Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht – Karls – University

Heidelberg

Presented by

M.Sc. Victoria Alicia Witte

born in: Oberhausen, Germany

Oral examination: 12.12.2025

A single - cell transcriptomics investigation of the adult *Platynereis dumerilii* mushroom bodies

Referees:

Dr. Hanh Vu

Prof. Dr. Steffen Lemke

Dedicated to my inner child.

You can unclench your jaw now.

“Once you kill a cow, you gotta make a burger”

- Lady Gaga

Summary

Candidate gene approaches have positioned the *Platynereis dumerilii* mushroom bodies as an interesting evolutionary case study. With resemblances to both the vertebrate telencephalon and to similarly structured insect mushroom bodies, the neurons making up these brain regions have been thought to stem from an ancestral cell type family. Single cell transcriptomics is enabling the investigation of such evolutionary questions based on a cell's *entire* repertoire of expressed genes. The present investigation used single *nucleus* RNA sequencing to characterise the neurons of this structure in adult *P.dumerilii*. This revealed a gene signature comprising known and novel genes that can now be compared to closely and more evolutionarily distant species. This approach also uncovered the stage - specific expression of a ciliary opsin - *cOpsin1* - and a possible *emx* paralog - *emx2* - in a subset of these neurons, which appear to undergo dramatic restructuring upon sexual metamorphosis. Incorporating publicly available datasets of similar stages further showed that this subset of neurons may reflect vertebrate - like mechanisms of adult neurogenesis influenced - either directly or indirectly - by *cOpsin1*. These findings are interpreted in light of recent work characterising a neurogenic region adjacent to the rhabdomeric eyes and responsible for their growth during sexual metamorphosis, a process that *also* seems to involve the expression of *cOpsin1* (Milivojev et al., 2024). In addition to this unexpected finding, the present work attempts to compare the overall mushroom body gene signature across distant species, finding *mef2c* to be a common gene shared among *P.dumerilii* and insect mushroom bodies, as well as cephalopod vertical lobe neurons. It also highlights how the diversity within Annelida can help resolve outstanding questions of whether these neurons represent an ancestral cell type and how future transcriptomic investigations on diverse annelid brains will provide macro- and microevolutionary insights.

Zusammenfassung

Kandidatengen-Ansätze haben die Pilzkörper von *Platynereis dumerilii* zu einem interessanten Fallbeispiel für die Evolutionsforschung gemacht. Aufgrund ihrer Ähnlichkeiten sowohl mit dem Telencephalon von Wirbeltieren als auch mit ähnlich strukturierten Pilzkörpern von Insekten wurde angenommen, dass die Neuronen, aus denen diese Teile des Gehirns bestehen, von derselben Zelltyp-Familie des letzten gemeinsamen Vorfahren abstammen. Die Einzelzell-Transkriptomik ermöglicht die Untersuchung solcher evolutionären Fragen auf der Grundlage des gesamten Repertoires der exprimierten Gene einer Zelle. In der vorliegenden Untersuchung wurden die Neuronen dieser Struktur in adulten *P. dumerilii* mittels Einzelkern-RNA-Sequenzierung charakterisiert. Dabei wurde eine Gensignatur aus bekannten und neuen Genen identifiziert. Dieser Ansatz deckte auch die stadienspezifische Expression des Sehpigments *cOpsin1* und eines emx-Paralogs – *emx2* – in einer Untergruppe dieser Neuronen auf, die sich nach der sexuellen Metamorphose dramatisch umstrukturieren. Die Einbeziehung verfügbarer Datensätze ähnlicher Stadien zeigte darüber hinaus, dass diese Untergruppe von Neuronen möglicherweise Mechanismen der adulten Neurogenese aufweist, die denen der Wirbeltiere ähneln und direkt oder indirekt durch *cOpsin1* beeinflusst werden. Diese Ergebnisse werden im Zusammenhang mit neuen Arbeiten interpretiert, die eine *c-opsin1*-abhängige neurogene Region beschrieben haben, die an die rhabdomerischen Augen angrenzt und für deren Wachstum während der sexuellen Metamorphose verantwortlich ist (Milivojev et al., 2024). Zusätzlich zu diesem unerwarteten Ergebnis versucht die vorliegende Arbeit, die Gensignatur der Pilzkörper zwischen entfernten Arten zu vergleichen, und stellt fest, dass *mef2c* ein gemeinsames Gen ist, das sowohl bei *P. dumerilii* als auch bei Insekten sowie bei Neuronen des Vertikallappens von Kopffüßlern vorkommt. Sie unterstreicht auch die Bedeutung vergleichender Einzel-Transkriptomik an verschiedenen Annelidengehirnen für die makro- und mikro-evolutionäre Forschung.

Acknowledgements

I cannot have done this without the constant support of the Arendt lab current and former members. These people are more than colleagues; they are friends, emotional support people, field coordinators, professional stunt drivers, and all around wizards. They have made these few years incredibly memorable. I'd like to thank Detlev for giving me the chance to explore so many different avenues during my time in the lab. Thank you for taking a chance on this small electrophysiologist wanting to explore new horizons. Thank you to my thesis committee members Hanh, Mike, and Steffen for their constant input and encouragement. Thank you Leslie and Phil as my first guides into the world of Platy… on the field no less. I probably should have run away after my first week, but I know good people when I meet them. Thank you for guiding me through this entire journey, through the chaotic fieldwork, decompression meetings, and world of single-cell sequencing, and thank you for building my confidence in the lab. Thank you to Ido, Anto, Tobi, and Luca for their constant support and becoming more than colleagues. Thank you to my student Gülce for becoming a friend and climbing buddy despite all the HCRs. Thank you my chaos corner pals, Kevin (iykyk) and Fabi, I think I focused more once you left, but my heart was emptier for it. Thank you Alvin for joining the chaos corner and for helping me grapple with the chaos that is the Platynereis species complex. I am almost sorry that I ever suggested genotyping the tube-dwelling worms we had in the lab. Thank you to Matt for his image analysis genius in the last stretch of my thesis, it turns out segmenting mushroom bodies is harder than we thought so enjoy the qualitative data for now. Thank you wine and thank you dogs. Thank you to my family for not telling me this was a bad idea, you are too proud of me. Lastly, thank you dear reader, let us swim through this together.

Table of Contents

Summary....5
Acknowledgements....9
Table of Contents....11
Abbreviation List....14
1. Introduction....15
1.1. How to define cell types....15
1.2. *Platynereis dumerilii* as a model....17
1.3. Aims of this thesis....19
2. A transcriptomic exploration of the maturing *P.dumerilii* brain....20
2.1. Contributions....20
2.2. Background....21
2.2.1. The *P.dumerilii* mushroom body as an interesting case study....21
2.2.2. Gene signature for *P.dumerilii* MB neurons....21
2.3. Materials and methods....24
2.3.1. Animal culture and sampling....24
2.3.2. Nuclei extraction....24
2.3.3. Single nuclei capture and library preparation....25
2.3.4. Single cell analysis....26
2.3.4.1. Quality Control and Seurat analysis - whole head dataset....27
2.3.4.2. Neuronal subset....28
2.3.4.3. Mushroom body subset....29
2.3.5. In situ validation....30
2.3.5.1. Whole - Mount Hybridization Chain Reaction (HCR)....30
2.3.5.2. Whole - Mount Immunofluorescence (IF)....32
2.3.5.3. Nuclear stains....33
2.3.5.4. Cryosectioning & Hematoxylin/Eosin stain....33
2.4.6. Confocal Imaging and Processing....34
2.4. Results....35
2.4.1. snRNAseq captures cell type diversity in the *P.dumerilii* brain....35
2.4.2. A distinct transcriptomic signature for MB neurons....36
2.4.3. Neuronal subset reveals three *cOpsin1* expression domains....37
2.4.4. Gene signatures of the MB subpopulations....40

2.4.5. Ventral MB expresses proneural genes in premature worms....41
2.4.6. Effect of *cOpsin1* on the ventral MB....41
2.4.7. Ventral MB *cOpsin1*-expression is stage specific....43
2.4.8. Restructuring of MBs upon sexual metamorphosis....44
2.5. Summary and perspectives....45
2.5.1. Summary....45
2.5.2. Linking *cOpsin1* and *P.dumerilii* sexual metamorphosis....51
2.5.3. A proposed *cOpsin1* - dependent scheme for MB restructuring....52
2.5.4. Future perspectives....53
3. Beyond candidate genes: Comparing *Platynereis dumerilii* mushroom body neurons to distant species....55
3.1. Contributions....55
3.2. Background....56
3.2.1. How to compare across large evolutionary distances....56
3.2.2. Selected comparison species....57
3.3. Materials and methods....58
3.3.1. Collapsing emapper predicted proteins into Orthology groups....58
3.3.2. Cross-species integration analysis....59
3.4. Results....61
3.4.1. Orthogroup - clustering maintains original clusters....61
3.4.2. *P.dumerilii* vs *O.vulgaris* vs *H.saltator* comparison....61
3.4.3. Marker Orthogroup identifiers....62
3.5. Summary and perspectives....64
4. Lessons from the field and genotyping the *Platynereis spp*. species complex....66
4.1. Contributions....66
4.2. Background....67
4.2.1. The annelid brain from a macro - and microevolutionary perspective....67
4.2.2. Classifying samples from the TREC expedition....69
4.3. Materials and methods....71
4.3.1. Annelid field collection....71
4.3.2. Genotyping....71
4.4. Results....75
4.5. Summary and perspectives....76
5. Conclusion....78
5.1. Insights into the maturing annelid mushroom bodies....78

5.2. Future perspectives 79
Bibliography 80
Appendix 98
Supplementary Figures 98
Supplementary Tables 99
Table of Buffer compositions 107
RMarkdown files 108

Abbreviation List

aNSC: adult neural stem cell
AMP: Amplification buffer
CCA: canonical correlation analysis
cPRCs: ciliary photoreceptors
DAPI: 4′6- Diamidin-2-phenylindol
Dpf: days post fertilisation
EdU: 5-ethynyl-2′-deoxyuridine
EtOH: ethanol
FASW: filtered artificial sea water
GPCR: G-coupled protein receptor
HB: homogenization buffer
HYB: hybridization buffer
Hpf: hours post fertilisation
IF: Immunofluorescence
KC: Kenyon cell
KO: knockout
MB: mushroom body
MeOH: methanol
MS: mechanosensory
NIT-GC2: Nitric oxide-activated guanylate cyclase 2
NS: neurosecretory
OG: Orthology group / Orthogroup
PBS: phosphate-buffered saline
PFA: paraformaldehyde
PK: proteinase K
PRC: photoreceptor
PTW: PBS with Tween-20
rPRCs: rhabdomeric photoreceptors
SMS: selected model species
snRNAseq: single nucleus RNA sequencing
SSCT: saline-sodium citrate buffer with Tween-20
VL: vertical lobe
WB: probe wash buffer
WM - HCR: whole - mount hybridization chain reaction

1. Introduction

In order to understand how the diversity of complex brains present in extant species arose in evolution, it is important to understand what the brain of a Bilaterian ancestor may have looked like. This can be explored by comparing the brains of diverse species at different levels: anatomical structures, cell types, and molecular mechanisms. This thesis will focus on cell types as the unit of comparison across species. Such a comparative approach can elucidate which neuronal cell types are common among extant brains, and which ones are species-specific, serving as a proxy to understand the composition of a hypothetical Urbilaterian brain that gave rise to present-day diversity (Arendt et al., 2019). In order for this approach to work, defining what is meant by a *cell type* is critical since it provides the unit of comparison. This definition can come in many forms, all with their own merits.

1.1. How to define cell types

While on the surface a trivial question, defining cell types, particularly neuronal cell types, is all but that. Historically, morphology has not only been able to classify the diverse neurons of the brain, but was critical in understanding that the nervous system was made up of individual connected units - conductive neurons - at all, rather than one continuum (Llinás, 2003; Ramón y Cajal, 1906; Ramón y Cajal & Azoulay, 1955). Fast forward to today, detailed cell morphology is still an incredibly useful classifier of cell types that can even be used by unsupervised neural networks to successfully group cells; matching groupings obtained based on gene expression values (Zinchenko et al., 2023). Single cell transcriptomics can now further expand the definition of a cell type to include all of the genes it expresses. Such methods make it possible to tease apart the regulatory mechanisms that in part determine a cell's function and aforementioned morphology. The concept of terminal selectors, meaning the transcription factors or complexes that define, in this case a neuron, by activating the gene program that will result in the expression of ion channels, receptors, and enzymes

needed for its function (Hobert, 2016), is useful for defining neuron types based on such regulatory programs. It is especially useful for comparing neuron types across long evolutionary distances (Arendt et al., 2016, 2019). **Figure 1** schematises the definition put forward in Arendt et al. (2019), which suggests that related cell types may share part of a regulatory program inherited from a common ancestor and can be considered a *cell type family*, with an expectation that older families will be seen across many species, while newer sub families or cell types will be phylum , genus, or even species - specific.

Reconciling these transcriptomic signatures with neuron morphology and function, however, has become a nuanced endeavour (Özel & Desplan, 2025). For example, some morphologically and functionally distinct neurons are actually transcriptomically similar, as seen in the zebrafish tectum (Shainer et al., 2025). Furthermore, the lifelong processes of development present an ever changing transcriptomic signature, adding another variable to consider in the quest to define cells by terminal selectors (Arendt et al., 2019; Oel et al., *in prep.*). High resolution connectomics is able to classify each bilateral pair of neurons in *P.dumerilii* as a different cell type, but this level of detail is only achieved when considering a small percentage of neurons, omitting differentiating neurons and those with few synapses (Jékely et al., 2024). Yet, such developmental cell states can, on the contrary, provide great comparative insights, as in the evolution of the telencephalon, where similarities in embryonic stages of sharks and mice make it possible to infer relatedness between neuron progenitor populations (Quintana-Urzainqui et al., 2025). In flies, certain cell types can *only* be distinguished by their transcriptomes in earlier developmental stages, becoming indistinguishable in adult stages (Li et al., 2017). Despite this nuance, comparisons between different nematode species have shown that transcription factor expression and neurotransmitter usage seem to remain stable among comparable neuronal cell types (Toker et al., 2024), information that can be obtained from single cell sequencing approaches. This does not seem to be the case for neurotransmitter receptors and neuropeptidergic signaling, which is more variable (Toker et al., 2024). On the other hand, terminal effector genes

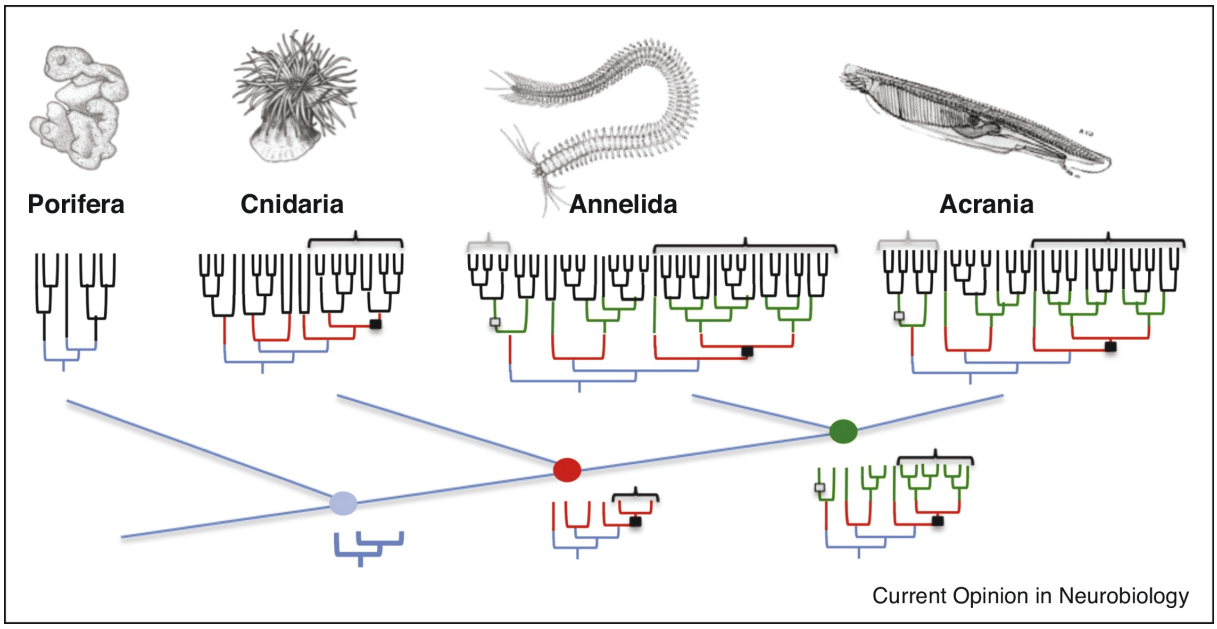


Figure 1. A hypothetical depiction of cell type evolution from Arendt et al. (2019). Blue branches depict oldest cell type families likely to be most conserved across long evolutionary distances. Black branches represent the species-specific cell types that are part of larger conserved cell type families.

(hormones, neuropeptides) were reported to indeed be rather constant between different teleost fish species (Shafer et al., 2022).

Bearing this nuance in mind, single cell transcriptomics remains a practical approach to define cell types based on their regulatory programs and draw evolutionary inferences about ancestral cell types and families, since it is hypothesized that across large distances cell type families rather than specific cell types will be conserved (Arendt et al., 2019; Musser et al., 2021). It is also an accessible approach for species yet to be functionally characterized in detail.

1.2. *Platynereis dumerilii* as a model

As previously mentioned, to draw evolutionary inferences about related neuron types and families present in the brains of extant bilaterians, it is important to sample a variety of species spanning different phyla. Well-established models - mouse, fish, fly - exist among the deuterostomes and ecdysozoans, and such models have been expanded to include lophotrocozoan species such as *Platynereis dumerilii* - a marine annelid with spiral cleavage (Arendt et al., 2021; Fischer et al., 2010). This model can be cultured in the lab and also collected from the wild during either its pelagic spawning (epitoke) stage or during its benthic tube-dwelling (atoke) stage (Fischer & Dorresteijn, 2004; Legras et al., 2023; Özpolat al., 2021). **Figure 2** depicts this life cycle (Legras et al., 2023). This life cycle consists of three dramatic metamorphoses described in detail by others (A. Fischer & Dorresteijn, 2004; A. H. Fischer et al., 2010; Hauenschild, 1966) and summarised as follows. The first occurs when the planktonic larvae settle and begin their benthic growth phase of eating and adding segments. The second, cephalic metamorphosis, occurs soon thereafter and involves the incorporation of the first trunk segment into the head, at which point the body plan is set until sexual metamorphosis. During this time the worm remains on the benthos in a self - spun tube, feeding and adding segments. Finally, at around 70 segments, gametes begin

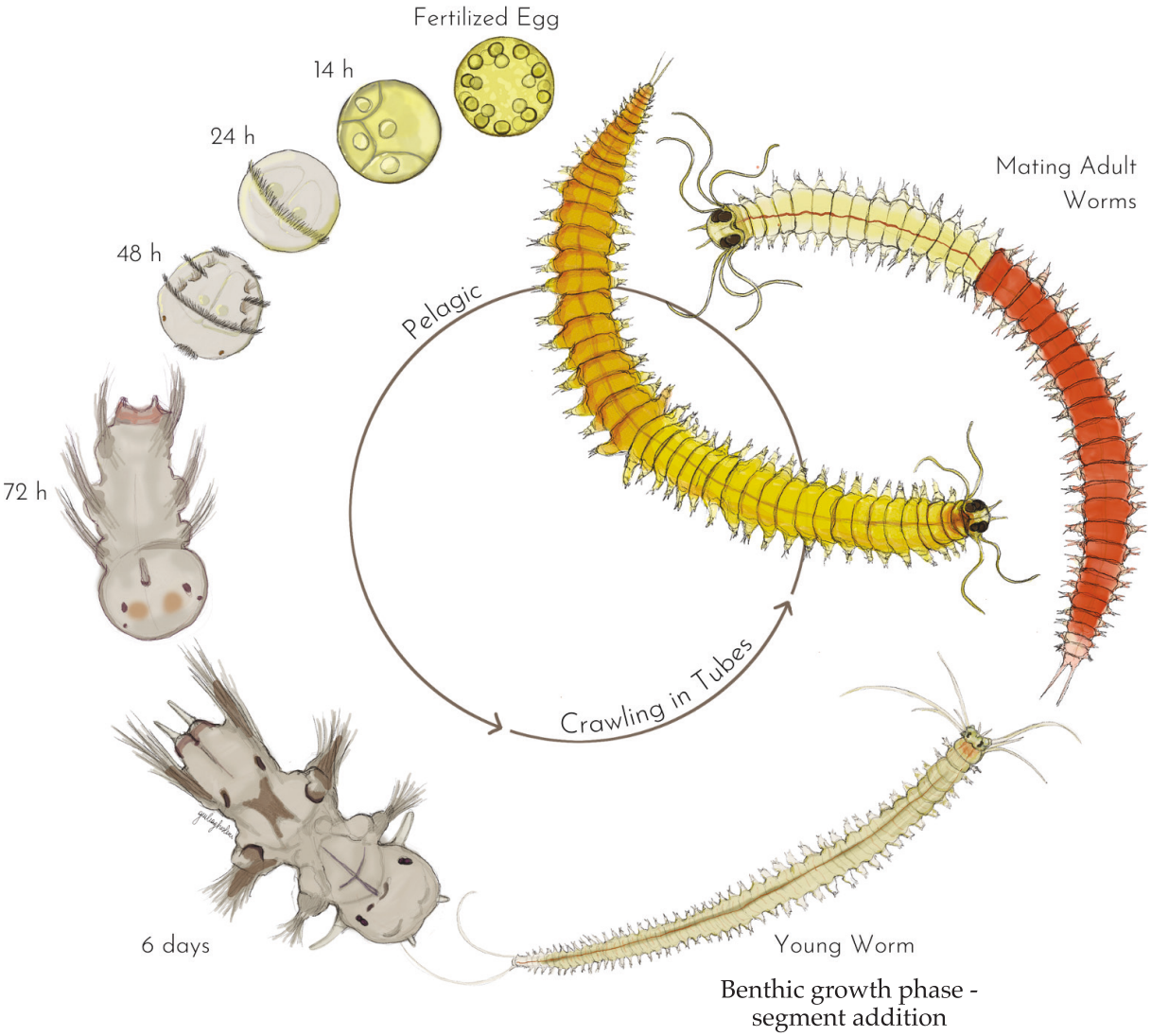


Figure 2. A schematic representation of the *P.dumerilii* life cycle. Artwork by Giulia Ghisleni from Legras et al. (2023). Note the pelagic - to - benthic transition followed by a return to a pelagic lifestyle after sexual metamorphosis.

maturing in the body cavity. Soon after, food intake stops and the final metamorphosis - sexual metamorphosis - takes place. This is when the rhabdomeric eyes grow in size, the dorsal chromatophores disappear, the females appear yellow in colour due to the oocytes they carry, while the males have a white anterior section and red posterior section indicative of sperm cells and muscle, respectively. Only at this point is it possible to *visually* distinguish reproductive male versus female worms, although earlier transcriptomic differences are apparent (Ribeiro et al., 2024).

Via an array of opsins and cryptochrome proteins, this maturation process occurs in time with the circadian and monthly lunar cycles, in order to synchronise spawning events with the new moon (Hauenschild, 1960; Pende et al., 2020; Poehn et al., 2022; Wulf et al., 2025; Zantke et al., 2013; Zurl et al., 2022). This final metamorphosis has long been known to be triggered via brain hormones, such that headless worms matured *faster,* indicating that the reduction of brain hormone release is needed for sexual metamorphosis to occur (Hauenschild, 1966). The hormone methylfarnesoate was later discovered as a key player in this process, showing the ability to directly suppress oocyte maturation and either directly or indirectly maintain a worm's premature - specific posterior regenerative ability (Schenk et al., 2016). Together with recent work showing that *gnRH* - like (*gonadotropin - releasing hormone*) peptides get upregulated in the brains of reproductive worms, and at critical points in the lunar cycle (Andreatta et al., 2020), it is clear that an interplay of metabolic signaling, lunar cycle, as well as seasonal changes (Veedin Rajan et al., 2021), influence the timing of sexual metamorphosis.

Overall, with a body plan resembling an early Cambrian fossil polychaete (Parry & Caron, 2019) and an intron - rich gene structure resembling that of humans (Raible et al., 2005), *P.dumerilii* is a useful model to include when exploring neuronal cell type evolution.

1.3. Aims of this thesis

In this thesis, I aim to broadly characterize the neuronal cell types of the maturing *P.dumerilii* brain using single nucleus transcriptomics, which will be the focus of Chapter 2. I focus on one neuronal cell type - the mushroom body interneurons - that, according to candidate gene approaches and previous functional studies, are likely involved in associative functions in the brain of *P.dumerilii* (Arendt et al., 2021; Chartier et al., 2018; Tomer et al., 2010; Vergara et al., 2021). With this gene signature in hand, in Chapter 3, I attempt to compare this cell type to other bilaterian species to explore whether it - or at least molecular mechanisms used by it - may indeed be ancestral. The fourth chapter of this thesis explores these insights from a microevolutionary perspective, by setting the foundation for further transcriptomic comparisons among the brains of annelids living in various environmental contexts. It touches on the diversity of brain structures that exists even within Annelida, and how future transcriptomic comparisons among the brains of diverse annelids - including those within the *Platynereis spp.* species complex - will further resolve outstanding questions of whether the cell types introduced in earlier chapters are indeed ancestral. For this to work, careful classification of the specimens of this species complex is an important first step.

2. A transcriptomic exploration of the maturing *P.dumerilii* brain

2.1. Contributions

As with this entire thesis, this chapter would not have been possible without the support of my EMBL colleagues and lab members. I did the downstream analyses myself with the guidance of Phil Oel, Tobias Gerber and Niko Papadopoulos and others. Initial 10x captures were done by Leslie Pan and Phil Oel. I joined the team for dissections, extractions and captures after the sn022/23 library. Library preparation was done by Leslie Pan. Tobias Gerber mapped sequencing reads to the Platynereis genome and disentangled multi-species captures. R analysis code was based on Seurat pipelines and custom functions from Phil Oel and Niko Papadopoulos, who, along with Tobias Gerber, guided my analysis. OpenAI technology assisted with quick code modifications or errors. I performed the relevant *in situ* validation, nuclear stains, and imaging myself with the guidance of Luca Santangeli, Matt Benton and members of the ALMF, as well as image processing tips from Phil Oel, Idoia Quintana - Urzainqui and Leslie Pan. Lastly, the preprint dataset from Milivojev et al. (2024) was incredibly helpful in the interpretation of my findings and reassuring to see that different labs, methods, and worm cultures led to reproducible results.

2.2. Background

2.2.1. The *P.dumerilii* mushroom body as an interesting case study

The *P.dumerilii* mushroom bodies (MB) are thought to be a centre for information integration and association since, functionally, they respond as part of a chemosensory circuit with a slight delay after stimulus onset (Chartier et al., 2018). They express patterning genes similar to the vertebrate pallium (Tomer et al., 2010; Vergara et al., 2021), and in older post-cephalic metamorphic developmental stages they structurally resemble insect mushroom bodies (Heuer et al., 2010, 2012; Tomer et al., 2010). More specifically, the anatomical structure of the MB - with dense nuclei forming dendritic connections in distinct calyces and axonal tracts exiting via dorsal and ventral peduncles - resembles those of insects both at 6dpf (Vergara et al., 2021) and more clearly in adult stages, where this region is easily distinguishable from surrounding nuclei (Heuer et al., 2010, 2012; Tomer et al., 2010). Regarding a transcriptomic signature, already at 48hpf the mediolateral as well as anterior-posterior anlage of the anterior brain - defined by *pax6*, *nkx2.1/2.2* and *otx* respectively - is similar to mouse pallium development, with the MBs developing from an *emx* - expressing subregion of the *pax6* region (Tomer et al., 2010). The similarities between early patterning genes of the vertebrate pallium and the annelid MB are depicted in **Figure 3A - C**. The structural and transcriptomic similarities to distant model Bilaterians make the neurons of the *P.dumerilii* MBs a good candidate to classify and compare using single nucleus RNA sequencing (snRNAseq).

2.2.2. Gene signature for *P.dumerilii* MB neurons

Candidate gene approaches offer good insight into what genes to look for when using single cell transcriptomic datasets to define the MB neurons. As mentioned above, in early development, the *P.dumerilii* MBs emerge from an *emx* and *pax6* expressing anterior brain region (Tomer et al., 2010). Later at 6 dpf (days post fertilisation), this region begins

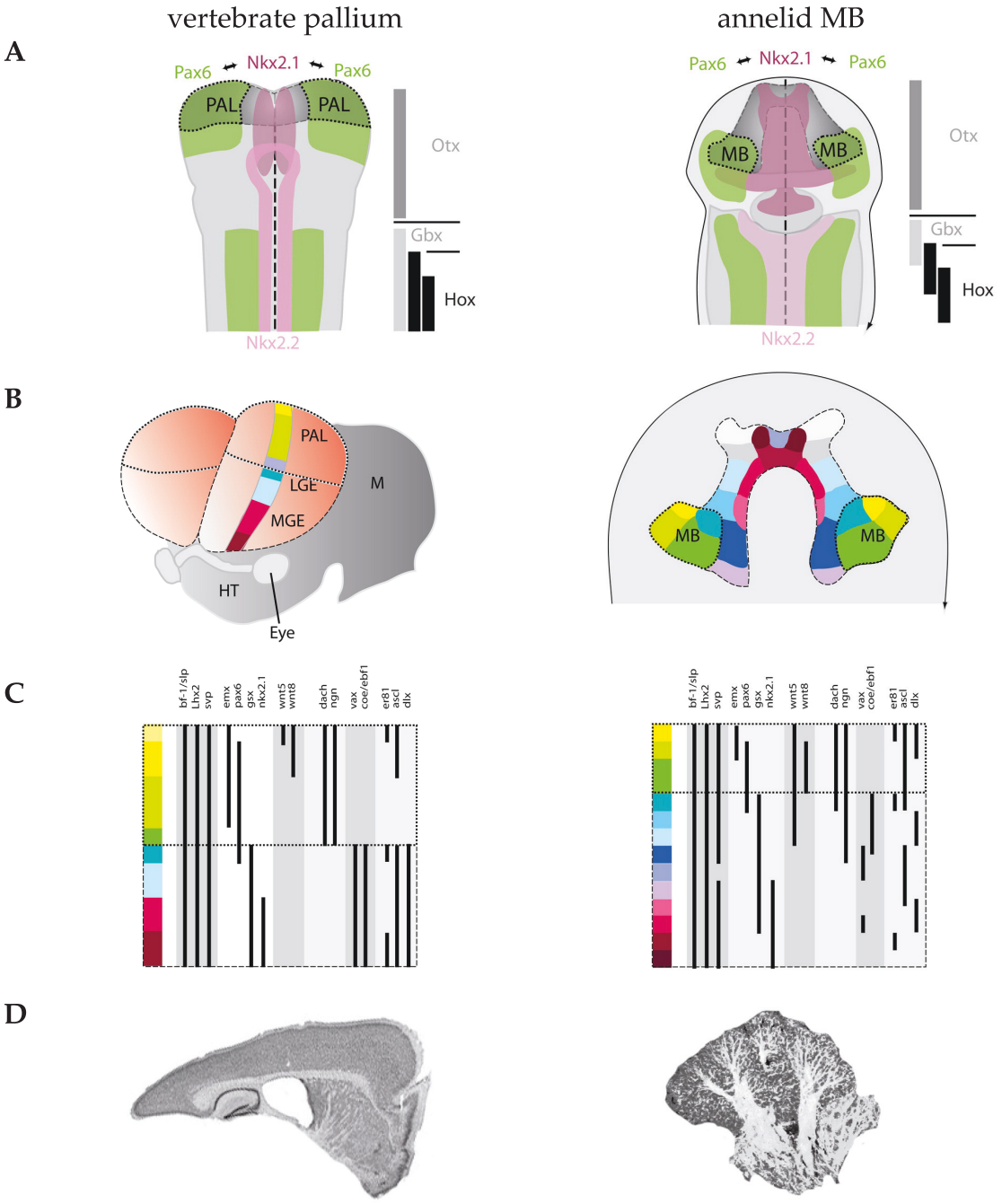


Figure 3. Schematic comparison of vertebrate pallium (left panel) development and annelid MB (right panel) development. Figure from Tomer et al. (2010) with minimal adaptations. (**A**) Anterior - posterior (*otx, gbx, hox*) and mediolateral (*pax6, nkx2.1, nkx2.2*) patterning genes are similar in both vertebrate and annelid early development, with the pallium and the MBs emerging from the anterior *pax6+* region. Regionalisation of the vertebrate telencephalon and annelid anterior brain as (**B**) a schematic representation and (**C**) colour - coded gene signatures matching B. Note the specificity of *pax6* and *emx* to both the vertebrate pallium and the annelid MBs. (**D**) Parasagittal sections showing the resulting divergent adult structures. Abbreviations: MB: mushroom bodies, PAL: pallium, LGE and MGE: lateral and medial ganglionic eminences, M: mesencephalon, HT: hypothalamus.

expressing *arx*, part of it remaining proliferative - labeled by EdU and expression of *soxb2* paralogs (Arendt et al., 2021; Vergara et al., 2021). There is also co-expression of *arx* and *ptf1* in a portion of the 6dpf MB - the dorsal calyx of the ventral peduncle - at the 6dpf stage (Vergara et al., 2021). Furthermore, these neurons express vesicular glutamate as well as acetylcholine transporters, indicating possible co-transmission (Arendt et al., 2021). They also express neurosecretory markers like prohormone convertase 2 (*phc2)* (Achim et al., 2015; Vergara et al., 2021).

The transcription factor signature - *arx, dlx, lhx6/8* and *nkx2.1* in part - of the 6dpf *P.dumerilii* MBs is also reminiscent of genes specifying broader vertebrate GABAergic interneuron families that populate the telencephalon (Arendt et al., 2019, 2021; Zeisel et al., 2018). Of these transcription factors, *arx* is responsible for the fate specification and migration of cortical GABAergic interneurons in vertebrates (Marsh et al., 2016; Tsuboi & Yoshihara, 2025). The *P.dumerilii* MB interneurons, however, consistently show cholinergic neurotransmitter identity, and possible co-transmission with glutamate, more reminiscent of vertebrate *cholinergic* striatal interneurons which do not show evidence of being determined by *arx*, but do show a role for *lhx8, nkx2.1* in their specification (Ahmed et al., 2019; Kljakic et al., 2017). Interestingly, the *arx+ cholinergic* interneurons are characteristic of the more closely related cephalopod associative vertical lobe region (Bidel et al., 2023; Styfhals et al., 2022).

Another characteristic marker, *ptf1a,* is known in vertebrates to specify inhibitory neurons outside of the telencephalon in the cerebellum and spinal cord (Glasgow et al., 2005; Hoshino et al., 2005; Leto et al., 2008), while also playing a role in neuronal maturation of the vertebrate forebrain (Fujiyama et al., 2018). Interestingly, when considering the whole 6 dpf larva, *ptf1a* seems to also specify GABAergic interneurons in other regions of the body outside the MBs (Oel et al., *in prep.*). Overall, it appears that the *P.dumerilii* MB neurons comprise gene signatures from different vertebrate interneuron populations, in addition to reflecting telencephalic patterning genes as described previously (Arendt et al., 2021; Tomer

et al., 2010; Vergara et al., 2021). Compared to insect MBs, Tomer et al. (2010) showed that in early development, while both *P.dumerilii* and *Drosophila* MBs share many patterning genes - including the transcription factor *tailless* (accession no.: GU169423.1) - the *emx* and *arx* expression seems to be uniquely shared between annelids and vertebrates.

With this information in mind, in this chapter I will explore whether these aforementioned transcriptomic signatures can pinpoint MB neurons in a snRNAseq dataset for further evolutionary comparisons based on entire transcriptomes. In order to enrich cells of the MB region I investigated the maturing *P.dumerilii* stages, which are amenable to head dissections, and produced a dataset of the adult brain across sexual metamorphosis. This will complement the thus far anatomical characterisations of adult annelid MBs (Heuer et al., 2010; Tomer, 2008), defining a transcriptomic signature for the neurons of this region for future comparative work. As I will show, this exploratory approach also revealed exciting insights into an adult neural stem cell (aNSC) - like state of the ventral MB neurons that persists into adult stages.

2.3. Materials and methods

2.3.1. Animal culture and sampling

Platynereis dumerilii were grown in rectangular plastic containers containing filtered artificial sea water (FASW) at 18°C under an artificial moon cycle - three weeks of darkness, followed by one week of artificial moonlight (Fischer & Dorresteijn, 2004; Hauenschild, 1960; Kuehn et al., 2019). Worms over 2 cm that still contained food in their digestive tract were selected for all experiments. For sequencing experiments these large premature worms were all over 55 segments. The heads of sexually mature males and females were obtained from worms taken from spawning events in the lab culture. Worms of similar size were sampled for *in situ* validation experiments. For *cOpsin1* stainings, I specifically sampled worms above and below 60 segments across a small time course spanning the shift from darkness to moonlight (day 1: moon off, day 2: moon turned on, day 3: one day with moon on). Sexually mature worms used for imaging experiments were selected based on colour change (**Figure 2**) as well as actively spawning worms to match the single nucleus captures.

2.3.2. Nuclei extraction

This protocol was adapted from Lamanna et al. (2023), the adapted version is recently reported in Quintana-Urzainqui et al. (2025). On different occasions, four premature worms over 55 segments, as well as four sexually mature worms (2 males and 2 females) collected from a spawning event, were anaesthetized in a 1:1 solution of 7.5% MgCl mixed with FASW (**Figure 4A**, schematic workflow). Snapshots were taken at a brightfield Leica MZ16 dissection scope in order to later count the segments using the Adobe Photoshop 2024 (25.12.0 release) counting tool. Heads were dissected above the pharynx, just behind the nuchal organs. In all except the one dissection (sn022/23), sensory appendages (cirri, antennae, and palps) were trimmed in order to enrich cells of the central brain. Immediately after dissection, head tissues were placed in 1.5ml Eppendorf LoBind tubes and flash-frozen

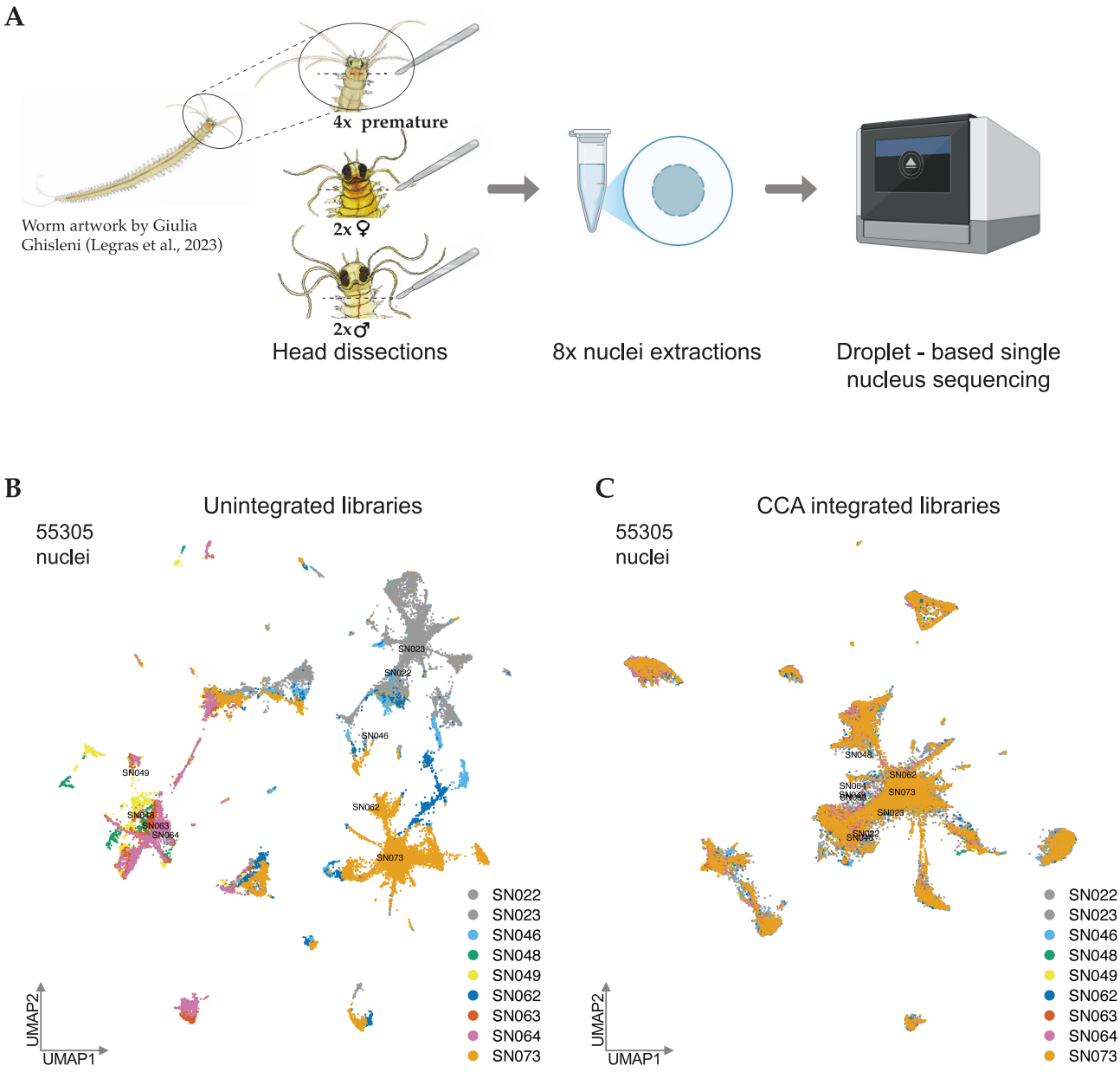


Figure 4. Experiment scheme and snRNAseq library integration. (**A**) Experimental schematic, worm artwork from Legras et al. (2023), remaining experimental scheme created with BioRender.com (**B**) Unintegrated UMAP dimensional reduction showing separation based on maturation stage (premature: SN022, SN023, SN046, SN062, SN073, sexually mature: SN048, SN049, SN063, SN064), and experimental batch effect especially in the case of library SN073. (**C**) UMAP dimensional reduction after Seurat CCA integration.

in liquid nitrogen for storage or immediate use. On ice, tissues were homogenised for 5 minutes in 300µl of Homogenization buffer (HB) using micro pestles (Axygen). All following steps were performed on ice and using a tabletop centrifuge (Eppendorf) at 4 °C. Any unlysed tissue was pelleted by centrifugation at 100 x g for 1 minute. The supernatant - containing nuclei - was transferred to a new tube and centrifuged at 400 x g for 4 minutes. The supernatant was discarded and the nuclei - containing - pellet was washed and resuspended in 300µl of fresh HB buffer. Nuclei were pelleted again at 400 x g for 4 minutes. The supernatant was again discarded and the pellet of nuclei was resuspended in 50 - 100µl of ice cold Dulbecco's PBS supplemented with 1µM DTT, 0.4 U/µl Murine RNase Inhibitor (NEB), and 0.2 U/µl Superase-In (ThermoFisher). This suspension was passed through a 40µm Cell strainer (Falcon™) to remove any aggregates. Nuclei quality and concentration was checked by staining 5µl of the suspension with SYBR Safe DNA stain (Invitrogen) and imaging on a Zeiss Axio Imager using a C-Chip hemocytometer (Neubauer improved).

2.3.3. Single nuclei capture and library preparation

Single nuclei capture and library preparation were performed using the Chromium Single Cell 3′ Gene Expression Kit (v3.1 chemistry) and the Chromium Controller (10x Genomics). Approximately 15,000 nuclei from the suspension were loaded per capture channel. Libraries SN046, SN062, SN063, SN064, nuclei were loaded as multi-species captures (15000 nuclei from each species) and demultiplexed later by mapping reads to respective genomes. cDNA amplification was performed with 15 cycles with a 3 minute extension time. The following steps were performed by Leslie Pan. Libraries were quantified using a Qubit fluorometer, and fragment size was checked on an Agilent Bioanalyzer. Libraries were sequenced with 250 million paired-end reads on an Illumina NextSeq 2000 using the following read configuration: 28 cycles for Read 1, 10 cycles each for the 15 and 17 indices, and 90 cycles for Read 2.

2.3.4. Single cell analysis

Raw sequencing data were mapped to version 1.0 of the *P.dumerilii* genome (Mutemi et al., 2025) using CellRanger v6.1.1 (v7.0.1 for library sn073) by Tobias Gerber. Individual Seurat objects were generated from the raw count matrices and basic filtering of extremely high or low RNA counts was performed. Libraries from multi-species captures were mapped to a pan- genome which comprised concatenated genomes of the species included - *Platynereis dumerilii* (Mutemi et al., 2025), *Pseudo - nitzschia multistriata* (Basu et al., 2017)*, Branchiostoma lanceolatum* (accession no.:GCA_900088365). Importantly, raw CellRanger count matrices were used without applying any prefiltering for multi-species runs for further processing. A mixed - species Seurat object was generated and a binarized presence-absence matrix across all pan genes was established. By counting the frequency of gene presence for each species, respectively, a species score was calculated. Each score was normalized by the number of total genes in the respective species. To balance out potential differences in species - specific gene detection rates, the obtained scores were normalised to the maximum value of each species - specific score across all nuclei generating values ranging from 0 to 1. Obvious species doublets were filtered out if the score was higher than 0.05 for all species in a nucleus. The top 10000 variable features for each species were calculated and used to cluster the mixed - species dataset. The normalised scores (calculated above) for each species were plotted on this object with DimPlot(), and clusters were subset into single-species datasets for downstream analysis by selecting clusters based on dominant species rate.

Doublets were removed from each library using Scrublet v0.2.3 and Python v3.9.18 on the EMBL Jupyter server (Wolock et al., 2019). Here, a Scrublet score was assigned by comparing each transcriptome of a given library to simulated doublets (summed counts of randomly selected pairs of transcriptomes) of that same library. A cell was considered a doublet if its score exceeded the Boxplot threshold set for that library - in other words, if a score was higher than the minimum outlier value calculated from the boxplot distribution of scores in

a given library, it was flagged as a doublet. The score and binary doublet (1) or singlet (0) information was saved in the libraries' metadata so that doublets could be removed in downstream Seurat analyses. Analyses were performed using R Software v4.3.2 on the EMBL RStudio server using the Seurat v5.0.1 (Hao et al., 2024), as well as some analyses - namely Niko Papadopoulos' custom cluster merging function - run in R v4.1.1 and Seurat v4.2.1 (Hao et al., 2021).

2.3.4.1. Quality Control and Seurat analysis - whole head dataset

Integration and QC. Non-integrated data first underwent preprocessing steps: normalisation, variable feature selection (top 10% of genes), scaling, and PCA analysis with 200 principal components. Upon inspection of Elbow and Scree plots, 80 principal components were used for downstream analysis. Clustering (FindClusters()) and UMAP embedding (RunUMAP()) were used to assess pre-integration dataset dimensionality. Libraries were integrated using Seurat's canonical correlation analysis (CCA) integration method for its ability to integrate despite substantial differences in gene expression (Hao et al., 2021), which is expected between premature and sexually mature states. For this integration, the IntegrateLayers() function with all genes and default parameters was used, mitigating large batch effects (**Figure 4B - C**). The integrated dataset was clustered as above using the Louvain algorithm and a resolution of 1, resulting in 30 clusters with varying predicted doublet contributions (**Figure 5A - B**). Barcodes flagged as doublets (described above) were removed. **Figure 5C** shows predicted doublet contribution to each cluster; doublet-dominated clusters (23 and 25) did not persist after filtering. Upon further inspection, two Louvain clusters (1 and 3 - mostly contributed to by libraries sn023 and sn073) seemed to be dominated by low features, low RNA counts, and mostly ribosomal reads (**Figure 6, 7**). This was mitigated by setting *nFeature* and *nCount* thresholds to a minimum of 500 and 1000 respectively, leaving a final dataset with 29542 nuclei. Clusters were coarsely annotated to the best of our knowledge, based on calculated marker genes using FindAllMarkers() with Seurat v4 parameters (logfc.threshold = 0.25, test.use = "wilcox", min.pct = 0.1).

