21	S20	p-value	pv.adj
0109 L2/3 IT CTX Glut_2	0.005	0.005	0.005	0.005	0.005	7.5e-08	1.4e-06
0074 L4/5 IT CTX Glut_1	0.005	0.005	0.005	0.072	0.008	1.1e-06	9.9e-06
0079 L4/5 IT CTX Glut_2	0.013			0.005	0.005	4.1e-05	2.5e-04
0455 L6 CT CTX Glut_5				0.005	0.007	3.9e-04	1.8e-03
0116 L2/3 IT CTX Glut_3	0.007	0.006		0.337		1.1e-03	4.0e-03
0092 L4/5 IT CTX Glut_4	0.005	0.067				3.0e-03	9.0e-03
0061 L5 IT CTX Glut_3					0.005	5.0e-03	0.01
0114 L2/3 IT CTX Glut_3	0.005					5.0e-03	0.01
0443 L6 CT CTX Glut_2				0.005		5.0e-03	0.01
0086 L4/5 IT CTX Glut_3	0.058	0.014				6.6e-03	0.012
0375 L5 ET CTX Glut_5		0.011				0.011	0.018
0026 L5/6 IT TPE-ENT Glut_2		0.015				0.015	0.022
0439 L6 CT CTX Glut_1				0.024		0.024	0.033
0007 IT EP-CLA Glut_1					0.027	0.027	0.035
0359 L5 ET CTX Glut_2		0.039				0.039	0.047
0003 CLA-EPd-CTX Car3 Glut_1				0.922	0.008	0.044	0.049
0453 L6 CT CTX Glut_5					0.061	0.061	0.065
0004 CLA-EPd-CTX Car3 Glut_1		0.064		0.481		0.14	0.14

Table S2. p-values for individual GABAergic cluster-section combinations at the cluster and supertype levels. `'pv.adj'` denotes the Banjamini-Hochberg corrected p-values.

Cluster/Supertype name	S24	S23	S22	S21	S20	p-value	pv.adj
0754 Pvalb Gaba_8				0.008	0.028	0.0021	0.02
0212 Pvalb Gaba_8				0.010	0.026	0.0024	0.02
0624 Vip Gaba_1	0.011					0.011	0.024
0641 Vip Gaba_5			0.013			0.013	0.024
0645 Vip Gaba_6	0.025				0.085	0.015	0.024
0740 Pvalb Gaba_3	0.026	0.039		0.841	0.017	0.0044	0.024
0741 Pvalb Gaba_3	0.007					0.007	0.024
0742 Pvalb Gaba_3	0.063	0.040				0.018	0.024
0747 Pvalb Gaba_4					0.017	0.017	0.024
0811 Sst Gaba_13				0.018		0.018	0.024
0173 Vip Gaba_1	0.013					0.013	0.024
0178 Vip Gaba_6	0.020				0.091	0.013	0.024
0226 Sst Gaba_13				0.018		0.018	0.024
0709 Lamp5 Gaba_1			0.029			0.029	0.035
0199 Lamp5 Gaba_1			0.068			0.068	0.077
0208 Pvalb Gaba_4					0.101	0.100	0.11
0207 Pvalb Gaba_3	0.017	0.994		0.930	0.126	0.13	0.13

Subclass	Section	p-value1	p-value2	Area(mm^2)	R(μm)	%	λ(1/mm^2)	Conf
046 Vip Gaba	20,24	7.9e-03	0.12	4.81	46	0.76	42	0.76
049 Lamp5 Gaba	22	0.005	0.63	2.70	59	2.47	30	0.77
052 Pvalb Gaba	20,21,23,24	1.1e-06	0.51	12.32	30	3.41	93	0.78
053 Sst Gaba	21	0.005	0.24	3.08	39	2.14	61	0.73

Table S3. All GABAergic subclasses display clustering effects. p-values across different brain sections are combined by Fisher's method and corrected for multiple testing using the Benjamini-Hochberg method. The first p-value column corresponds to the one-sided DCLF test for clustering. The second p-value column corresponds to the one-sided DCLF test for repulsion.

