Developmental Biology 533 (2026) 24–37

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/locate/developmentalbiology

Original research article

Gene expression cartography of a developing neuronal structure

Leonardo Tadini [a], Lilia Younsi [b], Isabel Holguera [a], Félix Simon [a], Maximilien Courgeon [c], Nikolaos Konstantinides [a,*]

[a] Université Paris Cité, CNRS, Institut Jacques Monod, F-75013 Paris, France
[b] Université Paris Cité, CNRS, INSERM, Institut Cochin, Paris, France
[c] Aix-Marseille Université & CNRS, IBDM-UMR7288 & Turing Centre for Living Systems, Marseille, France

ARTICLE INFO

Keywords:
Spatial transcriptomics
Drosophila optic lobe
Novosparc
Single-cell mRNA sequencing
Neurodevelopmental 3D-atlas
Developing nervous system

ABSTRACT

Brains are complex structures comprising thousands to billions of neurons that belong to thousands of different neuronal types. These neurons often come from different progenitor domains and have very diverse developmental histories, yet they need to find their precise locations in the brain and integrate into appropriate circuits. While a large number of brain single-cell mRNA sequencing atlases have described the neuronal part list of the brain, spatial transcriptomic studies have lagged behind in offering the spatial information that is essential to understand brain structures. Here, we use a gene expression cartography algorithm called Novosparc to reconstruct the spatial distribution of gene expression and cell type localization in a complex, yet tractable, developing brain structure, the *Drosophila* optic lobe. We generate a three-dimensional atlas of this structure (https://larva3dnovosparc.ijm.fr); this atlas allowed us to identify spatially patterned transcription factors that define neuronal types. Importantly, we identify caveats in the algorithm and we discuss limitations of the current implementation that should guide future algorithmic improvements. Altogether, this work provides an invaluable tool to test gene expression patterns and paves the way for the generation of three-dimensional atlases of more complex brain structures, which will enhance our understanding of how neurons with diverse developmental lineages can integrate to form a functional brain.

1. Introduction

The visual part of the *Drosophila* brain, *i.e.* the optic lobe, is a complex brain structure composed of ~38,500 neurons (Apitz and Salecker, 2014; Matsliah et al., 2024; Nériec et al., 2016) that belong to, at least, 150 different neuronal types (Kurmangaliyev et al., 2020; Özel et al., 2021) (although this number is likely to be much larger based on the recent connectomic data (Matsliah et al., 2024; Nern et al., 2025)), as well of ~5000 glia of more than 15 different types (Ferreira and Desplan, 2023; Lago-Baldaia et al., 2023). These neurons form synapses with each other in the four neuropils of the optic lobe, which are called the lamina, medulla, lobula, and lobula plate (Fig. 1A). The morphology and the transcriptome of the different cell types have been described extensively over the last thirty years (Fischbach and Dittrich, 1989; Konstantinides et al., 2018; Özel et al., 2021; Simon and Konstantinides, 2021), making the fly optic lobe one of the most comprehensively described adult neuronal structures.

Like the more complex vertebrate brain structures, the optic lobe neurons come from a variety of different developmental lineages during the third instar larval stage.

(a) Most of the neurons are born from a neuroepithelial sheet called the Outer Proliferation Center (OPC), which gives rise to the lamina neurons in the lateral side and the medulla neurons in the medial side (Fig. 1B) (Nériec et al., 2016). The OPC is further divided in six spatial domains defined by the expression of the transcription factors, Vsx, Optix, and Rx, subdivided into dorsal and ventral by the expression of salm and disco, respectively (Valentino and Erclik, 2022). Notably, the expression of signaling molecules further subdivide the epithelium into twelve different domains (Malin et al., 2024).
(b) Most of the lobula cortex neurons come from the Wg-expressing tips of the OPC (Bertet et al., 2014; El-Danaf et al., 2025), as well as another epithelial structure (the Inner Proliferation Center - IPC), which also gives rise to the most abundant neuronal cell

* Corresponding author.
E-mail address: nikos.konstantinides@ijm.fr (N. Konstantinides).

https://doi.org/10.1016/j.ydbio.2026.01.013
Received 8 September 2025; Received in revised form 19 January 2026; Accepted 27 January 2026
Available online 31 January 2026

type of the lobula plate (the T4/T5 neurons), as well as other C/T neurons (Apitz and Salecker, 2015).

(c) The rest of the lobula and lobula plate neurons come from the central brain neuroblasts (El-Danaf et al., 2025), which will not be discussed further in this study.

Besides the spatial patterning of the OPC neuroepithelium, the neuroblasts derived from the OPC are also patterned temporally (Konstantinides et al., 2022; Zhu et al., 2022). The combination of spatial and temporal patterning in combination with a Notch binary cell fate decision during the terminal division of an intermediate progenitor (the ganglion mother cell – GMC) is responsible for the generation of the different cell types of the medulla (Fig. 1B and C), which compose ~70 % of the neuronal diversity of the optic lobe (Apitz and Salecker, 2014).

Despite this complexity, the way neurodevelopment proceeds in the OPC leads to a very defined spatial organization of the brain: first, spatial patterning of the epithelium leads to the division of the OPC neurons into at least six distinct domains. Second, the neuroblasts are born from this neuroepithelial sheet in a wave that begins medially and extends laterally (Sato et al., 2016; Yasugi et al., 2010); this leads to the presence of neuroblasts of gradually different age along the mediolateral axis (Konstantinides et al., 2022; Li et al., 2013; Nériec et al., 2016). Therefore, despite its complexity, the developing optic lobe displays a remarkable spatial organization that has been extensively described (Ngo et al., 2017).

In an earlier study, we generated a single-cell sequencing atlas of the

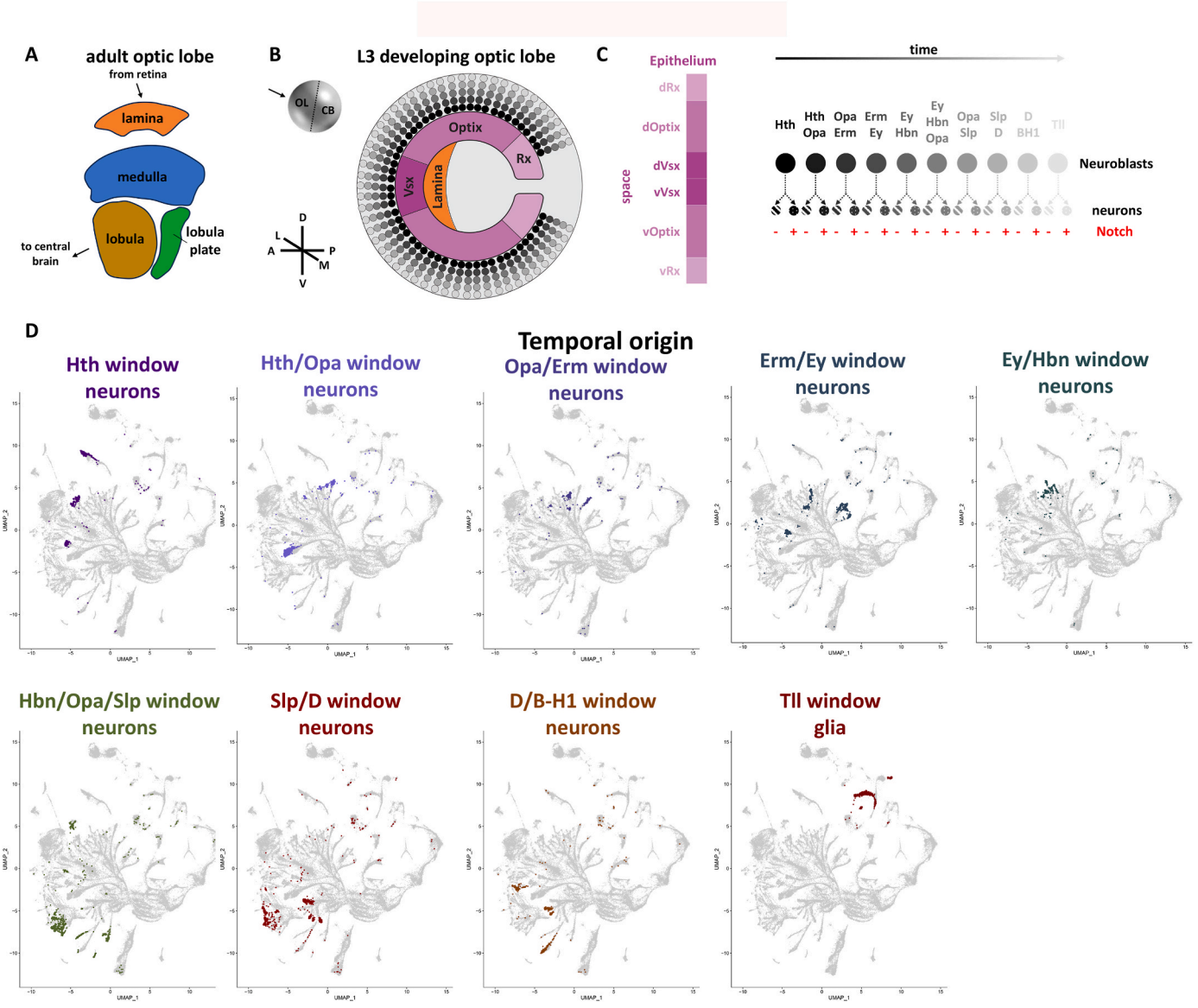


Fig. 1. The developing *Drosophila* optic lobe

(A) Schematic of the adult optic lobe. It consists of four neuropils, the lamina, medulla, lobula, and lobula plate. (B) Schematic of the developing optic lobe in the late third instar larval stage, as seen from a lateral view. The medulla neuroblasts (shades of grey) come from the medial side of the main Outer Proliferation Center (magenta), while the lamina ones (orange) come from the OPC lateral side. The OPC epithelium and neuroblasts are patterned spatially (shades of magenta) and temporally (shades of grey), respectively, by transcription factor expression as shown in Fig. 1C. (C) The main OPC gives rise to the majority of neuronal types of the optic lobe using an intersection of three mechanisms: first, the spatial patterning of the epithelium in six domains, second, the temporal patterning of neuroblasts in, at least, eleven temporal windows, and, third, a Notch-driven binary cell fate decision upon the division of the intermediate progenitors. (D) UMAP plots showing the localization of the different neuronal types that come from the respective temporal windows, as annotated before (Konstantinides et al., 2022). In this two-dimensional representation, neurons with the same temporal origin tend to cluster together. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

third instar larval developing optic lobe (Konstantinides et al., 2022) (Supplementary Fig. S1A). By sequencing ~40,000 single cells (1X coverage of the brain) and using trajectory analysis algorithms, we were able to identify most of the temporal transcription factors that are expressed in medulla neuroblasts. While this was not part of the study, it did not escape our attention that the two-dimensional representation of the single-cell sequencing atlas (in the form of a UMAP plot) retained clear spatial information, *i.e.* the lamina and the medulla were separated along the UMAP1 axis, while the lobula plate was distinguished along the UMAP2 axis (Supplementary Fig. S1A). Moreover, neuronal types with similar temporal origins were clustered together (Fig. 1D), while this was not immediately obvious for neurons from similar spatial origins (Supplementary Figure S1B) likely because of a more complex spatial origin of neuronal types as many cell types are generated from more than one spatial domain (Malin et al., 2024; Simon et al., 2025). While this information proved helpful, it lags considerably from a real three-dimensional spatial transcriptomic atlas.