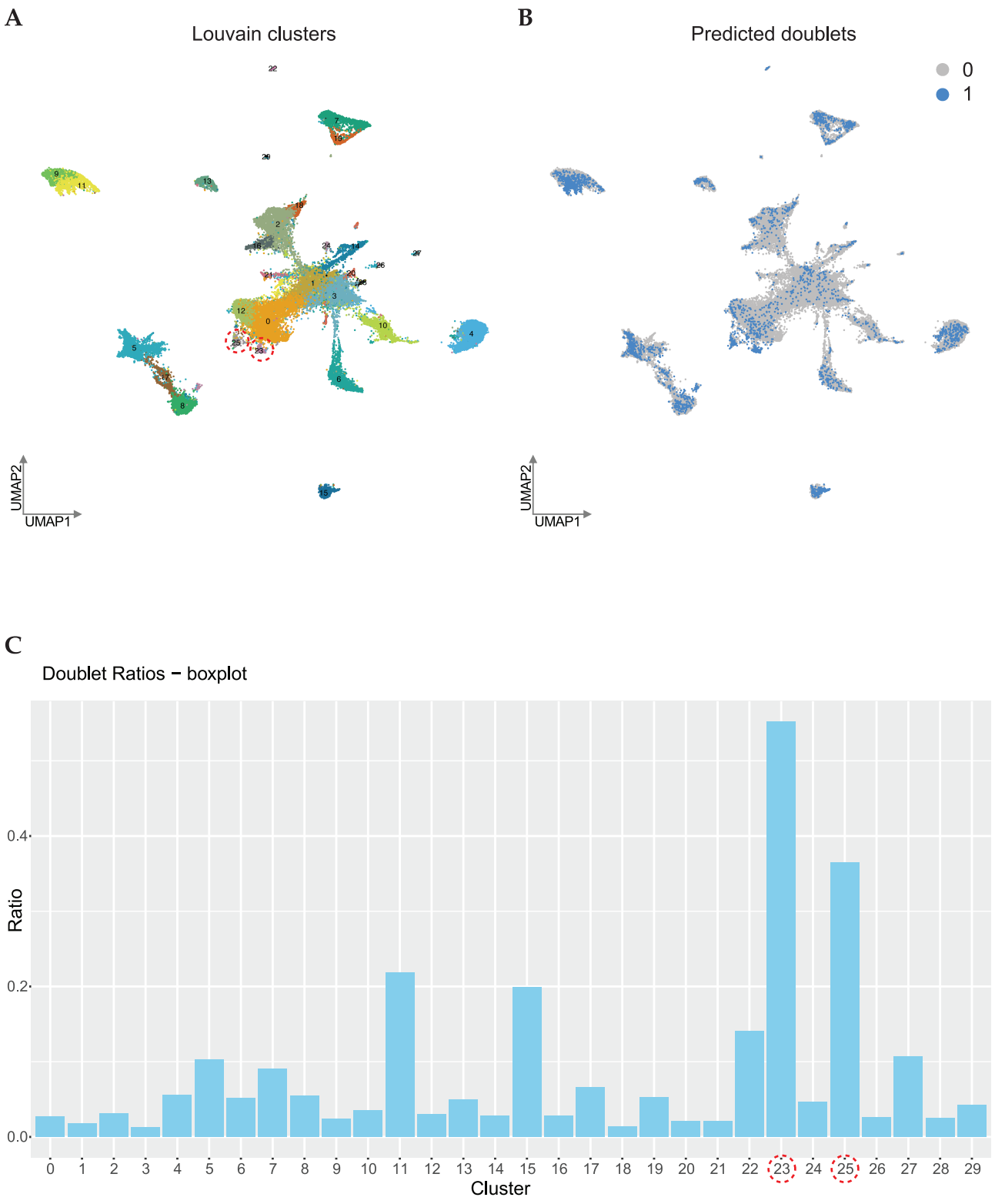


Figure 5. Contribution of predicted doublets. (**A**) Louvain clustering of integrated dataset produces 30 preliminary clusters. (**B**) Highlighted doublets - (1) in blue - based on the boxplot threshold for each library plotted on integrated UMAP. (**D**) Barplot showing ratio of predicted doublets across Louvain clusters.

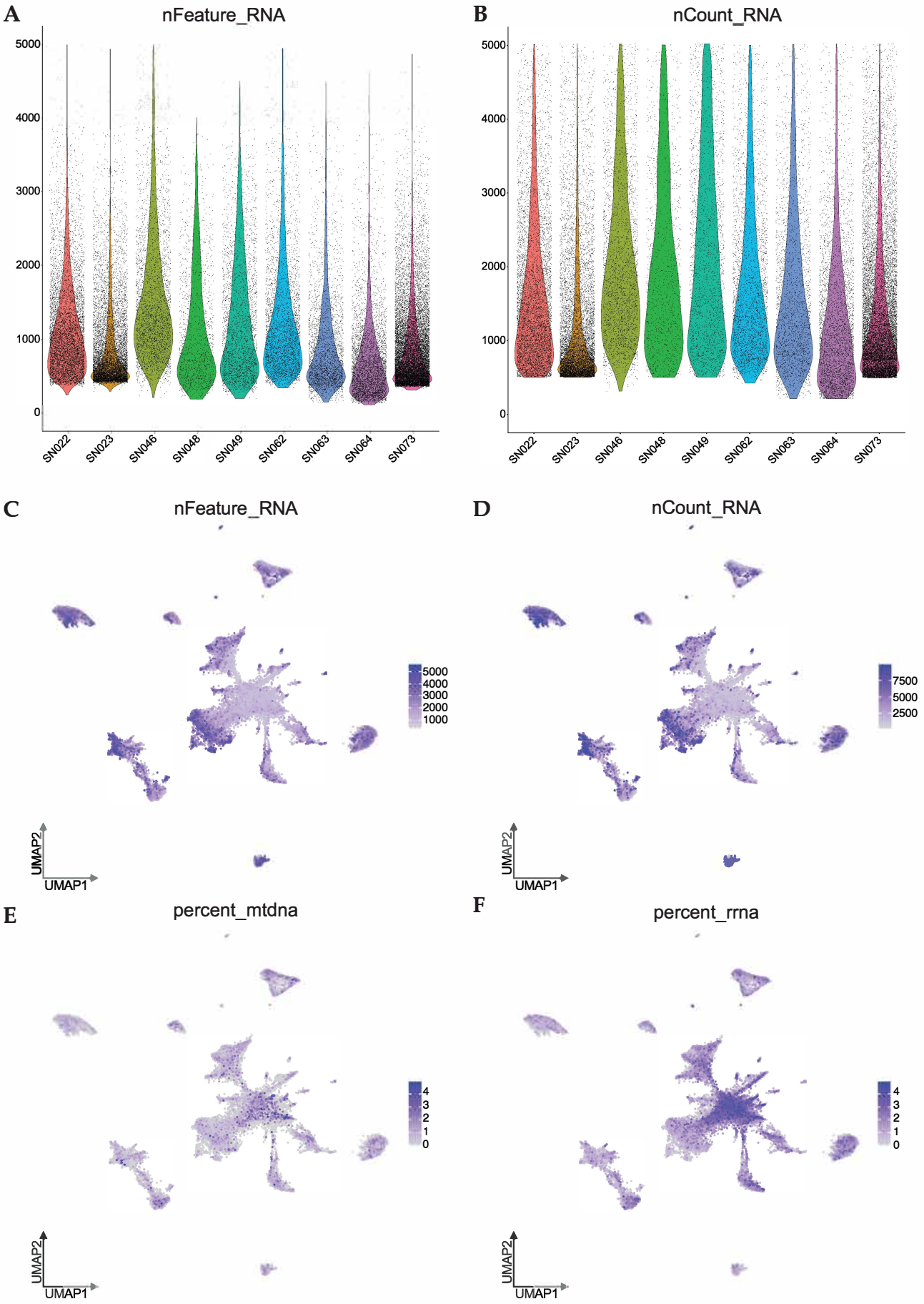


Figure 6. Additional quality control of snRNAseq dataset. Violin plots showing (**A**) number of features/genes and (**B**) RNA counts across libraries (premature: SN022, SN023, SN046, SN062, SN073, sexually mature: SN048, SN049, SN063, SN064). (**C**) Number of features/genes, (**D**) RNA count, (**E**) percentage of mitochondrial counts, and (**F**) percentage of ribosomal counts shown on integrated UMAP reduction from **Figure 4C**.

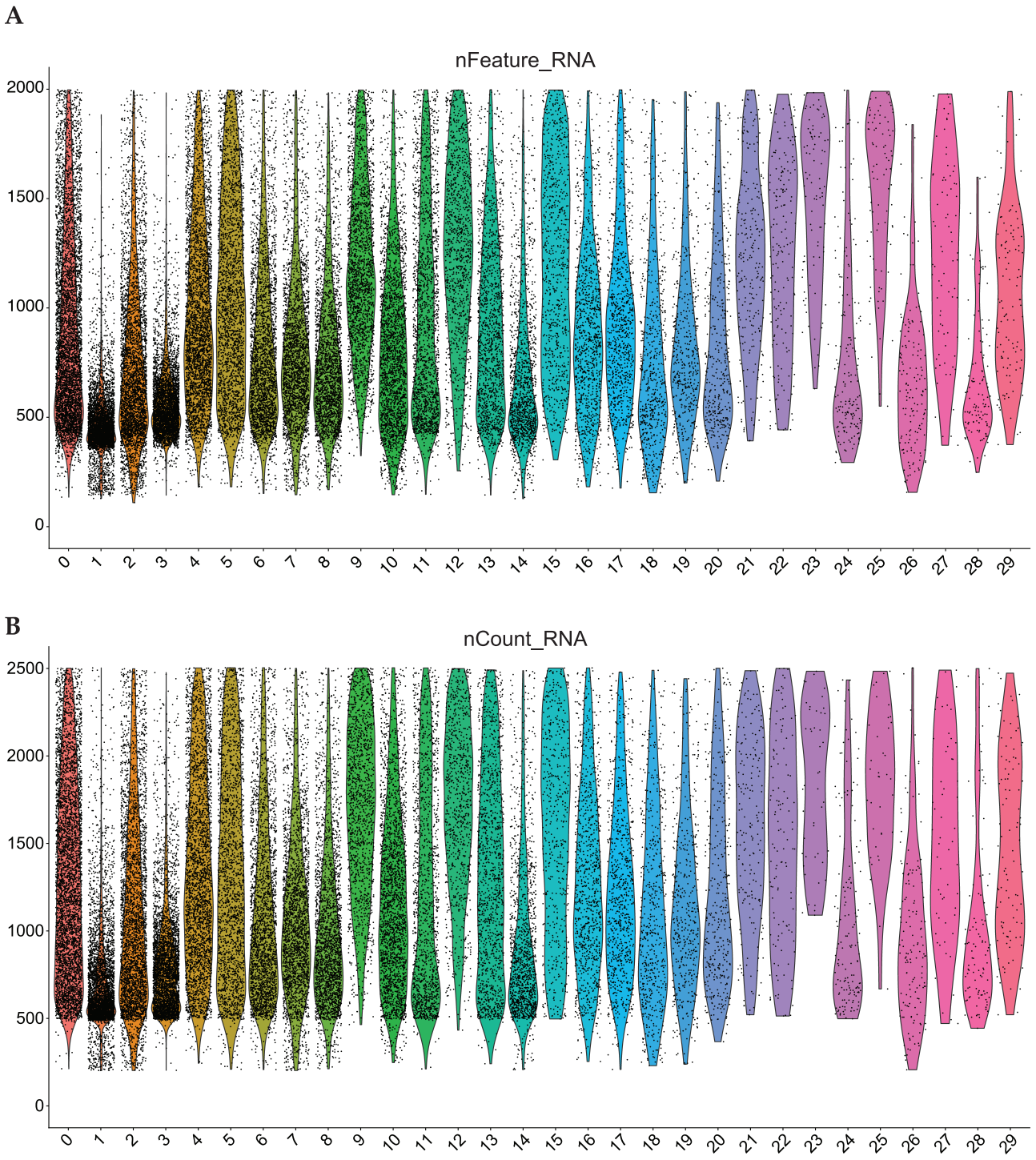


Figure 7. QC metrics across Louvain clusters. Violin plots showing (**A**) number of features/genes and (**B**) RNA count across Louvain clusters from **Figure 5A**. Note clusters 1 and 3 are dominated by nuclei with low number of features and RNA counts.

Neighbour-joining tree analysis. These clusters were grouped into cell type families by generating a neighbour-joining tree based on the average expression across all genes for each cluster. For this, the Seurat AverageExpression() function was run on the RNA assay of the object, followed by a Euclidean distance calculation between clusters. This distance matrix was used to build the neighbour-joining tree, using the nj() function, as well as the boot.phylo() function from the ape v5.5 package to assess branch support with 10000 bootstrap replicates. Bootstrap values were extracted and annotated with the following colour thresholds: black ≥80, orange 51–79, red ≤50.

Transfer annotations. The Seurat FindTransferAnchors() and TransferData() functions (Stuart et al., 2019), were used to transfer the cluster identities of a single *cell* atlas of the same species and tissue from Milivojev et al. (2024) to the present *nuclei* dataset. The reference published dataset was first filtered for the top 10,000 variable genes using the "vst" method, of which 9,974 were also present in my query nuclei dataset and used as anchoring features. Predicted identities were added to the query dataset metadata and cluster correspondence was compared using DimPlot() to overlay original and transferred labels on an integrated UMAP embedding of the single *nuclei* dataset.

2.3.4.2. Neuronal subset

Subclustering analysis was performed for the neuronal clade (9499 nuclei), selecting neuronal clusters based on marker genes and whether they formed a well - supported clade in the neighbour - joining tree. Subsetted data was split by library, and underwent normal pre-processing and re-integration as above (Section 2.3.4.1). In an attempt to achieve cell type resolution the neuronal subset was overclustered at a Louvain resolution of 4 and then iteratively merged, such that only clusters that exceeded 20 differentially expressed with a minimum average log fold change of 2, remained as individual clusters. This left 20 neuronal subclusters stemming from the original broader neuronal clades. Markers were calculated using the same parameters as the previous section.

2.3.4.3. Mushroom body subset

Mushroom body clusters from the single *nuclei* dataset as well as from the downloaded single *cell* datasets (multi - stage dataset, *cOpsin1* knockout premature dataset, EdU premature dataset) from Milivojev et al., (2024) were subsetted and integrated - across experiment and stage - using SCTransform v0.4.1 with glmGAMPOI v1.14.3 variance stabilization followed by Seurat CCA integration with dims set to 50, k.anchors set to 10, and k.weight set to 40 due to the lower number of cells/nuclei per condition. Standard log normalisation and scaling produced many library - specific clusters, hence the SCTranform method was tried as was done in Styfhals et al. (2022) who also integrated single *nucleus* and single *cell* data. Elbow and scree plots suggested 10-20 principal components captured much of the variance. Louvain clustering using 10 PCs at a resolution of 0.2 reproduced the three clusters found in the nuclei - only neuronal subset and was used for downstream analysis, although one library - specific cluster lacking specific markers was merged with a related larger cluster. Markers were calculated with the same parameters as in previous sections.

Differentially expressed genes between WT and *cOpsin1* knockout libraries downloaded from Milivojev et al. (2024) were performed using the Seurat FindMarkers() function for each of the three subclusters of this subset, as well as with the whole dataset, and with the remaining neurons in the dataset for comparison. Gene lists were filtered to retain the top 100 up and down regulated genes - based on average log fold change - in the *cOpsin1* knockout library compared to its matched premature control dataset and compared using the ggvenn() function from the ggvenn v0.1.10 package. Heatmaps, using the plot_heatmap() function from the Scillus v0.5.0 package, were generated for the top 40 up and downregulated DEGs for each subcluster, after filtering out unannotated and duplicated genes, such that shared DEGs appeared only once.

2.3.5. *In situ* validation

2.3.5.1. Whole - Mount Hybridization Chain Reaction (HCR)

This protocol is based on the HCR portion from Ćorić et al. (2023) and the v3.0 HCR protocol from Molecular Instruments (Choi et al., 2018). Published *P.dumerilii* coding sequences (e.g. *cOpsin1*, *arx*) or translated computationally-derived open-reading frames obtained from the current genome were submitted to Molecular Instruments for probe design. Using Geneious Prime ® 2024.0.5, probe sequences that produced BLAST (Altschul et al., 1990) hits in the genome that did *not* correspond to the gene of interest were excluded. Probe sequences are listed in **Table S1** in the Appendix.

Day 1 Fixation and storage. Worms of the desired size were anaesthetized in a 1:1 mixture of 7.5% (w/v) $MgCl_2$ and FASW. Heads were dissected just above the pharynx and transferred to 1.5 ml microcentrifuge tubes pre-filled with FASW on ice. FASW was then replaced with 4% Paraformaldehyde (PFA) (ProSciTech) in PTW (PBS + 0.1% Tween-20), and samples were fixed for 1 hour at room temperature under gentle shaking. Tissue was dehydrated through a series of increasing methanol (MeOH) concentrations in PTW for three minutes each (25% MeOH, 50% MeOH, 75% MeOH) followed by two 1-minute washes in 100% MeOH. Samples were stored at -20°C at least overnight.

Day 2 Tissue clearing and probe hybridization. All PTW washes were 5 minutes with 1ml PTW, unless noted as rinses in which 1ml PTW was added and immediately removed. On ice, samples were rehydrated in decreasing MeOH concentrations (75% MeOH, 50% MeOH, 25% MeOH) for 5 minutes each, followed by one PTW rinse and two PTW washes. Samples were treated with 100µg/ml proteinase K (PK) for 5 minutes at room temperature, then rinsed twice and washed twice with PTW on ice. Post-fixation was performed in 4% PFA in PTW on ice for 20 minutes, followed by one rinse and two washes in PTW. Tissue clearing was performed according to Pende et al., 2020. Briefly, tissue was immersed in 1ml of pre-chilled

acetone (Merck) for 30 minutes at -20°C, followed by one rinse and two washes in PTW. Tissue was then incubated at 37°C under gentle shaking in 1 - 2ml of clearing *Solution 1* (see Appendix for recipe) for 15-20 minutes or up to 60 minutes for sexually mature stages, then washed in PTW three times. Samples were incubated in 200µl pre-warmed Hybridisation Buffer (HYB) supplemented with Ribonucleoside vanadyl complexes (VRC) (Merck, R3380) at a concentration of 1:40 VRC:HYB for at least 30 minutes at 37°C. During this time the probe solution was prepared such that the final probe concentration (1µM stock) would be 4nM in the final 250µl volume of HYB/VRC solution. This solution was heated to 37°C, added to the samples (i.e. 50µl of probe - containing solution was added to the 200µl already in the tube) and incubated overnight at 37°C under gentle shaking.

Day 3 Amplification. Samples underwent four 15-minute washes at 37°C with 1ml *preheated* Probe Wash Buffer (Molecular Instruments). Next, at room temperature, samples were washed twice with 1ml 5x SSCT and equilibrated in 100µl Amplification Buffer (AMP) (Molecular Instruments) for 30 minutes. Meanwhile, hairpins - in individual tubes of a PCR strip - for the desired fluorophore combinations were heated to 95°C in a preheated thermocycler for 90 seconds and let cool to room temperature in the dark for 30 minutes. Enough of each hairpin (e.g. a two-gene combination requires 4 hairpins: B1h1-546nm, B1h2-546nm, B5h1-647nm, B5h2-647nm) was snap-cooled for a final concentration of 60nM per hairpin in a final amplification volume of 150µl. Cooled hairpins were added to the AMP to reach a final volume of 50µl. This mix was added to the 100µl of AMP already in the sample tube to reach the final 150µl volume per sample. Samples were incubated at RT overnight protected from light with under shaking.

Day 4 Amplification termination and mounting. At room temperature, samples were washed four times in 1ml 5x SSCT: twice for 5 minutes, once for an 1 hour and once more for 5 minutes. This was followed by a 25 minute incubation in 5µg/ml DAPI in 5x SSCT and two more 5 minute washes in 1ml 5x SSCT. As much of the 5x SSCT as possible was removed and

replaced, initially, with 5µl of SlowFade Glass mounting medium (Invitrogen, Refractive Index: 1.52) for 15 minutes to equilibrate. This was then topped up to around 50µl SlowFade and stored at 4°C short term or mounted on slides immediately. Heads were mounted using 0.2mm - 0.3mm sticky spacers (Sunjin lab) topped up with 12 - 21 µl of mounting medium.

2.3.5.2. Whole - Mount Immunofluorescence (IF)

This protocol is based on the DEEPclear tissue clearing combined with immunofluorescence protocol from Pende et al. (2020). Worm heads were dissected, fixed, dehydrated and rehydrated as in section 2.3.5.1. All washes were performed at room temperature in 1ml PTW for 5 minutes, except rinses, where 1ml of PTW was added and immediately removed. After rehydration, samples were washed three times in PTW. Each sample was then incubated in 1ml of pre-chilled acetone for 30 minutes at -20°C. After another three washes in PTW, samples were treated with a 75µg/ml PK solution for 6 minutes at room temperature. Samples were rinsed twice and then washed three times with PTW. Samples were incubated in 1ml clearing *Solution 1* for 15-30 minutes at 37°C in a preheated Thermomixer (Eppendorf) and then washed three times in PTW. Samples were blocked in 1% molecular grade bovine serum albumin (BSA) (3458509, Merck) in PTW for 1 hour at room temperature, then incubated in a 1:200 dilution of primary mouse Anti-acetylated tubulin antibody (T6793, Sigma) for 5 days at 4°C under gentle shaking. At room temperature, samples were rinsed once with PTW, followed by 4 longer PTW washes ranging between 30 minutes - 1 hour each. They were then incubated in goat anti-mouse - Alexa FluorTM 488 cross-adsorbed secondary antibody (A11001, Thermo) at a concentration of 1:200 for 5 days at 4°C under gentle shaking. Finally, samples were washed three times with PTW then stained with DAPI and mounted as in section 2.3.5.1.

2.3.5.3. Nuclear stains

I was lucky to get my hands on the same line of *cOpsin1* knockout worms used in Milivojev et al. (2024) from the Jekely lab at COS at the University of Heidelberg. Worm heads of different stages and conditions (WT premature n = 8, WT female n = 8, WT male n = 6, *cOpsin*KO premature n = 6, *cOpsin*KO female n = 6, *cOpsin*KO male n = 3) were dissected, fixed, and cleared as described in section 2.3.5.1. The PK digestion and post-fixation steps were omitted. After the final PTW wash, samples were incubated in NucSpot650 (Biotium, 41034) or DAPI for at least 25 minutes and mounted as described previously.

2.3.5.4. Cryosectioning & Hematoxylin/Eosin stain

Worm heads were dissected and fixed in 4% PFA in PTW at room temperature as in section 2.3.5.1, but for 30 minutes. After three 5 minute washes in 1x pPBS, heads were dehydrated in 20% sucrose in pPBS overnight at 4°C. Samples were then incubated for at least 1 hour in a 1:2 mixture of TissueTek® O.C.T Compound (OCT):20% sucrose, followed by a 30 minute incubation in 1:1 OCT:20% sucrose. Just before embedding the heads in cryomolds (TissueTek®), the samples were moved to 100% OCT. Once oriented in the OCT, the mold was pressed against dry ice until the OCT became opaque. Samples were stored at -20°C, or -80°C for longer term storage. Tissue was sectioned horizontally at 10µm on a Leica Cryostat, at a chamber temperature of -23 °C and an object temperature of -25 °C.
Slides with the cryosections were thawed overnight at room temperature. Remaining steps occurred at room temperature by Idoia's students Dori and Maite. Slides were washed in 1x PBS for around 3 hours to remove excess OCT, followed by two 5 minute washes in dH20. Slides were incubated in Mayer's Hematoxylin stain (Thermo) for 5 minutes and washed under running tap water for 10 minutes. This was followed by a 3 minute incubation in Eosin (1:4 dilution in 96% ethanol, with 0.5% glacial acetic acid). Slides were rinsed quickly, then washed twice for 5 minutes in 96% ethanol. This was followed by two 5 minute washes in 100% ethanol and two 5 minute washes in Xylene. Slides were left to dry for one minute before mounting in EuKitt (Sigma-Aldrich) and visualised on a Zeiss Axio Imager.

2.4.6. Confocal Imaging and Processing

Images were captured with a Leica Stellaris 8 confocal microscope with a white light laser from the Advanced Light Microscopy Facility (ALMF), or a Leica SP8 CSU confocal microscope for imaging of nuclear stains, using LAS X software (Leica). Images were processed using Fiji/ImageJ software v2.14.0 (Schindelin et al., 2012). Image adjustments were made per channel using Adobe Photoshop 2024 (25.12.0 release), such that the same adjustment curves were applied to the same channel across images being compared within a figure. Single focal planes are presented for MB images across depth, for negative controls, and when a detailed view of MB nuclei for stage comparisons is needed. Other images are maximum intensity z - projections for better visibility of regional expression.

2.4. Results

2.4.1. snRNAseq captures cell type diversity in the *P.dumerilii* brain

A total of 8 adult *P.dumerilii* heads were dissected and trimmed of sensory head appendages. Of these, 4 of premature stage - worms around 2mm with food still present in the gut -, 2 spawning male and 2 spawning female heads, were dissected and the extracted nuclei were run in 4 separate experiments with each tissue sample its own 10x reaction (1 premature tissue sample was captured in two separate 10x reactions as technical replicates - sn022/sn23) (**Figure 4A**). A first look at the single nuclei transcriptomic dataset after basic filtering (section 2.3.4.1) showed separation by reproductive stage (sexually mature vs premature), and by experiment - especially library sn073 (**Figure 4B**). This was mitigated upon integration with the Seurat CCA method (**Figure 4C**).

Following additional filtering and doublet removal (section 2.3.4.1), 29542 nuclei were left, giving rise to 31 clusters (**Figure 8A**) that were broadly annotated by myself and Phil Oel based on differentially expressed marker genes using the FindAllMarkers() function. By employing the Seurat TransferAnchors() function (Stuart et al., 2019), with my single-*nucleus* dataset as a query dataset, these broad annotations were also consistent with those reported in an independent single-*cell* dataset of the same species and stage (Milivojev et al., 2024). **Figure 8B** shows how the predicted cluster annotations, although broader, correspond with what was originally annotated in the present nuclei dataset. For some annotations - including the mushroom body neuron cluster - this served as a validation, for other cell types, whose gene-signature I was less-familiar with, this offered an alternative annotation I did not suspect (e.g. non-astrocyte glia cluster). These annotations were also supported by an unbiased neighbour-joining tree analysis to show the relationships between the different clusters based on their overall gene expression. Here, the neurons form a well supported clade, reproduced in over 80% of the bootstrap replicates (**Figure 9**).

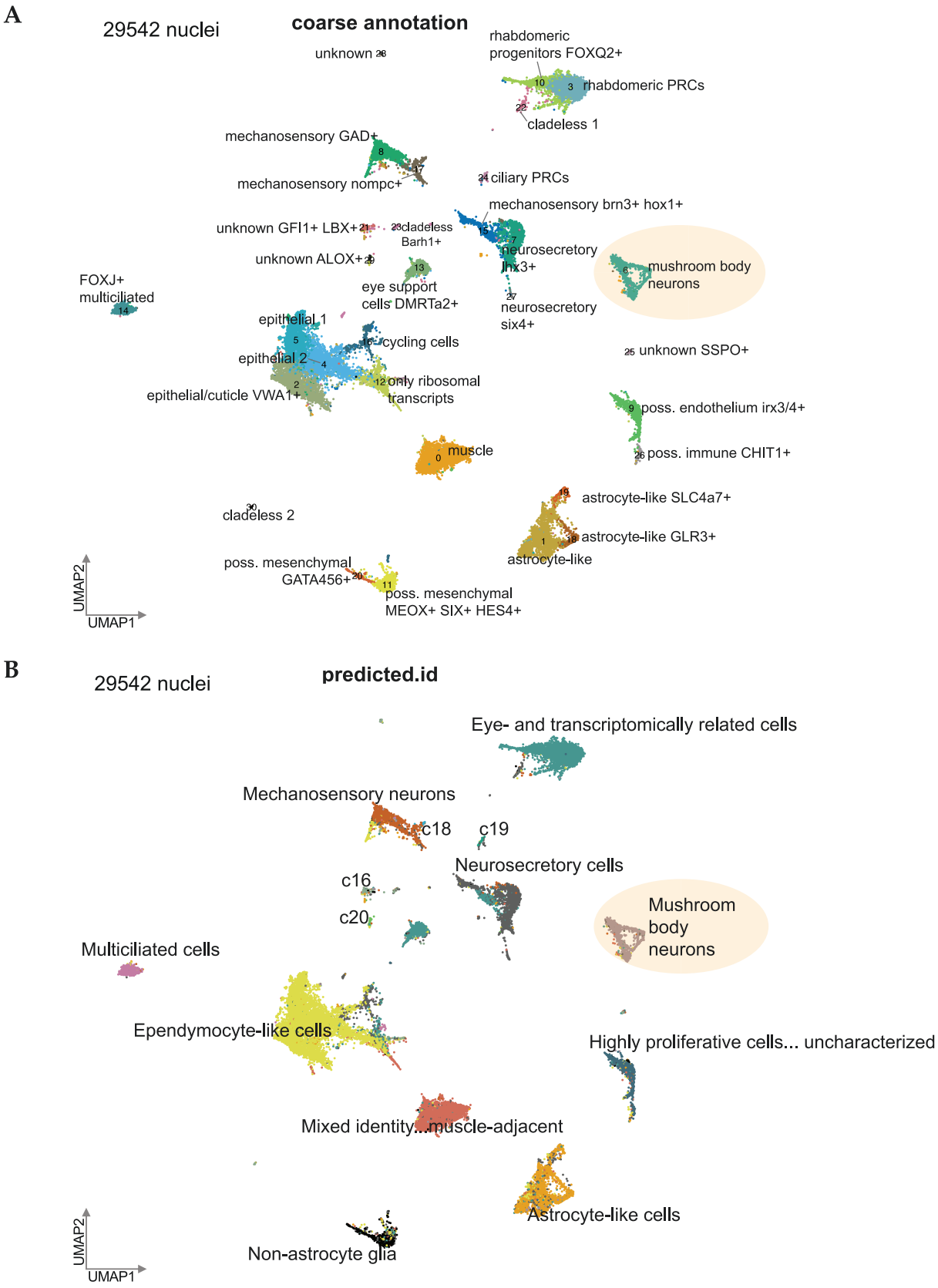


Figure 8. Post filtering UMAP annotated based on marker genes and by TransferAnchors() function. (**A**) Manual annotations after inspecting marker genes. (**B**) Predicted annotations transferred from preprint dataset (Milivojev et al., 2024). Note broad correspondence of annotations, including for mushroom body neurons, highlighted.

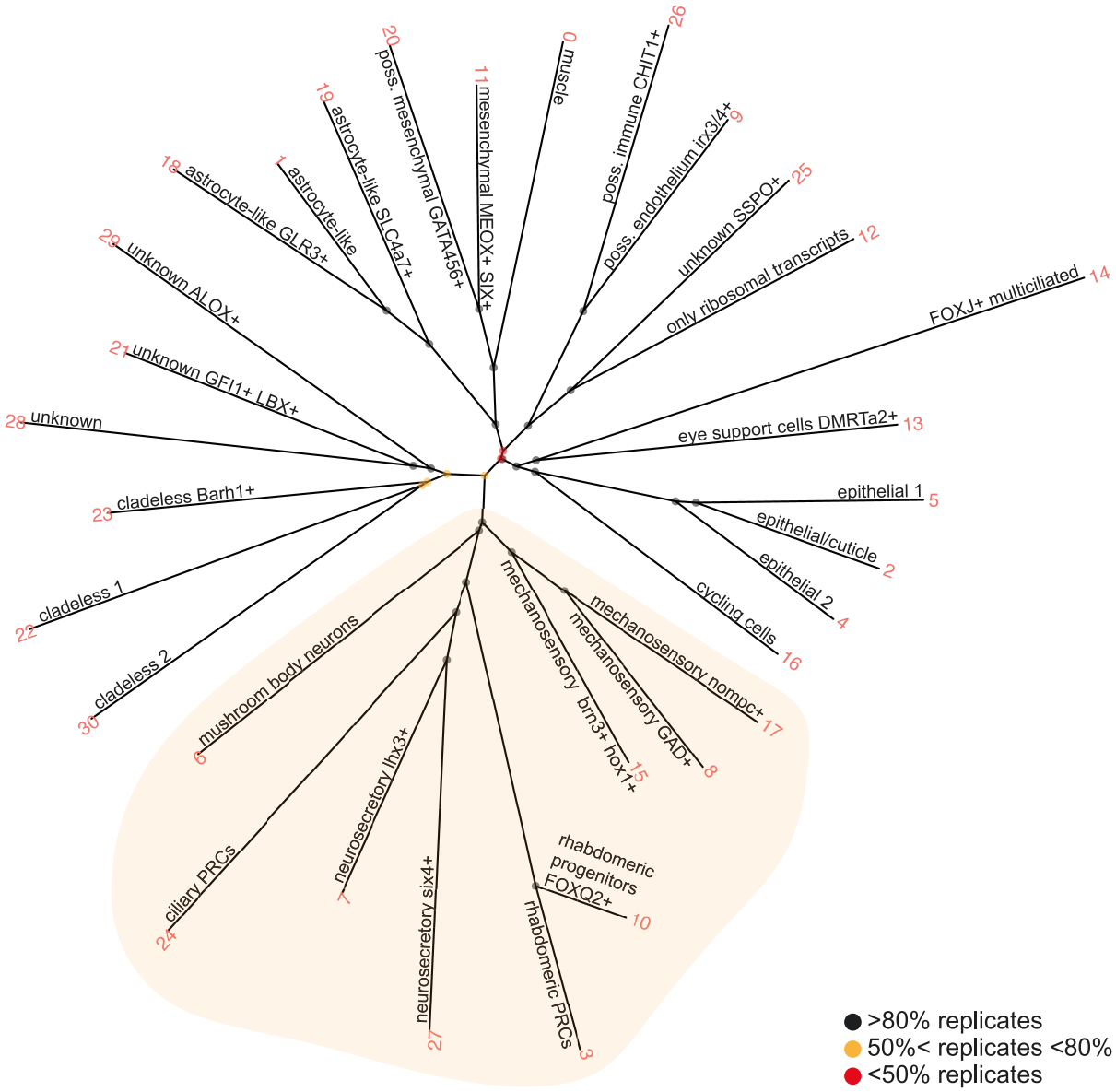


Figure 9. Unrooted neighbour-joining tree. Based on all clusters and all genes with 10k bootstrap replicates. Node colouring according to percentage of replicates supporting a given clade. Neuronal clade is highlighted, note >80% bootstrap values for all nodes of this clade.

2.4.2. A distinct transcriptomic signature for MB neurons

Figure 10 provides anatomical orientation of the MBs, helpful for the upcoming *in situ* validations. The transcriptomic analysis from the previous section revealed a clear MB cluster of neurons in the dataset. This gene signature included the expression of candidate marker genes *arx* (accession no.: GU169412.1) and *ptf1* (**Figure 11A**). These are primarily cholinergic neurons, expressing ChaT (choline acetyltransferase, accession no.: EF384218.1), while vGLUT+ (vesicular glutamate transporter,) more specifically marks other mechanosensory (MS) and neurosecretory (NS) clusters.

Beyond known markers, this cluster showed a specific expression of *pdfr* (receptor for pigment dispersing factor, accession no.: OL606759.1) and of newly identified markers *emx2* (empty spiracles homeobox) and the neuronal calcium sensor *ncald* (neurocalcin delta) thus far annotated by eggNOG-mapper (emapper) (Cantalapiedra et al., 2021). Note that gene names for published *P.dumerilii* genes will have associated accession numbers. Remaining gene names come from emapper - annotated protein sequences from the new genome and are denoted by their XLOC - location on the genome - code (Mutemi et al., 2025). Any paralog names, therefore, do not necessarily reflect the actual paralog use in *P.dumerilii.* For the purposes of this thesis, gene names in figures will contain the gene *XLOC* code, its emapper annotation/description if available, and the *P.dumerilii* gene name if known (e.g. *XLOC* | emapper | pdu name). A summary, **Table S2,** of genes mentioned in text is in the Appendix. In text, genes will be named by their published *P.dumerilii* name - if available - or the emapper annotation if not.

To confirm *in situ* whether this gene signature localises to the MB region of the adult annelid brain, HCR stainings of two of these markers - *arx, ptf1* - were performed on premature stages of *P.dumerilii.* This revealed that, indeed, the co-expression of these genes - thus far only shown in early developmental stages - localised to the dense nuclei of the MBs (**Figure**

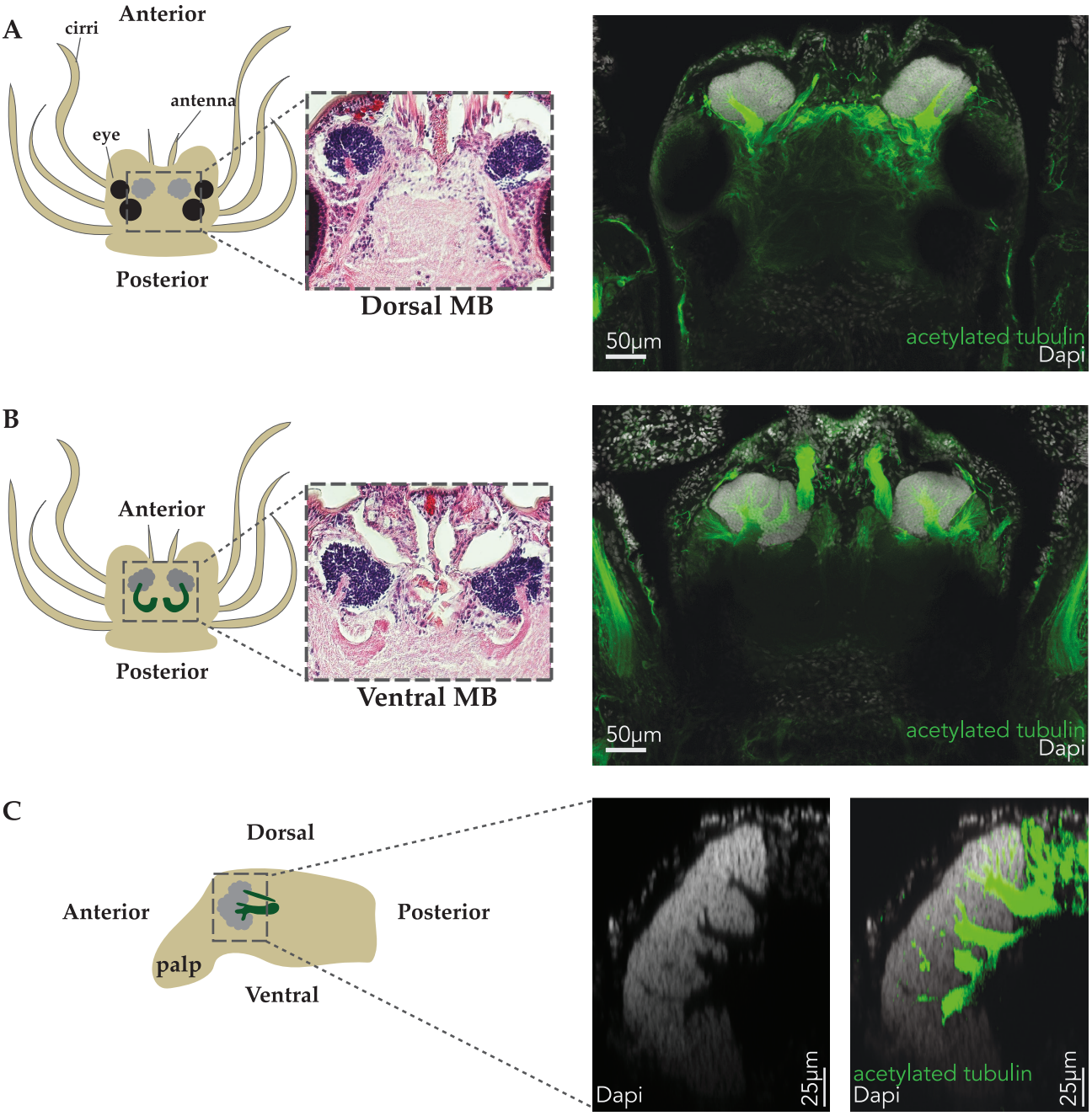


Figure 10. Premature *P.dumerilii* MB anatomy. Schematic, H&E stained cryosections, and whole-mount immunofluorescence images of (**A**) the dorsal region of the MB, immunofluorescence showing a 6µm z - projection, (**B**) the ventral region of the MB, immunofluorescence showing a 15µm z - projection, to capture the ventral MB branching. (**C**) Orthogonal projected image (134 slices) showing sagittal view through one MB. Z - step size is 1.5µm. Immunofluorescence images acquired with a confocal 20x oil objective.

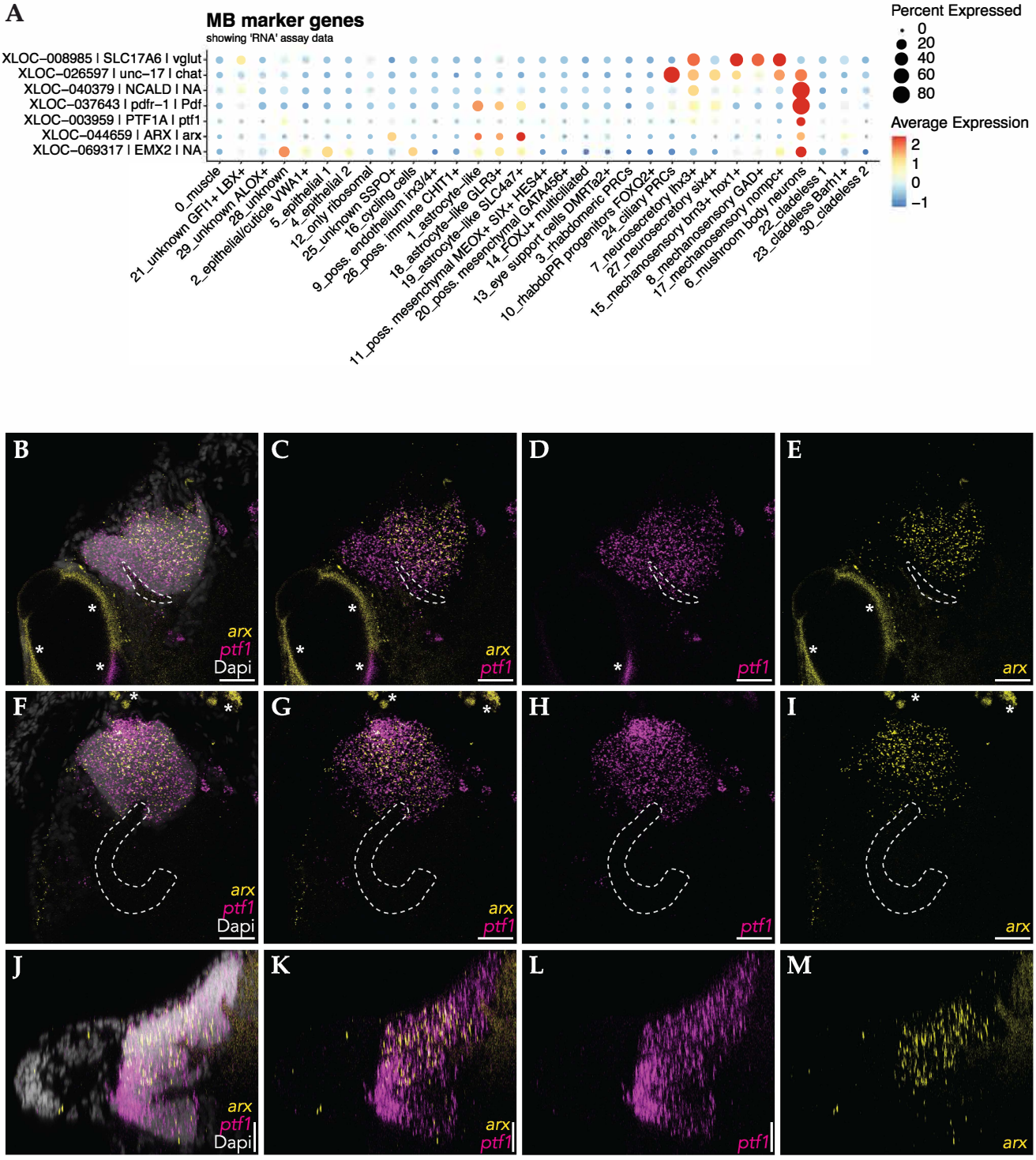


Figure 11. MB gene signature and whole-mount HCR validation showing *arx* and *ptf1* co-expression in the MB. (**A**) Dotplot showing MB marker genes. (**B - E**) Representative dorsal MB z -slice (**F - I**) Representative ventral MB z -slice. Top of MB images is anterior direction, left is lateral. (**J - M**) Orthogonal projected image showing sagittal view through one MB. Step size is 1μm, 168 z - slices included in the projected image. Top of the image panels is dorsal, the left is anterior. All images acquired with a confocal 20x oil objective. All scale bars 25μm. Dotted lines schematise dorsal and ventral MB peduncles. Asterisks mark regions of autofluorescence.

11B - M). The *ptf1* expression being broader than the *arx* expression, is something to keep in mind for later sections. Indeed, *arx* expression seems sparse or absent in the nuclei around the dorsal MB peduncle (**Figure 11B - E**). An older triple - staining on 10µm cryosections of a premature head show that *ncald* also localises well to this region (**Figure S1A - E**). The poor *ptf1* staining in this case is likely due to it being acquired in the very autofluorescent 488nm channel. Further representative HCR controls are also shown in **Figure S1F - Q**. Here, the general neuron marker, *elav* (accession no.: EF384209.1), is compared to an MB specific gene, *ptf1*(**Figure S1F - I**), a non-MB gene, *brn3* (accession no.: KC109636.1), shows no expression in the MB compared to the MB gene, *mef2c* (**Figure S1J - M**), and amplifier controls (relevant fluorophore-linked hairpin sequences mixed with mis-matching probes) revealed suspiciously overlapping signals in the tightly bundled axons of the MB peduncles that are likely stuck hairpins (**Figure S1N - Q**).

All things considered, the markers used to define the MB cluster in the snRNAseq dataset are also specific to this region *in situ.* While this chapter will now dive deeper into the subpopulations of cells within this cluster, this signature becomes important for broader evolutionary comparisons discussed in Chapter 3.

2.4.3. Neuronal subset reveals three *cOpsin1* expression domains

To explore subtler differences and cell populations that are not immediately apparent in the full atlas, clusters from the neuronal clade were subsetted, re-integrated and re-clustered at a high Louvain resolution. This over-clustered dataset was merged such that any clusters that did not have a minimum of 20 genes with an of average log fold change of at least 2 were iteratively merged, leaving 20 neuronal subclusters (**Figure 12A**).