Table S4. p-values for individual GABAergic cluster-section combinations at the subclass level.

Subclass name	S24	S23	S22	S21	S20	p-value
046 Vip Gaba	0.16				0.159	0.12
049 Lamp5 Gaba			0.63			0.63
052 Pvalb Gaba	0.085	0.998		1	0.318	0.51
053 Sst Gaba				0.24		0.24

Example case studies

Pvalb pair 0740 and 0742 We study the mixture of Pvalb 0740 and Pvalb 0742 as a single cluster. Since the corresponding set of points does not cover the entire section (Figure S5a), we define the initial observation window for this mixture pair as the union of the observation windows for the individual transcriptomic types.

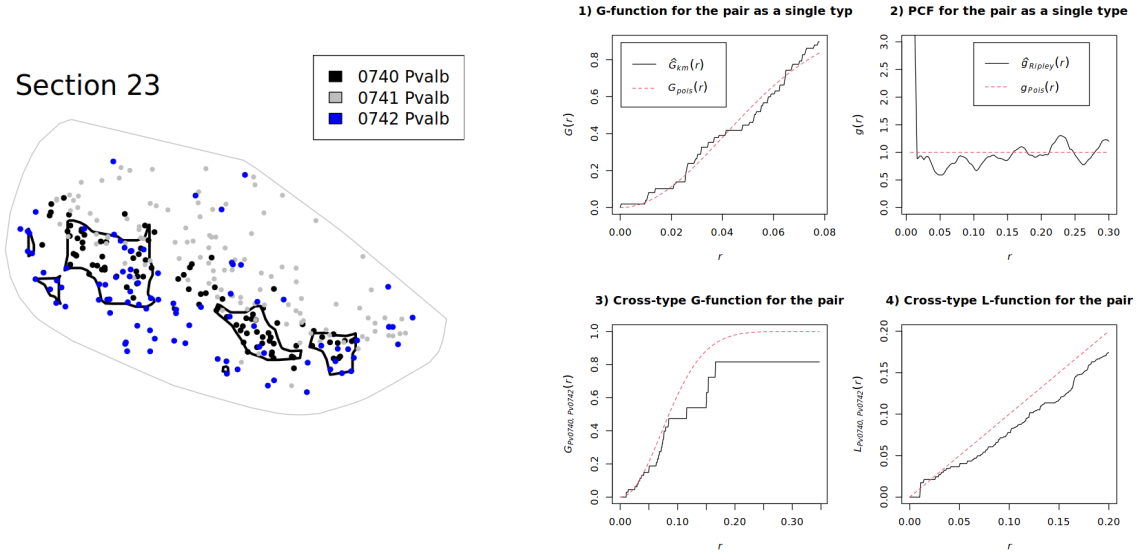
Figure S5b1 shows the nearest-neighbor distribution of the two types as a single cluster and demonstrates that the cells can get arbitrarily close to each other (p-value for self-avoidance: 0.99, 1-sided DCLF test), spatially cluster (weakly) in distances smaller than 50 microns, then their curve closely follows that of CSR. Moreover, we examine the pairwise correlation function for the mixture pair in Figure S5b2. Similarly, the pair correlation function plot is similar to that of the Pvalb subclass (no hardcore distance and no significant local repulsion). Intuitively, for the point process of each cluster in the observation window, the other one serves as noise that destroys its self-avoidance property. We further study their interaction by treating them as a marked point process inside the observation window where the marks are cluster types. Unsurprisingly, Figure S5b3 and Figure S5b4 shows that the two types appear to be spatially clustered at close distances ($\sim 50\mu m$) and repel each other at long distances, at the scale of cortical parcellations (Table S5).

Table S5. Table of cell counts by parcellation structures for mixture of Pvalb 0740 and Pvalb 0742.

