Olfactory responses of *Drosophila* are encoded in the organization of projection neurons

Kiri Choi[1], **Won Kyu Kim**[1], **Changbong Hyeon**[1*]

***For correspondence:**
hyeoncb@kias.re.kr (CH)

[1]School of Computational Sciences, Korea Institute for Advanced Study, Seoul 02455, Korea

Abstract The projection neurons (PNs), reconstructed from electron microscope (EM) images of the *Drosophila* olfactory system, offer a detailed view of neuronal anatomy, providing glimpses into information flow in the brain. About 150 PNs constituting 51 glomeruli in the antennal lobe (AL) are bundled together in the axonal extension, routing the olfactory signal received at the AL to the mushroom body (MB) calyx and the lateral horn (LH). Here we quantify the neuronal organization by inter-PN distances and examine its relationship with the odor types sensed by *Drosophila*. The homotypic PNs that constitute glomeruli in AL are tightly bundled and stereotyped in position; however, such a glomerular-like PN organization in AL is no longer sustained in the higher brain center. Instead, odor-type dependent clusters consisting of multiple different homotypes become more prevalent in MB calyx and LH. We find that there are statistically significant associations between the spatial organization among a group of homotypic PNs and certain stereotyped behavioral responses. In particular, pheromone-encoding homotypes are segregated from other neurons not only in LH but also in MB calyx. Our findings suggest that before neural computation in the inner brain, some of the olfactory responses are already encoded either by the homotype or in the spatial organization of PNs, illuminating that the labeled-line strategy is at work in the *Drosophila* olfactory system.

Introduction

Anatomical details of neurons inside the brain obtained based on a full connectome of the *Drosophila* hemisphere reconstructed from EM image datasets (***Zheng et al., 2018***), offer the detailed wiring diagram of the brain, shedding light on the origin of brain function. Out of the immense amount of the data, we study the second-order neurons, known as the projection neurons (PNs) of the olfactory system. It is the PNs that bridge between the olfactory receptor neurons (ORNs) in the antenna and maxillary palp and higher olfactory centers where neural computation occurs for fruit flies to sense and perceive the environment (***Hallem and Carlson, 2004***). Three neuropils, namely the antennal lobe (AL), mushroom body (MB) calyx, and lateral horn (LH), are the regions that abound with an ensemble of axonal branches of PNs and synapses (Figure 1). The PNs in AL constitute glomeruli that receive olfactory signals collected from ORNs of the same receptor type (***Gao et al., 2000***; ***Couto et al., 2005***). PNs projecting to MB calyx and LH relay the signals further inside the brain through synaptic junctions with the Kenyon cells (KCs) and lateral horn neurons (LHNs), respectively.

Since the seminal work by Ramón y Cajal (***y Cajal, 1911***), who recognized neurons as the basic functional units of the nervous system, there have been a series of attempts at classifying neu-

rons using different representations of neuronal morphologies and at associating the classified anatomies with their electrophysiological responses and functions (***Uylings and Van Pelt, 2002***; ***Scorcioni et al., 2008***; ***Jefferis et al., 2007***; ***Seki et al., 2010***; ***Gillette and Ascoli, 2015***; ***Lu et al., 2015***; ***Li et al., 2017***; ***Kanari et al., 2018***; ***Mihaljević et al., 2018***; ***Gouwens et al., 2019***; ***Laturnus et al., 2020***). Systematic and principled analyses of neuronal anatomy would be a prerequisite for unveiling a notable link between the PN organization and olfactory features. Several different metrics involving spatial projection patterns (***Jefferis et al., 2007***), electrophysiological properties (***Seki et al., 2010***; ***Gouwens et al., 2019***), topological characteristics (e.g. morphometrics) (***Uylings and Van Pelt, 2002***; ***Scorcioni et al., 2008***; ***Lu et al., 2015***; ***Mihaljević et al., 2018***; ***Gouwens et al., 2019***), intersection profiles (***Gouwens et al., 2019***), and NBLAST scores (***Jeanne et al., 2018***; ***Zheng et al., 2018***) have been utilized over the past decades. More recently, machine learning approaches have been used as a popular tool for classification tasks (***Vasques et al., 2016***; ***Buccino et al., 2018***; ***Mihaljević et al., 2018***; ***Zhang et al., 2021***).

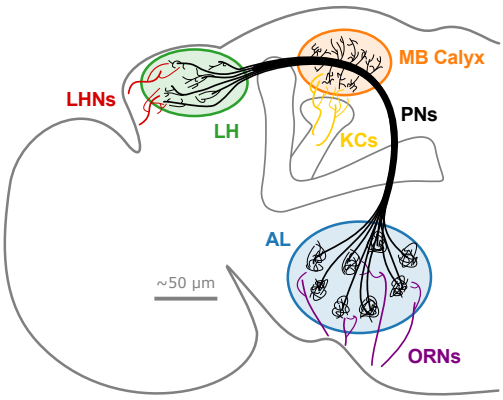


Figure 1. A schematic of the *Drosophila* olfactory system. PNs comprising each glomerulus in AL receive input signals from ORNs of the same receptor type and relay the signals to MB calyx and LH. PNs in MB calyx synapse onto KCs; and PNs in LH synapse onto LHNs.

Among a multitude of information that can be extracted from the neural anatomy associated with PNs, the inter-PN organization draws our attention. We calculate inter-PN distance matrices of PNs in the three neuropils and study them based on the 51 glomerular types to discuss how the inter-PN organization changes as the PNs extend from AL to MB calyx and LH. We carry out statistical analysis to unravel potential associations between the PN organization and the behavioral responses of *Drosophila* to external stimuli encoded to glomerular types. We find a statistically significant association between the structure and the function, such that certain odor types and behavioral responses are linked to a characteristic inter-neuronal organization. Overall, our findings suggest that the *Drosophila* olfactory system leverages the efficiency of the labeled-line design in sensory information processing (***Min et al., 2013***; ***Howard and Gottfried, 2014***; ***Andersson et al., 2015***; ***Galizia, 2014***).

Results and Discussions

Spatial organization of neurons inside neuropils

The pairwise distances $d_{\alpha\beta}$ (Equation 4) between all the possible pairs (α and β) of 111 uniglomerular PNs (uPNs) and the outcomes of $d_{\alpha\beta}$-based clustering analysis can be demonstrated in the form of a matrix (Figure S1) (see Methods for the distance metric and clustering method used).

Hierarchical clustering of PNs in AL identifies 39 clusters from C_1^{AL} to C_{39}^{AL}, which generally matches with the 51 glomerular types known for each PN based on image-based segmentation (***Zheng et al., 2018***; ***Grabe et al., 2015***) and coordinate-based probabilistic methods (***Bates et al., 2020***). The discrepancy in the number between 39 and 51 largely stems from a group of homotypic PNs merged into a single cluster (see Figure S2). A couple of homotypic PNs with poorly formed glomeruli are split and classified into nearby clusters as well (e.g., DL2d in Figure S2A).

The groups of PNs clustered in MB calyx and LH are visualized in Figures 2A and 2B with the colors indicating the odor types encoded by the individual uPNs, which will be discussed in detail later. In MB calyx, PNs are divided into 3 clusters, with C_1^{MB} and C_3^{MB} encompassing the posterior and an-

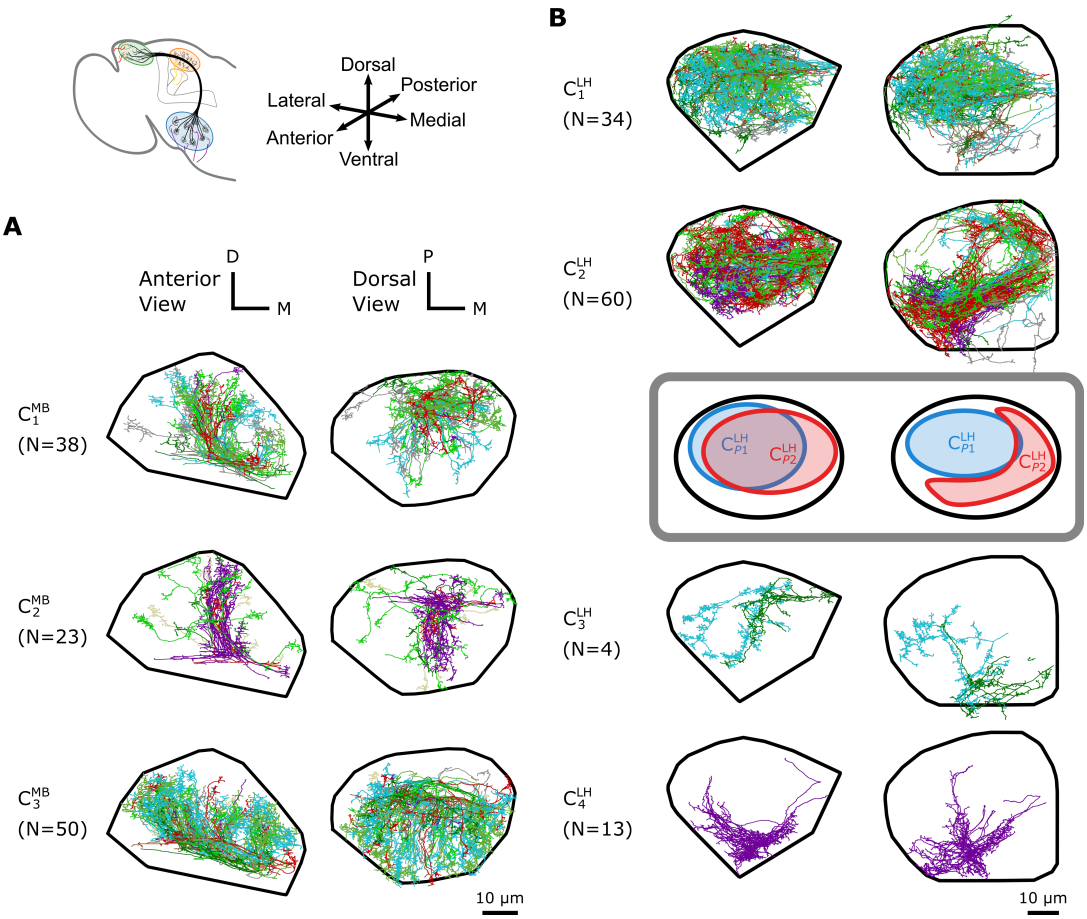


Figure 2. $d_{\alpha\beta}$-based clustering on PNs to (A) MB calyx and (B) LH, resulting in 3 and 4 clusters, respectively. (inset) A cartoon illustrating the relative position between clusters C_1^{LH} and C_2^{LH}. The individual projections are color-coded based on encoded odor types (Dark green: decaying fruit, lime: yeasty, green: fruity, gray: unknown/mixed, cyan: alcoholic fermentation, red: general bad/unclassified aversive, beige: plant matter, brown: animal matter, purple: pheromones). The first and second columns illustrate the anterior and the dorsal view, respectively (D: dorsal, M: medial, P: posterior). The black line denotes the approximate boundary of each neuropil.

terior regions of the neuropil, respectively, and C_2^{MB} somewhere in between and predominantly projecting towards the dorsal region (Figure 2A). The clusters C_1^{MB} and C_2^{MB} are vertically aligned at the central part of MB calyx, whereas the PNs in cluster C_3^{MB} extend their axonal branches over the anterior region.

Four clusters are identified in LH (Figure 2B). The clusters C_1^{LH} and C_2^{LH} include the majority of PNs, mainly projecting to the dorsal part of LH with the cluster C_2^{LH} enveloping C_1^{LH} (Figure 2B, inset). The cluster C_3^{LH} is composed only of 4 neurons. Finally, the cluster C_4^{LH} projecting to the anterior-ventral region displays clear segregation from the other neurons.

We use the Pearson's χ^2-test (See Methods for details of the independence test) to assess the likelihood of independence between two categorical data. First, the test between our $d_{\alpha\beta}$-based clustering output for AL and the glomerular labels (Table S1) yields $p \ll 0.001$ and the Cramér's $V_{\text{glo}}^{C^{\text{AL}}} = 0.835$, the latter of which is analogous to the correlation coefficient for the χ^2-test (see Methods), validating that our $d_{\alpha\beta}$-based clustering result (Figure S2A) is statistically significant. The PNs in the same glomerulus are indeed well bundled and segregated from other PNs belonging to different glomeruli. The χ^2-test of the glomerular labels against MB calyx cluster labels as well as against LH cluster labels also returned statistically significant results ($p \ll 0.001$ for both tests, $V_{\text{glo}}^{C^{\text{MB}}} = 0.699$, $V_{\text{glo}}^{C^{\text{LH}}} = 0.665$). The χ^2-tests between the cluster labels of AL and MB calyx or those of

AL and LH yield similar statistics ($p \ll 0.001$ for both tests, $V_{C^{\mathrm{AL}}}^{C^{\mathrm{MB}}} = 0.664$, $V_{C^{\mathrm{AL}}}^{C^{\mathrm{LH}}} = 0.752$). The mutual information between the same set of nominal variables, which is calculated to verify our χ^2-tests (see Methods), leads to similar conclusions (see SI and Table S2).

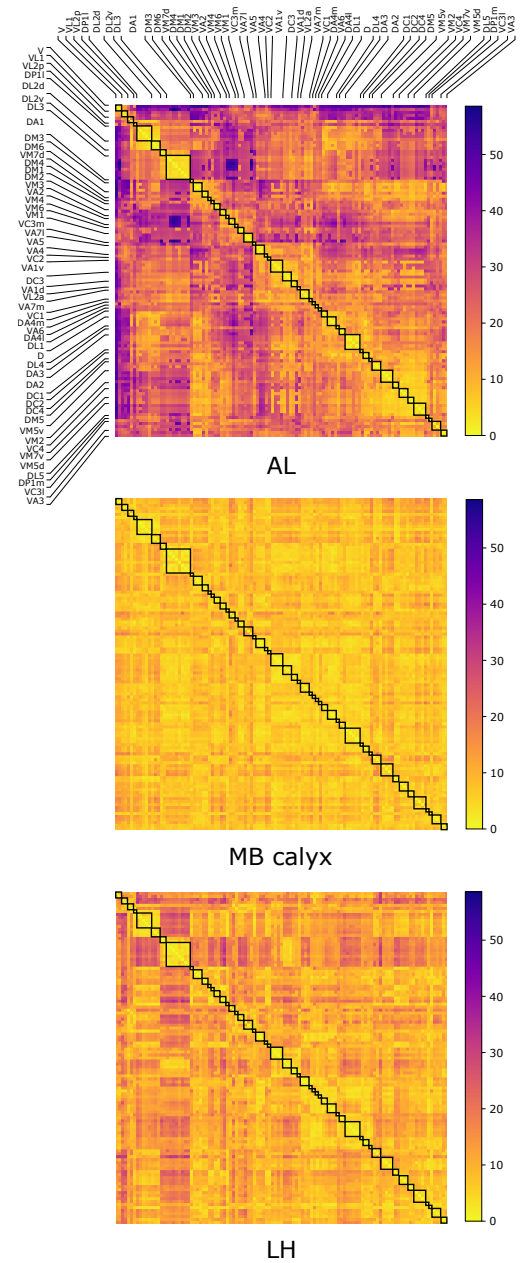


Figure 3. The three $d_{\alpha\beta}$ matrices representing the pairwise distances in units of μm between individual PNs in AL, MB calyx, and LH. The diagonal blocks represent the homotypic PNs comprising the 51 glomeruli, labeled at the edges.