Unexpectedly, three regions of *cOpsin1* (accession no.: AY692353.1) - a vertebrate - type ciliary opsin (Arendt et al., 2004) - expression were apparent (**Figure 12B**). First, in a small cluster corresponding to the ciliary photoreceptors of the brain. Despite the low *cOpsin1*

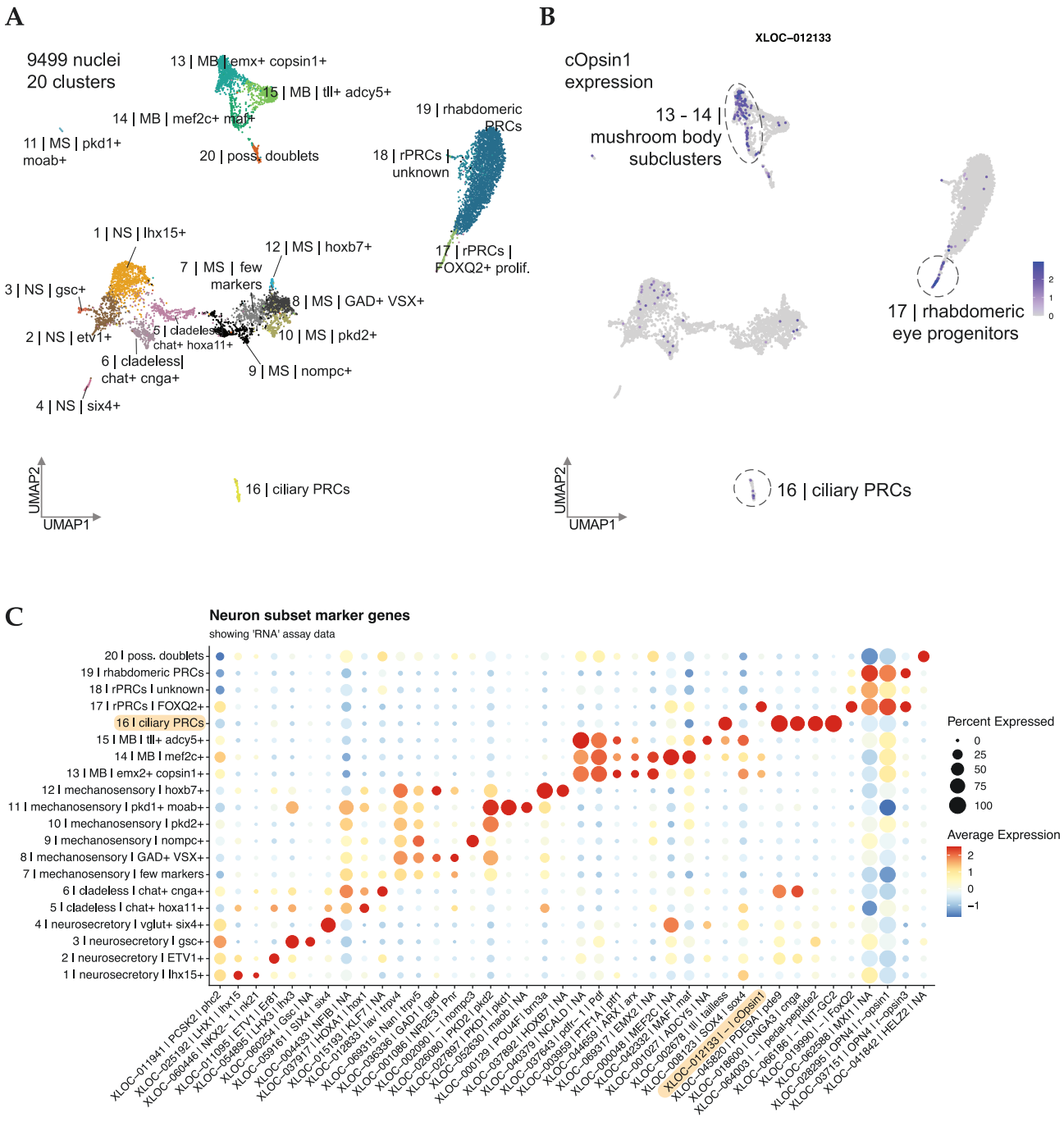


Figure 12. Neuronal subset and three regions of *cOpsin1* expression. (**A**) 20 subclusters left after subsetting, re-integrating and re-clustering neuronal clusters from the neuronal clade highlighted in **Figure 8**. (**B**) Feature plot of *cOpsin1* expression with three regions of expression circled. (**C**) Dotplot of known marker genes for broad neuron categories, including select transcription factors. Note *emx2* and *cOpsin1* expression in two of the MB subclusters. Abbreviations: MB: mushroom body, MS: mechanosensory, NS: neurosecretory, PRC: photoreceptor.

expression here, the specific expression of *pedal-peptide 2* (accession no.: KF515945.1), *pde9* (Achim et al., 2018), *cnga* (accession no.: KM199644.1), as well as *NIT-containing guanylate cyclase 2* (Jokura et al., 2023), provided a clear ciliary photoreceptor signature. Published sequences were blasted against the genome (Mutemi et al., 2025) to find the best XLOC match for visualising on the single nucleus dataset. This gene signature can be seen in **Figure 12C**, which plots known marker genes for the neuronal cell types, as well as specific transcription factors from the marker gene list calculated with FindAllMarkers(). For these reasons, despite the scarce *cOpsin1* expression, this cluster was annotated as the ciliary PRCs, and thus represents the first region of expected *cOpsin1* expression. Further *cOpsin1* expression was apparent in a subcluster of the rhabdomeric photoreceptors in line with the recently reported progenitor cells responsible for adult eye growth during sexual metamorphosis (Milivojev et al., 2024). Lastly, *cOpsin1* was surprisingly expressed in two MB subclusters that also specifically expressed *emx2* (**Figure 12B - C**).

This *emx2* gene, could be a possible paralog of the *emx* (accession no.: GU169418.1) published for *P.dumerilii,* since its protein sequence is grouped with vertebrate *emx* sequences in a phylogenetic tree analysis (**Figure 13**) using the EMBL Job Dispatcher MUSCLE alignment and Simple Phylogeny tools (Madeira et al., 2024), and visualised with phylo.io (Robinson et al., 2016). Furthermore, BLASTn analysis of the open-reading frames of *emx2* and of the published *P.dumerilii emx* aligned to separate regions of the new- yet to be annotated - HiFi assembled *P.dumerilii* genome (Mutemi et al., 2025), as well as different sequences of a *P.dumerilii* long read Iso-Seq transcriptome in the lab, that is incomplete for technical reasons, but able to provide additional annotations and insights for RNA sequencing experiments as was seen in Shields et al. (2021). Therefore, from a computational stand point, it appears that these are two separate *emx* paralogs.

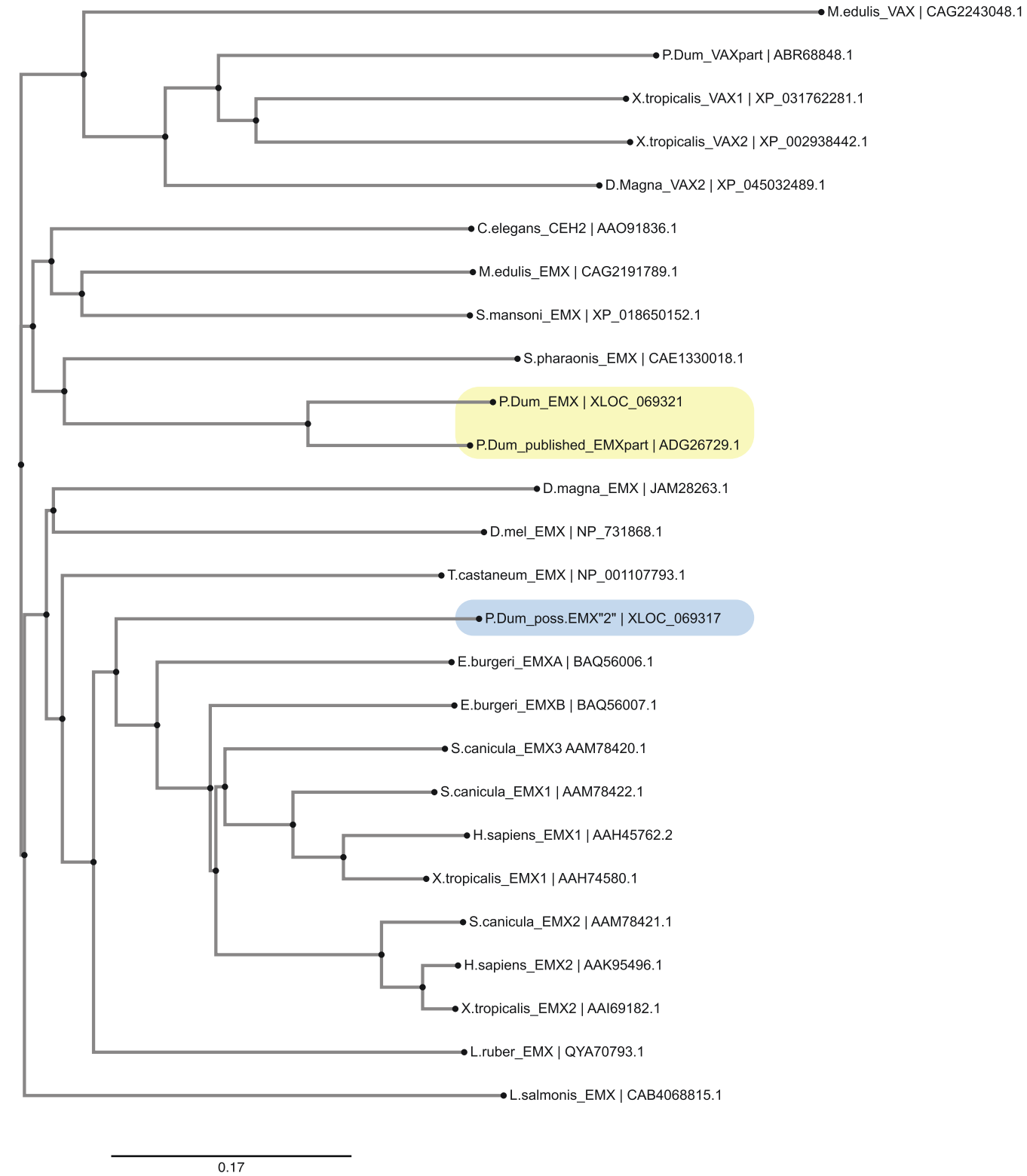


Figure 13. Phylogenetic tree of EMX protein sequences with VAX as outgroup. *P.dumerilii* sequences and associated gene location codes are highlighted. Yellow highlights the published *P.dumerilii* sequence and best blast hit from the new genome. Blue highlights the additional *P.dumerilii* sequence found in the genome and annotated as EMX2 by emapper. Tree visualised with Phylo.io (Robinson et al., 2016). Accession numbers listed for published sequences.

Next, I wanted to check if probes designed against this *emx2* gene were indeed localised to the MB region, or part of it. **Figure 14** shows that this is indeed the case. When compared to the known MB marker *arx, emx2* expression seems localised to the ventral regions of the MB. Interestingly, despite being a broad but sparse marker in the single nucleus dataset, *arx* expression again seems to be sparse or absent in the nuclei surrounding the dorsal peduncle (**Figure 14 A - D**), similar to what was seen in **Figure 11**.

Further validating this subpopulation, HCR co-staining of *emx2* and *cOpsin1* in premature worms also showed co-expression in the ventral region of the mushroom body (**Figure 15A - H**). This *cOpsin1* expression, however, seems tied to segment number, such that only worms over 60 segments expressed *cOpsin1* the ventral MB (**Figure 15I - L**), despite detectable *cOpsin1* expression in the presumed dorsal brain cPRCs (**Figure 15M - P**)**.** Interestingly, although in the snRNAseq dataset all premature worms (>55 segments) showed *cOpsin1* transcripts in the MB subcluster, HCR experiments showed that only 60% of worms over 60 segments showed *cOpsin1* expression in the MB, and in some cases, worms below this segment threshold also showed expression in the ventral MB (**Table 1**). Examples of these exceptional cases are shown in **Figure S2.** Once again the absence of *cOpsin1* in the MB is unlikely to be technical considering its expression in other parts of the head; in this case the growing rhabdomeric eye is depicted (**Figure S2I - J**). There was no correlation with the artificial moon phase and all worms were fixed at the same time of day. Speculation on this nuanced expression will be discussed in the Summary section. To better understand these MB subpopulations I wanted to integrate my data with published data that included an even earlier developmental stage, as well as a *cOpsin1* knockout premature stage.

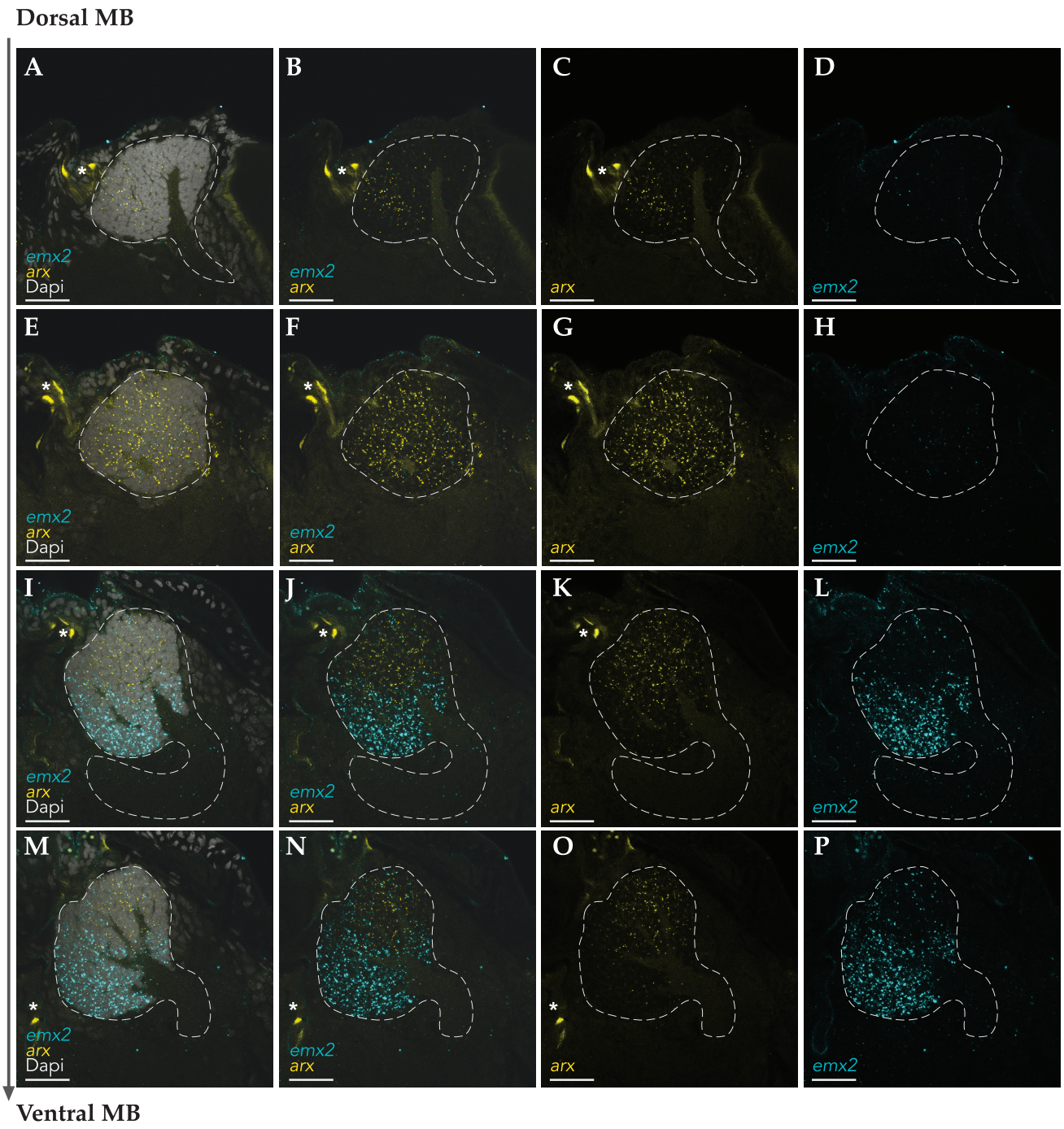


Figure 14. Patterns of co-expression of *emx2* and *arx* in the premature *P.dumerilii* MB from dorsal to ventral MB regions. (**A - D**) z33, (**E - H**) z52, (**I - L**) z80, (**M - P**) z95. Z step-size 1μm. Top of images is anterior and right is lateral. All scale bars 25μm. Images acquired with a confocal 40x oil objective. Autofluorescence marked by asterisks.

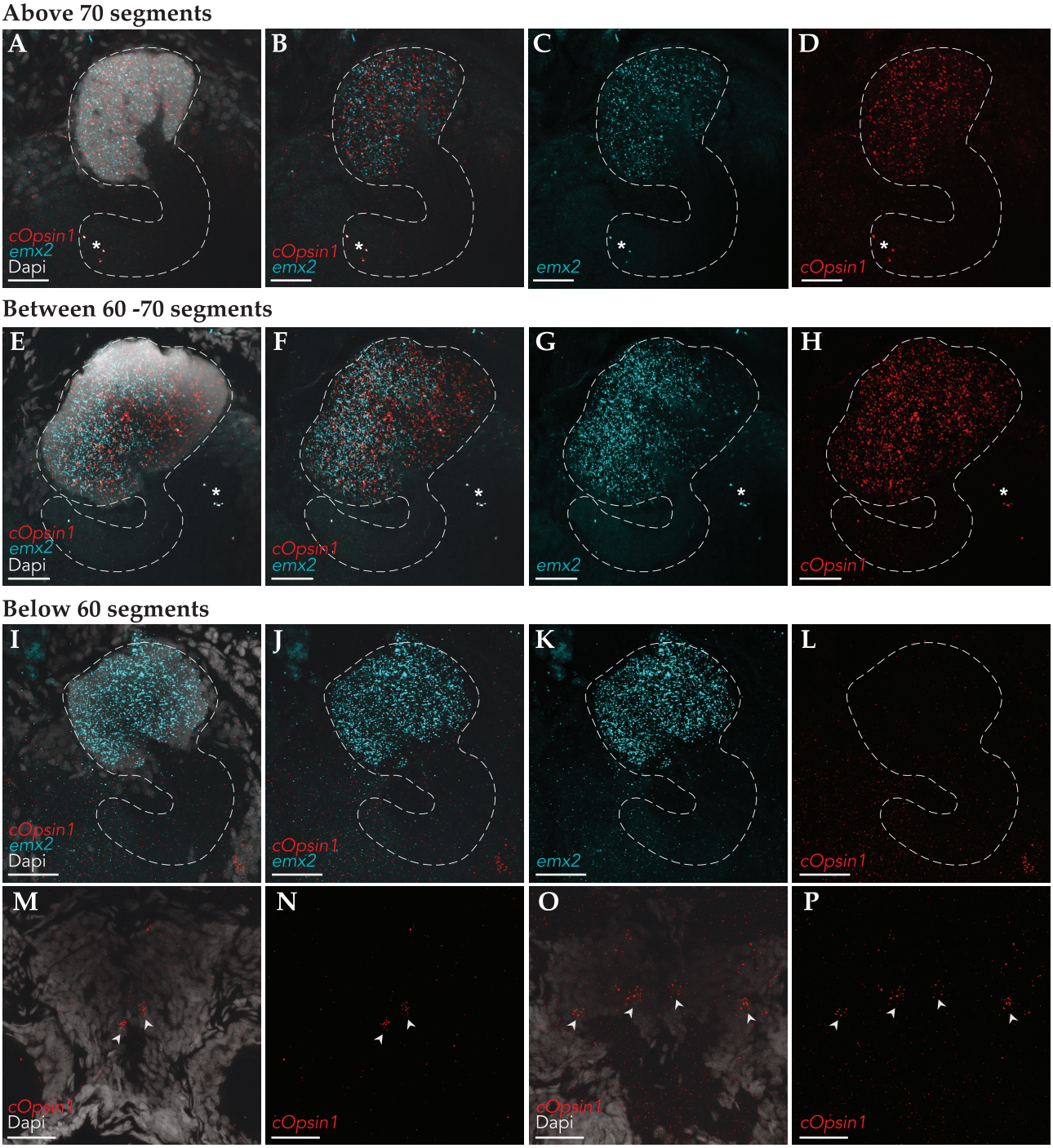


Figure 15. Co-expression of *emx2* and *cOpsin1* in the ventral MB of premature worms. (**A - D**) Ventral MB of a 73 segment worm showing co-expression of both genes, 5μm z-projection (1μm step - size). (**E - H**) Ventral MB of a 65 segment worm showing co-expression of both genes, 5μm z-projections (1μm step - size). (**I - L**) Ventral MB of a 51 segment worm showing expression of *emx2* but not *cOpsin1* in the ventral MB, 7.5μm z-projections (1.5μm step - size). Top of images is anterior and right is lateral in all MB images. (**M - P**) Same 51 segment worm as in previous panel showing that *cOpsin1* is expressed in the dorsal brain cPRCs, indicated by white arrow heads. (**M - N**) A 3μm z - projection (1.5μm step - size) of the dorsal brain. (**O - P**) A 6μm z - projection (1.5μm step - size), of the dorsal brain moving further ventrally. Bottom of dorsal brain images is anterior. All scale bars 25μm. Images acquired with a confocal 40x oil objective. Asterisks indicate suspiciously overlapping signals, likely stuck hairpins (**see Figure S1**).

Table 1. Summary of *cOpsin1* expression in the MB detected with WM-HCR according to segment number. Not all experiments were co-stainings of this gene combination, hence some cases of missing data for *emx2.*

Worm size	***Emx2* detected in MB**	***cOpsin1* detected in MB**
> 60 segments	8/8	4/7
51 - 60 segments	3/3	2/6
41 - 50 segments	-	0/1
30 - 40 segments	-	0/4

2.4.4. Gene signatures of the MB subpopulations

Towards this end, I subsetted and combined the MB clusters from my *nuclei* dataset with the subsetted MB cluster of the single *cell* datasets published in Milivojev et al. (2024), producing a combined MB subset. With a low Louvain cluster resolution, the three overarching subclusters also seen in the nuclei neuronal subset emerged: an *emx2*/*cOpsin1*+ subcluster (ventral MB), *mef2*/*maf*+ subcluster (possible transitional cell type), and an *emx2*- but *tll*+ (dorsal MB) subcluster (**Figure 16A - B**). Beyond this dichotomy (presence and absence of *emx2* expression), the ventral MB subcluster showed specific expression of serotonin receptors (*htr4, 7*), a possible neuropeptide Y receptor, an adrenergic receptor *adraa2b*, as well as collagens, a possible LIM-containing homeobox transcription factor, and voltage-gated ion channels like *hcn1*, a hyperpolarization activated cation channel (**Figure 16B**). The dorsal MB subcluster, defined additionally by the *tailless* (accession no.: GU169423.1) transcription factor and an adenylate cyclase, *adcy5,* showed a different repertoire of receptors including a possible neurokinin receptor - *tacr2* - , a potential histamine receptor - *hrh1* - , a potential octopamine receptor (*OctR*), and a dopamine receptor (*DRD2*).

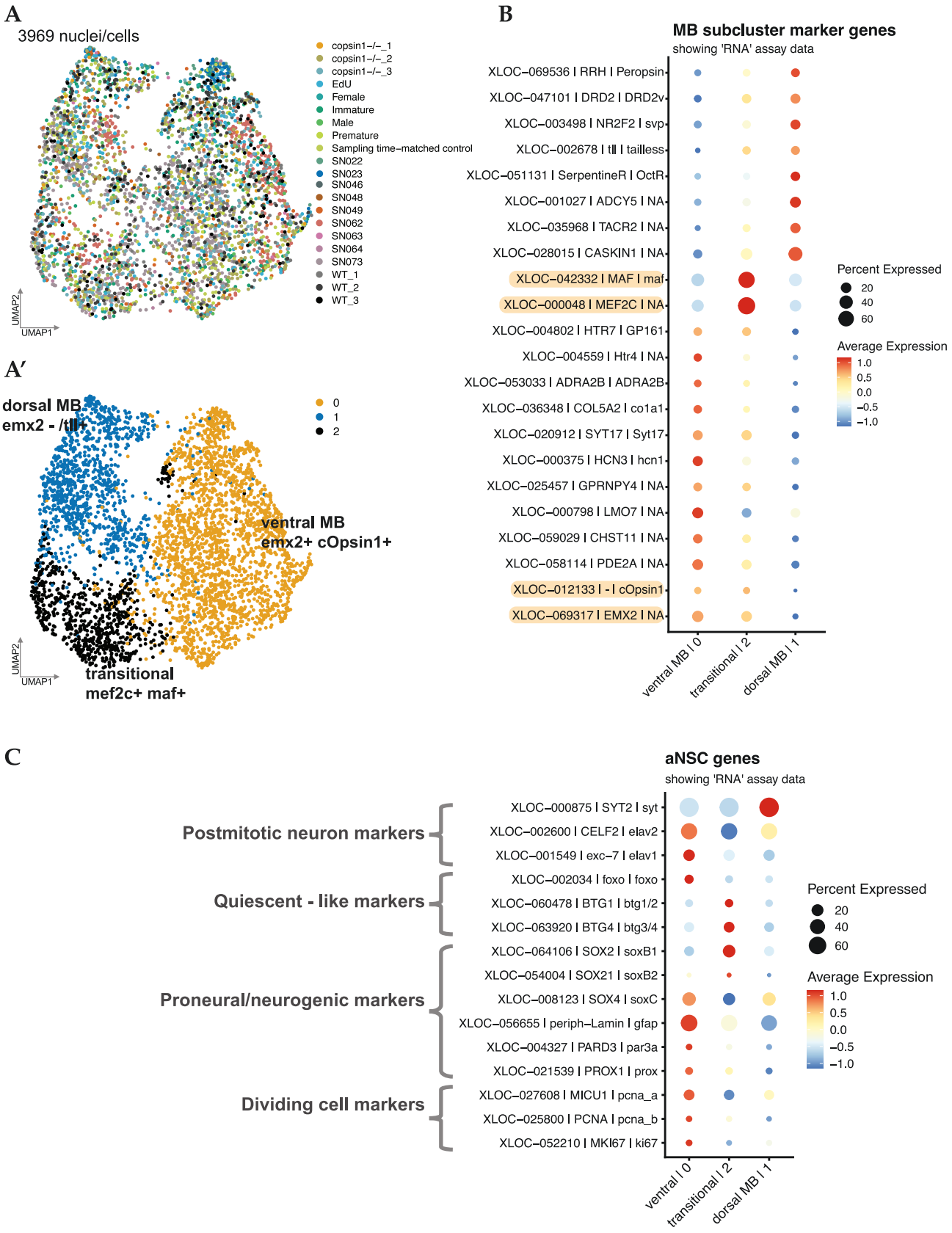


Figure 16. Combined single *nucleus* and single *cell* MB subset. (**A**) Integrated MB datasets, including premature and reproductive stages from the single nucleus dataset, as well as several single cell datasets from (Milivojev et al., 2024): immature, premature, reproductive dataset, premature - EdU+ dataset and time matched controls, as well as a premature *cOpsin1* knockout dataset and WT controls. (**A′**) Clustering of integrated MB dataset, with 10 PCs and Louvain resolution of 0.2 (**B**) Dotplot of selected marker genes for each subcluster. Highlighted markers of note: *emx2*, *cOpsin1*, *mef2c*, *maf*. (**C**) Expression of aNSC genes reported in Milivojev et al. (2024) across MB subclusters.

2.4.5. Ventral MB expresses proneural genes in premature worms

Interestingly, the third subcluster, while sharing many marker genes with the other subclusters, notably had *mef2c* and *maf* transcription factors among their top 20 calculated marker genes (**Figure 16B**). This was suggestive of a transitional cell population, especially considering that the brain of premature stages of *P.dumerilii* shows abundant EdU labeling and a gene signature reminiscent of aNSCs (adult neural stem cells) compared to reproductive stages (Milivojev et al., 2024). This proliferative - to - quiescent transition is somewhat reflected in the MB subclusters, with the ventral *emx2*+ MB cluster showing a proliferative and neurogenic signature, the likely transitional *mef2c/maf*+ cluster beginning to express some quiescence genes, and the dorsal MB cluster expressing postmitotic genes, although, quiescence and postmitotic markers are also apparent in the ventral MB cluster (**Figure16C**). Despite forming a clear cluster, this transitional *mef2c*+ population does not seem localised to a specific region of the MB *in situ* (**Figure 17A - F**). Although, *mef2c* is a broader marker considering both its broader expression in the snRNAseq neuronal subset (**Figure 12C**), and its expression in the dorsal brain (**Figure 17G - H**).

2.4.6. Effect of *cOpsin1* on the ventral MB

Leveraging the *cOpsin1* knockout (KO) subset (Milivojev et al., 2024) contained within this combined subset, I investigated what impact *cOpsin1* might have on these MB neurons. Top differentially expressed genes (DEGs) were ordered based on average log fold change since *p-values* can be deceptively small when using the FindMarkers() default method (Junttila et al., 2022). To see if DEGs, for the *cOpsin1* KO versus the WT, were specific to the MB clusters or a reflection of the general effect of the knockout, the top 100 DEGs for the entire dataset, as well as for the combined non-MB neurons, were compared to those of the MB subset (**Figure 18**). Indeed some of the genes upregulated in the MB of the knockout condition were also upregulated in the entire dataset including a collagen (*col4a2*), *mrlc* (non - muscle myosin (*mrlc*) implicated in cell division (Karess et al., 1991), the microtubule - associated gene *Janus*

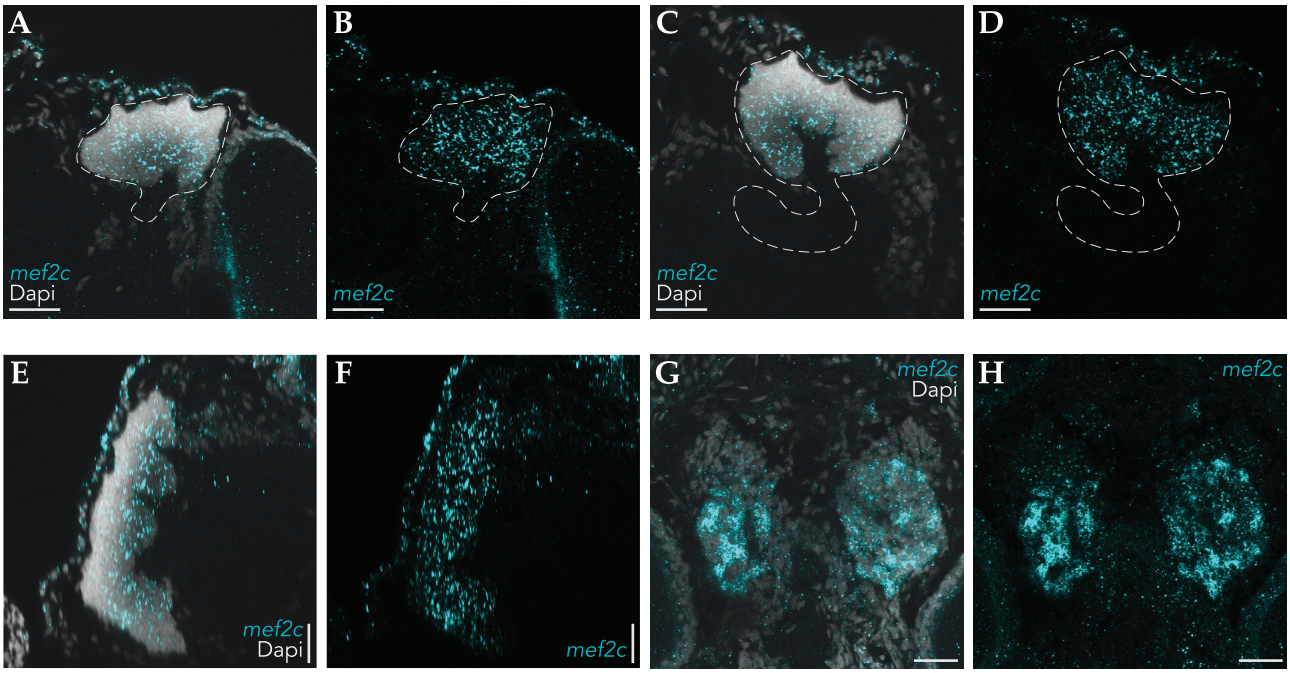


Figure 17. Expression of *mef2c* across MB and dorsal brain. (**A - B**) Dorsal MB, z47, and (**C - D**) ventral MB, z88. Top of images is anterior and right is lateral. (**E - F**) Orthogonal projected image (136 slices) showing sagittal view through one MB. Top of images is dorsal, left is anterior. (**G - H**) Expression of *mef2c* in the dorsal brain, bottom of images is anterior. All scale bars 25μm. Images acquired with a 40x oil objective with a z - step size of 1.5μm.

Kinase And Microtubule Interacting Protein 1/2/3 (*jkip3*), and even an annotated *creb - binding protein* (*CBP*) which may be acting as a coactivator for many developmental processes (Goodman & Smolik, 2000). The MB genes upregulated and shared only with other neurons were mostly genes encoding ribosomal proteins (**Figure 18B**). Considering the aNSC- like signature reported for premature versus reproductive worms (Milivojev et al., 2024), it is possible that the *cOpsin1* knockout worms are delayed in development and therefore show an overall tendency to express genes necessary for protein synthesis and cell division, characteristic of a more premature stage. This broad shift towards a neurogenic state is also reflected in the MB downregulated genes. In the knockout condition, the MB subclusters all shared the downregulation of transcription factors suggestive of neuronal differentiation, such as the *soxC* gene reported by Milivojec et al. (2024) and known for its role in neuronal differentiation in early *Platynereis* development (Kerner et al., 2009), as well as emapper-annotated genes like *mxl1* (**Figure 19, Figure 21A**). The latter is also expressed in rhabdomeric eye progenitors (**Figure 12C**) and is known for its role in switching from proliferative states to differentiation (Okada & Shi, 2018; Quéva et al., 1998; Yuan et al., 1998).

The individual subclusters showed specific downregulation of some of their marker genes, or emapper - annotated genes of the same type (e.g. transitional cluster - a *Maf* transcription factor annotated as *fosl2*) (**Figure 19**). Interestingly, the ventral MB *emx2+* subcluster showed downregulation of a possible cAMP responsive element binding protein - *crebzf* - reported to trigger both apoptosis or differentiation (Bodnarchuk et al., 2012; Valderrama et al., 2009), and the transitional *mef2c/maf* population showed a specific downregulation of a possible CCAAT enhancer binding protein - CEBPB (**Figure 19**), important for survival of newborn neurons in the adult hippocampus (Cortes-Canteli et al., 2011; Herold et al., 2011; Hodge & Hevner, 2011). On the other hand, all three subpopulations showed an upregulation of *gfap,* a proneural marker reported by Milivojev et al. (2024), further suggesting an overall maintained neurogenic state in the knockouts (**Figure 20, Figure 21B**). Interestingly, the ventral MB *emx2+* population showed specific upregulation of *arx,* and the transitional *mef2c/maf* cluster

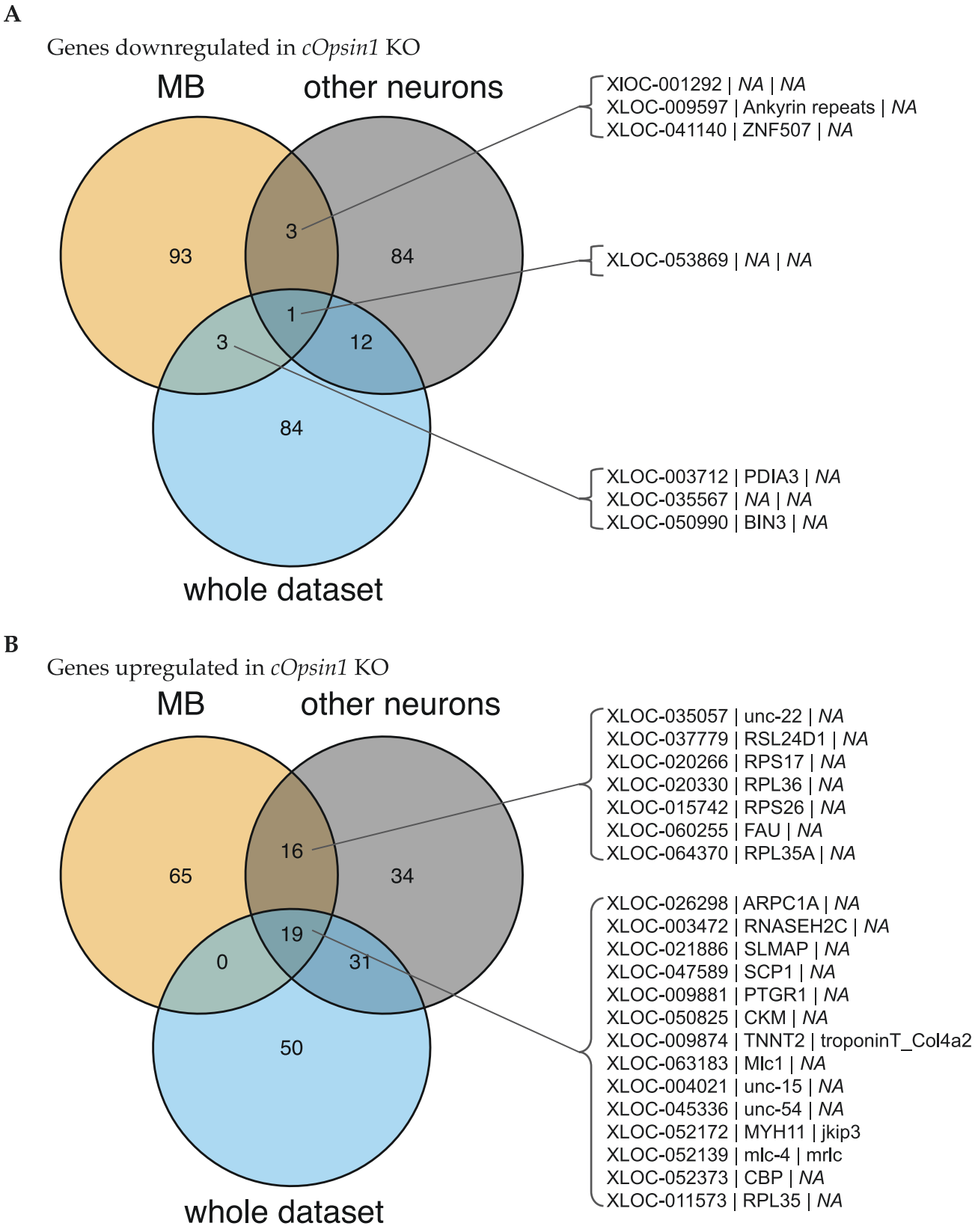


Figure 18. Shared and unique genes among top 100 DEGs between premature *cOpsin1* knockouts and WT worms across the entire dataset, non-MB neuron clusters, and MB neuron cluster. (**A**) Shared genes downregulated in the knockout. (**B**) Shared genes upregulated in the knockout dataset. Annotated genes shared with the MB cluster are listed.

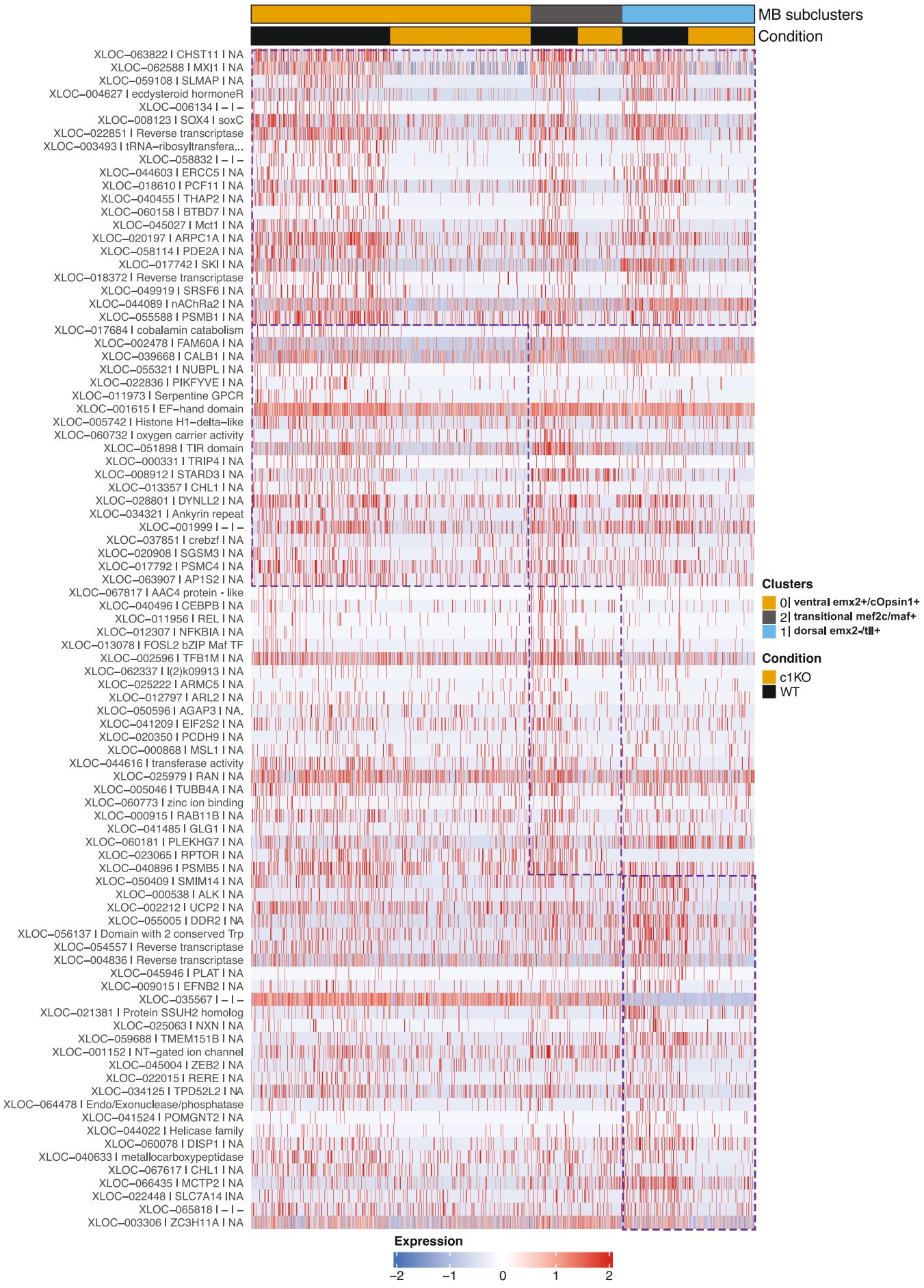


Figure 19. Heatmap of top annotated downregulated genes in the MB subclusters of the *cOpsin1* knockout dataset compared to premature WT controls.

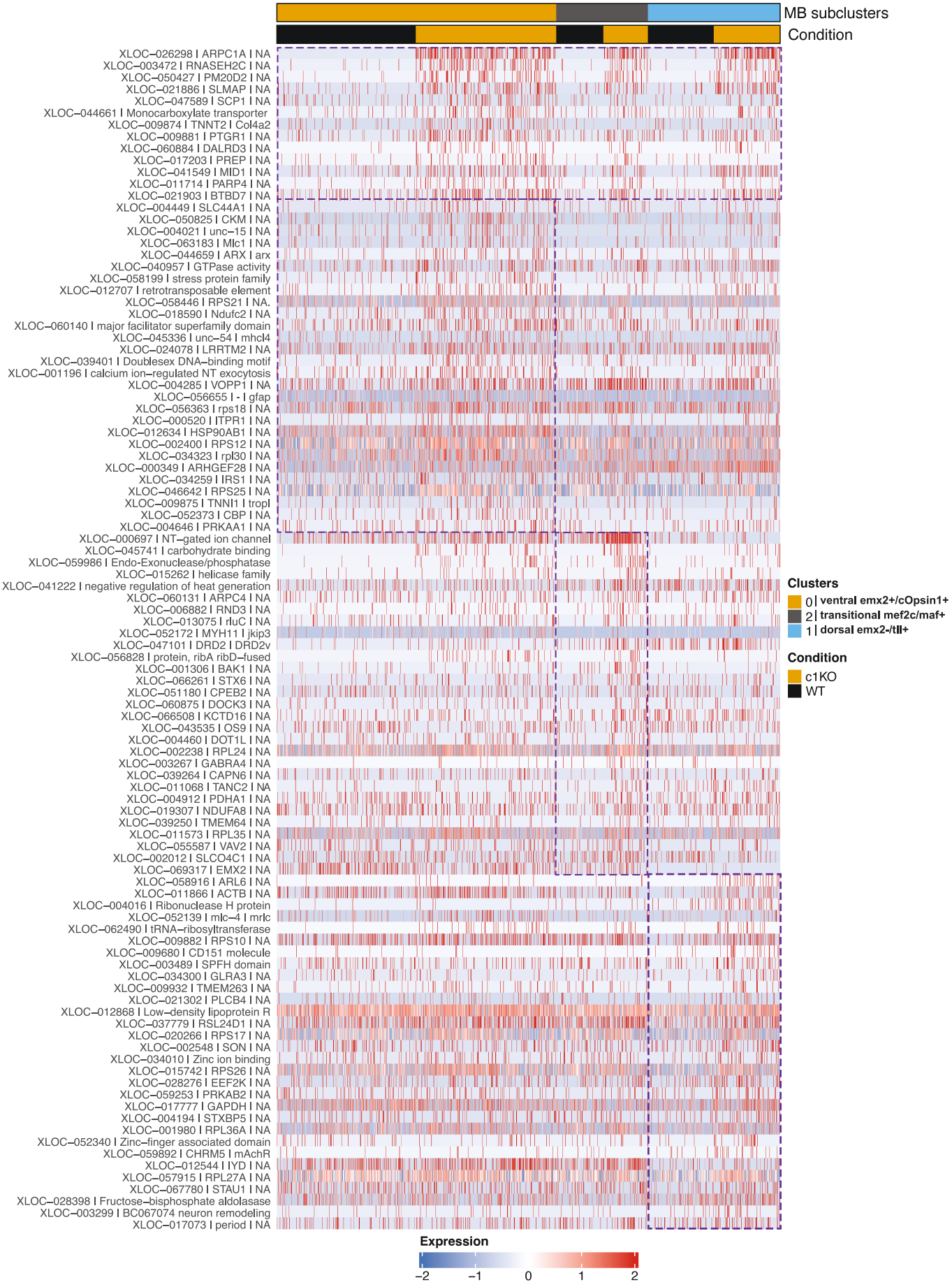


Figure 20. Heatmap of top annotated upregulated genes in the MB subclusters of the cOpsin1 knockout dataset compared to premature WT controls.

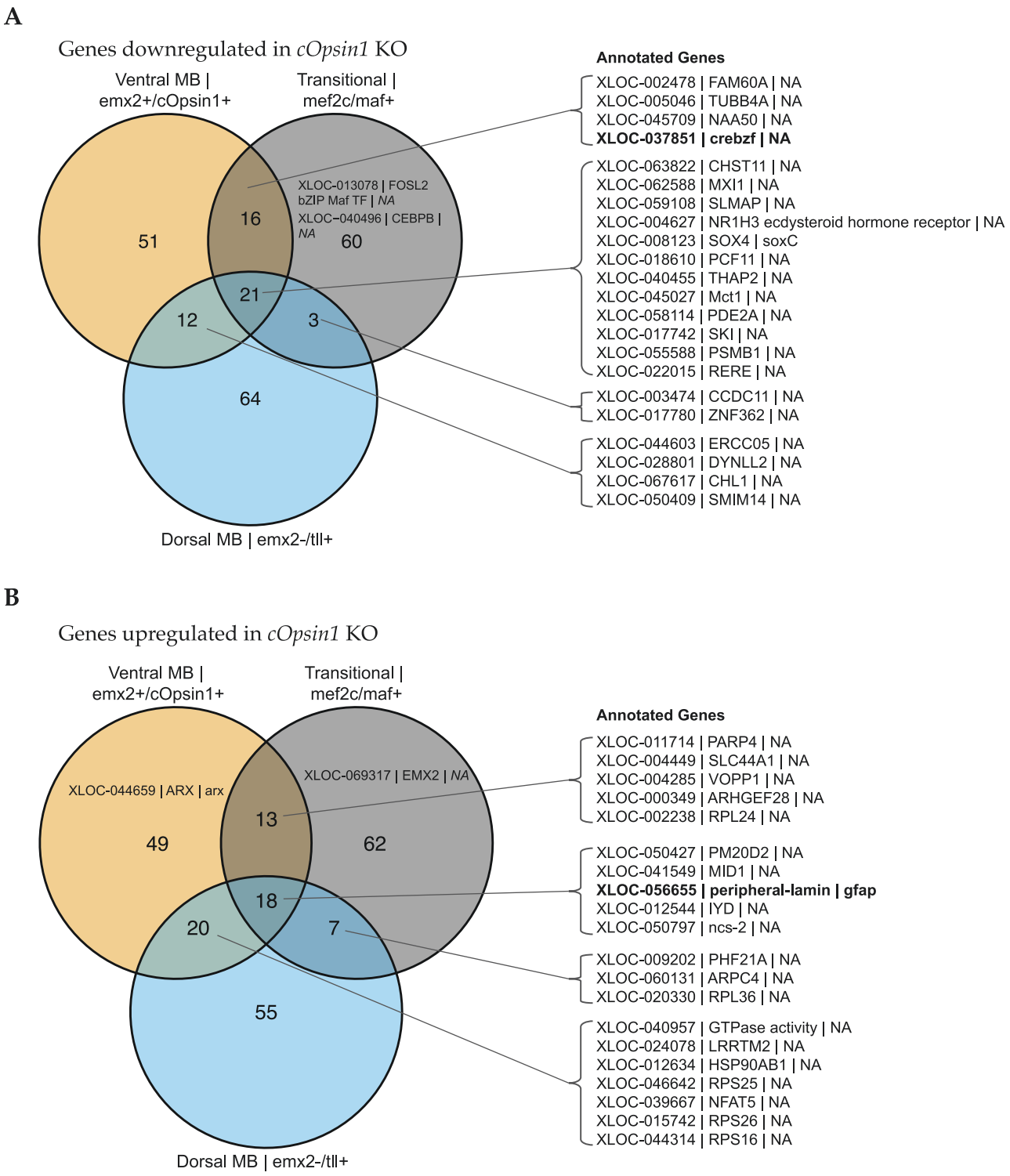


Figure 21. Shared and unique genes among top 100 DEGs between MB subclusters of premature *cOpsin1* knockouts and WT controls. Annotated genes listed for sets of shared DEGs. Noteworthy unique DEGs also listed (**A**) Venn diagram showing shared and unique genes among MB subclusters downregulated in the *cOpsin1* knockout. (**B**) Venn diagram showing shared and unique genes among MB subclusters upregulated in the *cOpsin1* knockout.

showed specific upregulation of *emx2* (**Figure 20**), suggesting that these ventral MB marker genes may be involved in maintaining a neurogenic state - possibly explaining why *arx* expression in the MB was less uniform - even absent around the dorsal peduncle (**Figure 11** and **14**). Genes involved in the *cOpsin1* cascade (*pde9, cnga*) of cPRCs were not apparent in this analysis, however they were also not present in the marker genes of any subcluster. It is possible that the link to *cOpsin1* lies in its ability to also directly affect the excitability of neurons via Gβγ signaling (Tsukamoto & Kubo, 2023).