Parcellation	Total	Pvalb 0740	Pvalb 0742
AUDd2/3	5	0	5
AUDd4	1	0	1
AUDp4	11	8	3
AUDp5	30	26	4
AUDpo2/3	6	2	4
AUDpo4	5	3	2
SSs2/3	30	3	27
SSs4	34	23	11
SSs5	9	8	1
TEa4	4	2	2
VISC4	5	5	0
VISC5	12	6	6
VISpor2/3	3	0	3

Vip supertype 0173 and cluster 0624 As noted in the main text, the supertype 0173 Vip is an exception of the tree hierarchy diagram. It has an ECC value of 2.54, although it displays significant repulsion (one-sided DCLF test, $p = 0.013$). We examine this supertype and its child cluster 0624 Vip in section 24.

Figure S6a shows the spatial distribution of supertype 0173 Vip neurons in sagittal section 24. For the supertype and its dominant cluster 0624 Vip (blue dots), we plot the corresponding windows for analysis (supertype: black outline, cluster: blue outline). The selected region for the supertype 0173 includes the selected region for the cluster 0624. The region for supertype 0173 is mainly covered by that of 0624, with uncovered regions consisting of 2 cells from

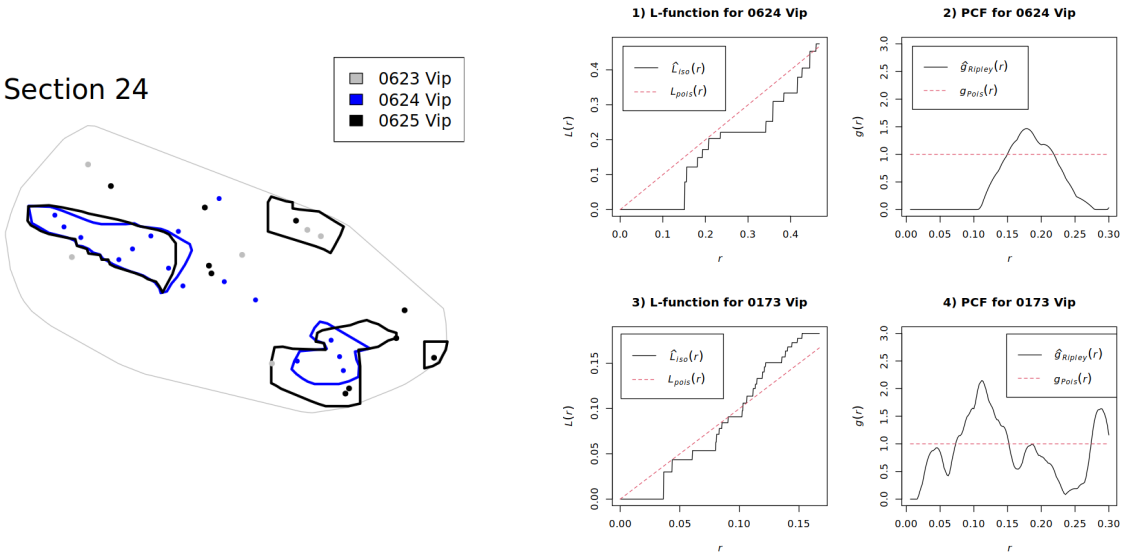


(a) Distribution of the Pvalb pair in section 23. **(b)** Summary functions for the Pvalb cluster pair.

Figure S5. Case study for Pvalb cluster pair 0740 and 0742

0623 Vip and 4 cells from 0625 Vip.

Figure S6b shows the L-functions and pairwise correlation functions for the supertype 0173 and cluster 0624, which demonstrates similarities between the supertype and the cluster in terms of the hardcore distance and the range of local repulsion. Since both cell types 0623 and 0625 have few samples in this observation window, and do not significantly disturb the self-avoidance behavior of 0624 Vip, we hypothesize that an oversplit in the reference hierarchy may contribute to the significance at this Vip supertype level.



(a) Distribution of the Vip supertype and clusters in section 24. **(b)** Summary functions for the Vip supertype and cluster.

Figure S6. Case study for Vip supertype 0173 and cluster 0624.