We also categorize the spatial organization of PNs in reference to the glomerular labels. The homotypic PNs constituting a glomerulus in AL give rise to the block diagonal pattern in $d_{\alpha\beta}$-matrix (Figure 3), indicating that the homotypic PNs are tightly bundled in AL and that such organizations are preserved in MB calyx and LH. However, clear differences are found in the off-diagonal part of $d_{\alpha\beta}$ (Figure 3). To make our analysis of $d_{\alpha\beta}$ matrices more quantitative and concise, we define the mean intra- and inter-homotypic PN distances as follows.

$$\bar{d}_{\mathrm{intra},X} \equiv \frac{1}{N} \sum_{\alpha,\beta \in X}^{N} d_{\alpha\beta} \tag{1}$$

and

$$\bar{d}_{\mathrm{inter},X} \equiv \frac{1}{N} \sum_{\alpha \in X, \beta \notin X}^{N} d_{\alpha\beta}, \tag{2}$$

where X denotes a homotype and N is the total number of PN pairs to be averaged. $\bar{d}_{\mathrm{intra},X}$ measures the degree of PNs in the homotype X being bundled. A smaller $\bar{d}_{\mathrm{intra},X}$ means that X-th homotypic PNs are more tightly bundled (Figure S3). Similarly, $\bar{d}_{\mathrm{inter},X}$, which measures the degree of packing (or segregation), is defined by the average distance between the neurons comprising the X-th homotype and others around them. Thus, a small value of $\bar{d}_{\mathrm{inter},X}$ indicates tight packing of X-th homotype with heterotypic PNs, while a large value means that the homotypic PNs comprising the homotype X are segregated from other homotypes (Figure S3).

The degree of bundling averaged over all homotypes, $\bar{d}_{\mathrm{intra}} (= N_X^{-1} \sum_X \bar{d}_{\mathrm{intra},X}) \approx 4\ \mu$m, is comparable over the three neuropils (blue bars in Figure 4A). On the other hand, from $\bar{d}_{\mathrm{inter}}$ defined as the mean inter-bundle distance averaged over all Xs, we find that homotypic PNs are segregated from others in AL, whereas the distinction between the homotypes is only weakly present in MB calyx (the orange bars in Figure 4A and the cartoon of Figure 4B).

Next, we take the ratio of mean intra- to inter-PN distances to quantify the degree of overlap between homotypes,

$$\lambda_X = \bar{d}_{\mathrm{intra},X} / \bar{d}_{\mathrm{inter},X}. \tag{3}$$

A large value of λ_X (particularly $\lambda_X > 1$) signifies that the PNs in the X-th homotype overlap more with others. The value $\lambda (= N_X^{-1} \sum_X^{N_X} \lambda_X)$ averaged over all Xs (red bars in Figure 4A) suggests that the extent of overlap between PNs is maximal in MB calyx and minimal in AL (Figure 4A,B).

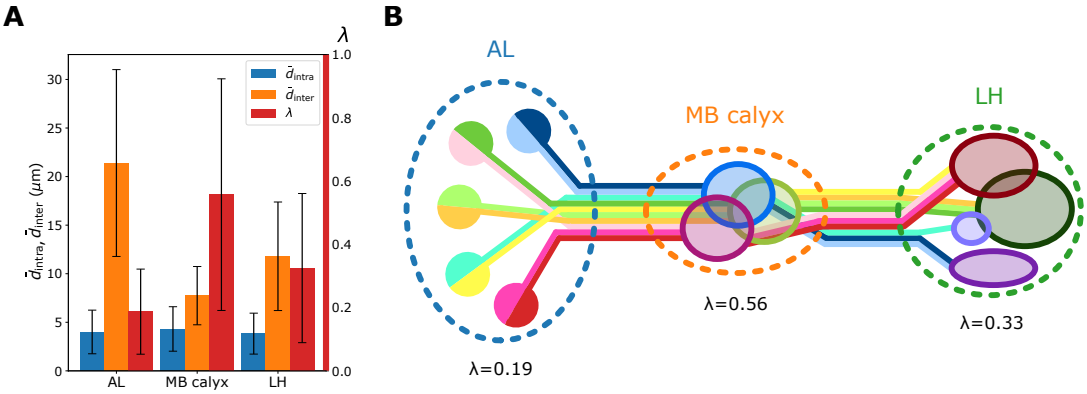


Figure 4. Organization of homotypic PNs in three neuropils. (A) Bar graphs depicting $\bar{d}_{\text{intra}}$ (blue, degree of bundling), $\bar{d}_{\text{inter},X}$ (orange, degree of packing), and the ratio between the two distances λ (red, degree of overlapping). Error bars depict the standard deviation. (B) Diagram illustrating the overall organization of PNs at each neuropil. Homotypic PNs are tightly bundled and segregated in AL. Several groups of homotypic PNs form distinct heterotypic spatial clusters at higher olfactory centers, extensively overlapping in MB calyx (see Figure2A).

Figure 5A and Figure 6 show individual values of λ_X in the three neuropils. We find the following features: (i) In AL, $\lambda_X \leq 0.4$ for all the glomeruli except VC3l and VL2p, indicating that homotypic PNs are tightly bundled and segregated from PNs in other glomeruli; (ii) In MB calyx, most of λ_X's exceed 0.4 and even the cases with $\lambda_X > 1$ are found (DA3, VC3m, VL2p), implying that the PNs are mixed with others; (iii) Although not as much as those in AL, many of PNs in LH are again well bundled and segregated (see Figure 6B, $\bar{d}_{\text{intra},X}$ and $\bar{d}_{\text{inter},X}$ in Figure S3).

The entire neuron morphologies of homotypic PNs from two glomeruli with the smallest ($X =$ DL3) and the largest ($X =$ DM5) λ_Xs in LH are visualized along with the other uPNs (gray) (Figure 5B). The homotype DL3, which seldom overlaps with others in AL and LH ($\lambda_{\text{DL3}} \approx 0.1 - 0.15$), displays a significant overlap in MB calyx ($\lambda_{\text{DL3}} \approx 0.75$). DL3 is tightly packed in AL and LH, whereas it is dispersed in MB calyx. Meanwhile, the homotype DM5, which is well segregated from others in AL ($\lambda_{\text{DM5}} \approx 0.2$), displays significant dispersion, in both MB calyx and LH ($\lambda_{\text{DM5}} \approx 0.6 - 0.7$).

Taken together, the organization of olfactory PNs differs greatly in the three neuropils. The clear homotype-to-homotype segregation in AL no longer holds in MB calyx. The $d_{\alpha\beta}$-based clustering suggests the presence of clusters made of multiple different homotypic PNs (Figure 4B). For some of the homotypes, the well-segregated organizations in AL are recovered when they reach LH (compare Figures 6A and B).

Associations between neuronal organization and olfactory features

Now we explore how the structural features identified from our clustering outputs are linked with odor types and valencies (behavioral responses). As briefly mentioned earlier, the color-codes in Figures 2A, 2B, 5, and 6 depict corresponding odor types encoded by each homotype, which follow the same categorical convention used by Bates *et al.* (***Bates et al., 2020***). The O and × represent the putative valence, which indicates whether *Drosophila* is attracted to or repelled from a specific chemical. For example, DA2 responds to geosmin, a chemical generated from harmful bacteria and mold, which evokes a strong repulsion in *Drosophila* (***Stensmyr et al., 2012***). Similarly, VM3 is suggested to encode repulsive odors, while VM2 and VM7d encode attractive odors (***Bates et al., 2020***). Overall, the following information is acquired from the literature (***Hallem et al., 2004***; ***Galizia and Sachse, 2010***; ***Badel et al., 2016***; ***Bates et al., 2020***) and labeled accordingly:

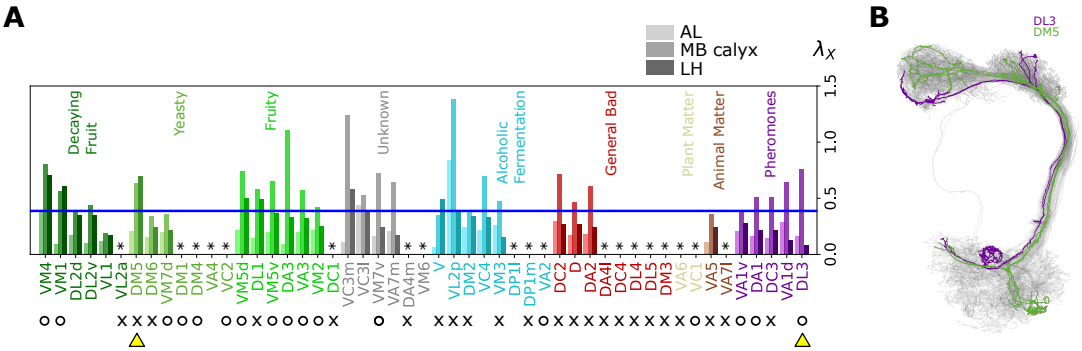


Figure 5. Degree of overlap between inter-homotypic PNs, λ_X (X = VM4, VM1, … , DL3). (A) The degree of overlap (λ_X) for X-th homotype in AL, MB calyx, and LH (from lighter to darker colors). The name of each homotype is color-coded based on the odor types associated with the glomerulus obtained from the literature and is sorted based on the value of λ_X for each odor type at LH. Asterisks ($*$) mark homotypes composed of a single neuron, whose intra-bundle PN distance is not available. O (attractive) and X (aversive) indicate the putative valence information collected from the literature. The blue horizontal line denotes $\lambda_X = 0.4$. (B) Two homotypes, DL3 (purple) and DM5 (green), which are indicated by yellow triangles in (A), are highlighted along with other uPNs (gray).

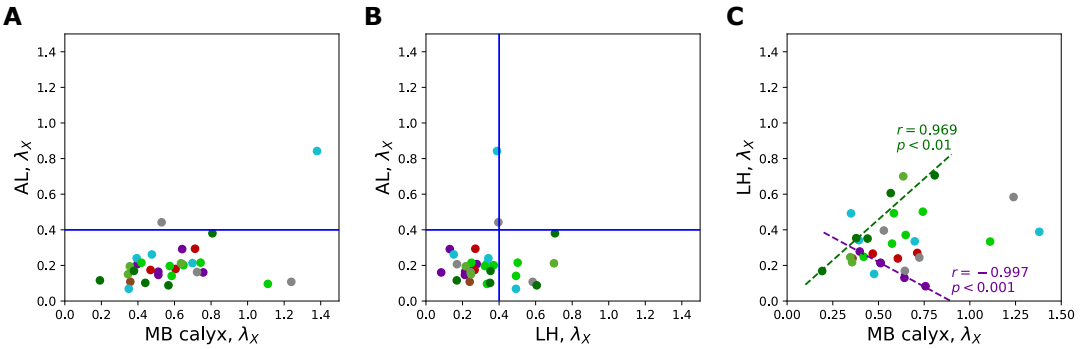


Figure 6. Scatter plots depicting the relationships between λ_Xs at two different neuropils; (A) AL versus MB calyx, (B) AL versus LH, (C) LH versus MB calyx. The color code is the same as in Figure5A. The blue lines in (A) and (B) denote $\lambda_X = 0.4$.

- DA1, DA3, DL3, DM1, DM4, VA1v, VA2, VA3, VC1, VC2, VM1, VM2, VM4, VM5d, VM5v, VM7d, VM7v (17 groups) encode attractive (O) odor.
- D, DA2, DA4l, DA4m, DC1, DC2, DC3, DC4, DL1, DL4, DL5, DM2, DM3, DM5, DM6, DP1m, V, VA5, VA6, VA7l, VA7m, VL2a, VL2p, VM3 (24 groups) encode aversive (×) odor.
- The remaining groups are characterized as either unknown, non-preferential, or conflicting valence information.

Collecting the glomerular labels of tightly bundled homotypic PNs with $\lambda_X < 0.4$ in LH (Figures 5A and 6), we explore the presence of any organizational tendency.

1. In LH, out of 32 glomeruli composed of multiple PNs ($2 \leq n \leq 8$), 25 glomeruli (DL2d, DL2v, VL1, DM6, VM7d, VM5v, DA3, VA3, VM2, VC3l, VM7v, VA7m, VL2p, DM2, VC4, VM3, DC2, D, DA2, VA5, VA1v, DA1, DC3, VA1d, DL3) satisfy the condition of $\lambda_X < 0.4$.
2. Homotypes VA1v, DA1, DC3, VA1d, and DL3 (colored in purple in Figures 2A, 2B, 5A, and 6) encode pheromonal stimuli involved with reproduction (***Grabe et al., 2016***; ***Bates et al., 2020***; ***Dweck et al., 2015***), and DL2d, DL2v, VL1, VM7d, VM5v, DA3, VA3, VM2, and VC4 encode odors presumed to be associated with identifying attractive food sources (***Couto et al., 2005***; ***Semmelhack and Wang, 2009***; ***Mohamed et al., 2019***; ***Bates et al., 2020***) (see Figure 5A). A previous work (***Grosjean et al., 2011***) has identified a group of glomeruli that co-process

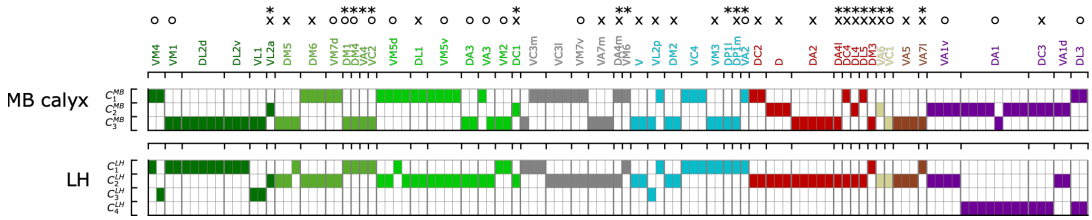


Figure 7. A diagram summarizing how the clusters of PNs in the MB calyx (3 clusters) and LH (4 clusters) are associated with the odor types (Dark green: decaying fruit, olive: yeasty, green: fruity, cyan: alcoholic fermentation, red: general bad/unclassified aversive, beige: plant matter, brown: animal matter, purple: pheromones, gray: unknown (***Bates et al., 2020***)). Asterisks ($*$) on the top mark homotypes composed of a single PN. O and X represent the putative valence information collected from the literature (O: attractive, X: aversive).

food stimuli and pheromones via olfactory receptor gene knock-in coupled with behavioral studies. The list of homotypes mentioned above is consistent with those glomeruli reported in Ref. ***Grosjean et al.*** (***2011***).