To reconstruct the spatial localization of neuronal cell types and the spatial gene expression of the developing fly optic lobe, we used Novosparc (Moriel et al., 2021; Nitzan et al., 2019), a gene cartography tool that has been previously used to spatially reconstruct single-cell mRNA sequencing atlases of tissues such as the stage 5 *Drosophila* embryo (Nitzan et al., 2019) and the mid-planula stage of *Nematostella* (He et al., 2023). Novosparc relies on the hypothesis that physically neighbouring cells share a similar gene expression profile. Novosparc reconstructs spatial gene expression by leveraging a geometric framework that maps single-cell transcriptomic data onto spatial coordinates, preserving the tissue's structural organization. It uses optimal transport-based algorithms to infer spatial positions of cells based on known spatial reference information.

Here, we used Novosparc to reconstruct in three dimensions gene expression emanating from the developing fly optic lobe single-cell mRNA sequencing atlas. To achieve this, we used pre-existing knowledge to divide the brain into different regions based on their developmental lineage (Ngo et al., 2017) to generate a three-dimensional object representing the developing brain in Blender, a free and open-source platform for 3D modeling and rendering (https://www.blender.org). We used existing single-cell mRNA sequencing datasets (Konstantinides et al., 2022; Simon et al., 2025) to identify spatially and temporally informative genes that we used as landmarks while running the algorithm. We validated the performance of the algorithm by confirming the expression of the landmark genes, as well as reconstructing the expression of genes for which the algorithm didn't have any information. We validated their reconstructed expression using immunohistochemistry, existing Gal4 lines and Hybridization Chain Reaction Fluorescence *in situ* hybridization (HCR-FISH) (Tsuneoka and Funato, 2020). Finally, we used the reconstructed gene expression to predict temporally and spatially regulated neuronal type-defining transcription factors. Our work generates a powerful tool for reconstructing gene expression in a developing brain, sets a pathway for the use of such approaches for more complex brains and highlights the importance that developmental lineage plays in their generation.

2. Results

To apply Novosparc, we required (a) a three-dimensional spatial framework onto which single cell transcriptomes could be mapped and (b) a set of marker genes to serve as spatial landmarks for this projection. To construct this framework we first generated a regionalized three-dimensional spatial reference that enabled precise alignment of marker gene expression.

2.1. Regionalization of the three-dimensional structure

As discussed earlier, the different neuronal sets are born from different neuronal lineages (OPC, IPC). Therefore, while these different neuronal sets are adjacent to each other, their transcriptomes differ significantly, reflecting their different developmental histories. Therefore, we first generated a three-dimensional structure corresponding to the developing optic lobe that reflected these different parts, using Blender. This was a half-sphere consisting of 41,287 vertices - corresponding to a similar number of cells in the L3 optic lobe (~38,500) and the respective single-cell mRNA sequencing dataset (41,392) (Fig. 2).

To regionalize this structure, we relied on a 3D model of the developing optic lobe that was published by Ngo et al., 2017 and is based on confocal imaging z-projections of fixed tissue (Fig. 2A). Based on this model, we divided the half-sphere into nine different regions: a) the OPC epithelium, b) the OPC neuroblasts, c) the IPC epithelium, d) the IPC neuroblasts, e) the posterior medulla cortex neurons (T and C cells), f) the lamina neurons (L cells), g) the medulla neurons, h) the lobula plate neurons (T4/T5 cells), and i) the lobula neurons (Fig. 2B). In addition to these regions, we further added three more regions for j) surface glia, k) GMCs (Fig. 2C), and l) lamina wide-field neurons (Lawf1 and Lawf2 – the only migrating neurons of the optic lobe (Suzuki et al., 2016a)), which were not represented in the 3D model.

As described earlier, different parts of the developing brain (such as the OPC) display an even finer spatial structure. Therefore, we further subdivided some of these twelve primary regions into subregions.

- We divided the OPC epithelium, neuroblasts, GMCs, and neurons into six spatial domains, corresponding to the dorsal and ventral Vsx (which account for 25 % of the epithelium), Optix (50 %), and Rx (25 %) domains. The respective size of each spatial domain was decided based on the actual size of these domains (Malin et al., 2024) (Figs. 1B and 2D).
- We further divided the OPC neuroblasts, GMCs, and progeny into 10 equally sized temporal domains, which correspond to the 10 temporal windows that we identified earlier (Konstantinides et al., 2022) (Fig. 1B and C and 2E), after consolidating the two Tll-expressing windows, one of which lacks identified progeny.
- We divided the posterior medulla cortex neurons into an equal number of intermingled C and T cells (Apitz and Salecker, 2018) (Fig. 2F),
- and the Lawf neurons into an equal number of Lawf1 and Lawf2 cells (Fig. 2F').

As a result, our final 3D reference model consists of 41,287 vertices grouped into 245 anatomically defined regions, each annotated with the binarized expression pattern of 43 landmark genes (Supplementary Table 1). We then asked whether the cell number of each of these domains corresponds to the cell number from the sc-mRNA-seq dataset. For this purpose, we assigned each of the sc-mRNA-seq atlas clusters to one of the 245 regions and compared the cell number of each of the regions in the Blender model and the estimated cell number from the sc-mRNA-seq dataset (Supplementary Table 1). With the exception of two domains (vRx Notch ON and Notch OFF neurons of the 9th temporal window) that were disproportionally depleted in Blender, the rest of the domains' cell numbers have a good correspondence between the 3D model and the sc-mRNA-seq dataset (Supplementary Fig. S1C).

2.2. Selection of appropriate marker genes and parameters

We then decided on spatially informative genes that would allow the algorithm to discriminate between the majority of the aforementioned 245 regions. For this purpose, we first removed from the dataset cells that were annotated to be part of the central brain (Konstantinides et al., 2022) (as these were not accounted for in the three-dimensional model in Blender) (Supplementary Figure S2A) and then selected 43 different genes based on different criteria.

- genes that would differentiate between neuroepithelium (*shg*, *dpn*), neuroblasts (*shg*, *dpn*, *ase*), GMCs (*ase*, *elav*), neurons (*elav*), and glia

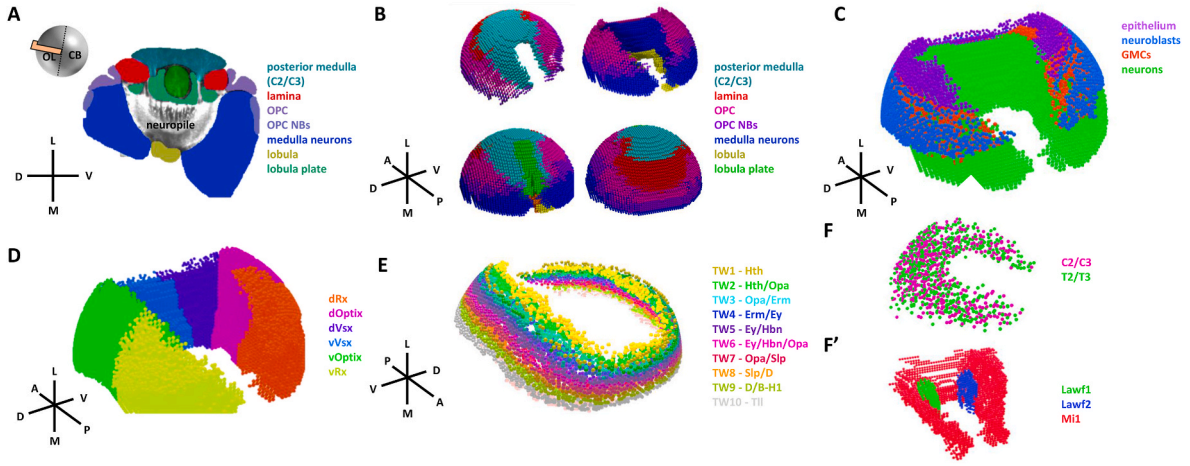


Fig. 2. Regionalization of the three-dimensional model
(A) Color-coded cross-section of the developing optic lobe. Region annotation comes from (Ngo et al., 2017). (B) The three-dimensional model of the optic lobe was divided into seven domains, i.e. the posterior medulla, lamina, OPC, OPC neuroblasts, medulla neurons, lobula, and lobula plate. (C) Visualization of the OPC epithelium (magenta), neuroblasts (blue), GMCs (red), and neurons (green) after the addition of an extra region of GMCs underneath the neuroblasts. (D) The OPC epithelium, neuroblasts, GMCs, and neurons were further divided into six domains, based on the spatial patterning of the epithelium. (E) The medulla neuroblasts and GMCs were further subdivided into ten different temporal windows (TW1-10). (F-F′) The posterior medulla was subdivided into two populations mixed in a salt-and-pepper manner, C2/C3 and T2/T3 cells, while the Lawf1 and Lawf2 cells were added adjacent to the Mi1 neurons. Abbreviations: D: dorsal, V: ventral, M: medial, L: lateral, A: anterior, P: posterior. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

(*repo*, *wrapper*) (Konstantinides et al., 2022) (Fig. 3A and Supplementary Figure S2B).
- genes that would discriminate between the different neuropils (Lobula plate: *acj6* and *Lim1*; Lamina: *eya*) (Konstantinides et al., 2022; Schilling et al., 2019) (Fig. 3C).
- spatial transcription factors genes that discriminate neuroepithelial cells from different spatial origin (*Vsx1*, *Optix*, *Rx*) (Erclik et al., 2017) and temporal transcription factors that discriminate neuroblast from distinct temporal window (*hth*, *Dll*, *erm*, *opa*, *ey*, *hbn*, *scro*, *slp1*, *D*, *B-H1*, *tll*) (Konstantinides et al., 2022) (Fig. 3B and Supplementary Figure S2C).
- genes that would differentiate between the different neuronal types. To do this, we wanted to identify transcription factors that would be expressed differentially in neurons based on their spatial, temporal, and Notch origins. We selected transcription factors that would discriminate as much as possible neurons with different Notch status (Notch-ON and Notch-OFF), different temporal origins (temporal windows: Hth, Hth/Opa, Opa/Erm, Erm/Ey, Ey/Hbn, Hbn/Opa/Slp, Slp/D, D/B-H1) and different spatial origins (spatial domains: ventral Dpp, Optix, and Vsx, and dorsal Dpp, Optix, and Vsx) (note that temporal identities refer to neuroblast expression windows, so neurons born in a given window do not necessarily maintain expression of the corresponding temporal factor). The transcription factors that were selected are the following: *aop*, *ap*, *bsh*, *dac*, *CG34340*, *Ets65A*, *fd59A*, *fkh*, *Hmx*, *kn*, *Lim3*, *oc*, *run*, *sim*, *Sox102F*, *svp*, *tj*, *toy*, *tup*, *vvl* (Simon et al., 2025)(Fig. 3D and Supplementary Figure S2D).