Mixture point process

Mixture of Strauss and Poisson point processes: We seek to characterize a mixture of a point pattern corresponding to a pure cell type, modeled by the Strauss process, and a point pattern corresponding to cells that are incorrectly assigned to that same type (even though they actually belong to multiple different cell types), modeled by the Poisson process. The theorem below formalizes this setup.

Edge effects: Point process statistics change at the boundaries of finite regions of interest. For instance, trivially,

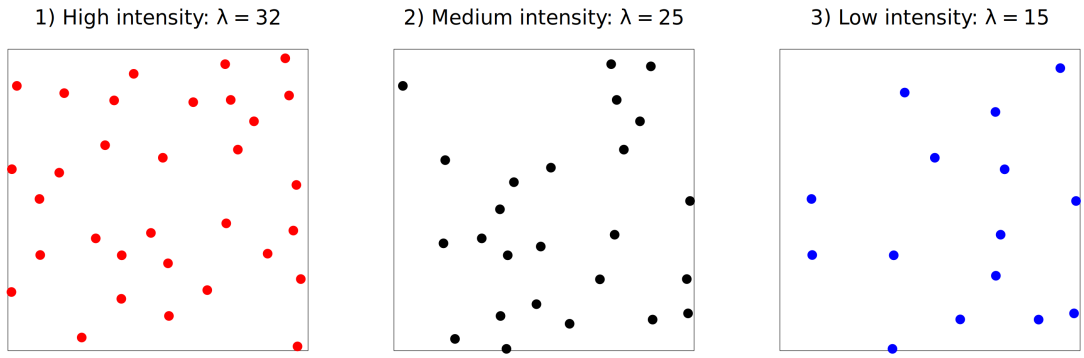


Figure S7. Sample hardcore process point patterns of high, med and low intensities.

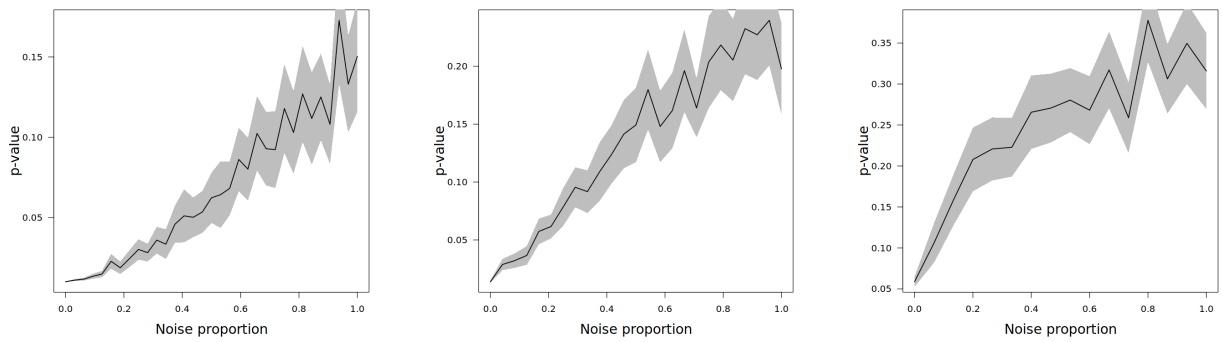


Figure S8. Deviation from CSR for mixtures starting from different hardcore processes. Left to Right: hardcore processes of high, medium and low intensities. In all cases, a point pattern generated from the Poisson point process is independently added to the point pattern of the hardcore process. Noise proportion denotes the ratio of Poisson points to hardcore points. y-axis reports the p-value obtained from the 1-sided DCLF test. We observe decrease in statistical significance with increase in noise proportion. The pattern becomes more robust to noise as the intensity of the baseline hardcore process increases.