2.4.7. Ventral MB *cOpsin1*-expression is stage specific

In light of the neurogenic gene signature of the ventral MB cluster, I investigated whether this expression changed across stages. Indeed the expression of *cOpsin1* was only detectable in the *premature* libraries in the combined MB subset, although it did not reflect obvious changes in cell type abundance (**Figure 22B**). Bearing in mind the different cell types per sample, **Figure 22C** shows that *emx2* expression is present across stages, diminishing in reproductive worms. It persists in *cOpsin1* knockout cells included from Milivojev et al. (2024). On the other hand, *cOpsin1* expression in the MB is restricted to the WT premature stage (**Figure 22D**). Interestingly, the premature EdU+ cells from the included single *cell* dataset (Milivojev et al., 2024), did not show *cOpsin1* expression (**Figure 22D,** panel 3), lending support to the fact that *cOpsin1* in these cells might lead to exit from an aNSC - like state.

Although there was not a clear shift in relative cell type abundance when looking only at the combined MB subset (**Figure 22B**), the *emx2+* ventral MB cluster and the *mef2c/maf+* transitional cluster were overrepresented in premature worms compared to reproductive worms when looking only at the snRNAseq neuronal subset from section 2.4.3, in contrast to the dorsal MB *emx2-/tll+* cluster (**Figure 23**, highlighted). This discrepancy may be due to the fact that the reproductive stages in the single *nuclei* dataset were from later - fast swimming spawning stages - compared to the single *cell* dataset - where only a colour change had taken

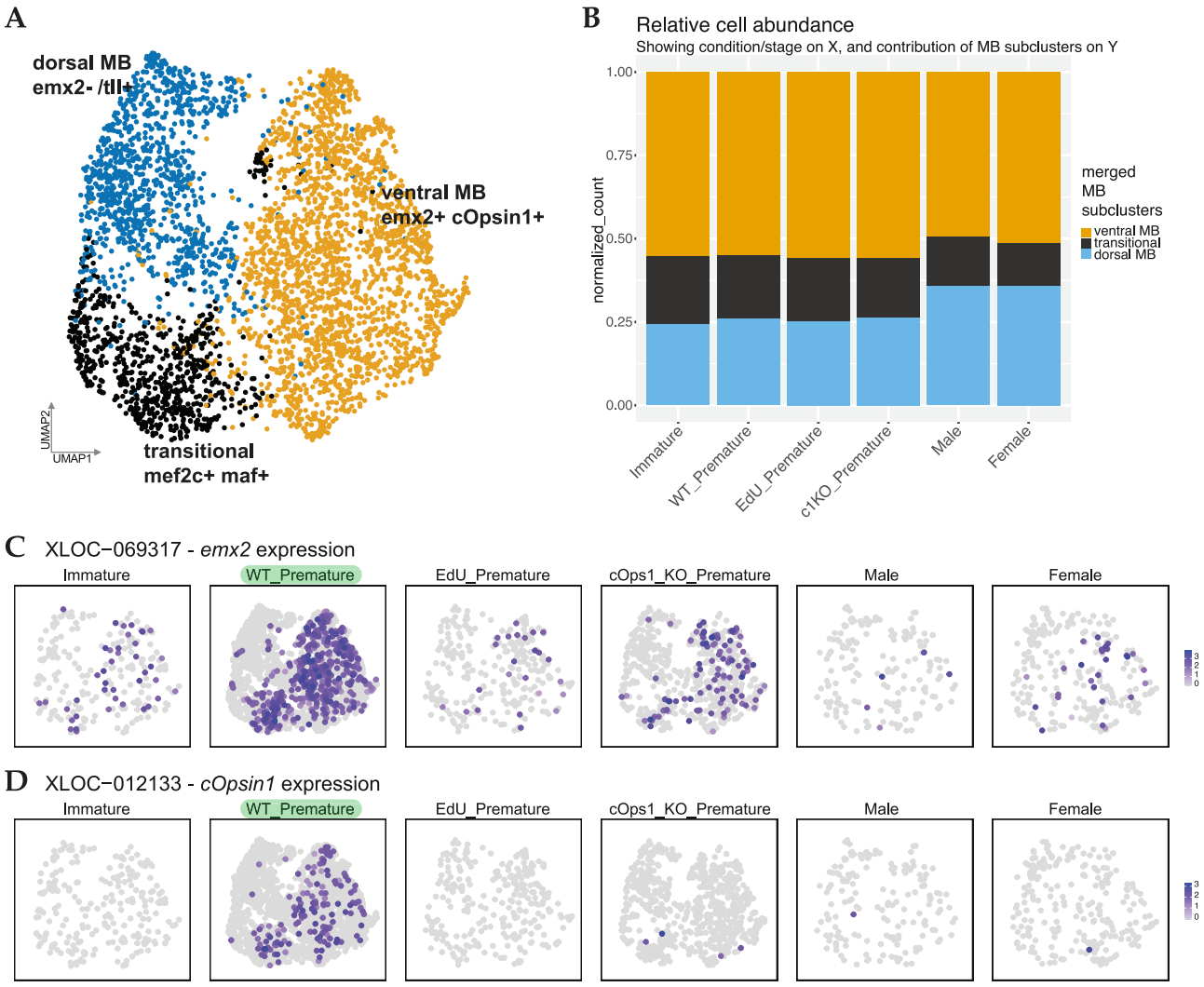


Figure 22. Stage - specific *cOpsin1* expression in the MB and relative subcluster abundance. (**A**) UMAP dimensional reduction of three MB subclusters. (**B**) Relative cell abundance of MB subclusters. Stage and condition along x-axis, with cell count relative to total number of cells from each stage/condition on the y-axis. (**C**) FeaturePlots showing *emx2* expression across stages and conditions. (**D**) FeaturePlots showing *cOpsin1* expression across stages and conditions.

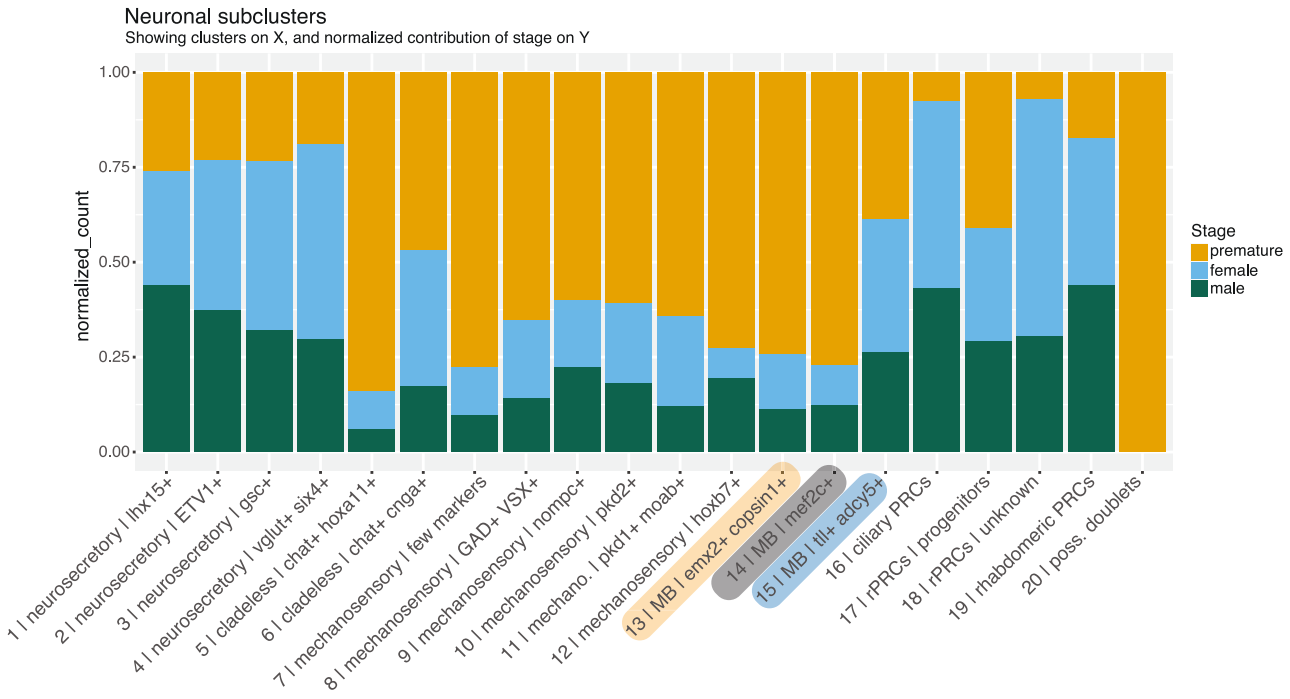


Figure 23. Relative cell type abundance across the entire neuronal subset (single nucleus dataset). Neuronal clusters annotated as in **Figure 12** shown on the x-axis. Nuclei count relative to the total number of nuclei from each stage on the y-axis. Bars coloured by stage.

place. The different normalisation methods used for the MB subset (SCTransform) compared to the neuronal subset (log normalised and scaled) integration might also have an effect. Interestingly, this change in relative cell abundance upon sexual maturation extends to other neuronal cell types (**Figure 23**). This includes the expected rhabdomeric PRCs (clusters 18 - 19) that are a cell type that has been shown to increase as a result of *cOpsin1+* progenitors (Milivojev et al., 2024). It also includes other neurosecretory cell types and, interestingly, the ciliary PRCs (cluster 16), whose UV detection capabilities may be important for the return to a pelagic swimming phase for spawning. Overall, the results of this section prompted me to explore whether these computational findings were reflected in anatomical differences between the MBs of premature and reproductive stages.

2.4.8. Restructuring of MBs upon sexual metamorphosis

Nuclear stains were used to assess whether morphological changes were observable in the dorsal and/or ventral MB as a result of sexual metamorphosis. Indeed, the ventral MB of reproductive worms (both males and females) appeared smaller/restructured compared to premature worms (**Figure 24**)**.** Reproductive worms were sampled after visible colour change distinguishing males and females but not yet indicative of worms ready to spawn (**Figure 24E - H** and **M - P**), as well as fast swimming spawning worms (**Figure 24I - L** and **Q - T**). This small sampling is indicative of a progressive restructuring specifically of the ventral MB. This corroborates what is seen in the relative nuclei abundances of the *emx2/cOpsin1+* ventral MB subcluster for reproductive stages from the previous section. Of WT MBs imaged (premature MBs, n = 16, and reproductive MBs, n = 26), 22 of the reproductive MBs showed this qualitative phenotype in the ventral MB. Of the *cOpsin1* knockout MBs imaged (premature MBs, n = 12, and reproductive MBs, n = 18), 14 of the MBs from reproductive stages showed this phenotype. The few reproductive MBs that appeared to be intact were all female worms having only undergone a colour change. All premature MBs were intact. Further quantification of size, and nucleus morphology is necessary to draw conclusions about the effect of the knockout, but representative images are in **Figure S3**.

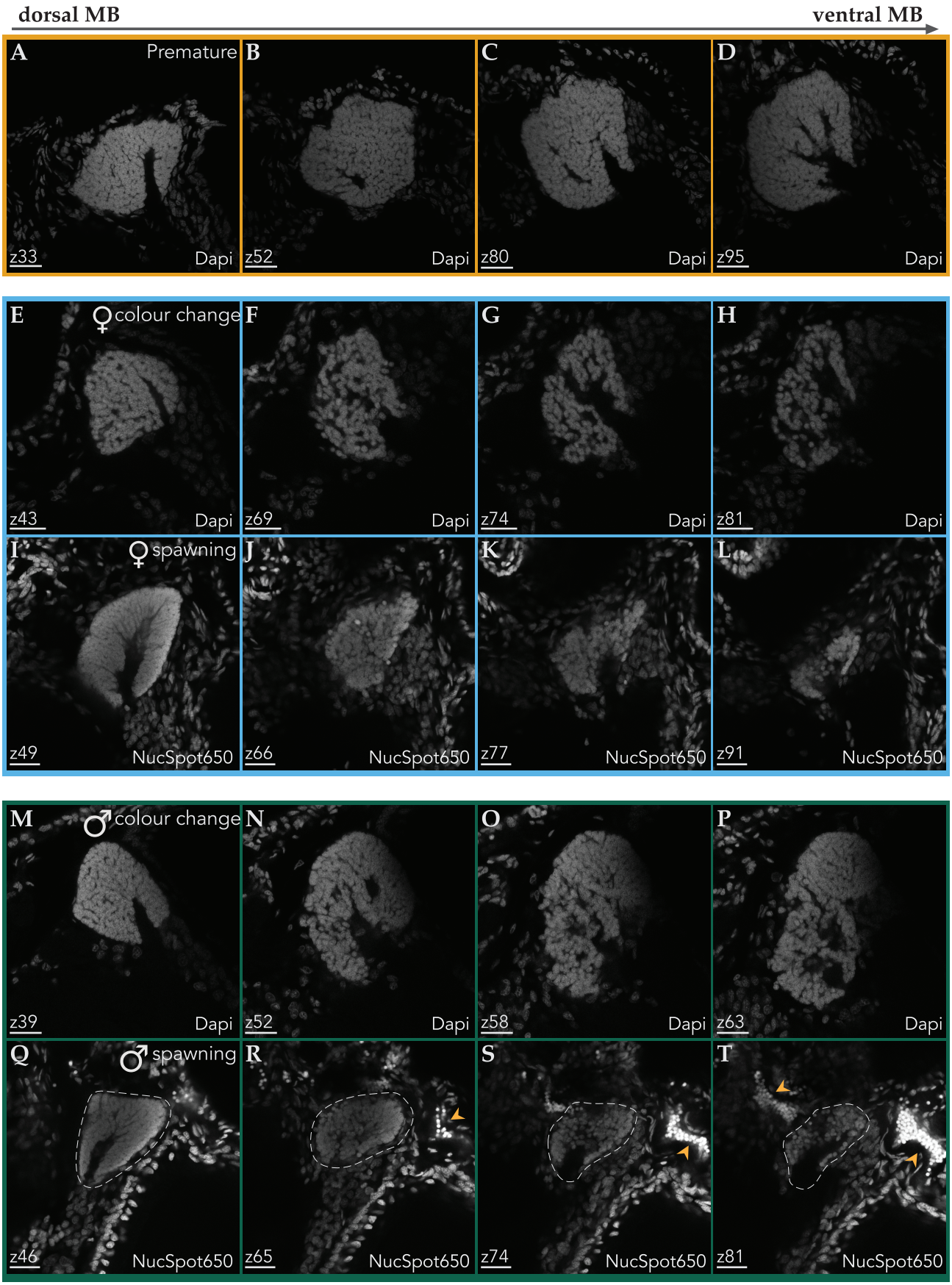


Figure 24. Restructuring of MB upon sexual metamorphosis visualised with nuclear stains. Dorsal (left panel) to ventral (right panel) MBs shown. (**A - D**) Premature MB, z - step size 1µm. (**E - H**) Colour changed female MB, z - step size 1.5µm. (**I - L**) Spawning female MB, z - step size 1µm. (**M - P**) Colour changed male MB, z - step size 1.5µm. (**Q - T**) Spawning male MB, z - step size 1µm. Arrowheads mark bright regions of male gametes stored in the head. For each image panel, top is anterior, right is lateral. All scale bars are 20µm. Images acquired with a confocal 40x oil objective.

2.5. Summary and perspectives

2.5.1. Summary

This chapter has shown that, in the quest to define annelid mushroom body cell types using single cell transcriptomics, an unexpected gene signature and an expanded view of well known *P.dumerilii* genes in the maturing MBs emerges.

2.5.1.1. A distinct transcriptomic signature for annelid mushroom bodies.

First, and most relevant for the next chapter, is that the intrinsic neurons of the annelid MB form a very well defined cluster in a snRNAseq dataset. This approach required few samples and yet was able to corroborate and complement recent single *cell* transcriptomic dataset of similar stages (Milivojev et al., 2024). The nuclei dataset was corroborative in that cell types and families, including the newly described rhabdomeric eye progenitors (Milivojev et al., 2024), could be recovered in both datasets. It was complementary in that the additional region of *cOpsin1* expression in the MB cluster was most evident in the nuclei dataset, as was its expression in the now annotated ciliary photoreceptor cluster. This may be due to cytoplasmic *cOpsin1* transcripts being hard to capture, since this opsin has been notoriously hard to stain with classic *in situ* hybridization techniques (Ayers et al., 2018). Investigating the translation dynamics of this gene might provide answers, since heavily translated genes tend to be covered in ribosomes, although how this might impact staining or sequencing methods is unclear (Paulet et al., 2017; VanInsberghe et al., 2021).

From the snRNAseq dataset, the gene signature of the mature annelid MBs, based on known and novel markers, indeed reflects an interneuron identity. Particularly, the maintained expression of *arx* is one indicator, known to be important in the development of cortical GABAergic interneurons and their migration from the medial ganglionic eminence in vertebrates (Kitamura et al., 2002; Marsh et al., 2016; Tsuboi & Yoshihara, 2025). Interestingly,

in postnatal stages, *arx* is also important for interneuron migration and differentiation from the subventricular zone (Tsuboi & Yoshihara, 2025) and in the function of parvalbumin interneurons (Joseph et al., 2020). The other known marker, *ptf1a,* is also an important player in interneuron specification outside of the cortex (Glasgow et al., 2005; Hoshino et al., 2005), possibly even for interneurons in other parts of the body in 6dpf *P.dumerilii* (Oel et al., *in prep.*). The new MB marker gene, *neurocalcin delta* (*ncald*), is also reported. It is part of the neuronal calcium sensor family, able to transduce intracellular calcium into diverse signaling pathways (Burgoyne, 2007), and has been shown in mice to increase in expression postnatally, even being implicated in the process of hippocampal adult neurogenesis in a recent study (Upadhyay et al., 2019).

The neurotransmitter signature of this cell type was cholinergic, similar to the case of cephalopod vertical lobe interneurons (Bidel et al., 2023; Styfhals et al., 2022). Considering that in vertebrates, both GABAergic and cholinergic telencephalic interneurons share a common developmental origin (Marín et al., 2000; Su-Feher et al., 2022), and that there exist interneurons with GABAergic and cholinergic co-transmission (Lozovaya et al., 2018) as well as evidence for cholinergic and glutamatergic co-transmission (Kljakic et al., 2017; Vergara et al., 2021), it is reasonable that this interneuron diversity stems from an ancestral cell type that gave rise to the annelid interneuron population as well as the interneuron diversity seen in other species. As the brains of more Lophotrocozoan species are investigated in such detail, including in the context of full body atlases, this evolutionary history of interneurons will be further uncovered.

Lastly, it is worth noting that at 6 dpf, spatially mapped *in situ* stainings revealed TF signatures for eight regions of the developing MB (Vergara et al., 2021). These divisions were not apparent in the present single nucleus dataset. This could be due to the fact that cells of different morphology or function - and anatomical position - can have very similar transcriptomes (Özel & Desplan, 2025; Shainer et al., 2025). Alternatively, cell types may have

distinguishable transcriptomes in early development but become indistinguishable in later life stages (Li et al., 2017). The latter is somewhat unlikely, since a whole-body atlas of *P.dumerilii* at 6dpf, shows the same two subclusters of the bonafide MB cluster - an *emx2+* half, and an *emx2-* half (Oel et al., *in prep.*).

2.5.1.2. Divergent gene signatures among MB neurons.

Differing gene signatures for the ventral and dorsal MB fits with previous functional and anatomical reports of a dorsal versus ventral dichotomy of the *P.dumeirlii* MBs (Chartier et al., 2018; Tomer, 2008). In the present investigation the ventral MB cluster specifically expressed receptors for serotonin and neuropeptide Y, as well as the adrenergic receptor *adrab2.* On the other hand, the dorsal MB cluster specifically expressed a possible neurokinin receptor (*tacr2*), a histamine receptor (*hrh1*) and octopamine and dopamine receptors. Many of these receptors have been implicated in olfactory learning in other MB - possessing animals. Serotonin receptors in the fly mushroom body are necessary for olfactory learning of reward stimuli (Ganguly et al., 2020; Lee et al., 2021), as are receptors for octopamine and dopamine (Kim et al., 2013; Wu et al., 2013). Interestingly, in mice, neuropeptide Y is able to signal hunger - via *NPY5R* - and leads to the preference of food-related odours (Horio & Liberles, 2021). The neuropeptide Y ortholog in flies also signals hunger, increasing appetitive memory formation (Krashes et al., 2009), suggesting an evolutionarily old system for signaling hunger. This offers one hypothesis, as to whether the ventral MB is responsible for integrating hunger - related cues, which might be especially relevant for *P.dumerilii*'s benthic growth phase and become less important upon the cessation of food intake associated with sexual metamorphosis. Work by Chartier et al. (2018) did show that the dorsal versus ventral MB responded differently to chemical stimuli, although only a few stimuli were tested and both dorsal and ventral MBs responded to the food - related cue of glutamate, with the dorsal MB even responding to sucrose.

An alternative explanation comes from larval connectomics work showing that the MBs are made of intrinsic interneurons as well as output neurons, the former projecting to the neuropil and the latter projecting directly onto the ventral nerve chord and mouth (Jékely et al., 2024). Input directly from sensory structures, like the antennae and palps is sent to the output neurons forming small sensory-motor circuits. While it is unclear whether these populations may represent the division seen in the present transcriptomic dataset, it provides an alternative possibility. Although, it is unclear how such circuits change in later developmental stages - after loss of cilia and after settlement and cephalic metamorphosis - in preparation for the worm's benthic growth phase. MB-intrinsic sensory neurons have also been reported in 3dpf and 6dpf stages (Jékely et al., 2024; Vergara et al., 2021). If these persist into adulthood and are present in the present dataset, they are likely hidden among the mechanosensory clusters of this atlas, not with the investigated interneuron clusters. Further characterization of the other cell types in this atlas as well as future circuit - level and behavioural studies for adult stages of *P.dumerilii* will be telling.

2.5.1.3. Maintained neurogenic potential in ventral MBs.

Besides these differences in effector genes, the clearest signature distinguishing the ventral MB cluster, was the specific expression of *emx2* and *cOpsin1* in premature worms, although the expression of *cOpsin1* in the MB was not apparent in some HCR stainings, despite being present in all premature *sn*RNAseq libraries. A technical reason for this is unlikely because *cOpsin1* expression was seen in other regions of the head - the cPRCs or growing rhabdomeric eyes - in the same stainings. While worms were fixed at the same time of day, a detailed investigation of the influence of circadian rhythmicity would be beneficial, although, bulk transcriptomics of the maturing brain have shown that circalunar phase has a surprisingly small impact on transcripts of the brain, and rather affects the proteome (Schenk et al., 2019). Therefore, a detailed investigation - with more samples - across segment sizes would also be useful, since segment number is a proxy, and indeed some worms with fewer than 60 segments appeared further along in the maturation process (large

eyes, large size). It is therefore likely that the threshold for *cOpsin1* expression in the MB, and likely its role in sexual metamorphosis, is influenced by other factors, such as metabolic state which is another signal that needs to be integrated for the maturation process to occur (Andreatta et al., 2020; Schenk et al., 2016). Nevertheless, this signature was clearly localised to the ventral MB in stainings, whose relative abundance was reduced in the single nucleus dataset for reproductive stages, and which appeared sparser in nuclear stains. This suggests that the neurons of the ventral MB undergo restructuring during sexual metamorphosis, either by acting as another neurogenic niche in the adult brain and/or via apoptosis.

Indeed proneural genes reported in Milivojev et al. (2024), showed higher expression in the ventral MB compared to the dorsal MB cluster, and genes that were broadly downregulated in the MB of the *cOpsin1* knockout dataset included the conserved *soxC* transcription factor implicated in early *P.dumerilii* neuronal differentiation (Kerner et al., 2009), as well as in vertebrate adult neuronal fate commitment and glial reprogramming (Mu et al., 2012). This coincided with an upregulation of *arx* in the ventral MB subcluster, known for its expression in the proliferative region of the MBs at 6dpf (Arendt et al., 2021), suggesting *arx* may play a role in maintaining this potential throughout the lifespan. It is also interesting that a distinct transitional population appears in the single cell dataset; one that is similar in expression to the ventral MB cluster but with *mef2c* and *maf* transcription factors among its top markers, which together have been reported to lead to parvalbumin interneuron fate - rather than somatostatin interneuron fate - from subventricular zone progenitors (Pai et al., 2020). This transitional cluster also showed a maintained expression of *emx2* in the knockout dataset (Milivojev et al., 2024), a gene that in the vertebrate brain seems to be expressed in adult neural stem cells (aNSCs) (Hong et al., 2007; Wei et al., 2011). Moreover, there was specific downregulation of *crebzf* in the ventral MB and *CEBPß* in this transitional population in the *cOpsin1* knockouts. While the former has been implicated in both differentiation or apoptosis (Bodnarchuk et al., 2012; Valderrama et al., 2009), the latter basic region-leucine zipper (bZIP) domain - containing transcription factor is known to ensure

survival of new neurons during adult neurogenesis in the vertebrate hippocampus (Cortes-Canteli et al., 2011; Herold et al., 2011; Hodge & Hevner, 2011). Lastly, and importantly, *cOpsin1* has been reported to be expressed in progenitor cells leading to rhabdomeric eye growth in *P.dumreilii* (Milivojev et al., 2024), forming a strong link between its expression and the morphological changes occurring in the *P.dumerilii* brain associated with sexual maturation.

The expression of an opsin in unconventional cell types is not surprising. In zebrafish and medaka, opsins have even been found to be expressed in cholinergic interneurons as well as motoneurons (Fischer et al., 2013). Such multi-sensory interneurons present in both vertebrates and annelids lend further evidence to the notion that they may stem from an ancestral cell type family. It would be interesting to see if *cOpsin1* might also be expressed in interneurons of other Lophotrocozoan species, including other annelids with comparable brain structures, and whether, in *P.dumerilii,* this opsin might also be expressed during another dramatic morphological transition - cephalic metamorphosis. Especially considering preliminary findings by my Master's student Gülce Serka, showing that *cOpsin1* is expressed in regions beyond the ciliary photoreceptors in worms ranging from 6dpf, 8dpf, 14dpf and 2 months old.

2.5.1.4. Anatomical restructuring of annelid mushroom bodies upon sexual metamorphosis.

Simple nuclear stains were able to highlight the dramatic restructuring of the MB in sexually mature worms compared to premature worms. While the nuclei surrounding the dorsal peduncle seemed intact, the ventral MB showed more sparse nuclei in line with tentative relative cell abundance changes in the single nucleus atlas. Whether this restructuring is a result of apoptosis, differentiation, or both, remains to be seen. Literature on the sexual maturation of *P. dumerilii* describes a metamorphosis that involves the dedifferentiation or transdifferentiation of many cell types (A. Fischer & Dorresteijn, 2004; A. H. Fischer et al., 2010; Hauenschild, 1956, 1966, 1974; Schenk et al., 2016). Together with recent lines of

evidence showing that posterior regeneration of premature worms (Stockinger et al., 2024), as well as the restructuring of musculature during sexual metamorphosis to support the behavioural changes associated with spawning (Dahlitz et al., 2023), are likely a result of dedifferentiation of existing cell types, it would not be surprising that similar remodeling mechanisms are at play in the MBs of maturing *P.dumerilii.*

2.5.2. Linking *cOpsin1* and *P.dumerilii* sexual metamorphosis

How then might *cOpsin1* impact this process? Qualitative inspection of the MBs suggests that knockout worms *do* undergo the restructuring of the ventral MB, meaning that perhaps the effect of *cOpsin1* is at the level of timing or synchronization. Further quantification is needed to reveal if there are changes in the relative size of the dorsal MB - which visually remains intact - and, importantly, how exactly the *cOpsin1* knockout affects this restructuring. These changes are likely triggered in part hormonally, as suggested by the receptor expression of the overall MB cluster in the snRNAseq dataset, including the receptor for the circadian neuropeptide PDF (pigment dispersing factor) (Arboleda et al., 2019; Shahidi et al., 2015). Furthermore, *P.dumerilii cOpsin1,* which responds to UVA/deep violet light, has been shown to regulate the enzymes responsible for synthesis of neurohormones including PDF and NPY (Tsukamoto et al., 2017; Veedin Rajan et al., 2021), for which receptors were expressed in the MB subclusters explored. While during pelagic larval stages, *cOpsin1* acts as a UV – sensing depth gauge as part of the cPRC circuit (Verasztó et al., 2018), during the benthic growth stage, this opsin remains important since it can sense the increased UVA light during summer time (Veedin Rajan et al., 2021). The expression of *cOpsin1* in progenitor cells of the rhabdomeric eyes that sit in a stem cell niche surrounding the lens (Milivojev et al., 2024), and now its expression in the ventral MB, further suggest a role for *cOpsin1* in timing certain aspects of sexual metamorphosis. Relative cell abundance analysis suggests that there are other cells of the brain that increase in abundance upon sexual metamorphosis, and anatomical comparisons of the MBs point to a

reduction in the neurons of the ventral MB. It therefore seems plausible that other aNSC-like niches exist in the brain of *P.dumerilii* that contribute to these cell type shifts.

Moreover, *P.dumerilii cOpsin1* is biased towards signaling in a Gβγ - dependent way, meaning it can directly hyperpolarize a neuron by opening inwardly rectifying potassium channels leading to fast self-inactivation (Tsukamoto & Kubo, 2023). It is also a bistable opsin, able to convert between *cis* - retinal and *trans* - retinal, and therefore allowing it to function without the need for a new retinal molecule in tissue without a large supply (Tsukamoto et al., 2017; Tsukamoto & Kubo, 2023; Wietek et al., 2024). The specific expression of *hcn1* - a hyperpolarization - activated, cyclic nucleotide - gated cation channel - in the *cOpsin1* expressing ventral MB subcluster could be an important link. This channel is known for sharpening the responses of vertebrate ciliary photoreceptors to light (Barrow & Wu, 2009), as well as the ability to alter membrane excitability of neurons implicating it in a variety of functions, including pacemaking and long - term potentiation (Bender & Baram, 2008; Biel et al., 2009). This effect on membrane excitability, as well evidence suggesting that such channels may directly conduct the ubiquitous second messenger, calcium (Michels et al., 2008), provides another way of linking the expression of this *cOpsin1* to downstream gene expression changes. While the exact downstream signaling needs further investigation, the restructuring in the MB likely results from either a direct effect of *cOpsin1 or* an indirect hormonal effect as mentioned above.

2.5.3. A proposed *cOpsin1* - dependent scheme for MB restructuring

Overall it appears that the maturing *P.dumerilii* MB has a subset of interneurons with maintained neurogenic potential. I propose a *cOpsin1* dependent scheme for their differentiation and/or apoptosis to be further investigated in future studies (**Figure 25**).
In sum, at *around* 60 segments - the exact threshold needs further investigation and is likely defined by multiple factors - the ventral MB begins expressing the UV - sensing opsin, *cOpsin1*. This opsin is able to detect seasonal changes in UVA/deep violet light intensity

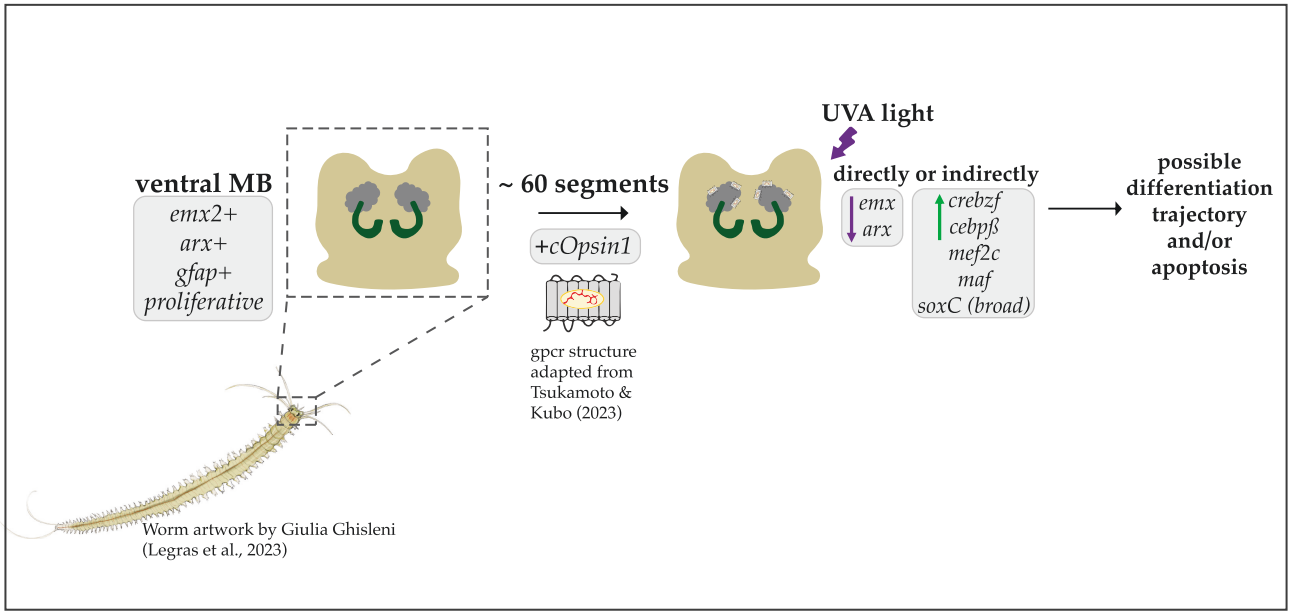


Figure 25. Summary scheme with segment - dependent *cOpsin1* expression in the MBs and resulting gene expression changes.

(Veedin Rajan et al., 2021) which, together with hormonal and metabolic signals (Andreatta et al., 2020; Schenk et al., 2016), might coordinate aspects of sexual metamorphosis. The *cOpsin1* expression in *emx2+* neurons of the ventral MB, may directly or indirectly lead to their differentiation and/or apoptosis. The expression of the *P.dumerilii soxC* neuronal differentiation marker and the transitional *mef2c+/maf+* population lends support to the former possibility, as does the expression of *cOpsin1* in the rhabdomeric eye progenitors - a neurogenic niche necessary for eye growth occurring during sexual maturation (Milivojev et al., 2024). However, if these neurons are using vertebrate - like mechanisms of adult neurogenesis, a balance of both processes is most likely (Ryu et al., 2016). Taken together, and since the *cOpsin1* knockout worms are capable of maturing, the role of this opsin in the MB neurons might be one facet of a timing mechanism that leads to the dramatic restructuring of the brain during sexual metamorphosis.

2.5.4. Future perspectives

While characterizing cell type signatures for the purpose of drawing evolutionary comparisons, I stumbled upon the unexpected expression of *cOpsin1* and *emx2* (a possible paralog for the developmental *P.dumerilii emx*) in the mushroom bodies of late stage worms. Single nucleus and single *cell* transcriptomics datasets revealed a potentially ancestral *emx* based mechanism of adult neural progenitor maintenance (Hong et al., 2007; Kobeissy et al., 2016) and a *soxC* based mechanism of *adult* neurogenesis (Kerner et al., 2009; Mu et al., 2012). If adult neurogenesis is indeed the mechanism at play, it will be important to confirm the occurrence of apoptosis and/or differentiation in the ventral mushroom body *in situ*, since a balance of both processes are likely implicated (Ryu et al., 2016). A trajectory analysis would offer further insights into whether these findings indeed represent a differentiation trajectory, although it is yet unclear what the endpoints may be since many neuron types change in abundance during this metamorphosis. A single cell sequencing time course experiment, sampling worm heads regularly between premature and reproductive stages would help in disentangling the fate of these ventral MB neurons. Lastly, with on-going

efforts to annotate the *P.dumerilii* genome (Mutemi et al., 2025) with detailed phylogenetic analyses, the true annotations of some of the genes implicated in this chapter and how similar they are to vertebrate orthologs, will be uncovered. This will further disentangle the mechanisms at play.

Indeed, the *cOpsin1* knockout is a lifelong knockout, it is unclear how this may affect other life stages especially during cephalic metamorphosis and the benthic growth phase. However, the expression of *cOpsin1* - at least in the MBs - appears to be something that is life stage specific, since it only appears during a small window before sexual metamorphosis according to the single nucleus dataset. This selective expression might suggest that dramatic changes to overall development are negligible in the knockout.

Overall, this chapter has shown how fruitful exploratory single cell transcriptomics can be in terms of uncovering unexpected gene expression patterns in known and novel cell types. What other levels of possible evolutionary conservation (or convergence) can be investigated with this dataset? The next chapter revisits the broader MB cell type annotation in an attempt to infer whether these interneurons might belong to an ancestral cell type family by comparing the current dataset to those of animals with either similar structures - insects - or those with neurons expressing similar candidate genes - cephalopods.

3. Beyond candidate genes: Comparing *Platynereis dumerilii* mushroom body neurons to distant species

3.1. Contributions

The analysis performed in this chapter was thanks to a pipeline conceived by Phil Oel to compare photoreceptor cell types across vertebrates, and who initially integrated the *P.dumerilii* brain dataset with several distant Ecdysozoan and vertebrate species. I wanted to use this pipeline and do a targeted integration of the Platynereis brain dataset with that of a chemosensory insect species and a more closely related Lophotrocozoan species. I used the *Harpegnathos saltator* dataset (Sheng et al., 2020) mined by Phil for the former aim. For the latter, I mined for another Lophotrocozoan species dataset that was easy to access and interpret. This led me to the *Octopus vulgaris* single-cell dataset (Styfhals et al., 2022).

3.2. Background

3.2.1. How to compare across large evolutionary distances

Drawing inferences about ancestral cell types by comparing cell types in extant species requires methods to bring the units of comparison into the same space. That means finding common ground between proteomes with very different use of paralogs, or species-specific paralog use. This chapter will not be a benchmarking chapter comparing different methods of doing so, but rather a test of one such method, to see if it might provide insights into whether the *P.dumerilii* mushroom body interneurons, as broadly defined in Chapter 2, might be related to neurons in the brains of distant species deemed similar thus far based on candidate gene approaches and/or anatomy (Arendt et al., 2021; Styfhals et al., 2022; Tomer et al., 2010). The overall goal of methods being developed towards this aim, is to remove the batch effect of *species* in order to compare cell types in a common space, something that becomes difficult the more distant the species being compared (Song et al., 2023).
One such method designed for whole-body atlases, SAMap, does so by iteratively determining protein sequence and mapping the single cell datasets into the same low dimensional space (Tarashansky et al., 2021). Here, complete cell type annotations are important, and for distant species with less-well annotated genomes, this missing information may be confounding (Song et al., 2023; Tanay & Sebé-Pedrós, 2021; Tarashansky et al., 2021). For *P.dumerilii* at least, paralogs are thus far inferred from emapper annotations. Future work still needs to be done to resolve which gene duplications are real versus those that may for now be computational artefacts. Furthermore, shorter genes (hormones, neuropeptides) are missing from the emapper annotation used at the time of this analysis, adding to the annotation sparsity.

Nevertheless, for the purposes of this chapter, the Seurat CCA approach will be used to attempt to integrate brain single cell datasets from distant species (Song et al., 2023). This

will be achieved by collapsing species-specific genes (as determined by emapper annotations) into orthology groups or orthogroups (OGs) so that all species considered will exist in the same "gene" space. This method has also been recently reported to corroborate findings by SAMap, such that both methods were able to show correspondence of Cajal-Retzius cells between the shark, mouse, and the salamander (Quintana-Urzainqui et al., 2025).

3.2.2. Selected comparison species

I first selected another Lophotrocozoan - *Octopus vulgaris* - because of the transcriptomic similarity of its vertical lobe interneurons and the *P.dumerilii* MBs (Bidel et al., 2023; Styfhals et al., 2022). Additionally, despite cephalopods being rather derived, they are more closely related to *P.dumerilii* than the other phyla discussed (ecdysozoans, vertebrates). Lastly, I wanted to include a species shown to have a transcriptomically similar cell type, but that did not have the conserved mushroom body structure described for annelids and insects. Beyond these reasons, data accessibility was a big factor in the datasets selected.

It was also worthwhile to include an insect comparison because of the similarity in both structure (Heuer et al., 2010, 2012) and in transcriptomic signature with insect MBs (Tomer et al., 2010). Furthermore, since the focus is on neurons of the MB, the *Harpegnathos saltator* dataset proved well - suited (Sheng et al., 2020). Authors showed that, compared to drosophila, the ant MB neurons - the Kenyon cells (KCs) - made up a much larger fraction of the total cells of the brain, both in the explored single cell datasets as well as anatomically (Sheng et al., 2020). Though these regions have evolved extensively in each lineage, the hypothesis is that the major neuron types underpinning them may have been present in a common ancestor, and may be reflected in some mixed - species clusters upon integration.

3.3. Materials and methods

3.3.1. Collapsing emapper predicted proteins into Orthology groups

In order to compare Seurat objects of these distant species, an estimated transcriptome of the last Bilaterian ancestor was generated, meaning that for each species proteome, genes were pooled to their orthology group (orthogroup or OG) - level gene identifiers. The *O.vulgaris* proteome was downloaded from UniProt (UP000515154**,** https://www.uniprot.org/proteomes/UP000515154), which was estimated from the *Octopus sinesis* genome also used in Styfhals et al. (2022). The *H.saltator* proteome was edited as described by Phil and came from work by Shields et al. (2021) (GSE172309, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE172309). The *P.dumerilii* proteome was estimated from its genome (Mutemi et al., 2025) and edited by Phil. For each proteome, the longest protein sequence per gene was kept. These curated proteomes were edited so gene names of the proteomes matched those of the single cell objects. These proteomes were run through emapper (v.2.1.3) using the *emapper.py -m mmseqs -i* command. The output included not only the original gene names but also orthogroup codes for different taxonomic levels. The Bilaterian orthogroup codes were used. From this annotation file, the top-scoring orthogroup match per gene was kept, and orthogroup identifiers corresponding to Bilateria were extracted. The count matrices of each original seurat object were pooled such that counts for genes falling into the same orthogroup were summed , resulting in an orthogroup-level count matrix used for the downstream cross-species analysis.

A schematic example of how this brings three species into a comparable gene space is shown in **Table 2** and **Table 3**, which use *mef2* (*myocyte enhancer factor 2*) genes as a simple case to show the schematic count matrices before and after collapsing by orthogroup, respectively. All gene codes are actual species gene codes that were annotated as a *mef2* paralog (**Table 2**) and the corresponding bilaterian OG identifier assigned by emapper (**Table 3**).

Table 2. Hypothetical count matrices before pooling counts into orthology groups (OGs). Species - specific gene codes are those annotated as a *mef2* gene according to emapper. Despite all receiving the same annotation, count matrices cannot be integrated because of species - specific gene names.

***Pdum*-gene codes**	**Cell 1_*pdum***
XLOC-000048	2
XLOC-049566	0

***Ovul*-gene codes**	**Cell 1_*ovul***
LOC115211031	2
LOC115211138	1

***Hsal*-gene codes**	**Cell 1_*hsal***
LOC105192578	1

Table 3. Combined count matrix after summing counts by orthogroups. For each species, all gene codes annotated as *mef2* - regardless of paralog - are pooled by their common bilaterian OG code, allowing datasets to be integrated.

OG gene identifiers	**Cell 1_*pdum***
3CRCH	2

OG gene identifiers	**Cell 1_*ovul***
3CRCH	3

OG gene identifiers	**Cell 1_*hsal***
3CRCH	1

This method ignores species-specific use of certain paralogs in more recently evolved cell types, however, for distant species, it serves as a good proxy for comparison, since such nuanced differences were likely absent in the last common - in this case bilaterian - ancestor.

3.3.2. Cross-species integration analysis

These modified count matrices were used to generate OG-pooled seurat objects using the CreateSeuratObject() function, then loading the metadata from the original seurat object. The *P.dumerilii* object alone was also clustered based on OG-level gene identifiers using 80 PCs and a Louvain resolution of 1, in order to ensure that similar neuronal clusters were recovered using this method. For the cross-species integration the OG-pooled seurat objects were integrated using the Seurat CCA method with 30 dimensions and 80 k.anchors. This was followed by standard clustering workflow, with 20 PCs based on Elbow plot inspection,

and a Louvain clustering at a low resolution of 0.2, to see if broad cell type families formed mixed - species Louvain clusters. Of note, the older version of this analysis (**Figure S4**) between only *P.dumerilii* and *O.vulgaris* datasets used 30PCs and a resolution of 0.5 for clustering. Low dimensional UMAP embedding was used to visualise the integrated objects. Markers were calculated with FindAllMarkers() with the same settings as in section 2.3.4.1.

3.4. Results

3.4.1. Orthogroup - clustering maintains original clusters

To ensure that using OG - level gene identifiers results in a clustering that matches what was described in Chapter 2, the *P.dumerilii* neurons were clustered independently using the same parameters from section 2.3.4.1, but using the OG - pooled seurat object. This resulted in 18 neuronal clusters (**Figure 26A**). By visualising this dataset with the broad annotations from **Figure 8A** and the more detailed neuronal annotations from **Figure 12A**, it is clear that the OG - level gene identifiers were sufficient to recover the same broad neuron families (**Figure 26B**) and most of the more detailed neuronal subclusters (**Figure 26C**) that were reported in Chapter 2.