All of the marker genes and their corresponding landmark expression can be found in Supplementary Table 2.

We also performed comprehensive tests across the relevant parameter space to determine how each parameter affects Novosparc's performance (Moriel et al., 2021), for example.

- the alpha parameter, which controls the interpolation between two modes of reconstruction: de novo spatial reconstruction ($\boldsymbol{\alpha = 0}$) and reconstruction based solely on the information provided by the reference atlas ($\boldsymbol{\alpha = 1}$),
- the number of genes used for the reconstruction: It is generally advisable to include only the most differentially expressed genes, as these are more likely to be spatially informative (*i.e.*, expressed in some cells but not others, ideally across a well-defined spatial domain). Based on the above criterion, *elav* was not used in the reconstruction as it was not among the most differentially expressed genes.
- the number of nearest neighbors, which determines the number of neighbors used to construct the k-nearest neighbors (kNN) graphs for cells and spatial locations, and
- the epsilon parameter, which is associated with the entropic regularization term. A low epsilon value leads to more localized mapping, while a higher epsilon value results in higher-entropy mappings, approaching a uniform distribution.

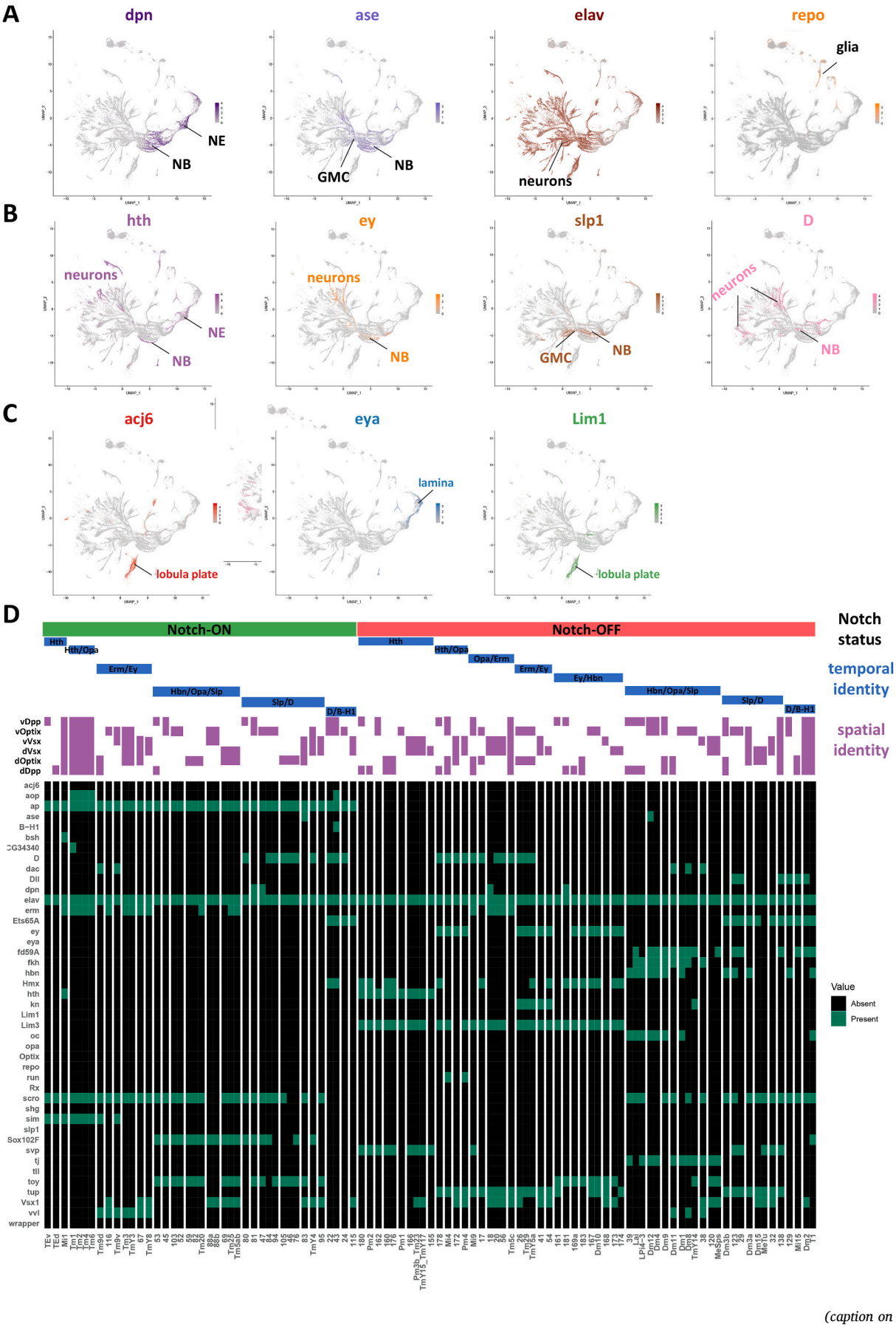
We measured the effect of each parameter by visually observing reconstructed gene expression across 20 z-stacks and calculating the entropy of the transportation of individual cells (lower entropy was associated with a more efficient mapping) and concluded in the use of the following parameter values: alpha = 0.5, number of genes = 5000, k neighbors = 15, and epsilon = 0.005. Details about parameter selection can be found in the Materials and Methods and in Supplementary Fig. S3.

2.3. Evaluation of Novosparc run

Using the above spatial reference, parameters, and markers, we reconstructed in three dimensions the sc-mRNA-seq dataset of the third instar larval developing optic lobe using Novosparc.

The first test for the performance of Novosparc was to evaluate the expression of the genes that were used as landmarks. Since the alpha parameter was set to 0.5, Novosparc was free to navigate between a de novo and a guided reconstruction. The reconstructed expression of the landmark genes can be visualized in Fig. 4 and Supplementary Figure S4. We compared the reconstructed expression with the actual expression of some of these marker genes using antibody staining (Fig. 4).

Among the genes that differentiate cell classes (neuroepithelium, neuroblasts, GMCs, neurons, and glia), the reconstructed expression of *dpn* is found in the OPC and IPC neuroepithelium (marked with asterisk) and the neuroblasts of the medulla and the lobula plate (Fig. 4A), as it is obvious also from the sc-mRNA-Seq (Fig. 3A). Notably, immunostaining against Dpn protein shows faint expression in the neuroepithelium and



(caption on next page)

Fig. 3. Marker gene selection
(A) UMAP plots showing the expression of cell class markers. Combinations of cell class markers, such as *dpn*, *ase*, *elav*, and *repo* can differentiate between neuroepithelium, neuroblasts, GMCs, neurons, and glia. (B) UMAP plots showing the expression of temporal transcription factors. Combinations of temporal transcription factors can differentiate between the different neuroblast and GMC temporal windows, as well as many neuronal types. (C) UMAP plots showing the expression of different neuropil neuron markers. Combinations of these markers can differentiate the lobula plate and lamina from the medulla neurons. (D) On/off expression heatmap of the different landmark genes (on the left side) that are expressed in neuronal types (indicated in the bottom of the heatmap) of different spatial, temporal, and Notch origin (indicated on top of the heatmap) at pupal stage P15. The temporal and spatial transcription factors that define the origin are expressed in progenitors (neuroepithelium and/or neuroblasts). Data come from the mixture modeling (Davis et al., 2020) used in (Simon et al., 2025).

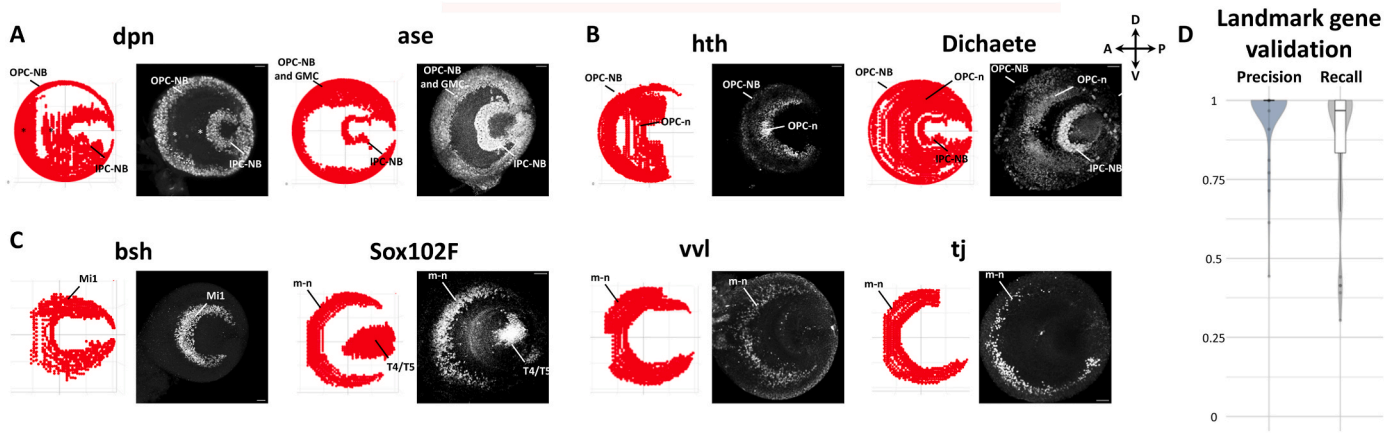


Fig. 4. Reconstructed marker gene expression
Comparison of reconstructed marker gene expression as seen in https://larva3dnovosparc.ijm.fr and as seen by antibody staining against the respective proteins. (A) *dpn* and *ase* are expressed as expected, the neuroblasts (medulla and lobula plate neuroblasts) and neuroblasts and medulla GMCs, respectively. Notably, *dpn* transcript can be detected in the OPC and IPC neuroepithelia (asterisks), in which protein expression is very faint. (B) *hth* can be detected in young medulla neuroblasts, as well as medulla neurons in both reconstructed gene expression and antibody staining. Similarly, *D* can be detected in medulla and lobula plate neuroblasts, as well as many medulla neurons in both representations. (C) Neuronal markers, such as *bsh, Sox102F*, *vvl*, and *tj* are detected in the same expression pattern in both reconstructed gene expression and antibody stainings. Abbreviations: m-NB: medulla neuroblasts, lp-NB: lobula plate neuroblasts, m-GMCs: medulla GMCs, m–n: medulla neurons. Orientation cross applies to all panels. (D) Precision and recall were calculated for landmark genes to evaluate the accuracy of the Novosparc reconstructions. Boxplots overlaid on violin plots summarize the distribution of precision and recall across all genes. For landmark genes, both precision and recall were consistently high (>95 %).

stronger expression in the neuroblasts. Reconstructed *ase* mRNA and immunostaining against Ase protein shows that the gene is expressed in neuroblasts and GMCs (Fig. 4A) that come from the OPC and IPC, while *shg* reconstructed mRNA expression can be found in the neuroepithelium and the neuroblasts, as expected (Supplementary Fig. S4A and Supplementary Fig. S2B). On the other hand, the reconstructed expression of *repo* can only be found in the glia that are born during the last division of the main OPC neuroblasts (Supplementary Fig. S4A) while it is expressed in all the glia of the optic lobe, which shows that migrating cells like glia are very hard to reconstruct with such approaches (discussed further in the "Limitations of this study" Discussion section). Wrapping glia express *wrapper* mRNA as expected (Supplementary Fig. S4A and Supplementary Fig. S2B).