3. DM6, VL2p, DM2, VM3, D, and DA2 are the homotypes that are likely associated with aversive food odors. DA2 responds to spoilage; VL2p, DM2, and VM3 to those generated from alcoholic fermentation; DM6 and D to those emitted by flowers (***Galizia and Sachse, 2010***; ***Bates et al., 2020***).
4. Many of the 25 homotypes with $\lambda_X < 0.4$ respond to various odors which can be described as kairomones, a type of odors emitted by other organisms (***Kohl et al., 2015***). This includes the pheromone encoding groups (DA1, DC3, VA1d, and DL3) and others such as DA2, VC3l, VL2p, and VA5, which respond to geosmin, 1-hexanol, 2-oxopentanoic acid (***Silbering et al., 2011***), and 2-methyl phenol, respectively (***Hallem et al., 2004***; ***Galizia and Sachse, 2010***).

The scatter plot of λ_X between MB calyx and LH (Figure 6C) indicates that there exist correlations in λ_X for homotypes associated with particular odor types. Most notably, the 'decaying fruit' group (DL2d/DL2v responds to butyric acid; VL1 to pyrrolidine; VM1 to ammonia; VM4 to phenylethylamine) demonstrates a strong positive correlation ($r = 0.969, p < 0.01$), suggesting that a trend of overlap in MB calyx carries over to LH (see dark green symbols and line in Figure 6C). Meanwhile, a strong negative correlation ($r = -0.997, p < 0.001$) between pheromone-responding PNs implies that a higher degree of overlap in MB calyx leads to a lower degree of overlap in LH.

Figure 7 recapitulates the cluster information from $d_{\alpha\beta}$-based analysis along with the homotypes, odor types (color-codes), and putative valence (attractive (O) and aversive (×) odors). A few points are worth making:

1. 19 out of 51 glomeruli are made of a single PN ($n = 1$, the asterisked glomeruli in Figures 5A and 7), which tends to be characterized with densely branched structures (see Figure S4). Among the 19 homotypes, 11 encode aversive stimuli (×), 5 encode attractive stimuli (O), and 3 have no known valence information (see Table S3). The prevalence of single PN homotypes encoding aversive stimuli is noteworthy.
2. In LH, cluster C_4^{LH}, located in the anterior-ventral region of the neuropil, is composed only of pheromonal stimuli encoding homotypic PNs, DA1, DC3, and DL3 (Figures 2B and 7), which is consistent with the ref. ***Jefferis et al.*** (***2007***). The other pheromonal stimuli encoding PNs, VA1v, and VA1d, are grouped into C_2^{LH}. In MB calyx, the majority of the PNs encoding pheromonal stimuli, except DL3, are clustered to C_2^{MB} (see Figures 2A and 7)
3. Along with the clusters of neurons visualized in Figures 2A and 2B, of particular note, are the clusters formed by several homotypes in the MB calyx and LH. A large portion of PNs in LH that encodes potentially aversive responses are grouped into cluster C_2^{LH}, which envelops cluster C_1^{LH} where mostly food-related homotypes converge (Figure 2B). Given that the synaptic communications established with KCs in MB calyx and with LHNs in LH are crit-

ical for neural computation in the inner brain, the specific type of PN organization in each neuropil should matter. It has been suggested that spatial convergence, segregation, and overlap of different homotypic PNs within neuropil influence the information processing in higher olfactory centers (***Grosjean et al., 2011***).

We validate the statistical significance of the associations between $d_{\alpha\beta}$-based clusters and odor type/valence using the independence test again. The spatial organization of PNs inside neuropils is dependent on odor types ($p \ll 0.001$ for all neuropils) and to a certain extent, odor valence (see Table S1). While the test suggests an association between clusters labels of AL and LH with odor valence information ($p \ll 0.001$), the same is not true for MB calyx where the p-value is not significant enough ($p = 0.147$). The test based on mutual information returns similar results (see SI and Table S2).

According to the previous studies (***Jefferis et al., 2007***; ***Liang et al., 2013***; ***Kohl et al., 2013***; ***Fişek and Wilson, 2014***), PNs in LH and LHN are highly stereotyped in terms of connectivity and response. PNs are spatially organized in AL and LH in association with the order type and valence functions (***Min et al., 2013***; ***Huoviala et al., 2020***). The presence of tightly bundled anatomy of PNs of homotypes ($\lambda_X < 0.4$) in both AL and LH (Figure 6B) may imply that the *Drosophila* olfactory system dedicates a part of the second-order neural circuit on behalf of the “labeled-line" design, which enables the organism to sense urgent chemical stimuli at the early stage of information processing without going through more sophisticated neural computation in the inner brain (***Howard and Gottfried, 2014***; ***Andersson et al., 2015***; ***Min et al., 2013***).

Concluding Remarks

The inter-neuronal organization revealed in this study and its association with odor type/valence are reminiscent of the generally accepted notion that *the form determines the function* in biology. Previously observed stereotypes of neurons in the *Drosophila* olfactory system were based on the differentiation between pheromones and non-pheromones (***Ruta et al., 2010***; ***Kohl et al., 2013***; ***Frechter et al., 2019***; ***Chakraborty and Sachse, 2021***), the whole-cell patch-clamp recording (***Seki et al., 2017***) and imaging studies suggestive of stimulus-dependent arrangement of neurons in LH (***Marin et al., 2002***; ***Wong et al., 2002***; ***Jefferis et al., 2007***). Many homotypic PNs that are spatially localized in LH can be associated with key survival features and a strong innate response (***Seki et al., 2017***). In this sense, the stereotyped localization of pheromone-encoding PNs in C_2^{MB} and C_4^{LH} is of great interest. Our study not only lends support to the existing studies pointing to the labeled-line strategy in the *Drosophila* olfactory system but also suggests that an even more sophisticated level of spatial organization that depends on the homotypes and odor should be present.

While not explored in this work, a host of local neurons (LNs) and multiglomerular PNs (mPNs) also constitute sophisticated neural circuits to regulate the signals received from ORNs (***Sudhakaran et al., 2012***; ***Bates et al., 2020***). A large portion of these mPNs is likely inhibitory (***Berck et al., 2016***; ***Tobin et al., 2017***), whereas the role of interneurons can be both inhibitory and excitatory (***Turner et al., 2008***; ***Wilson et al., 2004***). Electrophysiological measurements indicate that mPNs are narrowly tuned to a specific set of odor stimuli (***Berck et al., 2016***), which is significant given that PNs are generally thought to be more broadly tuned than presynaptic ORNs (***Wilson et al., 2004***). Together, of interest is our observation that most of the single PN homotypes, characterized with densely branched morphology, encode signals with aversive responses. Direct transmission of the associated signals across the three neuropils via a single PN might speed up the overall processing of the olfactory signals as well as reduce the energetic cost.

The *Drosophila* brain EM reconstruction project has evolved to its near completion since the EM image dataset was first released. The reconstruction of the majority of the *Drosophila* central brain as well as the corresponding connectome with detailed information of the individual synapses has become publicly available (***Scheffer et al., 2020***). Our analysis for the second-order neurons inside the *Drosophila* olfactory system can be translated to the brain of different organisms including the

central nervous system (CNS) of humans as well as other parts of the nervous system in *Drosophila*. According to the neurotransmitter map from a recent study (***Dolan et al., 2019***), however, sophisticated processes beyond neuronal anatomy are apparently at work in olfactory signal processing. Thus, to fully glean the functional implication of our findings, knowledge of the synaptic interface between PNs and higher brain centers is essential. The extension of our study to the synaptic interfaces between PNs and KCs will be of great interest for a better understanding of the olfactory signal processing beyond the implication of the neural anatomy studied here.

Materials and methods

Data preparation

We used the neuron morphology skeleton data of 162 *Drosophila* olfactory neurons reconstructed based on the TEMCA2 dataset (***Zheng et al., 2018***) traced from EM images. The neurons were extracted from the right hemisphere of the female *Drosophila*. Out of 162 neurons, three neurons (ID=203840, 2738003, 2738059) are dropped as they were identified not to be PNs. Six neurons in the dataset (ID=1549518, 2738042, 2738083, 2738261, 2970058, 2970073) did not project to any of the neuropils. When constructing the original morphological skeleton, we removed a small number of artifacts in the data, e.g., certain data points are dramatically offset from the expected trajectories, presumably due to a single misalignment. These data points were identified by a geometrical criterion that the distances between three consecutive points should be smaller than $10\ \mu$m.

The morphological information of each neuron is stored as a set of 3D coordinates with the connectivity specified with the parent nodes. Complete reconstruction of neuron morphology was made by connecting data points based on their parent-child relationship.

Next, we systematically demarcated the regions of AL, LH, and MB calyx. The density of data points projected to each axis was used for the identification since the neuropils are featured with a much higher density of data points than the projections connecting them. The boundaries defining each neuropil were systematically chosen from local minima that separate neuropils from rigid axons. Due to the unique structure of PNs, sometimes the projection along a given axis cannot fully differentiate two neuropils. To resolve this issue, projections along each axis were sampled while rotating the data points along the reference axes at 5° increments to obtain multiple snapshots. The densities were analyzed to choose the optimal degrees of rotation along the reference axes that can best segment the neuropils. We used the smallest average and deviation value of density at the local minima as the criteria to choose the optimal rotation. The process has been repeated for each neuropil to identify a set of boundaries along multiple transformed axes with various degrees of rotations that optimally confines each neuropil. This information has been combined to create a set of conditions per neuropil for segmentation. The resulting neuropils were confirmed through visual inspection. An overview of the segmentation process is available in Figure S5.

From the TEMCA2 dataset, we collected a total of 111 labeled uPNs constituting 51 glomeruli under this criteria. Other PNs were manually sorted into multiglomerular PNs or local neurons. We also identified PNs that encompass all three neuropils and those that do not.

Inter-PN distance

The "distance" between two PNs, α and β, with different lengths ($N_\alpha \leq N_\beta$) is quantified by calculating

$$d^2_{\alpha\beta} = \frac{1}{N_\alpha} \sum_{i=1}^{N_\alpha} \min \left[(\boldsymbol{r}^\alpha_i - \boldsymbol{r}^\beta_j)^2 \right]. \tag{4}$$

Equation 4 is evaluated for a pair of $\boldsymbol{r}^\alpha_i$ and $\boldsymbol{r}^\beta_j$ with $j = 1, \ldots, N_\beta$ that gives rise to the minimum value. This particular metric, which is conceptually similar to NBLAST (***Costa et al., 2016***) and to the one used by Zheng *et al.* (***Zheng et al., 2018***), ensures that $d_{\alpha\beta}$ is small when two PNs, α and β, are bundled together.

Morphological and spatial clustering of projection neurons

A method of hierarchical clustering was used to cluster the uPNs at each neuropil. The $d_{\alpha\beta}$ matrices are used to perform hierarchical clustering from which the neurons are sorted into a set of clusters. The number of clusters is decided based on the magnitude of the silhouette coefficient. The silhouette coefficient s_i of an i-th sample (neuron) in a cluster is calculated using

$$s_i = \frac{b_i - a_i}{\max\left(a_i, b_i\right)} \tag{5}$$

where a_i is the mean L_2-distance between a sample i and all other samples in the same cluster, and b_i is the shortest mean distance between a sample i and all samples in the nearest neighboring cluster. The optimal cluster number is determined by the highest average silhouette coefficient, resulting in 39 clusters in AL, 3 clusters in MB calyx, and 4 clusters in LH (Figure S6).

Pearson's χ^2-test of independence

The association between two categorical variables is assessed using Pearson's χ^2-test. For the test, a contingency table, which lists the categorical frequency of two variables, is created. For example, O_{ij} of the i- and j-th element of the contingency table shown below is the frequency counting the putative valence $i = 1$ (attractive), 2 (aversive), 3 (unknown), and the number of PNs in one of the 4 clusters in LH with $j = 1$ (C_1^{LH}), 2 (C_2^{LH}), 3 (C_3^{LH}), 4 (C_4^{LH})

	C_1^{LH}	C_2^{LH}	C_3^{LH}	C_4^{LH}	Total
Attractive	10	16	1	10	37
Aversive	7	32	1	3	43
Unknown	17	12	2	0	31
Total	34	60	4	13	111

Then the χ^2 value is evaluated based on the table using

$$\chi^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{6}$$

where R and C are the numbers of rows and columns, and O_{ij} and E_{ij} are the observed and expected frequencies of the event in the i-th row and j-th column, respectively, with E_{ij} being calculated from O_{ij} as

$$E_{ij} = N p_{i\cdot} p_{\cdot j}, \tag{7}$$

where $p_{i\cdot} = \sum_j^C O_{ij}/N$ and $p_{\cdot j} = \sum_i^R O_{ij}/N$ with N being the total count. Thus, E_{ij} is the frequency expected by assuming that the two categorical data are statistically independent. The Pearson's χ^2 test aims to check whether there is a significant difference between O_{ij} and E_{ij}.

In the χ^2-test, the p-values are estimated using $f_k(x)$, the χ^2-distribution with the degree of freedom $k = (R-1)(C-1)$. If the test returns a χ^2 value that gives rise to a p-value smaller than the defined significance level ($\alpha = 0.01$), the null hypothesis of independence between the two data sets should be rejected. As a result, the distribution of the categorical data is deemed significantly different from a randomly generated distribution, which concludes that the association between two sets of data is statistically significant.

For the above contingency table with $k = 6$, which leads to $\chi^2 \approx 27.3$ (Equation 6), we get a p-value much smaller than the significance level ($\alpha = 0.01$), $p = 1 - \int_0^{\chi^2} f_{k=6}(x)dx \approx 1.29\times10^{-4} \ll \alpha = 0.01$.

When Pearson's χ^2 statistics are available, one can calculate the Cramér's V with bias correction, a measure of association between two categorical variables, as follows.

$$V = \sqrt{\frac{\phi'^2/N}{\min(R'-1, C'-1)}} \tag{8}$$

where $\phi'^2 = \max\left(0, \chi^2/N - (R-1)(C-1)/(N-1)\right)$, $R' = R - (R-1)^2/(N-1)$, and $C' = C - (C-1)^2/(N-1)$. Similar to the Pearson correlation coefficient, the value V ranges between 0 and 1 where 0 indicates no association and 1 indicates a complete association between two categorical variables.

Mutual information

Mutual information (MI) I is used to verify the significance of association between nominal variables observed in Pearson's χ^2-test for independence. The MI measures the information transfer or the similarity between two data. The concept can be extended to clustering outputs to check how two different clustering labels from the same data are similar to each other. Traditionally, the MI between two jointly discrete variables A and B is given by

$$I(A;B) = \sum_{i=1}^{n_A}\sum_{j=1}^{n_B} P(A_i, B_j)\log\left[\frac{P(A_i, B_j)}{P(A_i)P(B_j)}\right] \tag{9}$$

where n_A (or n_B) is the number of clusters in A (or B). Numerically, the MI between two clustering outputs A and B is calculated by evaluating $P(A_i) = N_{A_i}/N$, $P(B_i) = N_{B_i}/N$, and $P(A_i, B_j) = N_{A_i \cap B_j}/N$ where N is the total count and $N_{A_i \cap B_j}$ is the number of elements common in both clusters A_i and B_j.