3.4.2. *P.dumerilii* vs *O.vulgaris* vs *H.saltator* comparison

Integrating this dataset with the octopus and ant datasets, prepared in the same way, produced several mixed-species clusters after low resolution Louvain clustering (**Figure 27A - B**). Cluster annotations are assigned based on original species - specific cell types contributing 1) to at least 5% of cells of the cluster *and* 2) making up at least 50% of the cells from that species in the given cluster. Neurons labeled as unstable in the octopus dataset were those deemed as such based on a Jaccard index below 0.6 (Styfhals et al., 2022), in future iterations of this analysis, I would remove these clusters, as they likely obscure the contribution of the different annotated neuron types to the Louvain clusters. For example, the octopus glutamatergic, GABAergic, peptidergic, and serotonergic neurons, are hidden within the larger neuronal Louvain clusters 1 and 2, despite being consistent with the other dominant contributing cell types like *P.dumerilii* neurosecretory neurons. Higher resolution clustering may also resolve this. The *P.dumerilii* rPRC - dominated clusters align with the fact that this was the only dataset containing eye photoreceptors (**Figure 27B**). Further species - specific clusters include the *H.saltator* hemocytes as well as the large glial cluster that is

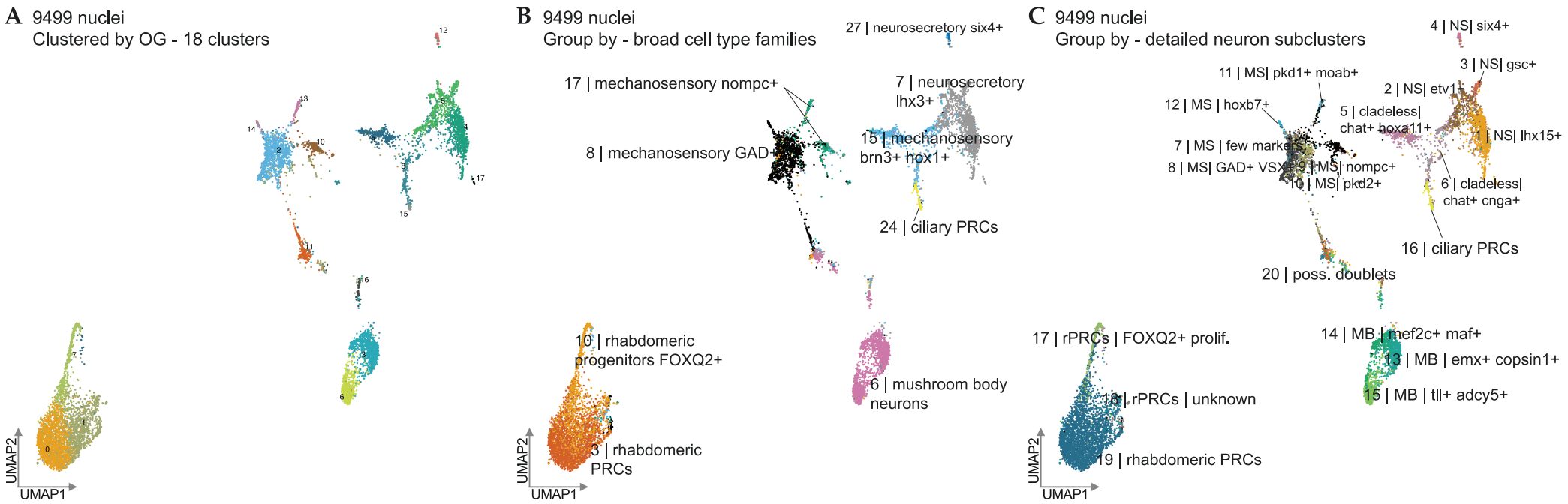


Figure 26. *P.dumerilii* neuron subset clustered based on orthogroup - level gene identifiers. (**A**) UMAP reduction showing 18 clusters emerging using 80PCs and a Louvain resolution of 1, as done for the gene - clustered neuronal dataset. (**B**) UMAP grouped by broad annotations from **Figure 8A**. (**C**) UMAP grouped by neuron annotations from **Figure 12A**. Abbreviations: MB: mushroom body, MS: mechanosensory, NS: neurosecretory, PRC: photoreceptor.

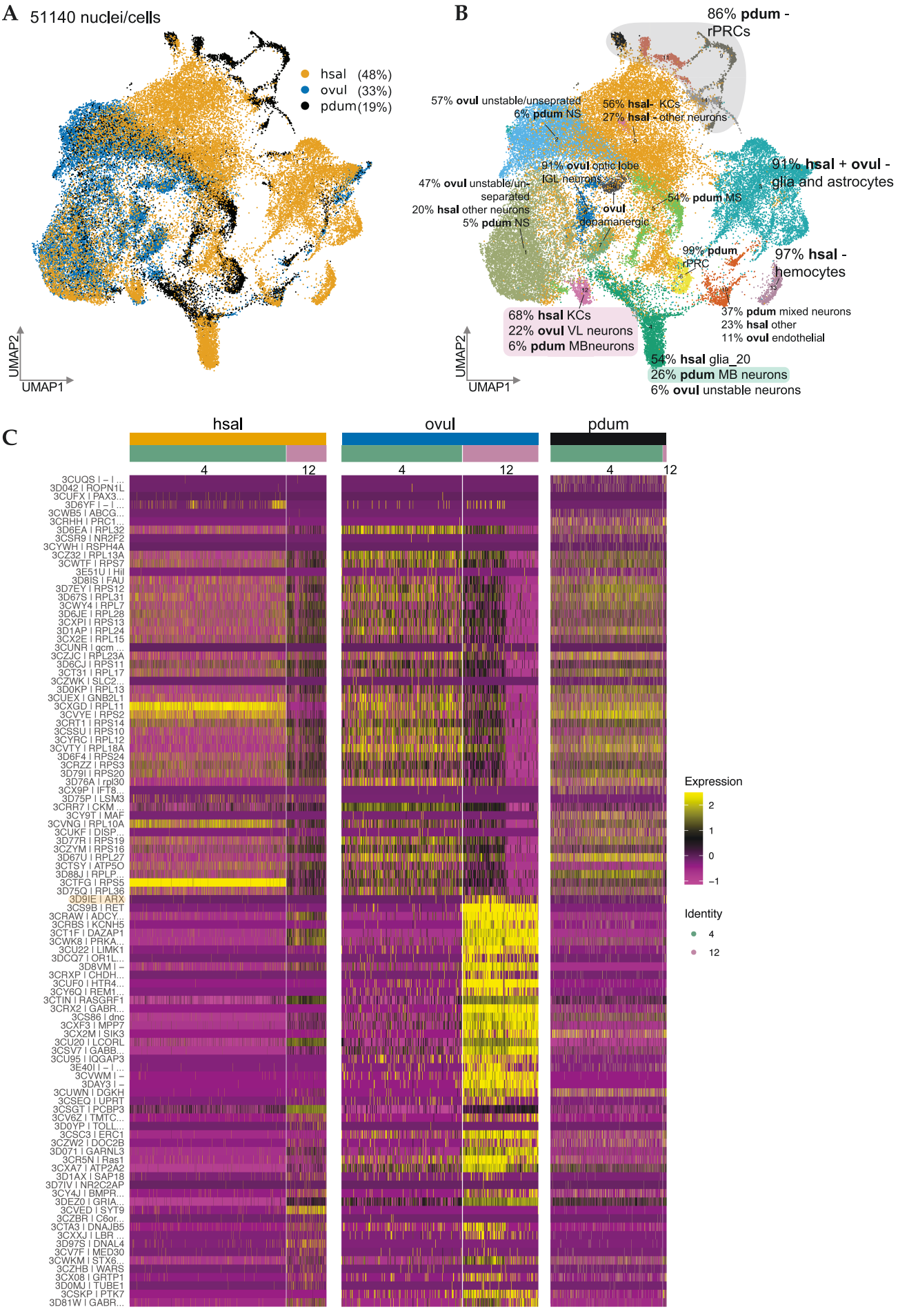


Figure 27. Three - species integration analysis. (**A**) Integrated UMAP dimensional reduction coloured by species - hsal (*H.saltator*), ovul (*O.vulgaris*), pdum (*P.dumerilii*). Percent of cells/nuclei contributed by each species is listed. (**B**) Low resolution Louvain clusters, labeled by the percent contribution of most prominent species - specific cell types. (**C**) Heatmaps showing top 50 (by average log fold change) OG markers for clusters 4 and 12 by species.

dominated by the octopus and ant cells because similar cell types from *P.dumerilii* were not included in the analysis (**Figure 27B**).

Interestingly, 96% of the cells contributing to Louvain cluster 12 came from mushroom body neurons (*P.dumerilii*), vertical lobe neurons (*O.vulgaris*), and class 1 Kenyon cells (*H.saltator*) (**Figure 27B**). This cluster contained 92% of the *O.vulgaris* vertical lobe neurons. It contained 5% of the total *H.saltator* KCs, with the remaining ant KCs either distributed or found in the *H.saltator* dominated cluster 0 (56% *H.saltator* KCs - 85% of the total KCs in the *H.saltator* dataset). While it only contains 2,5% of the *P.dumerilii* MB neurons, of the *P.dumerilii* neurons in this cluster, almost all of them came from the MB population (93%). The remaining *P.dumerilii* MB neurons were either distributed, some in the *H.saltator* KC - dominated cluster 0 (17%), and interestingly in cluster 4 (57%) which has an *H.saltator* glial cell type as its other dominant cell type. With the focus being on neurons of the MBs, especially of *P.dumerilii*, clusters 4 and 12 will be the focus of the marker OG analysis.

3.4.3. Marker Orthogroup identifiers

The top 50 (based on average log fold change) OG - markers for the integrated clusters 4 and 12 are shown in **Figure 27C**. The top markers for cluster 12 reflect the gene signature of the *O.vulgaris* vertical lobe neurons. This includes *arx* (3D9IE) which is also apparent in the small number of *P.dumerilii* - contributing cells, as well as LIMK1 (3CU22). While not among the top 50 markers, the OGs for *mef2* genes also appeared on this list (3D7N1, 3CRCH). Most other shared markers seem to be the OGs of effector genes like adenyl cyclases (3CRAW), guanine nucleotide exchange factors (3CTIN), and synaptotagmin (3CWKM).

The top markers for cluster 4 are mostly ribosomal proteins, even showing expression in a fraction of the *O.vulgaris* vertical lobe cells of cluster 12 (**Figure 27C**). The OG for a *maf* transcription factor, discussed in the previous chapter, appears on this list, although it is specific to the *P.dumerilii* cells of cluster 4. It is worth noting that the *H.saltator* glial cluster

that contributes 57% of the cells of the mixed - species cluster 4 was left unannotated in its original dataset, since it had glial markers but lacked specific markers (Sheng et al., 2020). Perhaps the neurogenic signature described in Chapter 2 drives the clustering of *P.dumerilii* MB neurons with this unannotated glial population, although the OG for *gfap* - a glial and proneural marker - was not among the marker list.

A previous version of this analysis, including only *P.dumerilii* and *O.vulgaris* datasets - before the additional filtering (section 2.3.4.1) of the *P.dumerilii* nuclei dataset - also resulted in two clusters (7 and 27) containing almost exclusively *O.vulgaris* vertical lobe and *P.dumerilii* MB neurons (**Figure S4A - A″**). Cluster 27 was dominated by the *O.vulgaris* vertical lobe gene signature, with many OG markers also expressed in the smaller number of contributing *P.dumerilii* MB neurons (e.g. *ret* - 3CS9B, *limk1* - 3CU22) (**Figure S4B**). While *arx* appeared among the top 50 OG markers for both clusters, the larger cluster 7 mirrored the three-species comparison, with many ribosomal proteins among the top markers expressed in the contributing cells of both species (**Figure S4B**). Also as in the three - species analysis, the *mef2c* OG (3CRCH) appeared as a marker for both clusters, just out of reach of the top 50.

3.5. Summary and perspectives

Overall, this approach, although promising, requires more detailed optimisation and comparison to other approaches in the literature. In future iterations, repeating this analysis with only the most confidently annotated cell types (including cell types with expected matches across datasets as well as those absent in some of the datasets as negative controls) could clarify the feasibility of this integration method with distant species. It will also benefit as the genome of *P.dumerilii* is further annotated. Even still, it does lend support to known MB genes being shared across these very diverse species, including *arx* - between cephalopods and annelids - as well as *mef2c,* common to both insect MB development, octopus vertical lobe development (Crittenden et al., 2018; Styfhals et al., 2022), and based on the previous chapter, also a marker for the annelid MBs. With further benchmarking by looking at more closely related - species, this approach may provide annotations for previously uncertain cell types. Importantly, it could also provide insights into the genes that are driving mixed - species clusters together, and whether these are terminal selector genes as proposed in Arendt et al. (2019), or effector genes involved in a cell's state or function. This was exemplified in the mixed species cluster containing *P.dumerilii* MB neurons and *H.saltator* unannotated glia, which seemed to be driven together by genes encoding ribosomal proteins. Bearing in mind the apparent neurogenic signature of *P.dumerilii* MBs in Chapter 2 *and* that the activation of aNSCs seems to be coupled to an increase in protein synthesis machinery (Baser et al., 2019; Gao et al., 2025; Llorens-Bobadilla et al., 2015), this result is not so surprising. Considering that the *H.saltator* dataset from Sheng et al. (2020) was exploring an ant caste transition from worker to queen, and found that the ensheathing - type glia increase during this transition, it is possible that glial progenitors exist in the dataset and that these end up clustered with the *P.dumerilii* neurogenic MB cells based on a shared increase in protein synthesis.

This brings up another caveat, as the evolutionary distance between compared species increases, determining comparable stages between them also becomes more difficult. In the present analysis, the octopus dataset was a brain dissection of the paralarval stage (Styfhals et al., 2022), in contrast to the adult annelid and ant stages used, which can influence which cell types are present or can be detected (Li et al., 2017). Species-specific peculiarities such as RNA editing mechanisms in cephalopods (Shoshan et al., 2021; Voss & Rosenthal, 2023) add further complexity when trying to interpret cell type similarities based on transcriptomic data alone, although detecting these edits with single - cell sequencing is an on - going endeavour (Xu et al., 2023). Nevertheless, it would still be worthwhile to compare the annelid brain to a basal chordate like amphioxus, since many of the transcriptomic signatures described in the previous chapter also resemble what is known from the vertebrate literature.

A better picture of ancestral and species-specific neuronal cell types will emerge as more species within the Lophotrocozoan phylum are investigated, in the same way that the brains of more and more vertebrate species are being compared using similar integration methods (Hain et al., 2022; Quintana-Urzainqui et al., 2025; Song et al., 2023; Woych et al., 2022). This way, the true diversity of bilaterian brains can be used to infer which neuronal cell types and families were likely present in a Bilaterian ancestor. Interestingly, the considerable diversity of MB - like structures among annelids alone, make it difficult to conclude whether the contributing cells can be considered the same type (Heuer et al., 2010), suggesting that already by comparing non - model annelids, a lot can be learned about whether these MB neurons are ancestral, at least within Annelida. While Chapter 4 will be a departure from this storyline, it will show how a species complex such as *Platynereis spp.* and the phylum of Annelida can offer insights into brain evolution at both the macro- and microevolutionary scale.

4. Lessons from the field and genotyping the *Platynereis spp.* species complex

4.1. Contributions

I tackled this thesis during a time where the lab was dominated by the TREC (Traversing European coastlines) expedition, part of EMBL's theme of Planetary Biology, and this chapter is a reflection of that. It was only possible thanks to the lab members who took on extra roles as the Selected Model Species (SMS) field team, especially Leslie Pan, Antonella Ruggiero, Emily Savage, Phil Oel, Tobias Gerber, Luca Santangeli, Cyril Cros, and Alvin Han who together were heavily involved in field sampling along with collaborators from the SMS team (Liz, Hanin, Jill and students). Additionally, Leslie Pan and Evgeniia Moschogianni came up with and optimized the initial in-field genotyping protocol. Along with joining 5 field expeditions, I have been part of the genotyping endeavour together with Alvin Han and Leslie Pan. Wet lab work was shared with Alvin throughout the process and Leslie provided constant input on experimental design and insights into sequencing technologies. I had no idea that fieldwork would be such a big part of my (and everyone's) time in the lab these past few years. This chapter is mostly a reflection of the shift in focus of the lab, and EMBL, towards studying model organisms in the context of their environment, and the associated challenges. While related to previous chapters, it is mostly a standalone dive into the adventure of fieldwork.

4.2. Background

As suggested in the previous chapters, the question of convergence or conservation is not so trivial to answer, especially when attempting to make inferences across very distant species. Broadening in depth investigations to diverse Lophotrocozoan species can only be beneficial in the quest to define and compare neuronal cell types on a transcriptomic basis. Annelida alone represents a diverse group of animals, and in depth anatomical investigations of their brains indicate that repeated loss of brain complexity - defined by clear sensory and integrative structures like eyes and mushroom bodies respectively - is something that has occurred multiple times in annelid evolution (Beckers et al., 2019; Helm et al., 2022; Heuer et al., 2010), and in some annelids is seen during the transition from a planktonic pelagic to a more sedentary benthic life stage (Beckers et al., 2023; Helm et al., 2022). This makes annelids a group of animals with diverse brain anatomies suited for investigating ancestral cell types, but also for linking changes in cell types or their expressed set of genes to an environmental context.

4.2.1. The annelid brain from a macro - and microevolutionary perspective

Among annelids, the mushroom bodies - with their dense nuclei and lobed axon outputs - are one indicator of brain complexity and are more commonly seen in errant annelids compared to sedentary - filter feeding ones (Heuer et al., 2010). However, while this bilateral mushroom body-like structure exists in many annelids, anatomical variability makes it difficult to know whether the contributing cells represent the same ancestral neuron type (Heuer et al., 2010). Comparing annelid brains at a transcriptomic level could tell us, on a *macroevolutionary* scale, if the MB cell types - and others - are shared despite differences in brain anatomy and organisation.

On the other hand, dramatic differences in lifestyle can appear even between morphologically similar sister species like *P.dumerilii* and *P. massiliensis*. As described in Chapter 2, *P.dumerilii* transitions between a pelagic planktonic stage, then grows in segment number during a tube - dwelling benthic phase, followed by a return to a pelagic stage for broadcast spawning. *P.massiliensis*, on the other hand, is a brooder that spends its whole life in a tube - dwelling phase, reflected in a loss of certain larval features like the larval eye and ciliary bands characteristic of trochophore larvae (Helm et al., 2015; Schneider et al., 1992). Interestingly, the brooder *P.massiliensis*, is more commonly found near acidic CO_2 vents, suggesting that this form of reproduction may be more protected (possibly by providing a protected microclimate or by the buffering of nearby O_2 - producing algae) and suitable for harsh environments (Calosi et al., 2013; Lucey et al., 2015; Wäge et al., 2017). Therefore, comparing members even within this species complex can tell us, on a *microevolutionary* scale, how the environment might be shaping the brains of nearly indistinguishable annelids at shorter timescales.

Chapter 2 showed that single nucleus RNA sequencing can successfully classify cell types of the annelid brain from a small number of samples, while remaining comparable to single *cell* datasets of similar resultant cell number but more input samples (Milivojev et al., 2024). This utility means that, for species not easily cultured in the lab, or for precious samples collected from the field, a more or less complete cell type complement is still achievable with this approach. As more annelid genomes are generated and annotated - with projects such as the Darwin Tree of Life (https://www.darwintreeoflife.org/) and the additional Platynereis genomes (i.e. *P.massiliensis, P.megalops*) sequenced by Mutemi et al. (2025), the goal of producing such comparative atlases for unconventional model species is becoming more attainable. However, due to the striking morphological similarity between members of, particularly the *Platynereis spp.* species complex, genotyping remains a necessary first step to characterise these samples before committing to expensive single cell experiments.

4.2.2. Classifying samples from the TREC expedition

The members of the *Platynereis spp.* species complex far exceeds the two sister species described above. In general, because of the physical isolation of brooding species, they tend to mix less and show more nucleotide diversity (Wäge et al., 2017). Work by Teixeira et al. (2022) has classified members of this "hyper - diverse" species complex into MOTUs (Molecular Operational Taxonomic Units) based on mtCO1 (~658bp), *16s* rDNA (~368bp), and *28s* rDNA (~420bp) sequences. Although these MOTUs are only molecular definitions of sister species, they are the categories that will be used to classify the collected expedition samples. The EMBL 2023 - 2024 TREC expedition (Traversing European Coastlines) provided an opportunity to collect a diversity of annelids - primarily from the *Platynereis spp.* species complex - along with metadata relating to anthropogenic activities, bioacoustics, microbial communities, and co-habitating species (e.g. seagrasses, anemones, sponges) to contextualise future findings. Sampling missions took place at 46 marine institutions, with sites yielding suspected members of the *Platynereis spp.* species complex indicated in **Figure 28A**. These annelids were collected from both coastal seagrass (*Posedonia oceanica*, *Zostera marina*), algae, and kelp. Mature animals were collected from spawning events during a parallel expedition following the moon cycle. The latter expedition will not be discussed in detail, although some reproductive worms from it were included in the genotyping development process.

This chapter will describe some of the field sampling that took place over the course of this expedition and the initial genotyping efforts to categorise this diverse species complex. Focus will be on the mitochondrial *cytochrome c oxidase subunit I* gene (CO1), a ~700bp fragment established for classifying diverse invertebrates (Folmer et al., 1994), with broader taxonomic classification being achieved upon modifications of the original primers (Elbrecht & Leese, 2017; Leray et al., 2013). For the present chapter, which will only explore collected polychaete samples, polychaete - specific CO1 primers (Carr et al., 2011; Teixeira et al., 2022)

proved to be successful in amplifying this CO1 fragment from members of the *Platynereis spp.* species complex as well as other non - *Platynereis* nereids. Combining these primers with internal CO1 primers from Leray et al. (2013) to produce shorter amplicons was also successful, setting the stage for future Next Generation Sequencing (NGS) approaches (Kennedy et al., 2020).

Figure 28. TREC sampling scheme and sample processing. (**A**) Map showing major sampling locations. Stars indicate sites where *Platynereis spp*. complex - like annelids were found. (**B**) Snapshots of collected substrate from the field. (**C**) Workflow of sample preparation - specifically for annelid samples. Adapted from figures by Leslie Pan. Icons from BioRender.com.

4.3. Materials and methods

4.3.1. Annelid field collection

Between 2023 and 2024 annelids were sampled in combination with other species during the EMBL TREC expedition. This would not have been possible without the support of local marine stations and divers as well as the TREC team - especially the SMS team. **Figure 28A** shows a map depicting the collection sites visited. Of these sites, worms morphologically characterised as belonging to the *Platynereis spp.* species complex were collected at 12 of them. The preliminary results from this chapter cover a subset of samples from Crete, Trieste, Naples, Ferrol, and Bilbao. Sampling was conducted in close collaboration with local divers who collected algae, seagrass, and kelp for the team to sort for annelids **Figure 28B.** Small samples of substrate were flash frozen for later identification. The sampling scheme and downstream processing for genotyping is shown in **Figure 28C**. Briefly, small annelids were placed directly into 200µl of DNA/RNA shield (Zymo, R1100-250) for storage. If enough specimens of a type and of appropriate size (~2cm) were collected, each worm was dissected into three pieces. First, and relevant for this chapter, the tail was cut and placed into 200µl of DNA/RNA shield. The head was removed above the pharynx when possible and flash frozen in liquid nitrogen for future comparative snRNAseq experiments. The trunk was also flash frozen for future microbiome studies. The specimens used for this optimization were chosen from diverse geographical locations known to harbour members of the *Platynereis spp.* species complex (Crete, Trieste, Naples, Ferrol, and Bilbao). Sequencing results were checked by aligning to known CO1 sequences, listed in **Table S3,** Appendix.

4.3.2. Genotyping

Back in the lab, 5 - 10 zirconia beads (ROTH, N039.1) were added to each tail sample for tissue homogenization. Using a Omni Bead Ruptor Elite bead mill homogenizer (Revvity), samples were homogenized for two 30s cycles at 2m/s. For a small number of samples a

hand-held pestle was used. Maintaining sample order, 50µl of the homogenate was transferred to 96 - well plates for easier pipetting. To extract DNA, 2µl of each homogenised sample was added to 48µl of QuickExtract™ DNA extraction solution (Lucigen), already prepared in a new 96 - well plate. Samples were mixed, centrifuged, then incubated in a thermocycler at 65 °C for 6 minutes then 98 °C for 2 minutes.

The ~700bp mitochondrial CO1 fragment was amplified using *polychaete* - specific LCOforward and HCOreverse primers - polyLCO, polyHCO (Carr et al., 2011; Teixeira et al., 2022) - each at a concentration of 10µM. Primer sequences are listed in the Appendix (**Table S4**). Each reaction was set up according to **Table 4**, and amplified with the cycling conditions in **Table 5**, which successfully amplified the CO1 regions of samples collected from diverse locations and consisting of specimens from within and outside the *Platynereis spp.* complex. The low annealing temperature was selected after testing a variety of temperatures with gradient PCR.

Table 4. Reaction set up using Q5® High-Fidelity 2X Master Mix (NEB, M0492S) or NEBNext High-fidelity 2X PCR Master Mix (NEB, M0541S).

Component	1 reaction (25µl)	1 reaction (50µl)	Final Concentration
Q5 2x High-fidelity master mix	12.5µl	25µl	1x
10µM fwd primer	2.5µl	5µl	1µM
10µM rev primer	2.5µl	5µl	1µM
DNA Template	1µl / sample	2µl / sample	<1000 ng
dH2O	6.5µl	13µl	up to final volume
Total Volume	25µl	50µl	-

Table 5. Cycling conditions for CO1 amplification able to amplify *Platynereis spp.* and non - *Platynereis spp.* CO1 regions.

Step	Temp	Time	# cycles
Initial denaturation	98°C	30s	1
Denaturation	98°C	10s	35
Annealing	**45°C**	30s	
Extension	72°C	20s	
Final extension	72°C	2min	1
Hold	4 °C	-	-

After checking for successful amplification on a standard agarose gel, or a 2% e-gel (Thermo, A42346), PCR products were cleaned using left - side size selection with SPRIselect magnetic beads (Beckman Coulter, B23318) according to the manufacturer protocol. Briefly, 0.65x the reaction volume of SPRIselect beads was added to each reaction, and mixed with a pipette 15 times. This solution was incubated for 5 minutes then placed on a magnetic rack in order to remove the supernatant. Samples were washed twice with 200µl of fresh 80% EtOH, incubating for 30 seconds each time. Tubes were spun down and placed back on the magnetic rack. Excess EtOH was removed, and samples were left to air dry for no longer than 2 minutes. To elute, tubes were removed from the magnetic rack and 15 - 50µl of nuclease - free water was added, mixed with a pipette and left to incubate for 5 minutes. After being placed back on a magnetic rack, eluant was transferred to a fresh 96 - well plate to check concentration and send for sequencing.

Staggered primers - meaning primers with added spacers to mitigate issues of low diversity of bases (Wu et al., 2015) - were also tested for the two halves of the CO1 region (3′ part - 319bp, 5′ part - 313bp) to keep amplicons small enough for future NGS applications (Leray et al., 2013). Reactions were set up using the PCR product from the full length CO1

amplification, using staggered versions of the polyLCO/HCO primers combined with staggered internal CO1 primers from Leray et al. (2013). Staggered primers are listed in the Appendix (**Table S4**); with 8 primers (4 forward, 4 reverse) targeting the 3′ half of the CO1 region, and 8 primers (4 forward, 4 reverse) targeting the 5′ half. Primers were mixed in equimolar amounts to a final total concentration of 10 µM. Reactions were set up as in **Table 4**, and amplified under the cycling conditions in **Table 6**.

Table 6. 2 - step PCR cycling conditions for amplification of two halves of the CO1 region using staggered primers.

Step	Temp	Time	# cycles
Initial denaturation	98°C	30s	1
Denaturation	98°C	10s	10
Annealing	54°C	30s	
Extension	72°C	20s	
Denaturation	98°C	10s	20
Annealing/extension	72°C	50s	
Final extension	72°C	2min	1
Hold	4°C	-	-

The PCR products were cleaned in the same way as above, except by adding a **1x** volume of SPRIselect beads due to the smaller amplicon sizes. **Figure 29** shows the successful amplification using these staggered primers.

For the purposes of this thesis, though, only the initial CO1 product from the first PCR of a few selected samples was sent for Sanger sequencing and will be discussed.

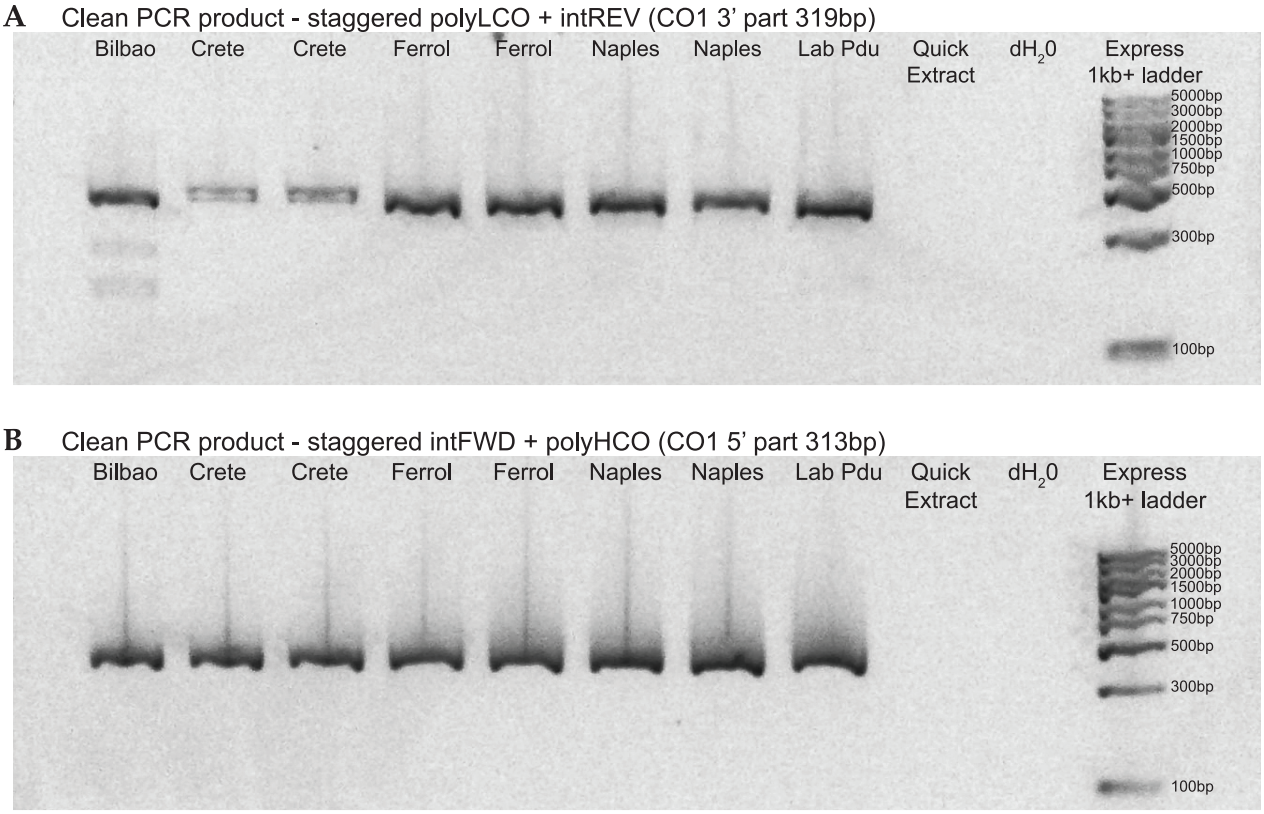


Figure 29. Amplification with staggered external and internal CO1 primers. Cleaned PCR products run on 2% e- gel. (**A**) 3′ CO1 fragment (319bp) across tail samples selected for optimisation. (**B**) 5′ CO1 fragment (313bp) across tail samples selected for optimisation.

4.4. Results

The Sanger sequencing results of the test samples were aligned to the CO1 sequences from Teixeira et al. (2022), using Geneious Prime ® 2024.0.5, with the built in Geneious alignment tool and default parameters, followed by the built in neighbour - joining tree tool with default parameters. **Figure 30** shows a summary figure of the diversity of MOTUs - described in Teixeira et al. (2022) - to which the tested expedition samples best aligned. All sequences prefixed with *TREC* come from the expedition, while selected annotated reference sequences are those from Teixeira et al. (2022). The full alignment and neighbour - joining tree showing the relationship between the MOTUs can be seen in **Figure S5**.

Samples belonging to the *Platynereis spp.* species complex, as well as non *Platynereis* nereidids (Pseudonereis, MOTU 11) were successfully characterised. With only a few of the collected polychaetes sequenced so far, it is clear there is a diversity of MOTUs even within a single sampling location in line with findings from Teixeira et al. (2022). A clear next step would be to link this diversity to TREC metadata (e.g. bioacoustics, algal substrate, anthropogenic impact), to better understand if the micro-environments that were sampled correlate with the species present. Reassuringly, one of the Naples samples along with a lab positive control sample (**Figure 30**, highlighted) both fell within MOTU 4 - since most *P.dumerilii* in lab culture are thought to originate from Naples (Hauenschild & Fischer, 1969). Broadly, the geographical locations for the TREC samples correspond to those where samples of the same MOTU were reported by Teixeira et al. (2022), with the exception of the Ferrol sample that aligned to MOTU 6, which was not reported in the Atlantic. However, the alignment shows many divergent SNPS compared to the reference MOTU 6 sequences (**Figure 30**), suggesting a possible technical error which will be clarified as more sequences are added.

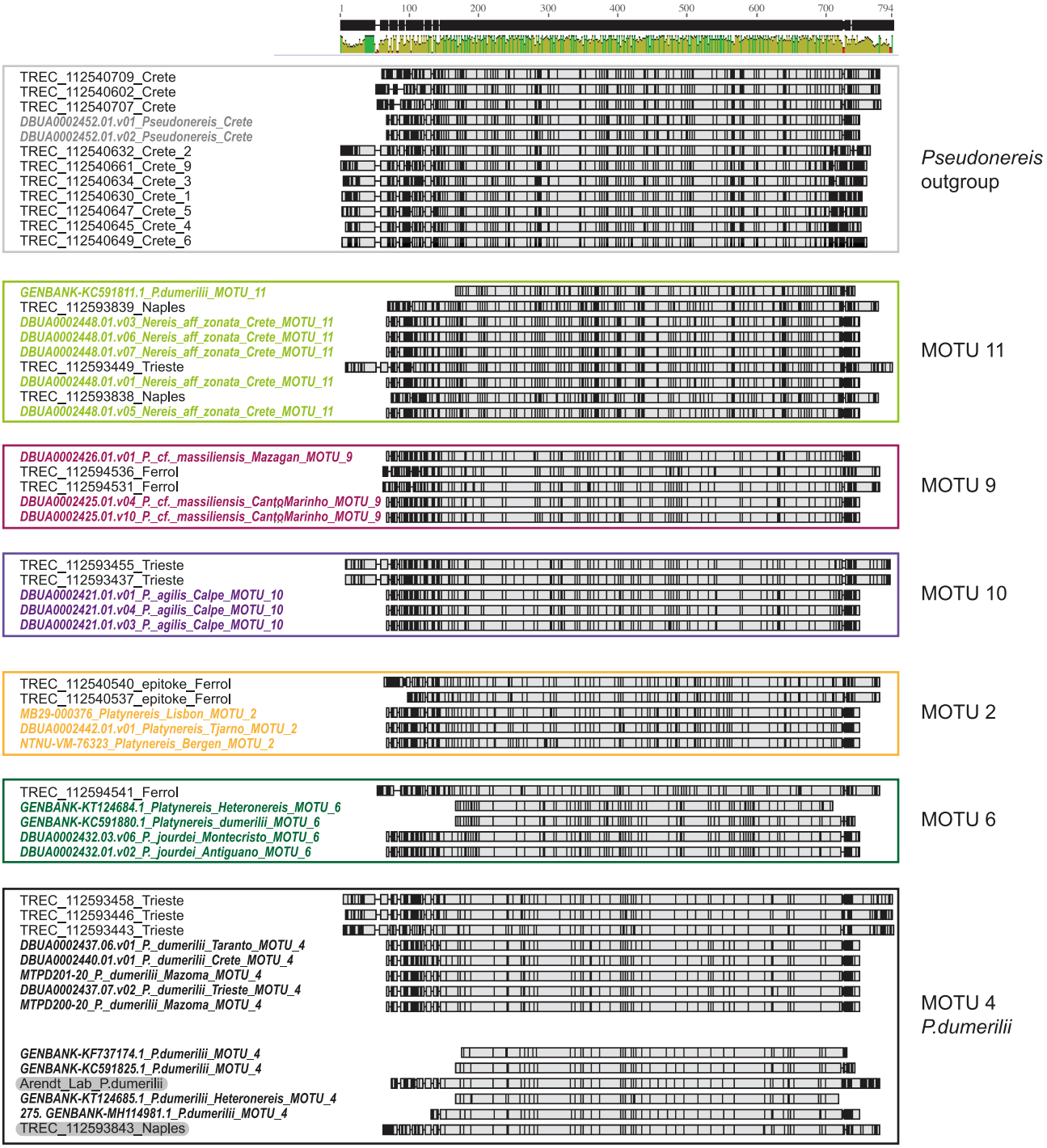


Figure 30. Multiple sequence alignment snapshots for the selection of tested polychaete samples. Expedition samples are formatted as "TREC_barcode_site" and are embedded with the most closely aligned reference CO1 sequences from Teixeira et al. (2022). Sequences are grouped by MOTU (Molecular Operational Taxonomic Units) classifications, based on the full neighbour - joining tree for this multiple sequence alignment, shown in **Figure S5**.

4.5. Summary and perspectives

This chapter showed that classic Sanger sequencing of only the CO1 region of samples collected on the TREC expedition was able to categorise polychaetes of the *Platynereis spp.* complex, as well as morphologically similar polychaetes collected from diverse locations visited during the TREC expedition. Of the TREC samples, 282 have been categorised morphologically as belonging to the *Platynereis spp.* complex and remain to be sequenced. Therefore an aim for the immediate future is to further increase the scalability of CO1 barcoding (Kennedy et al., 2020), in particular because it would be prudent to include the other primers - 16s rDNA (~368bp), and 28s rDNA (~420bp) - used to define MOTUs in Teixeira et al. (2022).

In the end, linking these findings to the associated metadata collected on the expedition as well as on - going microbiome work in the lab will help correlate the collected species diversity with an environmental context. So far from Trieste - on the Adriatic sea - both premature broadcast spawning (MOTU 4) specimens, as well as a subset of brooders (MOTU 10) were collected from the same geographical location (45.726179, 13.689958). Linking these results to the specific substrate from which worms were collected is an interesting next step, to see if both MOTUs lived on similar or different algal substrates. Contrast this to Ferrol, where the yet to be assigned MOTU 2 - which now can be more confidently classified as a broadcast spawner - was *only* found during the parallel mission collecting actively spawning mature worms. This MOTU was not among the tube - dwelling annelids collected from the substrate of the same location, where only the brooding *P.massiliensis* (MOTU 9) was detected. Though the number of samples tested so far is small, it will be interesting to see whether Atlantic broadcast spawners versus brooders are found on different substrates or depths compared to the Mediterranean sites, since the tides of the Atlantic might be another harsh environment that may favour brooding over broadcast spawning, specifically for worms living in more shallow intertidal zones.

This is only the beginning of this large multi-disciplinary project, with the focus over these past two years being on sample collection and characterisation. Nevertheless, the aim of this chapter was to emphasize the challenges in characterizing members of this diverse species complex, and that this complexity should not hinder future transcriptomic work - in this case for the characterisation of cell types - but should be considered when interpreting findings.

5. Conclusion

5.1. Insights into the maturing annelid mushroom bodies

By using single cell transcriptomics to define and expand the transcriptomic signature of *P.dumerilii* mushroom body neurons, I was also able to uncover new evolutionary insights about the interneurons of this region. From a species - specific perspective, besides the transcription factor markers known from larval stages - *arx, ptf1* (Vergara et al., 2021), the ventral part of the maturing annelid MB interneurons also expressed a possible paralog of *emx* - *emx2* -, a gene having undergone multiple duplication events in different lineages (Minguillón et al., 2002; Noro et al., 2015). This was paired with the stage - specific expression of a vertebrate - like opsin (Arendt et al., 2004) - *cOpsin1* - , which thus far, for adult worms, has only been reported in brain ciliary PRCs and progenitor cells of the growing rhabdomeric eyes (Milivojev et al., 2024). This expression pattern also adds further evidence supporting the ancestral nature of sensory interneurons (Fischer et al., 2013; Wulf et al., 2025). The functional implications of this opsin in *P.dumerilli* MBs might be an exit from a maintained neurogenic state, since the expression of *cOpsin1* seems to coincide with downregulation of *arx* and *emx2*, and the upregulation of *mef2c* and possible bzip transcription factors with known roles in vertebrate adult neurogenesis and plasticity (Cortes-Canteli et al., 2011; Herold et al., 2011; Hodge & Hevner, 2011) and apoptosis (Bodnarchuk et al., 2012; Valderrama et al., 2009). This seems to be accompanied by a broad increase in *soxC* expression, another vertebrate ortholog known for its role in neuronal differentiation in early *P.dumerilii* development (Kerner et al., 2009). Overall, this suggests that the neuronal restructuring in the adult *P.dumerilii* brain might use mechanisms mirroring adult neurogenesis in vertebrates under the influence - either directly or indirectly - of a ciliary opsin. From a broader comparative perspective, however tentative, it appears the overall MB transcriptomic signature bears similarities to cephalopod vertical lobe

neurons (Styfhals et al., 2022) and insect MBs (Sheng et al., 2020), with the *mef2c* orthology group as one common link - not yet described for *P.dumerilii* MBs. Although in the present analysis it is not clear whether these genes are actually driving cross - species clustering in the method used. Exploring whether this MB transcriptomic signature exists among other members of the more closely related *Platynereis spp*. species complex, as well as in more basal annelids, or annelids without a clear MB structure will be important to better understand if the evolutionary inferences being drawn indeed reflect an ancestral Bilaterian cell type.

5.2. Future perspectives

While Chapter 4 was a departure, it emphasised that for such transcriptomic comparisons to work they must extend to a broad array of species, of various evolutionary distances. A first proof of principle for any cross - species comparison for annelids, might be *P.dumerilii* and its sister species *P.massiliensis*, a brooder rather than a broadcast spawner, adapted to living in harsher environments (Lucey et al., 2015; Wäge et al., 2017). Exemplified in Chapter 4 and highlighted by Teixeira et al. (2022), classifying members of this species complex is not so trivial, for instance the reassigned *P.agilis*, is very similar in CO1 sequence to the brooder *P.massiliensis*, but was classified as a separate species based on morphological differences (e.g. less pigmentation, longer cirri). In fact, tube - dwelling worms in the Arendt lab, of unknown origin but labelled as *P.massiliensis*, have since been classified as *P.agilis* based on CO1 sequence, a distinction only possible with the detailed work of Teixeira et al. (2022). Further classification of the TREC samples collected from intertidal, subtidal, and from active spawning events, will provide the environmental context necessary for future deeper transcriptomic comparisons. I hope this thesis highlights the layers of complexity in tackling both macro- and microevolutionary questions, but the necessity for both levels in order to make evolutionary inferences. I also hope to highlight that, with the plethora of published datasets that exist - there is so much information contained within that remains to be discovered.

Bibliography

Achim, K., Eling, N., Vergara, H. M., Bertucci, P. Y., Musser, J., Vopalensky, P., Brunet, T., Collier, P., Benes, V., Marioni, J. C., & Arendt, D. (2018). Whole-Body Single-Cell Sequencing Reveals Transcriptional Domains in the Annelid Larval Body. *Molecular Biology and Evolution*, *35*(5), 1047–1062. https://doi.org/10.1093/molbev/msx336

Achim, K., Pettit, J.-B., Saraiva, L. R., Gavriouchkina, D., Larsson, T., Arendt, D., & Marioni, J. C. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nature Biotechnology*, *33*(5), 503–509. https://doi.org/10.1038/nbt.3209

Ahmed, N. Y., Knowles, R., & Dehorter, N. (2019). New Insights Into Cholinergic Neuron Diversity. *Frontiers in Molecular Neuroscience*, *12*. https://doi.org/10.3389/fnmol.2019.00204

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andreatta, G., Broyart, C., Borghgraef, C., Vadiwala, K., Kozin, V., Polo, A., Bileck, A., Beets, I., Schoofs, L., Gerner, C., & Raible, F. (2020). Corazonin signaling integrates energy homeostasis and lunar phase to regulate aspects of growth and sexual maturation in Platynereis. *Proceedings of the National Academy of Sciences*, *117*(2), 1097–1106. https://doi.org/10.1073/pnas.1910262116

Arboleda, E., Zurl, M., Waldherr, M., & Tessmar-Raible, K. (2019). Differential Impacts of the Head on Platynereis dumerilii Peripheral Circadian Rhythms. *Frontiers in Physiology*, *10*, 900. https://doi.org/10.3389/fphys.2019.00900

Arendt, D., Bertucci, P. Y., Achim, K., & Musser, J. M. (2019). Evolution of neuronal types and families. *Current Opinion in Neurobiology*, *56*, 144–152. https://doi.org/10.1016/j.conb.2019.01.022

Arendt, D., Musser, J. M., Baker, C. V. H., Bergman, A., Cepko, C., Erwin, D. H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M. D., & Wagner, G. P. (2016). The origin and evolution of cell types. *Nature Reviews Genetics*, *17*(12), 744–757. https://doi.org/10.1038/nrg.2016.127

Arendt, D., Tessmar-Raible, K., Snyman, H., Dorresteijn, A. W., & Wittbrodt, J. (2004). *Ciliary Photoreceptors with a Vertebrate-Type Opsin in an Invertebrate Brain*. *306*.