Among the spatial transcription factors, *Vsx1* expression was reconstructed in the central part of the OPC neuroepithelium, as well as in medulla neurons that predominantly, but not exclusively, come from this neuroepithelial region (asterisk), and *Optix* and *Rx* expression was reconstructed in the respective regions of the neuroepithelium (Supplementary Fig. S4B), as expected. We also tested for the reconstructed expression of the temporal genes *hth, D, erm, ey, slp1, hbn, opa,* and *Tll*: *hth* is expressed in the medulla NBs, as well as some medulla neurons, as also shown by antibody staining (Fig. 4B). Interestingly, *hth* transcript expression was also reconstructed in the progeny of the *Tll* temporal window (whereas the protein is absent), which we also verified by HCR-FISH (Supplementary Figure S4C). Similarly, *D* is expressed in medulla neuroblasts and neurons, as well as the IPC neuroblasts, as was already known (Apitz and Salecker, 2015) (Fig. 4B). *erm* is expressed very broadly as can also be confirmed by antibody staining, *ey* is expressed in medulla neuroblasts and many different medulla neurons, *slp1* is expressed exclusively in successive temporal windows of the medulla neuroblasts, *hbn* is expressed in medulla neuroblasts and neurons, C/T cells, and Lawf neurons, *opa* is expressed in medulla neuroblasts and GMCs in two temporal windows, and *tll* is expressed in old medulla neuroblasts, lobula plate neuroblasts, and the lamina (Supplementary Figure S4D).

We also looked for the spatial reconstruction of genes that differentiate different neuropil origin. As expected, *acj6* expression was assigned exclusively to the lobula plate T cells, *eya* to the lamina and lamina widefield cells (Lawf1 and Lawf2), and *Lim1* to T4/T5 cells and Lawf2 (Supplementary Fig. S4E).

Finally, we probed for the reconstructed expression of the neuronal type-specific genes, *bsh, Sox102F, vvl, tj, ap, dac, dll, Ets65A, fd59A, fkh, kn, oc, run, svp, toy,* and *tup*. As can be seen in Fig. 4C and Supplementary Figure S4F, their reconstructed expression matches to a large extent their actual expression as revealed by antibody stainings. It should be noted that, while Novosparc reconstructs transcript expression, antibody stainings reveal protein expression, which could be slightly different: for this reason, we also used the transcript expression in the UMAP plots to confirm the accuracy of the reconstructions. Overall, we evaluated the quality of the reconstruction of 31 marker genes, all of which (with the exception of *repo*) were reconstructed accurately (Supplementary Fig. S4F).

To quantitatively assess the accuracy of the spatial reconstruction, we compared the reconstructed expression of each landmark gene with its binarized reference expression pattern. For each of the 43 landmark genes, we calculated precision (proportion of cells predicted as expressing a gene that are truly expressing it in the reference atlas: TP/(TP + FP)) and recall (proportion of cells that truly express a gene in the

reference atlas that are correctly identified as such in the reconstruction: TP/TP + FN) by treating the reference atlas as ground truth (Supplementary Table 2) and the reconstructed Novosparc output as predictions. Specifically, for every vertex in the 3D model, we evaluated whether the reconstructed expression state (ON/OFF, based on the thresholds used in Fig. 4 and S4 – Supplementary Table 3) matched the corresponding binarized reference assignment. Both precision and recall were consistently high across all landmark genes, exceeding 95% in nearly all cases, indicating that the spatial reconstruction faithfully recapitulates the reference expression patterns, as evaluated also qualitatively (Fig. 4D and Supplementary Table 4).

2.4. Reconstruction of expression of non-reference genes

We then challenged the reconstruction with a more difficult task: to test the expression of genes that were not used as a reference. For this purpose, we decided to probe different types of genes beyond transcription factors, such as neuronal effector genes (in particular different cell surface molecules) and signaling molecules (in particular members of the Wnt family, which is well known to regulate neuronal diversification in the visual system (Bertet et al., 2014; Özel et al., 2021; Suzuki et al., 2016b)) (Fig. 5).

We first examined the cell type–specific expression of cell surface molecules. Because these proteins are predominantly localized at synapses rather than in cell bodies, antibody staining would not accurately reflect cell body expression for comparison with the Novosparc reconstruction. Therefore, we tagged DIP-β, DIP-ε, DIP-η, and DIP-θ with T2A-Gal4 using the "plug-and-play" system (Diao et al., 2015) to drive expression of GFP in the cell types, where these genes are expressed. *DIP-beta* is expressed in few cell types, as can be seen by the UMAP and the antibody staining, including Lawf1 and Lawf2, and other cells that likely originate from the Vsx1 domain, as their cell bodies are concentrated centrally (Fig. 5A). The 3D reconstruction of Novosparc shows perfectly this expression, but it misses *DIP-beta*-expressing cells that are found in the Optix domain (note that the processes of these cells are also seen by immunostaining using membrane GFP, for example in the medulla cortex or the lobula plug, and these will not be included in the Novosparc reconstruction, which only includes neuronal somas). Similarly, *DIP-epsilon* is also expressed in few cell types with an obvious enrichment in cells in the dorsal side of the optic lobe. The 3D reconstruction identifies this difference in the dorsoventral localization of *DIP-epsilon*-expressing cells, but missplaces it to the ventral side (Fig. 5A). *DIP-eta* is expressed across the entire medulla neuronal cortex with a slight but noticeable depletion in the neurons that come from the Vsx1 domain, which is noticeable in both antibody staining and 3D reconstruction (Fig. 5A). Finally, *DIP-theta* is also expressed in different neuronal types that are primarily localized in the dorsal side of the optic lobe. Novosparc reconstruction identifies this difference along the dorsoventral axis, however it seems less pronounced, as it identifies more cells in both the ventral and the dorsal side than the ones that are marked with the Gal4 line (Fig. 5A). We then looked for the localized expression of *Wnt2,* which is expressed in immature Mi1 cells (as indicated by the UMAP) and can be seen in few early-born neurons that span the medulla cortex, as indicated by both 3D reconstruction and Gal4 expression (Fig. 5B). In total, out of the 5 genes that we tested, 4 were reconstructed very well, while only *DIP-epsilon* expression was less accurate.

To extend our quantitative assessment of reconstruction performance beyond the landmark genes (Fig. 4D), we next evaluated the accuracy with which Novosparc reconstructs the spatial distribution of non-landmark genes. Establishing ground truth for these genes is inherently more challenging, as no binarized spatial reference is available. We therefore relied on the experimentally validated cell-type origins reported in (Simon et al., 2025), which define the expected spatial and temporal domains for 15 terminal selector genes (*ab, apt, Atf3, CG11085, CG11294, CG43689, dm, elB, foxo, ham, NK7.1, TfAP-2, tsh, zfh1, zfh2*) (Supplementary Fig. S5 and Supplementary Table 2). Using these annotations as ground truth, we compared the expected expression domains with the reconstructed spatial patterns generated by Novosparc. For each gene, we computed precision and recall using the same definitions applied to the landmark gene analysis (Fig. 5C and Supplementary Table 4). Overall, the reconstruction of non-landmark genes showed high precision (≈80 %), indicating that when Novosparc predicts a gene to be expressed in a given region it is very likely to be correct, while recall values (≈65 %) were somewhat lower, which is expected given the less well-defined nature of the available ground truth.

To further assess the predictive performance of the spatial reconstruction independently of the genes used to guide the mapping, we performed a Leave-One-Out Cross-Validation (LOOCV) analysis on the landmark gene set. In this framework, each landmark gene was iteratively excluded from the reference, Novosparc was rerun using the remaining genes, and the spatial expression of the withheld gene was quantitatively compared to its original reference pattern. This analysis revealed robust predictive performance, with average precision and recall values of approximately 80 % across landmark genes (Fig. 5D and Supplementary Table 4). Interestingly, and consistent with expectations (see Study limitations), the weakest performance was observed for genes primarily encoding spatial domain information: Vsx1 and Rx exhibited the lowest precision, while Rx also showed the lowest recall. These results indicate that although spatially restricted genes are inherently more challenging to predict when withheld from the reference, the reconstruction retains substantial predictive power across the landmark set.

2.5. Novosparc predicts complex regulation of candidate terminal selector expression

Terminal selectors are transcriptional regulators that coordinate the acquisition and maintenance of neuron-type-specific features, including synaptic connectivity, neurotransmitter identity and ion channel expression (Hobert, 2008, 2011). They have been best characterized in *C. elegans* but also in the *Drosophila* optic lobe (Özel et al., 2022; Simon et al., 2025). Concentric transcription factors are transcription factors regulated by temporal transcription factors (although usually under the regulation of spatial transcription factors and/or Notch, as well), which results in a concentric expression in the medulla cortex (Hasegawa et al., 2011). We recently noticed that concentric transcription factors are all predicted terminal selectors (Simon et al., 2025), highlighting the crucial role of temporal patterning in establishing the great majority of neuronal diversity in this neuronal structure. We wondered whether we could use the 3D reconstruction of the optic lobe to a) identify the expression pattern of all published candidate terminal selectors and categorize them and b) identify the relative spatial organization of concentric gene-expression domains with respect to the neuroblast temporal windows.

Indeed, by plotting the expression of all 95 candidate terminal selectors (Supplementary Figure S6).

- We noticed that about one third of them (31) form concentric rings in the medulla cortex, suggesting that they are under the regulation of temporal transcription factors,
- 18 of them have specific spatiotemporal patterns indicating that they are regulated by both spatial and temporal transcription factors,
- 9 show even more complex expression patterns raising the possibility that they are regulated by temporal and/or spatial transcription factors,
- 9 are mostly expressed in neurons of specific neuropils,
- 16 have restricted expression patterns, while
- 12 have broader, albeit specific, expression patterns.