The significance was assessed by comparing the observed I with the distribution of Is from randomly sampled variables. Specifically, the cluster label was randomly sampled 1000 times to generate a distribution of I under the assumption of independence. The value of observed I is considered significant if the approximated p-value is below 0.01 ($p < 0.01$).

Data Availability

All data generated during this study and Python script are available in DrosophilaOlfaction-main.zip included as the supporting file. They are also available at https://github.com/kirichoi/DrosophilaOlfaction.

Acknowledgements

We thank Dr. Ji Hyun Bak for helpful discussions. This study was supported by KIAS Individual Grants CG077001 (K.C.), CG076002 (W.K.K.), and CG035003 (C.H.). We thank the Center for Advanced Computation in KIAS for providing the computing resources.

Declaration of Interests

The authors declare no competing interests.

References

Andersson MN, Löfstedt C, Newcomb RD. Insect olfaction and the evolution of receptor tuning. Frontiers in Ecology and Evolution. 2015; 3:53.

Badel L, Ohta K, Tsuchimoto Y, Kazama H. Decoding of context-dependent olfactory behavior in Drosophila. Neuron. 2016; 91(1):155–167.

Bates AS, Schlegel P, Roberts RJ, Drummond N, Tamimi IF, Turnbull R, Zhao X, Marin EC, Popovici PD, Dhawan S, et al. Complete connectomic reconstruction of olfactory projection neurons in the fly brain. Curr Biol. 2020; 30(16):3183–3199.

Berck ME, Khandelwal A, Claus L, Hernandez-Nunez L, Si G, Tabone CJ, Li F, Truman JW, Fetter RD, Louis M, et al. The wiring diagram of a glomerular olfactory system. Elife. 2016; 5:e14859.

Buccino AP, Ness TV, Einevoll GT, Cauwenberghs G, Häfliger PD. A deep learning approach for the classification of neuronal cell types. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* IEEE; 2018. p. 999–1002.

y Cajal SR. Histologie du système nerveux de l'homme & des vertébrés: Cervelet, cerveau moyen, rétine, couche optique, corps strié, écorce cérébrale générale & régionale, grand sympathique, vol. 2. A. Maloine; 1911.

Chakraborty SD, Sachse S. Olfactory processing in the lateral horn of Drosophila. Cell and Tissue Research. 2021; p. 1–11.

Costa M, Manton JD, Ostrovsky AD, Prohaska S, Jefferis GS. NBLAST: rapid, sensitive comparison of neuronal structure and construction of neuron family databases. Neuron. 2016; 91(2):293–311.

Couto A, Alenius M, Dickson BJ. Molecular, anatomical, and functional organization of the Drosophila olfactory system. Curr Biol. 2005; 15(17):1535–1547.

Dolan MJ, Frechter S, Bates AS, Dan C, Huoviala P, Roberts RJ, Schlegel P, Dhawan S, Tabano R, Dionne H, et al. Neurogenetic dissection of the Drosophila lateral horn reveals major outputs, diverse behavioural functions, and interactions with the mushroom body. Elife. 2019; 8:e43079.

Dweck HK, Ebrahim SA, Thoma M, Mohamed AA, Keesey IW, Trona F, Lavista-Llanos S, Svatoš A, Sachse S, Knaden M, et al. Pheromones mediating copulation and attraction in Drosophila. Proc Natl Acad Sci. 2015; 112(21):E2829–E2835.

Fişek M, Wilson RI. Stereotyped connectivity and computations in higher-order olfactory neurons. Nat Neurosci. 2014; 17(2):280–288.

Frechter S, Bates AS, Tootoonian S, Dolan MJ, Manton J, Jamasb AR, Kohl J, Bock D, Jefferis G. Functional and anatomical specificity in a higher olfactory centre. Elife. 2019; 8:e44590.

Galizia CG. Olfactory coding in the insect brain: data and conjectures. Euro J Neurosci. 2014; 39(11):1784–1795.

Galizia CG, Sachse S. Odor coding in insects. The neurobiology of olfaction. 2010; 35:70.

Gao Q, Yuan B, Chess A. Convergent projections of Drosophila olfactory neurons to specific glomeruli in the antennal lobe. Nat Neurosci. 2000; 3(8):780–785.

Gillette TA, Ascoli GA. Topological characterization of neuronal arbor morphology via sequence representation: I-motif analysis. BMC Bioinfo. 2015; 16(1):1–15.

Gouwens NW, Sorensen SA, Berg J, Lee C, Jarsky T, Ting J, Sunkin SM, Feng D, Anastassiou CA, Barkan E, et al. Classification of electrophysiological and morphological neuron types in the mouse visual cortex. Nat Neurosci. 2019; 22(7):1182–1195.

Grabe V, Baschwitz A, Dweck HK, Lavista-Llanos S, Hansson BS, Sachse S. Elucidating the neuronal architecture of olfactory glomeruli in the Drosophila antennal lobe. Cell Rep. 2016; 16(12):3401–3413.

Grabe V, Strutz A, Baschwitz A, Hansson BS, Sachse S. Digital in vivo 3D atlas of the antennal lobe of Drosophila melanogaster. J Comp Neurol. 2015; 523(3):530–544.

Grosjean Y, Rytz R, Farine JP, Abuin L, Cortot J, Jefferis GS, Benton R. An olfactory receptor for food-derived odours promotes male courtship in Drosophila. Nature. 2011; 478(7368):236–240.

Hallem EA, Carlson JR. The odor coding system of Drosophila. TRENDS in Genetics. 2004; 20(9):453–459.

Hallem EA, Ho MG, Carlson JR. The molecular basis of odor coding in the Drosophila antenna. Cell. 2004; 117(7):965–979.

Howard JD, Gottfried JA. Configural and elemental coding of natural odor mixture components in the human brain. Neuron. 2014; 84(4):857–869.

Huoviala P, Dolan MJ, Love F, Myers P, Frechter S, Namiki S, Pettersson L, Roberts RJ, Turnbull R, Mitrevica Z, et al. Neural circuit basis of aversive odour processing in Drosophila from sensory input to descending output. BioRxiv. 2020; p. 394403.

Jeanne JM, Fişek M, Wilson RI. The organization of projections from olfactory glomeruli onto higher-order neurons. Neuron. 2018; 98(6):1198–1213.

Jefferis GS, Potter CJ, Chan AM, Marin EC, Rohlfing T, Maurer Jr CR, Luo L. Comprehensive maps of Drosophila higher olfactory centers: spatially segregated fruit and pheromone representation. Cell. 2007; 128(6):1187–1203.

Kanari L, Dłotko P, Scolamiero M, Levi R, Shillcock J, Hess K, Markram H. A topological representation of branching neuronal morphologies. Neuroinfo. 2018; 16(1):3–13.

Kohl J, Huoviala P, Jefferis GS. Pheromone processing in Drosophila. Current opinion in neurobiology. 2015; 34:149–157.

Kohl J, Ostrovsky AD, Frechter S, Jefferis GS. A bidirectional circuit switch reroutes pheromone signals in male and female brains. Cell. 2013; 155(7):1610–1623.

Laturnus S, Kobak D, Berens P. A systematic evaluation of interneuron morphology representations for cell type discrimination. Neuroinfo. 2020; 18(4):591–609.

Li Y, Wang D, Ascoli GA, Mitra P, Wang Y. Metrics for comparing neuronal tree shapes based on persistent homology. PLoS One. 2017; 12(8):e0182184.

Liang L, Li Y, Potter CJ, Yizhar O, Deisseroth K, Tsien RW, Luo L. GABAergic projection neurons route selective olfactory inputs to specific higher-order neurons. Neuron. 2013; 79(5):917–931.

Lu Y, Carin L, Coifman R, Shain W, Roysam B. Quantitative arbor analytics: unsupervised harmonic co-clustering of populations of brain cell arbors based on L-measure. Neuroinfo. 2015; 13(1):47–63.

Marin EC, Jefferis GS, Komiyama T, Zhu H, Luo L. Representation of the glomerular olfactory map in the Drosophila brain. Cell. 2002; 109(2):243–255.

Mihaljević B, Larrañaga P, Benavides-Piccione R, Hill S, DeFelipe J, Bielza C. Towards a supervised classification of neocortical interneuron morphologies. BMC Bioinfo. 2018; 19(1):1–22.

Min S, Ai M, Shin SA, Suh GS. Dedicated olfactory neurons mediating attraction behavior to ammonia and amines in Drosophila. Proc Natl Acad Sci. 2013; 110(14):E1321–E1329.

Mohamed AA, Retzke T, Chakraborty SD, Fabian B, Hansson BS, Knaden M, Sachse S. Odor mixtures of opposing valence unveil inter-glomerular crosstalk in the Drosophila antennal lobe. Nat Commun. 2019; 10(1):1–17.

Ruta V, Datta SR, Vasconcelos ML, Freeland J, Looger LL, Axel R. A dimorphic pheromone circuit in Drosophila from sensory input to descending output. Nature. 2010; 468(7324):686–690.

Scheffer LK, Xu CS, Januszewski M, Lu Z, Takemura Sy, Hayworth KJ, Huang GB, Shinomiya K, Maitlin-Shepard J, Berg S, et al. A connectome and analysis of the adult Drosophila central brain. Elife. 2020; 9:e57443.

Scorcioni R, Polavaram S, Ascoli GA. L-Measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. Nature protocols. 2008; 3(5):866–876.

Seki Y, Dweck HK, Rybak J, Wicher D, Sachse S, Hansson BS. Olfactory coding from the periphery to higher brain centers in the Drosophila brain. BMC Biol. 2017; 15(1):1–20.

Seki Y, Rybak J, Wicher D, Sachse S, Hansson BS. Physiological and morphological characterization of local interneurons in the Drosophila antennal lobe. J Neurophysiol. 2010; 104(2):1007–1019.

Semmelhack JL, Wang JW. Select Drosophila glomeruli mediate innate olfactory attraction and aversion. Nature. 2009; 459(7244):218–223.

Silbering AF, Rytz R, Grosjean Y, Abuin L, Ramdya P, Jefferis GS, Benton R. Complementary function and integrated wiring of the evolutionarily distinct Drosophila olfactory subsystems. J Neurosci. 2011; 31(38):13357–13375.

Stensmyr MC, Dweck HK, Farhan A, Ibba I, Strutz A, Mukunda L, Linz J, Grabe V, Steck K, Lavista-Llanos S, et al. A conserved dedicated olfactory circuit for detecting harmful microbes in Drosophila. Cell. 2012; 151(6):1345–1357.

Sudhakaran IP, Holohan EE, Osman S, Rodrigues V, Vijayraghavan K, Ramaswami M. Plasticity of recurrent inhibition in the Drosophila antennal lobe. Journal of Neuroscience. 2012; 32(21):7225–7231.

Tobin WF, Wilson RI, Lee WCA. Wiring variations that enable and constrain neural computation in a sensory microcircuit. Elife. 2017; 6:e24838.

Turner GC, Bazhenov M, Laurent G. Olfactory representations by Drosophila mushroom body neurons. J Neurophysiol. 2008; 99(2):734–746.

Uylings HB, Van Pelt J. Measures for quantifying dendritic arborizations. Network: Computation in Neural Systems. 2002; 13(3):397.

Vasques X, Vanel L, Villette G, Cif L. Morphological neuron classification using machine learning. Frontiers in neuroanatomy. 2016; 10:102.

Wilson RI, Turner GC, Laurent G. Transformation of olfactory representations in the Drosophila antennal lobe. Science. 2004; 303(5656):366–370.

Wong AM, Wang JW, Axel R. Spatial representation of the glomerular map in the Drosophila protocerebrum. Cell. 2002; 109(2):229–241.

Zhang T, Zeng Y, Zhang Y, Zhang X, Shi M, Tang L, Zhang D, Xu B. Neuron type classification in rat brain based on integrative convolutional and tree-based recurrent neural networks. Sci Rep. 2021; 11(1):1–14.

Zheng Z, Lauritzen JS, Perlman E, Robinson CG, Nichols M, Milkie D, Torrens O, Price J, Fisher CB, Sharifi N, et al. A complete electron microscopy volume of the brain of adult Drosophila melanogaster. Cell. 2018; 174(3):730–743.

Supplementary Information

Identifying association between two categorical data via mutual information

We verify our Pearson's χ^2-test of independence of two categorical variables by calculating the mutual information (see Methods). The mutual information between glomerular labels and $d_{\alpha\beta}$-based clustering output in AL is equal to $I(\text{glo}; C^{\text{AL}}) = 3.367$, which is significantly different (more than 4σ) from the mean of randomly sampled I distribution ($I(\text{glo}; C^{\text{AL}})_{rand} = 2.571 \pm 0.045$). This result is consistent with our χ^2-test, as the mutual information of observed variables is significantly larger than the mutual information under the assumption of random sampling, suggestive of statistically significant association between glomerular labels and AL cluster labels. In MB calyx and LH, the mutual information between glomerular labels and $d_{\alpha\beta}$-based clustering output is given by $I(\text{glo}; C^{\text{MB}}) = 0.983$ and $I(\text{glo}; C^{\text{LH}}) = 1.003$. Both have 4σ or larger deviation from the means of the randomly sampled I distribution ($I(\text{glo}; C^{\text{MB}})_{rand} = 0.547 \pm 0.043$, $I(\text{glo}; C^{\text{LH}})_{rand} = 0.740 \pm 0.041$).

The same method is applied to confirm that statistically significant association exists between odor type and the clustering outputs, with $I(\text{odor}; C^{\text{AL}}) = 1.829$, $I(\text{odor}; C^{\text{MB}}) = 0.368$, and $I(\text{odor}; C^{\text{LH}}) = 0.485$, all of which differ by more than 4σ from the means of the randomly sampled I distribution ($I(\text{odor}; C^{\text{AL}})_{rand} = 1.069 \pm 0.051$, $I(\text{odor}; C^{\text{MB}})_{rand} = 0.082 \pm 0.026$, $I(\text{odor}; C^{\text{LH}})_{rand} = 0.120 \pm 0.032$). For odor valence, we obtain $I(\text{val}; C^{\text{AL}}) = 0.900$, $I(\text{val}; C^{\text{MB}}) = 0.030$, and $I(\text{val}; C^{\text{LH}}) = 0.127$, where only $I(\text{val}; C^{\text{AL}})$ and $I(\text{val}; C^{\text{LH}})$ differ significantly from the means of the randomly sampled I distribution ($I(\text{val}; C^{\text{AL}})_{rand} = 0.400 \pm 0.048$, $I(\text{val}; C^{\text{MB}})_{rand} = 0.018 \pm 0.012$, $I(\text{val}; C^{\text{LH}})_{rand} = 0.028 \pm 0.017$). Overall, the conclusion drawn from the association study based on mutual information is identical to the Pearson's χ^2-test.

Table S1. Pearson's χ^2 tests of independence of variables. C^X indicates cluster labels from $d_{\alpha\beta}$-based clustering at X neuropil. Cramér's V values are displayed on each cell and the corresponding p-values are shown in parentheses. Bold entries are used for specifying statistically significant results.