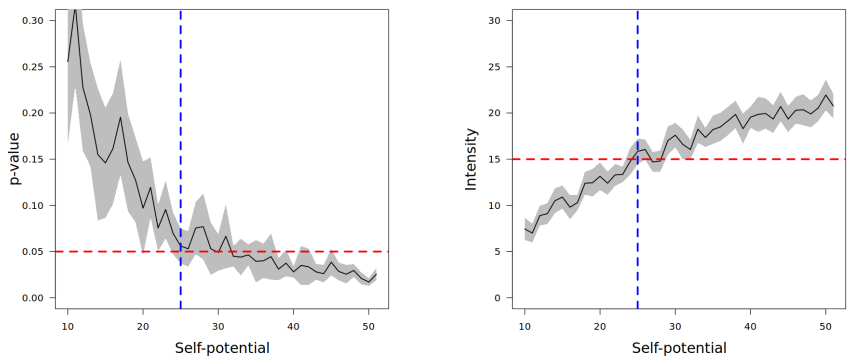


Figure S9. Effect of initial point process intensity. Left: Change in statistical significance with respect to intensity of the hardcore process. y-axis same as that of Figure S8. The p-value decreases to 0.05 at self-potential around 25, approximately corresponding to intensity around 15 (Right), which is taken as our benchmark intensity to be used for cell-type selection.

	Cluster Name	S20	S21	S22	S23	S24
1	0003 CLA-EPd-CTX Car3 Glut_1	0.64	0.68			
2	0004 CLA-EPd-CTX Car3 Glut_1		0.79		0.74	
3	0007 IT EP-CLA Glut_1	0.92				
4	0026 L5/6 IT TPE-ENT Glut_2				0.55	
5	0061 L5 IT CTX Glut_3	0.73				
6	0067 L4/5 IT CTX Glut_1				**0.47**	
7	0074 L4/5 IT CTX Glut_1	0.51	0.56	0.55	0.54	0.57
8	0079 L4/5 IT CTX Glut_2	0.62	0.58	**0.44**		0.51
9	0080 L4/5 IT CTX Glut_2					**0.40**
10	0086 L4/5 IT CTX Glut_3				0.92	0.89
11	0092 L4/5 IT CTX Glut_4				0.51	0.53
12	0109 L2/3 IT CTX Glut_2	0.85	0.86	0.81	0.82	0.80
13	0114 L2/3 IT CTX Glut_3					0.69
14	0116 L2/3 IT CTX Glut_3		0.51		0.71	0.72
15	0359 L5 ET CTX Glut_2				0.62	
16	0375 L5 ET CTX Glut_5				0.69	
17	0436 L6 CT CTX Glut_1		**0.33**	**0.32**		
18	0439 L6 CT CTX Glut_1	**0.50**	0.59			
19	0443 L6 CT CTX Glut_2		0.60			
20	0453 L6 CT CTX Glut_5	0.76				
21	0455 L6 CT CTX Glut_5	0.57	0.66			
1	0624 Vip Gaba_1					0.75
2	0641 Vip Gaba_5			0.71		
3	0645 Vip Gaba_6	0.83				0.74
4	0709 Lamp5 Gaba_1			0.73		
5	0740 Pvalb Gaba_3	0.69	0.62		0.80	0.59
6	0741 Pvalb Gaba_3					0.71
7	0742 Pvalb Gaba_3				0.85	0.73
8	0743 Pvalb Gaba_4	**0.48**				
9	0747 Pvalb Gaba_4	0.76				
10	0754 Pvalb Gaba_8	0.86	0.74			
11	0811 Sst Gaba_13		0.65			

Table S6. Filtering results for Glutamatergic and GABAergic clusters. The table shows the HC ratio (ratio of the number of high confidence cells to the total number of cells) per section from S20 (Section 20) to S24 (Section 24) for cell types passing the minimum intensity filter. Low HC ratio implies higher label noises. We highlight sections with HC ratio ≤ 0.5. In total, 18 Glutamatergic cell types and 10 GABAergic cell types are qualified in at least one section.