Importantly, we noticed that by looking into the spread of the

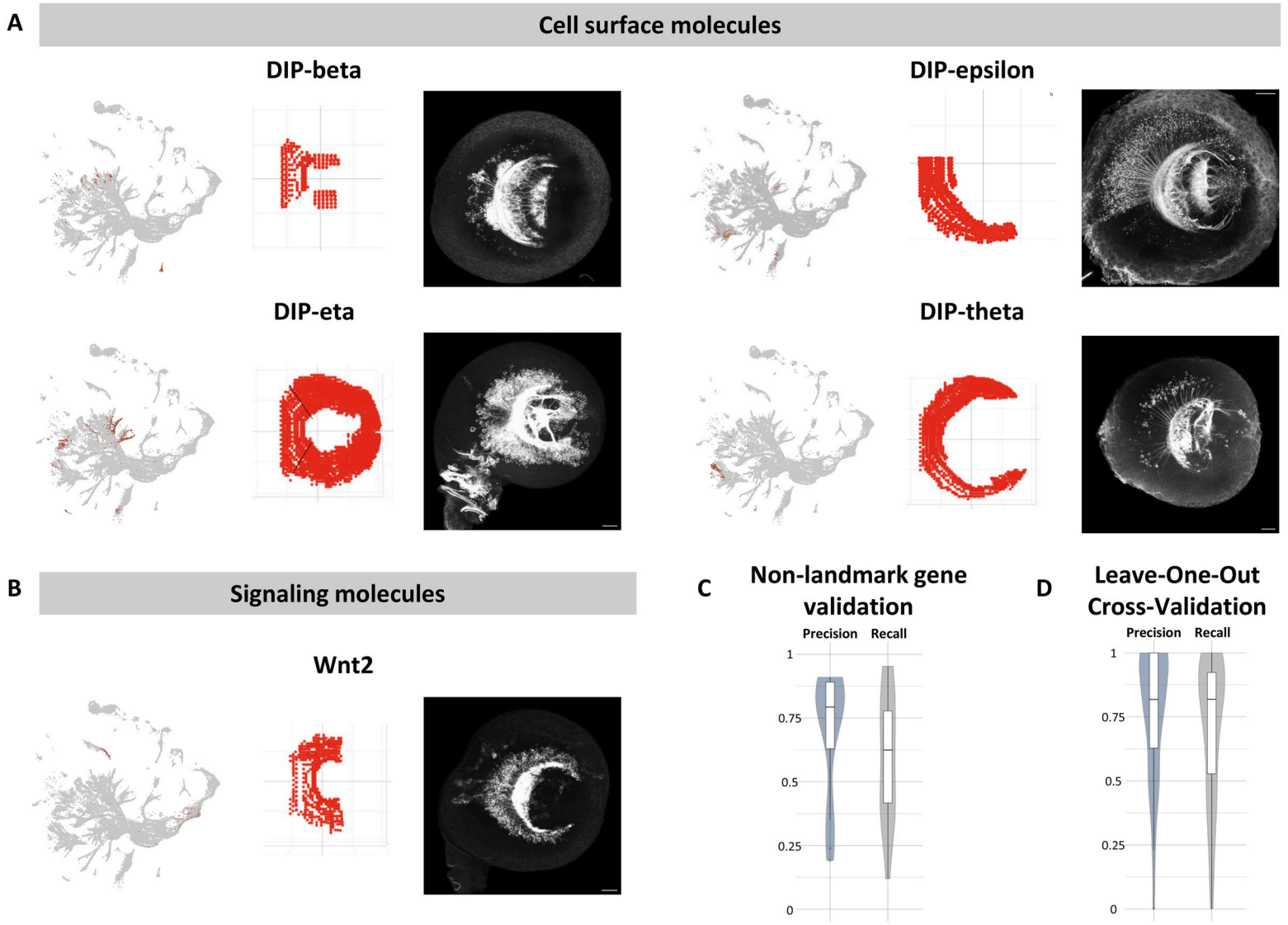


Fig. 5. Non-reference gene expression

(A) Cell surface molecule expression in the larval developing optic lobe, as can be seen in a UMAP plot, the Novosparc reconstruction, and by antibody staining of the respective T2A-Gal4 line driving myristoylated GFP. The reconstructed expression of *DIP-beta, DIP-eta*, and *DIP-theta* matches their actual expression pattern. The Novosparc reconstruction correctly captured the unilateral nature of *DIP-epsilon* expression, but assigned it to the ventral rather than the dorsal optic lobe. (B) *Wnt2* expression in the larval developing optic lobe, as can be seen in a UMAP plot, the Novosparc reconstruction, and by antibody staining of the Wnt2-Gal4 line driving GFP. The reconstructed expression of *Wnt2* matches its actual expression pattern. (C) Precision and recall were calculated for non-landmark genes to evaluate the accuracy of the Novosparc reconstructions. Boxplots overlaid on violin plots summarize the distribution of precision and recall across all genes. For non-landmark genes, precision is high (~80 %), while recall was lower (~65 %), reflecting a more challenging ground truth. (D) Precision and recall were calculated for each landmark gene using a Leave-One-Out Cross-Validation scheme, in which each gene was iteratively removed from the reference, reconstructed using Novosparc, and compared to its original spatial pattern. Boxplots overlaid on violin plots summarize the distribution of precision and recall across all landmark genes. Both precision and recall are high (approximately 80 %), indicating robust predictive performance of the spatial reconstruction.

reconstructed gene expression in the anteroposterior axis, we could estimate whether a gene was expressed in neurons that are generated during early, mid, or late temporal windows, as exemplified by the known early, mid, and late markers, *bsh*, *vvl*, and *hbn* respectively (Konstantinides et al., 2022) (Fig. 6A). The 3D reconstruction suggests that *tsh*, *dve*, and *ab* are also markers of early, mid, and late-born neurons, respectively. Similarly, looking at the distribution of the gene expression along the medulla cortex, we could predict the domains where each of the neurons that express these genes come from. For example, *bab2* is reconstructed in neurons that originate from the vVsx domain, *mirr* is expressed in neurons that come from the dVsx and dOptix domains, and *ara* in neurons that come from the dtOPC (Fig. 6B). Finally, genes that are under more complex regulation, such as *Sox21B* and *CG9932*, which are expressed in neurons from at least two different temporal windows can also be visualized (Fig. 6C).

In total, this shows that our 3D atlas can serve as a reliable hypothesis generator and can facilitate the study of different stages and different mechanisms of neuronal development.

2.6. A dedicated webpage to explore the three-dimensional atlas

To facilitate the use of this atlas, we made this three-dimensional reconstruction available to the scientific community via a dedicated webpage (https://larva3dnovosparc.ijm.fr). The webpage layout consists of several components.

- dropdown menus to select the genes to be visualized (up to 3 genes can be visualized at once).
- color pickers to choose the color for each gene.
- threshold inputs for each gene to adjust the range of values for visualization.
- axis filters (x, y, z) via sliders to perform virtual sections along each of the three axes.
- a download option to export the filtered dataset as a CSV file.

The main feature of the webpage is a 3D scatter plot that allows the visualization of the data points in space, colored by the presence or absence of selected gene expression. Users can blend colors for up to 3 genes and adjust the range of values to be visualized by setting minimum and maximum thresholds for each gene. The plot updates dynamically based on user input. The axes can be filtered according to the selected ranges, and colors are updated in real time. Finally, the user can download the filtered data in a CSV format. The basic functionalities of the webpage can be seen in Supplementary Figure S7.

3. Discussion

3.1. How to make a neuronal three-dimensional gene expression atlas?

Here, we present the pipeline that we followed to generate a three-dimensional gene expression atlas of a complex developing neuronal structure, the *Drosophila* optic lobe, using Novosparc. This computational approach allowed us to infer the spatial distribution of cell types and gene expression, providing a valuable resource for studying the organization of the developing nervous system. By leveraging single-cell mRNA sequencing data and spatial transcriptomic reconstruction, we identified spatially compartmentalized cell types and detected patterns of transcription factor expression. This work highlights the power of computational methods in augmenting single-cell sequencing datasets and provides insights into the spatial organization of neuronal development.

We also present the necessary steps needed to perform the reconstruction of a complex neuronal structure. First and foremost, one needs to take into account the developmental origin of the different neuronal types. While the hypothesis that cells that are close to each other are more likely to resemble each other transcriptionally is on average correct, it falls short when structures with different developmental origins juxtapose each other; cells with shared lineage tend to localize together, suggesting that spatial organization is not solely dictated by signaling cues but also by intrinsic lineage constraints. Therefore, to perform a 3D reconstruction with Novosparc, one should divide the tissue in regions based on their developmental origin. Then, one must identify markers that would separate these regions, as well as some of the cell types that were already known in the tissue of interest. Finally, we tested different parameters, although this had the least impact in the resulting reconstruction.

Finally, we provide further evidence that Novosparc is an efficient tool for 3D reconstruction. Using only 43 landmark genes, it can accurately reconstruct complex neuronal structures by leveraging the full transcriptome of single-cell data. While many modern spatial transcriptomics platforms provide high-plex or full-transcriptome coverage and integration methods can map scRNA-seq cells onto partial spatial references, Novosparc achieves spatial reconstruction without requiring a paired spatial dataset, making it a complementary approach when such references are limited or incomplete.

3.2. Study limitations

Novosparc also has limitations compared to other spatial transcriptomic approaches. First, it relies heavily on existing knowledge: in the absence of precise landmark expression information or developmental origin data, Novosparc was unable to satisfactorily reconstruct the complex developing optic lobe. Consequently, its capacity to identify novel “regions” that deviate from the rule that physical proximity correlates with transcriptomic similarity is limited. In the developing optic lobe, information is enriched for temporal patterning genes relative to spatial patterning genes, which is reflected in the reconstruction. This is further supported by the LOOCV results presented in Fig. 5D and Supplementary Table 4. Among the four genes with the lowest F1 scores, three are spatial transcription factors, Rx, Dll (both spatial and temporal), and Optix, while Vsx1 ranks ninth. Therefore, users of the reconstructed dataset should exercise caution when using it to identify potential novel regions. Notably, for all other genes besides Rx (where both precision and recall are low), either precision or recall is high; thus, careful adjustment of the threshold can improve the reconstructed expression.

Second, using Novosparc, we are restricting each cell to a single point. While this may not be a problem for different tissues, neurons are very complex cells with great neurite diversity, which often translates to localization of specific transcripts in different subcellular locations (Glock et al., 2021; van Oostrum and Schuman, 2025). This type of information is completely inaccessible in a Novosparc reconstruction and requires other spatial transcriptomic approaches for this complexity to be uncovered. Finally, Novosparc does not reconstruct absolute expression levels but rather reconstructs spatial gene expression by assigning probabilities of expression to different locations. This represents a limitation because it does not directly capture the quantitative differences in gene expression across the tissue. Instead, it provides a probabilistic estimate of where a gene is likely to be expressed based on spatial constraints and similarity to reference data. As a result, while Novosparc can reveal spatial trends and organization, it may not accurately reflect expression gradients or absolute transcript abundances, potentially limiting its application in cases requiring precise quantitative measurements.