Arendt, D., Urzainqui, I. Q., & Vergara, H. M. (2021). The conserved core of the nereid brain: Circular CNS, apical nervous system and lhx6-arx-dlx neurons. *Current Opinion in Neurobiology*, *71*, 178–187. https://doi.org/10.1016/j.conb.2021.11.008

Ayers, T., Tsukamoto, H., Gühmann, M., Veedin Rajan, V. B., & Tessmar-Raible, K. (2018). A Go-type opsin mediates the shadow reflex in the annelid Platynereis dumerilii. *BMC Biology*, *16*(1), 41. https://doi.org/10.1186/s12915-018-0505-8

Backfisch, B., Veedin Rajan, V. B., Fischer, R. M., Lohs, C., Arboleda, E., Tessmar-Raible, K., & Raible, F. (2013). Stable transgenesis in the marine annelid *Platynereis dumerilii* sheds new light on photoreceptor evolution. *Proceedings of the National Academy of*

Sciences, *110*(1), 193–198. https://doi.org/10.1073/pnas.1209657109

Barrow, A. J., & Wu, S. M. (2009). Low-Conductance HCN1 Ion Channels Augment the Frequency Response of Rod and Cone Photoreceptors. *The Journal of Neuroscience*, *29*(18), 5841–5853. https://doi.org/10.1523/JNEUROSCI.5746-08.2009

Baser, A., Skabkin, M., Kleber, S., Dang, Y., Gülcüler Balta, G. S., Kalamakis, G., Göpferich, M., Ibañez, D. C., Schefzik, R., Lopez, A. S., Bobadilla, E. L., Schultz, C., Fischer, B., & Martin-Villalba, A. (2019). Onset of differentiation is post-transcriptionally controlled in adult neural stem cells. *Nature*, *566*(7742), 100–104. https://doi.org/10.1038/s41586-019-0888-x

Basu, S., Patil, S., Mapleson, D., Russo, M. T., Vitale, L., Fevola, C., Maumus, F., Casotti, R., Mock, T., Caccamo, M., Montresor, M., Sanges, R., & Ferrante, M. I. (2017). Finding a partner in the ocean: Molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. *New Phytologist*, *215*(1), 140–156. https://doi.org/10.1111/nph.14557

Beckers, P., Gebhardt, T., & Helm, C. (2023). Loss of nervous system complexity – Morphological analyses shed light on the neuronal evolution in Myzostomida (Annelida). *Acta Zoologica*, *104*(4), 633–646. https://doi.org/10.1111/azo.12447

Beckers, P., Helm, C., Purschke, G., Worsaae, K., Hutchings, P., & Bartolomaeus, T. (2019). The central nervous system of Oweniidae (Annelida) and its implications for the structure of the ancestral annelid brain. *Frontiers in Zoology*, *16*(1), 6. https://doi.org/10.1186/s12983-019-0305-1

Bender, R. A., & Baram, T. Z. (2008). Hyperpolarization activated cyclic-nucleotide gated (HCN) channels in developing neuronal networks. *Progress in Neurobiology*, *86*(3), 129–140. https://doi.org/10.1016/j.pneurobio.2008.09.007

Bidel, F., Meirovitch, Y., Schalek, R. L., Lu, X., Pavarino, E. C., Yang, F., Peleg, A., Wu, Y., Shomrat, T., Berger, D. R., Shaked, A., Lichtman, J. W., & Hochner, B. (2023). Connectomics of the Octopus vulgaris vertical lobe provides insight into conserved and novel principles of a memory acquisition network. *eLife*, *12*, e84257. https://doi.org/10.7554/eLife.84257

Biel, M., Wahl-Schott, C., Michalakis, S., & Zong, X. (2009). Hyperpolarization-Activated Cation Channels: From Genes to Function. *Physiological Reviews*, *89*(3), 847–885. https://doi.org/10.1152/physrev.00029.2008

Bodnarchuk, T. W., Napper, S., Rapin, N., & Misra, V. (2012). Mechanism for the induction of cell death in ONS-76 medulloblastoma cells by Zhangfei/CREB-ZF. *Journal of Neuro-Oncology*, *109*(3), 485–501. https://doi.org/10.1007/s11060-012-0927-z

Burgoyne, R. D. (2007). Neuronal calcium sensor proteins: Generating diversity in neuronal Ca2+ signalling. *Nature Reviews Neuroscience*, *8*(3), 182–193. https://doi.org/10.1038/nrn2093

Calosi, P., Rastrick, S. P. S., Lombardi, C., de Guzman, H. J., Davidson, L., Jahnke, M., Giangrande, A., Hardege, J. D., Schulze, A., Spicer, J. I., & Gambi, M.-C. (2013). Adaptation and acclimatization to ocean acidification in marine ectotherms: An in

situ transplant experiment with polychaetes at a shallow CO2 vent system. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1627), 20120444. https://doi.org/10.1098/rstb.2012.0444

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, *38*(12), 5825–5829. https://doi.org/10.1093/molbev/msab293

Carr, C. M., Hardy, S. M., Brown, T. M., Macdonald, T. A., & Hebert, P. D. N. (2011). A Tri-Oceanic Perspective: DNA Barcoding Reveals Geographic Structure and Cryptic Diversity in Canadian Polychaetes. *PLOS ONE*, *6*(7), e22232. https://doi.org/10.1371/journal.pone.0022232

Chartier, T. F., Deschamps, J., Dürichen, W., Jékely, G., & Arendt, D. (2018). Whole-head recording of chemosensory activity in the marine annelid Platynereis dumerilii. *Open Biology*, *8*(10), 180139. https://doi.org/10.1098/rsob.180139

Choi, H. M. T., Schwarzkopf, M., Fornace, M. E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., & Pierce, N. A. (2018). Third-generation in situ hybridization chain reaction: Multiplexed, quantitative, sensitive, versatile, robust. *Development*, *145*(12), dev165753. https://doi.org/10.1242/dev.165753

Conzelmann, M., Williams, E. A., Krug, K., Franz-Wachtel, M., Macek, B., & Jékely, G. (2013). The neuropeptide complement of the marine annelid Platynereis dumerilii. *BMC Genomics*, *14*, 906. https://doi.org/10.1186/1471-2164-14-906

Ćorić, A., Stockinger, A., Schaffer, P., Roković, D., Tessmar-Raible, K., & Raible, F. (2023). A Fast And Versatile Method for Simultaneous HCR, Immunohistochemistry And Edu Labeling (SHInE). *Integrative and Comparative Biology*, *63*. https://doi.org/10.1093/icb/icad007

Cortes-Canteli, M., Aguilar-Morante, D., Sanz-Sancristobal, M., Megias, D., Santos, A., & Perez-Castillo, A. (2011). Role of C/EBPβ transcription factor in adult hippocampal neurogenesis. *PloS One*, *6*(10), e24842. https://doi.org/10.1371/journal.pone.0024842

Crittenden, J. R., Skoulakis, E. M. C., Goldstein, E. S., & Davis, R. L. (2018). Drosophila mef2 is essential for normal mushroom body and wing development. *Biology Open*, *7*(9), bio035618. https://doi.org/10.1242/bio.035618

Dahlitz, I., Dorresteijn, A., & Holz, A. (2023). Remodeling of the Platynereis Musculature during Sexual Maturation. *Biology*, *12*(2), 254. https://doi.org/10.3390/biology12020254

Denes, A. S., Jékely, G., Steinmetz, P. R. H., Raible, F., Snyman, H., Prud'homme, B., Ferrier, D. E. K., Balavoine, G., & Arendt, D. (2007). Molecular architecture of annelid nerve cord supports common origin of nervous system centralization in bilateria. *Cell*, *129*(2), 277–288. https://doi.org/10.1016/j.cell.2007.02.040

Elbrecht, V., & Leese, F. (2017). Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. *Frontiers in Environmental Science*, *5*. https://doi.org/10.3389/fenvs.2017.00011

Fischer, A., & Dorresteijn, A. (2004). The polychaetePlatynereis dumerilii (Annelida): A

laboratory animal with spiralian cleavage, lifelong segment proliferation and a mixed benthic/pelagic life cycle. *BioEssays*, *26*(3), 314–325. https://doi.org/10.1002/bies.10409

Fischer, A. H., Henrich, T., & Arendt, D. (2010). The normal development of Platynereis dumerilii (Nereididae, Annelida). *Frontiers in Zoology*, *7*(1), 31. https://doi.org/10.1186/1742-9994-7-31

Fischer, R. M., Fontinha, B. M., Kirchmaier, S., Steger, J., Bloch, S., Inoue, D., Panda, S., Rumpel, S., & Tessmar-Raible, K. (2013). Co-Expression of VAL- and TMT-Opsins Uncovers Ancient Photosensory Interneurons and Motorneurons in the Vertebrate Brain. *PLOS Biology*, *11*(6), e1001585. https://doi.org/10.1371/journal.pbio.1001585

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). *DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates*. *3*(5).

Fujiyama, T., Miyashita, S., Tsuneoka, Y., Kanemaru, K., Kakizaki, M., Kanno, S., Ishikawa, Y., Yamashita, M., Owa, T., Nagaoka, M., Kawaguchi, Y., Yanagawa, Y., Magnuson, M. A., Muratani, M., Shibuya, A., Nabeshima, Y., Yanagisawa, M., Funato, H., & Hoshino, M. (2018). Forebrain Ptf1a Is Required for Sexual Differentiation of the Brain. *Cell Reports*, *24*(1), 79–94. https://doi.org/10.1016/j.celrep.2018.06.010

Ganguly, A., Qi, C., Bajaj, J., & Lee, D. (2020). Serotonin receptor 5-HT7 in Drosophila mushroom body neurons mediates larval appetitive olfactory learning. *Scientific Reports*, *10*(1), 21267. https://doi.org/10.1038/s41598-020-77910-5

Gao, Y., Guo, L., Shi, G., Wang, R., Wang, X., & Lou, J. (2025). Emerging roles of ribosome translation in stem cells and stem cell therapy—A review. *Cell & Bioscience*, *15*(1), 71. https://doi.org/10.1186/s13578-025-01412-y

Glasgow, S. M., Henke, R. M., MacDonald, R. J., Wright, C. V. E., & Johnson, J. E. (2005). Ptf1a determines GABAergic over glutamatergic neuronal cell fate in the spinal cord dorsal horn. *Development*, *132*(24), 5461–5469. https://doi.org/10.1242/dev.02167

Goodman, R. H., & Smolik, S. (2000). CBP/p300 in cell growth, transformation, and development. *Genes & Development*, *14*(13), 1553–1577. https://doi.org/10.1101/gad.14.13.1553

Häfker, N. S., Holcik, L., Mat, A. M., Ćorić, A., Vadiwala, K., Beets, I., Stockinger, A. W., Atria, C. E., Hammer, S., Revilla-i-Domingo, R., Schoofs, L., Raible, F., & Tessmar-Raible, K. (2024). Molecular circadian rhythms are robust in marine annelids lacking rhythmic behavior. *PLOS Biology*, *22*(4), e3002572. https://doi.org/10.1371/journal.pbio.3002572

Hain, D., Gallego-Flores, T., Klinkmann, M., Macias, A., Ciirdaeva, E., Arends, A., Thum, C., Tushev, G., Kretschmer, F., Tosches, M. A., & Laurent, G. (2022). Molecular diversity and evolution of neuron types in the amniote brain. *Science*, *377*(6610), eabp8202. https://doi.org/10.1126/science.abp8202

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., … Satija, R. (2021).

Integrated analysis of multimodal single-cell data. *Cell*, *184*(13), 3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048

Hao, Y., Stuart, T., Kowalski, M. H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., & Satija, R. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology*, *42*(2), 293–304. https://doi.org/10.1038/s41587-023-01767-y

Hauenschild, C. (1956). Hormonale Hemmung der Geschlechtsreife und Metamorphose bei dem Polychaeten Platynereis Dumerilii. *Zeitschrift für Naturforschung B*, *11*(3), 125–132. https://doi.org/10.1515/znb-1956-0302

Hauenschild, C. (1960). Lunar Periodicity. *Cold Spring Harbor Symposia on Quantitative Biology*, *25*, 491–497. https://doi.org/10.1101/SQB.1960.025.01.051

Hauenschild, C. (1966). Der hormonale einfluss des Gehirns auf die sexuelle Entwicklung bei dem polychaeten *Platynereis dumerilii*. *General and Comparative Endocrinology*, *6*(1), 26–73. https://doi.org/10.1016/0016-6480(66)90108-0

Hauenschild, C. (1974). Normalisierung der geschlechtlichen Entwicklung kopfloser Fragmente junger ♀♀ vonPlatynereis dumerilii (Polychaeta) durch Behandlung mit konservierten Prostomien juveniler Individuen. *Helgoländer wissenschaftliche Meeresuntersuchungen*, *26*(1), 63–81. https://doi.org/10.1007/BF01613305

Helm, C., Adamo, H., Hourdez, S., & Bleidorn, C. (2015). An immunocytochemical window into the development of <I>Platynereis massiliensis</I> (Annelida, Nereididae). *The International Journal of Developmental Biology*, *58*(6-7–8), Article 6-7–8. https://doi.org/10.1387/ijdb.140081cb

Helm, C., Schwarze, G., & Beckers, P. (2022). Loss of complexity from larval towards adult nervous systems in Chaetopteridae (Chaetopteriformia, Annelida) unveils evolutionary patterns in Annelida. *Organisms Diversity & Evolution*, *22*(3), 631–647. https://doi.org/10.1007/s13127-022-00553-z

Herold, S., Jagasia, R., Merz, K., Wassmer, K., & Lie, D. C. (2011). CREB signalling regulates early survival, neuronal gene expression and morphological development in adult subventricular zone neurogenesis. *Molecular and Cellular Neuroscience*, *46*(1), 79–88. https://doi.org/10.1016/j.mcn.2010.08.008

Heuer, C. M., Kollmann, M., Binzer, M., & Schachtner, J. (2012). Neuropeptides in insect mushroom bodies. *Arthropod Structure & Development*, *41*(3), 199–226. https://doi.org/10.1016/j.asd.2012.02.005

Heuer, C. M., Müller, C. H., Todt, C., & Loesel, R. (2010). Comparative neuroanatomy suggests repeated reduction of neuroarchitectural complexity in Annelida. *Frontiers in Zoology*, *7*(1), 13. https://doi.org/10.1186/1742-9994-7-13

Hobert, O. (2016). Terminal Selectors of Neuronal Identity. In *Current Topics in Developmental Biology* (Vol. 116, pp. 455–475). Elsevier. https://doi.org/10.1016/bs.ctdb.2015.12.007

Hodge, R. D., & Hevner, R. F. (2011). Expression and actions of transcription factors in adult hippocampal neurogenesis. *Developmental Neurobiology*, *71*(8), 680–689. https://doi.org/10.1002/dneu.20882

Hong, S. M., Liu, Z., Fan, Y., Neumann, M., Won, S. J., Lac, D., Lum, X., Weinstein, P. R., & Liu, J. (2007). Reduced hippocampal neurogenesis and skill reaching performance in adult *Emx1* mutant mice. *Experimental Neurology*, *206*(1), 24–32. https://doi.org/10.1016/j.expneurol.2007.03.028

Horio, N., & Liberles, S. D. (2021). Hunger enhances food-odour attraction through a neuropeptide Y spotlight. *Nature*, *592*(7853), 262–266. https://doi.org/10.1038/s41586-021-03299-4

Hoshino, M., Nakamura, S., Mori, K., Kawauchi, T., Terao, M., Nishimura, Y. V., Fukuda, A., Fuse, T., Matsuo, N., Sone, M., Watanabe, M., Bito, H., Terashima, T., Wright, C. V. E., Kawaguchi, Y., Nakao, K., & Nabeshima, Y. (2005). Ptf1a, a bHLH Transcriptional Gene, Defines GABAergic Neuronal Fates in Cerebellum. *Neuron*, *47*(2), 201–213. https://doi.org/10.1016/j.neuron.2005.06.007

Jékely, G., Jasek, S., Gühmann, M., Bezares-Calderón, L. A., Williams, E. A., & Shahidi, R. (2024). *Whole-body connectome of a segmented annelid larva* (p. 2024.03.17.585258). bioRxiv. https://doi.org/10.1101/2024.03.17.585258

Jokura, K., Ueda, N., Gühmann, M., Yañez-Guerra, L. A., Słowiński, P., Wedgwood, K. C. A., & Jékely, G. (2023). Nitric oxide feedback to ciliary photoreceptor cells gates a UV avoidance circuit. *eLife*, *12*. https://doi.org/10.7554/eLife.91258.1

Joseph, D. J., Von Deimling, M., Hasegawa, Y., Cristancho, A. G., Ahrens-Nicklas, R. C., Rogers, S. L., Risbud, R., McCoy, A. J., & Marsh, E. D. (2020). Postnatal Arx transcriptional activity regulates functional properties of PV interneurons. *iScience*, *24*(1), 101999. https://doi.org/10.1016/j.isci.2020.101999

Junttila, S., Smolander, J., & Elo, L. L. (2022). Benchmarking methods for detecting differential states between conditions from multi-subject single-cell RNA-seq data. *Briefings in Bioinformatics*, *23*(5), bbac286. https://doi.org/10.1093/bib/bbac286

Karess, R. E., Chang, X., Edwards, K. A., Kulkarni, S., Aguilera, I., & Kiehart, D. P. (1991). The regulatory light chain of nonmuscle myosin is encoded by *spaghetti-squash*, a gene required for cytokinesis in Drosophila. *Cell*, *65*(7), 1177–1189. https://doi.org/10.1016/0092-8674(91)90013-O

Kennedy, S. R., Prost, S., Overcast, I., Rominger, A. J., Gillespie, R. G., & Krehenwinkel, H. (2020). High-throughput sequencing for community analysis: The promise of DNA barcoding to uncover diversity, relatedness, abundances and interactions in spider communities. *Development Genes and Evolution*, *230*(2), 185–201. https://doi.org/10.1007/s00427-020-00652-x

Kerner, P., Simionato, E., Le Gouar, M., & Vervoort, M. (2009). Orthologs of key vertebrate neural genes are expressed during neurogenesis in the annelid Platynereis dumerilii. *Evolution & Development*, *11*(5), 513–524. https://doi.org/10.1111/j.1525-142X.2009.00359.x

Kim, Y.-C., Lee, H.-G., Lim, J., & Han, K.-A. (2013). Appetitive Learning Requires the Alpha1-Like Octopamine Receptor OAMB in the Drosophila Mushroom Body Neurons. *Journal of Neuroscience*, *33*(4), 1672–1677.

https://doi.org/10.1523/JNEUROSCI.3042-12.2013

Kitamura, K., Yanazawa, M., Sugiyama, N., Miura, H., Iizuka-Kogo, A., Kusaka, M., Omichi, K., Suzuki, R., Kato-Fukui, Y., Kamiirisa, K., Matsuo, M., Kamijo, S., Kasahara, M., Yoshioka, H., Ogata, T., Fukuda, T., Kondo, I., Kato, M., Dobyns, W. B., … Morohashi, K. (2002). Mutation of ARX causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans. *Nature Genetics*, *32*(3), 359–369. https://doi.org/10.1038/ng1009

Kljakic, O., Janickova, H., Prado, V. F., & Prado, M. A. M. (2017). Cholinergic/glutamatergic co-transmission in striatal cholinergic interneurons: New mechanisms regulating striatal computation. *Journal of Neurochemistry*, *142*(S2), 90–102. https://doi.org/10.1111/jnc.14003

Kobeissy, F. H., Hansen, K., Neumann, M., Fu, S., Jin, K., & Liu, J. (2016). Deciphering the Role of Emx1 in Neurogenesis: A Neuroproteomics Approach. *Frontiers in Molecular Neuroscience*, *9*, 98. https://doi.org/10.3389/fnmol.2016.00098

Krashes, M. J., DasGupta, S., Vreede, A., White, B., Armstrong, J. D., & Waddell, S. (2009). A Neural Circuit Mechanism Integrating Motivational State with Memory Expression in Drosophila. *Cell*, *139*(2), 416–427. https://doi.org/10.1016/j.cell.2009.08.035

Kuehn, E., Stockinger, A. W., Girard, J., Raible, F., & Özpolat, B. D. (2019). A scalable culturing system for the marine annelid Platynereis dumerilii. *PLoS ONE*, *14*(12), e0226156. https://doi.org/10.1371/journal.pone.0226156

Lamanna, F., Hervas-Sotomayor, F., Oel, A. P., Jandzik, D., Sobrido-Cameán, D., Santos-Durán, G. N., Martik, M. L., Stundl, J., Green, S. A., Brüning, T., Mößinger, K., Schmidt, J., Schneider, C., Sepp, M., Murat, F., Smith, J. J., Bronner, M. E., Rodicio, M. C., Barreiro-Iglesias, A., … Kaessmann, H. (2023). A lamprey neural cell type atlas illuminates the origins of the vertebrate brain. *Nature Ecology & Evolution*, *7*(10), 1714–1728. https://doi.org/10.1038/s41559-023-02170-1

Lee, W.-P., Chiang, M.-H., Chang, L.-Y., Shyu, W.-H., Chiu, T.-H., Fu, T.-F., Wu, T., & Wu, C.-L. (2021). Serotonin Signals Modulate Mushroom Body Output Neurons for Sustaining Water-Reward Long-Term Memory in Drosophila. *Frontiers in Cell and Developmental Biology*, *9*. https://doi.org/10.3389/fcell.2021.755574

Legras, M., Ghisleni, G., Regnard, L., Dias, M., Soilihi, R., Celmar, E., & Balavoine, G. (2023). Fast cycling culture of the annelid model Platynereis dumerilii. *PLOS ONE*, *18*(12), e0295290. https://doi.org/10.1371/journal.pone.0295290

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T., & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, *10*(1), 34. https://doi.org/10.1186/1742-9994-10-34

Leto, K., Bartolini, A., & Rossi, F. (2008). Development of Cerebellar GABAergic Interneurons: Origin and Shaping of the "Minibrain" Local Connections. *The Cerebellum*, *7*(4), 523–529. https://doi.org/10.1007/s12311-008-0079-z

Li, H., Horns, F., Wu, B., Xie, Q., Li, J., Li, T., Luginbuhl, D. J., Quake, S. R., & Luo, L. (2017). Classifying Drosophila Olfactory Projection Neuron Subtypes by Single-Cell RNA Sequencing. *Cell*, *171*(5), 1206-1220.e22. https://doi.org/10.1016/j.cell.2017.10.019

Llinás, R. R. (2003). The contribution of Santiago Ramon y Cajal to functional neuroscience. *Nature Reviews Neuroscience*, *4*(1), 77–80. https://doi.org/10.1038/nrn1011

Llorens-Bobadilla, E., Zhao, S., Baser, A., Saiz-Castro, G., Zwadlo, K., & Martin-Villalba, A. (2015). Single-Cell Transcriptomics Reveals a Population of Dormant Neural Stem Cells that Become Activated upon Brain Injury. *Cell Stem Cell*, *17*(3), 329–340. https://doi.org/10.1016/j.stem.2015.07.002

Lozovaya, N., Eftekhari, S., Cloarec, R., Gouty-Colomer, L. A., Dufour, A., Riffault, B., Billon-Grand, M., Pons-Bennaceur, A., Oumar, N., Burnashev, N., Ben-Ari, Y., & Hammond, C. (2018). GABAergic inhibition in dual-transmission cholinergic and GABAergic striatal interneurons is abolished in Parkinson disease. *Nature Communications*, *9*(1), 1422. https://doi.org/10.1038/s41467-018-03802-y

Lucey, N. M., Lombardi, C., DeMarchi, L., Schulze, A., Gambi, M. C., & Calosi, P. (2015). To brood or not to brood: Are marine invertebrates that protect their offspring more resilient to ocean acidification? *Scientific Reports*, *5*, 12009. https://doi.org/10.1038/srep12009

Madeira, F., Madhusoodanan, N., Lee, J., Eusebi, A., Niewielska, A., Tivey, A. R. N., Lopez, R., & Butcher, S. (2024). The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Research*, *52*(W1), W521–W525. https://doi.org/10.1093/nar/gkae241

Marín, O., Anderson, S. A., & Rubenstein, J. L. R. (2000). Origin and Molecular Specification of Striatal Interneurons. *The Journal of Neuroscience*, *20*(16), 6063–6076. https://doi.org/10.1523/JNEUROSCI.20-16-06063.2000

Marsh, E. D., Nasrallah, M. P., Walsh, C., Murray, K. A., Nicole Sunnen, C., McCoy, A., & Golden, J. A. (2016). Developmental interneuron subtype deficits after targeted loss of Arx. *BMC Neuroscience*, *17*(1), 35. https://doi.org/10.1186/s12868-016-0265-8

Michels, G., Brandt, M. C., Zagidullin, N., Khan, I. F., Larbig, R., van Aaken, S., Wippermann, J., & Hoppe, U. C. (2008). Direct evidence for calcium conductance of hyperpolarization-activated cyclic nucleotide-gated channels and human native If at physiological calcium concentrations. *Cardiovascular Research*, *78*(3), 466–475. https://doi.org/10.1093/cvr/cvn032

Milivojev, N., Gamboa, C. L. V., Andreatta, G., Raible, F., & Tessmar-Raible, K. (2024). *A c-opsin functions in a ciliary-marginal zone-like stem cell region of an invertebrate camera-type eye* (p. 2024.08.19.608633). bioRxiv. https://doi.org/10.1101/2024.08.19.608633

Minguillón, C., Ferrier, D. E. K., Cebrián, C., & Garcia-Fernàndez, J. (2002). Gene duplications in the prototypical cephalochordate amphioxus. *Gene*, *287*(1), 121–128. https://doi.org/10.1016/S0378-1119(01)00828-9

Mu, L., Berti, L., Masserdotti, G., Covic, M., Michaelidis, T. M., Doberauer, K., Merz, K.,

Rehfeld, F., Haslinger, A., Wegner, M., Sock, E., Lefebvre, V., Couillard-Despres, S., Aigner, L., Berninger, B., & Lie, D. C. (2012). SoxC transcription factors are required for neuronal differentiation in adult hippocampal neurogenesis. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(9), 3067–3080. https://doi.org/10.1523/JNEUROSCI.4679-11.2012

Musser, J. M., Schippers, K. J., Nickel, M., Mizzon, G., Kohn, A. B., Pape, C., Ronchi, P., Papadopoulos, N., Tarashansky, A. J., Hammel, J. U., Wolf, F., Liang, C., Hernández-Plaza, A., Cantalapiedra, C. P., Achim, K., Schieber, N. L., Pan, L., Ruperti, F., Francis, W. R., … Arendt, D. (2021). Profiling cellular diversity in sponges informs animal cell type and nervous system evolution. *Science*, *374*(6568), 717–723. https://doi.org/10.1126/science.abj2949

Mutemi, K. N., Simakov, O., Pan, L., Santangeli, L., Null, R. W., Handberg-Thorsager, M., Vellutini, B. C., Peterson, K. J., Fromm, B., Larsson, T., Savage, E., Lopez, M. O., Hercog, R., Provaznik, J., Ordoñez-Rueda, D., Azevedo, N., Gazave, E., Vervoort, M., Tomancak, P., … Arendt, D. (2025). A genome resource for the marine annelid Platynereis spp. *BMC Genomics*, *26*(1), 665. https://doi.org/10.1186/s12864-025-11727-2

Nakama, A. B., Chou, H.-C., & Schneider, S. Q. (2017). The asymmetric cell division machinery in the spiral-cleaving egg and embryo of the marine annelid Platynereis dumerilii. *BMC Developmental Biology*, *17*(1), 16. https://doi.org/10.1186/s12861-017-0158-9

Noro, M., Sugahara, F., & Kuraku, S. (2015). Reevaluating Emx gene phylogeny: Homopolymeric amino acid tracts as a potential factor obscuring orthology signals in cyclostome genes. *BMC Evolutionary Biology*, *15*(1), 78. https://doi.org/10.1186/s12862-015-0351-z

Oel, P. A., Cros, C., Pan, L., Buglakova, A., Haury Parra, S., Santangeli, L., Gerber, T., Papadopoulos, N., Vergara, H. M., Puga, D., Disela, V., Bertucci, P. Y., Witte, V., Quintana-Urzainqui, I., Acevedo, J., Musser, J., Kreshuk, A., & Arendt, D. (n.d.). *A whole-body cell type atlas mapped into an electron microscopy volume of an annelid worm*.

Okada, M., & Shi, Y.-B. (2018). The balance of two opposing factors Mad and Myc regulates cell fate during tissue remodeling. *Cell & Bioscience*, *8*(1), 51. https://doi.org/10.1186/s13578-018-0249-8

Özel, M. N., & Desplan, C. (2025). Does a cell's gene expression always reflect its function? *Nature*. https://doi.org/10.1038/d41586-025-00088-1

Özpolat, B. D., Randel, N., Williams, E. A., Bezares-Calderón, L. A., Andreatta, G., Balavoine, G., Bertucci, P. Y., Ferrier, D. E. K., Gambi, M. C., Gazave, E., Handberg-Thorsager, M., Hardege, J., Hird, C., Hsieh, Y.-W., Hui, J., Mutemi, K. N., Schneider, S. Q., Simakov, O., Vergara, H. M., … Arendt, D. (2021). The Nereid on the rise: Platynereis as a model system. *EvoDevo*, *12*(1), 10. https://doi.org/10.1186/s13227-021-00180-3

Pai, E. L.-L., Chen, J., Fazel Darbandi, S., Cho, F. S., Chen, J., Lindtner, S., Chu, J. S., Paz, J. T., Vogt, D., Paredes, M. F., & Rubenstein, J. L. (2020). Maf and Mafb control mouse pallial interneuron fate and maturation through neuropsychiatric disease gene

regulation. *eLife*, *9*, e54903. https://doi.org/10.7554/eLife.54903

Parry, L., & Caron, J.-B. (2019). *Canadia spinosa* and the early evolution of the annelid nervous system. *Science Advances*, *5*(9), eaax5858. https://doi.org/10.1126/sciadv.aax5858

Paulet, D., David, A., & Rivals, E. (2017). Ribo-seq enlightens codon usage bias. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, *24*(3), 303–210. https://doi.org/10.1093/dnares/dsw062

Pende, M., Vadiwala, K., Schmidbaur, H., Stockinger, A. W., Murawala, P., Saghafi, S., Dekens, M. P. S., Becker, K., Revilla-I-Domingo, R., Papadopoulos, S.-C., Zurl, M., Pasierbek, P., Simakov, O., Tanaka, E. M., Raible, F., & Dodt, H.-U. (2020). A versatile depigmentation, clearing, and labeling method for exploring nervous system diversity. *Science Advances*, *6*(22), eaba0365. https://doi.org/10.1126/sciadv.aba0365

Platynereis dumerilii Mikroskop. Anatomie, Fortpflanzung, Entwicklung (with Hauenschild, C., & Fischer, A.). (1969). G. Fischer.

Poehn, B., Krishnan, S., Zurl, M., Coric, A., Rokvic, D., Häfker, N. S., Jaenicke, E., Arboleda, E., Orel, L., Raible, F., Wolf, E., & Tessmar-Raible, K. (2022). A Cryptochrome adopts distinct moon- and sunlight states and functions as sun- versus moonlight interpreter in monthly oscillator entrainment. *Nature Communications*, *13*(1), 5220. https://doi.org/10.1038/s41467-022-32562-z

Quéva, C., Hurlin, P. J., Foley, K. P., & Eisenman, R. N. (1998). Sequential expression of the MAD family of transcriptional repressors during differentiation and development. *Oncogene*, *16*(8), 967–977. https://doi.org/10.1038/sj.onc.1201611

Quintana-Urzainqui, I., Oel, P. A., Gerber, T., Pan, L., Papadopoulos, N., Serka, Z. G., Verbanac, A., Börsig, M., Torres-Sabino, D., Rollán-Delgado, I. C., Santangeli, L., Kaessmann, H., & Arendt, D. (2025). *Developmental landscape of shark telencephalon sheds light on the evolution of telencephalic cell types and the ancient origin of Cajal-Retzius cells* (p. 2025.05.08.652169). bioRxiv. https://doi.org/10.1101/2025.05.08.652169

Raible, F., Tessmar-Raible, K., Osoegawa, K., Wincker, P., Jubin, C., Balavoine, G., Ferrier, D., Benes, V., de Jong, P., Weissenbach, J., Bork, P., & Arendt, D. (2005). Vertebrate-Type Intron-Rich Genes in the Marine Annelid *Platynereis dumerilii*. *Science*, *310*(5752), 1325–1326. https://doi.org/10.1126/science.1119089

Ramón y Cajal, S. (1906). *Structure et connexions des neurones: Conférence de nobel faite à Stockholm le 12 décembre 1906.*

Ramón y Cajal, S., & Azoulay, L. (1955). Histologie du systeme nerveux de l'homme et des vertebres. In *Histologie du systeme nerveux de l'homme et des vertebres* (pp. 2v–2v). https://pesquisa.bvsalud.org/portal/resource/pt/biblio-1080560

Ribeiro, R. P., Null, R. W., & Özpolat, B. D. (2024). *Sex-biased gene expression precedes sexual dimorphism in the agonadal annelid Platynereis dumerilii* (p. 2024.06.12.598746). bioRxiv. https://doi.org/10.1101/2024.06.12.598746

Robinson, O., Dylus, D., & Dessimoz, C. (2016). Phylo.io: Interactive Viewing and Comparison of Large Phylogenetic Trees on the Web. *Molecular Biology and Evolution*, *33*(8), 2163–2166. https://doi.org/10.1093/molbev/msw080

Ryu, J. R., Hong, C. J., Kim, J. Y., Kim, E.-K., Sun, W., & Yu, S.-W. (2016). Control of adult neurogenesis by programmed cell death in the mammalian brain. *Molecular Brain*, *9*, 43. https://doi.org/10.1186/s13041-016-0224-4

Schenk, S., Bannister, S. C., Sedlazeck, F. J., Anrather, D., Minh, B. Q., Bileck, A., Hartl, M., von Haeseler, A., Gerner, C., Raible, F., & Tessmar-Raible, K. (2019). Combined transcriptome and proteome profiling reveals specific molecular brain signatures for sex, maturation and circalunar clock phase. *eLife*, *8*, e41556. https://doi.org/10.7554/eLife.41556

Schenk, S., Krauditsch, C., Frühauf, P., Gerner, C., & Raible, F. (2016). Discovery of methylfarnesoate as the annelid brain hormone reveals an ancient role of sesquiterpenoids in reproduction. *eLife*, *5*, e17126. https://doi.org/10.7554/eLife.17126

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682. https://doi.org/10.1038/nmeth.2019

Schneider, S., Fischer, A., & Dorresteijn, A. W. C. (1992). A morphometric comparison of dissimilar early development in sibling species of Platynereis (Annelida, Polychaeta). *Roux's Archives of Developmental Biology*, *201*(4), 243–256. https://doi.org/10.1007/BF00188755

Shafer, M. E. R., Sawh, A. N., & Schier, A. F. (2022). Gene family evolution underlies cell-type diversification in the hypothalamus of teleosts. *Nature Ecology & Evolution*, *6*(1), 63–76. https://doi.org/10.1038/s41559-021-01580-3

Shahidi, R., Williams, E. A., Conzelmann, M., Asadulina, A., Verasztó, C., Jasek, S., Bezares-Calderón, L. A., & Jékely, G. (2015). A serial multiplex immunogold labeling method for identifying peptidergic neurons in connectomes. *eLife*, *4*, e11147. https://doi.org/10.7554/eLife.11147

Shainer, I., Kappel, J. M., Laurell, E., Donovan, J. C., Schneider, M. W., Kuehn, E., Arnold-Ammer, I., Stemmer, M., Larsch, J., & Baier, H. (2025). Transcriptomic neuron types vary topographically in function and morphology. *Nature*, *638*(8052), 1023–1033. https://doi.org/10.1038/s41586-024-08518-2

Sheng, L., Shields, E. J., Gospocic, J., Glastad, K. M., Ratchasanmuang, P., Berger, S. L., Raj, A., Little, S., & Bonasio, R. (2020). Social reprogramming in ants induces longevity-associated glia remodeling. *Science Advances*, *6*(34), eaba9869. https://doi.org/10.1126/sciadv.aba9869

Shields, E. J., Sorida, M., Sheng, L., Sieriebriennikov, B., Ding, L., & Bonasio, R. (2021). Genome annotation with long RNA reads reveals new patterns of gene expression and improves single-cell analyses in an ant brain. *BMC Biology*, *19*(1), 254. https://doi.org/10.1186/s12915-021-01188-w

Shoshan, Y., Liscovitch-Brauer, N., Rosenthal, J. J. C., & Eisenberg, E. (2021). Adaptive Proteome Diversification by Nonsynonymous A-to-I RNA Editing in Coleoid

Cephalopods. *Molecular Biology and Evolution*, *38*(9), 3775–3788. https://doi.org/10.1093/molbev/msab154

Song, Y., Miao, Z., Brazma, A., & Papatheodorou, I. (2023). Benchmarking strategies for cross-species integration of single-cell RNA sequencing data. *Nature Communications*, *14*(1), 6495. https://doi.org/10.1038/s41467-023-41855-w

Stockinger, A. W., Adelmann, L., Fahrenberger, M., Ruta, C., Özpolat, B. D., Milivojev, N., Balavoine, G., & Raible, F. (2024). Molecular profiles, sources and lineage restrictions of stem cells in an annelid regeneration model. *Nature Communications*, *15*(1), 9882. https://doi.org/10.1038/s41467-024-54041-3

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, *177*(7), 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031

Styfhals, R., Zolotarov, G., Hulselmans, G., Spanier, K. I., Poovathingal, S., Elagoz, A. M., De Winter, S., Deryckere, A., Rajewsky, N., Ponte, G., Fiorito, G., Aerts, S., & Seuntjens, E. (2022). Cell type diversity in a developing octopus brain. *Nature Communications*, *13*(1), 7392. https://doi.org/10.1038/s41467-022-35198-1

Su-Feher, L., Rubin, A. N., Silberberg, S. N., Catta-Preta, R., Lim, K. J., Ypsilanti, A. R., Zdilar, I., McGinnis, C. S., McKinsey, G. L., Rubino, T. E., Hawrylycz, M. J., Thompson, C., Gartner, Z. J., Puelles, L., Zeng, H., Rubenstein, J. L. R., & Nord, A. S. (2022). Single cell enhancer activity distinguishes GABAergic and cholinergic lineages in embryonic mouse basal ganglia. *Proceedings of the National Academy of Sciences*, *119*(15), e2108760119. https://doi.org/10.1073/pnas.2108760119

Tanay, A., & Sebé-Pedrós, A. (2021). Evolutionary cell type mapping with single-cell genomics. *Trends in Genetics*, *37*(10), 919–932. https://doi.org/10.1016/j.tig.2021.04.008

Tarashansky, A. J., Musser, J. M., Khariton, M., Li, P., Arendt, D., Quake, S. R., & Wang, B. (2021). Mapping single-cell atlases throughout Metazoa unravels cell type evolution. *eLife*, *10*, e66747. https://doi.org/10.7554/eLife.66747

Teixeira, M. A. L., Langeneck, J., Vieira, P. E., Hernández, J. C., Sampieri, B. R., Kasapidis, P., Mucciolo, S., Bakken, T., Ravara, A., Nygren, A., & Costa, F. O. (2022). Reappraisal of the hyperdiverse Platynereis dumerilii (Annelida: Nereididae) species complex in the Northern Atlantic, with the description of two new species. *Invertebrate Systematics*, *36*(11), 1017–1061. https://doi.org/10.1071/IS21084

Tessmar-Raible, K., Raible, F., Christodoulou, F., Guy, K., Rembold, M., Hausen, H., & Arendt, D. (2007). Conserved Sensory-Neurosecretory Cell Types in Annelid and Fish Forebrain: Insights into Hypothalamus Evolution. *Cell*, *129*(7), 1389–1400. https://doi.org/10.1016/j.cell.2007.04.041

Toker, I. A., Ripoll-Sánchez, L., Geiger, L. T., Saini, K. S., Beets, I., Vértes, P. E., Schafer, W. R., Ben-David, E., & Hobert, O. (2024). *Molecular patterns of evolutionary changes throughout the whole nervous system of multiple nematode species* (p. 2024.11.23.624988). bioRxiv. https://doi.org/10.1101/2024.11.23.624988

Tomer, R. (n.d.). *Dissertation submitted to the Combined Faculties for the Natural Sciences and for*

Mathematics of the Ruperto-Carola University of Heidelberg, Germany for the degree of Doctor of Natural Sciences. 186.

Tomer, R., Denes, A. S., Tessmar-Raible, K., & Arendt, D. (2010). Profiling by Image Registration Reveals Common Origin of Annelid Mushroom Bodies and Vertebrate Pallium. *Cell*, *142*(5), 800–809. https://doi.org/10.1016/j.cell.2010.07.043

Tosches, M. A., Bucher, D., Vopalensky, P., & Arendt, D. (2014). Melatonin Signaling Controls Circadian Swimming Behavior in Marine Zooplankton. *Cell*, *159*(1), 46–57. https://doi.org/10.1016/j.cell.2014.07.042

Tsuboi, A., & Yoshihara, S. (2025). Arx revisited: Involved in the development of GABAergic interneurons. *Frontiers in Cell and Developmental Biology*, *13*. https://doi.org/10.3389/fcell.2025.1563515

Tsukamoto, H., Chen, I.-S., Kubo, Y., & Furutani, Y. (2017). A ciliary opsin in the brain of a marine annelid zooplankton is ultraviolet-sensitive, and the sensitivity is tuned by a single amino acid residue. *Journal of Biological Chemistry*, *292*(31), 12971–12980. https://doi.org/10.1074/jbc.M117.793539

Tsukamoto, H., & Kubo, Y. (2023). A self-inactivating invertebrate opsin optically drives biased signaling toward Gβγ-dependent ion channel modulation. *Proceedings of the National Academy of Sciences*, *120*(21), e2301269120. https://doi.org/10.1073/pnas.2301269120

Upadhyay, A., Hosseinibarkooie, S., Schneider, S., Kaczmarek, A., Torres-Benito, L., Mendoza-Ferreira, N., Overhoff, M., Rombo, R., Grysko, V., Kye, M. J., Kononenko, N. L., & Wirth, B. (2019). Neurocalcin Delta Knockout Impairs Adult Neurogenesis Whereas Half Reduction Is Not Pathological. *Frontiers in Molecular Neuroscience*, *12*, 19. https://doi.org/10.3389/fnmol.2019.00019

Valderrama, X., Rapin, N., Verge, V. M. K., & Misra, V. (2009). Zhangfei induces the expression of the nerve growth factor receptor, trkA, in medulloblastoma cells and causes their differentiation or apoptosis. *Journal of Neuro-Oncology*, *91*(1), 7–17. https://doi.org/10.1007/s11060-008-9682-6

VanInsberghe, M., van den Berg, J., Andersson-Rolf, A., Clevers, H., & van Oudenaarden, A. (2021). Single-cell Ribo-seq reveals cell cycle-dependent translational pausing. *Nature*, *597*(7877), 561–565. https://doi.org/10.1038/s41586-021-03887-4

Veedin Rajan, V. B., Häfker, N. S., Arboleda, E., Poehn, B., Gossenreiter, T., Gerrard, E., Hofbauer, M., Mühlestein, C., Bileck, A., Gerner, C., Ribera d'Alcala, M., Buia, M. C., Hartl, M., Lucas, R. J., & Tessmar-Raible, K. (2021). Seasonal variation in UVA light drives hormonal and behavioural changes in a marine annelid via a ciliary opsin. *Nature Ecology & Evolution*, *5*(2), 204–218. https://doi.org/10.1038/s41559-020-01356-1

Verasztó, C., Gühmann, M., Jia, H., Rajan, V. B. V., Bezares-Calderón, L. A., Piñeiro-Lopez, C., Randel, N., Shahidi, R., Michiels, N. K., Yokoyama, S., Tessmar-Raible, K., & Jékely, G. (2018). Ciliary and rhabdomeric photoreceptor-cell circuits form a spectral depth gauge in marine zooplankton. *eLife*, *7*, e36440. https://doi.org/10.7554/eLife.36440

Vergara, H. M., Pape, C., Meechan, K. I., Zinchenko, V., Genoud, C., Wanner, A. A., Mutemi, K. N., Titze, B., Templin, R. M., Bertucci, P. Y., Simakov, O., Dürichen, W., Machado, P., Savage, E. L., Schermelleh, L., Schwab, Y., Friedrich, R. W., Kreshuk, A., Tischer, C., & Arendt, D. (2021). Whole-body integration of gene expression and single-cell morphology. *Cell*. https://doi.org/10.1016/j.cell.2021.07.017

Voss, G., & Rosenthal, J. J. C. (2023). High-level RNA editing diversifies the coleoid cephalopod brain proteome. *Briefings in Functional Genomics*, *22*(6), 525–532. https://doi.org/10.1093/bfgp/elad034

Wäge, J., Valvassori, G., Hardege, J. D., Schulze, A., & Gambi, M. C. (2017). The sibling polychaetes Platynereis dumerilii and Platynereis massiliensis in the Mediterranean Sea: Are phylogeographic patterns related to exposure to ocean acidification? *Marine Biology*, *164*(10), 199. https://doi.org/10.1007/s00227-017-3222-x

Wei, B., Nie, Y., Li, X., Wang, C., Ma, T., Huang, Z., Tian, M., Sun, C., Cai, Y., You, Y., Liu, F., & Yang, Z. (2011). Emx1-expressing neural stem cells in the subventricular zone give rise to new interneurons in the ischemic injured striatum. *European Journal of Neuroscience*, *33*(5), 819–830. https://doi.org/10.1111/j.1460-9568.2010.07570.x

Wietek, J., Nozownik, A., Pulin, M., Saraf-Sinik, I., Matosevich, N., Gowrishankar, R., Gat, A., Malan, D., Brown, B. J., Dine, J., Imambocus, B. N., Levy, R., Sauter, K., Litvin, A., Regev, N., Subramaniam, S., Abrera, K., Summarli, D., Goren, E. M., … Yizhar, O. (2024). A bistable inhibitory optoGPCR for multiplexed optogenetic control of neural circuits. *Nature Methods*, *21*(7), 1275–1287. https://doi.org/10.1038/s41592-024-02285-8

Wolock, S. L., Lopez, R., & Klein, A. M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Systems*, *8*(4), 281-291.e9. https://doi.org/10.1016/j.cels.2018.11.005

Woych, J., Ortega Gurrola, A., Deryckere, A., Jaeger, E. C. B., Gumnit, E., Merello, G., Gu, J., Joven Araus, A., Leigh, N. D., Yun, M., Simon, A., & Tosches, M. A. (2022). Cell-type profiling in salamanders identifies innovations in vertebrate forebrain evolution. *Science*, *377*(6610), eabp9186. https://doi.org/10.1126/science.abp9186

Wu, C.-L., Shih, M.-F. M., Lee, P.-T., & Chiang, A.-S. (2013). An Octopamine-Mushroom Body Circuit Modulates the Formation of Anesthesia-Resistant Memory in Drosophila. *Current Biology*, *23*(23), 2346–2354. https://doi.org/10.1016/j.cub.2013.09.056

Wu, L., Wen, C., Qin, Y., Yin, H., Tu, Q., Van Nostrand, J. D., Yuan, T., Yuan, M., Deng, Y., & Zhou, J. (2015). Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. *BMC Microbiology*, *15*(1), 125. https://doi.org/10.1186/s12866-015-0450-4

Wulf, P. O., Häfker, N. S., Hofmann, K., & Tessmar-Raible, K. (2025). Guiding Light: Mechanisms and Adjustments of Environmental Light Interpretation with Insights from Platynereis dumerilii and Other Selected Examples. *Zoological Science*, *42*(1). https://doi.org/10.2108/zs240099

Xu, Y., Liu, J., Zhao, T., Song, F., Tian, L., Cai, W., Li, H., & Duan, Y. (2023). Identification and Interpretation of A-to-I RNA Editing Events in Insect Transcriptomes. *International*

Journal of Molecular Sciences, *24*(24), 17126. https://doi.org/10.3390/ijms242417126

Yuan, J., Tirabassi, R. S., Bush, A. B., & Cole, M. D. (1998). The C. elegans MDL-1 and MXL-1 proteins can functionally substitute for vertebrate MAD and MAX. *Oncogene*, *17*(9), 1109–1118. https://doi.org/10.1038/sj.onc.1202036

Zantke, J., Ishikawa-Fujiwara, T., Arboleda, E., Lohs, C., Schipany, K., Hallay, N., Straw, A. D., Todo, T., & Tessmar-Raible, K. (2013). Circadian and Circalunar Clock Interactions in a Marine Annelid. *Cell Reports*, *5*(1), 99–113. https://doi.org/10.1016/j.celrep.2013.08.031

Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L. E., La Manno, G., Codeluppi, S., Furlan, A., Lee, K., Skene, N., Harris, K. D., Hjerling-Leffler, J., Arenas, E., Ernfors, P., Marklund, U., & Linnarsson, S. (2018). Molecular Architecture of the Mouse Nervous System. *Cell*, *174*(4), 999-1014.e22. https://doi.org/10.1016/j.cell.2018.06.021

Zinchenko, V., Hugger, J., Uhlmann, V., Arendt, D., & Kreshuk, A. (2023). MorphoFeatures for unsupervised exploration of cell types, tissues, and organs in volume electron microscopy. eLife, 12, e80918. https://doi.org/10.7554/eLife.80918

Zurl, M., Poehn, B., Rieger, D., Krishnan, S., Rokvic, D., Veedin Rajan, V. B., Gerrard, E., Schlichting, M., Orel, L., Ćorić, A., Lucas, R. J., Wolf, E., Helfrich-Förster, C., Raible, F., & Tessmar-Raible, K. (2022). Two light sensors decode moonlight versus sunlight to adjust a plastic circadian/circalunidian clock to moon phase. *Proceedings of the National Academy of Sciences*, *119*(22), e2115725119. https://doi.org/10.1073/pnas.2115725119

R packages

Base R

R Core Team (2023). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Seurat

Stuart and Butler et al. Comprehensive Integration of Single-Cell Data. Cell (2019) [Seurat V3]

Hao and Hao et al. Integrated analysis of multimodal single-cell data. Cell (2021) [Seurat V4]

Hao et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nature Biotechnology (2023) [Seurat V5]

sctransform

Hafemeister, C., & Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology, 20*(1), 296. https://doi.org/10.1186/s13059-019-1874-1

Choudhary, S., & Satija, R. (2022). Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biology, 23*(1), 1–20. https://doi.org/10.1186/s13059-022-02631-1

glmGamPoi

Ahlmann-Eltze, C., & Huber, W. (2020). glmGamPoi: Fitting Gamma-Poisson generalized linear models on single cell count data. *Bioinformatics, 36*(24), 5701–5702. https://doi.org/10.1093/bioinformatics/btaa1009

Scillus

Xu, M., & Ge, Z. (2021). *Scillus: Seurat wrapper package enhancing the processing and visualization of single cell data* (Version 0.5.0) [R package]. https://CRAN.R-project.org/package=Scillus

Tidyverse

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A grammar of data manipulation* (Version 1.1.4) [R package]. https://CRAN.R-project.org/package=dplyr

Wickham, H. (2022). *stringr: Simple, consistent wrappers for common string operations* (Version 1.5.0) [R package]. https://CRAN.R-project.org/package=stringr

Plotting

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.