	C^{MB}	C^{LH}	Glomerular Labels	Odor Type	Odor Valence
C^{AL}	**0.664 (1.682E-09)**	**0.752 (2.190E-18)**	**0.835 (6.865E-134)**	**0.680 (1.001E-31)**	**0.680 (4.153E-10)**
C^{MB}		**0.401 (2.833E-07)**	**0.699 (1.736E-09)**	**0.556 (3.756E-11)**	0.113 (0.147)
C^{LH}			**0.665 (1.730E-11)**	**0.545 (7.665E-15)**	**0.312 (1.290E-4)**

Table S2. Mutual information (observed mutual information (top), randomly sampled mutual information (bottom) in each cell) from the association study. C^X is cluster labels from $d_{\alpha\beta}$-based clustering at X neuropil. Bold entries are used for observed I that are more than 4σ different from the randomly sampled I.

	C^{AL}	C^{MB}	C^{LH}	Glomerular Labels	Odor Type	Odor Valence
C^{AL}		**0.81** $\mathbf{0.400 \pm 0.047}$	**0.923** $\mathbf{0.569 \pm 0.049}$	**3.367** $\mathbf{2.568 \pm 0.045}$	**1.829** $\mathbf{1.069 \pm 0.051}$	**0.900** $\mathbf{0.401 \pm 0.0478}$
C^{MB}			**0.191** $\mathbf{0.028 \pm 0.017}$	**0.983** $\mathbf{0.545 \pm 0.042}$	**0.368** $\mathbf{0.081 \pm 0.027}$	0.030 0.019 ± 0.013
C^{LH}				**1.003** $\mathbf{0.740 \pm 0.042}$	**0.485** $\mathbf{0.123 \pm 0.032}$	**0.127** $\mathbf{0.028 \pm 0.016}$

Table S3. Statistics of the homotypes composed of a single PN (or multiple PNs) and the corresponding putative valence.

	Aversive	Attractive	Unknown	Total
Single PN Homotypes Count	11	5	3	19
Multiple PN Homotypes Count	13	12	7	32
Total	24	17	10	51

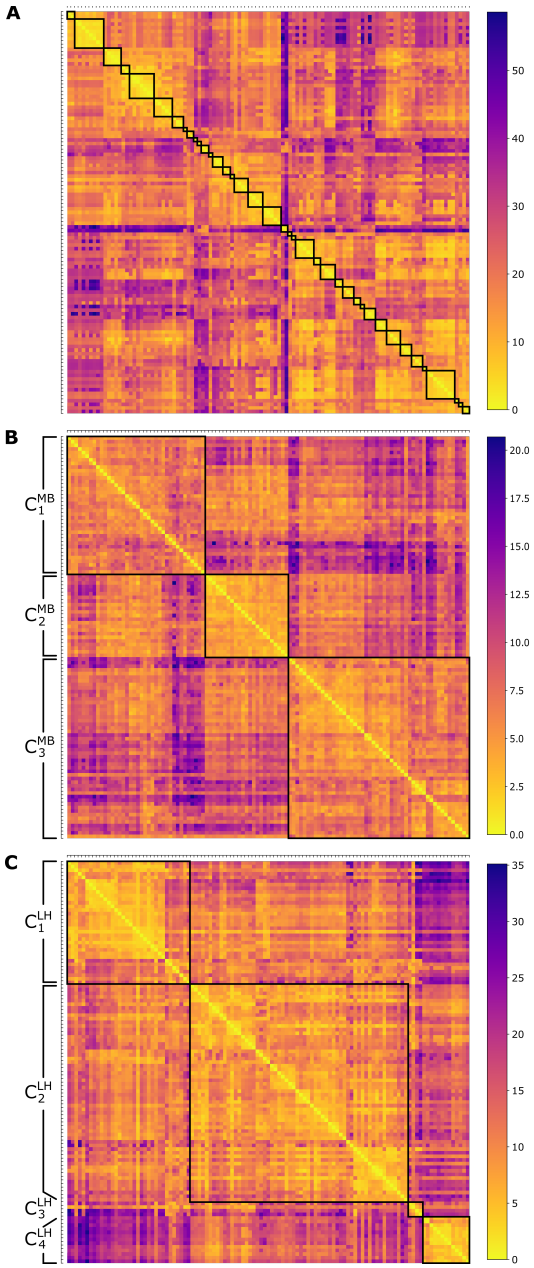


Figure S1. Three 111×111 matrices representing the inter-neuronal distances in units of μm between individual uPNs in (A) AL, (B) MB calyx, and (C) LH. The diagonal block represents each cluster (see Figure S2 for detailed information on the clustering label). The indices of PN in Figures S1A-C are sorted using the results from the clustering analysis.

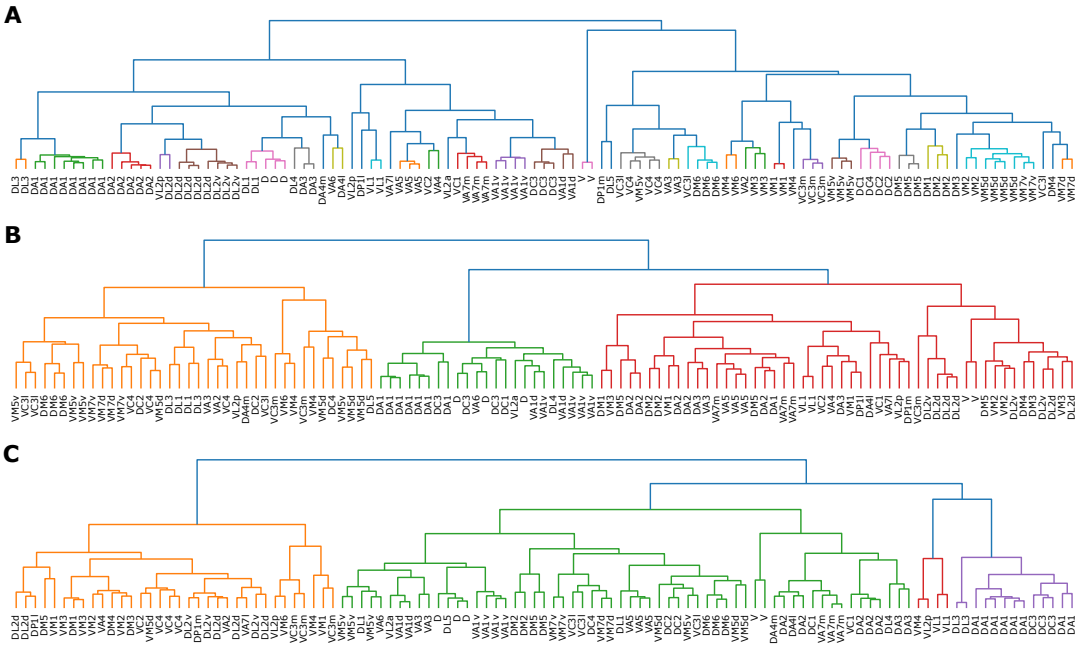


Figure S2. The dendrograms of $d_{\alpha\beta}$-based clustering on uPNs projections to (A) AL, (B) MB calyx, and (C) LH.

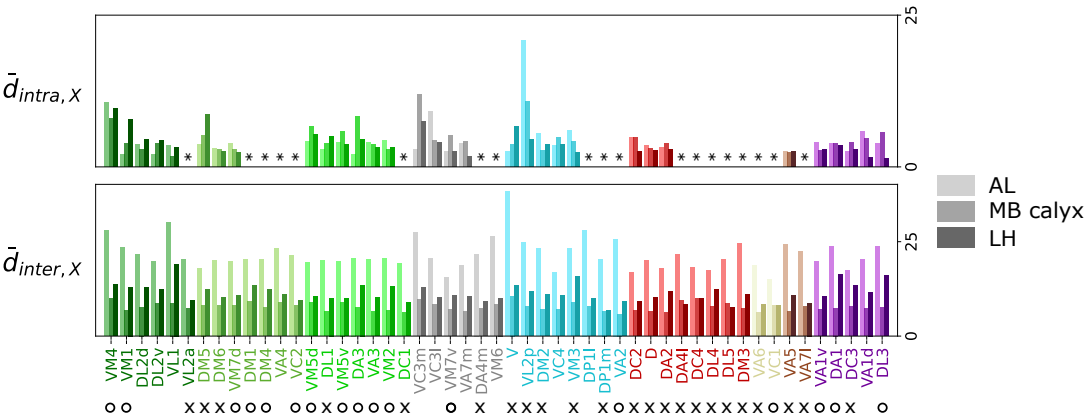


Figure S3. Comparison of the intra- ($\bar{d}_{\text{intra},X}$, degree of bundling) and inter-PN ($\bar{d}_{\text{inter},X}$, degree of packing) distances for X-th homotype in AL, MB calyx, and LH (from lighter to darker colors). The name of each homotype is color-coded based on the odor types obtained from the literature (Dark green: decaying fruit, lime: yeasty, green: fruity, gray: unknown, cyan: alcoholic fermentation, red: general bad/unclassified aversive, beige: plant matter, brown: animal matter, purple: pheromones). The homotypes are ordered based on both the odor type and intra-/inter-PN distance ratio at LH. Asterisks denote homotypes composed of a single neuron and thus the intra-PN distances are not available. O and X denote attractive and aversive odors, respectively.

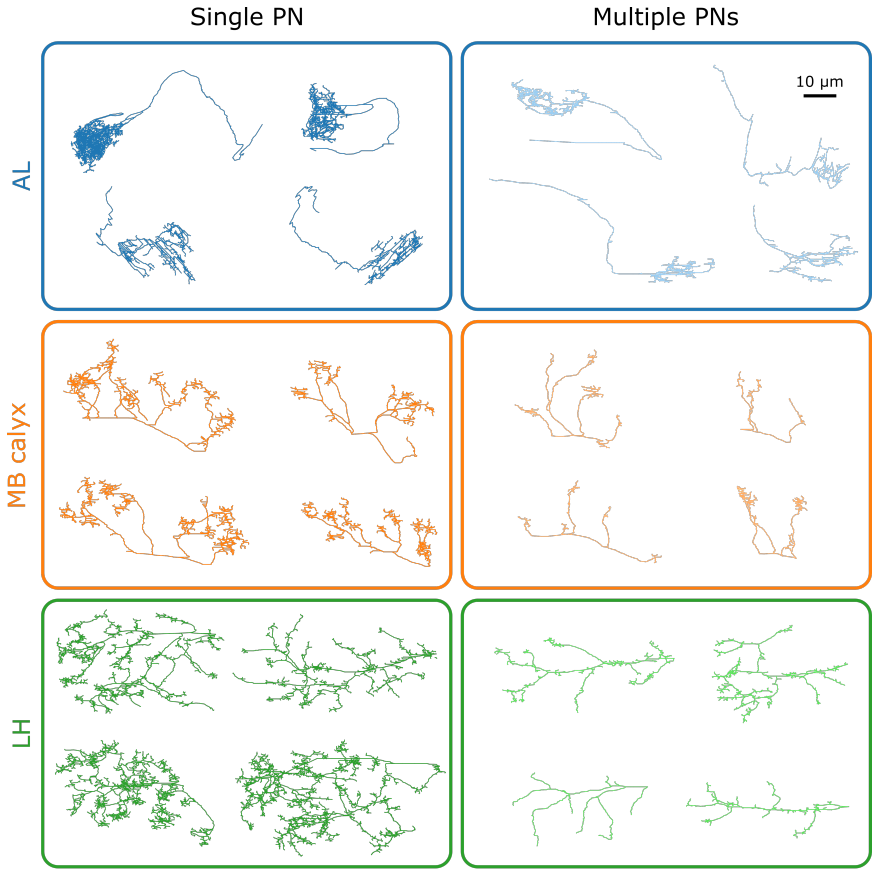


Figure S4. Selected morphologies of PN projections at each neuropil by homotypes with a single PN and homotypes with multiple PNs.