One final limitation of our study is that, while we tried to reconstruct the entire developing optic lobe, our focus was predominantly the neurons of the medulla neuropil that originate from the main OPC. This means that we have lower spatial resolution in lamina neurons and the neurons that come from the IPC, as well as glia, which were essentially only assigned to two regions: the last medulla temporal window and the wrapping glia region. Nonetheless, we do provide the pipeline, the know-how, and the code for future studies to elaborate further on either

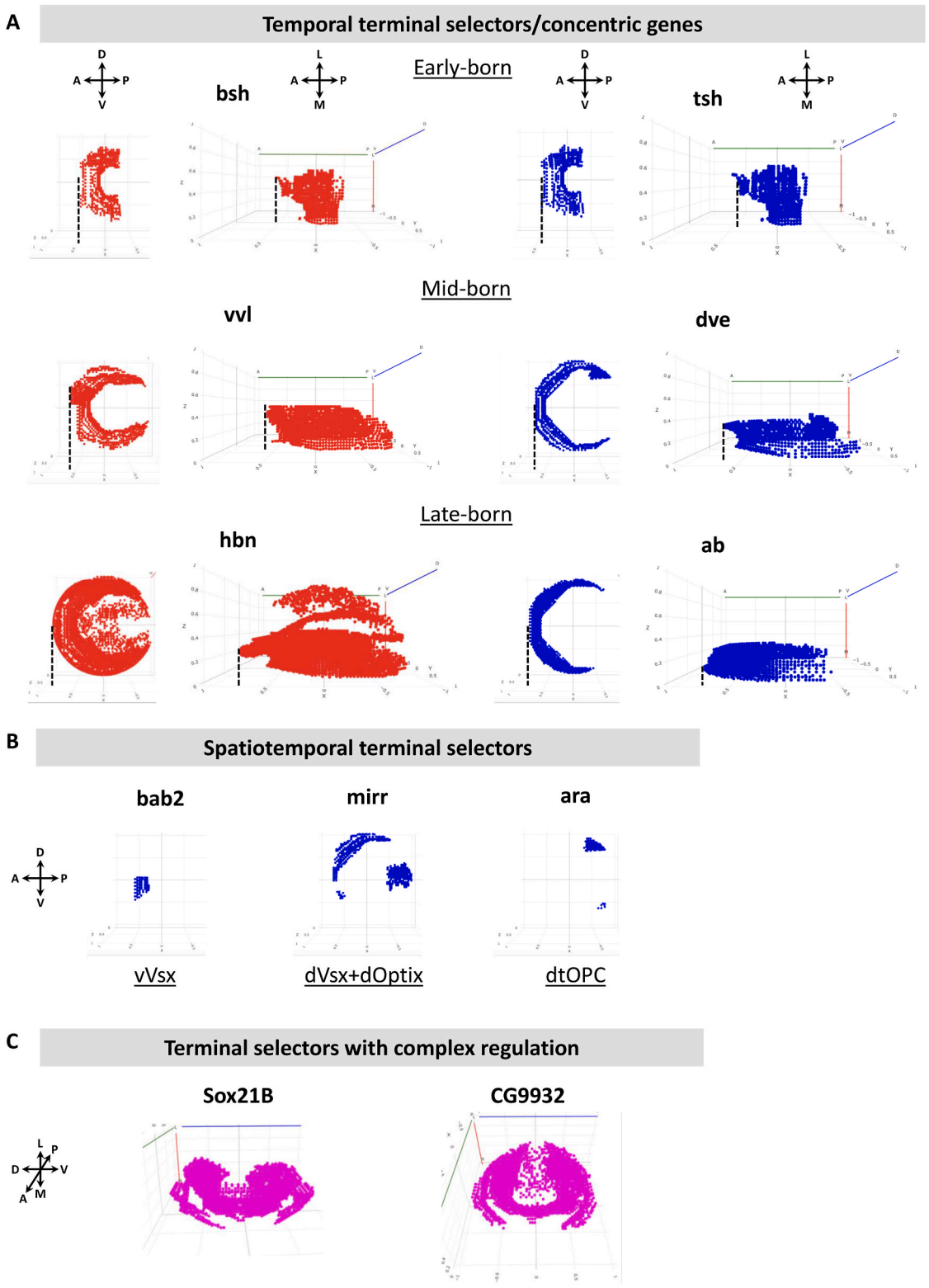


Fig. 6. Spatio-temporal regulation of terminal selectors
(A) Novosparc-reconstructed expression of temporally regulated terminal selectors. We use the 3D reconstruction to identify all concentric genes (whose expression is likely regulated by the temporal transcription factors) and to classify whether the neurons that express these genes are born early (such as *bsh*- and *tsh*-expressing neurons), mid (*vvl*- and *dve*-expressing neurons), or late (*hbn*- and *ab*-expressing neurons). (B) Novosparc-reconstructed expression of spatiotemporally regulated terminal selectors. We use the 3D reconstruction to identify all terminal selectors which are spatially restricted in the medulla cortex, which indicates that they originate from one (or more) of the spatial domains, such as *bab2* (vVsx domain), *mirr* (dVsx and dOptix domains), and *ara* (dtOPC). (C) Novosparc-reconstructed expression of terminal selectors with complex regulation. We identify terminal selectors whose expression pattern is more complex, such as *Sox21B*, which is expressed in neurons from two temporal windows in a spatially restricted manner, and *CG9932,* which is also is expressed in neurons from two distinct temporal windows.

of these two populations.

3.3. Using spatial reconstructions to generate new insights

Despite the above limitations, our spatial reconstruction provides a framework for deriving new biological insights from single-cell data. Using this reconstruction, we can (i) predict spatial and temporal regulation of neuronal genes to guide targeted functional experiments by prioritizing candidate transcription factors or signaling pathways acting in specific regions, and (ii) support comparative analyses across species, particularly to study the evolution of spatial patterning and cell-type specification mechanisms.

3.3.1. Temporal and spatial patterning of terminal selector expression

First, we were able to use the atlas to study terminal selector regulation. The 3D reconstruction of the *Drosophila* optic lobe confirms that terminal selector gene expression is closely linked to temporal transcription factor dynamics. Specifically, a significant fraction of candidate terminal selectors follows concentric expression patterns in the medulla cortex, suggesting regulation by temporal transcription factors. By analyzing gene expression along the anteroposterior axis, we could predict whether neurons expressing these genes originated from early, mid, or late temporal windows. Similarly, the anteroposterior spread of gene expression in the medulla cortex allowed us to infer neuroblast domain origins. These findings reinforce the idea that temporal transcription factors play a crucial role in coordinating neuronal diversity by guiding terminal selector expression. Of course, as we have recently shown (Simon et al., 2025), most of these genes are under more complex regulatory regime that requires input from spatial transcription factors and Notch. Furthermore, genes with complex regulatory patterns, such as *Sox21B* and *tup*, highlight the existence of multi-layered regulatory mechanisms. Our results demonstrate that the 3D spatial reconstruction of the optic lobe is a powerful tool for dissecting developmental gene regulation, providing a framework to investigate how temporal and spatial factors interact to shape neuronal identity.

3.3.2. Evolution of developing nervous systems

The know-how gained from reconstructing spatial gene expression can be applied to studying the evolution of developing nervous systems by enabling cross-species comparisons of spatial organization. With the increasing availability of single-cell sequencing datasets across diverse species, we now have an unprecedented opportunity to investigate how neurodevelopmental programs, such as spatial and temporal patterning (Filippopoulou and Konstantinides, 2023), have evolved. By using spatial reconstructions of single-cell data, we can identify conserved and divergent patterns of gene expression, cellular arrangements, and developmental trajectories across taxa. This approach can reveal how lineage constraints shape nervous system architecture and how novel structures emerge through evolutionary modifications (Roberts et al., 2022; Tosches, 2017). Additionally, comparing the spatial distribution of homologous genes across species can help disentangle the relative contributions of conserved regulatory networks and species-specific adaptations. The ability to reconstruct and analyze spatial gene expression in a three-dimensional context further enhances our capacity to study neuronal diversification at multiple levels, from gene regulatory dynamics to whole-brain patterning. As more single-cell and spatial datasets become available, comparative studies will provide deeper insights into the fundamental principles that govern nervous system development and its evolutionary transformations.

4. Materials and Methods

4.1. Drosophila husbandry and genetics

Fly stocks and crosses were maintained on standard food at 25 °C for experiments. The *Drosophila melanogaster* stocks used can be found in Supplementary Table 5.

4.2. HCR-FISH

The HCR protocol for *Drosophila* larval brains used was as specified previously (Ferreira et al., 2021).

4.3. Immunofluorescence

Late third instar larval *Drosophila* optic lobes were dissected in cold Schneider's media and transferred to Schneider's media on ice. Then they were fixed in 4 % paraformaldehyde for 20 min at room temperature. After washing with 0.5 % Triton-X diluted in PBS 1X (PBT) for 30 min, the brains were incubated at least 30 min in PBT with 5 % horse or goat serum (PBT-block) at room temperature. Then, the samples were incubated overnight with primary antibodies at 4 °C. After washing the primary antibodies twice for 30 min with PBT, the brains were incubated with the secondary antibodies overnight at 4 °C. The secondary antibodies were washed twice for 30 min with PBT, and the brains were mounted in Slowfade and imaged at a confocal microscope (Leica SP8) using a 63× glycerol objective. Images were processed in Fiji.

Details of the origin of all individual antibodies can be found in Supplementary Table 5.

4.4. Novosparc - blender

To build the atlas, we used the 3D software Blender. We imported CSV files of coordinates via the scripting tab in Blender and a Python script. Once imported, we assigned the xcoord, ycoord, and zcoord columns of the CSV file to the x, y, and z positions in Blender using Geometry Nodes. We first added a “Set Position” node between the input node and the output node. Then, we added a “Combine XYZ” node connected to the position socket of the “Set Position” node. We connected each socket of the “Combine XYZ” node (x, y, and z) to a “Named Attribute” node. We selected either xcoord, ycoord, or zcoord for each attribute, depending on which socket it was connected to. Finally, we applied the Geometry Nodes Modifier.

Layer by layer, we selected the points we wanted to associate with an atlas region and separated them into a new mesh, which we named according to the color associated with that region. After doing this for all the points, we exported the Blender file to an OBJ file that included all the meshes and their coordinates. Then, we created a new CSV file containing all the positions and added new columns indicating which mesh each position belonged to. For example, if a coordinate was in the "blue" mesh, we wrote 1 in the blue column; if not, 0. We applied the same method to generate all subregions.

4.5. Novosparc – R

Then, we prepared the published data for Novosparc. First, we needed to switch the published RDS object (GEO accession: GSE167266) to an AnnData object that contained the normalized but not scaled data in the correct location. An AnnData object has multiple layers; Novosparc accesses a specific layer of this object (the primary layer), so it is crucial to ensure that the desired data for reconstruction is stored in that layer. To achieve this, we read the.rds dataset into R as a Seurat object, which we then converted to an AnnData object, ensuring the appropriate data was placed in the primary layer. The script Conversion_To_AnnData_Only_Normalized can be found in the GitHub repository.

4.6. Novosparc - overview of the algorithm and workflow

The Novosparc algorithm is based on the hypothesis that there is a correspondence between the structure of positions in physical space and the structure of cells in gene expression space. Therefore, there is a

correspondence between pairwise distances of locations in physical space and pairwise distances of the cells in gene expression space. This is why the first step of the Novosparc algorithm is to compute cost matrices based on the construction of k-nearest-neighbors (kNN) graphs in both physical space and gene expression space. Using the first two inputs, the locations file (CSV file of coordinates) and the single-cell dataset, Novosparc computes two cost matrices: the cell-cell cost matrix and the location-location cost matrix. These consist of the pairwise distances between cells and locations, computed as the shortest path according to the kNN graph. As we use another input here, the reference atlas (locations file with marker genes' binarized expression), Novosparc computes an additional cost matrix: the cell-location cost matrix. This matrix captures the discrepancy between the expression of the marker genes in each cell and in each location of the target space.