Wilke, C. O. (2020). *cowplot: Streamlined plot theme and plot annotations for "ggplot2"* (Version 1.1.1) [R package]. https://CRAN.R-project.org/package=cowplot

Pedersen, T. L. (2023). *patchwork: The composer of plots* (Version 1.1.3) [R package]. https://CRAN.R-project.org/package=patchwork

Slowikowski, K. (2023). *ggrepel: Automatically position non-overlapping text labels with "ggplot2"* (Version 0.9.4) [R package]. https://CRAN.R-project.org/package=ggrepel

Yan, L. (2023). *ggvenn: Draw Venn diagram by "ggplot2"* (Version 0.1.10) [R package]. https://CRAN.R-project.org/package=ggvenn

Campitelli, E. (2023). *ggnewscale: Multiple fill and colour scales in "ggplot2"* (Version 0.4.9) [R package]. https://CRAN.R-project.org/package=ggnewscale

Garnier, S., Ross, N., Rudis, R., Camargo, A. P., Sciaini, M., & Scherer, C. (2023). *viridis: Colorblind-friendly color maps for R* (Version 0.6.4) [R package]. https://CRAN.R-project.org/package=viridis

Shenoy, A. (2021). *grafify: An R package for easy graphs, ANOVAs and post-hoc comparisons* (Version 5.0.0.1) [R package]. https://doi.org/10.5281/zenodo.5136508

Biostrings

Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2023). *Biostrings: Efficient manipulation of biological strings* (Version 2.70.0) [R package]. Bioconductor. https://doi.org/10.18129/B9.bioc.Biostrings

Trees

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics, 35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution, 3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Yu, G. (2022). *Data integration, manipulation and visualization of phylogenetic trees* (1st ed.). Chapman & Hall/CRC. https://doi.org/10.1201/9781003279242

Xu, S., Li, L., Luo, X., Chen, M., Tang, W., Zhan, L., Dai, Z., Lam, T. T., Guan, Y., & Yu, G. (2022). Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data. *iMeta, 1*(1), e56. https://doi.org/10.1002/imt2.56

Yu, G. (2020). Using ggtree to visualize data on tree-like structures. *Current Protocols in Bioinformatics, 69*(1), e96. https://doi.org/10.1002/cpbi.96

Yu, G., Lam, T. T. Y., Zhu, H., & Guan, Y. (2018). Two methods for mapping and visualizing

associated data on phylogeny using ggtree. *Molecular Biology and Evolution, 35*(12), 3041–3043. https://doi.org/10.1093/molbev/msy194

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution, 8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Appendix

Supplementary Figures

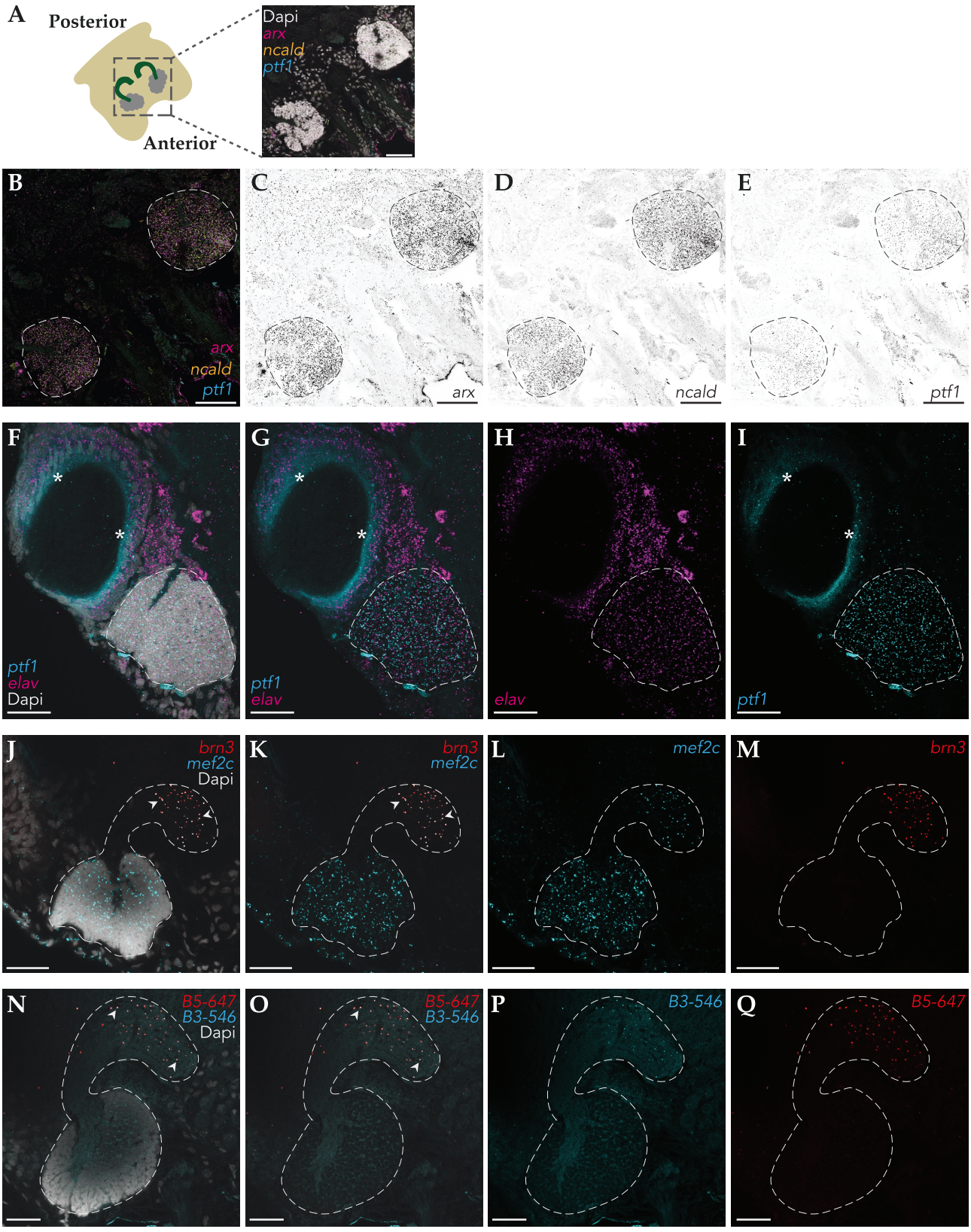
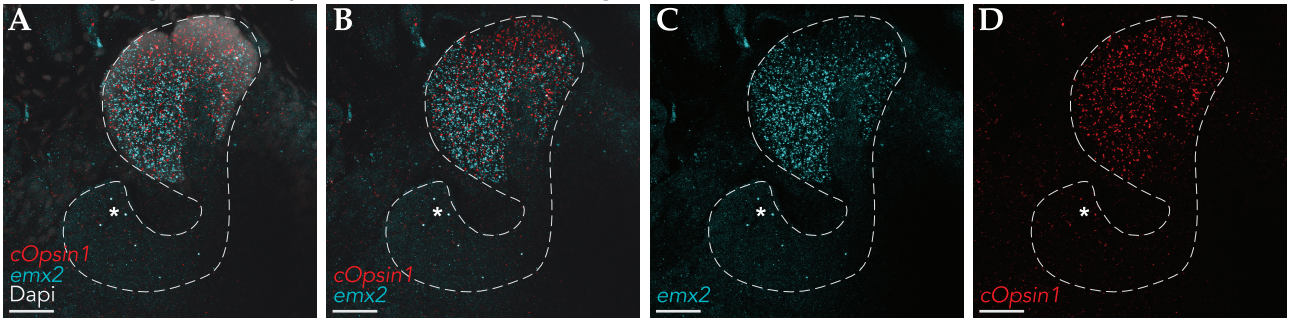
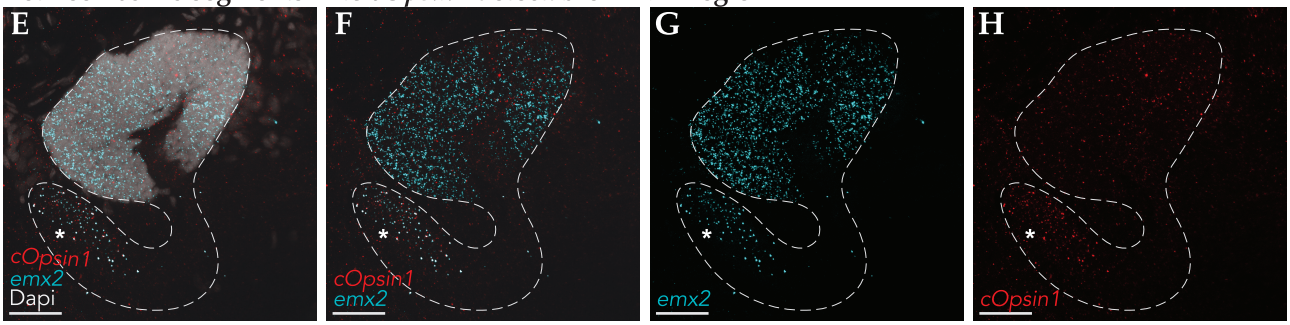


Figure S1. HCR control images and original validation on cryosections. (**A - E**) Triple - staining of MB markers *arx*, *ncald*, *ptf1* on 10μm cryosections, with orientation depicted in (**A**). 5μm z - projections (0.5μm step - size), 50μm scale bars. (**F - I**) General neuronal marker, *elav*, versus MB-specific gene, *ptf1*. Asterisks mark eye autofluorescence. 5μm z - projection (1μm step - size), 25μm scale bars. (**J - M**) Non-MB gene *brn3* co-stained with MB marker *mef2c*. Single focal plane, 25μm scale bars. (**N - Q**) Amplifier control. B3-546nm B5-647nm hairpins paired with a mismatching probe. Single focal plane, 25μm scalebars. Arrowheads point at examples of suspiciously overlapping signals in the lipid-rich peduncle, likely stuck hairpins. All images acquired with a confocal 40x oil immersion objective.

Below 60 segments - *cOpsin1* detectable in MB region



Between 60 -70 segments - No *cOpsin1* detectable in MB region



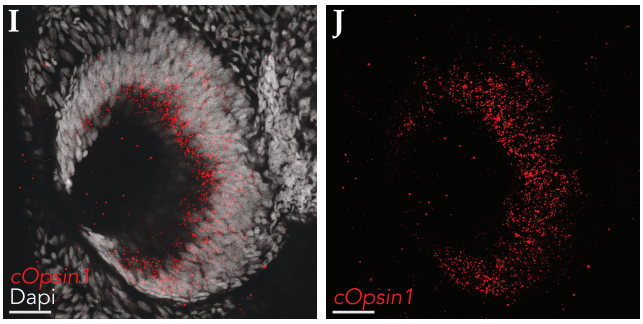


Figure S2. Exceptional cases of *cOpsin1* expression in ventral MB of premature worms. (**A - H**) 5μm z-projection (1μm step - size) of the ventral MB region. (**A - D**) Ventral MB of a 53 segment worm showing co-expression of *emx2* and *cOpsin1*. (**E - H**) Ventral MB of a 66 segment worm, where *cOpsin1* expression is not detected. (**I - J**) Growing rhabdomeric eye of the same 66 segment worm indeed shows *cOpsin1* expression, single focal plane. All scale bars 25μm. Images acquired with a confocal 40x oil objective. Asterisks indicate suspicious signals in the MB peduncles that will not be considered (see Figure S1).

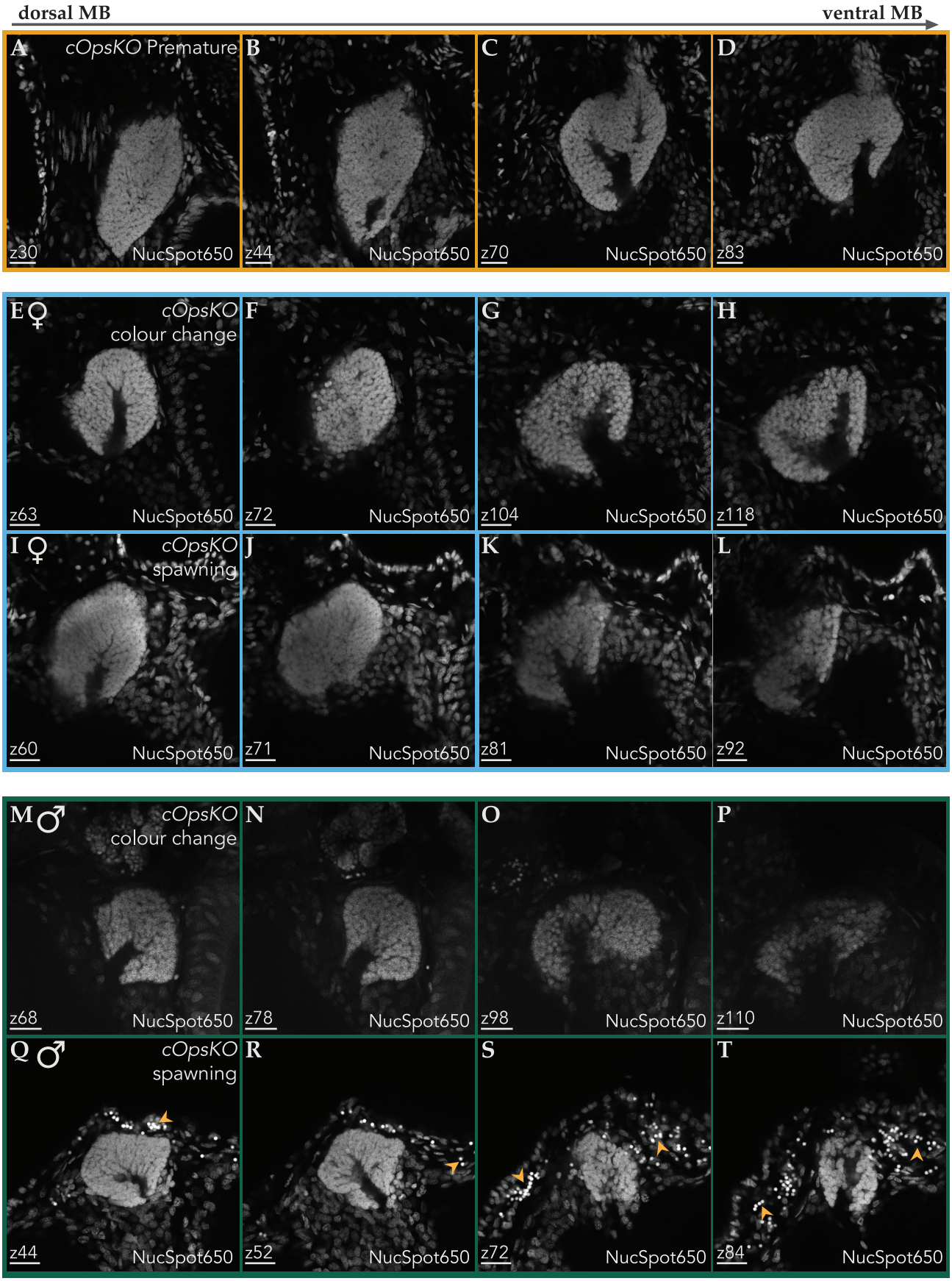


Figure S3. Restructuring of MB upon sexual maturation in *cOpsin1* knockout worms visualised with nuclear stains. Dorsal (left panel) to ventral (right panel) MBs shown. (**A - D**) Premature MB. (**E - H**) Colour changed female MB. (**I - L**) Spawning female MB. (**M - P**) Colour changed male MB. (**Q - T**) Spawning male MB. Arrowheads mark bright regions of male gametes stored in the head. For each image panel, top is anterior, right is lateral. All step sizes 1μm. All scale bars are 20μm. Images acquired with a confocal 40x oil objective.

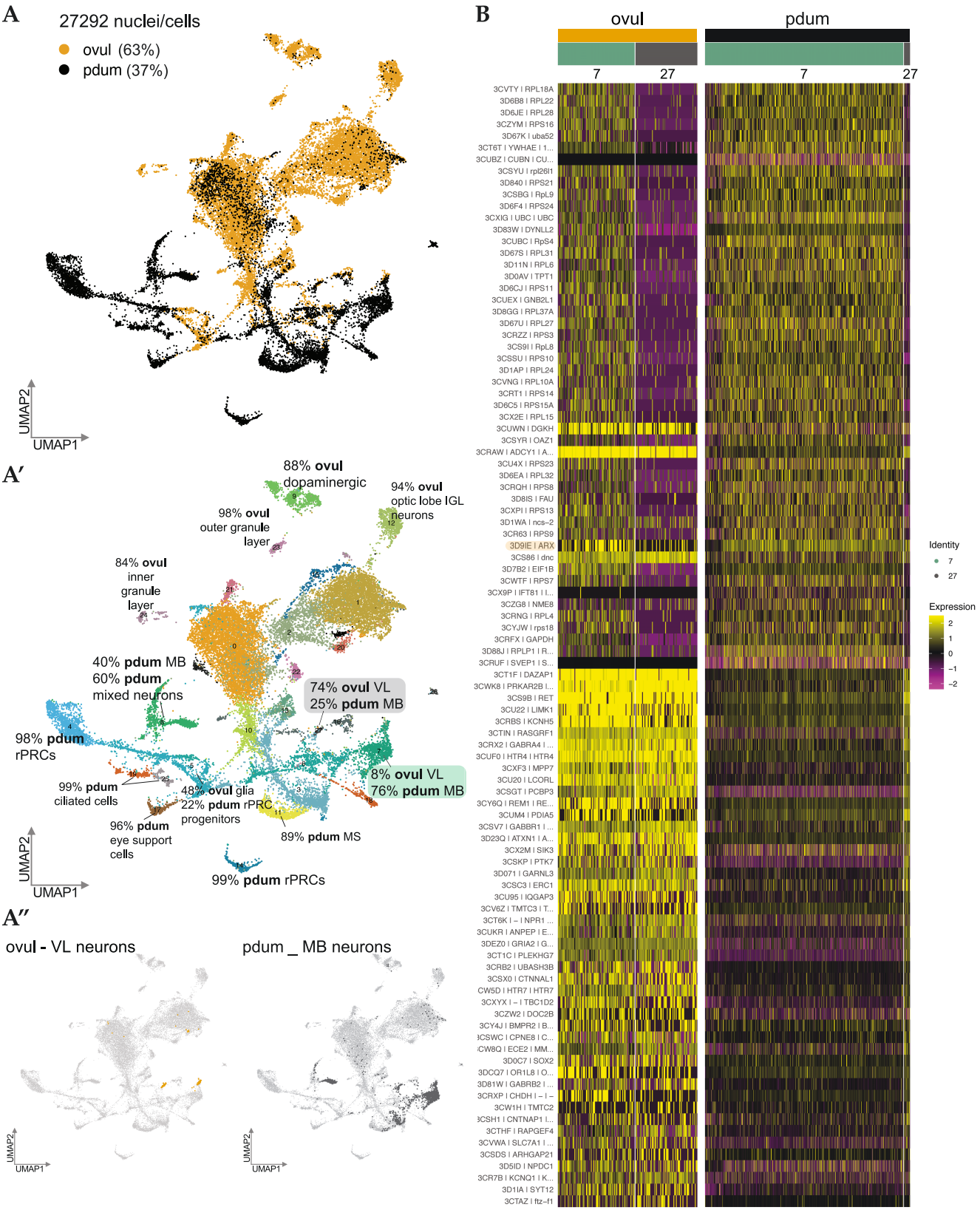
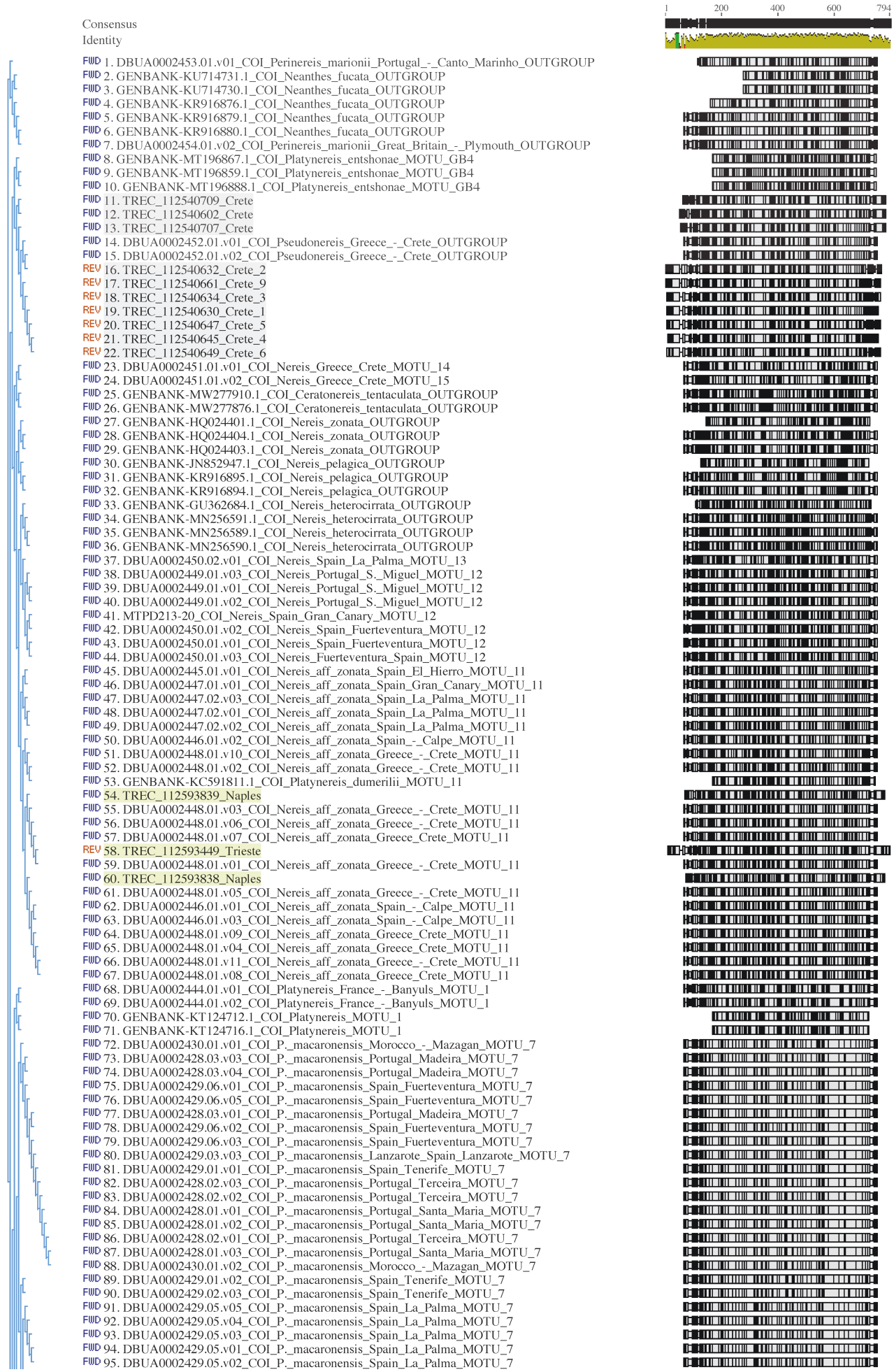
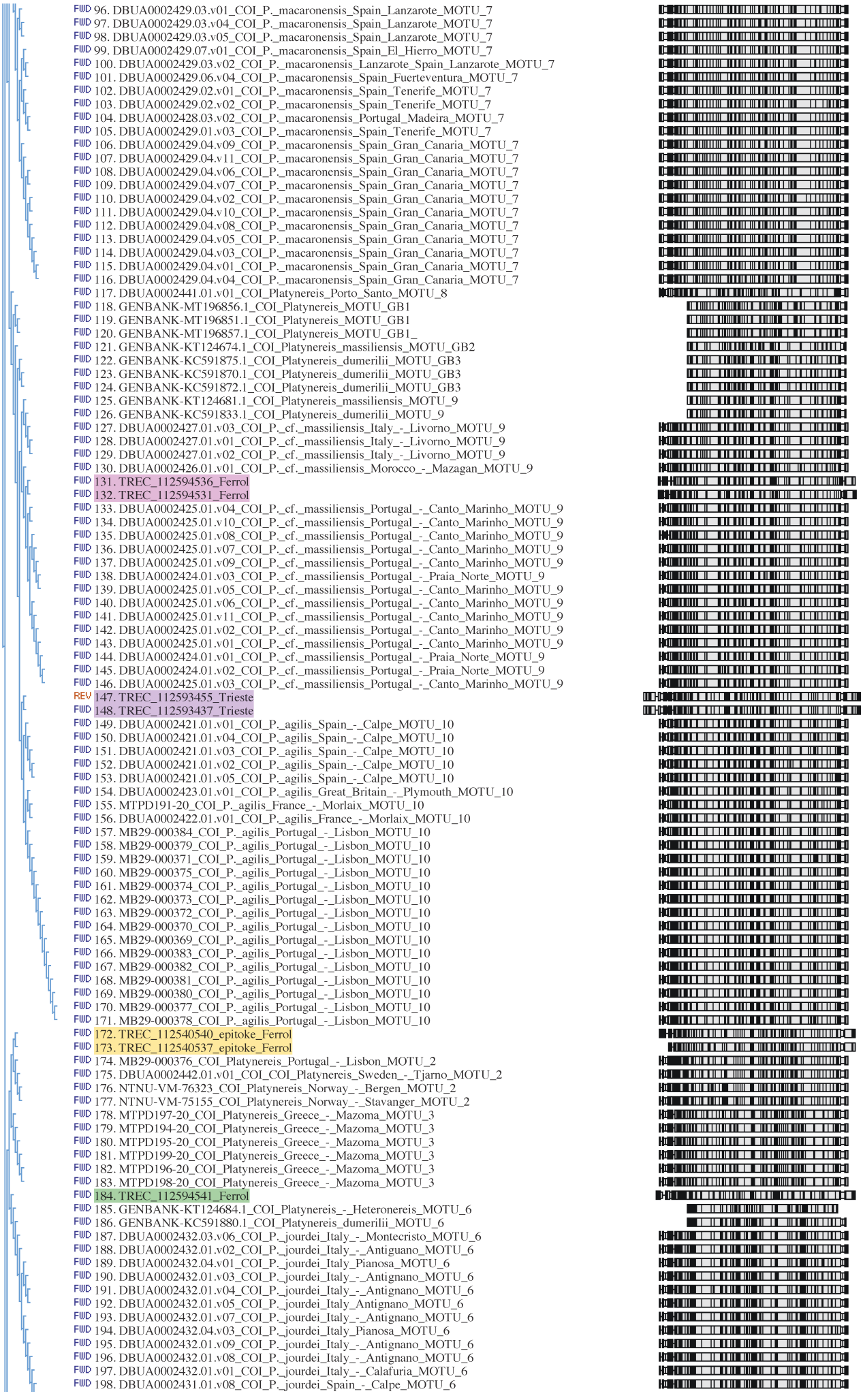


Figure S4. Cross - species integration of *P.dumerilii* and *O.vulgaris* datasets, done in 2023 on an older version of the *P.dumerilii* neuronal subset. (**A**) Integrated UMAP dimensional reduction coloured by species - ovul (*O.vulgaris*), pdum (*P.dumerilii*). Percent of cells/nuclei contributed by each species is listed. (**A′-A″**) Louvain clusters, labeled by percent contribution of the most prominent species - specific cell types. The bottom panel shows the distribution of *O.vulgaris* VL neurons and *P.dumerilii* MB neurons (**B**). Heatmaps showing top 50 (by average log fold change) OG markers for clusters 7 and 27 by species.


Consensus
Identity
1
200
400
600
794
FWD 1. DBUA0002453.01.v01_COI_Perinereis_marionii_Portugal_-_Canto_Marinho_OUTGROUP
FWD 2. GENBANK-KU714731.1_COI_Neanthes_fucata_OUTGROUP
FWD 3. GENBANK-KU714730.1_COI_Neanthes_fucata_OUTGROUP
FWD 4. GENBANK-KR916876.1_COI_Neanthes_fucata_OUTGROUP
FWD 5. GENBANK-KR916879.1_COI_Neanthes_fucata_OUTGROUP
FWD 6. GENBANK-KR916880.1_COI_Neanthes_fucata_OUTGROUP
FWD 7. DBUA0002454.01.v02_COI_Perinereis_marionii_Great_Britain_-_Plymouth_OUTGROUP
FWD 8. GENBANK-MT196867.1_COI_Platynereis_entshonae_MOTU_GB4
FWD 9. GENBANK-MT196859.1_COI_Platynereis_entshonae_MOTU_GB4
FWD 10. GENBANK-MT196888.1_COI_Platynereis_entshonae_MOTU_GB4
FWD 11. TREC_112540709_Crete
FWD 12. TREC_112540602_Crete
FWD 13. TREC_112540707_Crete
FWD 14. DBUA0002452.01.v01_COI_Pseudonereis_Greece_-_Crete_OUTGROUP
FWD 15. DBUA0002452.01.v02_COI_Pseudonereis_Greece_-_Crete_OUTGROUP
REV 16. TREC_112540632_Crete_2
REV 17. TREC_112540661_Crete_9
REV 18. TREC_112540634_Crete_3
REV 19. TREC_112540630_Crete_1
REV 20. TREC_112540647_Crete_5
REV 21. TREC_112540645_Crete_4
REV 22. TREC_112540649_Crete_6
FWD 23. DBUA0002451.01.v01_COI_Nereis_Greece_Crete_MOTU_14
FWD 24. DBUA0002451.01.v02_COI_Nereis_Greece_Crete_MOTU_15
FWD 25. GENBANK-MW277910.1_COI_Ceratonereis_tentaculata_OUTGROUP
FWD 26. GENBANK-MW277876.1_COI_Ceratonereis_tentaculata_OUTGROUP
FWD 27. GENBANK-HQ024401.1_COI_Nereis_zonata_OUTGROUP
FWD 28. GENBANK-HQ024404.1_COI_Nereis_zonata_OUTGROUP
FWD 29. GENBANK-HQ024403.1_COI_Nereis_zonata_OUTGROUP
FWD 30. GENBANK-JN852947.1_COI_Nereis_pelagica_OUTGROUP
FWD 31. GENBANK-KR916895.1_COI_Nereis_pelagica_OUTGROUP
FWD 32. GENBANK-KR916894.1_COI_Nereis_pelagica_OUTGROUP
FWD 33. GENBANK-GU362684.1_COI_Nereis_heterocirrata_OUTGROUP
FWD 34. GENBANK-MN256591.1_COI_Nereis_heterocirrata_OUTGROUP
FWD 35. GENBANK-MN256589.1_COI_Nereis_heterocirrata_OUTGROUP
FWD 36. GENBANK-MN256590.1_COI_Nereis_heterocirrata_OUTGROUP
FWD 37. DBUA0002450.02.v01_COI_Nereis_Spain_La_Palma_MOTU_13
FWD 38. DBUA0002449.01.v03_COI_Nereis_Portugal_S._Miguel_MOTU_12
FWD 39. DBUA0002449.01.v01_COI_Nereis_Portugal_S._Miguel_MOTU_12
FWD 40. DBUA0002449.01.v02_COI_Nereis_Portugal_S._Miguel_MOTU_12
FWD 41. MTPD213-20_COI_Nereis_Spain_Gran_Canary_MOTU_12
FWD 42. DBUA0002450.01.v02_COI_Nereis_Spain_Fuerteventura_MOTU_12
FWD 43. DBUA0002450.01.v01_COI_Nereis_Spain_Fuerteventura_MOTU_12
FWD 44. DBUA0002450.01.v03_COI_Nereis_Fuerteventura_Spain_MOTU_12
FWD 45. DBUA0002445.01.v01_COI_Nereis_aff_zonata_Spain_El_Hierro_MOTU_11
FWD 46. DBUA0002447.01.v01_COI_Nereis_aff_zonata_Spain_Gran_Canary_MOTU_11
FWD 47. DBUA0002447.02.v03_COI_Nereis_aff_zonata_Spain_La_Palma_MOTU_11
FWD 48. DBUA0002447.02.v01_COI_Nereis_aff_zonata_Spain_La_Palma_MOTU_11
FWD 49. DBUA0002447.02.v02_COI_Nereis_aff_zonata_Spain_La_Palma_MOTU_11
FWD 50. DBUA0002446.01.v02_COI_Nereis_aff_zonata_Spain_-_Calpe_MOTU_11
FWD 51. DBUA0002448.01.v10_COI_Nereis_aff_zonata_Greece_-_Crete_MOTU_11
FWD 52. DBUA0002448.01.v02_COI_Nereis_aff_zonata_Greece_-_Crete_MOTU_11
FWD 53. GENBANK-KC591811.1_COI_Platynereis_dumerilii_MOTU_11
FWD 54. TREC_112593839_Naples
FWD 55. DBUA0002448.01.v03_COI_Nereis_aff_zonata_Greece_-_Crete_MOTU_11
FWD 56. DBUA0002448.01.v06_COI_Nereis_aff_zonata_Greece_-_Crete_MOTU_11
FWD 57. DBUA0002448.01.v07_COI_Nereis_aff_zonata_Greece_Crete_MOTU_11
REV 58. TREC_112593449_Trieste
FWD 59. DBUA0002448.01.v01_COI_Nereis_aff_zonata_Greece_-_Crete_MOTU_11
FWD 60. TREC_112593838_Naples
FWD 61. DBUA0002448.01.v05_COI_Nereis_aff_zonata_Greece_-_Crete_MOTU_11
FWD 62. DBUA0002446.01.v01_COI_Nereis_aff_zonata_Spain_-_Calpe_MOTU_11
FWD 63. DBUA0002446.01.v03_COI_Nereis_aff_zonata_Spain_-_Calpe_MOTU_11
FWD 64. DBUA0002448.01.v09_COI_Nereis_aff_zonata_Greece_Crete_MOTU_11
FWD 65. DBUA0002448.01.v04_COI_Nereis_aff_zonata_Greece_Crete_MOTU_11
FWD 66. DBUA0002448.01.v11_COI_Nereis_aff_zonata_Greece_-_Crete_MOTU_11
FWD 67. DBUA0002448.01.v08_COI_Nereis_aff_zonata_Greece_Crete_MOTU_11
FWD 68. DBUA0002444.01.v01_COI_Platynereis_France_-_Banyuls_MOTU_1
FWD 69. DBUA0002444.01.v02_COI_Platynereis_France_-_Banyuls_MOTU_1
FWD 70. GENBANK-KT124712.1_COI_Platynereis_MOTU_1
FWD 71. GENBANK-KT124716.1_COI_Platynereis_MOTU_1
FWD 72. DBUA0002430.01.v01_COI_P._macaronensis_Morocco_-_Mazagan_MOTU_7
FWD 73. DBUA0002428.03.v03_COI_P._macaronensis_Portugal_Madeira_MOTU_7
FWD 74. DBUA0002428.03.v04_COI_P._macaronensis_Portugal_Madeira_MOTU_7
FWD 75. DBUA0002429.06.v01_COI_P._macaronensis_Spain_Fuerteventura_MOTU_7
FWD 76. DBUA0002429.06.v05_COI_P._macaronensis_Spain_Fuerteventura_MOTU_7
FWD 77. DBUA0002428.03.v01_COI_P._macaronensis_Portugal_Madeira_MOTU_7
FWD 78. DBUA0002429.06.v02_COI_P._macaronensis_Spain_Fuerteventura_MOTU_7
FWD 79. DBUA0002429.06.v03_COI_P._macaronensis_Spain_Fuerteventura_MOTU_7
FWD 80. DBUA0002429.03.v03_COI_P._macaronensis_Lanzarote_Spain_Lanzarote_MOTU_7
FWD 81. DBUA0002429.01.v01_COI_P._macaronensis_Spain_Tenerife_MOTU_7
FWD 82. DBUA0002428.02.v03_COI_P._macaronensis_Portugal_Terceira_MOTU_7
FWD 83. DBUA0002428.02.v02_COI_P._macaronensis_Portugal_Terceira_MOTU_7
FWD 84. DBUA0002428.01.v01_COI_P._macaronensis_Portugal_Santa_Maria_MOTU_7
FWD 85. DBUA0002428.01.v02_COI_P._macaronensis_Portugal_Santa_Maria_MOTU_7
FWD 86. DBUA0002428.02.v01_COI_P._macaronensis_Portugal_Terceira_MOTU_7
FWD 87. DBUA0002428.01.v03_COI_P._macaronensis_Portugal_Santa_Maria_MOTU_7
FWD 88. DBUA0002430.01.v02_COI_P._macaronensis_Morocco_-_Mazagan_MOTU_7
FWD 89. DBUA0002429.01.v02_COI_P._macaronensis_Spain_Tenerife_MOTU_7
FWD 90. DBUA0002429.02.v03_COI_P._macaronensis_Spain_Tenerife_MOTU_7
FWD 91. DBUA0002429.05.v05_COI_P._macaronensis_Spain_La_Palma_MOTU_7
FWD 92. DBUA0002429.05.v04_COI_P._macaronensis_Spain_La_Palma_MOTU_7
FWD 93. DBUA0002429.05.v03_COI_P._macaronensis_Spain_La_Palma_MOTU_7
FWD 94. DBUA0002429.05.v01_COI_P._macaronensis_Spain_La_Palma_MOTU_7
FWD 95. DBUA0002429.05.v02_COI_P._macaronensis_Spain_La_Palma_MOTU_7


FWD 96. DBUA0002429.03.v01_COI_P._macaronensis_Spain_Lanzarote_MOTU_7
FWD 97. DBUA0002429.03.v04_COI_P._macaronensis_Spain_Lanzarote_MOTU_7
FWD 98. DBUA0002429.03.v05_COI_P._macaronensis_Spain_Lanzarote_MOTU_7
FWD 99. DBUA0002429.07.v01_COI_P._macaronensis_Spain_El_Hierro_MOTU_7
FWD 100. DBUA0002429.03.v02_COI_P._macaronensis_Lanzarote_Spain_Lanzarote_MOTU_7
FWD 101. DBUA0002429.06.v04_COI_P._macaronensis_Spain_Fuerteventura_MOTU_7
FWD 102. DBUA0002429.02.v01_COI_P._macaronensis_Spain_Tenerife_MOTU_7
FWD 103. DBUA0002429.02.v02_COI_P._macaronensis_Spain_Tenerife_MOTU_7
FWD 104. DBUA0002428.03.v02_COI_P._macaronensis_Portugal_Madeira_MOTU_7
FWD 105. DBUA0002429.01.v03_COI_P._macaronensis_Spain_Tenerife_MOTU_7
FWD 106. DBUA0002429.04.v09_COI_P._macaronensis_Spain_Gran_Canaria_MOTU_7
FWD 107. DBUA0002429.04.v11_COI_P._macaronensis_Spain_Gran_Canaria_MOTU_7
FWD 108. DBUA0002429.04.v06_COI_P._macaronensis_Spain_Gran_Canaria_MOTU_7
FWD 109. DBUA0002429.04.v07_COI_P._macaronensis_Spain_Gran_Canaria_MOTU_7
FWD 110. DBUA0002429.04.v02_COI_P._macaronensis_Spain_Gran_Canaria_MOTU_7
FWD 111. DBUA0002429.04.v10_COI_P._macaronensis_Spain_Gran_Canaria_MOTU_7
FWD 112. DBUA0002429.04.v08_COI_P._macaronensis_Spain_Gran_Canaria_MOTU_7
FWD 113. DBUA0002429.04.v05_COI_P._macaronensis_Spain_Gran_Canaria_MOTU_7
FWD 114. DBUA0002429.04.v03_COI_P._macaronensis_Spain_Gran_Canaria_MOTU_7
FWD 115. DBUA0002429.04.v01_COI_P._macaronensis_Spain_Gran_Canaria_MOTU_7
FWD 116. DBUA0002429.04.v04_COI_P._macaronensis_Spain_Gran_Canaria_MOTU_7
FWD 117. DBUA0002441.01.v01_COI_Platynereis_Porto_Santo_MOTU_8
FWD 118. GENBANK-MT196856.1_COI_Platynereis_MOTU_GB1
FWD 119. GENBANK-MT196851.1_COI_Platynereis_MOTU_GB1
FWD 120. GENBANK-MT196857.1_COI_Platynereis_MOTU_GB1_
FWD 121. GENBANK-KT124674.1_COI_Platynereis_massiliensis_MOTU_GB2
FWD 122. GENBANK-KC591875.1_COI_Platynereis_dumerilii_MOTU_GB3
FWD 123. GENBANK-KC591870.1_COI_Platynereis_dumerilii_MOTU_GB3
FWD 124. GENBANK-KC591872.1_COI_Platynereis_dumerilii_MOTU_GB3
FWD 125. GENBANK-KT124681.1_COI_Platynereis_massiliensis_MOTU_9
FWD 126. GENBANK-KC591833.1_COI_Platynereis_dumerilii_MOTU_9
FWD 127. DBUA0002427.01.v03_COI_P._cf._massiliensis_Italy_-_Livorno_MOTU_9
FWD 128. DBUA0002427.01.v01_COI_P._cf._massiliensis_Italy_-_Livorno_MOTU_9
FWD 129. DBUA0002427.01.v02_COI_P._cf._massiliensis_Italy_-_Livorno_MOTU_9
FWD 130. DBUA0002426.01.v01_COI_P._cf._massiliensis_Morocco_-_Mazagan_MOTU_9
FWD 131. TREC_112594536_Ferrol
FWD 132. TREC_112594531_Ferrol
FWD 133. DBUA0002425.01.v04_COI_P._cf._massiliensis_Portugal_-_Canto_Marinho_MOTU_9
FWD 134. DBUA0002425.01.v10_COI_P._cf._massiliensis_Portugal_-_Canto_Marinho_MOTU_9
FWD 135. DBUA0002425.01.v08_COI_P._cf._massiliensis_Portugal_-_Canto_Marinho_MOTU_9
FWD 136. DBUA0002425.01.v07_COI_P._cf._massiliensis_Portugal_-_Canto_Marinho_MOTU_9
FWD 137. DBUA0002425.01.v09_COI_P._cf._massiliensis_Portugal_-_Canto_Marinho_MOTU_9
FWD 138. DBUA0002424.01.v03_COI_P._cf._massiliensis_Portugal_-_Praia_Norte_MOTU_9
FWD 139. DBUA0002425.01.v05_COI_P._cf._massiliensis_Portugal_-_Canto_Marinho_MOTU_9
FWD 140. DBUA0002425.01.v06_COI_P._cf._massiliensis_Portugal_-_Canto_Marinho_MOTU_9
FWD 141. DBUA0002425.01.v11_COI_P._cf._massiliensis_Portugal_-_Canto_Marinho_MOTU_9
FWD 142. DBUA0002425.01.v02_COI_P._cf._massiliensis_Portugal_-_Canto_Marinho_MOTU_9
FWD 143. DBUA0002425.01.v01_COI_P._cf._massiliensis_Portugal_-_Canto_Marinho_MOTU_9
FWD 144. DBUA0002424.01.v01_COI_P._cf._massiliensis_Portugal_-_Praia_Norte_MOTU_9
FWD 145. DBUA0002424.01.v02_COI_P._cf._massiliensis_Portugal_-_Praia_Norte_MOTU_9
FWD 146. DBUA0002425.01.v03_COI_P._cf._massiliensis_Portugal_-_Canto_Marinho_MOTU_9
REV 147. TREC_112593455_Trieste
FWD 148. TREC_112593437_Trieste
FWD 149. DBUA0002421.01.v01_COI_P._agilis_Spain_-_Calpe_MOTU_10
FWD 150. DBUA0002421.01.v04_COI_P._agilis_Spain_-_Calpe_MOTU_10
FWD 151. DBUA0002421.01.v03_COI_P._agilis_Spain_-_Calpe_MOTU_10
FWD 152. DBUA0002421.01.v02_COI_P._agilis_Spain_-_Calpe_MOTU_10
FWD 153. DBUA0002421.01.v05_COI_P._agilis_Spain_-_Calpe_MOTU_10
FWD 154. DBUA0002423.01.v01_COI_P._agilis_Great_Britain_-_Plymouth_MOTU_10
FWD 155. MTPD191-20_COI_P._agilis_France_-_Morlaix_MOTU_10
FWD 156. DBUA0002422.01.v01_COI_P._agilis_France_-_Morlaix_MOTU_10
FWD 157. MB29-000384_COI_P._agilis_Portugal_-_Lisbon_MOTU_10
FWD 158. MB29-000379_COI_P._agilis_Portugal_-_Lisbon_MOTU_10
FWD 159. MB29-000371_COI_P._agilis_Portugal_-_Lisbon_MOTU_10
FWD 160. MB29-000375_COI_P._agilis_Portugal_-_Lisbon_MOTU_10
FWD 161. MB29-000374_COI_P._agilis_Portugal_-_Lisbon_MOTU_10
FWD 162. MB29-000373_COI_P._agilis_Portugal_-_Lisbon_MOTU_10
FWD 163. MB29-000372_COI_P._agilis_Portugal_-_Lisbon_MOTU_10
FWD 164. MB29-000370_COI_P._agilis_Portugal_-_Lisbon_MOTU_10
FWD 165. MB29-000369_COI_P._agilis_Portugal_-_Lisbon_MOTU_10
FWD 166. MB29-000383_COI_P._agilis_Portugal_-_Lisbon_MOTU_10
FWD 167. MB29-000382_COI_P._agilis_Portugal_-_Lisbon_MOTU_10
FWD 168. MB29-000381_COI_P._agilis_Portugal_-_Lisbon_MOTU_10
FWD 169. MB29-000380_COI_P._agilis_Portugal_-_Lisbon_MOTU_10
FWD 170. MB29-000377_COI_P._agilis_Portugal_-_Lisbon_MOTU_10
FWD 171. MB29-000378_COI_P._agilis_Portugal_-_Lisbon_MOTU_10
FWD 172. TREC_112540540_epitoke_Ferrol
FWD 173. TREC_112540537_epitoke_Ferrol
FWD 174. MB29-000376_COI_Platynereis_Portugal_-_Lisbon_MOTU_2
FWD 175. DBUA0002442.01.v01_COI_COI_Platynereis_Sweden_-_Tjarno_MOTU_2
FWD 176. NTNU-VM-76323_COI_Platynereis_Norway_-_Bergen_MOTU_2
FWD 177. NTNU-VM-75155_COI_Platynereis_Norway_-_Stavanger_MOTU_2
FWD 178. MTPD197-20_COI_Platynereis_Greece_-_Mazoma_MOTU_3
FWD 179. MTPD194-20_COI_Platynereis_Greece_-_Mazoma_MOTU_3
FWD 180. MTPD195-20_COI_Platynereis_Greece_-_Mazoma_MOTU_3
FWD 181. MTPD199-20_COI_Platynereis_Greece_-_Mazoma_MOTU_3
FWD 182. MTPD196-20_COI_Platynereis_Greece_-_Mazoma_MOTU_3
FWD 183. MTPD198-20_COI_Platynereis_Greece_-_Mazoma_MOTU_3
FWD 184. TREC_112594541_Ferrol
FWD 185. GENBANK-KT124684.1_COI_Platynereis_-_Heteronereis_MOTU_6
FWD 186. GENBANK-KC591880.1_COI_Platynereis_dumerilii_MOTU_6
FWD 187. DBUA0002432.03.v06_COI_P._jourdei_Italy_-_Montecristo_MOTU_6
FWD 188. DBUA0002432.01.v02_COI_P._jourdei_Italy_-_Antiguano_MOTU_6
FWD 189. DBUA0002432.04.v01_COI_P._jourdei_Italy_Pianosa_MOTU_6
FWD 190. DBUA0002432.01.v03_COI_P._jourdei_Italy_-_Antignano_MOTU_6
FWD 191. DBUA0002432.01.v04_COI_P._jourdei_Italy_-_Antignano_MOTU_6
FWD 192. DBUA0002432.01.v05_COI_P._jourdei_Italy_Antignano_MOTU_6
FWD 193. DBUA0002432.01.v07_COI_P._jourdei_Italy_-_Antignano_MOTU_6
FWD 194. DBUA0002432.04.v03_COI_P._jourdei_Italy_Pianosa_MOTU_6
FWD 195. DBUA0002432.01.v09_COI_P._jourdei_Italy_-_Antignano_MOTU_6
FWD 196. DBUA0002432.01.v08_COI_P._jourdei_Italy_-_Antignano_MOTU_6
FWD 197. DBUA0002432.01.v01_COI_P._jourdei_Italy_-_Calafuria_MOTU_6
FWD 198. DBUA0002431.01.v08_COI_P._jourdei_Spain_-_Calpe_MOTU_6

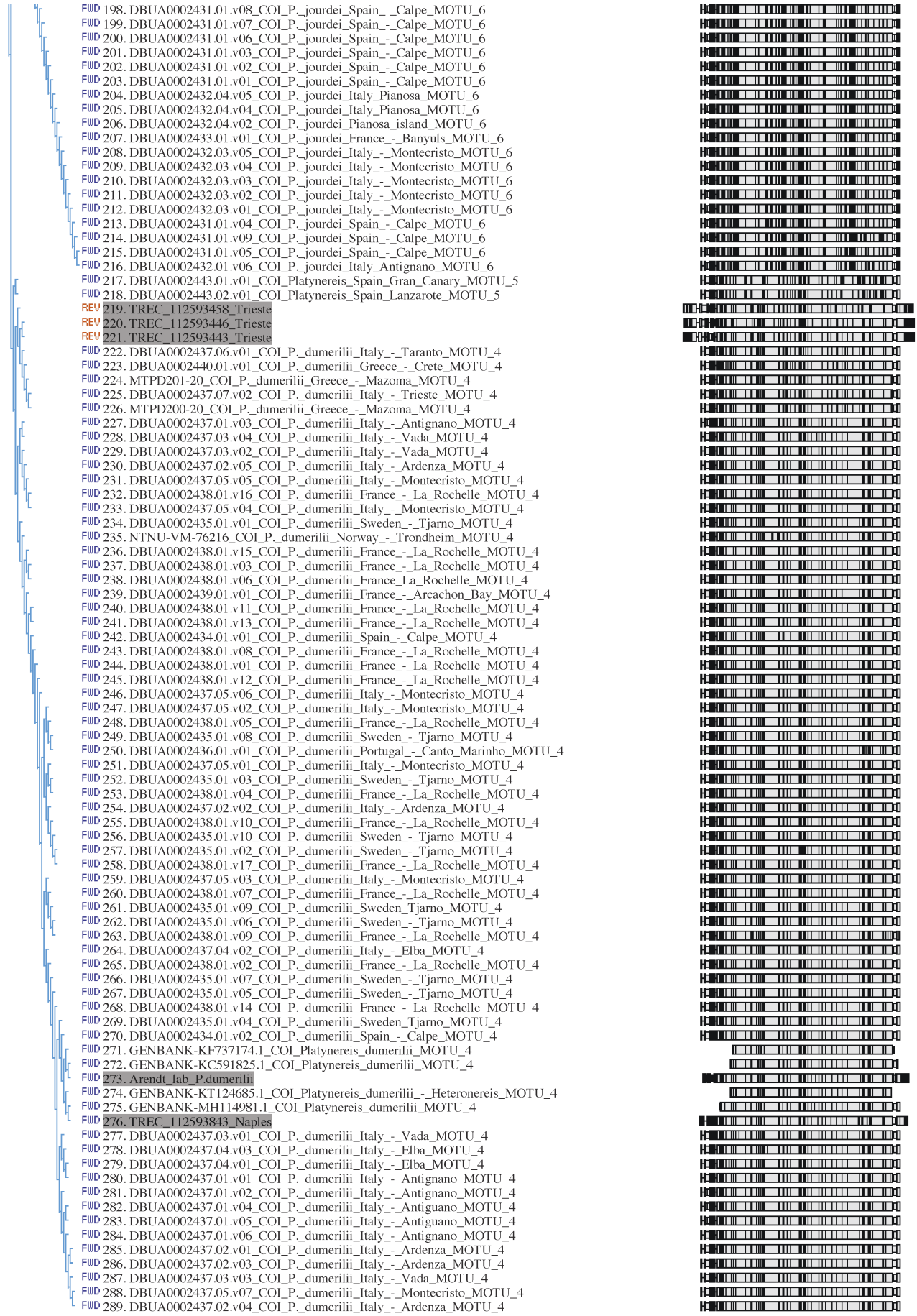


Figure S5. Full sequence alignment and neighbour - joining tree of TREC samples sequenced so far embedded among CO1 sequences from Teixeira et al. (2022).