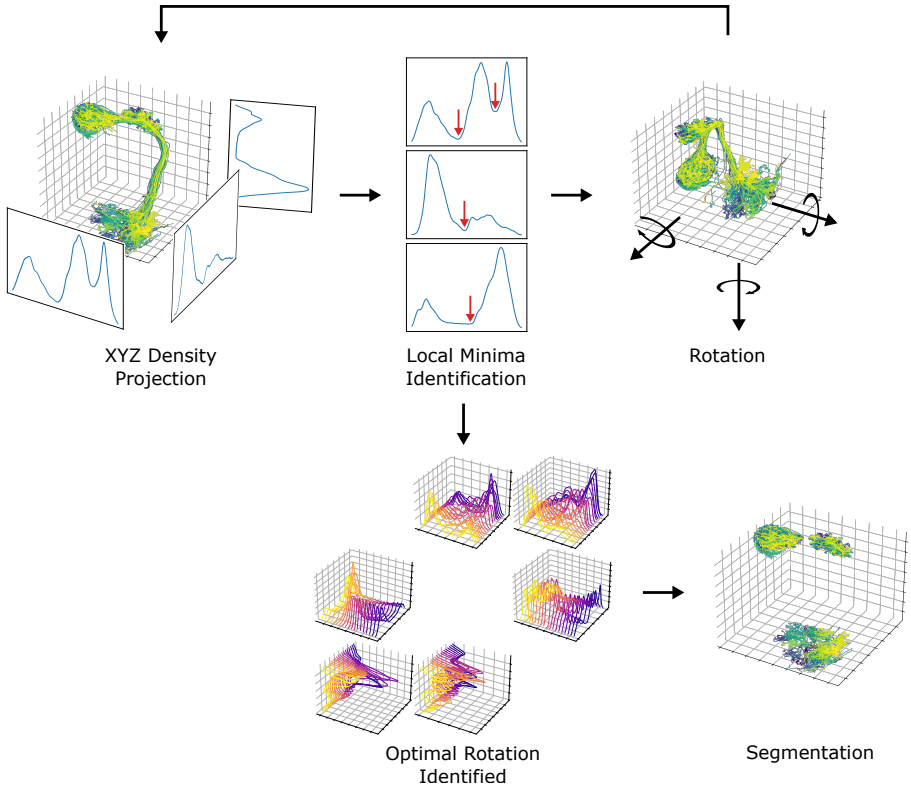
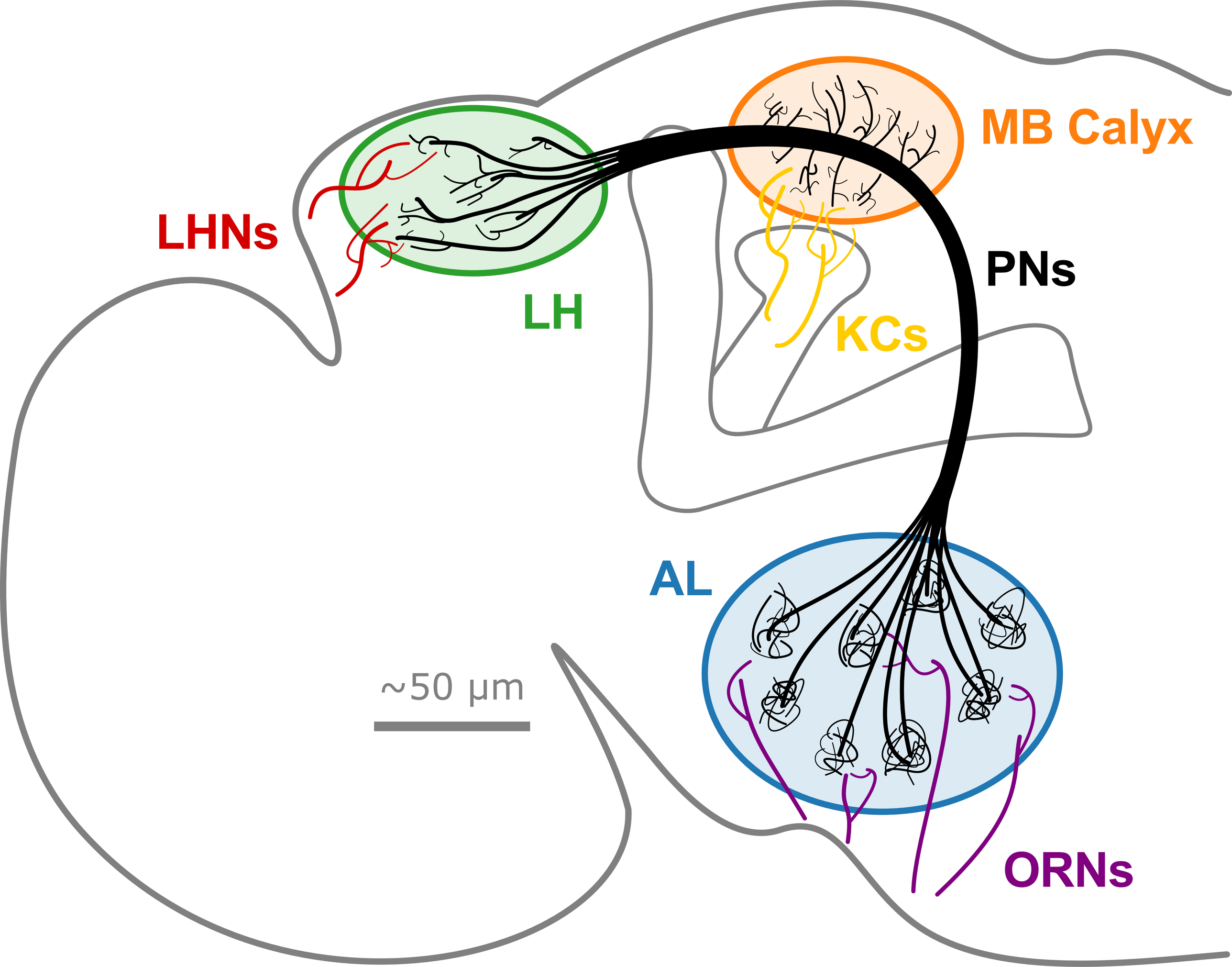
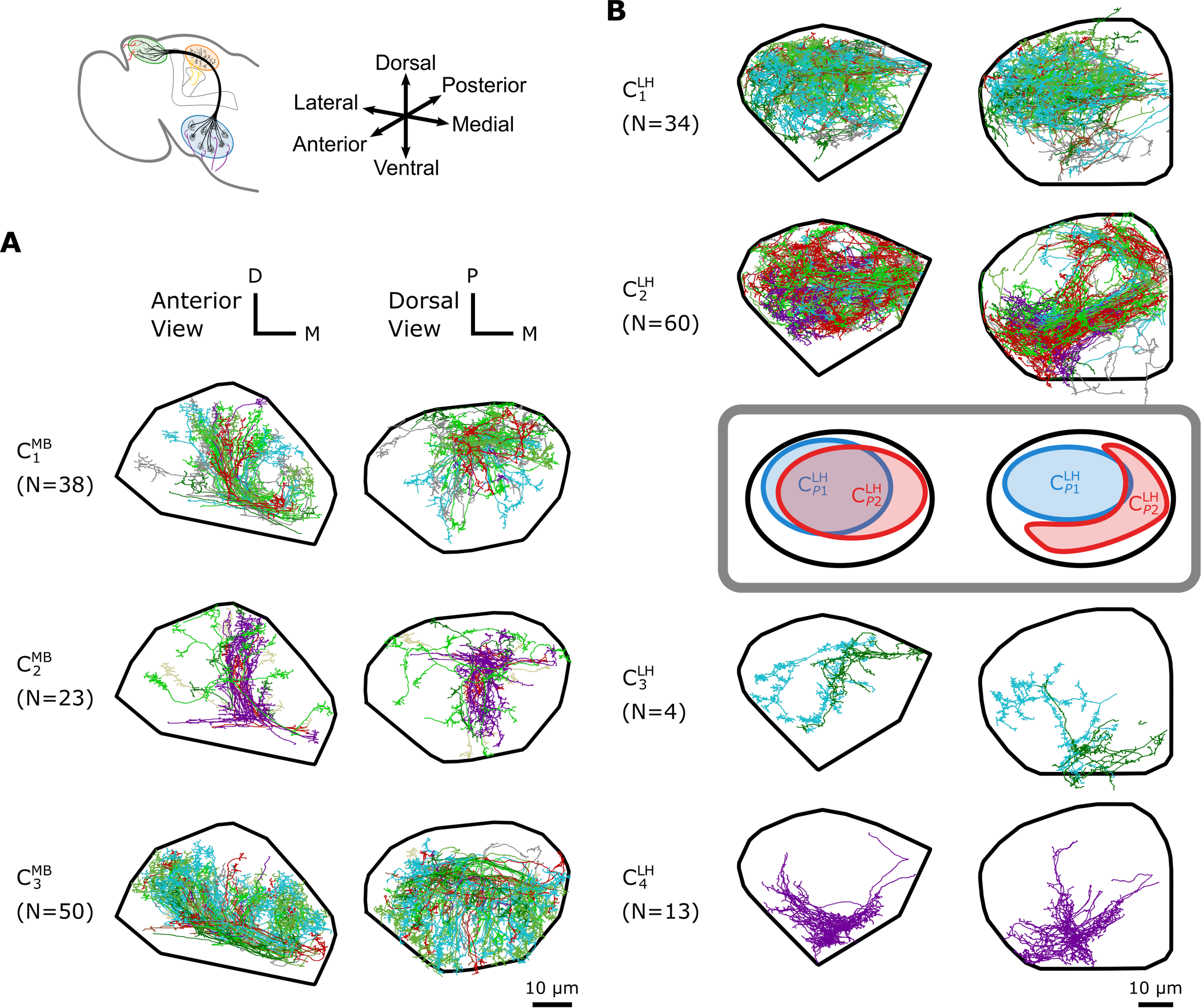
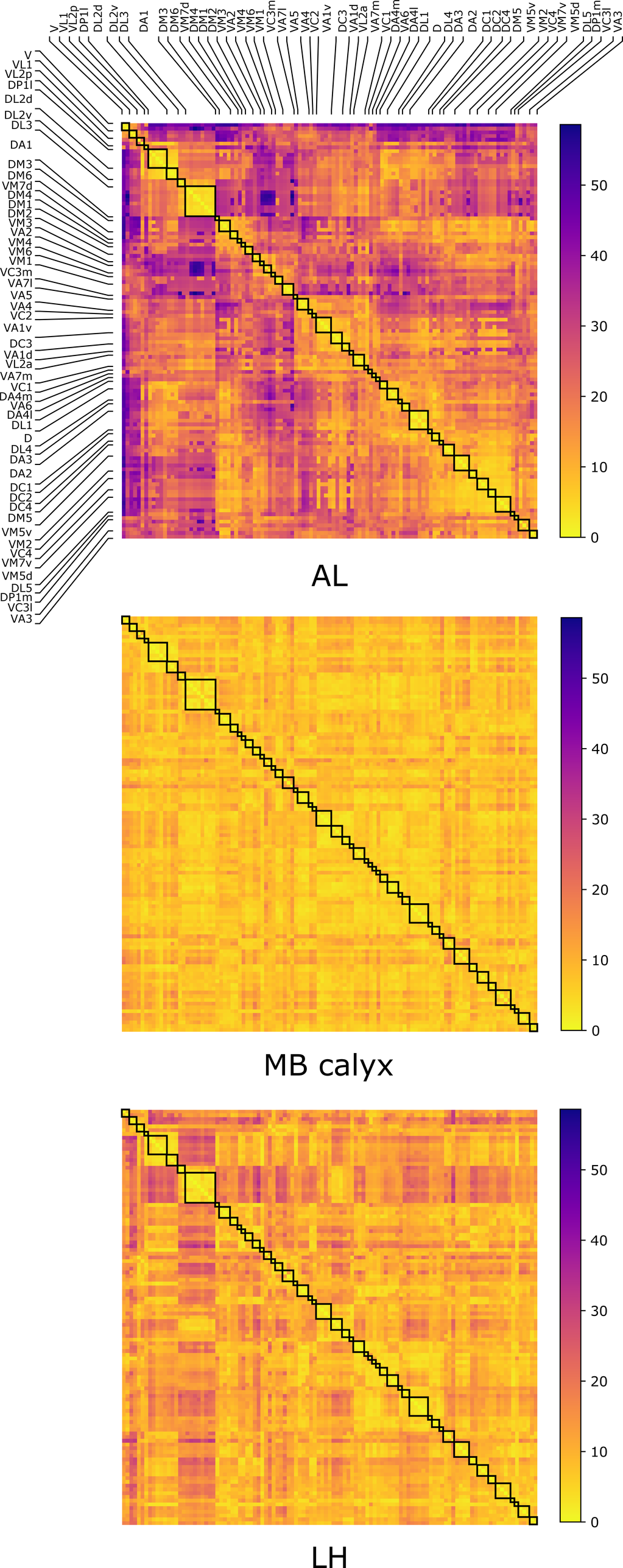


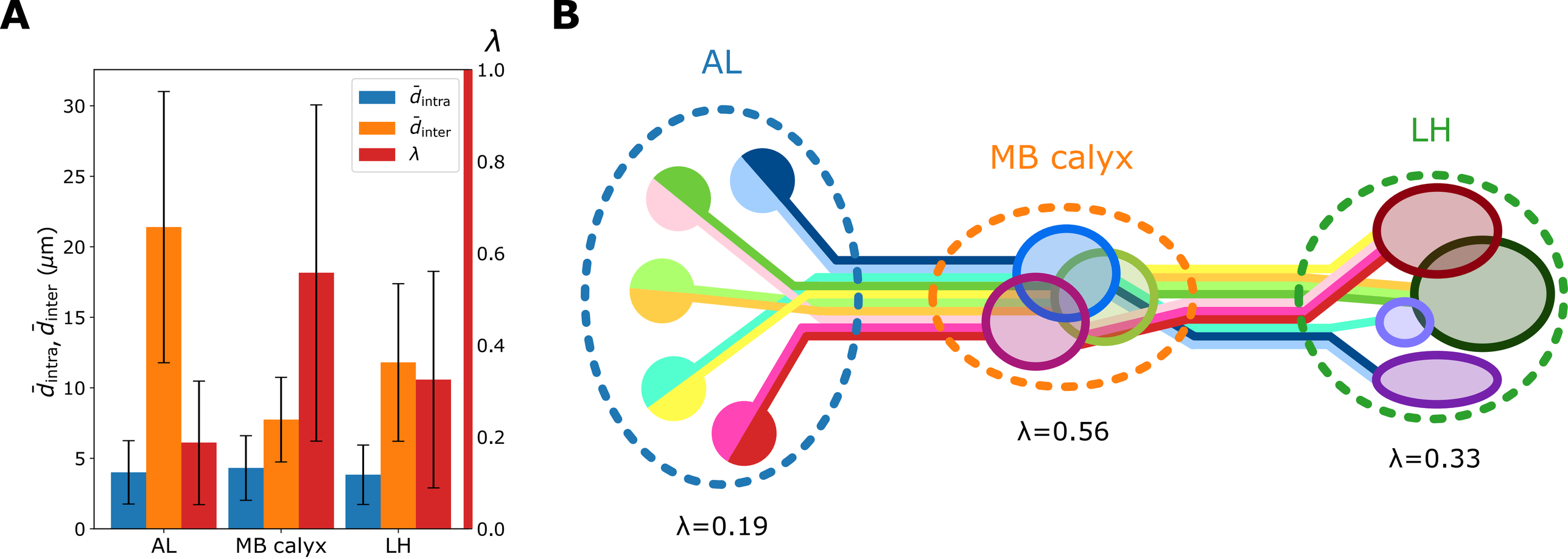
Figure S5. A diagram depicting the neuropil segmentation process. The data points from skeletal reconstruction are projected to each axes to generate distributions from which local minima are obtained. The process is repeated while rotating the PNs along each axes. A collection of histograms and corresponding local minima are surveyed to generate a set of optimal rotations and boundaries for individual neuropil. Resulting parameters are combined to form a collection of conditions to segment each neuropil.

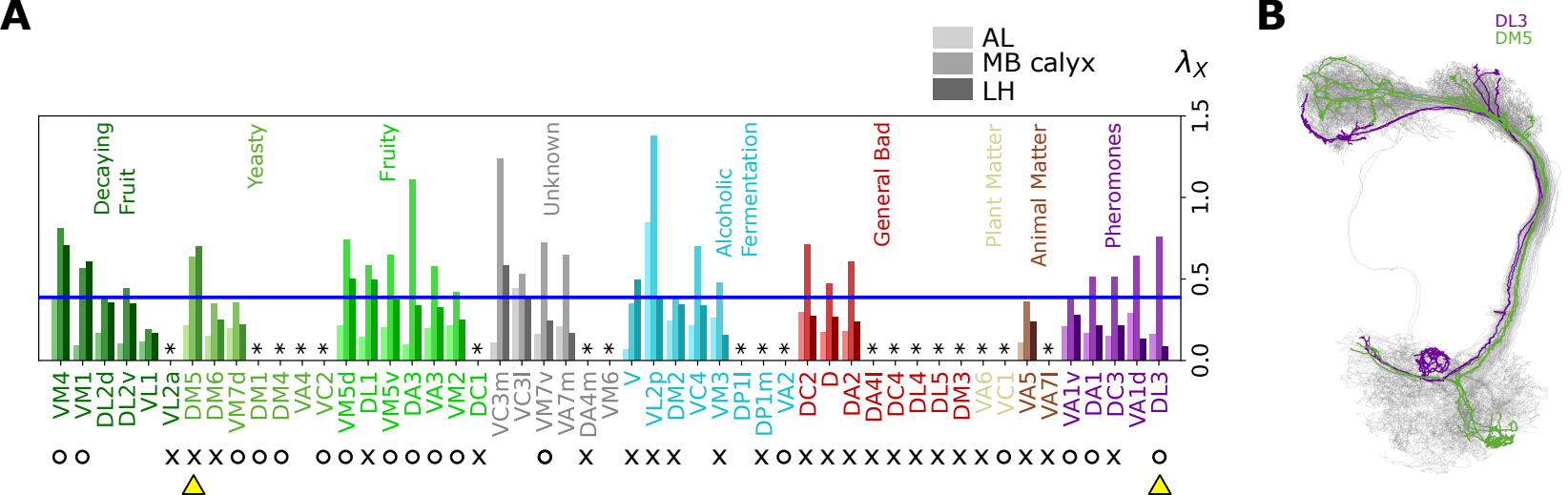
Figure S6. A plot depicting the average silhouette coefficient for different number of clusters for each neuropil. For $d_{\alpha\beta}$-based clustering, the optimal number of clusters maximizing the average silhouette coefficient is 39, 3, and 4 for AL, MB calyx, and LH, respectively.