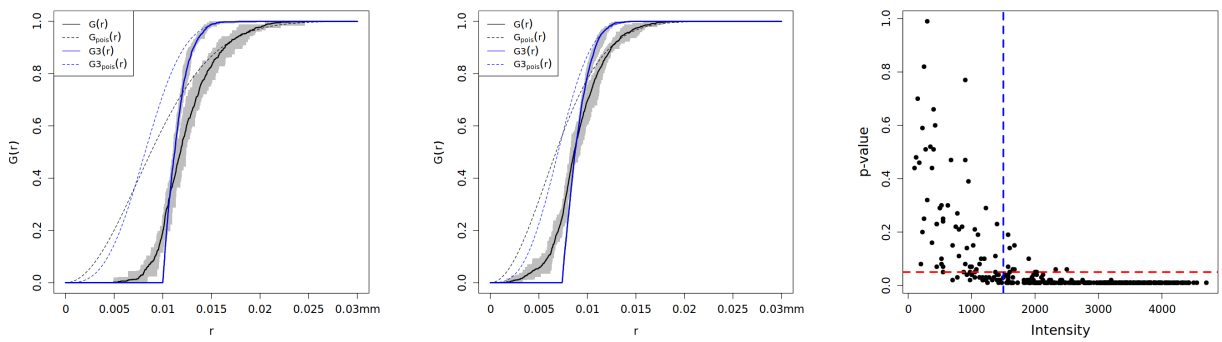


Figure S10. Simulation of 3D soma locations under spatial randomness. Nearest neighbor functions for the 3D and projected point patterns for average (left) and minimum (middle) soma radii used as hardcore distances. (Right) Decrease in statistical significance with respect to intensity of the projected point patterns. All units are in mm.

Cluster name	Section	Area(mm^2)	R(μm)	%	λ(1/mm^2)	Conf
0003 CLA-EPd-CTX Car3 Glut_1	20,21	1.52	29	9.02	278	0.81
0004 CLA-EPd-CTX Car3 Glut_1	21,23	1.51	42	5.41	120	0.85
0007 IT EP-CLA Glut_1	20	0.15	20	7.77	677	0.92
0026 L5/6 IT TPE-ENT Glut_2	23	0.44	49	0.00	133	0.70
0061 L5 IT CTX Glut_3	20	0.44	38	0.00	271	0.78
0074 L4/5 IT CTX Glut_1	20–24	1.81	24	4.37	535	0.75
0079 L4/5 IT CTX Glut_2	20,21,24	1.49	15	14.48	1201	0.74
0086 L4/5 IT CTX Glut_3	23,24	0.48	53	0.00	112	0.84
0092 L4/5 IT CTX Glut_4	23,24	0.57	28	5.61	398	0.74
0109 L2/3 IT CTX Glut_2	20–24	4.48	17	10.00	1125	0.88
0114 L2/3 IT CTX Glut_3	24	0.14	20	6.50	849	0.82
0116 L2/3 IT CTX Glut_3	21,23,24	0.45	15	13.84	1107	0.79
0359 L5 ET CTX Glut_2	23	0.10	41	0.00	165	0.83
0375 L5 ET CTX Glut_5	23	0.55	25	5.43	404	0.82
0439 L6 CT CTX Glut_1	21	0.48	24	7.51	446	0.76
0443 L6 CT CTX Glut_2	21	0.38	24	4.76	550	0.78
0453 L6 CT CTX Glut_5	20	0.11	19	10.91	498	0.80
0455 L6 CT CTX Glut_5	20,21	0.84	19	9.51	658	0.79

Table S7. Statistics of Glutamatergic clusters before region selection. For details, see Table 1.

the interaction strength is not well defined beyond the limits of the observation window. In the formal analysis below, we will "ignore edge effects". In practice, the edge effects can be corrected using various methods such as border correction, which focuses on an interior subregion of the observation window farther from the boundary than the interaction range. Moreover, the edge effects are negligible in the limiting cases of the self-potential $\beta \to \infty$ in a finite observation window or the observation window size $|W| \to \infty$ (in all dimensions) with a finite self-potential.

Theorem 1: Let $\boldsymbol{x_1}$ be a realization of a Strauss process with interaction strength $\gamma \geq 0$, interaction range R and self-potential β in a study region $W \in \mathbb{R}^d$. Let $\boldsymbol{x_0}$ be another point pattern in the same study region. We assume $\boldsymbol{x_0}$ is a realization of a Poisson process with intensity $a\beta$ where $a > 0$. The mixture process, defined as $x = \boldsymbol{x_1} \cup \boldsymbol{x_0}$ is also a Strauss process with the same self-potential β and interaction range R, but with a weaker interaction strength η.