Supplementary Tables

Table S1. Coding sequences (CDS) or open - reading frames extracted from the *P.dumerilii* genome submitted to Molecular Instruments for HCR probe design. Accession numbers for CDS in **Table S2** if the sequence is already published. Relevant for Chapter 2. Gene names are in the form XLOC | emapper annotation | *known Pdum name.*

Gene name	CDS	Probe sequences
XLOC_044659 | ARX | *arx*	AUggUAAUUgCUAAUUgCCAUUgUgACgUCACCAUg gAUCUAUCAAUUAAgAAACAACgCACAgCUUACgA CAUAgCAAgCCUAAUAggCCCCgAAgCUACAAAAgU UgAAAAAUggACUUCAUCAgAAAgCCCCCCUACUUC UCCCCCggggAgACCCAgCUCAAUUUCCCggCAggAU gACUUCgUCgCCAACUCUUCAgCCCCggAgCAUCCCC CUUCCCCCCCUUAgUgCUgAUgACCAAUACAAAAAA CUgAUgUCACCCACAAUgCAUCUggCUCAUgACgUC ACAgCUCgAAUCAUUCUAgCCAAUCACAgCCAAgAg UUgCAgUCACgUgAUUAUgCUUAUgACCAAUCACAA AggAggCAAAgUgUgAUgUCAUCAgACAgUgCUgCCA CCAgCCUgAgAAAUgAUgAUAUAgAUgAUgAUgAAC AgUUAUgUAUUgUggAUgAUgAUgAACCACUAAgUC CUgUAAAUACCAgCCCAgUgCAgCCCUCggCCUCgAU CCAggCACCgAAAgAAgUCAAAgUCCAggAAgACAA CUUAUCgAUCUCCgAUggCgAUggCgACUUggAUgAA AUgggAAAACgCAAACAAAgACgAUAUCgAACAACg UUCACCUCCUUCCAgUUAgAAgAAUUAgAAAgggCC UUCCAgAAgACgCACUACCCUgAUgUgUUCACUAgA gAggAACUggCCAUgAgAAUCAACCUAACCgAAgCg AgAgUUCAggUUUggUUUCAAAACAgACgAgCCAAg UggAggAAgAAggAgAAAgUggggCCACAAUCgCAUC CCUACggCCCUUUUgggggCggCCCUCCgACCgggCCA UUgggCAUCCCAgggUUgggCCAAAgACCCCUCggCC CACACACCACUUACACAgACUUgUUgAUCAAAUCU UACgAAAACCAUCUCgCUCAgAAgUUUggUggUCCA ggUAUgggUggAUUCUAUCCACCAggAUUAgggggAC CggUCgCUgCCAUgggAgUCUUUCCCggAUUgCCCUA UUCCggAUUgggUCAUCUgCAUCCggggAgUCACggA gCggCCgCAgUCgUAgCCggAAgCUUUCAgAgUCUCC UggCggCCAUgUCgAAUCCCgUCCCCggUggACAACC UCACUCUCACUCUCAUCCUCAUgCgCCCCCUgUgCA AgUUCCAACCUCgCCAAACgUAAACgUUgCUAUUCC CCAgggUUCgCCgAACCUAACACCACAAAAUUUACg gCggCCUggUUCggAgUCACCAgAUUCAAACAAAACg CCCUCCCCCCCACUCACCCCCUgUgUUgUACCCUCAg CAggACACCCCCAgAAgUgACgUCAUCAAAUCgAAU CCAgAAUCCCCCUCUgAgCAUgCUCAAgACAgACgA UCCUCUAgCAUAgCAgCAUUgCgCAUgCgUgCCAggg AgUAUgCCCUCAAgUUgCAgAUgggCCAACCAUgUg ACAgUgCAgUCUAUUgA	TCAATTTCCCggCAggATgACTTCgTCgCCAACTCTTCAgCCCCggAgCATC CATCTggCTCATgACgTCACAgCTCgAATCATTCTAgCCAATCACAgCCAAg CAACgCACAgCTTACgACATAgCAAgCCTAATAggCCCCgAAgCTACAAAAg AACTTATCgATCTCCgATggCgATggCgACTTggATgAAATgggAAAACgCA TTgCAgTCACgTgATTATgCTTATgACCAATCACAAAggAggCAAAgTgTgA gTgCAgCCCTCggCCTCgATCCAggCACCgAAAgAAgTCAAAgTCCAggAAg ATggTAATTgCTAATTgCCATTgTgACgTCACCATggATCTATCAATTAAgA CAgTTATgTATTgTggATgATgATgAACCACTAAgTCCTgTAAATACCAgCC CAAAgACgATATCgAACAACgTTCACCTCCTTCCAgTTAgAAgAATTAgAAA CCTTCCCCCCTTAgTgCTgATgACCAATACAAAAAACTgATgTCACCCACAA gAAAAATggACTTCATCAgAAAgCCCCCCTACTTCTCCCCCggggAgACCCA AAgTggAggAAgAAggAgAAAgTggggCCACAATCgCATCCCTACggCCCTT gggggACCggTCgCTgCCATgggAgTCTTTCCCggATTgCCCTATTCCggAT ggTCATCTgCATCCggggAgTCACggAgCggCCgCAgTCgTAgCCggAAgCT CACTCTCATCCTCATgCgCCCCCTgTgCAAgTTCCAACCTCgCCAAACgTAA ATgCgTgCCAgggAgTATgCCCTCAAgTTgCAgATgggCCAACCATgTgACA CTCgCTCAgAAgTTTggTggTCCAggTATgggTggATTCTATCCACCAggAT gTTgCTATTCCCCAgggTTCgCCgAACCTAACACCACAAAATTTACggCggC CTCggCCCACACACCACTTACACAgACTTgTTgATCAAATCTTACgAAAACC CAgAgTCTCCTggCggCCATgTCgAATCCCgTCCCCggTggACAACCTCACT ggTTCggAgTCACCAgATTCAAACAAAACgCCCTCCCCCCACTCACCCCCTg TCCCCCTCTgAgCATgCTCAAgACAgACgATCCTCTAgCATAgCAgCATTgC TTgTACCCTCAgCAggACACCCCCAgAAgTgACgTCATCAAATCgAATCCAg
XLOC_003959 | PTF1A | *ptf1*	ATGGATGGAGTGGACCTGGTAGCACGGACACTCATA GCGGACTACCCGGGCTACTCCTCCTACAACGACCTAG ACAGCTCCCTCAGTTCCGGAAGTTCCCCCTACGGATC TACCCTCGACGAGGACGAGGCCTTTTCCGGCAAAGT ACGACACCGACGGCGCCGACGGTGCCCCGAGCAGC AGGTCCACCAACGACAGGCGGCCAACCAGCGTGAA CGGCGCCGGATGCAGTCGATCAACGAGGCCTTCGAG GGCCTACGAGCGCACATCCCCACCCTGCCCTACGAG AAGCGACTCTCGAAAGTAGACACTCTCCGCTTAGCC ATCGGATATATCAGTTTCCTGACAGAACTTGTCCAGTC GGACGCTCACTCGAAGGAGAACATCGGCACTCAGTG TGGGGAACAGCCGAGGAAGGTCATTATCCATTGCCA CAGAGGCTACCAAGATGAAGGAGAAAGATACGGCCT CCCCCCGTTAGCAGGCCACTCGCTGTCTTGGACGGAT GAGAAGAAACCACTGAGGGGTCCGGGGAATGTCATG ACTGCTAAAATATGGACTCCAGAAGATCCAAGGAAA ATCGGAAATATGGACTGTCAAACAACTGTTCTGGATA CTTCCACGTTGGAATTATGA	ATggATggAgTggACCTggTAgCACggACACTCATAgCggACTACCCgggCT TCCTCCTACAACgACCTAgACAgCTCCCTCAgTTCCggAAgTTCCCCCTACg TCTACCCTCgACgAggACgAggCCTTTTCCggCAAAgTACgACACCgACggC CgACggTgCCCCgAgCAgCAggTCCACCAACgACAggCggCCAACCAgCgTg ATCggATATATCAgTTTCCTgACAgAACTTgTCCAgTCggACgCTCACTCgA CACAgAggCTACCAAgATgAAggAgAAAgATACggCCTCCCCCCgTTAgCAg CACTCgCTgTCTTggACggATgAgAAgAAACCACTgAggggTCCggggAATg ATgACTgCTAAAATATggACTCCAgAAgATCCAAggAAAATCggAAATATgg CCCACCCTgCCCTACgAgAAgCgACTCTCgAAAgTAgACACTCTCCgCTTAg

XLOC_040379 | NCALD | *NA*	ATGGGCCATGGCAGCTCAGTGGATCTGCGCCCGGAG GACCTCGTGAAACTGAAAGACCAGACTTACTTCCAA CAAAATGAAATTCAGACGTTGTACGACAAATTCATCT CGGAATTTCCATCTGGAATTATTACCAAGGAAGAATT CATTGAAATGTACCAGGACATGTTTCCCGAGTCAGAT GCCTCTCCGTTTGCGGAAAAAGTCTTCAGAGCGTATG ATACGGACGGAAACGGAGTGATAAACTTCCGAGAAT TCCTTTGCGCTTTGTCGGTCACCACCAGAGGGAGTGC AGAAGAAAAACTCGGCTGGAGTTTCAATTTGTACGA CGCAGACGGCGATGGGTATATTTCAAGAAAGGAAGC TACTGACATTTTGGCAGCGATATTCCGACTCCAAGGC AAAGCGGACATTCCAGGGAAAGCGGAAGAAGAGGC TCTCAATTTATTTTCATCTATCGACAAAAACAAGGAC GACCGTTTGTCTTTGAAAGAGTTTGTTTGCGGGGCCA AGACATCAGACGCCATTGCGCAGTTCCTCAGAATTGC CACAGAGGGCGAATAG	ATgggCCATggCAgCTCAgTggATCTgCgCCCggAggACCTCgTgAAACTgA gACCAgACTTACTTCCAACAAAATgAAATTCAgACgTTgTACgACAAATTCA TCggAATTTCCATCTggAATTATTACCAAggAAgAATTCATTgAAATgTACC gACATgTTTCCCgAgTCAgATgCCTCTCCgTTTgCggAAAAAgTCTTCAgAg TATgATACggACggAAACggAgTgATAAACTTCCgAgAATTCCTTTgCgCTT TCggTCACCACCAgAgggAgTgCAgAAgAAAAACTCggCTggAgTTTCAATT TACgACgCAgACggCgATgggTATATTTCAAgAAAggAAgCTACTgACATTT gCAgCgATATTCCgACTCCAAggCAAAgCggACATTCCAgggAAAgCggAAg gAggCTCTCAATTTATTTTCATCTATCgACAAAAACAAggACgACCgTTTgT TTgAAAgAgTTTgTTTgCggggCCAAgACATCAgACgCCATTgCgCAgTTCC
XLOC_069317 | EMX2 | *NA*	ATGCTGCCCGTGGTGCCCTCGGCCAAAAGGGCCAGT GGCTTCAGCATAGACTCGTTGATGTCGAAGGACAGG ACAAGTCCCAGGAGCCCTCCGACTGTCAGCTCAGTTC CAAGGGTGTCTTCGGCCTCCCCTCCTGCTTCTAGTTCG GTCCCGAGTTCAGCTCCTAGACTGAGTTTGAGTCCAG AGAGCATCTCGGCTGCTGCCGCTCATCATCACAGAGA ACGGGGGGAGTTGAGAGGGGAGATGCGGGGTGATG GCAGCAGACTGGATCCCCCCAGAGACAGAGACGGG CCCCGGGAGCAACACCATCCCCGTGATCACCCTGTCA GAGAACACACAATCGGAGGTTTGGGGTCAAGTTTTC ACCCTGTGCACCCAAGCTCTGCTTTACTGAGTGGGTT AAAGAACTTGCACGGGGGCTCAGGGCTAGGGCCGC ATCCAGAATCGTACCCAGAGGGGCTGGCTGCTCTTGG ACTGGCACCGGCAGCTCTCCAGGCCGGGGCTGCTGC GCACCTCCCCCCTGGCCTGACCTTTCCCGGGGGCCTC CCAGGCCCCCACCATGGACTACCTCACCACCCCGTGT TCCTTGGGGCCGCCCAAAGAGACCCCCTAACCCTCTA CCCCTGGCTACTCAGCAGACACGGCTCCTACTTCGGC CACCCCAGGCTAGGAGGTCACGAAGCAAACTTCCTG TTCCAACCCTTCCGGAAGCCGAAGCGGATCAGGACG GCCTTCAGTCCTTCTCAACTCTTGAGACTGGAACACG CTTTTGAAAAGAACCACTACGTTGTAGGCCAGGAGA GGAAGGATCTAGCTGCTAATCTTAGCTTAACGGAAAC ACAGGTCAGTCATATTGTAATATTGTTGTTTTCTTGCC TAATTTGA	TACTTCGGCCACCCCAGGCTAGGAGGTCACGAAGCAAACTTCCTGTTCCAAC ATGCTGCCCGTGGTGCCCTCGGCCAAAAGGGCCAGTGGCTTCAGCATAGACT TTGATGTCGAAGGACAGGACAAGTCCCAGGAGCCCTCCGACTGTCAGCTCAG CCAAGGGTGTCTTCGGCCTCCCCTCCTGCTTCTAGTTCGGTCCCGAGTTCAG AGAGAACGGGGGGAGTTGAGAGGGGAGATGCGGGGTGATGGCAGCAGACTGG CCCCCCAGAGACAGAGACGGGCCCCGGGAGCAACACCATCCCCGTGATCACC GTCAGAGAACACACAATCGGAGGTTTGGGGTCAAGTTTTCACCCTGTGCACC AGCTCTGCTTTACTGAGTGGGTTAAAGAACTTGCACGGGGGCTCAGGGCTAG CCGCATCCAGAATCGTACCCAGAGGGGCTGGCTGCTCTTGGACTGGCACCGG GCTCTCCAGGCCGGGGCTGCTGCGCACCTCCCCCCTGGCCTGACCTTTCCCG GGCCTCCCAGGCCCCCACCATGGACTACCTCACCACCCCGTGTTCCTTGGGG GCCCAAAGAGACCCCCTAACCCTCTACCCCTGGCTACTCAGCAGACACGGCT TTCCGGAAGCCGAAGCGGATCAGGACGGCCTTCAGTCCTTCTCAACTCTTGA CTAGCTGCTAATCTTAGCTTAACGGAAAACACAGGTCAGTCATATTGTAATAT CCTAGACTGAGTTTGAGTCCAGAGAGCATCTCGGCTGCTGCCGCTCATCATC
XLOC_000048 | MEF2c | *NA*	AUgggACgCAAAAAgAUCCAgAUAUCCAggAUUggA gACgAgCgAAACAgACAggUUACAUUCACAAAgAgA AAAUUUggCUUgAUgAAgAAAgCCUACgAgUUgAgC gUgCUgUgUgAUUgUgAgAUCgCCUUgAUCAUCUUC AACUCUgCCAACAAgCUAUUCCAgUAUgCCAgCACA gACAUggACAAggUgCUCCUCAAAUACACUgAgUAC AAUgAgCCCCAUgAgAgUAggACCAACAAAgAUAUC AUUgAggCgUUgCAUAAAAAAgAAAACAAAggACAA gAAAgCCCggAUgUUgACCAggACUCCCAgUACAUU CUCACCCCACggACUgAAgAgAAgUUUCAgAAAAUC ggACAAgAAUUUgAUAUgAUgAUgCACAggAAUgUC AUgAACAACAgAggUCUACCAggAUACCAgAACCAA ggAAUggCgggAAUgAUgCCCAUgUCAggACACUACC ACCAAggACACCCCggACAgACCCUgAUgCCgCCCCA CACCAUgUCACAACAAggAUCUCUUAgCCCACgACC CAACUCUACUggCggAAUgAUggACAUgAACAACAg UAACggAUACggAggAgCCggAUCgCCCAACCACCAC CCAAUggCCggAAgUCCUUCgCCCggAAUgAUgAACA gAUCgAAACACAgCCCACCgUCACCCggAUCCggACC gAgAAACAACUUgAgAgUggCCAUCCCCAACUCCCg UUCAgACCUgCCggUCAgUgAAgAUggUggACUCAgU gACgAgAgUUUggUggAUUCAgAUUAUCUCACgCAg CUUggCCCAgCCAgAgAUCUCUCUCUgAAUCACCUg AAUCgAAgUAgUCACAACUCUAUgggAACCCCAgCU gUCUCCAUggCAACCCCAAgCAUgACCggAUCUggA AACUUCUCUgggUUgUCCCCUUUUCCgggUgACUUC CCUUCAAUgAACAgCggCgAUUUACAUggACUAggA AACUUCAgCUCUUCgggACUUCUCUCAggACAUUgg UCgAAUCAgAUgUCUAUCACUUCAgCgCUACAACAC ggAAUgAACAACAACUCUAUgAACCUAUCAgUgAgC ACggCAAACAACggAAACAUgAgCAUCAAgAgCgAA CCgAUAUCACCgCCUCgCgAgUCgACCACgCCAACUA	ATgggACgCAAAAAgATCCAgATATCCAggATTggAgACgAgCgAAACAgAC gTgAgCACggCAAACAACggAAACATgAgCATCAAgAgCgAACCgATATCAC TCTATCACTTCAgCgCTACAACACggAATgAACAACAACTCTATgAACCTAT CTAggAAACTTCAgCTCTTCgggACTTCTCTCAggACATTggTCgAATCAgA TTgTCCCCTTTTCCgggTgACTTCCCTTCAATgAACAgCggCgATTTACATg AgAgATCTCTCTCTgAATCACCTgAATCgAAgTAgTCACAACTCTATgggAA CTCAgTgACgAgAgTTTggTggATTCAgATTATCTCACgCAgCTTggCCCAg AgAgTggCCATCCCCAACTCCCgTTCAgACCTgCCggTCAgTgAAgATggTg AACAgATCgAAACACAgCCCACCgTCACCCggATCCggACCgAgAAACAACT gCCggATCgCCCAACCACCACCCAATggCCggAAgTCCTTCgCCCggAATgA CCCAACTCTACTggCggAATgATggACATgAACAACAgTAACggATACggAg CAgACCCTgATgCCgCCCCACACCATgTCACAACAAggATCTCTTAgCCCAC ggAATggCgggAATgATgCCCATgTCAggACACTACCACCAAggACACCCCg ATgATgCACAggAATgTCATgAACAACAgAggTCTACCAggATACCAgAACC CTCACCCCACggACTgAAgAgAAgTTTCAgAAAATCggACAAgAATTTgATA AATgAgCCCCATgAgAgTAggACCAACAAAgATATCATTgAggCgTTgCATA TTCCAgTATgCCAgCACAgACATggACAAggTgCTCCTCAAATACACTgAgT

	gUCAgCAUCACCUgCgACCACCgUCAAACggACACA UgAgCgggCACgUUUCACCgCACCACAUgggCggCCA UgUUUCACCACAUCACAUgAgCggACACgUUUCACC gCACCACAUgAgUAACCAUAgCAACUCAUCgUCUCC AgUggggCUCAACggCggCACCCCAACggACUUUgAC gggCCgAgUAUgAAgCgCCCCCgCCUUgCCgAgUgggU ggCCAgUUAg	gTgCTgTgTgATTgTgAgATCgCCTTgATCATCTTCAACTCTgCCAACAAgC ggACACATgAgCgggCACgTTTCACCgCACCACATgggCggCCATgTTTCAC CATCACATgAgCggACACgTTTCACCgCACCACATgAgTAACCATAgCAACT
XLOC_012133 | - | *cOpsin1*	ATGGATGGGGAAAATCTCACAATTCCAAATCCAGTGA CAGAATTAATGGACACACCTATAAATTCCACTTATTTT CAAAATTTAAATGCAGAAACTGATGGTGGGAATCATT ACATCTACAACGCATTCACTGCCACAGACTACAACAT CTGTGCAGCGTACCTATTCTTCATTGCCTGTCTTGGCG TATCTTTGAACGTCTTAGTACTAGTTCTTTTCATCAAA GACAGAAAGCTGAGATCCCCTAATAATTTCCTGTATG TTAGTCTTGCCCTTGGTGATCTCTTAGTGGCCGTCTTT GGAACTGCTTTCAAGTTTATCATCACAGCCAGGAAA ACTTTACTCAGAGAAGAAGATGGCTTCTGCAAGTGG TATGGCTTCATCACCTACTTGGGAGGTCTTGCAGCCC TGATGACTCTTTCTGTAATTGCCTTTGTACGATGCTTA GCCGTCCTGAGGCTTGGAAGTTTCACGGGCCTCACC ACGAGAATGGGAGTGGCTGCGATGGCCTTCATTTGG ATATACTCCCTTGCCTTTACTTTGGCACCACTTCTTGG ATGGAATCATTACATACCTGAGGGTCTGGCCACCTGG TGTTCAATCGATTGGTTATCAGATGAAACATCCGACA AGTCTTACGTCTTTGCCATCTTCATATTTTGCTTTCTTG TTCCTGTGTTGATAATTGTGGTTTCGTATGGTCTCATAT ATGACAAAGTCAGAAAGGTGGCTAAAACTGGTGGAA GTGTGGCCAAAGCTGAACGGGAAGTCCTCAGGATGA CTTTGCTGATGGTCAGTCTCTTCATGCTTGCCTGGTCT CCTTATGCTGTCATCTGCATGCTTGCATCTTTCGGACC GAAGGATCTACTTCACCCTGTTGCAACGGTGATTCCA GCTATGTTTGCCAAATCCTCAACGATGTACAACCCAT TAATTTACGTCTTCATGAATAAGCAGTTCCGAAGATC ATTAAAGGTACTTCTCGGAATGGGGGTCGAGGATTTA AATTCAGAATCGGAAAGAGCCACAGGGGGCACTGC AACCAACCAGGTGGCTGCTACCTGA	TACgCggggATTCAgCATTTgCAgTgTgACTgCAAAgCgAggAggATTAATT ggATggggAAAATCTCACAATTCCAAATCCAgTgACAgAATTAATggACACA TTACATCTACAACgCATTCACTgCCACAgACTACAACATCTgTgCAgCgTAC CCTTggTgATCTCTTAgTggCCgTCTTTggAACTgCTTTCAAgTTTATCATC AgCCAggAAAACTTTACTCAgAgAAgAAgATggCTTCTgCAAgTggTATggC CATCACCTACTTgggAggTCTTgCAgCCCTgATgACTCTTTCTgTAATTgCC TgTACgATgCTTAgCCgTCCTgAggCTTggAAgTTTCACgggCCTCACCACg AATgggAgTggCTgCgATggCCTTCATTTggATATACTCCCTTgCCTTTACT ggCACCACTTCTTggATggAATCATTACATACCTgAgggTCTggCCACCTgg TTCAATCgATTggTTATCAgATgAAACATCCgACAAgTCTTACgTCTTTgCC CATATATgACAAAgTCAgAAAggTggCTAAAACTggTggAAgTgTggCCAAA TgAACgggAAgTCCTCAggATgACTTTgCTgATggTCAgTCTCTTCATgCTT CTggTCTCCTTATgCTgTCATCTgCATgCTTgCATCTTTCggACCgAAggAT ACTTCACCCTgTTgCAACggTgATTCCAgCTATgTTTgCCAAATCCTCAACg gTACAACCCATTAATTTACgTCTTCATgAATAAgCAgTTCCgAAgATCATTA ggTACTTCTCggAATgggggTCgAggATTTAAATTCAgAATCggAAAgAgCC AgggggCACTgCAACCAACCAggTggCTgCTACCTgAAgAggAAAACATACA CAgATCTACAAAATTCCAAAAATCAATTggCTgATgTCACAACTCATTgAAA ACTTAgTTAAATTCgCCgAAAAgTATTTgTTgAAAgCATTTTCTAgAATgAg TTAggAggATTTCATTTTTAAgAAgATTATAgCATAAggAgATCATATCACT CATCAAAgACAgAAAgCTgAgATCCCCTAATAATTTCCTgTATgTTAgTCTT CTTCATATTTTgCTTTCTTgTTCCTgTgTTgATAATTgTggTTTCgTATggT AgCTAAgCATAgACAATTATTTTTAgAgACTggTTAgTTCATACAgggCTAA
XLOC_001549 | exc-7 | *elav*	AUGCAGAUUACAGGUGGUAGCCCUAACCAAGAACCAG CCAUGAUGGACACAAACAGCCAGACAUCCCAACAGAA UGGAGGCUCCAACCAUUCGCCUGUCUCUCCUGGGGCG GAGGACAGUAAAACCAACCUGAUCGUGAAUUACCUGC CCCAGACCAUGACCCAAGAAGAAAUCCGCUCGCUCUU CUCCAGCAUAGGAGAGGUGGAGAGCUGUAAGCUCAU CCGAGACAAAGCAACAGAUCAGUCUUCAGGAACCUCA AGUUGUCAGAGUUUGGGAUAUGGCUUUGUCAAUUAC AAGAGACCAGAAGACGCAGAGAAGGCCAUCAACACAC UGAAUGGUCUGCGCCUACAGAACAAAACCAUAAAGG UCUCCUAUGCCAGACCCAGCAGUGAAAGCAUUAAAGG AGCUAACCUCUAUCUAAGCGGACUACCCAAAUCCAUG AGUGAACCGGAACUGCGAUCUUUAUUUUCUUCAUGU GGAUCAAUUAUAAACUGCAGAAUUCUCUGUGACAAC ACAACUGGGUUGUCCAAAGGGGUAGGGUUCAUCCGG UUCGACCAGAGAGUGGAAGCAGAGCGCGCCAUCAAAC AGCUGAACGGCAAGGUGCCCGAAGGAGCCACCGAACC CAUCACAGUGAAAUUCGCCAACGCUCCCAGCUCCAAC AAGAACCAACUGCCCAUCACGGCCAUGGCCACAUAUC UGUCGCCAACCCGCCGAUUCCUCGGACCCAUCCACCA UCCCGCUGGACGCUUCAGUUCUGCUUCCAGGUUCUCG CCGUUGGAUGGAGGCCUGAUGCCCAACACCCUGCUGU CCGGGAACGCCCUCAACGGAGCUGGUUGGUGCAUCUU CGUCUACAACCUGGCUCCCGAGACCGAGGAGAACGUC CUCUGGCAGUUGUUCGGACCAUUCGGAGCCGUGCAGA GUGUCAAAGUCAUCCGCGACUUCCAGACACAAAAGUG CAAAGGCUUCGGUUUUGUUACCAUGACAAACUAUGA CGAGGCACUCAUGGCCAUUCAGAGUCUCAAUGGAUAC ACCCUUGGCAACCGAGUAUUACAAGUCUCUUUCAAGA CUAAUAACCGGAAAUCUUAG	AAATCCATgAgTgAACCggAACTgCgATCTTTATTTTCTTCATgTggATCAA TTCATCCggTTCgACCAgAgAgTggAAgCAgAgCgCgCCATCAAACAgCTgA ATgCAgATTACAggTggTAgCCCTAACCAAgAACCAgCCATgATggACACAA gCggAggACAgTAAAACCAACCTgATCgTgAATTACCTgCCCCAgACCATgA AACACACTgAATggTCTgCgCCTACAgAACAAAACCATAAAggTCTCCTATg AgACCCAgCAgTgAAAgCATTAAAggAgCTAACCTCTATCTAAgCggACTAC TACACCCTTggCAACCgAgTATTACAAgTCTCTTTCAAgACTAATAACCggA CTCATCCgAgACAAAgCAACAgATCAgTCTTCAggAACCTCAAgTTgTCAgA ATAAACTgCAgAATTCTCTgTgACAACACAACTgggTTgTCCAAAggggTAg gAgACCgAggAgAACgTCCTCTggCAgTTgTTCggACCATTCggAgCCgTgC ggCAAggTgCCCgAAggAgCCACCgAACCCATCACAgTgAAATTCgCCAACg CCCAgCTCCAACAAgAACCAACTgCCCATCACggCCATggCCACATATCTgT gCTTCCAggTTCTCgCCgTTggATggAggCCTgATgCCCAACACCCTgCTgT gggAACgCCCTCAACggAgCTggTTggTgCATCTTCgTCTACAACCTggCTC AgTgTCAAAgTCATCCgCgACTTCCAgACACAAAAgTgCAAAggCTTCggTT gTTACCATgACAAACTATgACgAggCACTCATggCCATTCAgAgTCTCAATg

		AgCCAgACATCCCAACAgAATggAggCTCCAACCATTCgCCTgTCTCTCCTg CCAACCCgCCgATTCCTCggACCCATCCACCATCCCgCTggACgCTTCAgTT TTgggATATggCTTTgTCAATTACAAgAgACCAgAAgACgCAgAgAAggCCA
XLOC_000129 | POU4F | *brn3*	GGCAACCTCTTCGGGGGCCTGGATGACAGCCTCCTTGCC CGGGCCGAGGCTTTGGCGGCAGCTGACATTGGCAAGAA CTCCTCGCCGCCGCTGGTCAAGCCGGACTCCCTGTATGG CCACCCTGAGATGGGCACCCAGCACCCGGGGAATCCCC GCTCCCAGATGCCCCTGCATGGCTTCTCGGCCCCTGGGG ACGCCATGCTGGACCAGGTGTCTCTTGGGATGTCGATGC CCTTGACCGCCAGCATGGACCACACGGGCATCACCTCAC ACGGGCACTCCCACCAAATGTACACTCCCGTGTACAGCC ACCCCTCTAACGGAATGCACAATGCCCACCACCCAGGCC TCGGCCCCCCCCACCACCCGCACGCGGTACACCCCCACG GCGACGCCGACTGTGACCCTCGGGAGTTGGAGGCTTTTG CCGAACACTTCAAGCAGCGACGCATCAAGCTAGGAGTG ACCCAAGCCGACGTCGGCTCAGCCCTAGGCAAGCTCAA ACTTCCAGGGGTGGGCTCTCTGAGCCAGAGCACGATCTG CAGGTTCGAATCACTGACTTTGAGCCATAACAACATGAT AGCCCTTAAACCGGTGCTGCAGGCCTGGCTCGAAGAGG CCGAGGCAGCTGCTAGGGAGAGCAAGGCTAACGCTGAT GCCCTCGGGGGAGCAGAGAAGAAACGCAAGCGCACGT CGATCGCAGCCCCCGAGAAACGATCGCTGGAAGCTTAC TTCGCCGTGCAGCCACGCCCCTCGGGCGAAAAAATCGCT CAAATCGCTGAAAAGTTGGACCTCAAGAAAAATGTAGT CAGAGTGTGGTTCTGCAACCAAAGACAGAAGCAGAAAA GGATGAAATTCGCCTCAATGAACCACCCTGGTCAGCACG GAGGCCCCCACCACTA	CTTgggATgTCgATgCCCTTgACCgCCAgCATggACCACACgggCATCACCT CACgggCACTCCCACCAAATgTACACTCCCgTgTACAgCCACCCCTCTAACg gCCCCCgAgAAACgATCgCTggAAgCTTACTTCgCCgTgCAgCCACgCCCCT CACTTCAAgCAgCgACgCATCAAgCTAggAgTgACCCAAgCCgACgTCggCT gCCCTAggCAAgCTCAAACTTCCAggggTgggCTCTCTgAgCCAgAgCACgA gTgCTgCAggCCTggCTCgAAgAggCCgAggCAgCTgCTAgggAgAgCAAgg CACCCCCACggCgACgCCgACTgTgACCCTCgggAgTTggAggCTTTTgCCg TgCAggTTCgAATCACTgACTTTgAgCCATAACAACATgATAgCCCTTAAAC AACgCTgATgCCCTCgggggAgCAgAgAAgAAACgCAAgCgCACgTCgATCg ggCgAAAAAATCgCTCAAATCgCTgAAAAgTTggACCTCAAgAAAAATgTAg AgAgTgTggTTCTgCAACCAAAgACAgAAgCAgAAAAggATgAAATTCgCCT CTgTATggCCACCCTgAgATgggCACCCAgCACCCggggAATCCCCgCTCCC ggCAACCTCTTCgggggCCTggATgACAgCCTCCTTgCCCgggCCgAggCTT gCggCAgCTgACATTggCAAgAACTCCTCgCCgCCgCTggTCAAgCCggACT

Table S2. Genes discussed in Chapter 2 and sources if available. Gene Name column is in the form XLOC | emapper annotation | *known Pdum name if available.*

Gene Name	**Source / Accession number**
XLOC_044659 | ARX | *arx*	Tomer et al. (2010) GU169412.1
XLOC_003959 | PTF1A | *ptf1*	Vergara et al. (2021)
XLOC_040379 | NCALD | *NA*	-
XLOC_037643 | Pdfr-1 | *pdfr*	Häfker et al. (2024) OL606759.1
XLOC_026597 | unc-17 | *vAChT/ChAT*	Denes et al. (2007) EF384218.1
XLOC_008985 | SLC17A6 | vglut	Tomer et al. (2010) GU169424.1
XLOC_069321 | EMX1 | *emx*	Tomer et al. (2010) GU169418.1
XLOC_069317 | EMX2 | *NA*	-
XLOC_000048 | MEF2C | *NA*	-
XLOC_000375 | HCN3 | *hcn1*	-
XLOC_012133 | - | *cOpsin1*	Arendt et al. (2004) AY692353.1
XLOC_066186 | Adenylyl/guanylyl cyclase catalytic domain | *NIT-GC 2*	Jokura et al. (2023)
XLOC_045820 | PDE9A | *pde9*	Achim et al. (2018)
XLOC_018600 | CNGA3 | *cnga*	Tosches et al. (2014) KM199644.1
XLOC_064003 | - | *pedal-peptide2*	Conzelmann et al. (2013) KF515945.1
XLOC_002678 | tll | *tailless*	Tomer et al. (2010) GU169423.1
XLOC_000129 | POU4F | *brn3*	Backfisch et al. (2013) KC109636.1

XLOC_062588 | MXI1 | *NA*	-
XLOC_037851 | crebzf | *NA*	-
XLOC_040496 | CEBPB | *NA*	-
XLOC_008123 | SOX4 | *soxC*	Kerner et al. (2009) FN357282.1
XLOC_064106 | SOX2 | *soxB1*	ANS60443.1
XLOC_054004 | SOX21 | *soxB*	Kerner et al. (2009) FN357280.1 Stockinger et al. (2024)
XLOC_021539 | PROX1| *prox*	Kerner et al. (2009) FN357281.1
XLOC_056655 | - | *neurofilament (gfap* ortholog*)*	Milivojev et al. (2024)
XLOC_004327 | PARD3 | *par3a*	Nakama et al. (2017)
XLOC_027608 | MICU1| *pcna_a*	HF935038.1
XLOC_025800 | PCNA | *pcna_b*	-
XLOC_052210 | MKI67 | *ki67*	Milivojev et al. (2024)
XLOC_060478 | BTG1 | *btg1*	Milivojev et al. (2024)
XLOC_063920 | BTG4 | *btg2*	Milivojev et al. (2024)
XLOC_002034 | foxo | *foxo*	Milivojev et al. (2024)
XLOC_001549 | exc-7 | *elav*	Denes et al. (2007) EF384209.1
XLOC_002600 | CELF2 | elav2	Milivojev et al. (2024)
XLOC_000875 | SYT2 | syt	Tessmar-Raible et al. (2007) EF544397.1

Table S3. Reference CO1 sequences relevant for Chapter 4.

Name and source	Sequence
Lab_Pdu (Arendt Lab culture)	CTTGAGTCGGGCCTCCTAGGAACCTCTATAAGCCTCTTAATCCGGGCTGAACTAGGTCAA CCCGGATCGCTACTCGGGAGAGACCAACTATATAATACTATTGTTACAGCCCACGCATTC CTAATAATTTTTTTCTTAGTTATACCCGTAATAATCGGAGGGTTTGGCAATTGATTGGTG CCTTTAATATTAGGGGCCCCAGATATAGCATTCCCCCGATTAAATAACATAAGCTTCTGA CTTCTTCCCCCCTCTCTGACTCTTCTTCTTTCTAGGGCAGCAGTAGAAAAAGGAGTGGGT ACCGGCTGAACAGTCTATCCTCCTTTATCCAGTAATATTGCTCATGCTGGCCCTTCAGTA GACCTGGCAATCTTTTCTCTTCACCTAGCGGGGGTGTCCTCTATTATAGGGGCCTTAAAT TTTATTACCACAGTTATCAATATACGCTCAAAAGGACTAAAACTAGAACGTGTCCCTTTA TTTGTATGATCTGTAGTAATTACAGCGGTTCTTCTACTATTAAGGCTTCCAGTGTTAGCG GGTGCTATCACAATATTATTAACAGACCGAAACCTAAACACTGCGTTCTTTGATCCTGCT GGAGGGGGAGACCCAATCCTATACCAGCACTTGTTTTGGTTTTTTGGACCCTGGAAAATT
Pdu_partialCOI KP127954.1 (Lucey et al., 2015)	GCCTCCTAGGAACCTCTATAAGCCTCTTAATCCGGGCTGAACTAGGTCAACCCGGATCGC TACTCGGGAGAGACCAGCTATATAATACTATTGTTACAGCCCACGCATTCCTAATAATTT TTTTCTTAGTTATACCCGTAATAATCGGAGGGTTTGGCAATTGATTGGTGCCTTTAATAT TAGGGGCCCCAGATATAGCATTCCCCCGATTAAATAACATAAGCTTCTGACTTCTTCCCC CCTCTCTGACTCTTCTTCTTTCTAGGGCAGCAGTAGAAAAAGGAGTGGGTACCGGCTGAA CAGTCTATCCTCCTTTATCCAGTAATATTGCTCATGCTGGCCCCTCAGTAGACCTGGCAA TCTTTTCTCTTCACCTAGCGGGGGTGTCCTCTATTATAGGGGCCTTAAATTTTATTACCA CAGTTATCAATATACGCTCAAAAGGACTAAAACTAGAACGTGTCCCTTTATTTGTATGAT CTGTAGTAATTACAGCGGTTCTTCTACTATTAAGGCTTCCAGTGTTAGCGGGTGCTATCA CAATATTATTAACAGACCGAAACCTAAACACTGCGTTCTTTG
Pmass_Ischia_N3_01 KT124680.1 (Wäge et al., 2017)	CCGGATCACTACTGGGAAGGGATCAACTATATAATACAATCGTTACAGCCCATGCCTTCT TAATAATCTTTTTTCTCGTTATACCTGTAATAATTGGAGGATTTGGTAATTGACTAGTAC CTCTAATATTAGGGGCACCGGACATAGCCTTTCCTCGATTAAATAATATAAGGTTTTGAT TACTCCCCCCATCATTAACCCTTCTCTTATCTAGGGCCGCAGTAGAAAAGGGAGTCGGCA CTGGTTGAACAGTATACCCCCCATTAGCAAGTAATATTGCACACGCTGGCCCCTCAGTAG ATTTAGCTATTTTCTCCCTCCATCTAGCAGGGGTCTCCTCTATTATAGGAGCCCTAAATT TTATCACAACAGTCATCAATATACGATCAAAAGGACTAAAATTAGAACGCGTTCCTTTAT TTGTATGATCTGTAGTAATTACAGCAGTACTTCTTCTTTTAAGATTGCCAGTACTAGCAG GAGCTATTACTATGTTATTAACAGACCGCAACCTAAATACTGCGTTTTTTGATCCTGCAG GAGGAGGGGACCCCATTTTATACCAACA
Pmass_partialCOI KP127953.1 (Lucey et al., 2015)	GGCTATTAGGGACGTCTATGAGTCTCCTAATTCGGGCCGAATTAGGACAGCCCGGATCAC TACTGGGAAGGGATCAACTATATAATACAATCGTTACAGCCCATGCCTTCTTAATAATCT TTTTTCTCGTTATACCTGTAATAATTGGAGGATTTGGTAATTGACTAGTACCTCTAATAT TAGGGGCACCGGACATAGCCTTTCCTCGATTAAATAATATAAGGTTTTGATTACTCCCCC CATCATTAACCCTTCTCTTATCTAGGGCCGCAGTAGAAAAGGGAGTCGGCACTGGTTGAA CAGTATACCCCCCATTAGCAAGTAATATTGCACACGCTGGCCCCTCAGTAGATTTAGCTA TTTTCTCCCTCCATCTAGCAGGGGTCTCCTCTATTATAGGAGCCCTAAATTTTATCACAA CAGTCATCAATATACGATCAAAAGGACTAAAATTAGAACGCGTTCCTTTATTTGTATGAT CTGTAGTAATTACAGCAGTACTTCTTCTTTTAAGATTGCCAGTACTAGCAGGAGCTATTA CTATGTTATTAACAGACCGCAACCTAAATACTGCGTTTTTTG
Pagilis_partial CO1 OP347357.1 (Teixeira et al., 2022)	CACCTTATATTTTATCTTTGGCACCTGATCCGGACTACTAGGGACGTCTATAAGTCTCCTGATTCGGGCC GAATTAGGACAGCCCGGCTCACTACTGGGAAGGGATCAACTATATAATACAATCGTTACAGCCCACGCCT TCCTAATAATCTTTTTTCTCGTTATACCCGTAATAATTGGCGGATTTGGTAATTGACTAGTACCTTTAAT ATTAGGGGCACCGGACATAGCTTTTCCTCGATTAAATAATATAAGGTTTTGATTACTCCCCCCATCATTA ACCCTTCTCTTATCTAGGGCCGCAGTAGAAAAGGGAGTCGGCACTGGTTGAACAGTGTATCCCCCATTAG CAAGTAATATTGCACACGCTGGCCCCTCAGTAGATTTAGCTATTTTCTCCCTCCATCTAGCAGGGGTCTC ATCTATTATAGGAGCCTTAAATTTTATCACAACAGTCATTAATATACGATCAAAAGGACTAAAATTAGAA CGCGTTCCTTTATTTGTATGATCTGTAGTAATCACAGCAGTACTTCTTCTGTTAAGATTGCCAGTACTGG CAGGAGCTATTACTATATTACTAACAGACCGCAACCTAAATACTGCATTTTTTGACCCTGCAGGAGGAGG GGACCCCATTTTATACCAACATCTATTT

Table S4. Primer sequences used for CO1 barcoding (Chapter 4).

PrimerID	Primer Name	Sequence
LP403	PolyLCO (fwd)	GAYTATWTTCAACAAATCATAAAGATATTGG
LP404	PolyHCO (rev)	TAMACTTCWGGGTGACCAAARAATCA
LP448	mlCOIintF	GGWACWGGWTGAACWGTWTAYCCYCC
LP449	mlCOIintR	GGRGGRTASACSGTTCASCCSGTSCC
LP438	PolyLCO_1	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GAYTATWTTCAACAAATCATAAAGATATTGG
LP439	PolyLCO_N	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGN**GAYTATWTTCAACAAATCATAAAGATATTGG
LP440	PolyLCO_NN	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNN**GAYTATWTTCAACAAATCATAAAGATATTGG
LP441	PolyLCO_NNN	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNN**GAYTATWTTCAACAAATCATAAAGATATTGG
LP442	PolyHCO_1	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**TAMACTTCWGGGTGACCAAARAATCA
LP443	PolyHCO_N	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGN**TAMACTTCWGGGTGACCAAARAATCA
LP444	PolyHCO_NN	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNN**TAMACTTCWGGGTGACCAAARAATCA
LP445	PolyHCO_NNN	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNN**TAMACTTCWGGGTGACCAAARAATCA
LP458	mlCOIintF_1	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GGWACWGGWTGAACWGTWTAYCCYCC
LP459	mlCOIintF_N	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGN**GGWACWGGWTGAACWGTWTAYCCYCC
LP460	mlCOIintF_NN	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNN**GGWACWGGWTGAACWGTWTAYCCYCC
LP461	mlCOIintF_NNN	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNN**GGWACWGGWTGAACWGTWTAYCCYCC
LP462	mlCOIintR_1	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GGRGGRTASACSGTTCASCCSGTSCC
LP463	mlCOIintR_N	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGN**GGRGGRTASACSGTTCASCCSGTSCC
LP464	mlCOIintR_NN	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNN**GGRGGRTASACSGTTCASCCSGTSCC
LP465	mlCOIintR_NNN	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNN**GGRGGRTASACSGTTCASCCSGTSCC

Table of Buffer compositions

Name	**Composition**
10x pPBS	1.198 M NaCl 0.351 M $Na_2HPO_4 \cdot 2H_2O$ 0.025 M KH_2PO_4
PTW	1x pPBS 0.1% Tween-20
Homogenization buffer (2.5.2. Nuclei extraction)	250 mM sucrose 25 mM KCl 5 mM $MgCl_2$ 10 mM Tris-HCl pH 8.0 0.1% IGEPAL 1 µM DTT 0.4 U/µl Murine RNase Inhibitor (NEB) 0.2 U/µl Superase-In (ThermoFisher)
5x SSCT (2.5.5.1 WM-HCR)	5x sodium chloride sodium citrate (SSC) 0.1% Tween 20
Hybridization buffer (Molecular Instruments) (2.5.5.1 WM-HCR)	50% formamide 5× sodium chloride sodium citrate (SSC) 9 mM citric acid (pH 6.0) 0.1% Tween 20 50 µg/mL heparin 1× Denhardt's solution 10% dextran sulfate
Probe Wash Buffer (Molecular Instruments) (2.5.5.1 WM-HCR))	50% formamide 5× sodium chloride sodium citrate (SSC) 9 mM citric acid (pH 6.0) 0.1% Tween 20 50 µg/mL heparin
Amplification buffer (Molecular Instruments) (2.5.5.1 WM-HCR)	5× sodium chloride sodium citrate (SSC) 0.1% Tween 20 10% dextran sulfate
Clearing *Solution 1* (2.5.5.1 WM - HCR)	10% Tris(hydroxyethyl)ethylenediamine (THEED) (Sigma) 5% Triton (Roth) 25% Urea in dH2O (Sigma)

RMarkdown files

Code uploaded to: https://git.embl.org/grp-arendt/thesis_vw#