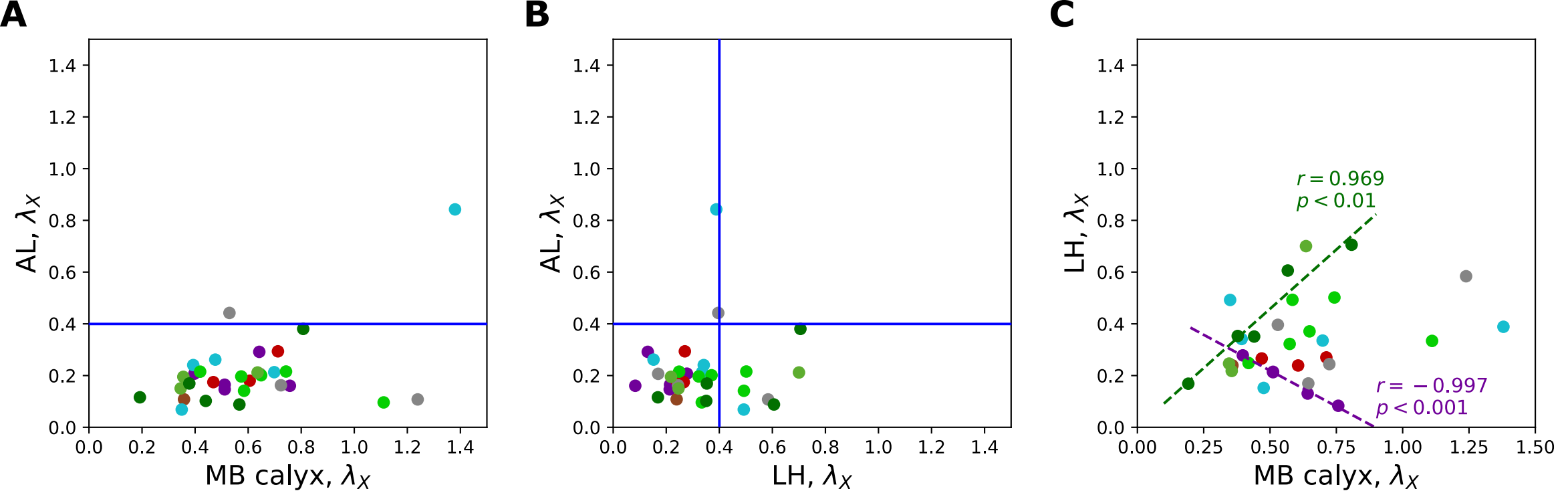

MB Calyx
LHNs
PNs
LH
KCs
AL
~50 µm
ORNs

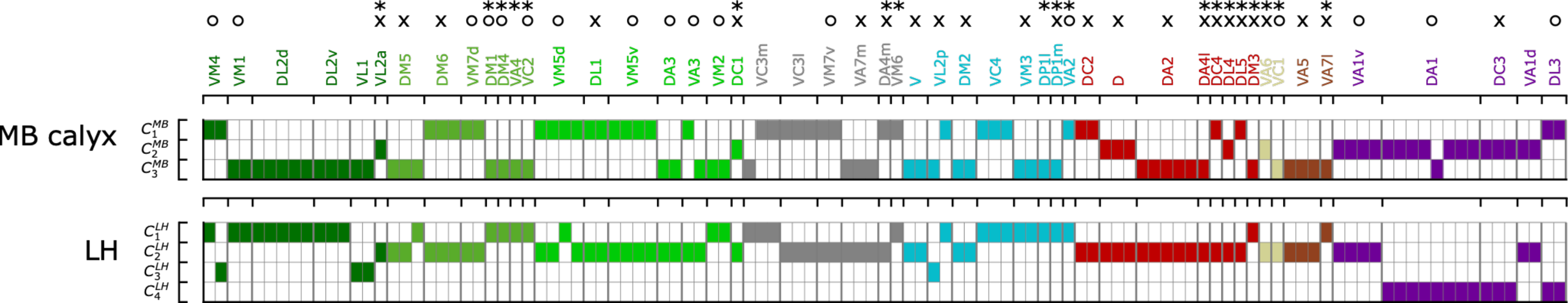

Dorsal
Posterior
Lateral
Medial
Anterior
Ventral
A
Anterior View
D
M
Dorsal View
P
M
C_1^{MB} (N=38)
C_2^{MB} (N=23)
C_3^{MB} (N=50)
10 µm
B
C_1^{LH} (N=34)
C_2^{LH} (N=60)
C_{P1}^{LH}
C_{P2}^{LH}
C_3^{LH} (N=4)
C_4^{LH} (N=13)
10 µm

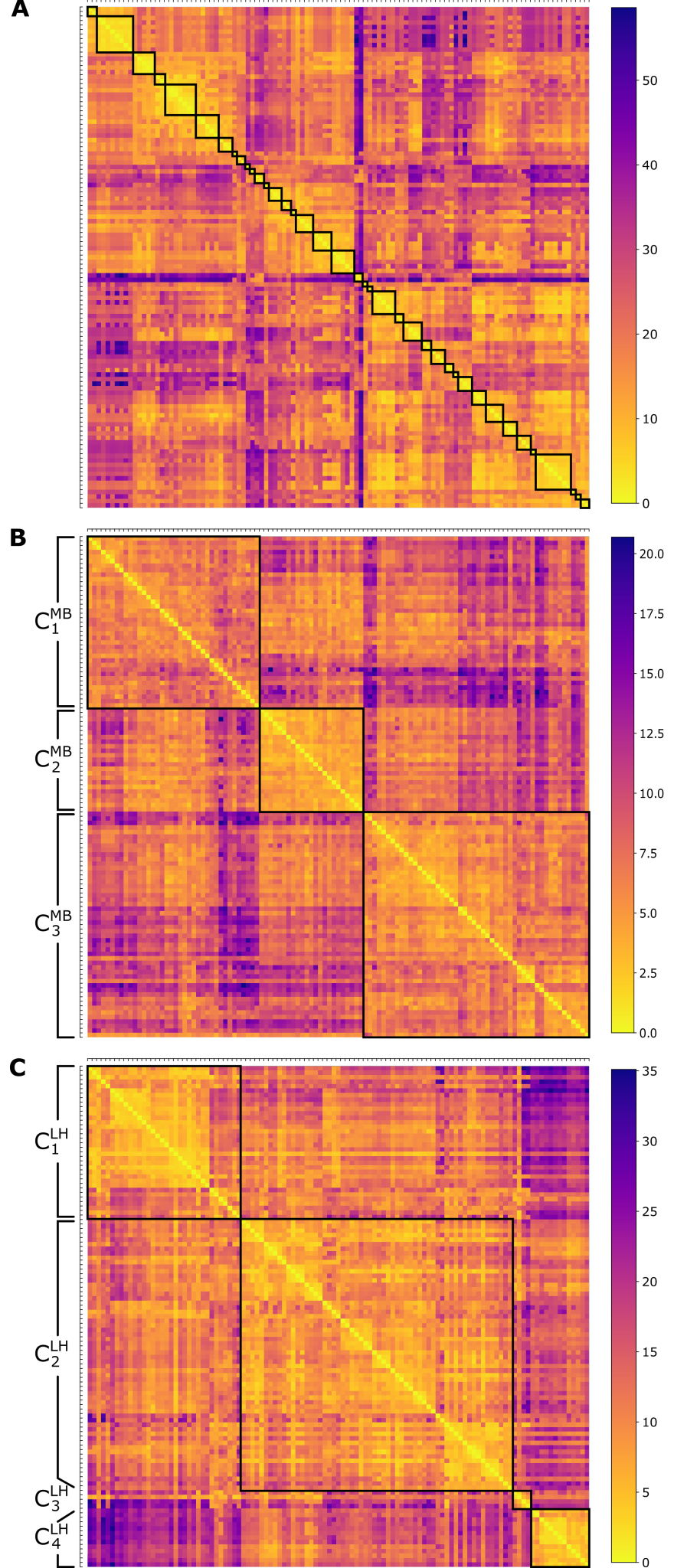

V
VL1
VL2p
DP1l
DL2d
DL2v
DL3
DA1
DM3
DM6
VM7d
DM4
DM1
DM2
VM3
VA2
VM4
VM6
VM1
VC3m
VA7l
VA5
VA4
VC2
VA1v
DC3
VA1d
VL2a
VA7m
VC1
DA4m
VA6
DA4l
DL1
D
DL4
DA3
DA2
DC1
DC2
DC4
DM5
VM5v
VM2
VC4
VM7v
VM5d
DL5
DP1m
VC3l
VA3
0
10
20
30
40
50
AL
MB calyx
LH

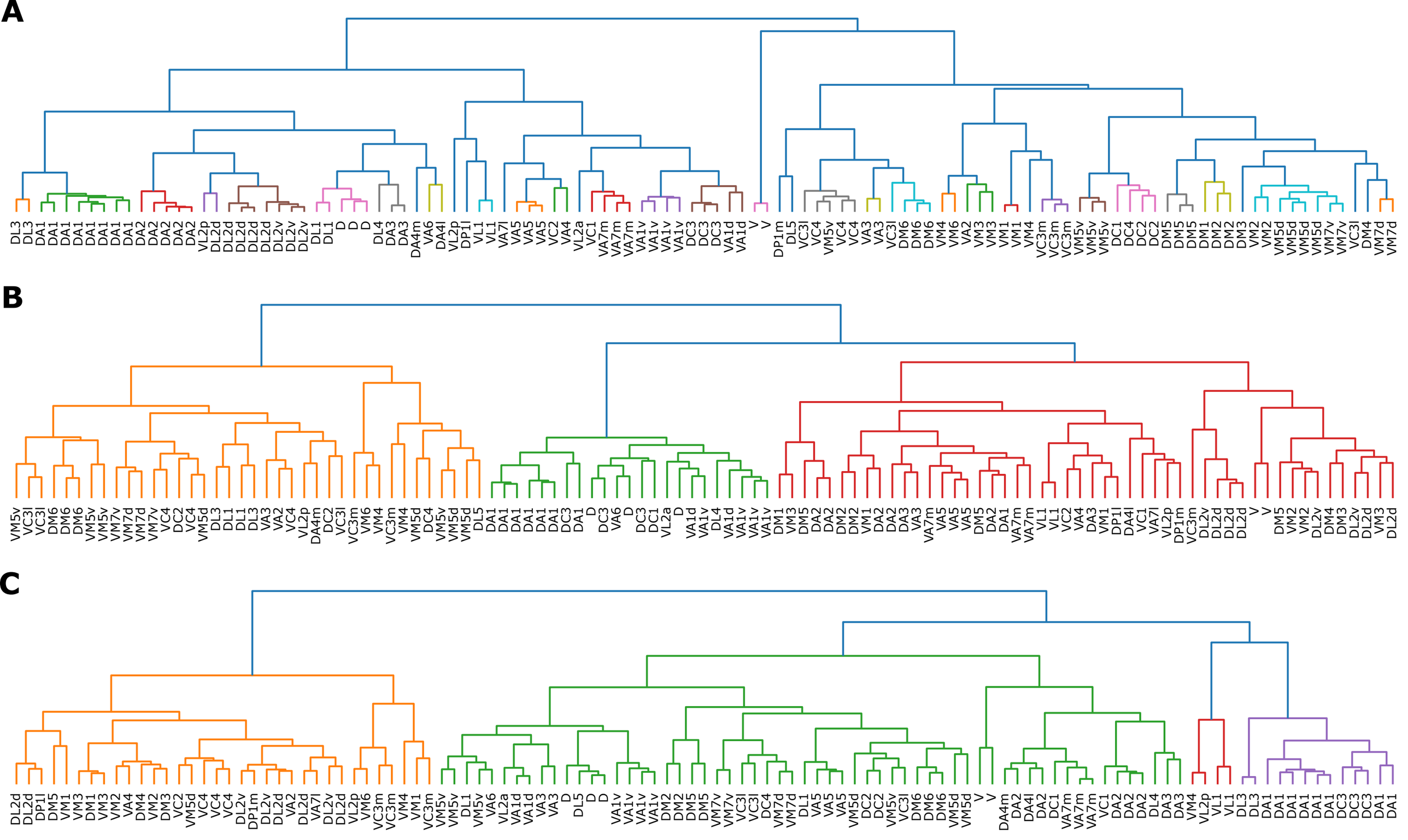

A
B
λ
1.0
0.8
0.6
0.4
0.2
0.0
30
25
20
15
10
5
0
$\bar{d}_{intra}$, $\bar{d}_{inter}$ (μm)
$\bar{d}_{intra}$
$\bar{d}_{inter}$
λ
AL
MB calyx
LH
AL
MB calyx
LH
λ=0.19
λ=0.56
λ=0.33