Moreover, when $\boldsymbol{x_0}$ and $\boldsymbol{x_1}$ are finite processes with n_0 and n_1 points respectively and $\alpha = n_0/n_1$ denotes the noise/label ratio,

$$\eta = -\log\left(\frac{e^{-\gamma} + \alpha^2 + \alpha + O(\epsilon^2)}{(\alpha+1)^2}\right)$$

for some small $\epsilon > 0$ with $\epsilon = (1 - e^{-\gamma})/g_1$, $g_1 \geq 0$.

Proof: Considering the mixture process as a pairwise interaction between two independent marked point processes, i.e. points belong to either the Poisson $\boldsymbol{x_0}$ or Strauss $\boldsymbol{x_1}$ processes, the mixture defines a *bivariate Strauss process.*[47] The conditional intensity of this mixture process is

$$\lambda((u,m)|\boldsymbol{x}) = \beta_m \exp\left(-\gamma_m \cdot t(u, \boldsymbol{x_1}, R)\right),$$

where $m \in \{0,1\}$, $\beta_0 = a\beta$, $\beta_1 = \beta$, $\gamma_0 = 0$, $\gamma_1 = \gamma$. When the mixture process is considered as an unmarked point process, the conditional intensity can be written as

$$\lambda(u|\boldsymbol{x}) = a\beta + \beta \exp\left(-\gamma \cdot t(u, \boldsymbol{x_1}, R)\right). \quad \textbf{(1)}$$

The conditional intensity of a Strauss process with an interaction strength η, self-potential β and interaction range R for the same mixture would be

$$\lambda(u|\boldsymbol{x}) = \beta \exp\left(-\eta \cdot t(u, \boldsymbol{x}, R)\right) \quad \textbf{(2)}$$

Solving Equations 1 and 2 for η reveals that the conditional intensity of the mixture can be put into the form of a Strauss conditional intensity:

$$\eta = \frac{-\log\left(a + \exp\left(-\gamma \cdot t(u, \boldsymbol{x_1}, R)\right)\right)}{t(u, \boldsymbol{x}, R)}. \quad \textbf{(3)}$$

Since $a > 0$ and $\gamma \geq 0$,

$$\eta \geq 0.$$

Moreover, since $\alpha > 0$ and $t(u, \boldsymbol{x_1}, R) \leq t(u, \boldsymbol{x}, R)$,

$$\eta < \gamma,$$

implying a weaker interaction strength.

While the expression for the interaction strength η in Equation 3 depends on the process $\boldsymbol{x_1}$ as well as $\boldsymbol{x}$, there is no underlying inhomogeneity in either of the non-interacting, independent mixture components. Therefore, ignoring edge effects, $\eta = \eta(R)$ and the mixture defines a valid Strauss process. To obtain such an expression, let n_0 and n_1 be fixed and $\alpha = n_0/n_1$. The K-function is defined as

$$K(r) = \frac{1}{\lambda}\mathbb{E}\{\#\text{point within distance } r \text{ of a randomly chosen point}\},$$

where λ is the intensity of $\boldsymbol{x_0} \cup \boldsymbol{x_1}$ (Section 4.3, (29)). Ignoring edge effects, the empirical K function is given by (Section 13.4, (28)),

$$\hat{K}(r) = \frac{1}{\lambda n}\sum_{u \in \boldsymbol{x}} t(u, \boldsymbol{x}, r) = \frac{2}{\lambda n} s(\boldsymbol{x}, r).$$

Assuming the interaction strength γ is fairly weak,[47] we write

$$e^{-\gamma} = 1 - g_1\epsilon$$

for some small $\epsilon > 0$ and g_1 is a non-negative constant. Since $\eta < \gamma$, we can write $e^{-\eta} = 1 - g_2\epsilon$ where $g_2 < g_1$ is also a non-negative constant.