Using these cost matrices, Novosparc computes the optimal transport cost matrix, which probabilistically assigns each cell to a location. This assignment depends on different parameters, such as the alpha parameter, which allows us to decide the importance we want to assign to the reference atlas. A value of alpha near 1 indicates high fidelity to the reference atlas, while a value near 0 reflects higher fidelity to the de novo reconstruction (with alpha = 0 being the de novo reconstruction).

The optimal transport matrix allows us to map the single-cell dataset to the locations file. We can then export the final file to a CSV file, with the first three columns being the coordinates (xcoord, ycoord, and zcoord) and all the other columns containing the gene expression data. Each row represents a cell, and each column (except for the first three) represents a gene.

The dataset used to compute the optimal transport matrix consisted of the gene expression values of the 5000 most variable genes, as using the most variable genes has been proven to improve the Novosparc reconstruction (Moriel et al., 2021).

4.7. Novosparc - fine tuning of novosparc parameters

When running Novosparc, there are several parameters that can be adjusted depending on the specifics of the reconstruction.

- **Alpha parameter**: This parameter controls the interpolation between two modes of reconstruction: de novo spatial reconstruction (α = 0) and reconstruction based solely on the information provided by the reference atlas (α = 1).
- **Number of genes**: This defines the number of genes used for the reconstruction. It is generally advisable to include only the most differentially expressed genes, as these are more likely to be spatially informative (i.e., expressed in some cells but not others, ideally across a well-defined spatial domain). However, the markers used for reconstruction are not always the most differentially expressed genes. For instance, if a marker is among the top 5000 differentially expressed genes, you must include at least the top 5000 genes in the reconstruction to incorporate that marker. Balancing this parameter is crucial for achieving accurate results.
- **Number of k-neighbors**: This parameter determines the number of neighbors used to construct the k-nearest neighbors (kNN) graphs for cells and spatial locations.
- **Epsilon**: This parameter is associated with the entropic regularization term. A low epsilon value leads to more localized mapping, while a higher epsilon value results in higher-entropy mappings, approaching a uniform distribution.

One of the challenges we encountered was measuring the reconstruction efficiency in an unbiased way. The goal of this fine-tuning was to run Novosparc with a set of different parameters and, based on the quality of the results, select the optimal parameters to use. To assess the efficiency of the reconstruction, we first chose a set of genes (*acj6, dpn, eya, toy, shg, elav, bsh, ap, hth, ey, slp1, tll, opa, tj, hbn, pxb, Optix, Vsx1, Rx,* and *wg*) and visualized their localization using the embedding function from Novosparc. The efficiency was initially evaluated through visual observation of the reconstructed gene expression across 20 z-stacks.

Additionally, for each set of parameters, the entropy of the transportation of individual cells was measured, and the different parameter sets were compared using histograms and boxplots (Supplementary Figure S3). Lower entropy was associated with a more efficient mapping. To ensure consistency, only one parameter was varied at a time. We established a default set of parameters: alpha = 0.5, number of genes = 2000, k neighbors = 15, and epsilon = 5e-3. We then varied each parameter individually.

4.7.1. Alpha parameter

The reconstructed images were quite similar, making it difficult to make a decision based solely on them. However, an analysis of individual cell entropies revealed a clear trend: entropy decreases as the alpha parameter increases, at least up to an alpha value of 0.5 (Supplementary Figure S3A). Based on this observation, we decided not to use alpha values of 0.1 and 0.3. Ultimately, we chose to proceed with an alpha value of 0.5, aiming for a reconstruction that incorporates a balance between de novo and marker-based approaches.

4.7.2. Epsilon parameter

This parameter was the most complicated to evaluate. Epsilon directly influences entropy, with lower epsilon values resulting in lower entropy. Also, using low epsilon values significantly increased computation time, with the code taking two days to run for epsilon 0.001 and even longer for epsilon 0.0005, compared to just 8–10 h for the other values.

As expected, lower epsilon values resulted in lower entropy values (Supplementary Figure S3B). We decided to keep epsilon = 0.005 for a balance between timing and reconstruction efficiency.

4.7.3. K-nearest neighbors

In the Novosparc code, there are two parameters defined for nearest neighbors: num_neighbors_s and num_neighbors_t, which are typically set to the same value, for example: num_neighbors_s = num_neighbors_t = 15. These parameters assume the existence of an expression "niche" or microenvironment around a cell, where expression differences within the niche are subtle.

- num_neighbors_s is chosen based on the number of immediate neighbors of a point, which depends on the grid's dimensionality. In our case, the grid is 3D, so this value falls between 17 and 26.
- num_neighbors_t can be adjusted based on factors such as the total number of cells, levels of noise, and the dimensionality of the grid. The Novosparc paper suggests that values for num_neighbors_t between 3 and 15 result in robust spatial embedding.

The entropy analysis showed almost no differences when using different k-nearest neighbors parameters (Supplementary Figure S3C).

4.7.4. Number of genes

Upon reviewing the images, we observed that the reconstructions most closely resembling reality were those using higher gene counts, specifically 5000 and 8000 genes. This observation was further supported by entropy analysis, which showed a decreasing entropy value as the number of genes increased (Supplementary Figure S3D). Based on this analysis, we decided to use 5000 genes for the reconstruction.

4.8. Novosparc – final code

The code used to run Novosparc can be found in the GitHub repository (**Novosparc_End**).

4.9. Novosparc - website

To visualize the data a website was developed; the code is available in GitHub (**Web App**). This Python script implements a web application for visualizing and interacting with the Novosparc reconstruction. The application, built using Dash, allows users to select genes, apply thresholds, and choose colors to represent gene expression levels in a 3D scatter plot. Additionally, users can filter the data along the x, y, and z axes and download the filtered dataset.

Key Features.

1. Google Cloud Storage Integration: The script begins by downloading a CSV file from a Google Cloud Storage bucket using the google.cloud library. The file contains reconstructed single-cell data with coordinates and gene expression levels, is saved to the local file system and then read into memory in chunks to manage large files. This file contains the results of the Novosparc code. Several prior operations were performed on this file: first, reuniting the xyz data, which is not initially included in the file. To achieve this, the xcoord, ycoord, and zcoord columns from the atlas are merged with the file. Next, the data in the file is linearly normalized to a range of 0–1, ensuring that no values have more than three decimal places. This normalization facilitates visualization in the code and reduces the size of the CSV file. The code used for normalization is available in GitHub (**Linear_Normalization**); the code used for reducing the numbers of decimals is also available (**Less_Decimals**).
2. Data Preprocessing: The data is loaded and processed, with coordinates (xcoord, ycoord, zcoord) and gene expression values extracted. The script prepares options for gene and color selection, allowing the user to overlay gene expression values with different color schemes (e.g., greyscale, yellow, green, etc.).
3. Dash Web Interface: The application layout consists of several components:
   - Dropdown menus to select the first and optionally the second gene.
   - Color pickers to choose the color for each gene.
   - Threshold inputs for each gene to adjust the range of values for visualization.
   - A suppression threshold for filtering out low-expression data.
   - Axis filters (x, y, z) via sliders to zoom in on specific regions of the 3D space.
   - A download button to export the filtered dataset as a CSV file.
4. Dynamic 3D Plot: The main feature of the website is a 3D scatter plot that visualizes the data points in space, colored by the selected genes. Users can blend colors for three genes and adjust the color intensity by setting minimum and maximum thresholds for each gene. Because Novosparc outputs continuous probabilistic expression values, each gene displays a distinct distribution of reconstructed expression across the 3D model. To generate binarized visualizations, users must therefore select a gene-specific threshold that reflects this underlying distribution. On the accompanying website, the full expression distribution for each gene is displayed, allowing users to choose an appropriate cutoff based on the shape and range of reconstructed values. The default lower threshold of 0.5 serves as a reasonable starting point for most genes, but users are encouraged to adjust this value to best capture the spatial domain of interest. This interactive approach ensures that visualization remains flexible and transparent, while allowing researchers to explore how different thresholds influence the apparent spatial pattern.
5. Interactivity: The plot updates dynamically based on user input. The axes are filtered according to the selected ranges, and colors are updated in real time. The second gene visualization can be toggled on/off via a checkbox.
6. Data Download: The app includes a feature to download the filtered data, which is triggered by a button. When clicked, the filtered dataset is converted to CSV format and made available for download.

This Dash-based application provides an interactive platform for visualizing and analyzing the Novosparc reconstructed data. Users can customize gene selection, color schemes, and visualization thresholds, and easily export the filtered data for further analysis.

4.10. Quantification of precision and recall for landmark and non-landmark gene reconstructions

To assess reconstruction accuracy, we quantified precision and recall for both the landmark genes used to guide the Novosparc optimization and an independent set of non-landmark genes.

For landmark genes, reconstructed expression values were extracted directly from the Novosparc output. Cells were classified as **predicted positive** when their reconstructed expression exceeded the visualization threshold defined for that gene. In addition, we implemented a **Leave-One-Out Cross-Validation (LOOCV)** procedure to evaluate predictive performance: each landmark gene was iteratively removed from the reference set, Novosparc was rerun using the remaining genes, and the predicted spatial pattern of the withheld gene was compared quantitatively to its reference pattern. This procedure was repeated for all landmark genes.

For non-landmark genes, we generated a reference expression atlas based on Simon et al. (2025) (Supplementary Fig. S5). Coordinates were rounded to three decimal places to ensure correspondence with the reconstructed dataset. For each gene, reconstructed expression values (thresholded as above) were matched to the reference atlas based on 3D coordinates. A cell was considered a **true positive (TP)** when both reconstruction and atlas indicated expression of that gene at the same coordinate.

For each gene, precision and recall were computed as.

- **Precision** $= TP/(TP + FP)$,
- **Recall** $= TP/(TP + FN)$,

where FP correspond to reconstructed-positive cells not expressing the gene in the atlas, and FN correspond to atlas-positive cells not recovered in the reconstruction. These metrics were computed independently for each gene and summarized as distributions across all landmark or non-landmark genes. Violin plots overlaid with boxplots were used for visualization.

CRediT authorship contribution statement

Leonardo Tadini: Writing – review & editing, Methodology, Investigation, Conceptualization. **Lilia Younsi:** Writing – review & editing, Methodology, Investigation. **Isabel Holguera:** Writing – review & editing, Methodology, Investigation. **Félix Simon:** Writing – review & editing, Methodology, Investigation. **Maximilien Courgeon:** Writing – review & editing, Methodology, Investigation. **Nikolaos Konstantinides:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization.

Data and resource availability

All data, related code, code to generate Figures, as well as the final resource are available in GitHub (https://github.com/NikosKonst/NovoSparcLarvaDm), GEO (under accession number GSE167266), and in the dedicated webpage (https://larva3dnovosparc.ijm.fr).