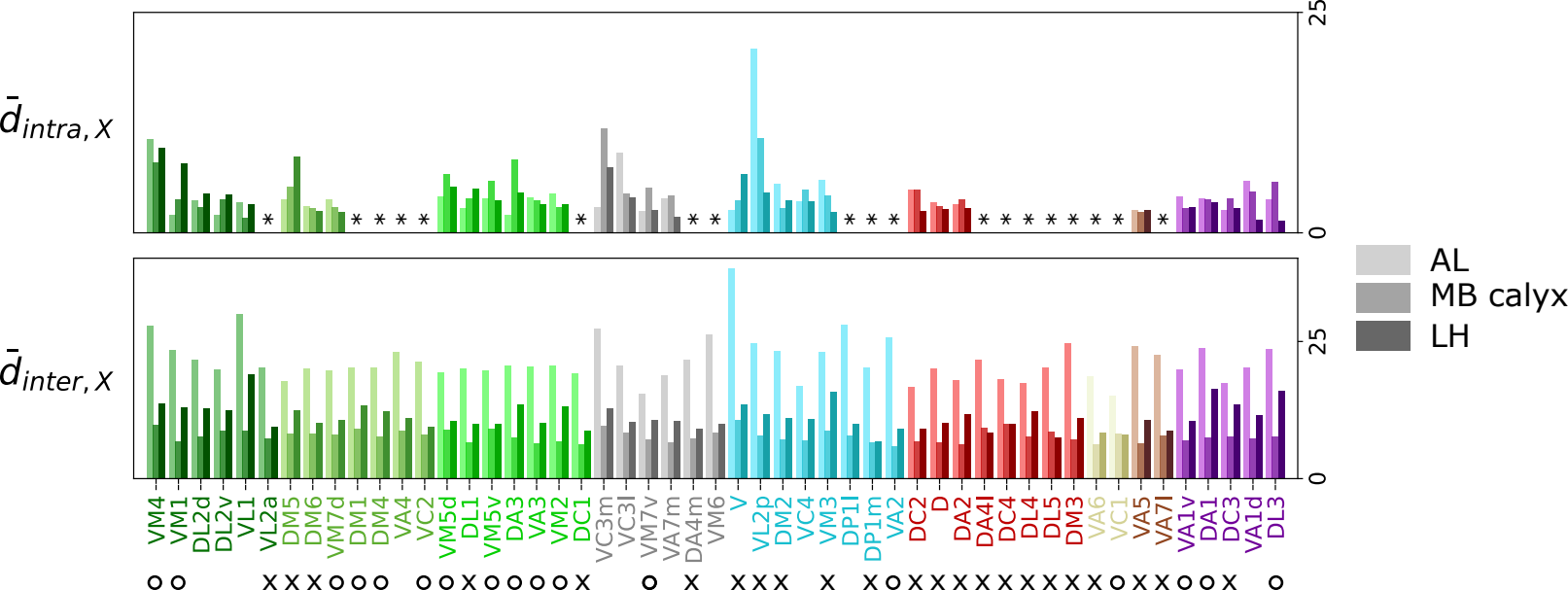

A
AL
MB calyx
LH
λ_X
1.5
1.0
0.5
0.0
Decaying Fruit
Yeasty
Fruity
Unknown
Alcoholic Fermentation
General Bad
Plant Matter
Animal Matter
Pheromones
VM4 VM1 DL2d DL2v VL1 VL2a DM5 DM6 VM7d DM1 DM4 VA4 VC2 VM5d DL1 VM5v DA3 VA3 VM2 DC1 VC3m VC3l VM7v VA7m DA4m VM6 V VL2p DM2 VC4 VM3 DP1l DP1m VA2 DC2 D DA2 DA4l DC4 DL4 DL5 DM3 VA6 VC1 VA5 VA7l VA1v DA1 DC3 VA1d DL3
B
DL3
DM5

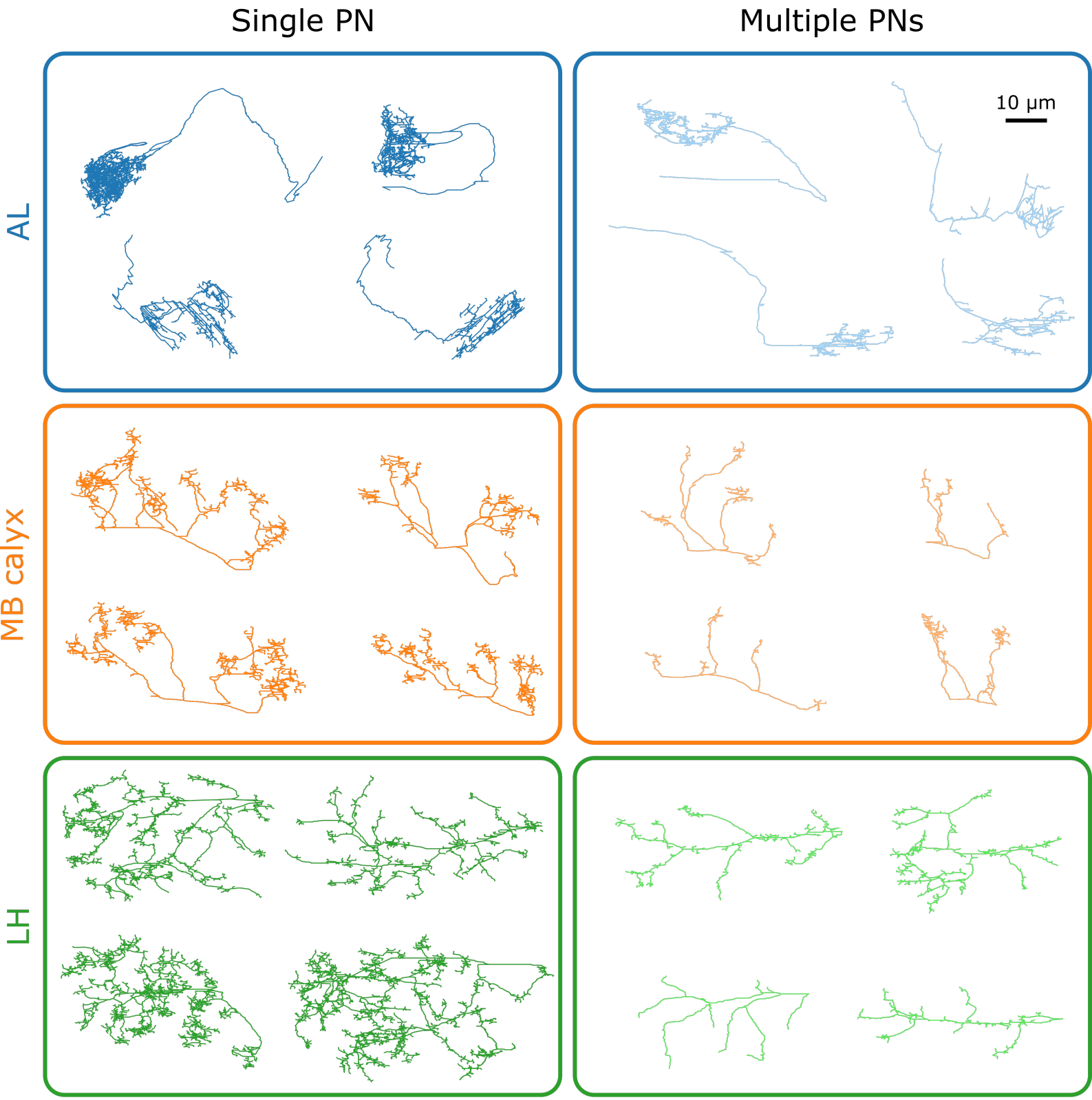

A
B
C
AL, λ_X
MB calyx, λ_X
LH, λ_X
$r = 0.969$
$p < 0.01$
$r = -0.997$
$p < 0.001$

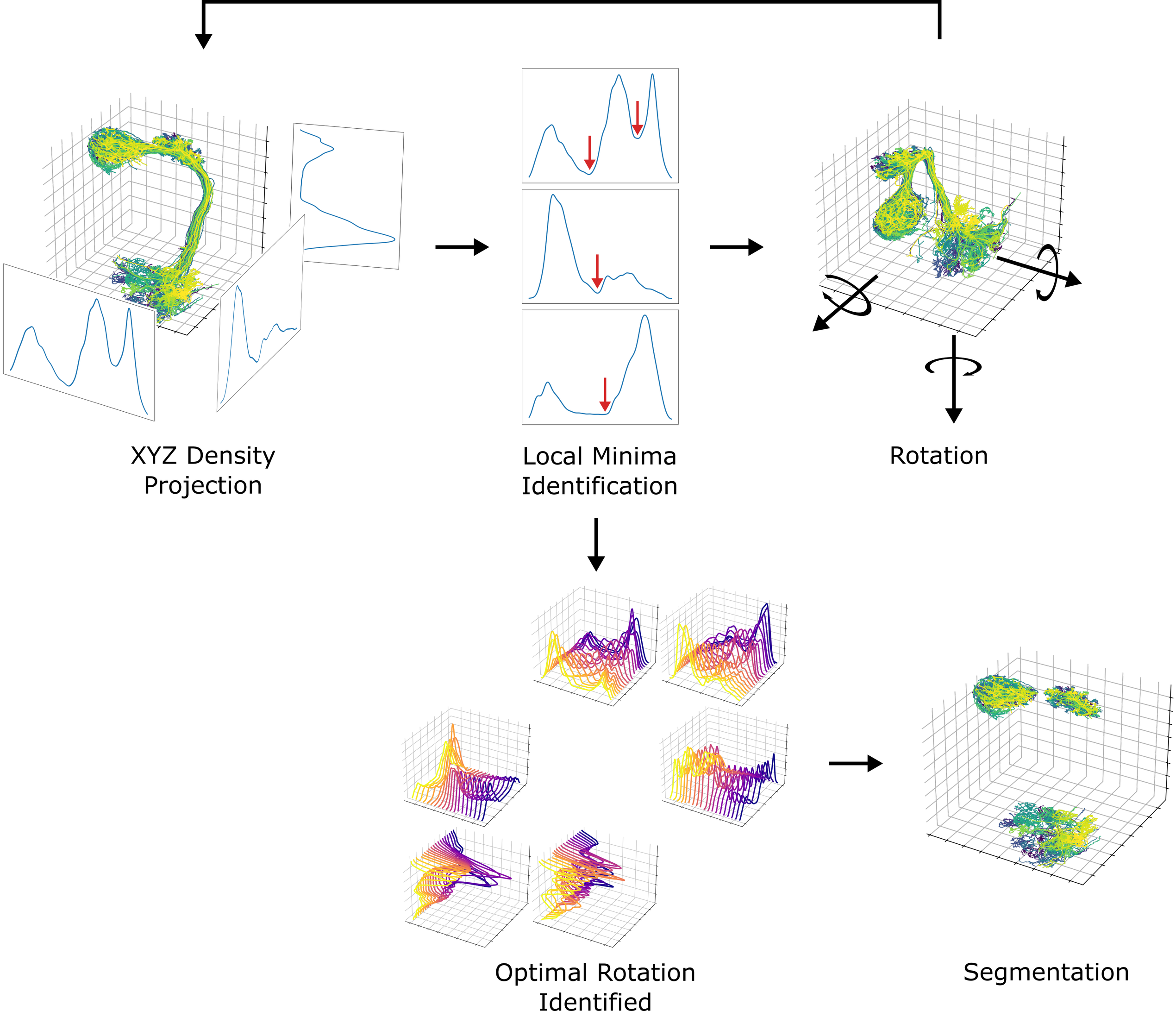

VM4
VM1
DL2d
DL2v
VL1
VL2a
DM5
DM6
VM7d
DM1
DM4
VA4
VC2
VM5d
DL1
VM5v
DA3
VA3
VM2
DC1
VC3m
VC3l
VM7v
VA7m
DA4m
VM6
V
VL2p
DM2
VC4
VM3
DP1l
DP1m
VA2
DC2
D
DA2
DA4l
DC4
DL4
DL5
DM3
VA6
VC1
VA5
VA7l
VA1v
DA1
DC3
VA1d
DL3
MB calyx
LH


A
50
40
30
20
10
0
B
C_1^{MB}
C_2^{MB}
C_3^{MB}
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
C
C_1^{LH}
C_2^{LH}
C_3^{LH}
C_4^{LH}
35
30
25
20
15
10
5
0


A
DL3 DL3 DA1 DA1 DA1 DA1 DA1 DA1 DA2 DA2 DA2 DA2 VL2p DL2d DL2d DL2d DL2d DL2v DL2v DL2v DL1 DL1 D D D DL4 DA3 DA3 DA4m VA6 DA4l VL2p DP1l VL1 VL1 VA7l VA5 VA5 VA5 VC2 VA4 VL2a VC1 VA7m VA7m VA7m VA1v VA1v VA1v VA1v DC3 DC3 DC3 VA1d VA1d V V DP1m DL5 VC3l VC4 VM5v VC4 VC4 VA3 VA3 VC3l DM6 DM6 DM6 VM4 VM6 VA2 VM3 VM3 VM1 VM1 VM4 VC3m VC3m VC3m VM5v VM5v VM5v DC1 DC4 DC2 DC2 DM5 DM5 DM5 DM1 DM2 DM2 DM3 VM2 VM2 VM5d VM5d VM5d VM5d VM7v VM7v VC3l DM4 VM7d VM7d
B
VM5v VC3l VC3l DM6 DM6 DM6 VM5v VM5v VM7v VM7d VM7d VM7v VC4 DC2 VC4 VM5d DL3 DL1 DL1 DL3 VA3 VA2 VC4 VL2p DA4m DC2 VC3l VC3m VM6 VM4 VC3m VM4 VM5d DC4 VM5v VM5d VM5d DL5 DA1 DA1 DA1 DA1 DA1 DC3 DA1 D DC3 VA6 D DC3 DC1 VL2a D VA1d VA1v DL4 VA1d VA1v VA1v VA1v DM1 VM3 DM5 DA2 DA2 DM2 DM2 VM1 DA2 DA2 DA3 VA3 VA7m VA5 VA5 VA5 DM5 DA2 DA1 VA7m VA7m VL1 VL1 VC2 VA4 DA3 VM1 DP1l DA4l VC1 VA7l VL2p DP1m VC3m DL2v DL2d DL2d DL2d V V DM5 VM2 VM2 DL2v DM4 DM3 DL2v DL2d VM3 DL2d
C
DL2d DL2d DP1l DM5 VM1 VM3 DM1 VM3 VM2 VA4 DM4 VM2 DM3 VC2 VM5d VC4 VC4 VC4 DL2v DP1m DL2v DL2d VA2 DL2d VA7l DL2v DL2d VL2p VM6 VC3m VC3m VM4 VM1 VC3m VM5v VM5v DL1 VM5v VA6 VL2a VA1d VA1d VA3 VA3 D DL5 D D VA1v VA1v VA1v VA1v DM2 DM2 DM5 DM5 VM7v VM7v VC3l VC3l DC4 VM7d VM7d DL1 VA5 VA5 VA5 VM5d DC2 DC2 VM5v VC3l DM6 DM6 DM6 VM5d VM5d V V DA4m DA2 DA4l DA2 DC1 VA7m VA7m VA7m VC1 DA2 DA2 DA2 DL4 DA3 DA3 VM4 VL2p VL1 VL1 DL3 DL3 DA1 DA1 DA1 DA1 DA1 DA1 DA1 DC3 DC3 DC3 DA1 DA1


$\bar{d}_{intra,X}$
$\bar{d}_{inter,X}$
25
0
AL
MB calyx
LH
VM4
VM1
DL2d
DL2v
VL1
VL2a
DM5
DM6
VM7d
DM1
DM4
VA4
VC2
VM5d
DL1
VM5v
DA3
VA3
VM2
DC1
VC3m
VC3l
VM7v
VA7m
DA4m
VM6
V
VL2p
DM2
VC4
VM3
DP1l
DP1m
VA2
DC2
D
DA2
DA4l
DC4
DL4
DL5
DM3
VA6
VC1
VA5
VA7l
VA1v
DA1
DC3
VA1d
DL3


Single PN
Multiple PNs
10 µm
AL
MB calyx
LH


XYZ Density Projection
Local Minima Identification
Rotation
Optimal Rotation Identified
Segmentation

Average Silhouette Coefficients
0.35
0.30
0.25
0.20
0.15
0
20
40
60
Number of Clusters
AL
MB calyx
LH