The K-function for the Strauss process with interaction strength η can be approximated to first-order in the interaction strength[47] as

$$K(r) = \begin{cases} e^{-\eta}\pi r^2, & \text{for } r \leq R \\ \pi r^2 - (1 - e^{-\eta})\pi R^2, & \text{for } r > R \end{cases} + \mathcal{O}(\epsilon^2).$$

In this way, we can approximate $s(\boldsymbol{x}, r)$, the number of unordered pairs within a distance $r \leq R$ in $\boldsymbol{x}$ as

$$s(\boldsymbol{x}, r) = \frac{1}{2}\lambda n \cdot e^{-\eta}\pi r^2 + O(\epsilon^2).$$

Similarly, for the Strauss component $\boldsymbol{x_1}$, the number of unordered pairs within a distance $r \leq R$ is

$$s(\boldsymbol{x_1}, r) = \frac{1}{2}\lambda_1 n_1 \cdot e^{-\gamma}\pi r^2 + O(\epsilon^2)$$

where λ_1 is the intensity of $\boldsymbol{x_1}$.

The K-function has a simple form for the Poisson point process: $K(r) = \pi r^2$, and the number of r-close pairs is given by

$$s(\boldsymbol{x_0}, r) = \frac{1}{2}\lambda_0 n_0 \cdot \pi r^2$$

where λ_0 is the intensity of $\boldsymbol{x_0}$.

To compute the number of r-close pair in the mixture, we consider the cross-type K-function between the two components $\boldsymbol{x_0}$ and $\boldsymbol{x_1}$. Since $\boldsymbol{x_0}$ and $\boldsymbol{x_1}$ are independent processes and the mixture process results from a set-theoretic union of $\boldsymbol{x_0}$ and $\boldsymbol{x_1}$, the cross-type K-function for the random superposition model (Section 5.3, (29)) is given by

$$K_{0,1}(r) = K_{1,0}(r) = \pi r^2,$$

where

$$K_{i,j}(r) = \frac{1}{\lambda_j}\mathbb{E}\{\#\text{type } j \text{ points within distance } r \text{ of a type } i \text{ point}\}.$$

If we denote the r-close pairs of component $\boldsymbol{x_0}$ and $\boldsymbol{x_1}$ as $s(\boldsymbol{x_0}, \boldsymbol{x_1}, r)$, i.e.,

$$s(\boldsymbol{x_0}, \boldsymbol{x_1}, r) = \sum_{x_i \in \boldsymbol{x_0}} \sum_{x_j \in \boldsymbol{x_1}} \mathbb{1}(||x_i - x_j|| \leq r),$$

we have

$$s(\boldsymbol{x_0}, \boldsymbol{x_1}, r) = \frac{1}{4}(\lambda_1 n_0 + \lambda_0 n_1)\pi r^2.$$

Since $\boldsymbol{x_0}$ and $\boldsymbol{x_1}$ are point patterns inside the same region W, $\lambda = n/|W|$, $\lambda_0 = n_0/|W|$ and $\lambda_1 = n_1/|W|$. Noting that

$$s(\boldsymbol{x}, r) = s(\boldsymbol{x_0}, r) + s(\boldsymbol{x_1}, r) + s(\boldsymbol{x_0}, \boldsymbol{x_1}, r)$$

for $r \leq R$, we obtain an approximate explicit expression for the interaction parameter η as

$$\eta = -\log\left(\frac{e^{-\gamma} n_1^2 + n_0^2 + n_1 n_0 + O(\epsilon^2)}{n^2}\right) \tag{4}$$

$$= -\log\left(\frac{e^{-\gamma} + \alpha^2 + \alpha + O(\epsilon^2)}{(\alpha+1)^2}\right). \tag{5}$$

■

Remark: Clearly, when $\alpha = 0$, $\eta = \gamma$ and as $\alpha \to \infty$, $\eta \to 0$. When the baseline process has large interaction strength s.t. $\gamma \gg 1$ and $\alpha = 1$, η is approximately $\log(2)$.