Funding

This work is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 949500) and the HORIZON-WIDERA-2023-ACCESS-02 grant no. 101159925 - SCENTINEL. I.H. is supported by a Marie Skłodowska-Curie Postdoctoral Fellowship (101154260).

Declarations of interest

None.

Acknowledgements

We thank the members of the Konstantinides lab and, in particular, Elisavet Iliopoulou for valuable discussions. We also thank Nikos Karaiskos and Shuonan He for help with the implementation of Novosparc and Volker Hartenstein for sharing images of the developing fly brain. We thank Claude Desplan, Jennifer Malin, and Dafni Hadjieconomou for critical reading of the manuscript. Finally, we thank Joel Marchand for informatic assistance during the implementation of the project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ydbio.2026.01.013.

Data availability

All data and related code are available in GitHub (https://github.com/NikosKonst/NovoSparcLarvaDm), GEO (under accession number GSE167266), and in the dedicated webpage https://larva3dnovosparc.ijm.fr.

References

Apitz, H., Salecker, I., 2014. A challenge of numbers and diversity: neurogenesis in the *Drosophila* optic lobe. J. Neurogenet. 28, 233–249.
Apitz, H., Salecker, I., 2015. A region-specific neurogenesis mode requires migratory progenitors in the Drosophila visual system. Nat. Neurosci. 18, 46–55.
Apitz, H., Salecker, I., 2018. Spatio-temporal relays control layer identity of direction-selective neuron subtypes in Drosophila. Nat. Commun. 9 (1), 1–16, 2018 9.
Bertet, C., Li, X., Erclik, T., Cavey, M., Wells, B., Desplan, C., 2014. Temporal patterning of neuroblasts controls Notch-mediated cell survival through regulation of Hid or Reaper. Cell 158, 1173–1186.
Davis, F.P., Nern, A., Picard, S., Reiser, M.B., Rubin, G.M., Eddy, S.R., Henry, G.L., 2020. A genetic, genomic, and computational resource for exploring neural circuit function. eLife 9.
Diao, F., Ironfield, H., Luan, H., Diao, F., Shropshire, W.C., Ewer, J., Marr, E., Potter, C.J., Landgraf, M., White, B.H., 2015. Plug-and-play genetic access to drosophila cell types using exchangeable exon cassettes. Cell Rep. 10, 1410–1421.
El-Danaf, R.N., Kapuralin, K., Rajesh, R., Simon, F., Drou, N., Pinto-Teixeira, F., Özel, M. N., Desplan, C., 2025. Morphological and functional convergence of visual projection neurons from diverse neurogenic origins in Drosophila. Nat. Commun. 16 (1), 1–15, 2025 16.
Erclik, T., Li, X., Courgeon, M., Bertet, C., Chen, Z., Baumert, R., Ng, J., Koo, C., Arain, U., Behnia, R., et al., 2017. Integration of temporal and spatial patterning generates neural diversity. Nature 541, 365–370.
Ferreira, A.A.G., Desplan, C., 2023. An Atlas of the developing Drosophila visual System Glia and subcellular mRNA localization of transcripts in single cells. bioRxiv, 552169, 2023.08.06.
Ferreira, A., Sieriebriennikov, B., Whitbeck, H., 2021. HCR RNA-FISH protocol for the whole-mount brains ofDrosophila and other insects v1. protocols.io.
Filippopoulou, K., Konstantinides, N., 2023. Evolution of patterning. FEBS J.
Fischbach, K.F., Dittrich, A.P., 1989. The optic lobe of Drosophila melanogaster. I. A Golgi analysis of wild-type structure. Cell Tissue Res. 258, 441–445.
Glock, C., Biever, A., Tushev, G., Nassim-Assir, B., Kao, A., Bartnik, I., Dieck, S. tom, Schuman, E.M., 2021. The translatome of neuronal cell bodies, dendrites, and axons. Proc. Natl. Acad. Sci. U. S. A. 118, e2113929118.
Hasegawa, E., Kitada, Y., Kaido, M., Takayama, R., Awasaki, T., Tabata, T., Sato, M., 2011. Concentric zones, cell migration and neuronal circuits in the Drosophila visual center. Development 138, 983–993.
He, S., Shao, W., Chen, S. Cynthia, Wang, T., Gibson, M.C., 2023. Spatial transcriptomics reveals a cnidarian segment polarity program in Nematostella vectensis. Curr. Biol. 33, 2678–2689.e5.
Hobert, O., 2008. Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. Proc. Natl. Acad. Sci. U. S. A. 105, 20067–20071.
Hobert, O., 2011. Regulation of terminal differentiation programs in the nervous system. Annu. Rev. Cell Dev. Biol. 27, 681–696.
Konstantinides, N., Kapuralin, K., Fadil, C., Barboza, L., Satija, R., Desplan, C., 2018. Phenotypic convergence: distinct transcription factors regulate common terminal features. Cell 174, 622–635 e13.
Konstantinides, N., Holguera, I., Rossi, A.M., Escobar, A., Dudragne, L., Chen, Y.C., Tran, T.N., Jaimes, A.M.M., Özel, M.N., Simon, F., et al., 2022. A complete temporal transcription factor series in the fly visual system. Nature 604, 316–322.
Kurmangaliyev, Y.Z., Yoo, J., Valdes-Aleman, J., Sanfilippo, P., Zipursky, S.L., 2020. Transcriptional programs of circuit assembly in the drosophila visual System. Neuron.
Lago-Baldaia, I., Cooper, M., Seroka, A., Trivedi, C., Powell, G.T., Wilson, S.W., Ackerman, S.D., Fernandes, V.M., 2023. A Drosophila glial cell atlas reveals a mismatch between transcriptional and morphological diversity. PLoS Biol. 21, e3002328.
Li, X., Erclik, T., Bertet, C., Chen, Z., Voutev, R., Venkatesh, S., Morante, J., Celik, A., Desplan, C., 2013. Temporal patterning of Drosophila medulla neuroblasts controls neural fates. Nature 498, 456–462.
Malin, J.A., Chen, Y.C., Simon, F., Keefer, E., Desplan, C., 2024. Spatial patterning controls neuron numbers in the Drosophila visual system. Dev. Cell 59, 1132–1145. e6.
Matsliah, A., Yu, S., Kruk, K., Bland, D., Burke, A.T., Gager, J., Hebditch, J., Silverman, B., Willie, K.P., Willie, R., et al., 2024. Neuronal parts list and wiring diagram for a visual system. Nature 634 (8032), 166–180, 2024 634.
Moriel, N., Senel, E., Friedman, N., Rajewsky, N., Karaiskos, N., Nitzan, M., 2021. NovoSpaRc: flexible spatial reconstruction of single-cell gene expression with optimal transport. Nat. Protoc. 16, 4177–4200.
Nériec, N., Desplan, C., Neriec, N., Desplan, C., 2016. From the eye to the brain: development of the drosophila visual System. Curr. Top. Dev. Biol. 116, 247–271.
Nern, A., Loesche, F., Takemura, S.Y., Burnett, L.E., Dreher, M., Gruntman, E., Hoeller, J., Huang, G.B., Januszewski, M., Klapoetke, N.C., et al., 2025. Connectome-driven neural inventory of a complete visual system. Nature 641 (8065), 1225–1237, 2025 641.
Ngo, K.T., Andrade, I., Hartenstein, V., 2017. Spatio-temporal pattern of neuronal differentiation in the Drosophila visual system: a user's guide to the dynamic morphology of the developing optic lobe. Dev. Biol. 428, 1–24.
Nitzan, M., Karaiskos, N., Friedman, N., Rajewsky, N., 2019. Gene expression cartography. Nature 576, 132–137.
Özel, M.N., Simon, F., Jafari, S., Holguera, I., Chen, Y.C., Benhra, N., El-Danaf, R.N., Kapuralin, K., Malin, J.A., Konstantinides, N., et al., 2021. Neuronal diversity and convergence in a visual system developmental atlas. Nature 589, 88.
Özel, M.N., Gibbs, C.S., Holguera, I., Soliman, M., Bonneau, R., Desplan, C., 2022. Coordinated control of neuronal differentiation and wiring by sustained transcription factors. Science 378.
Roberts, R.J.V., Pop, S., Prieto-Godino, L.L., 2022. Evolution of central neural circuits: state of the art and perspectives. Nat. Rev. Neurosci. 23, 725–743.
Sato, M., Yasugi, T., Minami, Y., Miura, T., Nagayama, M., 2016. Notch-mediated lateral inhibition regulates proneural wave propagation when combined with EGF-mediated reaction diffusion. Proc. Natl. Acad. Sci. U. S. A. 113, E5153–E5162.
Schilling, T., Ali, A.H., Leonhardt, A., Borst, A., Pujol-Martí, J., 2019. Transcriptional control of morphological properties of direction-selective T4/T5 neurons in Drosophila. Development 146.
Simon, F., Konstantinides, N., 2021. Single-cell transcriptomics in the Drosophila visual system: advances and perspectives on cell identity regulation, connectivity, and neuronal diversity evolution. Dev. Biol. 479, 107–122.
Simon, F., Holguera, I., Chen, Y.-C., Malin, J., Valentino, P., Njoo-Deplante, C., El-Danaf, R.N., Kapuralin, K., Erclik, T., Konstantinides, N., et al., 2025. Establishment of terminal selector combinations in optic lobe neurons. bioRxiv, 578975, 2024.02.05.
Suzuki, T., Hasegawa, E., Nakai, Y., Kaido, M., Takayama, R., Sato, M., 2016a. Formation of neuronal circuits by interactions between neuronal populations derived from different origins in the Drosophila visual Center. Cell Rep. 15, 499–509.
Suzuki, T., Trush, O., Yasugi, T., Takayama, R., Sato, M., 2016b. Wnt signaling specifies anteroposterior progenitor Zone identity in the Drosophila visual Center. J. Neurosci. 36, 6503–6513.
Tosches, M.A., 2017. Developmental and genetic mechanisms of neural circuit evolution. Dev. Biol. 431, 16–25.
Tsuneoka, Y., Funato, H., 2020. Modified in situ hybridization chain reaction using short hairpin DNAs. Front. Mol. Neurosci. 13, 75.
Valentino, P., Erclik, T., 2022. Spalt and disco define the dorsal-ventral neuroepithelial compartments of the developing Drosophila medulla. Genetics 222.
van Oostrum, M., Schuman, E.M., 2025. Understanding the molecular diversity of synapses. Nat. Rev. Neurosci. 26, 65–81.
Yasugi, T., Sugie, A., Umetsu, D., Tabata, T., 2010. Coordinated sequential action of EGFR and notch signaling pathways regulates proneural wave progression in the Drosophila optic lobe. Development 137, 3193–3203.
Zhu, H., Zhao, S.D., Ray, A., Zhang, Y., Li, X., 2022. A comprehensive temporal patterning gene network in Drosophila medulla neuroblasts revealed by single-cell RNA sequencing. Nat. Commun. 13.